

Annals of Computer Science and Information Systems
Volume 14

**Proceedings of the First International
Conference on Information Technology and
Knowledge Management**

December 22–23, 2017. New Delhi, India



**Ajay Jaiswal, Vijender Kumar Solanki,
Zhongyu (Joan) Lu, Nikhil Rajput (eds.)**



Annals of Computer Science and Information Systems, Volume 14

Series editors:

Maria Ganzha,

Systems Research Institute Polish Academy of Sciences and Warsaw University of Technology, Poland

Leszek Maciaszek,

Wroclaw University of Economy, Poland and Macquarie University, Australia

Marcin Paprzycki,

Systems Research Institute Polish Academy of Sciences and Management Academy, Poland

Senior Editorial Board:

Wil van der Aalst,

Department of Mathematics & Computer Science, Technische Universiteit Eindhoven (TU/e), Eindhoven, Netherlands

Frederik Ahlemann,

University of Duisburg-Essen, Germany

Marco Aiello,

Faculty of Mathematics and Natural Sciences, Distributed Systems, University of Groningen, Groningen, Netherlands

Mohammed Atiquzzaman,

School of Computer Science, University of Oklahoma, Norman, USA

Barrett Bryant,

Department of Computer Science and Engineering, University of North Texas, Denton, USA

Ana Fred,

Department of Electrical and Computer Engineering, Instituto Superior Técnico (IST—Technical University of Lisbon), Lisbon, Portugal

Janusz Górski,

Department of Software Engineering, Gdansk University of Technology, Gdansk, Poland

Mike Hinchey,

Lero—the Irish Software Engineering Research Centre, University of Limerick, Ireland

Janusz Kacprzyk,

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Irwin King,

The Chinese University of Hong Kong, Hong Kong

Juliusz L. Kulikowski,

Natęcz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland

Michael Luck,

Department of Informatics, King's College London, London, United Kingdom

Jan Madey,

Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw, Poland

Stan Matwin,

Dalhousie University, University of Ottawa, Canada and Institute of Computer Science, Polish Academy of Science, Poland

Michael Segal,

Ben-Gurion University of the Negev, Israel

Andrzej Skowron,

Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw, Poland

John F. Sowa,

VivoMind Research, LLC, USA

Editorial Associate:

Katarzyna Wasielewska,

Systems Research Institute Polish Academy of Sciences, Poland

Paweł Sitek,

Kielce University of Technology, Kielce, Poland

T_EXnical editor: Aleksander Denisiuk,

University of Warmia and Mazury in Olsztyn, Poland

Proceedings of the First International Conference on Information Technology and Knowledge Management

Ajay Jaiswal, Vijender Kumar Solanki,
Zhongyu (Joan) Lu, Nikhil Rajput (eds.)



POLSKIE TOWARZYSTWO INFORMATYCZNE
POLISH INFORMATION PROCESSING SOCIETY

Annals of Computer Science and Information Systems, Volume 14
Proceedings of the First International Conference on Information
Technology and Knowledge Management

ART: ISBN 978-83-949419-0-1

USB: ISBN 978-83-949419-1-8

WEB: ISBN 978-83-949419-2-5

ISSN 2300-5963

DOI 10.15439/978-83-949419-2-5

© 2018, Polskie Towarzystwo Informatyczne

Ul. Solec 38/103

00-394 Warsaw

Poland

Contact: secretariat@fedcsis.org

<http://annals-csis.org/>

Cover photo:

Miłosz Kulawiak,

Elbląg, Poland

Also in this series:

Volume 13: Communication Papers of the 2017 Federated Conference on Computer
Science and Information Systems, **ISBN WEB: 978-83-922646-2-0, ISBN USB: 978-83-922646-3-7**

Volume 12: Position Papers of the 2017 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-922646-0-6, ISBN USB: 978-83-922646-1-3**

Volume 11: Proceedings of the 2017 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-946253-7-5, ISBN USB: 978-83-946253-8-2,**

ISBN ART: 978-83-946253-9-9

Volume 10: Proceedings of the Second International Conference on Research in
Intelligent and Computing in Engineering, **ISBN WEB: 978-83-65750-05-1,**

ISBN USB: 978-83-65750-06-8

Volume 9: Position Papers of the 2016 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-60810-93-4, ISBN USB: 978-83-60810-94-1**

Volume 8: Proceedings of the 2016 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-60810-90-3, ISBN USB: 978-83-60810-91-0**

Volume 7: Proceedings of the LQMR Workshop, **ISBN WEB: 978-83-60810-78-1,**

ISBN USB: 978-83-60810-79-8

Volume 6: Position Papers of the 2015 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-60810-76-7, ISBN USB: 978-83-60810-77-4**

Volume 5: Proceedings of the 2015 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-60810-66-8, ISBN USB: 978-83-60810-67-5**

Volume 4: Proceedings of the E2LP Workshop, **ISBN WEB: 978-83-60810-64-4,**

ISBN USB: 978-83-60810-63-7

Volume 3: Position Papers of the 2014 Federated Conference on Computer Science and
Information Systems, **ISBN WEB: 978-83-60810-60-6, ISBN USB: 978-83-60810-59-0**

Volume 2: Proceedings of the 2014 Federated Conference on Computer Science and
Information Systems, **WEB: ISBN 978-83-60810-58-3, USB: ISBN 978-83-60810-57-6,**

ART: ISBN 978-83-60810-61-3

DEAR Reader, it is our pleasure to present to you the proceedings of the first International Conference on Information Technology and Knowledge Management (ICITKM-17) held at Shaheed Sukhdev College of Business Studies, University of Delhi, New Delhi, India on 22-23 December 2017. The conference provided the international forum to the researchers and practitioners from both academia as well as industry to meet and share the cutting-edge development in the field of technology and management.

The ICITKM-17 was organized by Shaheed Sukhdev College of Business Studies (SSCBS) and chaired by Dr. Poonam Verma, Principal SSCBS, while Dr. Ajay Jaiswal and Dr. Vijender Kumar Solanki, Associate Professor, Department of Computer Science & Engineering, CMR Institute of Technology (Autonomous), Hyderabad, India acted as the Chair of the organizing committee and program chair respectively.

Chief Guest Prof. Dr. Yogesh Singh, Vice Chancellor, Delhi Technology University, Delhi inaugurated ICITKM-17 by lighting the lamp of knowledge and wisdom along with Keynote Speakers Prof. R. K. Agrawal, SC&SS, Jawahar Lal Nehru University, Delhi, Mr. Arif Khan, Vice President & Head of System Development, VoerEir AB, Dr. Poonam Verma and Dr. Ajay Jaiswal.

The conference invited the research papers on the various research topics in the fields of Information Technology and Management such as Machine learning, Computer Networks, Data Mining and Databases, Artificial Intelligence & Cryptography, Pattern Recognition, Image Processing, Knowledge Management, Marketing Management, Operations Management, Organizational Behavior, Human Resource Management, Financial Management etc. The ICITKM-2017 also gave the chance to authors to submit their extended version paper in reputed journals indexed in Web of Science, Scopus, ESCIE & DBLP.

The conference started with inaugural speech by Prof. Yogesh Singh in which he stressed on collaborative research and need of entrepreneurial skills in youth. Next session had keynote speech by Prof. R. K. Agrawal, giving insight to Big Data, Machine Learning and Deep learning. Day 1 of the conference also had two technical sessions dedicated for paper presentations. First session was chaired by Dr. Rashmi Aggarwal, Professor and Head, Department of Computer Application, Manav Rachna International University and Dr. Vijender Kumar Solanki. Second Session was chaired by Dr. Shivani Aggarwal, KIET, Ghaziabad, UP, and Dr. Abhishek Pandey, Assistant professor, ACERC Ajmer.

Day 2 of the conference started with technical session chaired by Mr. Vinod Nangia, Director of New Business Development, Honeywell International, Phoenix, AZ, USA. It was followed by Keynote session by Prof. Muneesh Kumar, Head, Department of Financial Studies, University of Delhi titled "DNA of Emerging Enterprise". Next session was the panel discussion on "How to publish in high impact factor journals?" Following panelist shared their views on the topic.

Prof. Muneesh Kumar, Head, Department of Financial Studies, University of Delhi

Prof. Sanjay Sehgal, Dept. of Financial Studies, Univ. of Delhi

Prof. N.R. Bhanumurthy, National Institute of Public Finance and Policy (NIPFP), Govt. of India

Prof. S S Lodha – Southern Connecticut State University, New Haven, CT

This session gave researchers very useful tips for publishing their research papers in high impact factor journals. It was followed by keynote session by Prof. S. S. Lodha on educational leadership.

In valedictory function Dr. Poonam Verma gave away the certificates to participants. She praised the efforts put up by the organizing team consisting of following members from the college:

Dr. Ajay Jaiswal, Dr. Sameer Anand, Dr. Kumar Bijoy, Dr. Rishi Rajan Sahay, Dr. Abhishek Tandon, Mr. Abhimanyu Verma, Mr. Onkar Singh, Dr. Gurjeet Kaur, Mr. Vipin Rathi

The day ended with thanks to the faculty, Staff and students for their magnanimous support in organizing the conference.

The more detail of ICITKM-17 could be seen at Facebook pages & conference portal:

Official Web link: <http://www.fpr.org.in>

Facebook Link: <http://www.facebook.com/icitkm>

The two day conferences ended with the joy, learning, interaction and many more memories which could not be possible to write in text.

Editors

Ajay Jaiswal, Shaheed Sukhdev College of Business Studies (SSCBS), University of Delhi, New Delhi, India

Vijender Kumar Solanki, CMR Institute of Technology (Autonomous), Hyderabad, TS, India

Zhongyu (Joan) Lu, University of Huddersfield, UK

Nikhil Rajput, Ramanujan college, University of Delhi, New Delhi, India

Proceedings of the First International Conference on Information Technology and Knowledge Management

December 22–23, 2017. New Delhi, India

TABLE OF CONTENTS

FIRST INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND KNOWLEDGE MANAGEMENT

Call For Papers	1
Architecture-based Optimal Software Reliability Allocation under uncertain preferences	3
<i>Anu G. Aggarwal, Vibha Verma, Sameer Anand</i>	
An Innovative B2C E-commerce Websites Selection using the ME-OWA and Fuzzy AHP	13
<i>Anu G. Aggarwal, Akash Sharma</i>	
An Intuitionistic Approach for Ranking OTA Websites under Multi Criteria Group Decision Making Framework	21
<i>Anu G. Aggarwal, Himanshu Sharma, Abhishek Tandon</i>	
Clustered Comparative Analysis of Security Sensor Discrimination Data	29
<i>Neeraj Bhargava, Aakanksha Jain, Abhishek Kumar, Pramod Singh Rathore, Anuja Bansal</i>	
Detection of Malicious Executables Using Rule Based Classification Algorithms	35
<i>Neeraj Bhargava, Aakanksha Jain, Abhishek Kumar, Dac-Nhuong Le</i>	
Financial Inclusion in India and PMJDY: A Critical Review	39
<i>Kumar Bijoy</i>	
Vehicular Ad-Hoc Network for Smart Cities	47
<i>Bhawna Chaudhary, Sheetal Singh</i>	
Reliability Modeling of OSS Systems based on Innovation-Diffusion Theory and Imperfect Debugging	53
<i>Neha Gandhi, Neha Gondwal, Abhishek Tandon</i>	
Usability Of An E-Commerce Website Using Information Mining and Artificial Intelligence	59
<i>Pankaj Gupta, Bharat Bhushan Sagar</i>	
E-Assessment Tools for Programming Languages: A Review	65
<i>Sugandha Gupta, Anamika Gupta</i>	
Importance of Text Data Preprocessing & Implementation in RapidMiner	71
<i>Vaishali Kalra, Rashmi Aggarwal</i>	
A study of Optimal Testing Resource Allocation Problem for Modular Software with Change Point	77
<i>Gurjeet Kaur, Anu G. Aggarwal, Aman Kedia</i>	

An Adaptive Approach Digital Image Watermarking based on Discrete Wavelet Transform (DWT)	85
<i>Pramod Kumar, Prakarti Triwedi, Surbhi Sharma</i>	
An Adaptive Approach for image adaptive watermarking using Elliptical curve cryptography (ECC)	89
<i>Naveen Kumar, Prakarti Triwedi, Pramod Singh Rathore</i>	
Design of Intelligent PD Controller for Water Supply in Healthcare Systems	93
<i>Mohan Debarchan Mohanty, Mihir Narayan Mohanty</i>	
Detection of Arrhythmia using Neural Network	97
<i>Saumendra Kumar Mohapatra, Hemanta Kumar Palo, Mihir Narayan Mohanty</i>	
An Efficient Load Balancing Algorithms in Stream Processing With the Cloud Computing Environment	101
<i>Jyoti Patharia, Rakesh Rathi</i>	
The Use of Deep Learning in Speech Enhancement	107
<i>Rashmirekha Ram, Mihir Narayan Mohanty</i>	
A Survey on Advanced Approaches of EHR in inter-related data using Machine Learning	113
<i>T. V. M. Sairam, R. Rajalakshmi</i>	
Result Analysis Suite: A Completely Automated Result Analysis Solution	121
<i>Pankaj Sambyal, Anamika Rustagi, Sonia Rani, Chinab Bhudhiraja, Bhawna Sharma</i>	
An Approach towards economical hierarchic Search over Encrypted Cloud	125
<i>Sreenivas Sasubilli, Kumar Attangudi Perichiappan Perichappan, P. Srinivas Kumar, Abhishek Kumar</i>	
A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation	131
<i>Gopinadh Sasubilli, Uday Shankar Sekhar, Ms. Surbhi Sharma, Ms. Swati Sharma</i>	
A Detailed Study of EEG based Brain Computer Interface	137
<i>Geeta Sharma, Neha Sharma, Tanya Singh, Rashmi Agrawal</i>	
A comparative analysis of promethee, ahp and topsis aiding in financial analysis of firm performance	145
<i>Aditi Sharma, Gurjeet Kaur, Jatin Bansal</i>	
A Perspective Approach (OABC) Algorithm using Square Odd Routing for minimized Energy Consumption	151
<i>Akrati Sharma, V. P. Sharma, Hitesh Sharma</i>	
Importance of Search Engine Marketing in the Digital World	155
<i>Arokia R. Terrance, Shruti Shrivastava, Asmita Kumari</i>	
Application of ASIP in Embedded Design with Optimized Clock Management	159
<i>Mood Venkanna, Rameshwar Rao, P. Chandra Sekhar</i>	
Authors index	165

First International Conference on Information Technology and Knowledge Management 2017

SHAHEED Sukhdev College of Business Studies, University of Delhi in association with Ramanujan College, University of Delhi and Foundation for Policy Research Group proudly invites academicians, researchers and industry professionals to participate in the International Conference on “Information Technology and Knowledge Management”, which is scheduled on December 22 and 23 2017. The conference will be consisting of Academician, research lab and industry experts talks. It will provide a networking opportunity for association, collaboration and future research for participants .

All the accepted, registered and presented paper will be published as a volume of Annals of Computer Science and Information Systems (ISSN- 2300-5963). The Volume will be having an ISBN number and each paper will be assigned DOI number. The conference papers will be indexed in Cross Ref, BazEkon, Open Access Library, Academic Keys, Journal Click, PBN, ARIANTE. The conference papers will also be submitted for possible indexing to Web of Science, Index Copernicus, SCOPUS and DBLP.

The extended version of accepted paper will be invited to submit in reputed journals as listed in publication.

TOPICS OF INTEREST

Topics of interest for submission include, but are not limited to:

- Knowledge Management
- Data Mining & Warehousing
- Decision Support System
- Machine Learning
- Network Security
- Image Processing
- Smart Cities
- Robotics
- Internet of Things
- Neural Network
- Multimedia
- Big Data Analytic
- Knowledge Management
- Marketing Management
- Operations Management
- Organizational Behavior
- Human Resource Management
- Financial Management
- Statistical Applications
- Policy Framework
- Public Administration
- Management Information System

TECHNICAL PROGRAM COMMITTEE

- Dr. Vijay Bhasker Semwal, IIIT-Dharwad, Karnatka, India
- Dr. Rashmi Agarwal , M R International University, Faridabad, Haryana
- Dr. Gloria Jeanette Rincon Aponte, University Cooperativa de Colombia, Colombia
- Prof. Vishwanath Bijalwan , Institute of Technology, Gopeshwer, U.K
- Dr. Nguyen Cuong, Quang Nam University, Vietnam
- Dr. S Bhattacharya, Institute of Technology & Science, Ghaziabad, U.P
- Dr. Eddie Custovic, La Trobe University, Australia
- Dr. Rishi Rajan Sahay, Shaheed Sukhdev College of Business Studies, University of Delhi
- Dr. Abhishek Tandon, Shaheed Sukhdev College of Business Studies, University of Delhi
- Dr. Manik Sharma, DAV University, Jhalandar, Punjab, India
- Dr. Niranjana Murthy, M S Ramaiah Institute of Technology, Bangalore
- Dr. Mihir Mohanty, Siksha O Anushandhan University, Oddisha
- Dr. K. Saravanan, Anna University, Tirunelveli Campus, Tamilnadu
- Dr. Jagat Narayan Giri, Mangalmay Institution, Greater Noida
- Dr. Sushil Dixit, LBSIM, New Delhi
- Dr. Sharvari Tamane, MGM College of Engineering, Aurangabad, MS
- Dr. Akhilesh Sharma, Manipal University, Jaipur, Rajasthan
- Dr. Anish Gupta, Amity University, Greater Noida, UP
- Dr. Vijay Pithadia, Degree College, Gujrat
- Dr. Ajay Kumar Sinha, SIBM, Hyderabad
- Dr. Astha Pareek, The IIS University, Jaipur, Rajasthan
- Dr. Swati Shinde, PCCOE, Pune
- Dr. Nimit Gupta, Fortune Institute of International Business, Delhi
- Prof. Shaik Mahaboob Hussain, Vishnu Institute of Technology, AP
- Anoop.V.S, IIITM-Kerala

PROGRAM CHAIR

- Dr. Vijender Kumar Solanki, CMR Institute of Technology, Hyderabad, India
- Dr. Nikhil Kumar Rajput, Ramanujan College, University of Delhi, Delhi

GENERAL CHAIR

- Dr. Poonam Verma, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
- Dr. S.P.Agarwal, Ramanujan College, University of Delhi, Delhi

PUBLICITY CHAIR

- Dr. Shivani Agarwal, KIET, Ghaziabad, UP, India
- Prof. Abhishek Kumar, ACERC, Ajmer, Rajasthan, India

ORGANIZING CHAIR

- Dr. Ajay Jaiswal, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
- Dr. Sameer Anand, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
- Dr. Kumar Bijoy, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
- Mr. Onkar Singh, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
- Mr. Vipin Kumar Rathi, Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi

Architecture-based Optimal Software Reliability Allocation under uncertain preferences

Anu G. Aggarwal¹, Vibha Verma², Sameer Anand³

¹Department of Operational Research, University of Delhi, Delhi, India

²Department of Operational Research, University of Delhi, Delhi, India

³Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi, India

¹anuagg17@gmail.com, ²vibhaverma.du.aor@gmail.com, ³sananddu@gmail.com

Abstract—Reliability Allocation in an essential task of the software development process. Increasing complexities in software structure and demand for bug free software has made Reliability Allocation a mandatory task during design and planning phase. So far in the literature several methods and models have been discussed for achieving the reliability target based on user's and developer's point of view. The crucial question that arises is 'How to allocate reliability for a Software system in an uncertain atmosphere where developer's preferences are subjective in nature?'. In this paper, we have proposed the software reliability allocation problem incorporating the decision maker's subjective uncertain preferences using Ordered Weighted Averaging (OWA) approach based on Fuzzy Analytical Hierarchical Process (FAHP). Parameter determination using FAHP through architectural hierarchy of the software system helps in interaction of user's assessment with the software engineers and programmers outlook. The OWA technique ensures complete use of available information and also avoids any kind of biasedness in reliability allocation due to overestimation of developer's inclinations. The proposed MEMV-OWA (Maximum Entropy Minimum Variance) operator is a bi-objective mathematical programming problem that maximizes entropy (deployment of information) along with minimization of the variance in weighting vector in an uncertain environment. Reliability allocation procedure for software system using the anticipated process has been discussed in detail. Also precise demonstration of the procedure has been done with an application example.

Index Terms—Software Reliability Allocation, Architecture-based hierarchy, MEMV-OWA, FAHP, Uncertain preferences

I. INTRODUCTION

THE INCREASING importance of the software systems in control and management of critical functions of various machines in daily life has steered the demand for quality software both at individual and the organizational level. Quality refers to a reliable software i.e. no bug reported by the user during the execution period under stated conditions. Buggy software can lead to loss of time, money and human lives also. The objective of error free software can be attained by maximizing the reliability therefore it is of greater importance for management to perform precise reliability allocation during the planning and design phase of the software development process. Effective reliability allocation is necessary since achieving the target reliability for a software

system involves a lot of cost. It can be defined as method of allocating the pre-defined reliability goal among the subsystems.

The concept of reliability allocation was initially used for hardware systems. Later it was adopted as an important aspect in software reliability as well. Most of the researchers in the past decades have used the architecture-based models for achievement of target reliability of a software system subject to constraints of budget, delivery time etc. Kubat [11] gave one of the initial models for software reliability allocation; he proposed a stochastic model for minimization of cost subject to the reliability constraint. Halendar et. al [9] gave RCCM (Reliability Constrained Cost Minimization) Model for the reliability allocation. Rani and Misra [14] proposed economic model that minimized cost (development and failures cost) for the optimal allocation of reliability. In these papers the reliability was allocated based on one or two criteria like minimization of cost or maximization of reliability. Later on number of researchers proposed multi-attribute based allocation methodologies by taking into consideration key software characteristics for e.g. functionality, criticality, availability, complexity, reliability etc. Zahedi and Ashrafi [20] proposed analytical hierarchical process (AHP) based software reliability allocation with objective of maximizing the user's utility based on cost and price. Leung [18] proposed the operational profile based reliability problem by using AHP and software utility function to form the software utility function defined in terms of reliability measures. Chatterjee et al. [4] established a system hierarchy combining users and developers view and adopted Fuzzy Analytical Hierarchical process (FAHP) to determine the weights required for reliability allocation. Yue et al. [18] used Dempster-Shafer theory for reliability allocation among the Multi-software in multimedia systems subject to the budget constraints.

AHP is a multi-criteria decision making framework to tackle decision maker's preferences for the alternatives. Zadeh [19] extended this into fuzzy environment to take care of uncertainty and vagueness involved in real life application.

Hence due to possible subjective preferences it is better to collect the expert's opinion in terms of Fuzzy numbers. FAHP helps to overcome the non-clarity in ratings provided by experts by combining it with Fuzzy numbers and aggregate the levels of hierarchy with each other, by incorporating the view of user as well as the one's involved in the software development process. In this paper we have proposed integrated framework combining FAHP with OWA to take care of uncertain preference of decision makers namely users, developer and programmer.

The OWA operator was first proposed by Yager [16] as tool for formation of overall decision function after aggregation of the multi-criteria. The OWA weight vector can aggregate the information provided by decision maker in imprecise and complete manner. Fuller and Majinder [7] proposed MV-OWA (Minimum variance-ordered weighted averaging) operator for obtaining weighting vector by minimizing variance for a given level of orness. O' Hagan [13] proposed ME-OWA (Maximum Entropy-ordered weighting averaging) operator for getting the weighting vector by maximizing entropy under the constraint of orness value.

OWA operator has been widely used in the field of hardware reliability for apportionment of target reliability goal among the subsystems. Chang et. al [3] did the reliability allocation for a fighter aircraft airborne radar system using the ME-OWA method. They also conducted a comparative case study to show how ME-OWA operator overcomes the shortcomings of Average weighting allocation method. Feng et. al [6] used the ME-OWA based on AHP for reliability allocation among the modules of the coal mining machine. Chen et. al [5] proposed MEMV-OWA technique based on AHP for reliability allocation for fuel cell vehicle.

In this paper we extend the concept of MEMV-OWA operator in the field of software reliability for this we have considered the architecture-based hierarchy that relates the user's view point to the software engineer's view point and further to the view of developer for a software system [4][20]. The objective is to achieve the target reliability by allocating reliability among the modules which are at the lowest level in the hierarchy. The model parameters i.e. the weights for reliability allocation are obtained through MEMV-OWA (Maximum Entropy Minimal Variance – Ordered Weighted Averaging) operator based on FAHP.

In real world always there is always possibility of some sort of uncertainty involved in preferences made and hence it is difficult for users, engineers and developers to assign precise value for their preferences. This gave the need of moving from crisp judgments to fuzzy judgements. Firstly FAHP is used for determination of weights of modules and then the proposed MEMV-OWA operator is used at developer's level of hierarchy to consider the imprecise preferences of the decision maker which may be subjective in nature. The weighting vector for modules obtained after applying the integrated

MEMV-OWA operator allocates the reliability using all the information available avoiding any kind of biasedness in making fuzzy judgements.

The paper is divided into six sections. Section II describes the MEMV-OWA operator in detail followed by description of Fuzzy Analytical hierarchical process in section III. In section IV the integrated framework of MEMV-OWA operator and FAHP for reliability allocation have been explained. In section V the proposed methodology is illustrated by an example pertaining to a multi-module software system. Finally section VI presents conclusion of the paper.

II. MATHEMATICAL BACKGROUND OF MEMV-OWA OPERATOR

A MEMV-OWA operator is used to determine the weights for reliability allocation to overcome the uncertainty involved in developer's preferences and biasedness in determination of weights. The weights are obtained by solving a bi-objective optimization model of minimizing the variance and maximization of entropy. The operator helps to maximize the use of available information (by maximization of entropy) along with reduction of biasedness in developer's inclination for different alternatives (by minimization of variance). Before we discuss the MEMV-OWA operator we will briefly describe OWA operator.

An **OWA operator** of dimension m is a function $L: \mathbb{R}^m \rightarrow \mathbb{R}$ with associated weighting vector $W = (w_1, w_2, \dots, w_m)^T \in \mathbb{R}^m$ such that it satisfies the following properties,

$$w_1 + w_2 + \dots + w_m = 1 \quad ; \quad 0 \leq w_j \leq 1 \quad j = 1, 2, \dots, m \quad (1)$$

and

$$L(k_1, k_2, \dots, k_m) = \sum_{j=1}^m w_j b_j \quad (2)$$

where b_j is the j th largest element of the aggregates k_1, k_2, \dots, k_m arranged in the descending order ($b_1 \geq b_2 \geq \dots \geq b_m$) before calculation of scalar product with W .

Yager [16] proposed characteristic measures related with weight vector of an OWA operator namely the orness of the aggregation and measure the entropy of the aggregated objects. Also Yager [17] introduced operator called the measure of variance. Now we discuss these characteristic in detail.

Orness Measure: It is defined as [16]

$$orness(W) = \sum_{j=1}^m \frac{m-j}{m-1} w_j = \gamma \quad (3)$$

and

$$orness(W) - \gamma \in [0, 1]; \quad andness = 1 - \gamma \quad (4)$$

for the weighting vector W . In decision maker's view it helps to represent the relationship between the attributes. For

instance γ being close to zero represents a higher andlike operation implying that decision maker is highly non-committal. On contrary if orness(W) is nearer to 1 then the relationship of aggregates is orlike and decision maker is highly optimistic. Similarly if decision maker has moderate assessment for attributes than $\gamma = 0.5$ and the weights will be $\frac{1}{m}$.

Entropy Measure: It is also called as measure of dispersion and is defined as [16]

$$disp(W) = -\sum_{j=1}^m w_j \ln w_j \quad (5)$$

$disp(W)$ is the degree of information utilized related to the attributes in an unspecified environment.

- If $w_j = 1$ and $w_i = 0$ ($i \neq j$) the entropy of W is minimum ($disp(W) = 0$) and hence one of the two attributes is used for the aggregation process.
- Also if $w_j = \frac{1}{m}$, $j = 1, 2, \dots, m$ the entropy of weighting vector W is maximum ($disp(W) = \ln(m)$) and hence all attributes are used in the aggregation.

Variance Measure: It is defined as [17]

$$D^2(W) = \frac{1}{m} \sum_{j=1}^m [w_j - E(W)]^2 = \frac{1}{m} \sum_{j=1}^m w_j - \frac{1}{m^2} \quad (6)$$

It determines the stretch in values of weighting vector W for a given value of orness and thus helps in avoiding overestimation of any single attribute over other attributes. This makes involvement of attributes in decision making a totally fair process.

In this paper we used the anticipated characteristic of OWA operator to form a bi-objective mathematical programming program that incorporates in itself the benefits of these characteristics [5].

Firstly we discuss the ME-OWA operator proposed by O'Hagan[13]. This operator maximizes the entropy under orness value constraint. The approach was as follows:

$$Max \quad -\sum_{j=1}^m w_j \ln w_j \quad (7)$$

$$subject \ to \ \sum_{j=1}^m \frac{m-j}{m-1} w_j = \gamma \quad (0 \leq \gamma \leq 1) \quad (8)$$

$$\sum_{j=1}^m w_j = 1, \quad 0 \leq w_j \leq 1, \quad j = 1, 2, \dots, m \quad (9)$$

Then Fuller and Majinder [7] solved the above optimization problem using Lagrange multiplier method and determined the optimal weighting vector. The optimal weighting vector can be calculated by using the underneath equations:

$$\ln w_j = \frac{j-1}{m-1} \ln w_m + \frac{m-j}{m-1} \ln w_1 \quad (10)$$

$$w_j = \frac{m-1}{\sqrt{w_1^{m-j} w_m^{j-1}}} \quad (11)$$

and

$$w_m = \frac{((m-1)\gamma - m)w_1 + 1}{(m-1)\gamma + 1 - mw_1} \quad (12)$$

then

$$w_1 [(m-1)\gamma + 1 - mw_1]^m = ((m-1)\gamma)^{m-1} [(m-1)\gamma - m]w_1 + 1 \quad (13)$$

Also Fuller and Majinder [8] proposed the MV-OWA operator to minimize variance of W the weighting vector associated with the attributes under the orness value constraint. The approach is as followed:

$$Minimize \quad \frac{1}{m} \sum_{j=1}^m w_j^2 - \frac{1}{m^2} \quad (14)$$

$$subject \ to \ \sum_{j=1}^m \frac{m-j}{m-1} w_j = \gamma, \quad (0 \leq \gamma \leq 1) \quad (15)$$

$$\sum_{j=1}^m w_j = 1, \quad 0 \leq w_j \leq 1, \quad j = 1, 2, \dots, m \quad (16)$$

Now we mathematically depict the MEMV-OWA model used in the present paper:

$$Maximize \quad -\sum_{j=1}^m w_j \ln w_j \quad (17)$$

$$Minimize \quad \frac{1}{m} \sum_{j=1}^m w_j^2 - \frac{1}{m^2} \quad (18)$$

$$subject \ to \ \sum_{j=1}^m \frac{m-j}{m-1} w_j = \gamma, \quad (0 \leq \gamma \leq 1) \quad (19)$$

$$\sum_{j=1}^m w_j = 1, \quad 0 \leq w_j \leq 1, \quad j = 1, 2, \dots, m \quad (20)$$

This is bi-objective problem with orness level as the constraint. The above problem can be solved in two steps. Firstly ideal point method (IPM) is employed to translate the bi-objective to a single objective and then lagrange multiplier method is used to determine the weights. Chen et. Al [5] has solved this problem considering that the two objectives given in equation (7),(14) that have their optimal values as g_1^0 and g_2^0 . Then IPM can be implemented as:

$$Minimize: \quad ||G - G^0|| = |g_1^0 - g_1| + |g_2^0 - g_2| - (g_1^0 - \frac{1}{m} \sum_{j=1}^m w_j^2 - \frac{1}{m^2}) + (g_2^0 - \sum_{j=1}^m w_j \ln w_j) \quad (21)$$

Applying lagrange multiplier method assuming ≥ 3 , we get;

$$L(W, \lambda_1, \lambda_2) = (g_1^0 - \frac{1}{m} \sum_{j=1}^m w_j^2 - \frac{1}{m^2}) + (g_2^0 - \sum_{j=1}^m w_j \ln w_j) + \lambda_1 (\sum_{j=1}^m \frac{m-j}{m-1} w_j - \gamma) + \lambda_2 (\sum_{j=1}^m w_j - 1) \quad (22)$$

where λ_1 and λ_2 are real numbers.

The weighting vector for $m = 4$ calculated by Chen et. al. [5] for different orness levels has been shown in Table 1.

Table 1: Weighting Vector for Maximum Entropy Minimum Variance Ordered Weighted Averaging Operator

w_j	0.5	0.6	0.7	0.8	0.9	1
w_1	0.2500	0.3466	0.4580	0.5898	0.7568	1
w_2	0.2500	0.2732	0.2800	0.2619	0.1945	0
w_3	0.2500	0.2138	0.1659	0.1067	0.0407	0
w_4	0.2500	0.1664	0.0961	0.0416	0.0080	0

III. MATHEMATICAL BACKGROUND OF FAHP

The proposed MEMV-OWA approach is based on FAHP since reliability allocation is to be implemented on a software system with a structured architectural hierarchy. The hierarchy has been discussed later in the paper. Firstly FAHP is used to allocate the reliability among the attributes through the proposed hierarchy and then MEMV-OWA operator is applied to deal with the uncertain preferences of the software developer.

AHP introduced by Thomas Saaty [15] is a multi-criteria decision framework based on the crisp judgements, but since the real world problems encompass a lot of uncertainty and complexity, there is a need to move towards fuzzy environment.

Now we discuss briefly the mathematical concepts involved in the study of FAHP:

Fuzzy Set: A fuzzy set assigns to each object of its class a grade ranging between 0 and 1 through a membership function. Let A be the fuzzy subset of universal set Z for $z \in Z$. Then $\mu_A(z) \in [0,1]$ called membership function represents membership of z to A .

Fuzzy Number theory: The theory of Fuzzy set was initially given by Zadeh [19] to incorporate in itself the ambiguity involved in decisions due to inaccuracy in the data and the system.

Definition: A Fuzzy number ' A ' is the normal and convex subset of the universe fuzzy set Z i.e. mathematically $\forall z_1, z_2 \in Z$ and $\forall \lambda \in [0,1]$ [19]

$$\mu_A(\lambda z_1 + (1 - \lambda)z_2) \geq \min(\mu_A(z_1), \mu_A(z_2))$$

where $\mu_A(z) \in [0,1]$ is the membership function of A .

Triangular Fuzzy number: A Fuzzy number $A = (l, m, u)$ is defined as Triangular Fuzzy number (TFN) (Kauffmann and Gupta 1991) if its membership function $\mu_A(z)$ is equal to:

$$\mu_A(z) = \begin{cases} \frac{z-l}{m-l} & l \leq z \leq m \\ \frac{u-z}{u-m} & m \leq z \leq u \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

Graphically TFN has been presented in Figure 1.

The operation laws of for addition, multiplication and inverse are defined as follows:

$$\begin{aligned} (l_1, m_1, u_1) + (l_2, m_2, u_2) &= (l_1 + l_2, m_1 + m_2, u_1 + u_2) \\ (l_1, m_1, u_1) \times (l_2, m_2, u_2) &= (l_1 \times l_2, m_1 \times m_2, u_1 \times u_2) \\ (l_1, m_1, u_1)^{-1} &= \left(\frac{1}{u_1}, \frac{1}{m_1}, \frac{1}{l_1}\right) \end{aligned}$$

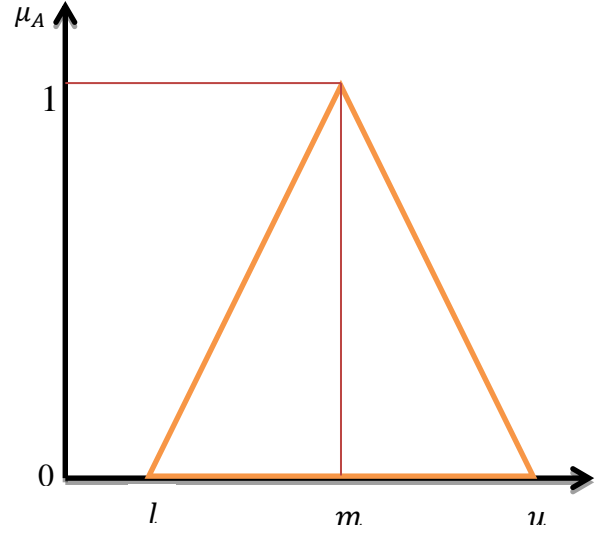


Figure 1: Triangular Fuzzy Number

The preferences of decision makers are expressed through triangular fuzzy numbers. The fuzzy scale is shown in Table 2. Then the goal is to finally obtain relative weights of modules and thus allocate reliability. The FAHP is used for the purpose of combining the user, engineers and development view of assessment. The steps involved in FAHP are:

- 1) Develop a hierarchy after defining the goal of the problem.
- 2) Establishment of fuzzy judgement matrix and the fuzzy weighting vector by employing pairwise comparisons among elements.
- 3) Relative weights of elements are aggregated and optimal one is selected by obtaining the overall rating of the element.

Table 2: Fuzzy Scale

Linguistic Scale	Triangular scale	Triangular fuzzy reciprocal scale
Just Equal	(1,1,1)	(1,1,1)
Equally Important	(1/2,1,3/2)	(2/3,1,2)
Weakly Important	(1,3/2,2)	(1/2,2/3,1)
Moderately Important	(3/2,2,5/2)	(2/5,1/2,2/3)
Moderately More Important	(2,5/2,3)	(1/3,2/5,1/2)
More Important	(5/2,3,7/2)	(2/7,1/3,2/5)
Strongly Important	(3,7/2,4)	(1/4,2/7,1/3)
Strongly More Important	(7/2,4,9/2)	(2/9,1/4,2/7)

It is important to test the consistency of all the pairwise comparison matrices for calculation of weights. It can be done easily by following the defuzzification process. The triangular fuzzy matrix is firstly defuzzified and then on the same lines as for consistency check in traditional AHP's consistency is checked for the defuzzified matrix. The matrix is accepted and we proceed for weight vector calculation if consistency is less than 0.1. For a given TFN $A = (l, m, u)$ the defuzzified real

number is $= \frac{l+2m+u}{4}$. Using the obtained real numbers the consistency is calculated according to underneath equations:

$$CI = \frac{\lambda_{max} - k}{k - 1}$$

$$CR = \frac{CI}{RI}$$

where k is the order of the pairwise matrix, RI is random index and CI is consistency index.

In this paper we follow the Chang's extent analysis [2] for the determination of parameters although many other ways have been developed in literature for the purpose. Let $Z = (z_1, z_2, \dots, z_m)$ be an object set and $X = (x_1, x_2, \dots, x_n)$ be the goal set. Following the Chang's extent analysis approach each object is taken one by one and the analysis is performed over every goal individually. Therefore n values of extent analysis are obtained for each object. i.e. $A_{x_i}^1, A_{x_i}^2, \dots, A_{x_i}^n$ where $A_{x_i}^j$ are TFN's, i is $1, 2, \dots, m$ and $j = 1, 2, \dots, n$.

The steps of performing the Chang's extent analysis are:

1. The Fuzzy synthetic extent value of i th object for n extent values is given as

$$S_i = \sum_{j=1}^n A_{x_i}^j \times \left[\sum_{i=1}^m \sum_{j=1}^n A_{x_i}^j \right]^{-1} \quad (24)$$

where

$$\sum_{j=1}^n A_{x_i}^j = \left(\sum_{j=1}^n l_j, \sum_{j=1}^n m_j, \sum_{j=1}^n u_j \right),$$

$$\sum_{i=1}^m \sum_{j=1}^n A_{x_i}^j = \left(\sum_{i=1}^m l_i, \sum_{i=1}^m m_i, \sum_{i=1}^m u_i \right) \text{ and}$$

$$\left[\sum_{i=1}^m \sum_{j=1}^n A_{x_i}^j \right]^{-1} = \left(\frac{1}{\sum_{i=1}^m u_i}, \frac{1}{\sum_{i=1}^m m_i}, \frac{1}{\sum_{i=1}^m l_i} \right)$$

2. The degree of possibility of $A_1 \geq A_2$ is given as:

$$V(A_1 \geq A_2) = \sup_{z_1 \geq z_2} [\min(\mu_{A_1}(z_1), \mu_{A_2}(z_2))] \quad (25)$$

$$V(A_i \geq A_2) = 1 \quad \text{if } \mu_{A_1}(z_1) = \mu_{A_2}(z_2) = 1 \quad (26)$$

Since A_1 and A_2 are fuzzy numbers (normal and convex by nature), Therefore we have:

$$V(A_1 \geq A_2) = 1 \quad \text{iff } m_1 \geq m_2 \quad (27)$$

$$V(A_1 \geq A_2) = hgt(A_1 \cap A_2) \mu_{A_1}(d) \quad (28)$$

where d represents the ordinate of point D and V is given as (Figure 2)

$$= \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} \quad (29)$$

It is the point of highest intersection between μ_{A_1} and μ_{A_2} .

Both $V(A_1 \geq A_2)$ and $V(A_2 \geq A_1)$ are required for the comparison.

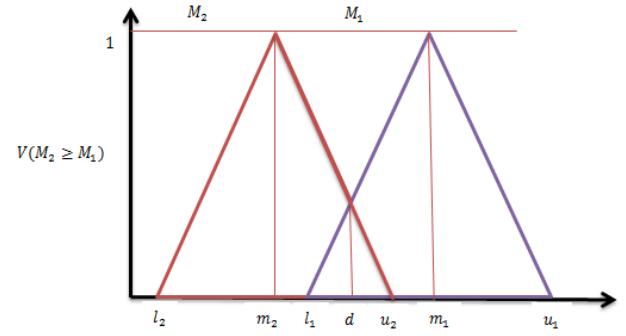


Figure 2: The intersection between M_1 and M_2 .

The possibility degree for a fuzzy number to be greater than l fuzzy numbers $A_i (i = 1, 2, \dots, l)$ is given as:

$$V(A \geq A_1, A_2, \dots, A_l) = V[(A \geq A_1) \text{ and } (A \geq A_2) \text{ and } \dots (A \geq A_l)] = \min V(A \geq A_i) \quad i = 1, 2, \dots, l \quad (30)$$

It is assumed that $d'(A_i) = \min V(S_i \geq S_j)$ for $l = 1, 2, \dots, m \quad l \neq i$.

Therefore weight vector is $W' = (d'(A_1), d'(A_2), \dots, d'(A_m))^T$.

The obtained weight vector needs to be normalized. The normalized weights are determined by dividing the each value of vector W' by the sum of values in W' . Mathematically the normalized weighting vector is given as:

$$= \left(\frac{d'(A_1)}{d'(A_1) + d'(A_2) + \dots + d'(A_m)}, \frac{d'(A_2)}{d'(A_1) + d'(A_2) + \dots + d'(A_m)}, \dots, \frac{d'(A_m)}{d'(A_1) + d'(A_2) + \dots + d'(A_m)} \right)$$

IV. METHODOLOGY OF RELIABILITY ALLOCATION

There is always a possibility of some sort of uncertainty, subjectivity and incompleteness involved in the preferences made by developers in real situations. Therefore it is not just enough to use traditional AHP's for reliability allocation. Practically it is difficult for users, engineers and developers to assign a specific value of importance for functions, programs and modules respectively. Therefore one moved to fuzzy conclusion from the crisp conclusion.

In this paper MEMV-OWA operator based on FAHP has been used for reliability allocation of a software system. Firstly the weights for the modules at the fifth level of hierarchy are derived using FAHP then MEMV-OWA operator is applied at the developer's level i.e. at the fourth level of hierarchy to combat the uncertainty involved in the developer's decision during making preferences for modules. Step by step procedure for the methodology is as follows:

- 1) Set reliability target for the software system.
- 2) Build the Software hierarchical model for the reliability apportionment process.
- 3) Construct the fuzzy pairwise comparison matrix for each level in hierarchy.

- 4) Calculate the λ_{max} (maximum Eigen value) and the corresponding normalized Eigen vectors after defuzzification of the Triangular Fuzzy numbers.
- 5) Check the consistency of the comparison matrix. If consistency not satisfied (i.e. $CI > 0.1$) then repeat steps 3 and 4.
- 6) On the basis of orness value γ obtain the OWA weight $W = (w_1, w_2, \dots, w_m)$ at developers level.
- 7) Obtain the value of OWA operator L using equation (2).
- 8) At last the reliability allocated to module is calculated according to the equation (37).

Architectural Hierarchy of the Software System

In our model we have developed the hierarchy based on [4][20] that links the user's view about functionalities to the software engineer's view of programs and further it is linked to the developer's or programmer's view of modules. The hierarchy uses the top-down approach for the reliability apportionment by combining the users, software engineers and programmer's assessment.

The first level of hierarchy is the overall system reliability target (R) based on user's view of various functionalities of software. The software system is developed to perform certain functions (given at second level of hierarchy) as per user's expectation. This level represents the user's view in the software system. The functions are denoted by $F_i (i = 1, 2, \dots, f)$ and third level of hierarchy consists of programs designed by the software engineers (SE) to accommodate the various functions specified by the user. It is denoted by P_j 's ($j = 1, 2, \dots, p$). While this level is from the SE's point of view. At the fourth level i.e. the last level of hierarchy is the modules written by programmers representing programmers view. It is denoted by $M_k (k = 1, 2, \dots, m)$. The hierarchy has been depicted diagrammatically through Figure 3. In the network we can observe that a program and a module may be connected to more than one functions and program respectively i.e. there exist many-to-many connections between them. In other words we can say that a program designed can be beneficial in performing one or more than one function simultaneously. Similarly a module can also be connected to one or more than one programs. The modules are independent in nature but they may have submodules with one-to-one connection.

The aim is to find the relative weights of each function F_i at the second level, program P_j at the third level and module M_k at the fourth level of hierarchy from the knowledge of reliability goal set by the user for the software system based on their assessment of the functions. In this paper we have calculated the alternative weights of modules considering the uncertain preferences of the developer while making decisions using Minimum Variance Maximum Entropy – Ordered weighted averaging operator.

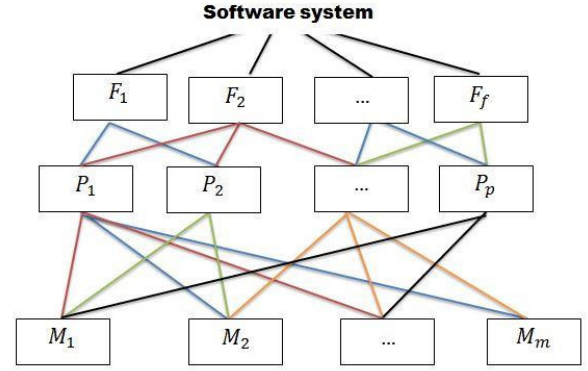


Figure 3: Architectural hierarchy of the software system

Relative Reliability Weight of Module

Once the architectural hierarchy has been formed and the reliability target is set, the next step is to form pairwise comparison matrix based on user's opinion. The data has been collected in terms of fuzzy numbers. Let it be given as follows:

$$\begin{bmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \dots & (l_{1f}, m_{1f}, u_{1f}) \\ (l_{21}, m_{21}, u_{21}) & (1,1,1) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ (l_{n1}, m_{n1}, u_{n1}) & (l_{n2}, m_{n2}, u_{n2}) & \dots & (1,1,1) \end{bmatrix} \quad (31)$$

where $(l_{ij}, m_{ij}, u_{ij}) = \left(\frac{1}{u_{ji}}, \frac{1}{m_{ji}}, \frac{1}{l_{ji}} \right)$ for $i, j = 1, 2, \dots, n$ and $n \neq j$

At the second level users were asked to compare the functions (F_i 's) pairwise. The total number of pairwise evaluation will be $\frac{f!}{2 \times (f-2)!}$ where ' f ' the total number of functions specified by the user. Similarly we go for the other levels of the hierarchy. The results of comparison of functions are stored in $A_f = [a_{fij}]$ a $f \times f$ matrix. The element a_{fij} gives the importance of i element over the j th element, the diagonal elements are obviously (1,1,1) and the lower part of the matrix are inverse of the corresponding upper triangular matrix elements. As mentioned earlier the relative weights at this stage are derived using Chang's extent analysis [1996].

Let relative weights for Function be denoted as $(WF_1, WF_2, \dots, WF_f)$. Allocated reliability for the function provided they are executed every time is given by [1].

$$RF_i = (R)^{WF_i} \quad (32)$$

Similarly the input matrix A_p and A_m are formed for programs and modules by taking opinion from the system engineer and programmer. Let the relative weight matrix for the program may be given as follows:

$$\begin{bmatrix} WP_1^1 & WP_2^1 & \dots & WP_p^1 \\ WP_1^2 & WP_2^2 & \dots & WP_p^2 \\ \dots & \dots & \dots & \dots \\ WP_1^f & WP_2^f & \dots & WP_p^f \end{bmatrix} \quad (33)$$

WP_j^i 's are the relative weights at the software engineer's level indicating weight for the j th program of the i th function. At the developer's level of hierarchy we need p number of pairwise matrices for comparison. The target reliability at this level will be RF_i for the programs connected to the function i . For the i th function the relative weights of related programs are given by

$$WP^i = (WP_j^i) \quad \forall i = 1, 2, \dots, f \text{ and } j \in P \\ \subset \{1, 2, \dots, p\} \text{ program related to } F_i\}$$

Then the reliability allocated to the programs is

$$RP_j^i = (RF_i)^{WP_j^i} \quad \forall i = 1, 2, \dots, f \quad j \in P \subset \\ \{1, 2, \dots, p\} \text{ program connected to } F_i\} \quad (34)$$

Doing this we obtain different reliability values of same program, since a function can be connected to more than one function. Therefore it will be best to choose the one with highest reliability.

$$RP_j = \text{Maximum}(RP_j^i) \quad \forall j = 1, 2, \dots, p \text{ and } i \in F \\ \subset \{1, 2, \dots, f\} \\ \text{for all functions in which } j\text{th program is called}\} \\ (35)$$

Here RP_j is the allocated reliability of the j th program. Similar approach is used for other levels of the hierarchy. At fourth level the reliability value allocated to modules is calculated.

We further calculate the weights of modules as follows:

$$WM^j = (WM_k^j) \quad \forall j = 1, 2, \dots, p \text{ and } k \in M \subset \\ \{1, 2, \dots, m\} \text{ for the module related to } P_j\}$$

Hence the reliability allocated to the programs is

$$RM_k^j = (RP_j)^{WM_k^j} \quad \forall j = 1, 2, \dots, p \quad k \in M \subset \\ \{1, 2, \dots, m\} \text{ module connected to } P_j\} \quad (36)$$

Doing this we obtain different reliability values of same module, since a module can be connected to more than one program. Therefore it will be best to choose the one with highest reliability.

$$RM_k = \text{Maximum}(RM_k^j) \quad \forall k = 1, 2, \dots, m \text{ and } j \in P \\ \subset \{1, 2, \dots, p\} \\ \text{for all programs in which } k\text{th module is called}\} \\ (37)$$

Hence RF_j , RP_k and RM_l are allocated reliability values of the functions, programs and modules respectively.

Relative weight of modules applying MEMV-OWA operator at program level of hierarchy

In section two of this paper we have shown the weight vector of MEMV-OWA operator against different orness values ranging from 0.5 to 1 (Table 1). Now these weights are used to obtain the final weights assigned to each module at the lowest level of hierarchy. The OWA function L_i is given as:

$$L_i(WP_1, WP_2, \dots, WP_p) = \sum_{k=1}^p w_k \times b_{ik} \quad (38)$$

where L_i is the weight of the i th module and b_{ik} is the k th largest element in vector $(WM_1, WM_2, \dots, WM_m)$ for a program. Similarly OWA weights for other modules are calculated. The obtained weight incorporates in itself the developer's uncertain preferences which are always subjective to outside fact.

V. APPLICATION EXAMPLE OF PROPOSED RELIABILITY ALLOCATION MODEL

We exhibit the FAHP based MEMV-OWA technique of Reliability apportionment with an example. Let us consider software architecture with three functions (F_1, F_2, F_3). The software engineer and programmers established four programs (P_1, P_2, P_3 and P_4) and six modules ($M_1, M_2, M_3, M_4, M_5, M_6$) respectively to accomplish the functions expected by the users. The linkages between functions, programs and modules are as follows:

- 1) Software engineers designs the program to execute the functions expected by the user. P_1, P_2 and P_4 serve the function F_1 , P_1, P_2 and P_3 serve function F_2 and P_2, P_3 and P_4 serve function F_3 .
- 2) Further developers or programmers use six modules to implement the 4 programs. P_1 appeals for modules M_1, M_2 and M_5 , P_2 appeals for modules M_2, M_3 and M_4 , P_3 appeals for modules M_4, M_5 and M_6 and program P_4 appeals for M_1, M_3, M_6 .

The target reliability for the software system is set to 0.90 based on the necessities of users. The total number of triangular fuzzy pairwise comparison matrices at different levels of hierarchy is 1, 3 and 4 respectively. After obtaining weights of modules using FAHP we apply the MEMV-OWA operator at software developer's level to find the weights for modules and thus allocate the reliability.

Based on the fuzzy scale given in Table 2 the fuzzy pairwise comparison matrices have been constructed according to views of users, SE (Software Engineers) and developers.

Reliability allocation for Functions:

Fuzzy matrix obtained for system after the comparison of functions is given by:

$$\begin{matrix} & F_1 & F_2 & F_3 \\ \begin{matrix} F_1 \\ F_2 \\ F_3 \end{matrix} & \begin{pmatrix} (1,1,1) & (3/2,2,5/2) & (3/2,2,5/2) \\ (2/5,1/2,2/3) & (1,1,1) & 2/3,1,2) \\ (2/5,1/2,2/3) & (1/2,1,3/2) & (1,1,1) \end{pmatrix} \end{matrix}$$

Now we use the FAHP method discussed earlier to determine the relative weights for the functions and thus the corresponding reliabilities. The weights and reliabilities so determined are as follows:

The normalized weighting vector is $WF = (0.523077, 0.271111, 0.205805)$ and reliabilities allocated to functions using expression (32) are:

$$\begin{aligned} RF_1 &= (R)^{WF_1} = 0.946380697 \\ RF_2 &= (R)^{WF_2} = 0.9718309859 \\ RF_3 &= (R)^{WF_3} = 0.97854954 \end{aligned}$$

The reliability allocated to a function is used as the reliability goal to allocate reliability among the programs associated with that particular function.

Reliability Allocation for Programs:

Fuzzy Pairwise comparison matrix for programs associated with function F_1 are :

$$\begin{matrix} & P_1 & P_2 & P_4 \\ \begin{matrix} P_1 \\ P_2 \\ P_4 \end{matrix} & \begin{pmatrix} (1,1,1) & (3/2,2,5/2) & (1,3/2,2) \\ (2/5,1/2,2/3) & (1,1,1) & (2/5,1/2,2/3) \\ (1/2,2/3,1) & (3/2,2,5/2) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weighting vector is $WF_1 = (0.428928, 0.142145, 0.428928)$ and reliabilities allocated to programs of function F_1 using expression (34) are:

$$\begin{aligned} RP_1 &= (RF_1)^{WP_1} = 0.976635514 \\ RP_2 &= (RF_1)^{WP_2} = 0.9921962096 \\ RP_4 &= (RF_1)^{WP_4} = 0.976635514 \end{aligned}$$

For Function F_2

$$\begin{matrix} & P_1 & P_2 & P_3 \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \end{matrix} & \begin{pmatrix} (1,1,1) & (2/3,1,3/2) & (2/9,1/4,2/7) \\ (2/3,1,3/2) & (1,1,1) & (2/5,1/2,2/3) \\ (7/2,4,9/2) & (3/2,2,5/2) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WF_2 = (0.41629, 0.291855, 0.291855)$

$$RP_1 = (RF_2)^{WP_1} = 0.988175676$$

$$\begin{aligned} RP_2 &= (RF_2)^{WP_2} = 0.991694352 \\ RP_3 &= (RF_2)^{WP_3} = 0.991694352 \end{aligned}$$

For function F_3

$$\begin{matrix} & P_2 & P_3 & P_4 \\ \begin{matrix} P_2 \\ P_3 \\ P_4 \end{matrix} & \begin{pmatrix} (1,1,1) & (2/3,1,3/2) & (2/7,1/3,2/5) \\ (2/3,1,3/2) & (1,1,1) & (2/5,1/2,2/3) \\ (5/2,3,7/2) & (3/2,2,5/2) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WF_3 = (0.423036, 0.288482, 0.288482)$

$$\begin{aligned} RP_2 &= (RF_3)^{WP_2} = 0.99086758 \\ RP_3 &= (RF_3)^{WP_3} = 0.993762994 \\ RP_4 &= (RF_3)^{WP_4} = 0.99372994 \end{aligned}$$

Since a program written is for accomplishment of more than one function of the software so the final reliability allocated to a program using expression (35):

$$\begin{aligned} RP_1 &= \text{Maximum}(0.976635514, 0.988175676) = 0.988175676 \\ RP_2 &= \text{Maximum}(0.9921962096, 0.991694352) = 0.992196209 \\ RP_3 &= \text{Maximum}(0.991694352, 0.993762994) = 0.993762994 \\ RP_4 &= \text{Maximum}(0.976635514, 0.993762994) = 0.993762994 \end{aligned}$$

Weight calculation for Modules:

We only determine weights for modules using FAHP method because reliability allocation for modules is to be done using the weights obtained from MEMV-OWA technique.

For program P_1 the fuzzy pairwise comparison matrix is given as:

$$\begin{matrix} & M_1 & M_2 & M_5 \\ \begin{matrix} M_1 \\ M_2 \\ M_5 \end{matrix} & \begin{pmatrix} (1,1,1) & (2/7,1/3,2/5) & (2/7,1/3,2/5) \\ (2/3,1,3/2) & (1,1,1) & (2/5,1/2,2/3) \\ (5/2,3,7/2) & (3/2,2,5/2) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WP_1 = (0.3520129, 0.3520129, 0.295943)$

For Program P_2

$$\begin{matrix} & M_2 & M_3 & M_4 \\ \begin{matrix} M_2 \\ M_3 \\ M_4 \end{matrix} & \begin{pmatrix} (1,1,1) & (2,5/2,3) & (5/2,3,7/2) \\ (1/3,2/5,1/2) & (1,1,1) & (1/2,2/3,1) \\ (2/7,1/3,2/5) & (1,3/2,2) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WP_2 = (0.248739, 0.248739, 0.502523)$

For program P_3

$$\begin{matrix} & M_4 & M_5 & M_6 \\ \begin{matrix} M_4 \\ M_5 \\ M_6 \end{matrix} & \begin{pmatrix} (1,1,1) & (2/7,1/3,2/5) & (1,3/2,2) \\ (5/2,3,7/2) & (1,1,1) & (7/2,4,9/2) \\ (1/2,2/3,1) & (2/9,1/4,2/7) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WP_3 = (0.3520129, 0.3520129, 0.295943)$

For program P_4

$$\begin{matrix} & M_1 & M_3 & M_6 \\ \begin{matrix} M_1 \\ M_3 \\ M_6 \end{matrix} & \begin{pmatrix} (1,1,1) & (2/9,1/4,2/7) & (1,3/2,2) \\ (7/2,4,9/2) & (1,1,1) & (7/2,4,9/2) \\ (1/2,2/3,1) & (2/9,1/4,2/7) & (1,1,1) \end{pmatrix} \end{matrix}$$

The normalized weights are $WP_4 = (0.345638, 0.345638, 0.308725)$

Consistency of each pairwise matrix has been checked by using defuzzification process discussed earlier. The essential details of the consistency have been shown in Table 3.

Table 3: Consistency check for pairwise comparison matrices

Matrix	λ_{max}	n (order of the matrix)	CI	RI	CR
System	3.0764180613	3	0.0382093	0.58	0.065877639
F_1	3.058563675	3	0.029282838	0.58	0.050485927
F_2	3.098245614	3	0.049122088	0.58	0.084693255
F_3	3.063174114	3	0.031586998	0.58	0.054460342
P_1	3.02875	3	0.01437526	0.58	0.024784931
P_2	3.06779661	3	0.033901652	0.58	0.058451124
P_3	3.02875	3	0.01437526	0.58	0.024784931
P_4	3.0417101	3	0.020855165	0.58	0.035957182

The weight vector for different modules corresponding to a program obtained using FAHP is shown in Table 4.

Table 4: Modular weights obtained using FAHP

Allocated Reliability to program	Module	P_1	P_2	P_3	P_4
	M_1	0.3520129	—	—	0.345638
	M_2	0.3520129	0.248739	—	—
	M_3	—	0.248739	—	0.345638
	M_4	—	0.502523	0.3520129	—
	M_5	0.295943	—	0.3520129	—
	M_6	—	—	0.295943	0.308725

Now MEMV-OWA operator is applied at the software developer's level in the hierarchy to obtain the weights for the modules and thus allocate the reliability. The weights for the OWA operator are given in Table 1 for different orness values. We have assigned weight zero for modules which are not connected to the program and then evaluated the relative weights of the modules for different orness values i.e. from 0.5 to 1. The weights obtained at different orness levels are shown in Figure 4, the graphical representation compares modular weights obtained at different orness levels. This clearly

implies that the subjective attitude of the decision maker affects the weights assigned for allocation.

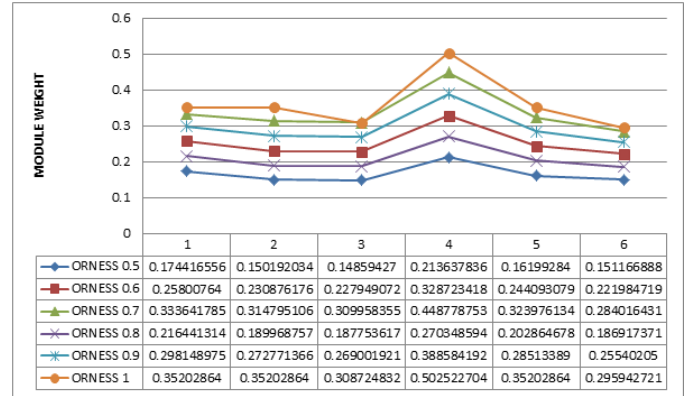


Figure 4: Comparison of Module weights for different orness levels

Now using the weights obtained after application of MEMV-OWA operator we find the reliability allocated to each module under different orness i.e. under different possible perspectives of the developer in an uncertain environment. The graphical representation of allocated reliability is shown in Figure 5. We can see that it is important to take into consideration the developer's behavior. The developer should neither be over-optimistic nor under optimistic. Reliability allocated for orness value is 1 whereas it is maximum for the average optimism behavior i.e. considering every alternative to be equally important.

After obtaining the relative weights for module we allocate reliability using expression (36). Since a module is serving more than one program, therefore maximum obtained reliability is allocated to the module using expression (37).

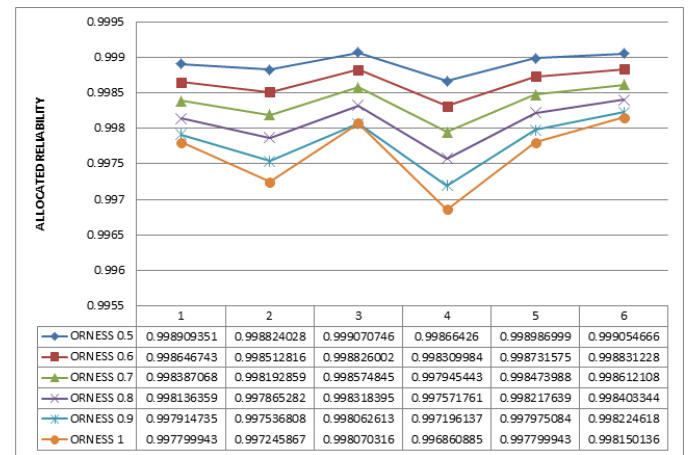


Figure 5: Comparison of Reliability Allocated at different orness levels

VI. CONCLUSIONS

This paper is contributing to the reliability allocation literature of software systems by introduction of a FAHP based MEMV-OWA technique by considering the uncertain preferences of the developer during the planning and design phase. The bi-objective MEMV-OWA technique problem

provides an unbiased way for the allocation of reliability among the modules by minimizing the variance and makes use of maximum information by maximizing the entropy. During reliability allocation in addition to the user's opinion at function level we have also incorporated the system engineers and programmers view at program and module level respectively. An example with the application of methodology has been demonstrated which makes the steps involved more clear. At different orness levels the weight and reliability allocated has been calculated to illustrate the sensitivity of optimal reliability allocation solution to the different level of agreement among the developer's preferences.

REFERENCES

- [1] K.K. Aggrawal and Y. Singh, "Software reliability Apportionment Using Analytical hierarchical process", ACM SIGSOFT software engineering notes, Vol 20, Issue 5, pp. 56-61, 1995.
- [2] D.Y. Chang, "Applications of the extent analysis method on fuzzy AHP", European Journal of Operational Research, Vol 95, Issue 3, pp. 649-655, 1996.
- [3] Yung-Chia Chang, Kuie-Hu Chang and Cheng-Shih Liaw, "Innovative reliability allocation using the maximal entropy ordered weighted averaging method", Computers and Industrial engineering, Vol 57, Issue 4, pp. 1274-1281, 2009.
- [4] S. Chatterjee, J.B. Singh and A. Roy, "A structure-based software reliability allocation using fuzzy analytical hierarchical process", International journal of systems science, Vol 46, Issue 3, pp. 513-525, 2015.
- [5] T. Chen, S. Zheng, H. Liao and J. Feng, "A Multi-Attribute reliability Allocation Method considering Uncertain Preferences", Quality and Reliability engineering international, Vol 32, pp. 2233-2244, 2016.
- [6] Y. Feng, Z. Hong, J. Cheng, G. Tian and H. Zhang, "Environment friendly reliability allocation for product platform based on expert measurement and ICN", Computers and Electrical Engineering, Vol 64, pp. 132-144 2017.
- [7] R. Fuller and P. Majinder, "An analytical approach for obtaining maximal entropy OWA operator weights". Fuzzy Sets and Systems, Vol 124, Issue 1, pp. 53-57, 2001.
- [8] R. Fuller, and P. Majinder, "On obtaining minimal variability OWA operator weights", Fuzzy sets and systems, Vol 136, pp. 203-215, 2003.
- [9] M. E. Halendar, M. Zaho and N. Ohlsson, "Planning models for software reliability and cost", IEEE Trans. Software engineering, Vol 24, Issue 6, pp. 424-434, 1998.
- [10] A. Kaufmann and M. M. Gupta, "Introduction to Fuzzy arithmetic: Theory and Applications", New York: Van Nostrand Reinhold 1986.
- [11] P. Kubat, "Assessing reliability of modular software", Operational research letters, Vol 8, Issue 1, pp.35-41, 1989.
- [12] Leung, Y.W. (1997). Software reliability allocation under an uncertain operational profile, Journal of Operational Research Society, Vol. 48, No. 4 pp. 401-411.
- [13] M. O'Hagan, "Aggregating template or rule antecedents in real time expert systems with fuzzy set logic", in: Proceedings 22nd Annual IEEE Asilomar Conference Signals, Systems and Computers, Pacific Grove, CA, Piscataway, Vol. 8, 1988.
- [14] I. Rani and R. B Misra, "Economic allocation of target reliability in modular software systems", in: Proceedings of the annual Reliability and Maintainability Symposium, pp 428-432, 2005.
- [15] T. L. Satty, "The Analytical Hierarchy Process: Planning, Priority Setting, resource Allocation", McGraw-Hill, Newyork, 1980.
- [16] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", IEEE Transactions on Systems, Man and Cybernetics, Vol. 18, Issue 1, 183-190, 1988.
- [17] R. R. Yager, "On inclusion of variance in decision making under uncertainty". International journal of uncertainty, Fuzziness and Knowledge-based systems, Vol. 4, Issue 5, pp. 401-419, 1996.
- [18] F. Yue, Goufu Zhang, Zhaopin Su, Yang Lu and Ting Zhang, "Multi-software reliability allocation in multimedia systems with budget constraints using Dempster-Shafer theory and improved differential evolution", Neurocomputing, Vol. 169, pp. 13-22, 2015.
- [19] L. A. Zadeh, "Fuzzy Sets", Information and control, Vol. 8, Issue 3, pp. 338-353, 1965.
- [20] F. Zahedi and N. Ashrafi "Software reliability allocation based on structure, utility, price and cost", IEEE Trans. Software engineering, Vol. 17, Issue 4, pp. 401-411, 1991.

An Innovative B2C E-commerce Websites Selection using the ME-OWA and Fuzzy AHP

Anu G. Aggarwal, Aakash

*Department of Operational Research, University of Delhi, Delhi, India
{anuagg17, aakash.du.or50}@gmail.com*

Abstract—Today internet has emerged as a huge marketplace of products and services for meeting needs of more than a million customers worldwide. It provides users a platform to access information globally in electronic form as well as in terms of business transaction, such as, e-payments, e-orders and e-booking etc. The advent of the internet has led to the establishment of electronic commerce. Today a large number of B2C e-commerce websites are available, which makes it difficult not only for the customers to find right product at right price, but also for a company to choose a better site for selling its product. Thus, there is need to rank e-commerce websites in B2C electronic commerce. The objective of this paper is to rank e-commerce websites on the basis of success factors, namely, System Quality, Content Quality, Usages, Trust, Customer Support, Online Customer Feedback and Personalization. Here we have used a two stage approach combining maximum entropy-ordered weighted averaging aggregation (ME-OWA) with fuzzy Analytic Hierarchy Process (FAHP) for choosing the best B2C e-commerce website.

Index Terms—Customer Oriented Success Factors, E-commerce, Fuzzy Set, Fuzzy AHP, Fuzzy ME-OWA

I. INTRODUCTION

Now A days, business organization use e-commerce system to communicate with their customers and to provide online business transactions. With the deep penetration, e-commerce is in use. To sustain the cut-throat competition market, it has become essential for business to direct towards online business platforms. Now combination of business and internet has become necessary for B2C e-commerce organizations to coordinate with their customer and stockholders. Business-to-consumer (B2C) e-commerce website is a used to communicate with customers and to provide business transactions. The e-commerce has grown continuously all over the world, its growth has occurred in both the countries either it developed or it is developing, which is the reason for increase in the online shopping community in the world [1].

The success of any business organization depends on the customer satisfaction level. Similarly, the success of B2C e-commerce also depends upon the customer satisfaction towards e-commerce websites. In this paper, we propose a framework based on maximum entropy ordered weighting averaging operator (ME-OWA) and fuzzy AHP to select the best B2C e-commerce website on the bases of customers oriented success factors. The ME-OWA is used for calculat-

ing the orness weights of evaluation criteria and fuzzy AHP is used for calculating the weight vector for alternatives respect to each criterion. The weights so obtained are used to rank the alternative through ordered weighted averaging aggregation (OWA) operator.

This paper is organized as follows: Section 2 gives the detailed research background. The methodology used in this paper has been described in the Section 3. The application part of the proposed methodology is shown in the Section 4. Section 5 concludes the paper.

II. RESEARCH BACKGROUND

More than two decades ago DeLone and McLean initially proposed information system (IS) success model [2]. This is the iconic model in the area of success measurements that has found vast application in many domains which includes business education, human resource management, etc. The major dimensions of information system success were system quality, information quality, use, user satisfaction, individual impact, and organizational impact. This model provides a basis for evaluating the performance of information systems. Later in 2001, Molla and Licker (M&L) shown that IS success model can be used for identifying the success of e-commerce system [3]. They highlighted the significance of customer satisfaction and their experience, towards e-commerce system. M&L defined trust as additional factor. The main factors of M&L's model were content quality, system quality, use and user satisfaction, support & service, and trust. As compared to DeLone and McLean's (D&M) IS success model, the M&L's model used content quality instead of information quality, user satisfaction for e-commerce customers, service in the place of service quality, and trust added as a new factor in B2C e-commerce environment.

DeLone and McLean updated their IS success model after a decade in 2002 with additional success factors intention to use, use and net benefits, and further extended in 2003. The main elements included by them were information quality, system quality user satisfaction, intention to use, use, net benefits, and service quality [4] [5]. DeLone and McLean demonstrated that D&M's updated IS success model could be used for evaluating the e-commerce success without any changes [6]. This model is based on the updated D&M's IS

success model in 2002 and extended in 2003. The success dimension of DeLone and McLean (2004) model used for evaluating the e-commerce success are system quality, information quality, service quality, usage, user satisfaction, and net benefits [6].

Today's B2C e-commerce world has many features which can affect the B2C e-commerce success such as: customer feedback and personalization. Customer feedback is a crucial factor now days, measured in terms of online reviews, ratings [7], and helpfulness [8]. Personalization is defined as the customization of products and services in accordance to the user needs [9]. This factor depends on a number of attributes such as individual preferences, cross-selling and up-selling for evaluating the success of B2C e-commerce. Therefore, there is a growing need to extend the previous B2C e-commerce success model with these latest B2C e-commerce success factors. Overall, we have included seven success criteria where five criteria namely system quality "(SQ)", content quality "(CQ)", Usage "(US)", Trust "(TR)", customer support "(CS)" are same as specified by D&M [6].

The two additional success criteria included in the proposed model are customer feedback "(CF)" and personalization "(PR)". At present a large number of B2C e-commerce websites are offering wide range of products with many promotional schemes and pricing policies, which makes it difficult not only for the customers to find the right product at the right price, but also for a supplier to choose a best site for selling its product. Thus, selection of B2C e-commerce websites is crucial and it is important for both customers and company. The main focus of this paper is to select best B2C e-commerce website on the basis of our extended B2C e-commerce success model with seven success criteria. As, this problem is multicriteria decision making problem, therefore, we use MCDM techniques, namely OWA operator, ME-OWA, and fuzzy AHP. The detailed methodology of these techniques is explained in the next section.

III. METHODOLOGY

The methodology used in this paper is described as follows:

(1) *The OWA*: The ordered weighted averaging aggregation operator (OWA) is a part of multi-criteria family, and it is given by Yager [10]. The main purpose of using OWA operators is to take full benefits of the available information and to remove the overestimation of decision maker's priorities. The ordered weighted averaging operators used in this paper are as follows:

Definition 1. An OWA operator of dimension n is defined as mapping F

$$F: Z^n \rightarrow Z \text{ (where } Z = [0,1])$$

with weight vector $(W) = (w_1, w_2, \dots, w_n)^T$.

such that

$$I. \quad w_i \in (0,1) \quad \forall i = 1,2, \dots, n.$$

$$II. \quad w_1 + w_2 + \dots + w_n = 1$$

and where

$$F(a_1, a_2, \dots, a_n) = w_1 b_1 + w_2 b_2 + \dots + w_n b_n = \sum_{i=1}^n w_i b_i = WB \quad (1)$$

where B be the ordered argument vector given by $B = (b_1, b_2, \dots, b_n)$ and b_i is the i^{th} largest element in the set $\{a_1, a_2, \dots, a_n\}$, which are arranged in descending order (*i.e.*, $b_1 \geq b_2 \geq \dots \geq b_n$) before calculating the scalar product with weight vector W .

Definition 2. Let F be an OWA operator and W be the weight vector given by $W = [w_1, w_2, \dots, w_n]^T$, then the degree of "Orness" associated with F is defined as:

$$orness(W) = \frac{1}{(n-1)} \sum_{i=1}^n (n-i) w_i = \beta \quad (2)$$

Here $\beta \in [0,1]$. If the value of "*orness*(W)" *i.e.* β is close to one then relationship between multiple attributes shows higher *orlike* level. *i.e.* $0.5 \leq \beta \leq 1$, while if it is close to zero then higher *andlike* level. *i.e.* $andness(W) = 1 - \beta$. Oring the multiple attributes means the decision maker is maximally optimistic but anding the multiple attributes means the decision maker is maximally noncommittal. Thus, as we move weight up, we increase the *orness*, while moving weight down causes decrease in the *orness*. Specifically, if $w_i = 1/n \quad \forall i = 1,2, \dots, n$, then the value of $orness(W) = 0.5$. In that case, the decision maker has a medium assessment.

Definition 3. For a given weight vector $W = (w_1, w_2, \dots, w_n)^T$, the measure of *entropy* (or *dispersion*) is defined as:

$$dispersion(W) = - \sum_{i=1}^n w_i \ln w_i \quad (3)$$

In particular, if $w_i = 1$ for some i , and $w_j = 0 \quad (\forall j \neq i)$, then the *dispersion* of weight vector W is minimum *i.e.* $dispersion(W) = 0$, it means that only one attribute is considered in the aggregation process and if $w_i = 1/n \quad \forall i = 1,2, \dots, n$; then *dispersion* of weight vector W is maximum *i.e.* $dispersion(W) = \ln n$. It means that all the attributes are considered in the aggregation process.

(2) *The ME-OWA*: The ME-OWA operator was proposed by O'Hagan [11], it maximizes the entropy of weight vector (W) subject to orness of weight vector (W). The mathematical formulation for this problem is given as follows:

$$\begin{aligned} &\text{maximize } -\sum_{i=1}^n w_i \ln w_i \\ &\text{s. t. } \frac{1}{(n-1)} \sum_{i=1}^n (n-i) w_i = \beta, \quad 0 \leq \beta \leq 1 \quad (4a) \\ &\quad \sum_{i=1}^n w_i = 1, \quad 0 \leq w_i \leq 1, \quad i = 1, \dots, n. \quad (4b) \end{aligned}$$

Fullér and Majlender used Lagrange multiplier method for determining the optimum weight vectors with respect to maximum entropy [12]. The associated weight $w_i, i = 1, 2, \dots, n$ corresponding to multiple attributes are given in Eqs.(5a)-(5c):

$$\begin{aligned} \ln w_j &= \frac{j-1}{n-1} \ln w_n + \frac{n-j}{n-1} \ln w_1 \\ \text{or } w_j &= \sqrt[n-1]{w_1^{n-j} w_n^{j-1}}, \text{ for } 1 \leq j \leq n \quad (5a) \end{aligned}$$

where n is the total number of attributes, and

$$w_n = \frac{((n-1)\alpha - n)w_1 + 1}{(n-1)\alpha + 1 - nw_1} \quad (5b)$$

$$\text{then } w_1[(n-1)\alpha + 1 - nw_1]^n = ((n-1)\alpha)^{n-1} [((n-1)\alpha - n)w_1 + 1] \quad (5c)$$

where the optimum value of w_1 should satisfy Eq. (4b) and is calculated using Eq. (5c), and then the optimal value of w_n is obtained from Eq. (5b). Remaining weights are determined by Eq. (5a). For illustration, the weight vectors of ME-OWA operator for $n = 7$ are given in Table 1.

Table 1. The weight vectors (w_i) of ME-OWA operator for $n = 7$.

	β					1
	0.5	0.6	0.7	0.8	0.9	
w_1	0.1429	0.2158	0.3096	0.4353	0.6227	1
w_2	0.1429	0.1852	0.2236	0.2497	0.2354	0
w_3	0.1429	0.1589	0.1614	0.1433	0.0889	0
w_4	0.1429	0.1364	0.1165	0.0821	0.0336	0
w_5	0.1428	0.1171	0.0842	0.0471	0.0127	0
w_6	0.1428	0.1004	0.0608	0.0270	0.0049	0
w_7	0.1428	0.0862	0.0439	0.0155	0.0018	0

(3) *The fuzzy set theory:* The fuzzy set theory was proposed by Zadeh to tackle the vagueness and uncertainty in decision making for enhancing precision [13]. So, the vague data may be represented by using fuzzy numbers which may be further subjected to mathematical functions in fuzzy domain. The fuzzy numbers are represented by membership function $\mu_{\tilde{M}}(x)$ ranging between 0 and 1. A triangular fuzzy numbers (TFNs) is shown in Fig. 1.

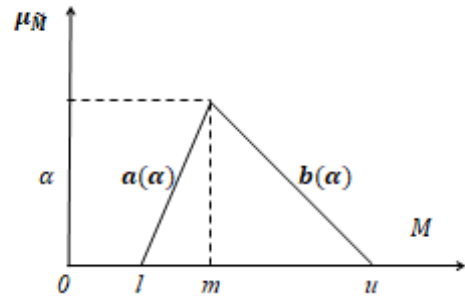


Fig. 1: Triangular Fuzzy Number (TFN)

A TFN is represented by (l, m, u) , where $l, m,$ and u are the smallest, most promising, and largest possible values respectively. The linear representation of TFN can be shown a membership function.

$$\mu_{\tilde{M}}(x) = \begin{cases} 0, & x < l \\ \frac{x-l}{m-l}, & l \leq x \leq m \\ \frac{\mu-x}{\mu-m}, & m \leq x \leq \mu \\ 0, & x > \mu \end{cases}$$

The membership function for a TFN with linear representation of left and right side as is given by: $a(\alpha) = l + \alpha(m - l); b(\alpha) = u - \alpha(\mu - m); \alpha \in [0, 1]$ where $a(\alpha)$ and $b(\alpha)$ are left and right side representation of a TFN respectively.

(4) *The Fuzzy AHP:* The analytical hierarchical process (AHP) is a decision making technique, which is used to tackle complex multi-criteria problems involving both qualitative and quantitative decisions variables [14]. It is used for determining the priorities among the different criteria, alternative comparison with respect to each criterion, and determining the preference for each alternative. The detailed procedure is as follows: firstly, decision makers have to construct a hierarchal structure for their problem in which goal, criteria and alternative phases, are placed at different levels. After the hierarchical structure, the pair wise comparison matrix will be constructed through decision makers' judgment on the importance of each criterion. The importance is measured using Saaty's fundamental scale [14]. Next, we check the consistency of the decision matrices and then calculate the relative importance weights for criteria at each level of hirerachy using pair-wise comparison matrices. The final result is on relative basis by comparing the importance of one alternative to another.

The AHP is usually used to judge intangibles, but it does not fully include the importance of qualitative aspects because its fundamental scale cannot indicate the humans' way of thinking. According to Kwong and Bai, experts' judgments are affected by imprecision and uncertainty. Thus, the use of precise and definite numbers as a linguistic judgment is not much logical [15].

To overcome this difficulty, fuzzy AHP technique is used for

optimal selection problem. Fuzzy AHP approach is the combination of triangular fuzzy number (TFNs) and analytical hierarchy process (AHP). It converts the linguistic judgments into the organized triangular fuzzy numbers (TFNs) for constructing fuzzy pair-wise comparisons matrices. These pair-wise matrices are used to get the weights of criterion and to rank the alternatives. Although there are a number of methods to tackle pair-wise matrices in fuzzy AHP such as: Buckley [16], Chang [17], Lee [18] etc. Out of all these, mostly used method is extent analysis method given by Chang [17]. This method is simple and effective to calculate the relative weights of the criteria and alternatives. In this paper, we have used Chang's extent analysis method to get the crisp weights of criteria and alternatives. The conversion scale, used for converting the linguistic judgments into the TFNs [19], is given in Table 2.

Table 2: Triangular Fuzzy Comparison scale

Linguistic scale	Triangular Fuzzy scale	Triangular fuzzy Reciprocal scale
Just equal	(1,1,1)	(1,1,1)
Equally important	(1/2,1,3/2)	(2/3,1,2)
Weakly important	(1,3/2,2)	(1/2,2/3,1)
Strongly more important	(3/2,2,5/2)	(2/5,1/2,2/3)
Very strongly more important	(2,5/2,3)	(1/3,2/5,1/2)
Absolutely more important	(5/2,3,7/2)	(2/7,1/3,2/5)

(5) *Chang's extent analysis method:*

The outlines of this method for fuzzy AHP are as follows: Let the object set be $X = \{x_1, x_2, \dots, x_n\}$ and the goal set be $U = \{u_1, u_2, \dots, u_n\}$. Each object is taken and extent analysis for each goal g_i is performed respectively. Thus, m extent analysis values for each object can be obtained using the following signs:

$$M_{g_i}^1, M_{g_i}^2, \dots, M_{g_i}^k, \quad i = 1, 2, \dots, n \quad (6)$$

where all the $M_{g_i}^j (j = 1, 2, \dots, k)$ are triangular fuzzy numbers (TFNs) with values l , m , and u .

The steps for calculating the relative weights of criteria and alternatives using fuzzy AHP through Chang's extent analysis are as follows:

Step 1: The fuzzy pairwise comparison matrix $A = (a_{ij})_{n \times k}$ is constructed, where a_{ij} is the relative importance of i^{th} entry over j^{th} entry in pairwise comparison using triangular fuzzy numbers (TFNs), whose parameters are l_{ij} , m_{ij} , and u_{ij} . They are the least, the most, and the largest possible value, respectively, i.e. $a_{ij} = (l_{ij}, m_{ij}, u_{ij})$ and a_{ij} satisfies with following

$$l_{ij} = \frac{1}{l_{ji}}, m_{ij} = \frac{1}{m_{ji}}, u_{ij} = \frac{1}{u_{ji}}.$$

Step 2: The fuzzy synthetic extent's value respect to i^{th} object is described as:

$$S_i = \sum_{j=1}^k M_{g_i}^j \otimes \left[\sum_{i=1}^n \sum_{j=1}^k M_{g_i}^j \right]^{-1} \quad (7)$$

For calculating $\sum_{j=1}^k M_{g_i}^j$, the fuzzy addition method of m extent analysis values for an individual matrix such that:

$$\sum_{j=1}^k M_{g_i}^j = (\sum_{j=1}^k l_j, \sum_{j=1}^k m_j, \sum_{j=1}^k u_j), \quad i = 1, 2, \dots, n \quad (8)$$

and to obtain $[\sum_{i=1}^n \sum_{j=1}^k M_{g_i}^j]^{-1}$, perform the fuzzy addition operation of $M_{g_i}^j (j = 1, 2, \dots, k)$ values such that

$$\sum_{i=1}^n \sum_{j=1}^k M_{g_i}^j = (\sum_{i=1}^n l_i, \sum_{i=1}^n m_i, \sum_{i=1}^n u_i) \quad (9)$$

And then calculate the inverse's vector in Eq. (4) as follows:

$$[\sum_{i=1}^n \sum_{j=1}^k M_{g_i}^j]^{-1} = \left(\frac{1}{\sum_{i=1}^n u_i}, \frac{1}{\sum_{j=1}^k m_j}, \frac{1}{\sum_{j=1}^k l_j} \right) \quad (10)$$

Step 3: If $M_1 = (l_1, m_1, u_1)$ and $M_2 = (l_2, m_2, u_2)$ then the degree of possibility of $M_2 \geq M_1$ is described as:

$$V(M_2 \geq M_1) = \sup_{y \geq x} [\min(\mu_{M_1}(x), \mu_{M_2}(y))] \quad (11)$$

and can be equivalently expressed as follows:

$$V(M_2 \geq M_1) = \begin{cases} 1, & \text{if } m_2 \geq m_1 \\ 0, & \text{if } l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & \text{otherwise} \end{cases} \quad (12)$$

where d is the ordinate of the highest intersection point D between μ_{M_1} and μ_{M_2} as shown in Fig. 2.

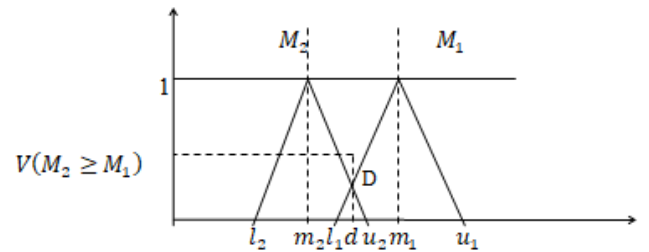


Fig. 2: Intersection of M_1 and M_2

Here, we require the values of $V(M_1 \geq M_2)$ and $V(M_2 \geq M_1)$ for comparing M_1 and M_2 .

Step 4: The degree of possibility for a convex fuzzy number to be greater than k convex fuzzy numbers $M_i (i = 1, 2, \dots, q)$ can be defined as:

$$V(M \geq M_1, M_2, \dots, M_q) = V[(M \geq M_1) \text{ and } (M \geq M_2) \text{ and } \dots \text{ and } (M \geq M_q)] = \min V(M \geq M_q), \quad i = 1, 2, \dots, q. \quad (13)$$

Now, assume that

$$d'(A_i) = \min V(S_i \geq S_q) \quad (14)$$

For $k = 1, 2, \dots, n; q \neq i$. Then the weight vector is given by

$$W' = (d'(A_1), d'(A_2), \dots, \dots, d'(A_n))^T$$

Step 5: After normalized, is given as follows:

$$W = (d(A_1), d(A_2), \dots, \dots, d(A_n))^T \tag{16}$$

where W is non-fuzzy number.

The triangular fuzzy conversion scale used in this paper is shown in Table 2. In the next section, we present a numerical application of the integrated fuzzy AHP with ME-OWA, framework.

IV. NUMERICAL ILLUSTRATION

Here objective is to select the best B2C e-commerce websites on the basis of B2C e-commerce success criteria. B2C e-commerce experts were invited to obtain the fuzzy pair-wise matrices, which are required to calculate the priority weights for each criterion as described. The application is based on the following steps:

Step 1: Construct the hierarchy for the problem

The hierarchy of the problem is demonstrated in Fig. 3. The hierarchy has three levels. The goal of the selection process defined as Best B2C e-commerce website at the top level, seven criteria at the second level, and five alternatives of B2C e-commerce websites are at the lowest level in the hierarchy.

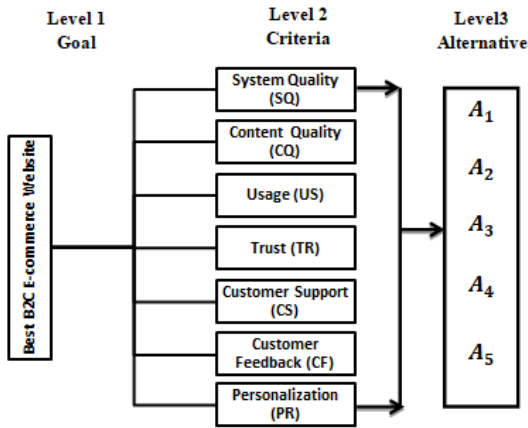


Fig 3: The hierarchy of the problem

Step 2: Construct the fuzzy pairwise comparison matrix M for criteria

The fuzzy pairwise comparison matrices are obtained from B2C e-commerce experts. Using the method described in Section 3.5, the following pair-wise comparison matrices and weight vectors for alternatives respect to each criterion are obtained:

$$M_{SQ} = \begin{bmatrix} (1,1,1) & (1,3/2,2) & (1/2,1,3/2) & (1,3/2,2) & (1,3/2,2) \\ (1/2,2/3,1) & (1,1,1) & (1/2,1,3/2) & (1,3/2,2) & (1,3/2,2) \\ (2/3,1,2) & (2/3,1,2) & (1,1,1) & (1,3/2,2) & (3/2,2,5/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (1/2,1,3/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (2/5,1/2,2/3) & (2/3,1,2) & (1,1,1) \end{bmatrix}$$

$$W_{SQ} = (0.2458, 0.2185, 0.2458, 0.1397, 0.1502)$$

$$M_{CQ} = \begin{bmatrix} (1,1,1) & (1,3/2,2) & (1/2,1,3/2) & (1,3/2,2) & (2,5/2,3) \\ (1/2,2/3,1) & (1,1,1) & (1/2,1,3/2) & (1,3/2,2) & (3/2,2,5/2) \\ (2/3,1,2) & (2/3,1,2) & (1,1,1) & (1,3/2,2) & (3/2,2,5/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (2,5/2,3) \\ (1/3,2/5,1/2) & (2/5,1/2,2/3) & (2/5,1/2,2/3) & (1/3,2/5,1/2) & (1,1,1) \end{bmatrix}$$

$$W_{CQ} = (0.2932, 0.2412, 0.2611, 0.1972, 0.0074)$$

$$M_{US} = \begin{bmatrix} (1,1,1) & (1/2,1,3/2) & (3/2,2,5/2) & (2,5/2,3) & (5/2,3,7/2) \\ (2/3,1,2) & (1,1,1) & (1/2,1,3/2) & (3/2,2,5/2) & (2,5/2,3) \\ (2/5,1/2,2/3) & (2/3,1,2) & (1,1,1) & (1,3/2,2) & (1/2,1,3/2) \\ (1/3,2/5,1/2) & (2/5,1/2,2/3) & (1/2,2/3,1) & (1,1,1) & (2,5/2,3) \\ (2/7,1/3,2/5) & (1/3,2/5,1/2) & (2/3,1,2) & (1/3,2/5,1/2) & (1,1,1) \end{bmatrix}$$

$$W_{US} = (0.4287, 0.3333, 0.1835, 0.0546, 0)$$

$$M_{TR} = \begin{bmatrix} (1,1,1) & (1/2,1,3/2) & (1,3/2,2) & (1,3/2,2) & (5/2,3,7/2) \\ (2/3,1,2) & (1,1,1) & (1,3/2,2) & (1,3/2,2) & (3/2,2,5/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (1,3/2,2) & (1,3/2,2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (5/2,3,7/2) \\ (2/7,1/3,2/5) & (2/5,1/2,2/3) & (1/2,2/3,1) & (2/7,1/3,2/5) & (1,1,1) \end{bmatrix}$$

$$W_{TR} = (0.3152, 0.2780, 0.1942, 0.2126, 0)$$

$$M_{CS} = \begin{bmatrix} (1,1,1) & (1/2,1,3/2) & (1,3/2,2) & (1,3/2,2) & (1,3/2,2) \\ (2/3,1,2) & (1,1,1) & (1,3/2,2) & (1,3/2,2) & (1,3/2,2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (1,3/2,2) & (1/2,1,3/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (5/2,3,7/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (2/3,1,2) & (2/7,1/3,2/5) & (1,1,1) \end{bmatrix}$$

$$W_{CS} = (0.2370, 0.2370, 0.1789, 0.2134, 0.1337)$$

$$M_{CF} = \begin{bmatrix} (1,1,1) & (5/2,3,7/2) & (2,5/2,3) & (1,3/2,2) & (3/2,2,5/2) \\ (2/7,1/3,2/5) & (1,1,1) & (1/2,1,3/2) & (1,3/2,2) & (1,3/2,2) \\ (1/3,2/5,1/2) & (2/3,1,2) & (1,1,1) & (1,3/2,2) & (1/2,1,3/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (2,5/2,3) \\ (2/5,1/2,2/3) & (1/2,2/3,1) & (2/3,1,2) & (1/3,2/5,1/2) & (1,1,1) \end{bmatrix}$$

$$W_{CF} = (0.5387, 0.1969, 0.0463, 0.1823, 0.0358)$$

$$M_{PR} = \begin{bmatrix} (1,1,1) & (1,3/2,2) & (1/2,1,3/2) & (1,3/2,2) & (3/2,2,5/2) \\ (1/2,2/3,1) & (1,1,1) & (3/2,2,5/2) & (1,3/2,2) & (1,3/2,2) \\ (2/3,1,2) & (2/5,1/2,2/3) & (1,1,1) & (1,3/2,2) & (1/2,1,3/2) \\ (1/2,2/3,1) & (1/2,2/3,1) & (1/2,2/3,1) & (1,1,1) & (5/2,3,7/2) \\ (2/5,1/2,2/3) & (1/2,2/3,1) & (2/3,1,2) & (2/7,1/3,2/5) & (1,1,1) \end{bmatrix}$$

$$W_{PR} = (0.2987, 0.2857, 0.0278, 0.2470, 0.1490)$$

Step 3: Calculate and check the consistency of pair wise comparison matrices

For checking the consistency of all the fuzzy pairwise comparison matrices, we convert fuzzy matrices into crisp comparison matrices through centroid defuzzification technique called center of gravity [10]. Formula for converting the TFNs is (Wang & Elhag, 2007):

$$a_{ij} = (l_{ij} + m_{ij} + u_{ij})/3, \text{ where } a_{ij} = (l_{ij}, m_{ij}, u_{ij}) \tag{17}$$

Consistency index (CI) and consistency ratio (CR) need to be calculated to check the consistency of the decision matrices. Here CI and CR are given by:

$$CI = (\lambda_{max} - n)/(n - 1) \tag{18}$$

$$CR = CI/RI \tag{19}$$

where λ_{max} is the largest eigenvalue of the pairwise matrix, n is the dimension of the matrix. The random Index (RI) value proposed by Satty [14] for the different number of criterion (n) is shown in Table 3.

Table 3: Relationship between random index (RI) and number of criteria (n)

N	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45

Information related to the consistency of each fuzzy pairwise matrix is shown in Table 4. From Table 4, we can observe that information given by fuzzy pairwise comparison matrix is satisfactory and meet the basic requirements of consistency (as $CR < 0.1$).

Table 4: Weights vectors for multiple criteria with λ_{max} , CI and CR

	SQ	CQ	US	TR	CS	CF	PR
A_1	0.2458	0.2932	0.4287	0.3152	0.2370	0.5387	0.2987
A_2	0.2185	0.2412	0.3333	0.2780	0.2370	0.1969	0.2857
A_3	0.2458	0.2611	0.1835	0.1942	0.1789	0.0463	0.0278
A_4	0.1397	0.1972	0.0546	0.2126	0.2134	0.1823	0.2470
A_5	0.1502	0.0074	0.0000	0.0000	0.1337	0.0358	0.1409
λ_{max}	5.1025	5.1045	5.2734	5.1439	5.2786	5.3081	5.3376
CI	0.0256	0.0261	0.0684	0.0360	0.0696	0.0770	0.0844
CR	0.0229	0.0233	0.0610	0.0321	0.0622	0.0688	0.0754

Step 4: Specify $\beta = orness$ according to Table 1, and then obtain the OWA weight vector $W = (w_1, w_2, \dots, w_n)$

According to *definition 2*, the value of β i.e. *orness* in B2C e-commerce website selection is determined on the basis of B2C e-commerce experts' uncertain preferences. If the expert is highly optimistic on the selection of B2C e-commerce websites, then the value of β (*orness*) is set to be 1. And if the expert is in the moderate state, then the value of β (*orness*) is set to be 0.5. The weight vector $W = (w_1, w_2, \dots, w_n)$ of ME-OWA for a specific value of β is given in Table 1. These values are used for aggregating the information of multiple B2C e-commerce success factors as attributes for B2C e-commerce websites. In this problem, the β (*orness*) is set to 0.8 because the experts' preference on the selection of B2C e-commerce websites are moderately optimistic. From Table 1, we get the weight vector as

$$W = (w_1, w_2, w_3, w_4, w_5, w_6, w_7)$$

$$W = (0.4353, 0.2497, 0.1433, 0.0821, 0.0471, 0.027, 0.0155)$$

In the next step, we calculate the OWA operators $F_i \forall i = 1, 2, \dots, n$, with respect to each alternative.

Step 5: Calculate the OWA operator F for alternatives

The maximum entropy-ordered weighted averaging aggregation (ME-OWA) operator of A_1 as F_1 is calculated as follows:

$$\begin{aligned} F_1(SQ, CQ, US, TR, CS, CF, PR) &= \sum_{i=1}^7 w_i b_i \\ &= 0.4353 \times 0.5387 + 0.2497 \times 0.4287 + 0.1433 \times 0.3152 \\ &\quad + 0.0821 \times 0.2987 + 0.0471 \times 0.2932 + 0.0270 \times 0.2458 \\ &\quad + 0.0155 \times 0.2370 = 0.4354 \end{aligned}$$

where b_i is the i^{th} largest element in the set $\{SQ, CQ, US, TR, CS, CF, PR\}$.

Similarly, the OWA operators for A_2, A_3, A_4 , and A_5 as $F_2 = 0.2962, F_3 = 0.2280, F_4 = 0.2207$, and $F_5 = 0.1230$.

Table 5: The ranking order of alternatives

Alternative	Priority	The ranking order
A_1	0.4354	1
A_2	0.2962	2
A_3	0.2280	3
A_4	0.2207	4
A_5	0.1230	5

Now from Table 5, we can observe that F_1 is dominating all the F_i 's. Therefore, A_1 is the best selection over other alternative present in the Table 5.

V. CONCLUSION

E-commerce is all about doing business electronically. In the era of internet technology, brick and mortar stores are fast being replaced by online stores, which have no geographical boundaries, save time and money, and provide an edge over the market competitors. Although e-commerce has many types but the major share of economy is encapsulated by the business-to-consumers (B2C) e-commerce. The B2C e-commerce success depends on a number of factors related with website quality and customer satisfaction.

In this paper, we present and extended e-commerce success model with includes customer oriented success criteria such as system quality (SQ), content quality (CQ), Usage (US), Trust (TR), customer support (CS), customer feedback (CF) and personalization (PR). These criteria have been used to rank a number of e-commerce websites using an integrated FAHP and ME-OWA approach. The framework used in this paper helps in capturing the uncertainty in the experts' opinion and preferences. The outcome of the paper gives the direction to the B2C e-commerce customers for finding right product at right price as well as for companies to choose the best site for selling its product on the web. In the future research, we may extend our framework for optimal online advertising planning and optimal budget allocation problem.

REFERENCES

- [1] Cyr, D. and C. Bonanni. Design and e-loyalty across cultures in electronic commerce. in Proceedings of the 6th international conference on Electronic commerce. ACM, (2004).
- [2] DeLone, W.H. and E.R. McLean, Information systems success: The quest for the dependent variable. Information systems research. 3(1), pp. 60-95. (1992).
- [3] Molla, A. and P.S. Licker, E-commerce systems success: An attempt to extend and respecify the DeLone and MacLean model of IS success. J. Electron. Commerce Res. 2(4), pp. 131-141. (2001).
- [4] DeLone, W.H. and E.R. McLean. Information systems success revisited. in System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on. IEEE, (2002).
- [5] DeLone, W.H. and E.R. McLean, The DeLone and McLean model of information systems success: a ten-year update. Journal of management information systems. 19(4), pp. 9-30. (2003).
- [6] DeLone, W.H. and E.R. Mclean, Measuring e-commerce success: Applying the DeLone & McLean information systems success model. International Journal of Electronic Commerce. 9(1), pp. 31-47. (2004).
- [7] Chevalier, J.A. and D. Mayzlin, The effect of word of mouth on sales: Online book reviews. Journal of marketing research. 43(3), pp. 345-354. (2006).

- [8] Ghose, A. and P.G. Ipeiritis, Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*. 23(10), pp. 1498-1512. (2011).
- [9] Andreou, A.S., et al., Key issues for the design and development of mobile commerce services and applications. *International Journal of Mobile Communications*. 3(3), pp. 303-323. (2005).
- [10] Yager, R.R., On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on systems, Man, and Cybernetics*. 18(1), pp. 183-190. (1988).
- [11] O'Hagan, M. Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. in *Signals, Systems and Computers*, 1988. Twenty-Second Asilomar Conference on. IEEE, (1988).
- [12] Fullér, R. and P. Majlender, An analytic approach for obtaining maximal entropy OWA operator weights. *Fuzzy Sets and Systems*. 124 (1), pp. 53-57. (2001).
- [13] Zadeh, L.A., Fuzzy sets. *Information and control*. 8(3), pp. 338-353. (1965).
- [14] Saaty, T.L., *The Analytic Hierarchy Process*, NY. McGraw-Hill, USA. Cook WD and Seiford LM.(1978). Priority ranking and consensus formation, *Management Science*.24, pp. 1721-1732. (1980).
- [15] Kwong, C.-K. and H. Bai, Determining the importance weights for the customer requirements in QFD using a fuzzy AHP with an extent analysis approach. *Iie Transactions*. 35(7), pp. 619-626. (2003).
- [16] Buckley, J.J., Fuzzy hierarchical analysis. *Fuzzy sets and systems*. 17(3), pp. 233-247. (1985).
- [17] Chang, D.-Y., Applications of the extent analysis method on fuzzy AHP. *European journal of operational research*. 95(3), pp. 649-655. (1996).
- [18] Lee, S.-H., Using fuzzy AHP to develop intellectual capital evaluation model for assessing their performance contribution in a university. *Expert systems with applications*. 37(7), pp. 4941-4947. (2010).
- [19] Bozbura, F.T., A. Beskese, and C. Kahraman, Prioritization of human capital measurement indicators using fuzzy AHP. *Expert Systems with Applications*. 32(4), pp. 1100-1112. (2007).

An Intuitionistic Approach for Ranking OTA Websites under Multi Criteria Group Decision Making Framework

Anu G. Aggarwal¹, Himanshu Sharma², Abhishek Tandon³

¹ Department of Operational Research, University of Delhi, Delhi, India

² Department of Operational Research, University of Delhi, Delhi, India

³ Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi, India

¹anuagg17@gmail.com, ²himanshusharma.du.or@gmail.com, ³abhishektandon@sscbsdu.ac.in

Abstract—The transformations from approaches based on crisp set towards fuzzy set were introduced to include the uncertainty experienced in decision making. But the problem of hesitation about any alternative still prevailed among the decision makers. Later, an extension of fuzzy sets, the intuitionistic fuzzy sets, filled the gap between the degree of satisfaction and the degree of dissatisfaction encountered by the decision makers. Also, the biasness noticed while comparing a number of alternatives based on multiple criteria by a single person, necessitated for a multi criteria group decision making (MCGDM) approach. In this paper, we propose a hybrid model that combines the MCGDM techniques of intuitionistic fuzzy analytical hierarch process (IFAHP) and intuitionistic fuzzy multi objective optimization on the basis of ratio analysis (IFMOORA). The IFAHP technique is used to find the criteria weights and IFMOORA technique is used for overall ranking of OTA websites. To validate the proposed model, we apply it to rank online travel agencies (OTA) websites on the basis of a number of success factors. The model developed is flexible and can be further applied to various selection problems.

Keyword—Intuitionistic fuzzy sets, Intuitionistic fuzzy MOORA, Intuitionistic fuzzy AHP, Online travel agency

I. INTRODUCTION

THESE days, the decision making process has shifted from a single criterion to multiple conflicting criteria. This has popularized the use of multi criteria decision making (MCDM) approaches by researchers in various extant studies. Earlier, these studies were based on the crisp or exact inputs from the decision makers or experts, but it failed to incorporate the uncertainty and ambiguity faced by decision makers. This loophole was corrected by extending the decision making inputs based on crisp sets to fuzzy sets [19]. Fuzzy sets are efficient in handling the vagueness and impreciseness faced by experts in decision making, by specifying some degree of membership to each alternative. Still, the techniques based on fuzzy sets were not able to consider the degree of hesitation encountered by the decision makers. This necessitated for extending these sets to intuitionistic fuzzy sets (IFS) introduced by Atanassov [2]. This enabled the decision maker to include both his preferences and non-preferences in decision making simultaneously. MCDM approaches evaluate each alternative over all criteria based on the inputs of a single expert, which is a biased

approach. Thus, multi criteria group decision making (MCGDM) approaches were introduced to obtain the best solution based on information provided by multiple decision makers. It has been noticed that even though the decision makers may have same objective, their perception towards the problem may vary [13]. Earlier each decision maker was allocated equal weights, but later weighing them according to their expertise was found to be more consistent approach [15]. For calculation purpose, the inputs from each decision maker are combined using various aggregation operators, to get an aggregated intuitionistic fuzzy decision matrix.

Analytic Hierarchy Process (AHP) is a flexible and intuitive MCDM technique developed by Saaty [14]. The method transforms a complex problem into a hierarchical structure, which makes the problem organized and easy to analyze. It makes use of a pair-wise comparison scale to compare objects and find consistency of the problem. However, AHP is mainly used for evaluation purposes and not for ranking as the latter may demand for a large number of pair-wise comparisons and also via additive aggregation it may lose vital information [8]. In order to tackle these demerits, we use IFAHP to obtain criteria weights. Some of the previous researches which have used the technique for similar purposes are: Supplier selection [5], information system outsourcing [6], and many more.

Multi objective optimization on the basis of ratio analysis (MOORA) is another MCDM technique for optimizing two or more criteria simultaneously based on certain constraints. This method segregates all the criteria into beneficial (which are to be maximized) and non-beneficial (which are to be minimized), which are then used to rank alternatives or selecting the best from the available options. A distinguishing feature of MOORA as compared to its outranking counterparts is the use of ratio system for normalization procedure. Thus, it is considered to be a non-subjective method and more robust than existing MCDM techniques as it make use of cardinal or most recent data. In this paper we use IFMOORA for outranking purpose. The use of IFMOORA for ranking purpose is a new concept and till date it has been used previously for selection problem [17].

The growth in internet technology has led to an increase in purchase of goods and services by making use

of e-commerce websites. The service sector has boomed with these advancements, especially, the online travel agency (OTA) websites usage has shown an increasing trend over the years. OTAs provide the customers a single platform for searching and booking airline tickets and hotel rooms through multiple vendors [9]. Though, much of the studies conducted in hospitality sector are related to hotels, less research has been conducted on OTAs. The success of an e-commerce website depends on the experience of previous visitors and the word-of-mouth spread by them. Here, we consider the following criteria: content quality, trust, support, service quality, customization [11] [12]. We add another criterion ‘electronic word-of-mouth’, which is similar to the concept of ‘word-of-mouth’ but extended over the internet. The word-of-mouth is a social communication, which acts as a powerful tool in the spread of information about the products or services, by previous purchasers. They have a high impact on potential customers as people believe others rather than the words communicated through the marketers [7].

In this paper, we combines the AHP and MOORA techniques extended over IFS under group decision making, to rank online travel agency (OTA) websites. Some of the earlier studies which have IFS for evaluation or ranking purposes are given by Table 1. The rest of the paper is structured as follows: Section 2 discusses the IFS based methods. The model is proposed in Section 3. Section 4 discusses a numerical illustration to validate the proposed model. The conclusion and discussions are given in Section 5.

Table 1: List of Previous Researches

Techniques	Authors	Purpose of the Study
IFGDM-Averaging operators	[16]	Evaluation of agro ecological regions
IFGDM-TOPSIS	[3]	Supplier selection
IFGDM-VIKOR	[6]	Information system outsourcing
IVIFGDM-Extended TOPSIS	[18]	Partner selection
IFGDM-TOPSIS	[1]	Machine Selection
IFAHP-IFMOORA	[17]	Project selection
IFAHP-IFAD	[5]	Supplier selection
IFGDM-TOPSIS	[4]	Hazardous waste management

II. PRELIMINARIES

A. Intuitionistic Fuzzy Set Theory

Here, we discuss some of the concepts for working on intuitionistic sets which will be used throughout the paper.

Definition 1: (Intuitionistic fuzzy set) Let Y be a crisp set and let $I \subset Y$ be a fixed set. Then the IFS \tilde{I} in Y is defined as,

$$\tilde{I} = \{(y, \mu_I(y), \vartheta_I(y)) | y \in Y\} \quad (1)$$

where $\mu_I: Y \rightarrow [0,1]$ and $\vartheta_I: Y \rightarrow [0,1]$. $\mu_I(y)$ indicate the membership degree and $\vartheta_I(y)$ the non-membership degree of $y \in Y$ in I , respectively. Also, $0 \leq \mu_I(y) + \vartheta_I(y) \leq 1$, for all $y \in Y$. The hesitancy degree of y is denoted by $\pi_I(y)$, and calculated as $\pi_I(y) = 1 - \mu_I(y) - \vartheta_I(y)$.

Definition 2: An intuitionistic preference relation K on the set $Y = \{y_1, y_2, \dots, y_n\}$ is represented by a matrix $(k_{is})_{n \times n}$, where

$k_{is} = \langle (y_i, y_s), \mu(y_i, y_s), \vartheta(y_i, y_s) \rangle \forall i, s = 1, 2, \dots, n$. Let us assume $k_{is} = (\mu_{is}, \vartheta_{is})$, where $\mu_{is} = \mu(y_i, y_s)$ and $\vartheta_{is} = \vartheta(y_i, y_s)$ denote the degree to which object y_i is preferred to the object y_s . Also, $\pi(y_i, y_s) = 1 - \mu_I(y_i, y_s) - \vartheta_I(y_i, y_s)$ indicates the hesitancy degree.

Definition 3: Intuitionistic fuzzy arithmetic operations

$$k_{is} \oplus k_{tl} = (\mu_{is} + \mu_{tl} - \mu_{is} \mu_{tl}, \vartheta_{is} \vartheta_{tl}) \quad (2)$$

$$k_{is} : k_{tl} = \left(\frac{\mu_{is}}{\mu_{tl}}, \frac{\vartheta_{is} - \vartheta_{tl}}{1 - \vartheta_{tl}} \right) \quad (3)$$

$$k_{is} \otimes k_{tl} = (\mu_{is} \mu_{tl}, \vartheta_{is} + \vartheta_{tl} - \vartheta_{is} \vartheta_{tl}) \quad (4)$$

Definition 4: Elements of a perfect multiplicative consistent intuitionistic preference relation $\hat{I} = (\hat{k}_{is})_{n \times n}$ matrix are denoted as $\hat{k}_{is} = (\hat{\mu}_{is}, \hat{\vartheta}_{is})$. For $s > i + 1$,

$$\hat{\mu}_{is} = \frac{\sqrt[s-1]{\prod_{t=i+1}^{s-1} \mu_{it} \mu_{ts}}}{\sqrt{\prod_{t=i+1}^{s-1} \mu_{it} \mu_{ts} + \prod_{t=i+1}^{s-1} (1 - \mu_{it})(1 - \mu_{ts})}} \quad (5)$$

$$\hat{\vartheta}_{is} = \frac{\sqrt[s-1]{\prod_{t=i+1}^{s-1} \vartheta_{it} \vartheta_{ts}}}{\sqrt{\prod_{t=i+1}^{s-1} \vartheta_{it} \vartheta_{ts} + \prod_{t=i+1}^{s-1} (1 - \vartheta_{it})(1 - \vartheta_{ts})}} \quad (6)$$

For $s = i + 1$ we take $\hat{k}_{is} = k_{is}$ and for $s < i$ we take $\hat{k}_{is} = (\hat{\vartheta}_{si}, \hat{\mu}_{si})$.

Definition 5: An intuitionistic preference relation R is acceptable if it satisfy the consistency condition,

$$d(I, \hat{I}) < \epsilon \quad (7)$$

where $d(I, \hat{I})$ is the distance measure between the given intuitionistic relation I and its corresponding perfect multiplicative consistent intuitionistic preference relation \hat{I} given by,

$$d(I, \hat{I}) = \frac{1}{2(n-1)(n-2)} \sum_{i=1}^n \sum_{s=1}^n (|\hat{\mu}_{is} - \mu_{is}| + |\hat{\nu}_{is} - \nu_{is}| + |\hat{\pi}_{is} - \pi_{is}|) \quad (8)$$

where ε is the consistency threshold.

Definition 6: Let the importance of g th decision maker be an IFN, $D_g = \{\mu_g, \nu_g, \pi_g\}$. The relative importance of decision maker, λ_g is given by,

$$\lambda_g = \frac{(\mu_g + \pi_g \left(\frac{\mu_g}{\mu_g + \pi_g} \right))}{\sum_{g=1}^l (\mu_g + \pi_g \left(\frac{\mu_g}{\mu_g + \pi_g} \right))} \quad (9)$$

So that $\sum_{g=1}^l \lambda_g = 1$

Definition 7: Intuitionistic fuzzy weighted averaging operator (IFWA)

$$IFWA_{\lambda}(k_{is}^g) = \left(1 - \prod_{g=1}^l (1 - \mu_{is}^g)^{\lambda_g}, \prod_{g=1}^l (\nu_{is}^g)^{\lambda_g} \right) \quad (i, s = 1, \dots, n) \quad (10)$$

Definition 8: Intuitionistic fuzzy averaging operator (IFA)

$$IFA(k_{is}) = \left(1 - \prod_{i=1}^n (1 - \mu_{is})^{\frac{1}{n}}, \prod_{i=1}^n (\nu_{is})^{\frac{1}{n}} \right) \quad (i, s = 1, \dots, n) \quad (11)$$

Detecting the object from the video is the first step to tracking an object. Detection of an interesting moving object can be achieved by different existing techniques such as frame differencing, optical flow, background subtraction, segmentation, point detectors [3] as shown in Figure 2.

B. Intuitionistic Fuzzy Analytic Hierarchy Process

The steps for performing the IFAHP technique are as follows:

Step 1: Let the set of alternatives be $H = \{H_1, H_2, \dots, H_n\}$ and criteria be $J = \{J_1, J_2, \dots, J_m\}$. The pair-wise comparison matrix of criteria and performance value matrix of alternatives are formed using IFNs given by Table 2.

Step 2: Form the group of decision makers, $D = \{D_1, D_2, \dots, D_l\}$ and determine the importance degree for each. Table 2 is used for assigning the importance weights to decision makers.

Step 3: Construct the intuitionistic preference relation matrices I_{1j}^g and I_2^g . The former indicates the performance value matrix of alternatives formed from the g th decision maker for j th criterion, and the latter presents the pair-wise comparison matrix of criteria established from g th DM. Construct a perfect multiplicative consistent intuitionistic preference relation \hat{I}_{1j}^g and \hat{I}_2^g , which are acceptable if

Table 2: Conversion of 1-9 scale to 0.1-0.9 scale

1-9 scale	0.1-0.9 scale	Linguistic Terms
1/9	0.1	Extremely not important (ENI)/beginner
1/7	0.2	Very strongly not important (VSNI)
1/5	0.3	Strongly not important (SNI)/practitioner
1/3	0.4	Moderately not important (MNI)
1	0.5	Equally important (EI)/proficient
3	0.6	Moderately important (MI)
5	0.7	Strongly important (SI)/expert
7	0.8	Very strongly important (VSI)
9	0.9	Extremely important (EXI)/master
Other values between 1/9 and 9	Other values between 0 and 1	Intermediate values used to present compromise

$d(I_{1j}^g, \hat{I}_{1j}^g) < \varepsilon$ and $d(I_2^g, \hat{I}_2^g) < \varepsilon$. Considering the conditions provided by Satty [14], the acceptable consistency threshold (ε) should be less than 0.1.

Step 4: Compute the relative importance of each decision maker using equation (9). Then, construct the aggregated intuitionistic fuzzy decision matrices \hat{I}_{1j} and \hat{I}_2 by applying IFWA operator given in equation (10).

Step 5: Construct the normalized decision matrices $\hat{I}_{1j}^{\check{}}$ and $\hat{I}_2^{\check{}}$ using the arithmetic operations given by equations (2) and (3).

Step 6: The criteria weight vector $W_2 = (W_{21}, W_{22}, \dots, W_{2m})$ is established by making use of IFA operator, given by equation (11), on the Normalized decision matrix $\hat{I}_2^{\check{}}$.

C. Intuitionistic Fuzzy Multi Objective Optimization based on Ratio Analysis

The steps of performing the IFMOORA technique are as follows:

Step 1: Construct the matrix $\check{I} = (W_{11}, W_{12}, \dots, W_{1m})$. The values of W_{1j} are computed by averaging the normalized matrix rows obtained by equation (11). The aggregated weighted intuitionistic fuzzy decision matrix \hat{R} is obtained as,

$$\hat{I} = (W_{21} W_{11}, W_{22} W_{12}, \dots, W_{2m} W_{1m}) \quad (12)$$

Arithmetic operators given by equations (2) and (4) are used for constructing \hat{I} .

Step 2: The criteria under consideration are divided into two groups as benefit (Ben_i) and cost (cos_i), and then their sum is computed for each alternative using equations,

$$Ben_i = \sum_{j=1}^q (\hat{\mu}_{is}, \hat{\nu}_{is})_j = (\hat{\mu}_{Ben_i}, \hat{\nu}_{Ben_i}) \quad i=1,2,\dots,n \quad (13)$$

$$cos_i = \sum_{j=q+1}^m (\hat{\mu}_{is}, \hat{\nu}_{is})_j = (\hat{\mu}_{cos_i}, \hat{\nu}_{cos_i}) \quad i=1,2,\dots,n \quad (14)$$

where q indicates the number of benefit criteria and remaining $m-q$ denotes the number of cost criteria

Step 3: Defuzzify the sum of benefits and costs by using below equations respectively,

$$Defuz_i(Ben) = \frac{1 - \hat{\nu}_{Ben_i}}{1 - \hat{\mu}_{Ben_i}} \quad i=1,2,\dots,n \quad (15)$$

$$Defuz_i(cos) = \frac{1 - \hat{\nu}_{cos_i}}{1 - \hat{\mu}_{cos_i}} \quad i=1,2,\dots,n \quad (16)$$

Step 4: Calculate the contribution value of each alternative using below equation and rank the alternatives in accordance to these values,

$$Contrib_i = Defuz_i(Ben) - Defuz_i(cos) \quad (17)$$

The $Contrib_i$ values may be positive or negative. The ranking of $Contrib_i$ represents the final contribution of each alternative. The highest $Contrib_i$ value represents the best alternative.

III. PROPOSED MODEL

The proposed model for ranking OTA websites on the basis of success factors is composed of two approaches namely IFAHP and IFMOORA. The phases of the model are discussed below and represented by Figure 1:

Phase 1: Construction of Hierarchy

In this phase, the OTA websites and the criteria which will be used in ranking are determined and the decision hierarchy is formed. The objective of the proposed model is placed at the top level, the success factors at the middle, and the OTA websites to be ranked at the bottom.

Phase 2: Criteria Weight Calculation using IFAHP

In this phase, decision matrix is constructed to obtain criteria weights. The elements in the matrices are filled using the pair-wise comparison values provided by the experts, obtained through a structured interview. Next, the aggregated intuitionistic fuzzy decision matrix is constructed. Then we normalize decision matrix, which in turn ultimately provide us with the final criteria weights.

Phase 3: Rank Websites using IFMOORA

Ranking websites is done using the IFMOORA technique. Firstly, the decision matrix is formed by rating each alterna-

tive against each criterion which is provided by experts in OTA industry. Then the sum of costs and benefits are obtained using aggregated weighted intuitionistic fuzzy decision matrix. This helps to evaluate the contribution value. The alternative corresponding to highest contribution value is considered to be the best.

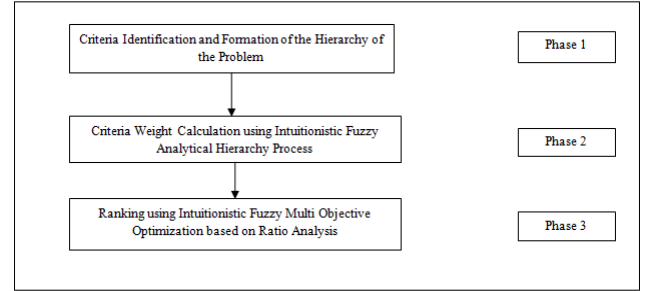


Figure 1: The Proposed Model

IV. NUMERICAL ILLUSTRATION

This study is conducted with the objective to rank OTA websites on the basis of success criteria. The numerical is modeled into three phases, as discussed in the previous section. The phases are performed as follows:

Phase 1: Construction of Hierarchy

In this phase, we convert our problem into a hierarchical structure. The goal of the problem is to find the best OTA website, which takes the highest level of the hierarchy. The success criteria or the factors form the middle level of the hierarchy and are stated as: Electronic word-of-mouth (J1), Service Quality (J2), Customization (J3), Content Quality (J4), Trust (J5), and Support (J6). The alternatives namely: H1, H2, H3, and H4 form the lowest level of the hierarchy. The structure is represented by Figure 2.

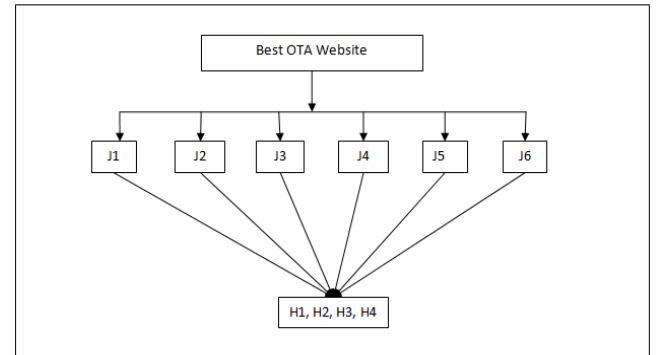


Figure 2: Hierarchy of the Problem

Phase 2: Criteria Weight Calculation using IFAHP

The set of alternatives is,

$$H = \{H1, H2, H3, H4\}$$

The set of criteria is,

$$J = \{J_1, J_2, J_3, J_4, J_5, J_6\}$$

Two decision makers are considered where the former is evaluated as an ‘expert’ and latter is evaluated as ‘proficient’. The set of decision maker is,

$$D = \{D_1, D_2\}$$

The pair-wise comparison matrices for criteria as given by D_1 and D_2 are given in Table 3.

Table 3: Pair-wise Comparison Matrix for Criteria

Criteria Weights		J1	J2	J3	J4	J5	J6
J1	D1	EI	EI	MI	EI	MI	MI
	D2	EI	EI	MI	SI	SI	MI
J2	D1	EI	EI	MI	MI	MI	MI
	D2	EI	EI	MI	MI	MI	MI
J3	D1	VSNI	VSNI	EI	MI	MI	VSI
	D2	VSNI	VSNI	EI	MI	MI	MI
J4	D1	SNI	VSNI	VSNI	EI	EI	MI
	D2	SNI	VSNI	VSNI	EI	EI	MI
J5	D1	MNI	VSNI	MNI	EI	EI	MI
	D2	SNI	VSNI	MNI	EI	EI	MI
J6	D1	VSNI	VSNI	VSNI	VSNI	VSNI	EI
	D2	VSNI	VSNI	MNI	VSNI	VSNI	EI

The perfect multiplicative consistent intuitionistic preference relation matrix is obtained for each decision maker with inputs obtained using equations (5) and (6). Using the distance formula provided by equation (8), we found that $d(I_2^1, \hat{I}_2^1) < 0.1$ and $d(I_2^2, \hat{I}_2^2) < 0.1$. This proves that the preference relation matrices are consistent.

The performance value matrices of alternatives formed from the k th decision maker for first criterion are represented by Table 4.

Table 4: Pair-wise Comparison Matrix for Alternatives over Criteria

	FOR J1	H1	H2	H3	H4
H1	D1	EI	EI	MI	MI
	D2	EI	MI	MI	MI
H2	D1	EI	EI	MI	MI
	D2	VSNI	EI	MI	MI
H3	D1	MNI	VSNI	EI	MI

	D2	VSNI	VSNI	EI	EI
H4	D1	MNI	VSNI	VSNI	EI
	D2	MNI	VSNI	EI	EI
	FOR J2	H1	H2	H3	H4
H1	D1	EI	EI	MI	MI
	D2	EI	MI	MI	MI
H2	D1	EI	EI	EI	MI
	D2	VSNI	EI	MI	MI
H3	D1	VSNI	EI	EI	EI
	D2	VSNI	VSNI	EI	MI
H4	D1	VSNI	VSNI	EI	EI
	D2	MNI	VSNI	VSNI	EI
	FOR J3	H1	H2	H3	H4
H1	D1	EI	MI	MI	MI
	D2	EI	EI	MI	MI
H2	D1	VSNI	EI	MI	MI
	D2	EI	EI	MI	MI
H3	D1	VSNI	VSNI	EI	MI
	D2	MNI	VSNI	EI	MI
H4	D1	MNI	VSNI	EI	EI
	D2	MNI	VSNI	VSNI	EI
	FOR J4	H1	H2	H3	H4
H1	D1	EI	MI	MI	MI
	D2	EI	EI	MI	MI
H2	D1	VSNI	EI	MI	MI
	D2	EI	EI	EI	MI
H3	D1	VSNI	VSNI	EI	MI
	D2	VSNI	EI	EI	EI
H4	D1	MNI	VSNI	VSNI	EI
	D2	VSNI	VSNI	EI	EI
	FOR J5	H1	H2	H3	H4
H1	D1	EI	EI	MI	MI
	D2	EI	MI	MI	MI
H2	D1	EI	EI	MI	MI
	D2	VSNI	EI	MI	MI
H3	D1	MNI	VSNI	EI	MI
	D2	VSNI	VSNI	EI	MI

H4	D1	MNI	VSNI	VSNI	EI
	D2	MNI	VSNI	VSNI	EI
	FOR J6	H1	H2	H3	H4
H1	D1	EI	EI	MI	MI
	D2	EI	MI	MI	MI
H2	D1	EI	EI	EI	MI
	D2	VSNI	EI	MI	MI
H3	D1	VSNI	EI	EI	EI
	D2	VSNI	VSNI	EI	EI
H4	D1	VSNI	VSNI	EI	EI
	D2	MNI	VSNI	EI	EI

Using the distance formula provided by equation (8), we found that $d(I_{1j}^g, \hat{I}_{1j}^g) < 0.1 \forall j=1, \dots, 6; g=1, 2$. Thus, all the matrices are consistent. Also, we have considered that $D1 = \{0.7, 0.2, 0.1\}$

$$D2 = \{0.5, 0.4, 0.1\}$$

Then, their corresponding weights are computed using equation (9) and $\lambda_1 = 0.58$ and $\lambda_2 = 0.42$.

Next we compute the aggregated intuitionistic fuzzy decision matrix using equation (10). Each value obtained is divided by summation of each criterion, and we obtain the normalized decision matrix. The criteria weights are obtained using the IFA operator given by equation (11). The weights are represented by Table 5.

Table 5: Criteria Weight Vector W_2

Criteria	Weight
J1	(0.58, 0.16, 0.26)
J2	(0.55, 0.20, 0.25)
J3	(0.46, 0.35, 0.19)
J4	(0.22, 0.64, 0.14)
J5	(0.25, 0.60, 0.15)
J6	(0.14, 0.72, 0.14)

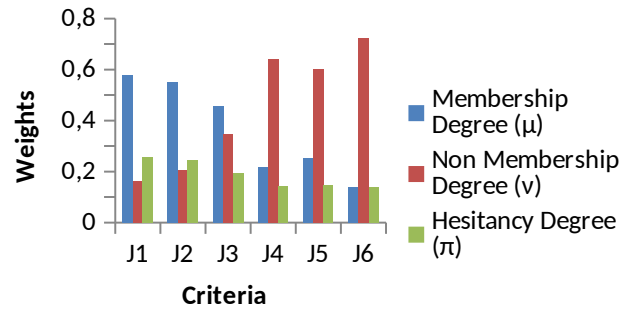


Figure 3: Criteria Weight Result

Phase 3: Rank Websites using IFMOORA

Combining the weights obtained using the decision matrices provided by $I_{1j}^g \forall j=1, \dots, 6; g=1, 2$, we compute the decision matrix evaluating each alternative on each criterion. The matrix is represented by Table 6,

Table 6: Decision Matrix

	J1	J2	J3	J4	J5	J6
H1	(0.53, 0.32, 0.15)	(0.51, 0.32, 0.17)	(0.54, 0.30, 0.16)	(0.53, 0.29, 0.18)	(0.54, 0.30, 0.16)	(0.50, 0.34, 0.16)
H2	(0.48, 0.39, 0.13)	(0.44, 0.48, 0.08)	(0.46, 0.40, 0.14)	(0.44, 0.45, 0.11)	(0.48, 0.37, 0.15)	(0.43, 0.51, 0.06)
H3	(0.32, 0.54, 0.14)	(0.40, 0.54, 0.06)	(0.31, 0.57, 0.12)	(0.38, 0.53, 0.09)	(0.34, 0.49, 0.17)	(0.25, 0.71, 0.04)
H4	(0.22, 0.66, 0.12)	(0.27, 0.63, 0.10)	(0.24, 0.65, 0.11)	(0.24, 0.65, 0.11)	(0.16, 0.68, 0.16)	(0.19, 0.73, 0.08)

Multiplying the value of Table 5 and 6, we obtain aggregated weighted intuitionistic fuzzy decision matrix.

Next, we obtain the sum of benefits and costs criteria. The electronic word-of-mouth (J1) represents how efficient the website is in providing the customers a space to give feedback and interact with others. It consists of attributes such as: online reviews and star ratings; and is considered as a benefit criterion [7]. Service quality (J2) consists of attributes: responsiveness, reliability, assurance, and empathy; that convey the level of service provided by the website, which is to be maximized. Customization (J3) represents the potential of the website to provide personalized goods and service, consisting of attributes: up-selling, cross-selling. Previous researches have shown that much of customization is not beneficial from customers' perspective, since enough of complementary options may hamper their purchasing decision [10]; so is considered a cost criterion. The content quality (J4) consists of attributes such as: up-to-datedness, preciseness, timeliness, accuracy, and is considered to be beneficial criterion. The attributes security and privacy make up the criterion trust (J5), and is considered to be a benefi-

cial criterion. Support (J6) consists of payment alternatives, frequently asked questions, account maintenance, and payment alternatives; and is considered to be a benefit criterion [12]. The sum of benefit and cost criteria is obtained.

Defuzzifying the sum of benefit and cost criteria using equations (15) and (16). Also, calculation of contribution of each alternative using equation (17) gives the final ranking. The results of the ranking are provided by Table 7,

Table 7: Ranking Result

Alternatives	$Defuz_i(Ben)$	$Defuz_i(cos)$	$Contrib_i$	Rank
H1	1.23	0.58	0.65	4
H2	1.16	0.48	0.68	2
H3	1.05	0.33	0.72	1
H4	0.92	0.26	0.66	3

As seen from Table 7, ‘H3’ is the best alternative since it has the highest $Contrib_i$ value.

V. CONCLUSION

The intuitionistic fuzzy set concept was introduced to represent the hesitation or vagueness experienced by decision makers in decision making. Also, the previous studies based on various MCDM techniques using inputs from a single decision maker were found to be biased. Thus, recent studies have extended to group decision making (GDM). In this paper, we propose an MCGDM model that combines the technique of intuitionistic fuzzy Analytic Hierarchy Process (IFAHP) and Intuitionistic fuzzy multi objective optimization based on ratio analysis (IFMOORA). A numerical illustration based on ranking of OTA websites on the basis of success factors is performed. The IFAHP is used to obtain the success criteria weights and the IFMOORA technique is used to obtain the final ranking of the OTA websites. The model is found to be flexible and can be applied to various real life problems in future. For the future studies, various MCDM approaches can be combined in intuitionistic environment to perform evaluation and ranking problems in various fields.

REFERENCES

[1] D. Aloini, R. Dulmin, V. Mininno, “A Peer IF-TOPSIS based Decision Support System for Packaging Machine Selection”, Expert Systems with Applications, Vol. 41, Issue 5, 2014.

[2] K. T. Atanassov, “Intuitionistic fuzzy sets”, Fuzzy Sets and Systems, Vol. 20, Issue 1, 1986.

[3] F. E. Boran, S. Genç, M. Kurt, D. Akay, “A Multi-criteria Intuitionistic Fuzzy Group Decision Making for Supplier Selection with TOPSIS Method”, Expert Systems with Applications, Vol. 36, Issue 8, 2009.

[4] G. Büyüközkan, F. Gocer, “An Intuitionistic Fuzzy MCDM Approach for Effective Hazardous Waste Management”, In: Intelligence Systems in Environmental Management: Theory and Applications, Springer International Publishing, 2017.

[5] G. Büyüközka, F. Göçer, “Application of a New Combined Intuitionistic Fuzzy MCDM Approach based on Axiomatic Design Methodology for the Supplier Selection Problem”, Applied Soft Computing, Vol. 52, 2017.

[6] K. Chatterjee, M. B. Kar, S. Kar, “Strategic Decisions using Intuitionistic Fuzzy VIKOR Method for Information System (IS) outsourcing”, In: Computational and Business Intelligence (ISCB), IEEE, 2013.

[7] R. E. Goldsmith, “Electronic word-of-mouth”, In: Electronic Commerce: Concepts, Methodologies, Tools, and Applications, IGI Global, 2008.

[8] Z. Güngör, G. Serhadlıoğlu, S.E. Kesen, “A Fuzzy AHP Approach to Personnel Selection Problem”, Applied Soft Computing, Vol. 9, Issue 2, 2009.

[9] E. C. Ku, Y. W. Fan, “The Decision Making in Selecting Online Travel Agencies: An Application of Analytic Hierarchy Process”, Journal of Travel & Tourism Marketing, Vol. 26, Issue 5-6, 2009.

[10] H. H. Lee, E. Chang, “Consumer Attitudes toward Online Mass Customization: An Application of Extended Technology Acceptance Model”, Journal of Computer Mediated Communication, Vol. 16, Issue 2, 2011.

[11] S. W. Lin, H. P. Fu, “Uncovering Critical Success Factors for Business-to-Customer Electronic Commerce in Travel Agencies”, Journal of travel & tourism marketing, Vol. 29, Issue 6, 2012.

[12] Molla, P. S. Licker, “E-commerce Systems Success: An Attempt to Extend and Respecify the Delone and McLean Model of IS Success”, J. Electron. Commerce Res., Vol. 2, Issue 4, 2001.

[13] I. J. Pérez, F. J. Cabrerizo, E. Herrera-Viedma, “Group Decision Making Problems in a Linguistic and Dynamic Context”, Expert Systems with Applications, Vol. 38, Issue 3, 2011.

[14] T. L. Saaty, “Analytic Hierarchy Process”, John Wiley & Sons Ltd, 1980.

[15] E. N. Weiss, V.R. Rao, “AHP Design Issues for Large-Scale Systems”, Decision Sciences, Vol. 18, Issue 1, 1987.

[16] Z. Xu, “Intuitionistic Preference Relations and their Application in Group Decision Making”, Information Sciences, Vol. 177, Issue 11, 2007.

[17] M. Yazdani, “New Intuitionistic Fuzzy Approach with Multi-objective Optimisation on the Basis of Ratio Analysis Method”, International Journal of Business and Systems Research, Vol. 9, Issue 4, 2015.

[18] Z. Yue, “An Intuitionistic Fuzzy Projection-based Approach for Partner Selection”, Applied Mathematical Modelling, Vol. 37, Issue 23, 2013.

[19] L. A. Zadeh, “Fuzzy Sets”, Information and Control, Vol. 8, Issue 3, 1965.

Clustered Comparative Analysis of Security Sensor Discrimination Data

Prof. Neeraj Bhargava¹, Aakanksha Jain²,
Abhishek Kumar³, Pramod Singh Rathore⁴, Anuja Bansal⁵

¹*School of system sc. & engg. Professor MDS University Ajmer, India*

²*M.Tech Scholar at MDS University Ajmer, India*

³*Assistant Professor, Computer Science Department ACERC, Ajmer*

⁴*Assistant Professor, Computer Science Department ACERC, Ajmer*

⁵*M.Tech Scholar at MDS University Ajmer, India*

¹*profneerajbhargava@gmail.com, ²Jain1994aakanksha@gmail.com,*

³*Abhishekkmr812@gmail.com, ⁴pramodrathore88@gmail.com, ⁵anujabansal07@gmail.com*

Abstract—Security alarm is used to protect from burglary (theft), property damage and from intruders. These security alarms consists sensors and alerting device to indicate the intrusion. Clustering is data mining technique which is used to analyzing the data. In this paper we discuss about different clustering algorithm like DBSCAN, Farthest first. These algorithms are used to evaluate the different number of clusters with the sensor discrimination data base. In any organization Sensor security has many types of security alarm. It may be glass breaking alarm, smoke heat and carbon monoxide alarm, and it may be false alarm. Our aim is to compare the different algorithms with the sensors data to find density clusters i.e. which type of data will provide dense cluster of useful alarm condition. This evaluation will also detect the outliers within data such as empty alarms.

Index Terms—Acoustic; Nuisance; Detector; Farthest First Algorithm.

I. INTRODUCTION

SECURITY sensor alarm is a device that gives us indication of abnormal situation within the system. It can sense the environment situation and provide the signal by alarming for protection. In Glass braking alarms, shock sensors is installed in glass pane. When glass is break a shock frequency generated through the glass, surrounding walls and through the window frame to send the frequency to device that generates the alarm of glass breaking.

Smoke heat and carbon monoxide alarm are not turn off for a second anytime. In this two metal plates are used to generate electric current. When a smoke is entering in chambers, it detects the small amount of smoke and sound the alarm for indicates the situation. In security sensor some time may be false alarm are causes due to extra precaution i.e. if small particle of smoke is detected by chamber alarm device then it sounds the false alarm.

DBSCAN is density based clustering [5][18] algorithm with noise. It discover cluster in different size and shapes. [16]DBSCAN is able to identify class in spatial database. It is more powerful clustering [1] algorithm then other algorithm to solving anomaly detection problem. DBSCAN is

use in all fields because it determines outlier accurately. In DBSCAN clusters are dense regions separated by low dense regions.

Fa farthestFirst Algorithm is a Clustering approach which selects center groups by choosing data point on basis of their distance from each other. Algorithm works in similar way as K-means do as follows:

(i) All of first it selects the centroid.

(ii) And then it begins the cluster assignment [12].

Di differences of FarthestFirst with K-mean are:

(a) FartestFirst works in single pass only to make a set cluster of selected data points, need not to update regularly after each iteration.

(b) Centroid calculation is done with actual data not with geometric aspects.

Significant advantage of choosing FartestFirst over K-mean is it provides good performance in centroid selection because distance calculated between centroid is optimal distance. Here term centroid can be defined as a location of cluster centers in attribute set space. These centroids can move freely and are not supposed to bind in set of value of instances

II. LITERATURE SURVEY

(1). *Passive infrared detector*: PIR detector is commonly used which is us in household and small business environments. When it determines the heat energy, it started the work [6]. It does not generate its own energy and does not detect motion. It is designed according to wall or ceiling mounted.

(2). *Infra sound detector*: Infrasound detector works when it detects the sound waves at the frequency below 20 hertz. Infrasound is untreatable for human ear, its sound waves traveled along hundred kilometers. This infrasound alarm give indication of earthquake, gravity waves etc. infrasound sensor, an order frequency filter, an analog to digital converter, and for analyze the record signal MCU is consist in the infrasound detector to give us result.

(3). *Motion sensor*: Motion sensor are use to detect [7] the movement using various technologies. In motion sensor infrared, ultrasonic, vibration and contact technology is used to detect the movement. These detection technologies are used to reduce the false alarm.

(4). *Ultrasonic detector*: Ultrasonic alarm or detector emits the waves by transmitter at frequency between 15 kHz to 75 kHz. These waves is inaudible to human hear due to its high frequency, by using sound waves it determine the distance for an object. To measure the distance this device transmit the signals or sound waves, then sound waves are reflected by solid surface and detected by device.

(5). *Vibration sensor*: These devices are designed to detect an attack on structure itself. These devices are placed on barriers. When vibration is occurring, it breaks the current flow by which alarm is started [10]. Its sensor is reliable and there are low chances of generating false alarm.

III. RELATED WORK

DBSCAN Clustering: Density based [3] clustering algorithm measure the density of a region with spatial datasets. DBSCAN depends on two arguments that are epsilon and Minpts. Minpts argument is used to reduce the single-link effect and epsilon [2] distinguishes the radius of the search circle. DBSCAN [8] has the ability to find the complete cluster surrounding with another cluster but not connected with that cluster [13][14]. It uses only two parameter to ordering of the points in the database to recognize the arbitrary pattern, different size clusters and filters noise.

FarthestFirst Clustering: Farthest-first is use to speed up the clustering. But for fast performance it need less move and adjustment. It works like k-mean algorithm in some ways. It choosing data points on basis of their distance from each other to select a center point. In farthest-first algorithm each successive point is as far as possible from the set of previously-selected points [9].

IV. EXPERIMENTAL EVALUATION

The dataset contains the possible values collected from different kind of sensor from any commercial organisation. Substance group is the class attributes which have values: (a) Group A which deals with all smoke, heat and Carbon monoOxide detection related sensors data.

(b)Acoustic is category which deals with all alarms associated with glass breaking events.

(c) Nuisance category includes all False and absent alarm.

Evaluation of attributes with class variable is as shown below:

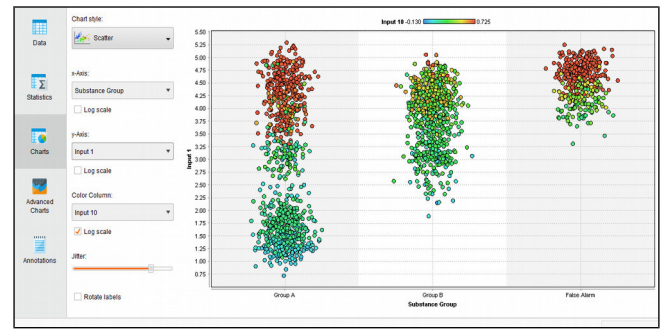


Figure 1: Evaluation of category 1 data with class variable

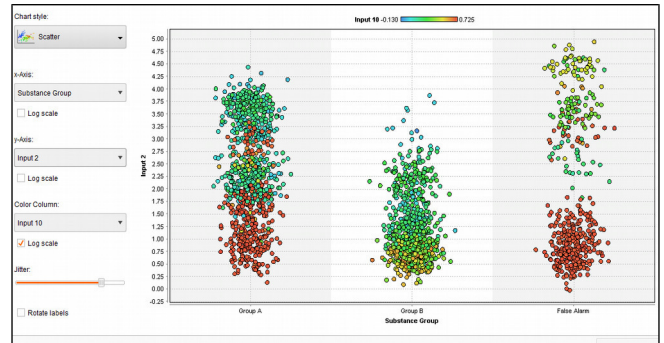


Figure 2: Evaluation of category 2 data with class variable

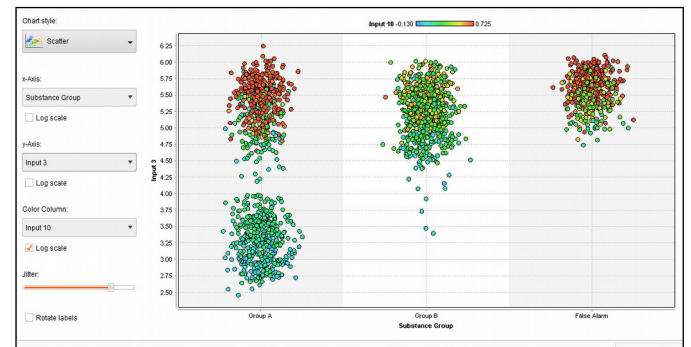


Figure 3: Evaluation of category 3 data with class variable

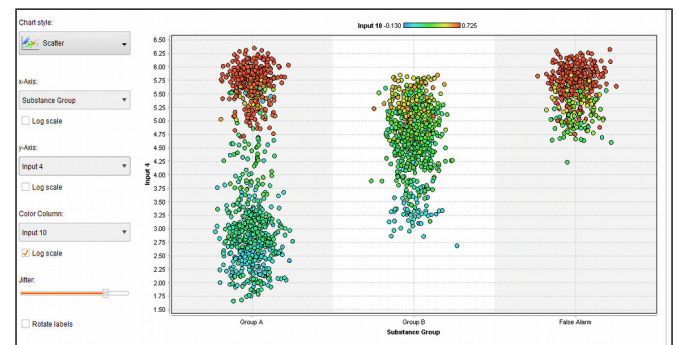


Figure 4: Evaluation of category 4 data with class variable

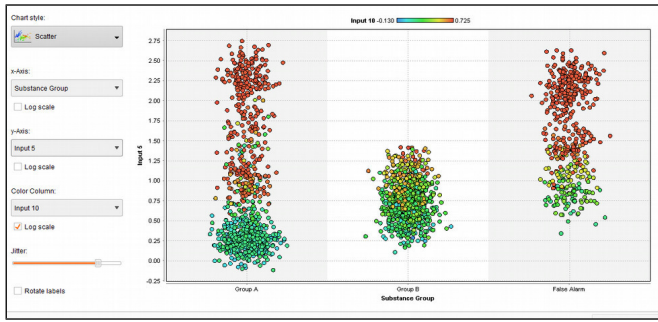


Figure 5: Evaluation of category 5 data with class variable

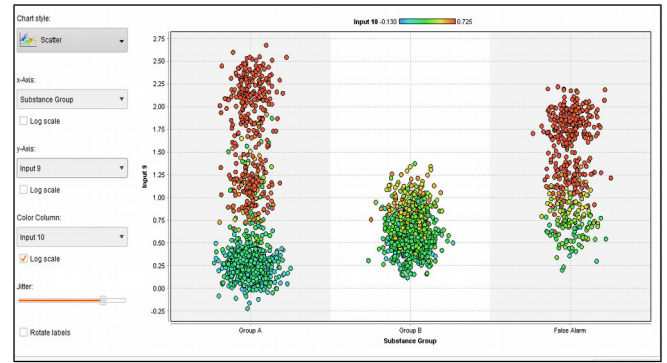


Figure 9: Evaluation of category 9 data with class variable

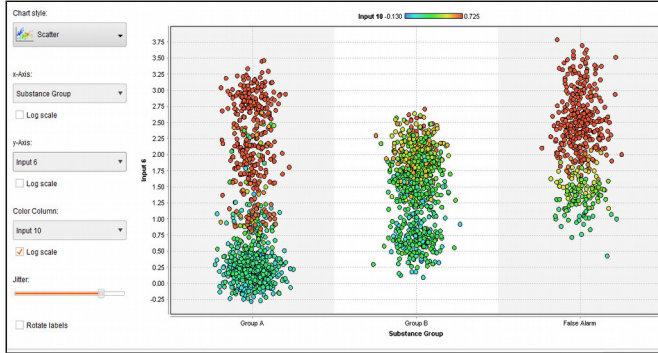


Figure 6: Evaluation of category 6 data with class variable

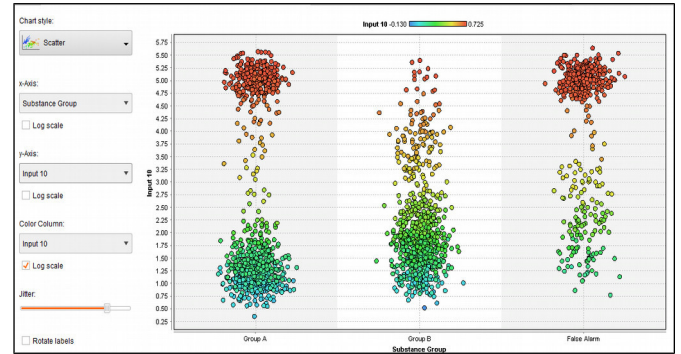


Figure 10: Evaluation of category 10 data with class variable

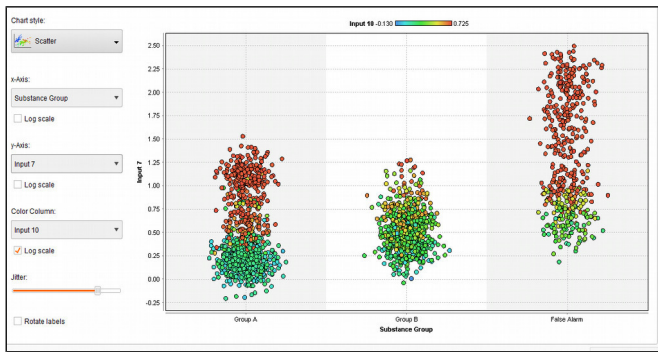


Figure 7: Evaluation of category 5 data with class variable

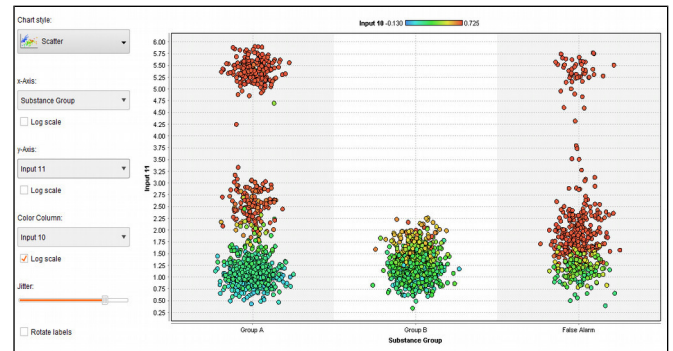


Figure 11: Evaluation of category 11 data with class variable

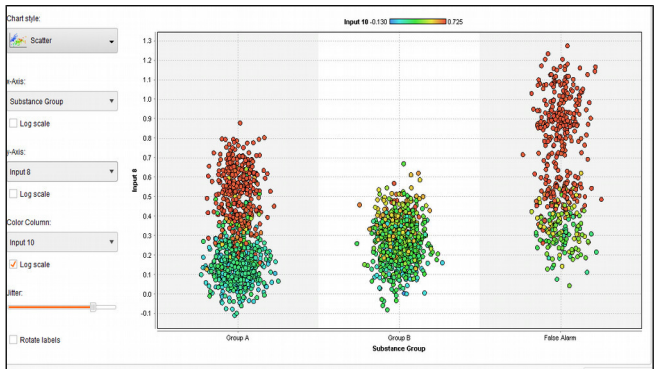


Figure 8: Evaluation of category 6 data with class variable

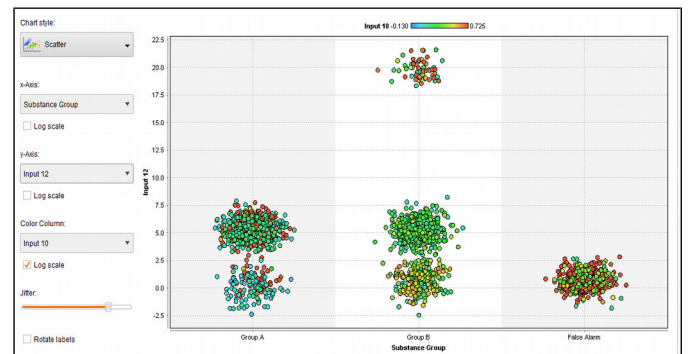


Figure 12: Evaluation of category 12 data with class variable

(1). *Experimental evaluation using Dbscan Algorithm:* Dbscan algorithm makes clusters by iteratively [4] checking neighbor elements of each data points within dataset [11]. In case nearby elements are more than minPts, a new cluster formed with O as core object. In second phase objects which are density-reachable with core object get merged into it [14]. The process terminates when no new point can be added to any cluster. Lower density points are concluded as noise thus gets omitted.

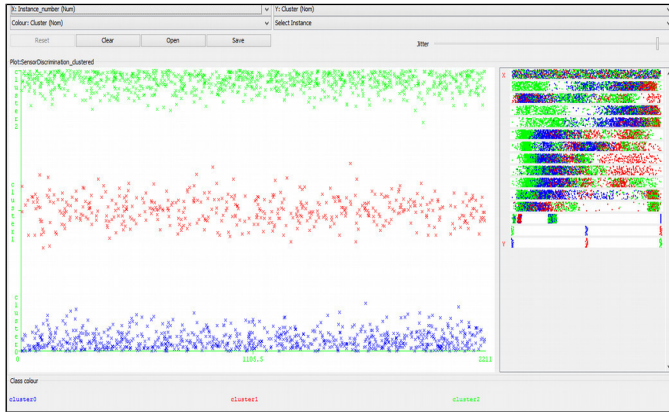


Figure 13: Experimental Evaluation using DBSCAN algorithm (Cluster assignment of each group)

(2) *Experimental evaluation using FarthestFirst Algorithm:* Farthest-First algorithm may be computed by greedy algorithm to maintain the distance between each point and previously selected points.

It works in following steps:-

(a) Firstly it initializes the empty sequence by sequencing of selected points. And also initialize the distances between each point and selected points to infinity.

(b) While all points have not been selected, the following steps are repeated:

- Find a point p by scanning the list of not-yet-selected points.
- Then remove p from the not-yet-selected point and add it to end of sequence of selected points.
- Now replace the distance stored for q by the minimum of its old value for each remaining not-yet-selected points.

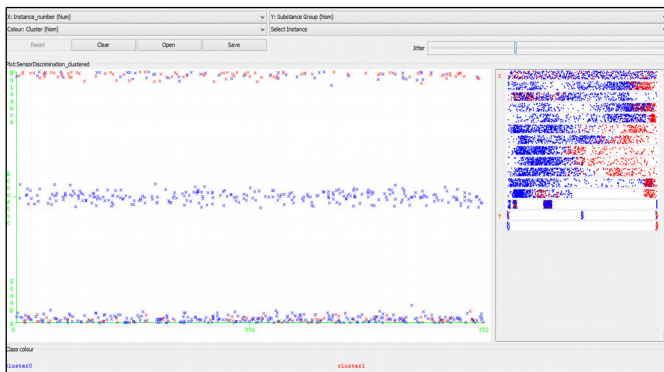


Figure 14: Experimental Evaluation using FarthestFirst algorithm (Cluster assignment in 2 groups)

V. RESULT ANALYSIS

As experimental evaluation shows how clusters were made for each input category with class attribute, how dense or noisy data are there with respect to category selection in term of class attribute. Now result is analyzed by comparing generated clusters with Dbscan and farthestfirst. Result of Dbscan consist of 3 clusters C0,C1 and C2,each have 799,456 and 947 data points 36%, 21% and 43% of total dataset respectively ,whereas result of FarthestFirst consist of 2 clusters Cluster 0 and Cluster1 have 1955 and 257 data points 88% and 12% respectively of total provided dataset. Another factor is build time which is 3.13 seconds for Db-scan and FarthestFirst take no time in building model.

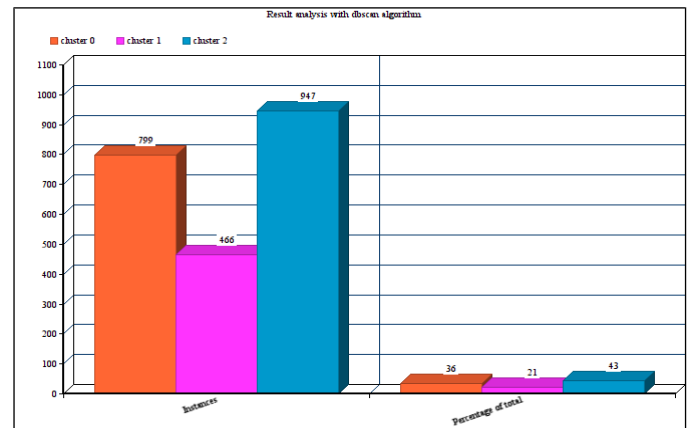


Figure 15: Result analysis with Dbscan algorithm

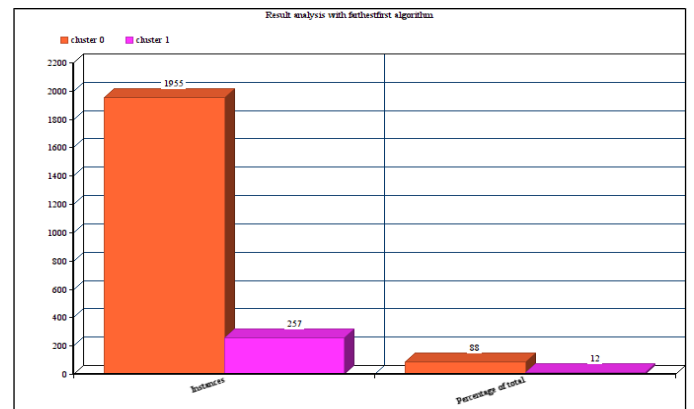


Figure 16: Result analysis with FarthestFirst algorithm

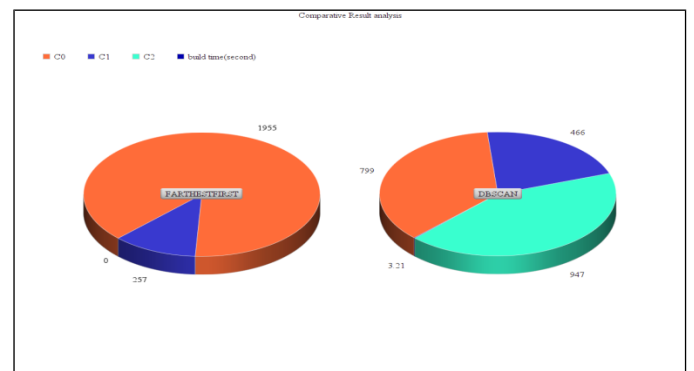


Figure 17: Comparative analysis of Dbscan with FarthestFirst

VI. CONCLUSION AND FUTURE WORK

Comparative analysis of results obtained after implementing Dbscan and FarthestFirst algorithm it is clear from the generated graph that results of Dbscan algorithm as compare to FarthestFirst are better in term of distinguishing the class variable clustering, three clusters generated will very clear assignments of data points into the corresponding cluster. Each set of input is also appraised with class variable, which clears the cluster assignment each data point of input group. We can easily analyze our desire set of the cluster and deport outliers. Both algorithms above are being analyzed using the trained set parameter based on its class attribute. For further future use this set of the clustering algorithm can be implemented in the large-sized organization to explore data detail obtained from sensors of that organization i.e we can categories type of alarm and on that basis harshness of alarm event can be decided and appropriate action can be quickly taken. Each group of input will contain each department's sensor data, which can individually get evaluated with a class variable i.e severity of sensor's alarm

REFERENCES

- [1] Ajin V W ,Lekshmy D Kumar "Big Data and Clustering Algorithms "International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016.
- [2] Yuchao Zhang, Hongfu Liu, Bo Deng "Evolutionary Clustering with DBSCAN" 2013 Ninth International Conference on Natural Computation (ICNC) 978-1-4673-4714-3/13/\$31.00 ©2013 IEEE.
- [3] Bi-Ru Dai, I.-Chang Lin. Efficient Map/Reduce-Based DBSCAN Algorithm with Optimized Data Partition [C]. IEEE Cloud, 2012: 59-66.
- [4] Tariq Ali, Sohail Asghar, Naseer Ahmed Sajid," Critical Analysis of DBSCAN Variations" 978-1-4244-8003-6/10\$26.00©2010IEEE.
- [5] S. R. Pande, S. S. Sambare and V. M Thakre, Data Clustering using data mining techniques, International Journal of advanced research in computer and communication engineering. 2012;142(4):87-91.
- [6] M. Akhil Jabbers, Dr. Prirti Chandra and Dr. B. L. Deekshatulu, Heart disease prediction system using associative classification and genetic algorithm, International conference on emerging trends in electrical, Electronics and communication technologies. 2012; 25(1): 78-90
- [7] P. Venkateswara Rao and A.Ramamohan Reddy, Shrimp disease detection using Back Propagation Neural networks, International Journal of Pharma and BioSciences. 2016; 7:3.
- [8] A. O. Quintana, "Clustering: Density-based Clustering", Universidad San Pablo, p. 4, July 2014 [Online]. Available: <http://biolab.uspceu.com/clustering/Density-based.pdf>.
- [9] Liang Sun and Shinichi Yoshida, A Novel Support Vector and K-Means based Hybrid Clustering Algorithm, Proceedings of the IEEE International Conference on Information and Automation. 2010; 64(4):20-23.
- [10] Yunseop (James) Kim, Member, IEEE, Robert G. Evans, and William M. Iversen "Remote Sensing and Control of an Irrigation System Using a Distributed Wireless Sensor Network" IEEE Transactions on Instrumentation and Measurement, Vol. 57, No. 7, July 2008
- [11] Dr.A.V.Senthil Kumar "Heart Disease Prediction Using Data Mining preprocessing and Hierarchical Clustering" International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), Vol. 4, No.6 Pages: 07 - 18 (2015)
- [12] A. O. Quintana, "Clustering: Density-based Clustering", Universidad San Pablo, p. 4, July 2014 [Online]. Available: <http://biolab.uspceu.com/clustering/Density-based.pdf>.
- [13] P. Shrivastava and H. Gupta, "A Review of Density-Based Clustering in Spatial Data", International Journal of Advanced Computer Research, vol. 2, no. 3, 2012.
- [14] A. Smith, V. Sucharita, P. Sowjanya and B. Geetha Krishna "A Predictive Model For Heart Disease Using Clustering Techniques" International Journal of Pharma and Bio Sciences, ISSN 0975-6299, Int J Pharm Bio Sci 2017 July; 8(3): (B) 529 -534
- [15] Rajkumar, A. and G. S. Reena, " DiagnosisOf Heart Disease Using DataminingAlgorithm." Global Journal of computer science and Technology, 2010. Vol. 10(Issue 10).
- [16] Varun Kumar, Nisha Rather,"Knowledge Discovery from Database using an Integration of clustering and Classification", IJACSA, vol 2 No. 3, PP. 29-33, March 2011.
- [17] Ritu Chauhan, Harleen Kaur, M.Afshar Alam, "Data Clustering Method for Discovering Clusters in Spatial Cancer Databases", International Journal of Computer Applications (0975 – 8887) Volume 10– No.6, November 2010.

Detection of Malicious Executables Using Rule Based Classification Algorithms

Dr. Neeraj Bhargava¹, Aakanksha Jain², Abhishek Kumar³, Dr. Dac-Nhuong Le⁴

¹School of system sc. & engg. Professor MDS University Ajmer, India

²M.TECH Scholar at MDS University Ajmer, India

³Assistant Professor, Computer Science Department ACERC, Ajmer

⁴Deputy-Head, Faculty of Information Technology, Haiphong University Haiphong, Vietnam

¹profneerajbhargava@gmail.com, ²Jain1994aakanksha@gmail.com, ³Abhishekkmr812@gmail.com,

⁴Nhuongld@hus.edu.vn.com

Abstract—Machine Learning class rule has varied packages together with classification, clustering, will understand association rules furthermore and is capable of the method an enormous set of the information set as measure supervised or unsupervised learning data. The paper deals with statistics mining sort set of rules on virus dataset created records from varied anti-virus logs. The work deals with classifications of malicious code per their impact on user's system & distinguishes threats on the muse in their connected severity; these threats are therefore named as malicious possible from varied sources, on various running structures. During this paper, the generated output is that the listing of records summarizing however because it ought to be the classifier algorithms are ready to predict the authentic magnificence of the days at a lower place the chosen take a look at module. The operating model deals with predicting the outliers of the threat datasets and predicts the optimum results supported analysis victimization the chosen rule. The work illustrates implementation of the algorithms corresponding to half, JRIP and RIDOR in additional economical manner because it relies on virus-log datasets to come up with A level of accuracy to the classification results.

Index Terms—Threats, Rule-based Classification, Prediction of Severity, Moderate, Malicious Executable, Danger, Normal.

I. INTRODUCTION

(1) INTRODUCTION to JRIP Algorithm (Rule-based Classification algorithm): JRIP sometimes called as RIPPER is one of popular classifier algorithm [7][5]. In JRIP instances of the dataset are evaluated in increasing order, for given dataset of threat a set of rules are generated. JRIP (RIPPER) algorithm treats each dataset of given database and generates a set of rules including all the attributes of the class. Then next class will get evaluated and does the same process as previous class, this process continues until all the classes have been covered.

(2) Introduction to PART Algorithm (Rule-based Classification algorithm): Full form of PART is Projective Adaptive Resonance Theory [4]. PART is refined method of rule generation [6]. After rule generation entire tree generated, the best tree is selected and its leaves are translated into rules. - PART support all type of classes like Binary and Nominal class and supports all type of attributes.

(3) Introduction to RIDOR Algorithm (Rule-based Classification algorithm): This algorithm is an implementation of a Ripple-Down Rule learner. The RIPPER algorithm directly extracts best rules from the provided dataset.

The Ripper algorithm completes its process in following phases:

a) Growth. b) Pruning. c) Optimization. d) Selection.

While the generation of Rules in Growth PARTs the manslayer algorithmic rule will usually decision as greedy algorithmic rule i.e. it avariciously adds attributes in rules being generated till the stopping criteria of the rule.

Incremental pruning is finished in Pruning PART, the i.e. algorithmic rule permits pruning of attribute sequences until fulfillment of pruning metrics.

The third PART suggests that the improvement stage optimizes every rule by followings a pair of steps:

1) Greedy addition of attributes in original rule

2) Grow a replacement rule severally with growing and prune PART as mentioned within the paragraph. When growing new rule victimization choice (last PART of Ripper Algorithm) PART best rule is chosen, and different not chosen rules are deleted.

II. LITERATURE SURVEY

Automated analysis operates on vast solicitation of detected malware threats and reduces the human effort in analysis of anti-malware [8] [16]. Another works are able to discuss malicious codes are dynamically analyzed by any machine driven system then analyzes some cognition performed classification system that is generated by analysis. SVM (Support Vector Machine) classifier uses these samples to coach itself so SVM will proactively notice malicious threats. Once we value the results on basis of quality of classification [15] and speed of experimental execution, it demonstrates smart results on the machine driven system. As per the author, this machine-driven analysis is enough ready to notice whether or not a file is infected or not infected, as a result of supported classification system footprints the file are often mechanically blacklisted.

The planned classifier will discover previously undiscovered malware. However, it cannot discriminate between safe and malicious threat files.

III. RELATED WORK

(1) *Adware*: Adware stands for Advertising-Supported software system. Adware generates by itself on websites once a user needs to access any video audio or the other quite data, it seems as advertising material and pop-ups. It quickly generates a commercial. Adware keeps track [9] on user activity and steals their browsing and different data. Most of the Adware's don't seem to be thought-about as a dangerous threat; it usually comes underneath low-risk threats.

(2) *Trojan Horse*: Also known as "Trojan", it interprets itself as any simple file to end users, so if they download it they are actually downloading [14] a malware. Trojan produces the effect of repudiation and Elevation of privileges to end-user. Any malicious [8] PARTY can remotely access that infected computer. Once access is obtained from a contaminative computer, the attacker can possibly to steal data like end-user login-detail; financial transactions can also access victim's electronic money. Further modification in files, installation [14] of other malware, keep track on user's activity, keylogging can also be done.

(3) *Virus*: A virus is a type of threat which is can itself and also can spread itself to other systems in a network. They can attach to programs and executable codes, when a system user accesses any of infected programs, whenever the user of that network access that program or code attachment infect those systems also. [11] Viruses spread through vulnerabilities in web apps. Viruses can spoof the information, harm client computer may be the whole network. Viruses some time generate botnets and infer user account information.

(4) *Worm*: Computer worms explore operating system vulnerabilities. They generate Payloads which are actually programming code to produce a harmful effect on host computer [10]. Basically, the worm can be defined as a replicating computer threat which produces a harmful effect on the system by slowing it down and many other annoying effects. Computer worm can be viewed as computer virus but it distinguishes itself with its self-replication characteristics [12] and spread independently means it needs not to be activated or access by running a program, opening a file, etc). Worms mostly spread by email attachments

IV. METHODOLOGY AND IMPLEMENTATION

(1) *Implementation using JRIP*: Here below figure 1 shows the practical implementation of JRIP algorithm in weka. Implemented results show that there could be 13 rule constructed by using JRIP, which are as follows Severity will be Moderate if malicious threat name is AdwareAUNPS and if source isdemo_version And threat names are Linux/Gates or HTML/Iframe.gen.w, and if threat name is Ransom_Fakecry,

Exploit_SWF.bde, Browext_Ink. If category is Adware then Severity will be Normal

Severity will be Normal if category is Worm and threat name is Conficker virus

Otherwise, Severity will be Danger

```

Classifier output
Scheme:weka.classifiers.rules.JRIP -F 3 -H 2.0 -O 2 -S 1
Relation: Virus
Instances: 187
Attributes: 6
  name
  effect
  category
  source
  detected_by
  Severity
Text mode:10-fold cross-validation
=== Classifier model (full training set) ===
JRIP rules:
=====
(name = AdwareAUNPS) => Severity=Moderate (12.0/0.0)
(source = demo_version) and (name = Linux/Gates) => Severity=Moderate (2.0/0.0)
(source = demo_version) and (name = HTML/Iframe.gen.w) => Severity=Moderate (2.0/0.0)
(name = Ransom_Fakecry) => Severity=Moderate (5.0/0.0)
(name = Exploit_SWF.bde) => Severity=Moderate (2.0/0.0)
(name = Browext_Ink) => Severity=Moderate (2.0/0.0)
(category = Adware) => Severity=Normal (12.0/4.0)
(name = Accep3773) => Severity=Normal (8.0/0.0)
(category = Worm) and (name = Conficker/Virus) => Severity=Normal (4.0/0.0)
(category = Worm) and (detected_by = .exe) => Severity=Normal (12.0/3.0)
(category = Worm) and (detected_by = .txt) => Severity=Normal (5.0/1.0)
(name = Backdoor_Files) => Severity=Normal (3.0/0.0)
=> Severity=Danger (104.0/6.0)
Number of Rules : 13
  
```

Figure (1): Prediction Rule Generated Using JRIP Rule Classifier

(2) *Implementation using PART*: Here below figure (2) shows the sensible implementation of the half algorithmic program in Weka. Enforced results show that there can be 14rule created by victimization JRIP, that area unit as follows If threat names areBackdoor_FFBMand ABAPRivpasA, SQLSlammer, ILOVEYOU, StormWorm then severity can Danger. If threat name is RunBooster then severity is traditional. If threat class is Adware then severity is traditional. If threat class is Malware then severity is Danger

If threat class is Trojan then severity is Moderate

```

Classifier output
PART decision list:
-----
name = Backdoor_FFBM: Danger (25.0)
name = RunBooster: Normal (15.0/4.0)
name = ABAPRivpasA: Danger (12.0/2.0)
name = AdwareAUNPS: Moderate (12.0)
name = SQLSlammer: Danger (15.0/4.0)
name = CryptoLocker: Danger (9.0)
name = ILOVEYOU: Danger (9.0)
name = StormWorm: Danger (9.0)
category = Adware: Normal (12.0/2.0)
category = Malware: Danger (10.0)
name = ACTSpooftask: Normal (8.0/1.0)
name = Accep3773: Normal (8.0)
category = Trojan: Moderate (15.0/4.0)
: Danger (29.0/7.0)
Number of Rules : 14
  
```

Figure (2): Prediction Rule Generated Using PART Rule Classifier

(3) *Implementation using RIDOR*:

Here below figure 3 shows the practical implementation of RIDOR algorithm in weka. Implemented results show that there could be 7rule constructed by using JRIP, which are as follows

Severity will be Normal Except if category is Virus then Severity will be Danger

and if name is Backdoor_FFBM then Severity will be Danger and if category is Malware then Severity will be Danger and if detected_by .txt) then Severity will be Danger and if name is AdwareAUNPS) then Severity will be Moderate and if category is Worm and source is email_attachment) then Severity will be Danger

```

Classifier output
==== Run Information ====
Scheme: java.classifiers.rules.Ridor -F 3 -O 1 -W 2.0
Relation: Virus
Instances: 137
Attributes: 6
  name
  affect
  category
  source
  detected_by
  severity
Test mode: 10-fold cross-validation
==== Classifier model (full training set) ====
Muggle DOM Rule Scanner(Ridor) rules
-----
Severity = Normal (131.0/133.0)
Except (category = Virus) => Severity = Danger (17.0/0.0) (7.0/0.0)
Except (name = Backdoor_JRIP) => Severity = Danger (18.0/0.0) (9.0/0.0)
Except (category = Malware) => Severity = Danger (13.0/0.0) (6.0/0.0)
Except (detected_by = .exe) => Severity = Danger (9.0/0.0) (4.0/0.0)
Except (name = Administrator) => Severity = Malware (9.0/0.0) (6.0/0.0)
Except (category = Worm) and (source = email_attachments) => Severity = Danger (4.0/0.0) (2.0/0.0)

Total number of rules (incl. the default rule): 7
Time taken to build model: 0.02 seconds
==== Misclassified cross-validation ====
    
```

Figure (3): Prediction Rule Generated Using Rule Classifier

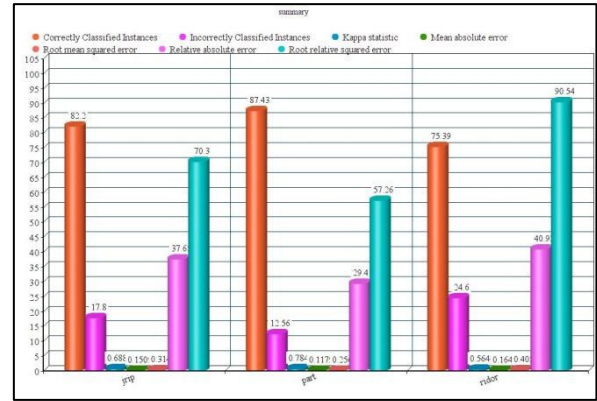


Figure (4): Summary of result

V. RESULT ANALYSIS

Algorithm	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Root relative squared error
JRIP	82.20	17.8	0.6888	0.1509	0.314	37.65	70.30
PART	87.43	12.56	0.7848	0.1179	0.256	29.41	57.26
RIDOR	75.39	24.60	0.5644	0.164	0.405	40.92	90.54

Table 1: Summary of result

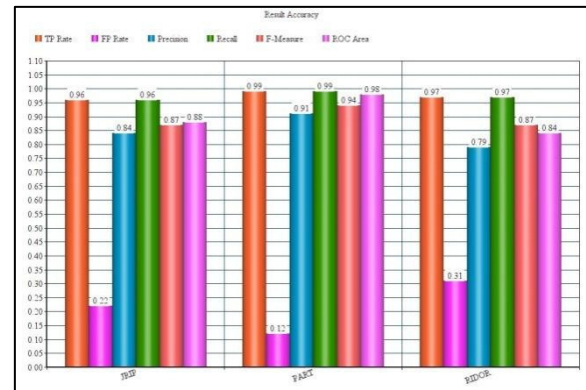


Figure (5): Comparison of results obtained through JRIP, PART and RIDOR algorithm

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	No. of Rules	Time taken to build model
JRIP	0.96	0.22	0.84	0.96	0.87	0.88	13	0.7sec
PART	0.99	0.12	0.91	0.99	0.94	0.98 ^v	14 ^v	0.5sec
RIDOR	0.97	0.31	0.79	0.97	0.87	0.84	7	0.2sec ^v

Table 2: Detail Accuracy of result by class

Here symbol 'v' means "better result"
 Symbol '*' means "worse result"
 And blank symbol means can't say whether a result is better or worse than the base algorithm.

- Here we take **JRIP** as base Algorithm.

VI. CONTRIBUTION

We propose a strategy that performs well on chosen dataset of malicious threats; on the premise of this experiment we will extend the scale of info i.e. as well as multiple networks knowledge log, a module will be created victimization this system i.e. analysis victimization 3 formula RIDOR, JRIP and half as whole to perform prediction of laptop threats in a very dynamic manner. The planned module can generate set of rules once process given dataset of threats detected that were detected In last decades, these rules will be wont to produce virus signatures to be wont to predict malicious threat samples in period of time. Rules generated by RIDOR half and JRIP, includes all potential rules, which will be generated for deleting malicious behavior of suspicious threat and outliers are shown here that shows that severity is considerably low i.e. those code of line will be treated as non-malicious class codes, and module can keep eye on those line of code that is underneath suspicious class. The planned model is going to be able to observe malicious behavior, intrusive advertisements, spying tools, phishing activities, and speedy replication of bound code in addition.

VII. CONCLUSION AND FUTURE WORK

Here in work model, we have a tendency to predict the severity of threats by imposing a information of threats in rule-based classification formula, here as established in Table two PART set of rules manufacture the upper lead to the period of rules as we have a tendency to take JRIP because the base formula. Here JRIP manufacture thirteen rules and half manufacture fourteen policies and RIDOR manufacture seven policies. Here the results of half formula proves conclusion over JRIP and RIDOR formula. Thus in term of rule generation, half represents the most effective result. currently come back to our next motive The Prediction of severity is as follows: if the class is malware is that the threat in ten times, if the class is malware it's miles sometimes risk and completely different policies like if the decision of threat is backdoor_ffbm is consistently hazarding. Severity may well be regular except class is virus i.e. severity may well be each moderate and danger and if decision = adwareaunps then severity could be moderate. Assessment on basis of consequences PART of half manufacture higher effects than JRIP and RIDOR within the period of mythical monster space enclosed and vary of rules.

As elite methodology perform well on the chosen dataset of malicious threats, in future we will extend the scale of information i.e. together with multiple networks knowledge log, a module may be made mistreatment this technique i.e analysis mistreatment 3 formula RIDOR, JRIP and half during a whole to perform prediction of pc threats during a dynamic manner.

REFERENCES

- [1] W. Nor Haizan W. Mohamed, Mohd Najib Mohd Salleh, Abdul Halim Omar "A Comparative Study of Reduced Error Pruning Method in Decision Tree Algorithms, 2012 IEEE International Conference on Control System, Computing and Engineering", 23–25 Nov. 2012, Penang, Malaysia.
- [2] Quinlan R. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers; 1993.
- [3] Thorsten Lehr, Jing Yuan, Dirk Zeumer, Supriya Jayadev, and Marylyn D Ritchie Rule based classifier for the analysis of gene-gene and gene-environment interactions in genetic association studies Published online 2011 Mar 1. doi: 10.1186/1756-0381-4-4
- [4] Pinto C M A, Machado J A T. Fractional Dynamics of Computer Virus Propagation. Mathematical Problems in Engineering, 2014: 259-305.
- [5] Ripple Down Rule learner (RIDOR) Classifier for IRIS Dataset V. Veeralakshmi et al. / International Journal of Computer Science Engineering (IJCE) ISSN : 2319-7323 Vol. 4 No.03 May 2015
- [6] Himadri Chauhan, Vipin Kumar, Sumit Pundir and Emmanuel S. Pilli 2013 International Symposium on Computational and Business Intelligence "A Comparative Study of Classification Techniques for Intrusion Detection", department of Computer Science and Engineering Graphic Era University Dehradun India DOI: 10.1109/ISCBI.2013.16
- [7] Neeraj Bhargava, Sonia Dayma, Abhishek Kumar, Pramod Singh IEEE Sponsored 3rd International Conference on Electronics and Communication Systems (ICECS 2016) An Approach for Classification using Simple CART Algorithm in Weka, MDS University Ajmer India.
- [8] George Cabau, Magda Buhu, Ciprian Opris: "Malware Classification Using Filesystem Footprints" a Bitdefender Technical University of Cluj-Napoca
- [9] Hengli Zhao, Ming Xu, Ning Zhong, Jingjing Yao, and Q. Ho, "Malicious Executables Classification
- [10] Based on Behavioral Factor Analysis," presented at the 2010 International Conference on e-Education, e-Business, e-Management and e-Learning, Sanya, China, 2010
- [11] "Malware Behavioral Analysis System." TWMAN
- [12] F. Cohen, "Computational aspects of computer viruses" Computers & Security, vol. 8, no. 4, pp. 297–298, 1989.
- [13] J. Stewart, "Behavioural malware analysis using Sandnets," Computer Fraud & Security, vol. 2006, no. Issue, pp. 4-6, December 2006.
- [14] Microsoft, "File system minifilter drivers," 2016. Available: [https://msdn.microsoft.com/enus/library/windows/hardware/ff540402\(v=vs.85\).aspx](https://msdn.microsoft.com/enus/library/windows/hardware/ff540402(v=vs.85).aspx)
- [15] Hi-Juan Jia, Yan-yan Yang, Na Guo Zhengzhou: "Research on Computer Virus Source Modeling with Immune Characteristic" Normal University, Zhengzhou Henan 450044
- [16] Muroya Y, Enatsu Y, Li H. Global stability of a delayed IRS computer virus propagation model. International Journal of Computer Mathematics, 2014, 91(3):347-367.
- [17] C. Developers, "Cuckoo sandbox - open source automated malware analysis," 2016. [Online]. Available: <https://media.blackhat.com/us-13/US-13-Bremer-Mo-Malware-Mo-Problems-Cuckoo-Sandbox-WP.pdf>

Financial Inclusion in India and PMJDY: A Critical Review

Kumar Bijoy

Shaheed Sukhdev College of Business Studies – University of Delhi

Abstract—The recent developments in banking and insurance have transformed the financial system, however, it is restricted only to certain segments of the society, excluding others. i.e. “financial exclusion”. People with low incomes, cannot access mainstream financial products such as bank accounts, credit, remittances and payment services, financial advisory services, insurance facilities, etc. This situation is detrimental for a nation’s economy as micro savings are not properly channeled into the productive sectors of the economy. This study explores to find the current status of financial inclusion in India, analyze the government’s effort through various policies like MGNREGA, DBT, PMJDY etc. The study is based on secondary information from the Governmental committees’ reports and primary information from six states of India before the launch of PMJDY. The post PMJDY status is reviewed only through the secondary information published in Government reports. It is found that PMJDY has increased the number of Bank accounts through diluting KYC for marginalized and deprived citizens but transactions in the account is very limited. A good number of accounts are dormant in the want of sufficient income to be routed through that account or lack of savings to be deposited. It is recommended to policy makers to concentrate on the creation of job opportunities for all to have complete financial inclusion in India.

Index Terms—Financial Inclusion, PMJDY, KYC, Financial Services, Sufficient income

I. INTRODUCTION

FINANCIAL Stability of an economy depends on three important pillars: Financial Inclusion, Financial Literacy and Consumer Protection. Financial Inclusion, broadly defined, refers to universal access to a wide range of financial services by all sections of the society in general and vulnerable groups such as weaker sections and low income groups at a reasonable cost. These include not only banking products but also other financial services such as insurance and equity products. Every one’s participation in formal financial service sector is the root of financial inclusion.

A well-established financial system is the effective circulatory system of economy. It empowers individuals to access the various facilities of the system in more efficient manner and contribute maximum to the system where-ever possible. This makes the system robust and economic shocks proof. Success of financial inclusion lies in achievements of three

steps: 1. Bank Account to at least one member of a household; 2. Regular deposit and withdrawals under that bank account and 3. The Bank Account is used for transactions payment. In addition, system should take care of facility and security aspects so that individuals should be free from all worries with respect to product pricing (interest rates on deposits and loans, transaction fee, penalty on delay etc.), ease of availability of products, ease of transactions and security at every steps. Financial Inclusion should help vulnerable groups such as low income groups, weaker sections, etc., to increase incomes, acquire capital, manage risk and work their way out of poverty. Various steps are being taken by the government over the years but still the problem of financial exclusion looms large and is the root cause of persistent mass poverty in India. This may be due to the skewed effort of the government i.e. focusing mainly on the supply side and ignoring the demand side of financial inclusion. In order to achieve this demand side efforts need to be taken including improving human and physical resource endowments, enhancing productivity, mitigating risk and strengthening market linkages. However, the primary focus is on improving the delivery systems, both conventional and innovative.

A. Global Scenario

The United Nations (UN) in its landmark research work titled “Building Inclusive Financial Sectors for Development”¹ (2006), popularly known as the Blue Book, had raised the basic question: “why are so many bankable people unbanked?” As per the Blue Book an inclusive financial sector, would provide access to credit for all “bankable” people and firms, to insurance for all insurable people and firms and to savings and payments services for everyone. “Financial inclusion, thus, has become an issue of worldwide concern and relevant for all (under-developed, developing and developed nations). According to GLOBAL FINANCIAL DEVELOPMENT REPORT 2014 more than 2.5 billion adults—about half of the world’s adult population—do not have a bank account. While some of these people exhibit no demand for accounts, most are excluded because of barriers such as cost, travel distance, and amount of paperwork. Enterprise surveys in 137 countries find that only 34 percent of firms in developing economies have a bank loan, whereas the share is 51 percent in developed economies. In develop-

ing economies, 35 percent of small firms identify finance as a major constraint, while only 16 percent in developed economies do so.

B. Developed Nations:

USA passed The Community Reinvestment Act, 1977 and Matched Savings Scheme (MSS) 1997 to Prohibit discrimination by banks against low and moderate income neighborhoods; to make mortgage loans to lower-income households. The Act also ensures that banks are rated every three years on their efforts in meeting community credit needs. **France** introduced Banking Act, 1984 which provides right to any person with French nationality to open an account with any bank and if refused the aggrieved person can apply to the Banque de France to designate a bank that should open an account. Further, French Banker's Association (Basic Banking Service Charter of 1992) committed to provide affordable account, Cash Card, Free access to a cash machine, Distance payment facilities, Bank Statement and Negotiable number of cheques to all citizens. **United Kingdom** in 1997 established: Social Exclusion Unit (SEU) to reduce social exclusion; Policy Action Teams (PATs) to look in an integrated way at the problems of poor neighborhoods; Financial Inclusion Task Force to ensure access to banking, affordable credit and access to face-to-face money advice and Financial Inclusion Fund to make the drive feasible. **Australia** through Australian Bankers Association (ABA) Code of Practice, 1995 and Rural Transformation Centre Programme (RTCP) is providing banking and other transaction services to communities without banking facilities. For this they are using existing stores and post offices or stand-alone centres and installing Electric Point of Sale (EPOS) equipment in post offices. **Belgium** has Charter of Basic Banking Services, 1996 which provide a basic bank account with no minimum balance and without overdraft facilities but enabled with credit transfers, direct debits, and deposit and withdrawal facilities. In **Canada**, Access to Basic Banking Services Regulations, 2003 enables Personal bank accounts to all Canadians regardless of employment or credit history and with minimum identification requirements and arranges for Banks/FIs to encash government cheques at no charge.

Note:

¹ The book is a result of a project undertaken by the UN Department of Economic and Social Affairs (DESA) and the UN Capital Development Fund (UNCDF) to analyze the obstacles to financial inclusion and to report on efforts to overcome those obstacles in various countries.

Developing and Underdeveloped Nations are also trying their best to finish financial exclusion. **Kenya** introduced M-Pesa to use the mobile phone company as a conduit for keeping and transacting money. It is Bank without a Bank. However, it can't use that money and the customers don't get any interest. Safaricom-Vodafone launched M-PESA in 2007 creating this parallel banking ecosystem. The

transactions, capped at \$500, happen in real time through a wide network of agents. Since May 2010, Safaricom has ventured into providing deposit facility too by tying up with Equity Bank. The "M-KESHO" savings account, as it is called, too can be accessed from a customer's mobile phone. **South Africa** being one of the most expensive banking markets asked a collective of five banks to launch 'no-frill' bank account scheme in 2004, the Mzansi account has no monthly fees. Five transactions in a month are free. According to a 2009 report by the non-profit Fin Mark Trust, there are 6 million Mzansi accounts in a country of 32 million. **Brazil** introduced Branchless banking through agents called banking correspondents (BCs). In 1997, as many as 40 million Brazilians out of the total 62 million did not have access to any financial services. Over the last decade, the model has grown rapidly and 150,000 BCs account for about 62 percent of the total number of points of service in the financial system. It is the world's largest such network and the total number of bank accounts has doubled between 2000 and 2008, from 63.7 million to 125.7 million. **Philippines** with two-third of population below poverty line and only 20% of the population have access to bank accounts, has adopted an approach of extensive use of Electronic money and Mobile Banking for Financial Inclusion in no bank area. E-money is used for transactions and can be recharged from cash in cash out joint. **China** addressed the issue of financial exclusion by introducing China Union Pay, a domestic payment network which would accelerate adoption of credit and debit cards in China in 2002. The total transactions through CUP has reached USD 1.9 trillion and are accepted in over 150 countries across the globe. The support from government, support from local banks, sound technological standards and most important, right value proposition for the ignored regional banks is the key to the success of CUP.

C. Indian Scenario:

First major step towards financial inclusion started with nationalization of banks in 1969 followed by establishment of Priority Sector Lending Banks in 1971 and Regional Rural Banks in 1975. Further banking penetration in rural India was tried through establishing National Bank for Agricultural and Rural Development in 1982. In addition to these institutional architecture, structural arrangements were initiated in 1992 by launching Self Help Groups Bank Linkage Program. Establishment of Small Industries Development Bank of India foundation for Micro Credit in the year 2000 and of Micro unit development & refinance agency ltd. (Mudra) in 2016 are new milestones towards the same.

Committees on Financial Inclusion: Indian Government has kept on addressing the issues of financial exclusion through various committees since independence. Some of them were directly addressing the issues of exclusion whereas others are creating the favorable condition for inclusion. Some of the committees are - **A C Shah Committee:** NBFC; **A Ghosh Committee:** Frauds & Malpractices

In Banks; **B Sivaraman Committee:** Institutional Credit For Agricultural & Rural Development; **B Venkatappaiah Committee:** All India Rural Credit Review; **Bhagwati Committee:** Unemployment; **Chatalier Committee:** Finance To Small Scale Industry; **DR Gadgil Committee:** Agricultural Finance; **Godwala Committee:** Rural Finance; **Khusrau Committee:** Agricultural Credit; **Mahalanobis Committee:** Income Distribution; **Narsimham Committee:** Financial System; **Purshottam Das Committee:** Agricultural Finance And Cooperative Societies; **R S Saria Committee:** Agricultural Finance And Cooperative Societies; **RN Malhotra Committee:** Reforms In Insurance Sector; **RN Mirdha Committee:** Cooperative Societies; **RV Gupta Committee:** Agricultural Credit Delivery; **UK Sharma Committee:** Lead Bank Scheme (Review); **Usha Thorat Panel:** Financial Inclusion; **Venketaiya Committee:** Review Of Rural Financing System; **YV Reddy Committee:** Reforms In Small Savings; **Dr. C. Rangarajan:** Estimation of Savings and Investment; **Deepak Mohanty Committee:** Medium-term Path on Financial Inclusion. The final suggestions of all these are to make India a more financially inclusive country.

Prime Minister Jan Dhan Yojna (PMJDY):

However, a transformational decision was taken by the Prime minister of India in the form of Pradhan Mantri Jan Dhan Yojana (PMJDY) on 28 August 2014. The main objective of this scheme is to ensure universal access to banking facilities with at least one basic bank account for every household. The scheme is based on “Sab ka sath sab ka vikas” i.e. inclusive growth through ‘Mera Khata Bhagya Vidhata’. Under the scheme, account holders will be provided zero-balance bank account with RuPay debit card and in addition an accidental insurance cover of Rs 1 lakh. Those who open accounts by January 26, 2015 over and above the Rs1 lakh accident, they will be given life insurance cover of Rs 30,000. Six months of opening of the bank account, holders can avail Rs 5,000 overdraft facility. The PMJDY scheme in addition to bank accounts offer overdraft facility, Insurance products and direct benefit transfer.

II. REVIEW OF LITERATURE:

Jeromi (2007) highlighted the issue of financial exclusion from formal sector and proliferation of informal financial institutions and money lenders in Kerala and invited the attention towards the social, political and economic issues arising out of this situation. **C Rangarajan (2008)** through his committee report emphasized the need to modify the credit and financial services delivery system to achieve greater inclusion. He further suggested that while banks and other financial institutions can also take some efforts on their own to improve the absorptive capacity of the clients, it is equally important for Government at various levels to initiate actions to enhance the earnings capacity of the poorer sections of the society. The two together can bring about the desired change of greater inclusion quickly. **Raghuram G. Rajan**

(2009) through his committee on financial sector reforms discussed in Chapter 3 (of report), the issue of financial inclusion and proposes a paradigm shift in the way we see inclusion. Instead of seeing it as an expanding credit, they urge to see it as expanding access to financial services, such as payments services, savings products, insurance products, and inflation-protected pensions. The Committee advocates a national goal of ensuring in three years that 90 per cent of households, if they so desire, have access to a deposit account and to the payments system, and that government transfers under various schemes be implemented through this system. **Misra (2010)** studied the various models for financial inclusion in different countries. **Malegam (2011)** through his committee report on Microfinance suggested the number of steps to strengthen financial inclusion in the country. **Arulmurugan et. al. (2013)** tried to study the various efforts for financial inclusion in India whereas **Sharma & kukreja (2013)** in their study focused on the relevance of financial inclusion in strengthening the India’s position in relation to other countries economy. **Krishna kumar & Vijaya kumar (2013)** focused in their study on effectiveness of Financial Inclusion products and programs. **Thapar (2013)** in her study concluded that though the banks are complying with RBI norms in terms of opening branches within areas of at least 2000 population, offering no frills account, kisan credit card, simplifying KYC norms, but still a lot of effort is to be put in for financial inclusion progress. **Nachiket Mor (2014)** as mandated to develop a comprehensive monitoring framework to track the progress of the financial inclusion and deepening efforts on a nationwide basis and suggested many steps to strengthen the process of inclusion. **Singh & Nisha (2014)** in their study established a direct relationship between human development and financial inclusion. They further highlighted the importance of physical infrastructure for connectivity & information and Government policies for financial inclusion.

A. Research Problem/ Hypothesis/ Objectives

The review of previous studies clearly indicate that various steps are being taken by the government over the years but still the problem of financial exclusion looms large. Most of the rural India doesn’t have access to banks, and even if they have after PMJDY they don’t have enough income and awareness/financial literacy to use the financial services.

Our objective is to study the state of financial inclusion in India through secondary and primary study and point out the weak links in the current structure and also come up with ways to counter these weak links. We aim to cover both demand and supply side problems.

III. DATA AND METHODOLOGY

Data is collected both from secondary and primary sources.

I. Secondary source:

Information on measures taken by the government in the past 20 years and progress on governmental schemes in the light of prevailing regulatory environment as well as data of micro-finance and micro-insurance in the country and some other developed and emerging economies. All the information is collected through published sources. Key sources are:

- [1] Government key advisory group reports like Nachiket Mor committee report, etc.
- [2] Private sector consultancy companies reports like reports by BCG, Mckinsey, etc.
- [3] Data published by banks at various levels like RBI, Lead Bank, NABARD.
- [4] Data Published by World Bank
- [5] Reports on successful financial inclusion in other countries

II. Primary Source:

Information through direct interaction with people and getting response recorded from various locations of India. A questionnaire is developed (see annexure at the end) to get first-hand information on extent of financial inclusion in India. Questions included - Personal Information; Basic access to financial system; Usage of Banks; Usage of Insurance.

A. Methodology

The study is based on the survey conducted in the 6 states namely Himachal Pradesh, Haryana, Delhi, Uttar Pradesh, Bihar and Jharkhand and views of people (belonging to very low income group) in villages around Shimla, Panipat, Sonapat, Mathura, Vrindavan, Kanpur, Biddupur (near Patna), and Dhanbad are recorded through questionnaires. All respondents are compulsorily of Below Poverty Line (BPL) category. Then the data collected through questionnaire is analyzed using SPSS software. Statistical results are analyzed to find out the real picture of financial inclusion in India.

The information available from secondary sources (various reports) and primary survey before PMJDY and secondary sources after PMJDY is used to draw conclusion and suggest some solution for improving financial inclusion conditions.

IV. RESEARCH FINDINGS:

As the majority of the rural population is still not included in the inclusive growth, the concept of financial inclusion becomes a challenge for the Indian economy. Since early 2000s, many concerted measures are initiated by the Reserve Bank of India and Government of India in favor of financial inclusion. Measures such as Self Help Group-Bank Linkage program (SHG-BLP), use of business facilitators and correspondents (BC-BF), easing of Know Your Customer (KYC) norms, Direct Benefit Transfer (DBT), sepa-

rate plan for urban financial inclusion, use of mobile technology, bank branches and ATMs, opening and encouraging 'no-frill-accounts' and emphasis on financial literacy have played a significant role for increasing the use of formal sources for availing loan/ credit. Measures initiated by the government include, opening customer service centers, credit counselling centers, Kisan Credit Card (KCC), Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) Scheme and Aadhar Scheme. These renewed efforts are more focused than the earlier measures which were more general in nature having a much wider scope.

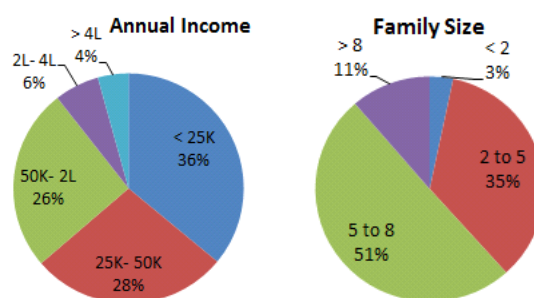
Below is a list of major schemes adopted by the Government and the RBI to promote financial inclusion. The list also gives the major pros and cons of each scheme.

The G20 Financial Inclusion Action Plan (FIAP) (2017)

is one of the most important guiding documents for the financial inclusion movement as they move rapidly to expand affordable, effective, and safe financial services to the two billion people and 200 million businesses that remain excluded from the financial system. The Global Partnership for Financial Inclusion (GPII) suggests the implementation of the G20 Financial Inclusion Action Plan (FIAP) by its four Subgroups: (1) SME Finance, (2) Regulation and Standard-Setting Bodies, (3) Financial Consumer Protection and Financial Literacy, and (4) Markets and Payment Systems. This is the most recent action plan for financial inclusion at global level.

Primary Study results:

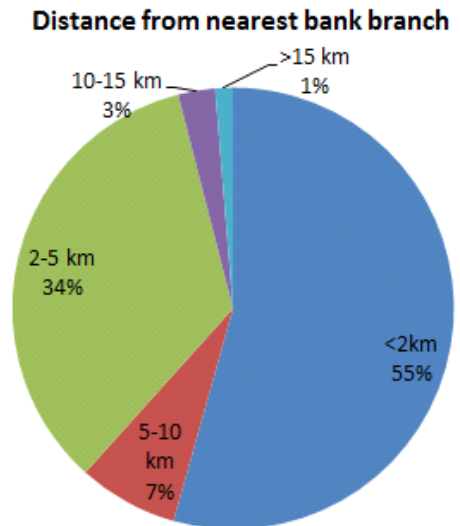
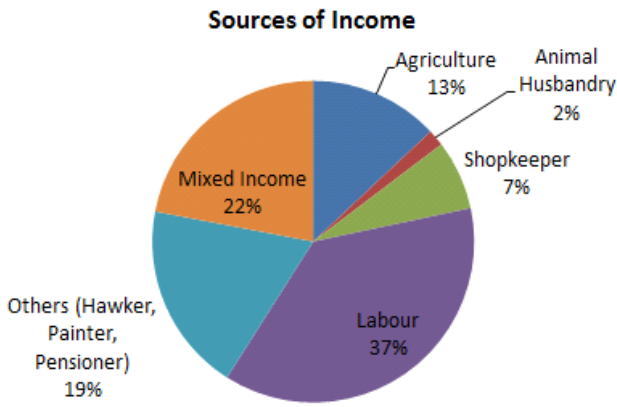
The research conducted in the 6 states of Himachal Pradesh, Haryana, Delhi, Uttar Pradesh, Bihar and Jharkhand and views of more than 600 people in the villages around Mathura, Vrindavan, Shimla and Dhanbad were recorded through questionnaires.



The **average family size was 7** members with almost 50% of the people interviewed being either illiterate or educated only up to class 8. Respondents were mostly agricultural or construction laborers usually earning an income less than Rs. 25000 per annum. **41% people did not had any bank account** whereas 50% had an account at some bank branch within 2 km radius from their homes. The primary reason for not having bank account was **lack of sufficient and stable income**.

Table: Summary of Schemes operational in India (Secondary data)

Scheme	Program	Findings	Remarks
Kissan Credit Card and General Credit Card	<ul style="list-style-type: none"> Includes Rashtriya Krishi Bima Yojana cover Gives farmers freedom to use cash in any manner (no need to submit bill receipts) 	<ul style="list-style-type: none"> Heavy subsidy by the government. (Rs 30,000 crore waiver in 2008) Many farmers default in hope of availing waivers Use of loan amount for consumption instead of agriculture 	In case of GCC Credit up to Rs 25000 without security offered
BC-BF Model	<ul style="list-style-type: none"> Allows door-step delivery of financial services Deposit, withdrawal and even remittance services provided Relaxed KYC norms 	<ul style="list-style-type: none"> High attrition rate of BCs Lack of trust between banks and agents Banks held responsible for BCs misdeeds Govt. placed cap on interest Govt. placed cap on service area (5km urban, 30km rural) Daily reporting norm- an impediment Commission structure cannot cover costs 	
Ultra-Small Branches	<ul style="list-style-type: none"> Tackles 'last-mile' connectivity problem 	<ul style="list-style-type: none"> Limited services offered on limited days 	50,000 such branches setup by march 2013
SHG-BLP	<ul style="list-style-type: none"> High recovery rate of loans Reduced dependency on informal sources of debt 	<ul style="list-style-type: none"> Interest rates are high Ever-greening of loans Size of loans given to members is generally too small to undertake any meaningful activity Wide regional disparity (48% of SHGs in Southern Region) 	
Direct Benefit Transfer (DBT)	<ul style="list-style-type: none"> Removes middlemen in cash disbursement If Aadhaar authentication equated to KYC, work is reduced significantly If online Aadhaar authentication is adopted, number of service points for banks increase 	<ul style="list-style-type: none"> Aadhaar enrollment camps can be leveraged to open bank accounts for people Not all beneficiaries have bank accounts Requires connectivity at all times for processing transactions 	
USSD Mobile Banking	<ul style="list-style-type: none"> Provides banking facilities like Money Transfer, Bill Payments, Balance Enquiries, Merchant payments etc 	<ul style="list-style-type: none"> Works only with updated GSM phones 	Unstructured Supplementary service data technology used. Is like SMS, but without cost for the user
RRBs	<ul style="list-style-type: none"> Enhances reach of formal banking system in rural areas 	<ul style="list-style-type: none"> High transaction cost due to large number of small accounts Low interest income due to concessional rates charged 	
No Frills Accounts- Directive by RBI	<ul style="list-style-type: none"> Provides access to all to enter formal banking system 	<ul style="list-style-type: none"> Accounts are created and left unused Providing this service is a loss making activity for banks 	
MGNREGA	<ul style="list-style-type: none"> Undertake useful development projects 	<ul style="list-style-type: none"> Misappropriation of funds by middlemen Fraud (by those availing the scheme) Heavy burden on exchequer (Rs 34,000 crore in 2014-15 budget) 	

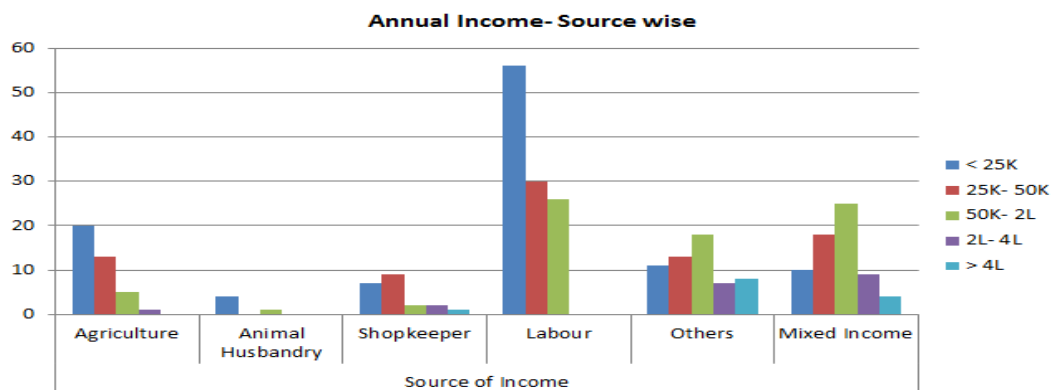
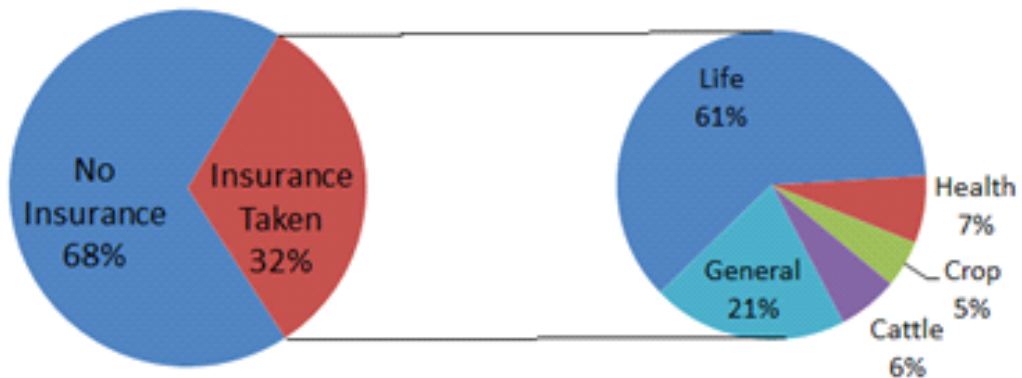


They faced multiple problems while taking a loan from banks and remain deprived from the bank loans. Some of the prominently sited reasons were:

- Cumbersome document requirements (cited by 18%)
- Harassment from the bank officials for obtaining a loan (7%)
- collateral requirements
- Lack of trust
- Afraid of the bank’s harsh loan recovery processes

They had **difficulty in availing bank loans without help from a middleman** who then charged a commission of **10%** on the loan amount. **58%** owned and used a mobile out of which only **3%** used their mobile for transaction purposes. Most of the users of transactions through mobile were from Delhi and Haryana sample. **72%** said they were not associated with any welfare scheme and were of the view

Insurance



that funds from such government welfare schemes are meant to be for Above Poverty Line (APL) people in the administration of the village. They also told that KCC loans were being availed only by farmers who had large agricultural land holdings.

With low income, the surveyed respondents could hardly save any money. Hence, any financial product including deposits or insurance is a luxury for them and they can't afford for it. Low insurance penetration in the villages with only 30% people surveyed having any kind of insurance. Out of these, most had a life insurance policy with LIC. The biggest problem is lack of sufficient income and large family size. People are aware of the benefits of insurance. They lack trust on the banking system. Commoners are deprived of the benefits of welfare schemes by powerful people in their own villages.

PMJDY for Financial Inclusion: Till 1st November 2017, 305.6 million beneficiaries banked so far under the scheme with ₹672.6472 billion balance in the accounts. 126 thousand Bank Mitras delivering branchless banking services in Sub-Service Areas and 229.7million Rupay Debit Cards issued to beneficiaries. One of the survey reported in English daily indicate that "The dominant reason for opening PMJDY accounts is to deposit savings, which is cited by 52% of account holders. The next highest reason (23%) is to avail of insurance attached to the accounts, followed by the desire to avail of direct benefit transfers and subsidies (16%)." These are very good for creating an environment for Financial Inclusion but until unless arrangement for regular income is made for all specially for BPL families along with the awareness campaign, complete and self-sustainable inclusion will not be possible.

V. SUMMARY:

This study based on secondary and primary data concludes that there is a strong structural exclusion which has divided our economy into people with no income, low income and high income. There is a need to strengthen the financial infrastructure for delivery of financial services at the doorstep of all. However, the study finds through primary survey that the low income groups are even not able to utilize the existing financial structure due to lack of sufficient sustainable income. Government efforts can only open bank accounts for the poor but cannot make them active accounts. PMJDY has made remarkable penetration in terms of access of banking service and insurance products but until employment is generated for all and a guaranteed basic income is available on regular basis this menace of financial exclusion will not end. Hence, there is a need to alleviate poverty for achieving true financial inclusion.

A. Conclusion:

Our primary study and secondary data analysis show the main stumbling block in achieving financial inclusion is the **Earning Capability**. Thus, the focus of our study is towards

3S Income Generation Model:

- There should be a Self-generating flow of income
- The income should be Sufficient
- The income should be Sustainable

This can be facilitated through:

- Infrastructure Development of warehouse facilities, irrigation facilities, movable market (e-choupals), movable banks, mobile banks etc.
- Government schemes like Jan Dan Yojna to increase the reach of banking facilities and provide benefits of insurance.
- Skill Development with focus on workforce of key industries and artisans in general
- Financial Awareness needs to be increased amongst the financially excluded by way of special initiatives

We, thus propose for a complete inclusion campaign to be in place, which will phase out subsidy over the years.

B. Recommendations

After extensive analysis of secondary and primary data, we identified the weak links in the current financial structure and suggested some ways in which these shortcomings can be corrected. Since financial inclusion is a macro level issue most of our recommendations are directed towards the central government. Some of them are as follows:

1. Enable the economy to create an opportunity for sufficient and stable income for all
2. Availability of Financial infrastructure like banks, ATMs, within reach (may be through the support of technology like internet and mobile banking).
3. Leakage of financial benefits in transit between policy desks to public pocket.
4. Awareness and availability of the financial products like Deposits, Insurance, Pension fund, overdrafts, bank loans etc.
5. Relaxation in 'KYC' norms to make the financial products and services available conveniently to small and marginal farmers and low-income households at reasonable cost with adequate protection progressively supplemented by social cash transfer.
6. Encouragement and motivation to use the financial products on regular basis.

Government is actively engaged in addressing the point number 2 to 6 of the above recommendations through Jan-Dhan Yojna, Aadhar Card and Mobile banking (JAM); DBT but the biggest bottle neck is first point i.e. creation of job opportunities. MGNREGA tried in very limited and in-

effective way to attend that issue. Make –in –India and National Skill Development Corporation (NSDC) might bring some change at this front.

C. Scope for further study:

This Study talked about the importance of financial inclusion and highlighted various policies that have been adopted in India to increase the same. Our study shows that Indian Government has taken various steps towards Financial Inclusion but the same haven't bore fruition vis-a-vis the efforts. Given the diversity and complexities in the demography of India, the successful global financial inclusion models can't be applied in isolation. Thus, the study suggests to take a modular approach towards increasing Financial Inclusion in India, especially in the most financially backward states like Odisha, Bihar, Jharkhand, Chhattisgarh etc.

REFERENCES:

- [1] Arulmurugan P, Karthikeyan P, Devi N., (2013). Financial Inclusion in India: A Theoretical Assessment. *Indian Journal of Applied Research*. 3 (3).
- [2] Jeromi, P.D., (2007). Regulation of Informal Financial Institutions: A Study of Money Lenders in Kerala. *Reserve Bank of India Occasional Papers*. 28 (1)
- [3] Krishnakumar, R & Vijayakumar, L. (2013). The Effectiveness of the Financial Inclusion Products. IRACST – International Journal of Commerce, Business and Management (IJCBM), ISSN: 2319–2828 Vol. 2, No.6, December 2013, pp 457-466.
- [4] Kumar Kabir, Merchan Camilo (2013). *What Role Startups Play In Financial Inclusion?* Available at: <http://www.cgap.org/blog/what-role-start-ups-financial-inclusion>.
- [5] Kumar Pankaj, Golait Ramesh, (2009). Bank Penetration and SHG-Bank Linkage Programme: A Critique. *Reserve Bank of India Occasional Papers*. 29 (3), pp.1-15
- [6] Misra, Udit (2010). *Financial Inclusion Models From Around The World*. Available at: <http://india.forbes.com/article/financial-inclusion/financial-inclusion-models-from-around-the-world/19292/1>.
- [7] Malegam, Y. H.(2011), Report of the Sub-Committee of the Central Board of Directors of Reserve Bank of India to Study Issues and Concerns in the MFI Sector.
- [8] Nachiket Mor (2014), Report of the Committee on Comprehensive Financial Services for Small Business and Low Income Households, *Reserve Bank of India*.
- [9] Porkodi S,Aravazhi D.,(2013). Role Of Microfinance & SHG in Financial Inclusion.*International journal of marketing,financial services & management research*,2(3).
- [10] Rangarajan,C.,Committee (2008), Report of the Committee on Financial Inclusion, Government of India.
- [11] Raghuram G. Rajan (2009), “A Hundred Small Steps - Report of the Committee on Financial Sector Reforms”.
- [12] Sarma Mandira, Pais Jesim, (2009). Financial Inclusion and Development: A Cross Country Analysis . *ICRIER Working Paper*.
- [13] Sengupta, Santanu (2009). Speeding Financial Inclusion Through BC-BF Model: Issues, Challenge & Oppurtunities.
- [14] Shah, Ajay.(2010), Report Of The Expert Committee On Harnessing The India Post Network For Financial Inclusion, Government of India.
- [15] Sharma Anupama, kukreja Sumita, (2013). An Analytical Study: Relevance of Financial Inclusion For Developing Nations. *Research Inveny: International Journal Of Engineering And Science*, pp.15-20.
- [16] Singh, Ravindra. & Nisha, (2014). Financial Inclusion and Economic Growth in India – A Theoretical Framework, *International Research Journal of Management Sociology and Humanity (IRJMSH)*, Volume 5, Issue 5, online ISSN 2277-9809, pp 131-138.
- [17] Thapar, Ashima, (2013). A study On The Effectiveness of The Financial Inclusion Program in India. *VSRD International Journal Of Business & Management*. 3 (6), pp.211-216.

Vehicular Ad-Hoc Network for Smart Cities

Bhawna Chaudhary¹, Sheetal Singh²

¹*School of Computer Systems & Sciences, Jawaharlal Nehru University, India*

²*Ramanujan College, University of Delhi, India*

¹*bhawna.0101@gmail.com, ²sheetal.2109@yahoo.co.in*

Abstract—The rapid increase in urban population is alleviating various kinds of problems such as long hours traffic-jams, pollution which is making cities life insecure and non-livable. The notion of a smart city is proposed to improve the quality of life. Smart cities are emerging to fulfill the desire for the safety of its users' and secure journeys over in the urban scenario by developing the smart mobility concept. At the same time, Vehicular Ad-hoc networks are widely accepted to attain such idea, by providing safety and non-safety applications. However, VANET has its own challenges from node mobility to location privacy. This paper discusses the application areas, security threats and their consequences of VANET into the smart city.

Keyword—Ad-Hoc Networks, VANET, Smart city, Security, Privacy

I. INTRODUCTION

THE growing need for vehicular ad-hoc networks (VANET), in which vehicles can communicate with each other, with or without the help of infrastructure on a temporary basis. The purpose of creating such network is to reduce the traffic delays and to make safe traveling for its users. In a typical VANET environment, vehicles directly communicate with other is known as V2V communication and with RSU is known as V2I communication. Each vehicle is equipped with a hardware OBU that has computational and communication capabilities.[1]. Apart from OBU, these smart vehicles are integrated with micro sensors, embedded systems, and GPS. As per dedicated short range communication (DSRC) standard, a vehicle needs to periodically broadcast the traffic and safety-related messages known as beacons. These beacons contain four-tuple information, i.e., the speed of the vehicle, location, direction and traffic events briefing accident or road scenarios. This beacon travels in the network carrying data loaded by the sender vehicle to others moving in the same region. For example, A vehicle can carry aware future traffic about the real-time traffic situations that would help other drivers to take early action in response to an unexpected situation. Due to these attractive features, this technology is considered as a mandate pillar in developing the smart city project.

VANET applications can be categorized into four main classes: Safety (time-critical and life-critical applications), Traffic Management (provide traffic information, prevent

traffic jams), Enhanced Driver Comfort and Maintenance and is described below:

- **Safety Applications:** Proactive measures for violation of traffic signals, stop sign and intersection collision; warning for the emergency vehicle coming, breakdown and wrong way driver; and can track a stolen vehicle, etc. are included in this category.
- **Traffic Management Applications:** These applications comprise of area access control, traffic flow control, electronic toll payment and rental car processing, etc. for the complete movement of the traffic on the roads.
- **Enhanced Driver Comfort Applications:** The applications under this category involve updated route guidance and navigation, parking spot locator, point-of-interest notification and map download/update/GPS correction, etc. for the driver's assistance while moving on the road.
- **Maintenance Applications:** This category includes wireless diagnostics, safety recall notice, and information about software update/flashing, etc.

Vehicular Ad-Hoc Networks (VANETs) permits Dedicated Short Range Communications (DSRC) of vehicles in the 5.9 GHz band, defined in the IEEE 802.11p standard. They support ITS with both Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications for applications in both near and far environment. In such a way, VANETs are a technology that enables a unified framework for integrating traditional ITS applications, Advanced Driver Assistance Systems (ADAS), Advanced Traveller Information Systems (ATIS), and Advanced Traffic Management Systems (ATMS).

A smart city has no universally accepted definition, but the motive behind developing such cities in every country is to enhance physical, social and economic infrastructure. In simple words, a city becomes smart when it start adopting the smart use of public resources provisions yield by the government, increase the quality of the services profound to its citizens and decrease the operational cost of public administration[2].The solutions implemented till date are installations of sensors and cameras in the public domain to encapsulate the data and use them to utilize in new services,

such as management of traffic lights, assistance in searching parking slots in crowded areas or help in flow management of public transportation. All such applications, not only improve the quality of life of citizen's but also diminish operational cost and lead to financial gains for the economy. Nowadays, specific focus in developing smart city demands smart mobility on the road, which includes enhancing traffic conditions, travel efficiency, vehicle safety and drivers/ passengers comfort while traveling. Readily availability of internet gives liberty to subscribe a bundle of services and helps to access real-time information about road conditions and facilities. For example, nearest petrol pumps/ Gas station, hospital, restaurant, etc. Such information can be easily accessed by VANET, but require perpetual network, i.e., still a big concern. The main contribution of this paper can thus be summarized as follows:

- (a) A study, which emphasizes for enabling VANET applications in the smart city project.
- (b) Also, discuss the benefits of implementing VANET applications and their possible consequences.

This article is structured as follows: Section 2 presents the contribution of VANET into Smart cities. Section 3, describes why smart cities need to introduce VANET like infrastructure?. Section 4, gives the overview of security and privacy challenges after implementing VANET. Finally, section 4, concludes the work.

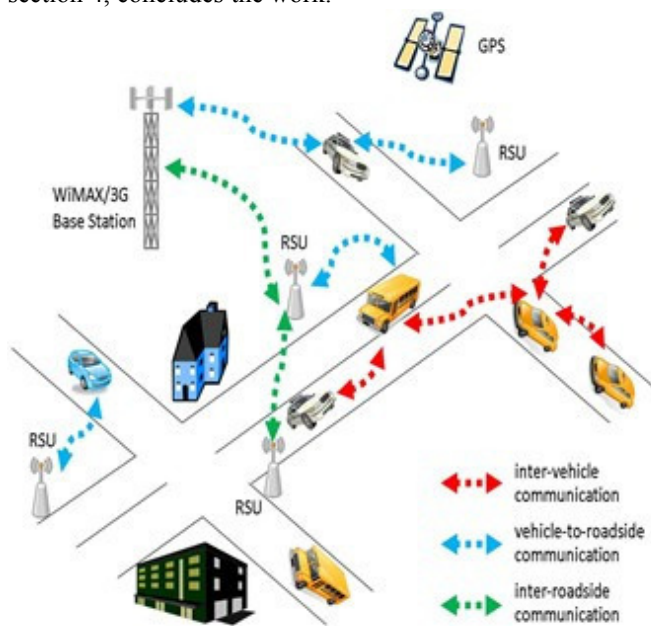


Figure 1. VANET Communication

II. RELATED WORK ON VEHICULAR AD-HOC NETWORKS IN SMART CITY

Traditionally cities (TC) were very simple in structure and developed for a thin population though the continuously increasing movement of rural population into urban triggers

the idea of developing smart-cities (SC). Current scenario of the city is very complicated as it has mix-cultured population, various categories of modes of transport, various communication technologies and utilities. Thus, to improve the shape of the current situation of metro cities the idea of a smart city has been gaining lots of attention from researchers and government. Though, debates are still not over its attribute set and standard needs. It is inevitable that these requirements cannot be satisfied without using Information and Communication Technology (ICT), which is already helping most of the city in proper utilization of resources. The extensive use of ICT in the smart city plays a vital role in collecting and delivering information and knowledge, by affecting the quality of life for its citizens by providing facilities such as e_services, a more in-depth involvement of citizens in the city governance and proactive step thanks to e_democracy and e_participation[3].

ICT acts as a digital nervous system that obtains data from heterogeneous sources such as parking spaces, traffic signals, security cameras and school thermostat, etc. The role played by ICT helps in decision-making planning and controlling activities within the automatic routine process. A perfect blending of right data and right policies can help peak hours traffic run smoothly in cities [4].

In [5], the importance of ICT in healthcare sector has been discussed i.e. using ICT; diagnosis of diseases and prevention can be made remotely by healthcare department. Apart from that observation of patients from hospitals can be done on demand. This movement from TC to SC can make learning system much forward, and it can help in improving the capitalizing system education policies. Also, it can create more opportunities for students and teachers using ICT tools. By using these tools, learning can become interactive and more research can be done[6]. The cities that are proposed to develop as SC have different sizes regarding the area, for this reason we cannot have a standard approach to apply the technologies. Another significant contribution was mentioned in [7] where the authors monitor the traffic with a novel approach for urban scenarios in which they implement collision detection and smart traffic management applications with a centralized and strongly infrastructure approach. In [7] this paper, it is suggested that initially, we should do the pilot study on smaller area sized cities which is not only cost-effective but also helps in calculating the outcomes fairly. Environment sustainability of a city is always an essential dimension as it may help to figure out the available green spaces which reflect the quality of life of citizens. Therefore, it is drawn that implementation of SC idea in such cities will help to lower the marginal cost if further improvements are required [8].

In [6], a survey explain the multifunctional data-driven intelligent transportation system, which collects a significant amount of data from various resources: Vision-Driven ITS (data gather from video sensors and used recollection include vehicle and pedestrian identification); Multisource-

Driven ITS (e.g. inductive-loop detectors, laser radar and GPS); Learning-Driven ITS (adequate assumption of the happening of accidents to enhance the safety of pedestrians by reducing the impact of vehicle collision); and Visualization-Driven ITS (helps in decision makers quickly for identifying unusual traffic patterns and then take necessary measures).

But, it requires a vast amount of memory to store the videos. Also in some complicated situations, as shown in figure 2, there are some problems regarding object reorganization.

In such a situation it becomes gruesome to identify each vehicle and perhaps to figure out the centroid of every vehicle. Hence it enhances problems in traffic density calculation. Another problem is while doing object extraction, if the color of the vehicle and the color of background become same then it becomes difficult to identify the object uniquely. Again we have to keep video data which is very large.



Figure 2. Complex scenario of traffic

III. REQUIREMENT OF VANET INFRASTRUCTURE IN SMART CITY

Smart cities are meant to enhance the performance of urban services through tight coupling of several sectors using Internet and Communication Technologies. The purpose of developing smart city is to satisfy requirements that can change in real-time, depending on the events appearing in the city. This section presents the requirement of VANET application into the smart city.

(1) *Traffic Management*: The expected number of people living in India is 1.34 billion, and 32.8 percent of this total population lives in urban areas. Today, cities are responsible for more than 75% of waste production, 80% of emissions and 75% of energy utilization. The road transport produces more than 50% of its total CO₂ emissions. This statistics profoundly point out the demand of implementing a technology that can make city transportation system more sustainable [7]. The first and foremost challenge to which developing a smart city is to resolve its traffic related problems. In

metro cities, the road is the only transport medium which creates adverse effects on overall traffic situation and environment safety level. In such complex environment, VANET can play a crucial role in improving traffic sustainability by controlling the system more efficiently and reduce energy consumption. It can be applied to the smart city infrastructure to help citizens, companies, and city government in the decision-making process, by providing real-time information about road conditions. VANET can resolve traffic situation in many different ways. For example, the GPS enabled vehicle can gather data from the network and able to predict the expected delays in a defined route and also suggest alternative less time taking routes to its users. Prior mentioned, both type of V2v and V2I network add-on their benefits into the network. V2V solution mainly focuses on safety, thanks to applications such as emergency braking system. Whereas V2I applications improve traffic flows by setting the most appropriate speed based on the future status communicated by smart traffic lights (i.e., red, yellow, green)[9].

(2) *Parking Management*: Parking problem becomes one of the major issues in the city transportation management since the spatial resource availability in the cities are limited and parking cost is increasing rapidly. In cities, people spend unnecessary time on searching for free parking spaces which not only consumes energy also cause chronic stress in life. According to the recent research significantly dealing with the parking problem states that the traffic flow peak produced by searching parking facilities can elevate 25-40% and on an average create delays of approximately 7.8 minutes[10]. It has been observed that 30% of the vehicles in the cities are struggling for the parking spot, with a consequent proportion of CO₂ emissions.

(3) *Collision Avoidance*: The VANET safety application set gathers information from other vehicles or sensors or both, for the safety management decisions. These decisions could involve a wide range of safety messages such as emergency braking, collision avoidance, intersection avoidance, alternate routes, etc. In highly populated areas, improving intersection collision systems help in reducing rate of roads accidents; that frequently occurs on the T-intersections or blind intersections. This type of communication involves V2I infrastructure of VANET which is primarily designed for cities scenario [11]. The sensors collect information from the vehicles moving towards the intersection and from the RSU's installed in that location; if sensors detect any possibility of a collision or any hazardous situation, an alert message is sent to all the vehicles driving towards that intersection.

(4) *Smart Policing*: The applications designed for road safety can be significantly proven useful in smart cities to perform the smart-policing such as the surveillance sensors installed at traffic signal can send a warning message to the drivers about a dangerous situation. On the contrary, if any vehicle breaks the rule the captured images can help the traffic police to trace out the vehicle's plate. In another situation, A warning message can be used to inform the

oncoming vehicle for the stop sign by recalculating its distance from the signal concerning its speed [12]. After receiving the message, the driver can control the speed and does not violate the sign which in turn avoid accidents. Hence, VANET not only assists the police with its applications but also makes them secure and livable by watching the roads round the clock.

(5) *Management Requirement*: In our traditional cities government is entirely responsible for every action. Limited transparency, fragmented accountability, different city division and leakage of resources are some fundamental characteristics of regular government. But in case of SC, we need e-governance which can monitor the whole city remotely. To cover the entire city's traffic and vehicles problems, we have technological solutions provided by VANET like SOS services that help in emergency cases to send a signal to the nearest infrastructure point directly. Alternatively, it depends upon the vehicles in range repeating the signal and delivering it to the most adjacent infrastructure.

To resolve the stated problems, VANET can apply to the smart city project by using the sensors, wireless communication technologies and efficient applications can be developed to assist the drivers with the information of free parking spots. Which reduces driver's frustration and negative impact on the city's traffic. In some countries, this application is already adopted and proven very convenient. More recent papers investigated the opportunities enabled by VANET technologies for car parking systems [13]. The car parking system was object area in the previous frame. Tracking by contour tracking method can be performed using two different approaches. (i) The state space models to model the contour shape and its motion. (ii) Minimizing the contour energy using direct minimization techniques like gradient descent. Made up of three layers: sensors to detect the occupancy of individual parking spots, communication technologies to collect the information from sensors, and an application layer to give (near) real-time information to the drivers. The application layer is crucial to assure a good user experience. After implementation, drivers can avail the benefits by using their smartphones or internet to make the slot reservation in a particular area parking space. Then, different technologies can be used to recognize each car at entry points (e.g., RFID, Bluetooth) and to trigger automatic reservation checking and parking payment [14].

IV. SECURITY AND PRIVACY CHALLENGES:

The awareness that smart cities are exposed to various security and privacy challenges is rapidly increasing, and lots of research focus to find feasible solutions to make our cities guarded is conducted. Security is of foremost priority in smart cities as business opportunities strongly rely on rules and regulation designed by the government to maintain privacy and trust in the technology. Before the market introduction of VANETs attractive applications and to increase

demand, citizens should be confident that these data exchanges will remain secure and confidential. Also, their personal information is kept private and not revealed to unwanted entities. In this section, we enlighten the security and privacy threats of VANETs that could be vulnerable in smart cities.

Though it has been suggested by some researchers to introduce VANET like applications into the smart city and there is no uncertainty that these features can make lives better. But on the other hand, their use makes people's life vulnerable and expose them to additional threats. For example, Smart parking is susceptible to many risks and precisely hardware attacks, which can compromise the physical devices or their communication interfaces by using jammers. To regulate and govern in such environment, a central controlling body is required. Another issue concerns the loss of privacy and confidentiality of citizens. For making a reservation in a parking slot, driver's information needs to be shared with controlling authority or central authority which will save the given credentials into a vast database. If a malicious user compromises the information available in database, confidential credential of a driver may expose or by finding out the source and destination of a person, a malicious user can figure out the possible route of the driver.

In the V2R environment, vehicles rely on the information provided by RSU's to run the network efficiently. These RSU's are responsible for sending safety messages to the vehicles and placed on the road-side with minimal security restrictions. The second type of possible attack is RSU spoofing; a malicious user can spoof the safety messages and mislead the drivers to the wrong direction. In the case of emergency situations, if a node injects malign flows into the VANET that will eventually degrade the ability of a vehicle to forward the packets containing information to neighbor nodes since RSU's are considered reliable, can create havoc in the city. Consequently, safety messages will not be relayed to traffic authorities, possibly leading to fatal accidents [15].

Another prospect is that an adversary may get access to GPS system and spoof the device, failing which can lead to faking positioning and unexpected outcomes may occur. These types of attacks like jamming attack or replay attack whenever happen in the SC not only affect a single entity but raise the question to the authorities regarding safety and security of its citizens[16]. Such things point the need for proactive measures that must be taken care of before initiating the project. As [14] says "technology gives comfort at the big cost."

Privacy is an essential factor for the public's acceptance and successful deployment of VANETs in SC as people are increasingly concerned about Big Brother enabling technologies and demanding their rights to privacy. In vehicle context, it can be attained when two related goals are satisfied (untraceability and unlinkability). This service ensures the user can preserve control of personal data and his/her lo-

cation. This service also secures other information related to the vehicle such as the identity of the driver, the driving behavior, Electronic License Plate (ELP), the speed of the vehicle, internal car sensor data, the past and present location of the vehicle from unauthorized parties. Therefore privacy can be of categorizing into various types such as (a) User Data Privacy (b) Location Privacy (c) Route Privacy.

Finally, by examining these possibilities from a different perspective, it can be concluded as it would be favorable to implement such novel security mechanisms, which can be successful where other conventional solutions have failed. By using various available technologies like LTE, GPRS, Zigbee, UMTS and Bluetooth, users remain connected in smart cities [6]. Every technology has an independent and specific security policy which clashes with the other technology policy. These raise a situation where policy conflicts may occur, so we need a standard privacy policy for a single category of technologies. For example, LTE networks use IPSec for its backhaul protection, on the other hand, Public Key Infrastructure (PKI) is used to prevent VANET from malicious users. Hence, maintaining and controlling these diverse set of security applications is not a simple task. To resolve this problem, we need a central controller, who poses powers to deploy an extensive collection of security policies by eliminating the overlap occurs between the technologies, and also helps to optimize the available network resources [14]. The central controller knows that he can efficiently manage the entanglement of globally available security policies and thus leads to enhance the overall network performance by designing compatible platform

V. CONCLUSION:

Establishing connectivity of vehicles in a smart city is required to provide support to the drivers in alarming situations. Precisely, VANETs are globally accepted as a foundation for safety, minimizing traffic related issues and exciting infotainment applications available for drivers, passengers, and walkers in the smart city. The expected contribution of VANETs are to provide information to the drivers about upcoming potential threats present in their surroundings, also gives alternate route information. The Wireless communication nature of VANETs applications makes its users vulnerable to the city, as we have discussed. Another issue that needs to be concerned is that a user should be well aware of the handling methods of such technologies. Mostly, this paper provides the understanding of how the VANET can be

useful and susceptible to the smart cities at the same moment. To resolve vulnerabilities, we need government contribution to regulate and apply the standards designed by standardization bodies and research institutions along with car manufacturers, to built secure and safe smart cities.

REFERENCES

- [1] S. A.-Sultan, M. M. Al-Doori, A. H. Al-Bayatti, H. Zedan, "A comprehensive survey on the Vehicular Ad Hoc Network", *Journal of Network and Computer Application*, Vol.37, pp.380-392, Jan 2014.
- [2] R. Mangaracina, A. Perego, G. Salvadori, A. Tumino, "A comprehensive view of ITS for urban smart mobility", *International Journal of Logistics*, Vol. 20, No.1, pp 39-52, 2017.
- [3] R. P. Dameri, "Using ICT in Smart City", Springer international publishing, pp 45-63, 2017.
- [4] H. Chourabi et al, "Understanding smart cities: An integrated framework", In system sciences(HICSS), International conference held in Hawaii, pp 2289-2297, 2012.
- [5] T. V. Ramachandra, Shwetmala, "Emission from India's transport sector: Statewide Synthesis", *Atmospheric Environment*, Vol.43, Issue 34, pp 5510-5517, Nov 2009.
- [6] A. D. Maio, M. R. Palatella, R. Sona, L. Lamorte, X. V. Lajosana et al., "Enabling SDN in Vanets : What is the impact on security?," *Sensors* 2016.
- [7] A. F.Santamaria, C. Sottile, A.Lupia, P. Rimando, "An efficient traffic management protocol based in IEEE 802.11p standard", *International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2014)*, pp 634-645, IEEE.
- [8] P. Neirotti, A. D. Marco, A. C. Caigliano, G. Mangano, F. Scorrano, "Current trends in smart city initiative: Some stylised facts", *Cities* 38, pp 25-36, Jun 2014.
- [9] E. L.Glasear, J. D.Gottlieb, "Urban resurgence and the consumer city.", *Urban Studies*, Vol.43, Issue 8, pp 1275-1299, 2006.
- [10] Z. Ji, I. Ganchev, M. O'Droma, L. Zhao, X. Zhang, "A Cloud-Based Parking Middleware for IoT Based Smart Cities", *Sensor*, Vol.14, Issue 12, pp 22372-22393, Nov 2014.
- [11] T. B. Hodel and S. Cong, "Parking space optimisation services, uniformed web application architecture", *ITS world congress proceedings*, 2003.
- [12] Z. Y. Rawashdeh and S. M. Mahmud, "Intersection collision avoidance system architecture." In *consumer communications and networking conference, CCNC 2008*, pp 493-497.
- [13] G. Yan, S. Olariu and M.C. Weigle, "Providing VANET security through active position detection.", *Computer communications*, Vol.31, Issue 12, pp 2883-2897, 2008.
- [14] B. Mokhtar, M. Azab, "Survey on Security Issues in Vehicular Ad Hoc Networks", *Alexandria Engineering Journal*, Vol.54, Issue 4, pp 1115-1126, Dec 2015.
- [15] J.-H. Shin, H. -B. Jun, "A study on smart parking guidance algorithm", *Transport research part C : Emerging technologies*, Vol.44, pp 291-317, July 2014.
- [16] P. C. Chishere, S. Magrini, "population growth in european cities: weather matters but only nationally", *Regional studies*, Vol.40, Issue1, pp 23-27, 2006.
- [17] A. M. Malla, R. K. Sahu, "Security attacks with an effective solution for dos attacks", *International Journal of Computer Applications*, Vol. 66, Issue 22, Jan 2013.
- [18] Y. C. Hu, H. J. Wang, "Location privacy in wireless networks," *Proceeding of ACM SIGCOMM Asia Workshop*, 2005.

Reliability Modeling of OSS Systems based on Innovation-Diffusion Theory and Imperfect Debugging

Neha Gandhi¹, Neha Gondwal², Abhishek Tandon³

¹Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, India

²Department of Operational Research, University of Delhi, India

³Shaheed Sukhdev College of Business Studies, University of Delhi, India

¹nehagandhi1990@gmail.com, ²neha28gondwal@gmail.com, ³abhishektandon86@gmail.com

Abstract—Open Source Software (OSS) has obtained widespread popularity in last few decades due to the exceptional contribution of some well established ones like Apache, Android, MySQL, LibreOffice, Linux etc. not only in the field of information technology but also in other sectors such as research, business and education. These systems are characterized by a huge shift in development pattern they adopt in comparison to proprietary software. Reliability modeling for such systems therefore is a growing area of research now days. Number of users adopting and working on refinement of such systems post-release play an indispensable role in their reliability growth. In this paper, we have proposed a software reliability growth model (SRGM) based on Non-homogeneous Poisson process (NHPP) based on number of users, under the phenomenon of Imperfect Debugging. The renowned Bass Model from Marketing based on the Theory of Diffusion of Innovation is used to depict the user growth phenomenon. Various fault content functions are considered in proposed models to represent imperfect debugging conditions and their performance is evaluated on fault dataset of GNOME 2.0. Four goodness-of-fit criteria namely Coefficient of Determination, Mean Square Error, Predictive Ratio Risk, and Predictive Power are used to calculate the estimation accuracy of all the proposed models and it has been observed that prediction capabilities of models based on imperfect debugging phenomenon is better than model assuming perfect debugging situation.

Index Terms—Software Reliability Growth Model; Open Source Software; Imperfect debugging; Fault content function; Diffusion of innovation.

I. INTRODUCTION

WITH the advancements in field of technology, a visual expansion can be seen in the software industry. Open source software (OSS) has revolutionized the development trend of software in past few decades. Ubiquitous acceptance of OSS has become the sole reason for a huge inclination of developers towards these software systems. OSS can be defined as the software for which the source code is shared to learn, modify, and extend the software under some licensing guidelines [1]. Some of the characteristics of OSS due to which traditional SDLC models cannot be applied to OSS are described as follows:

A. Characteristics of OSS

- **Unclear Requirements:** OSS development doesn't witness a dedicated requirement elicitation phase where requirements are documented as in case of closed software. Here, development just starts with a single developer or a small group with a random idea. Requirements are not properly framed and freeze before development.
- **Unstable Team Size:** On contrary to closed source software a dedicated team of fixed number of people work on a project, OSS doesn't has fixed number of people working on it. It varies with time and attractiveness of software.
- **Minimal Testing Effort:** In OSS development, the main focus is given to implementation of idea as opposed to closed source software where sincere efforts and significant time is spent on testing phase after development completes.
- **No Deadlines:** Since development is totally dependent on voluntary participation of developers across the world, no strict deadlines on deliverables can be imposed.

As observed in above characteristics, during software development life cycle (SDLC) of OSS, it doesn't undergo an exhaustive testing phase as opposed to proprietary software where extensive testing efforts are spent before release of software. Because OSS cannot be tested rigorously for its functionality in such a limited duration of time, it becomes significant to study its reliability growth during its user phase. Software Reliability is the probability of software to perform its operations without any fail for a specified time interval in specified conditions [1]. Reliability growth of software can be modeled by a software reliability growth model (SRGM). An SRGM is a mathematical representation that depicts the software fault detection process as a function of different factors and parameters e.g. CPU time, Testing effort expenditures, Number of test cases, Code coverage etc. In the process of simulating software fault process with an SRGM, it is often assumed that each time a fault is encountered, the responsible fault is removed with certainty.

However this assumption is quite impractical as correction of a fault involves modifications in original source code which may lead to addition of new faults. In this paper, we propose an SRGM in which phenomenon of software fault detection is modeled with respect to user growth function on real life failure dataset (GNOME 2.0) under imperfect debugging conditions.

II. LITERATURE REVIEW

Numerous studies have been done in past to model reliability growth phenomenon for OSS and various SRGMs have been proposed. Goel-Okumoto model [3], Yamada S-Shaped model [4], Musa-Okumoto model [5] are some of the traditional NHPP based software reliability growth models for closed source software systems. Paulson et al. [6] performed an empirical study on open and closed source software to quantitatively investigate and validate the perceptions about these software. Rossi et al. [7] discussed the pattern of occurrence of faults in various OSS and thus make reliability predictions for future. Yamada and Yamaguchi [8] discussed a statistical process control method for OSS to determine the stability and thus getting estimation of development time needed to attain desired level of reliability. Li et al. [9] proposed reliability analysis model and used the model to predict optimal version-updating for OSS. Tamura and Yamada [10] combined software growth modeling with neural network approaches to estimate the reliability of OSS. Yang et al. [11] have given an SRGM for reliability estimation of multi-release OSS. Several other studies also have performed analysis on reliability of open source software [12, 13, 15]. Various studies demonstrated the reliability analysis of software under imperfect debugging scenarios. Kapur et al. [16] proposed frameworks to derive SRGMs in the presence of some realistic processes like imperfect debugging and error generation. Pham [17] developed a cost model with imperfect debugging conditions considering penalty cost due to delays in software delivery and random length of software development. Chu-Ti Lin [18] investigated the effects of imperfect debugging in modeling reliability via a simulation based model. In our paper, we model software fault process for OSS in its functional phase relating the usage factor with reliability modeling and comparing its performance under different fault content functions used for imperfect debugging element in model.

III. MODEL FOUNDATIONS

The SDLC followed by OSS is unlike normal commercial software. For OSS, the testing effort spent is almost negligible in comparison to commercial software and therefore voluntary participation from developers across the world is crucial in refinement of quality of OSS. Studies done in past in direction of reliability assessment of OSS assumed that fault detection rate for OSS in operational phase will follow a hump-shaped curve. Initially, it grows with the growth in volunteer participation and reaches the highest point. This is

due to the craze in developers for the new product in market. And then starts decreasing with the decrease in number of volunteers with time. This may be due their fall in interest for same product over time or due to the introduction of some new product in market. Similar trend is observed in user growth pattern of an Innovation. According to Rogers (1962), An Innovation can be characterized by following key features.

A. Characteristics of Innovation

1. *Relative Advantage*: It is the measure of improvement innovation brings over its competitive option. OSS brings lot of advantages in terms of cost benefit, vast resource knowledge base; lack of delivery deadlines saves the product from quality compromises.
2. *Compatibility*: It is the level of compatibility the innovation has with the current lifestyle of people. OSS doesn't need a huge lifestyle change over its closed source competitors. A user of commercial software can seamlessly adopt OSS.
3. *Complexity*: It refers to the level of difficulty an adopter of innovation will face to learn or use innovation. Various communities are there on internet for discussion of problems and challenges on OSS. They serve perfect to carry any discussion on OSS and pass on the reviews of the product to other potential users.
4. *Trialability*: It refers to the degree to which the innovation can be explored and tested by potential users. OSS is released in beta versions for the trial of users. Moreover after the release source code is also made available so that it can be expanded and customized as per needs.
5. *Observability*: It is the extent to which results of innovation are accessible to group of potential adopters. The important features of OSS like cost effectiveness, easy access and sharing of code etc are easily observable by people using it like educational institute, employees in corporate sector, freelancers etc.

Due to the presence of above characteristics in OSS, it can be considered as an Innovation. The usage growth process for an innovation can be best explicated with the renowned Innovation Diffusion Model of marketing (Bass, 1969). Since OSS is an Innovation it is justified to apply this model to describe user growth for an OSS.

B. Diffusion of Innovation

According to Rogers [19], the process of propagation of innovation i.e. new idea or new product, among the members of the communication system over a time frame is known as diffusion. As OSS is an Innovation, This theory perfectly applies to it. Among the potential volunteers of OSS, initially it is adopted by ones who are opinion leaders i.e. people who adopt it based on their own interest in software. They are also known as Innovators. Another group of people who later start using OSS based on word-of-mouth from innovators are known as Imitators. To represent the usage growth factor in our proposed SRGM we have used The Bass Model [20].

Bass Model

The Bass Model is a mathematical representation of process of diffusion. It depicts the concept of adoption of a product (OSS) among the potential users i.e. Innovators and Imitators through a mathematical expression. According to Bass, the diffusion process is defined by the following equation,

$$\frac{dN(t)}{dt} = \left(p + q \frac{N(t)}{N} \right) (\bar{N} - N(t)) \tag{1}$$

Where, $N(t)$ depicts cumulative number of adopters at

time t , $p(\bar{N} - N(t))$ represents those adopters (Innovators)

who are not influenced by number of users already adopted therefore p denotes the coefficient of innovation. The term

$q \frac{N(t)}{N} (\bar{N} - N(t))$ represents those users (Imitators)

who are influenced by number of previous adopters and thus q depicts the coefficient of imitation. Solution of equation (1) under the initial condition of $N(0)=0$ results in equation (2) which is given as follows:

$$N(t) = \bar{N} \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \tag{2}$$

C. Model Development

Notations

t	Calendar time
m, m(t)	Expected number of faults removed in time interval (0, t]
N, N(t)	Cumulative number of users in time interval (0,1]
a	Constant, the number of initial faults in a software
b	Constant, fault removal rate
p	Constant, coefficient of innovation
q	Constant, coefficient of imitation
\bar{N}	Constant, total number of potential users of the software
α	Proportion of error generation

Assumptions

- Software fault process is a NHPP phenomenon.
- The number of failures during operational phase is

dependent upon the number of faults remaining in the software.

- As soon as any deviation from expected behaviour is encountered, it is considered as a failure and the fault that causes that failure is located. There is a possibility that at the time of fixing of a fault, few additional faults may get introduced.
- The number of faults removed is a function of number of users working on that software.
- The number of users is assumed to be a function of time and they are represented by the diffusion model given by Bass [20].

Based on above assumptions, the failure phenomenon can be illustrated as follows:

$$\frac{dm}{dt} = \frac{dm}{dN} \frac{dN}{dt} \tag{3}$$

The components on the right side of the equation are discussed as follows:

Component-1

This component describes the rate of detection of faults with respect to users. The rate at which failures occur depends upon the number of faults remaining in the software. As per this assumption the differential equation for fault removal can be written as,

$$\frac{dm}{dN} = b(a(N) - m) \tag{4}$$

where, $a(N)$ denotes the fault content function with respect to number of users.

Component-2

The concept of user growth in adoption of OSS is the focus of the paper. The Innovation- Diffusion Model proposed by Bass [20] is used to describe user growth phenomenon in this study. According to Bass, the process of adoption of innovation in the potential population comprises of innovators and imitators and is defined by the equation (1) and is used to represent this component in equation (3). The solution of this equation for initial condition $N(0)=0$ is given by equation (2).

D. Fault Content functions

In this paper, we have assumed fault content function as a function of number of users working on OSS in its operational phase and is denoted by $a(N)$. The reason behind this is as the number of users working on an OSS increase, the more are the changes and modifications done in original code and hence the fault content also increases. Different fault content functions for corresponding to the cases of perfect and imperfect debugging are discussed as follows,

Perfect Debugging: The process of detection of faults and their rectification is termed as Debugging. Debugging

process that does not incorporate additional faults in the software is referred to as perfect debugging.

Case 1: In this case the fault content function represented by $a(N)$ is assumed to be a constant. i.e. no new faults are introduced during debugging process. Therefore here,

$$a(N) = a \quad (5)$$

Imperfect Debugging: Imperfect debugging is the phenomenon where new faults get introduced while correcting the previous ones. Various fault content functions pertaining to this phenomenon are discussed hereby,

Case 2: In this case, we considered number of faults to be a linear function of number of users.

$$a(N) = a(1 + \alpha N) \quad (6)$$

Case 3: In this case, we consider an exponential fault content function using Yamada et al. [21], it means faults are introduced exponentially with respect to number of users.

$$a(N) = a e^{\alpha N} \quad (7)$$

Case 4: Here, we have adopted rate of introduction of new faults as a function of the number of faults already removed in the software.

$$a(N) = a + \alpha m(N) \quad (8)$$

Case 5: Under this case of imperfect debugging, we assume that the new faults can be introduced exponentially per detected fault (Pham & Zhang [22]). Here C is assumed as constant.

$$a(N) = c + a(1 - e^{-\alpha N}) \quad (9)$$

Here, $N = N(t)$ represents the number of users of OSS up till time t .

(5) Proposed Models

The proposed models obtained by combining both components of Equation (3) are explained as follows:

SRGM 1: We obtained following model under above stated case 1.

$$m_1(t) = a(1 - e^{-bN}) \quad (10)$$

SRGM 2: This model is obtained under case 2 which represents imperfect debugging.

$$m_2(t) = a \left(\alpha N + \left(1 - \frac{\alpha}{b} \right) (1 - e^{-bN}) \right) \quad (11)$$

SRGM 3: Under case 3, following expression is obtained:

$$m_3(t) = \frac{ab}{\alpha + b} (e^{\alpha N} - e^{-bN}) \quad (12)$$

SRGM 4: For case 4, we got the following model:

$$m_4(t) = \frac{a}{1 - \alpha} (1 - e^{-b(1 - \alpha)N}) \quad (13)$$

SRGM 5: The following model is based on above discussed case 5:

$$m_5(t) = (a + c) (1 - e^{-bN}) - \frac{ab}{b - \alpha} (e^{-\alpha N} - e^{-bN}) \quad (14)$$

For SRGM 1- SRGM 5, $N(t)$ is given by equation (2).

IV. NUMERICAL STUDY

For the purpose of parameter estimation, we have used the real time failures data set of a very renowned OSS i.e. GNOME 2.0 provided by Li et al. (2011) as given in Table 1.

Table 1: Detected faults in GNOME 2.0 release

Day	Detected Bugs	Day	Detected Bugs	Day	Detected Bugs	Day	Detected Bugs
1	6	7	8	13	6	19	1
2	5	8	4	14	8	20	1
3	3	9	8	15	6	21	1
4	2	10	3	16	2	22	2
5	5	11	2	17	2	24	3
6	5	12	1	18	1		

Estimation of proposed models is performed with the Least Squares Principle. The estimation results of the models proposed in this study as described in Equation 10-14 on the GNOME 2.0 dataset are presented in Table 2.

Table 2: Estimated Parameters

Proposed model	a	b	\bar{N}	p	q	α	c
SRGM 1	94.445	0.053	50.922	0.014	0.143	-	-
SRGM 2	81.349	0.420	37.104	0.003	0.117	0.004	-
SRGM 3	80.736	0.827	58.038	0.01	0.112	0.008	-
SRGM 4	84.537	0.259	57.586	0.003	0.112	0.001	-
SRGM 5	87.243	0.721	30.237	0.002	0.117	0.009	80.068

Also, we have performed the comparison of our proposed SRGMs using four goodness-of-fit criteria namely Coefficient of Determination (R^2), Mean Square Error (MSE), Predictive Ratio Risk (PRR), and Predictive Power (PP). The expression and interpretation of these is discussed may be found in Ref. [23].

The goodness-of-fit values obtained for the above mentioned criteria for the five proposed SRGM are summarised in Table 3.

Table 3: Goodness of fit Comparison Table

Proposed Model	R^2	MSE	PRR	PP
SRGM 1	0.994	4.336	4.050	3.395
SRGM 2	0.994	4.081	3.635	3.132
SRGM 3	0.994	4.024	3.543	3.080
SRGM 4	0.994	4.051	3.521	3.072
SRGM 5	0.994	4.077	3.648	3.137

The goodness-of-fit curves obtained corresponding to SRGM 1, SRGM 2, SRGM 3, SRGM 4, and SRGM 5 are shown in Figure 1-5.

V. DISCUSSION ON RESULTS

From Table 2, It can be observed that the value of q which represents the coefficient of imitation is always greater than p which is the coefficient of innovation. This observed relation between p and q is consistent with the Bass Diffusion Model and illustrates the importance of word of mouth on the decision making of used by the potential users. From the Table 3, it can also be concluded that the all the SRGMs based on bass growth function fits data well. Moreover, from the values of goodness-of-fit criteria obtained it can be inferred that models built over imperfect debugging assumption (SRGM 2-SRGM 5) outperform the model depicting perfect debugging situation (SRGM 1) which explains the fact that imperfect debugging models are close to realistic situation.

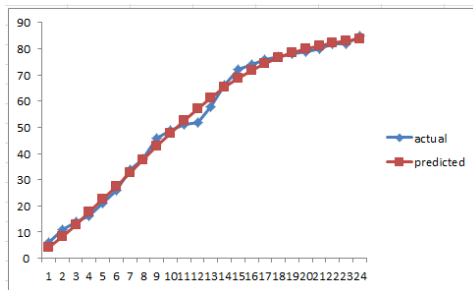


Figure 1: goodness-of-fit curve for SRGM 1

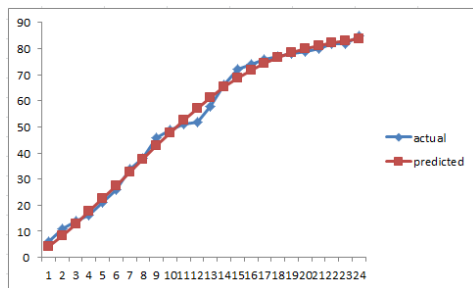


Figure 2: goodness-of-fit curve for SRGM 2

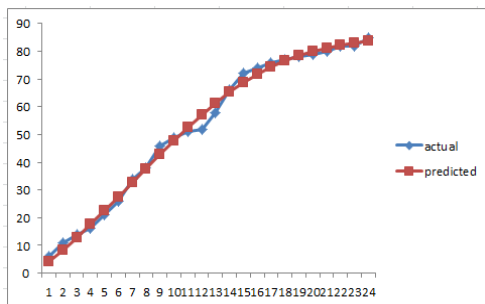


Figure 3: goodness-of-fit curve for SRGM 3

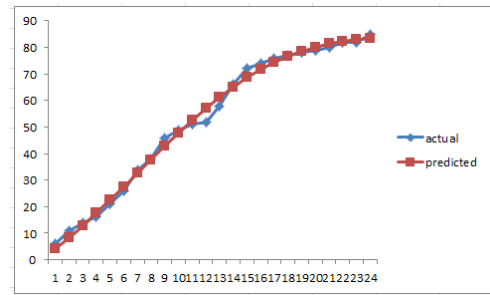


Figure 4: goodness-of-fit curve for SRGM 4

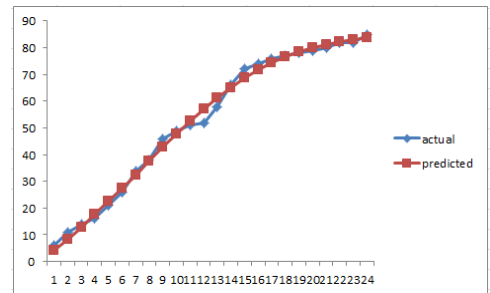


Figure 5: goodness-of-fit curve for SRGM 5

VI. CONCLUSION AND FUTURE SCOPE

In this study, we have related user growth with reliability growth phenomenon for OSS. Bass Innovation Diffusion model has been used to represent user growth function. Five SRGMs have been proposed using various fault content functions to model perfect and imperfect debugging phenomenon. The mean value functions for the proposed SRGMs are used and parameters' estimates are calculated using least square estimation technique. Performance of all the proposed models is compared using four goodness-of-fit methods: R^2 , MSE, PRR, and PR and it is found that imperfect debugging models gave better estimation results as compared to model with assumption of perfect debugging. In future studies the functions to model imperfect debugging can be extended to include the concept of change point or randomness in user growth. For user growth we have used Bass Innovation-Imitation model. In future, we may extend our work for multi release OSS.

REFERENCES

- [1] W. S. Jawadekar, "Software Engg", Tata McGraw-Hill Education, 2004.
- [2] P. Kapur, H. Pham, A. Gupta, P. Jha, "Software reliability assessment with OR applications", Springer, 2011.
- [3] A. L. Goel, K. Okumoto, "Time-dependent error-detection rate model for software reliability and other performance measures", IEEE transactions on Reliability, Vol. 28, Issue 3, 1979, 206-211.
- [4] S. Yamada, M. Ohba, S. Osaki, "S-shaped reliability growth modeling for software error detection", IEEE Transactions on reliability Vol. 32, Issue 5, 1983, 475-484.
- [5] J. D. Musa, K. Okumoto, "A logarithmic Poisson execution time model for software reliability measurement". In Proceedings of the 7th international conference on Software engineering, IEEE Press, 1984, 230-238.

- [6] J. W. Paulson, G. Succi, A. Eberlein, "An empirical study of open-source and closed-source software products", *IEEE Transactions on Software Engineering*, Vol. 30, Issue 4, 2004, 246-256.
- [7] B. Rossi, B. Russo, G. Succi, "Modelling failures occurrences of open source software with reliability growth", *Open Source Software: New Horizons*, 2010, 268-280.
- [8] S. Yamada, M. Yamaguchi, "A method of statistical process control for successful open source software projects and its application to determining the development period", *International Journal of Reliability, Quality and Engineering*, Vol. 23, Issue 5, 2016.
- [9] X. Li, Y. F. Li, M. Xie, S. H. Ng, "Reliability analysis and optimal version-updating for open source software", *Information and Software Technology*, Vol. 53, Issue 9, 2011, 929-936.
- [10] Y. Tamura, S. Yamada, "A component-oriented reliability assessment method for open source software", *International Journal of Reliability, Quality and Safety Engineering*, Vol. 15, Issue 1, 2008, 33-53.
- [11] J. Yang, Y. Liu, M. Xie, M. Zhao, "Modeling and analysis of reliability of multi-release open source software incorporating both fault detection and correction processes", *Journal of Systems and Software*, Vol. 115, 2016, 102-110.
- [12] C. Rahmani, A.H. Azadmanesh, L. Najjar, "A Comparative Analysis of Open Source Software Reliability", *JSW*, Vol. 5, Issue 12, 2010, 1384-1394.
- [13] C. Rahmani, H. Siy, A. Azadmanesh, "An experimental analysis of open source software reliability", Department of Defense/Air Force Office of Scientific Research, 2009.
- [14] Y. Tamura, S. Yamada, "Comparison of software reliability assessment methods for open source software, in: *Parallel and Distributed Systems*", 2005. Proceedings. 11th International Conference on, Vol. 2, IEEE, 2005, 488-492.
- [15] Y. Zhou, J. Davis, "Open source software reliability model: an empirical approach", *ACM SIGSOFT Software Engineering Notes*, Vol. 30, 2005, 1-6.
- [16] P. Kapur, H. Pham, S. Anand, K. Yadav, "A unified approach for developing software reliability growth models in the presence of imperfect debugging and error generation", *IEEE Transactions on Reliability* Vol. 60, Issue 1, 2011, 331-340.
- [17] H. Pham, "A software cost model with imperfect debugging, random life cycle and penalty cost", *International Journal of Systems Science* Vol. 27, Issue 5, 1996, 455-463.
- [18] C. T. Lin, "Analyzing the effect of imperfect debugging on software fault detection and correction processes via a simulation framework", *Mathematical and Computer Modeling*, Vol. 54, Issue 11, 2011, 3046-3064.
- [19] E. M. Rogers, "Diffusion of innovations", Simon and Schuster, 2010.
- [20] F. M. Bass, "A new product growth for model consumer durables", *Management science*, Vol. 15, Issue 5, 1969, 215-227.
- [21] S. Yamada, K. Tokuno, S. Osaki, "Imperfect debugging models with fault introduction rate for software reliability assessment", *International Journal of Systems Science*, Vol. 23, Issue 12, 1992, 2241-2252.
- [22] H. Pham, X. Zhang, X.,"NHPP software reliability and cost models with testing coverage", *European Journal of Operational Research*, Vol. 145, Issue 2, 2003.
- [23] N. Gandhi, N. Gondwal, A.G. Aggarwal, A. Tandon, "Estimating Reliability for OSS: An approach with change-point in operational phase", In proceedings of 6th International Conference on Reliability, Infocom Technologies and Optimization, IEEE, 2017, 251-255.

Usability of an E-Commerce Website Using Information Mining and Artificial Intelligence

Pankaj Gupta¹, Bharat Bhushan Sagar²

Computer Science & Engineering, Birla Institute of Technology, Mesra: Ranchi, India
¹pgupta@bitmesra.ac.in, ²drbbsagar@gmail.com

Abstract—Everyday a number of people are launching new websites of which many are e-Commerce websites. E-Commerce website means business and to have business they have to be useful to the customer. So, it is very important for the e-Commerce site to get the trust of the customer. Or in other words the trustworthiness of an e-Commerce site is very important to have business. Usability can be increased by enhancing trust. How much a customer trusts a website becomes a fundamental question which beckons for an answer. More a customer trusts a website, more he will be using the website. This article includes a study on website usability, trust evaluation and how to improve trust to improve the usability of a website. Here we find the different factors that affect the quality of an e-commerce site. Based on these factors the trustworthiness of an e-commerce site is calculated using Matlab.

Index Terms—E-commerce, Website, Information Mining, Usability, Trust, Artificial Intelligence, Matlab.

I. INTRODUCTION

THE NATURE of a site can be surveyed in various ways. To date, investigations of sites have concentrated on site ease of use. So, usability is a very important concept. Ease of use is a subjective rating of how simple a site is to utilize, and on account of web based business, how likely it is that a potential client will be changed over to a genuine one. Usability can be increased by enhancing trust [2]. An estimation of ease of use comprises of a few unique variables, however comes down to how well a site gives its guests what they are truly searching for. So, usability of a site simply means whether a customer is finding a site to be useful or not.

1. *Factors That Decides Website Usability* [3]:
 - a. How individuals visit destinations?
 - b. Why and when they take off?
2. Simplicity and highlight decision
3. Flow
4. Feedback
5. Navigation and data engineering
 - a. Category and connection names
 - b. Link plan

- c. Click-capacity and saw affordance
 - d. Persistent route
 - e. Pop-ups
 - f. Opening new program windows
6. Page format
7. **Enhancing believability and trust**
8. Loyalty: Enticing clients to return
 - a. Registration
 - b. Value-included highlights for rehash clients
9. Information rummaging
 - a. User consideration
 - b. Scrolling
 - c. Information fragrance
10. Search
 - a. Query strings
 - b. How individuals utilize internet searcher comes about pages (SERP)
 - c. Improving site-particular pursuit
11. Content ease of use
 - a. How clients read on the web?
 - b. Writing for the Web
 - c. Low-education clients
 - d. Presentation of content: clarity and coherence

This article includes a study on website usability, trust evaluation and how to improve trust to improve the usability of a website [1].

Acceptance Leads To Usability: One approach to see whether a web based business webpage will be acknowledged by a client is to apply Technology Acceptance Model (TAM). This model expresses that the accomplishment of a framework can be controlled by client acknowledgment of the framework measured by two factors: the Perceived Usefulness (PU) of the framework and Perceived usability

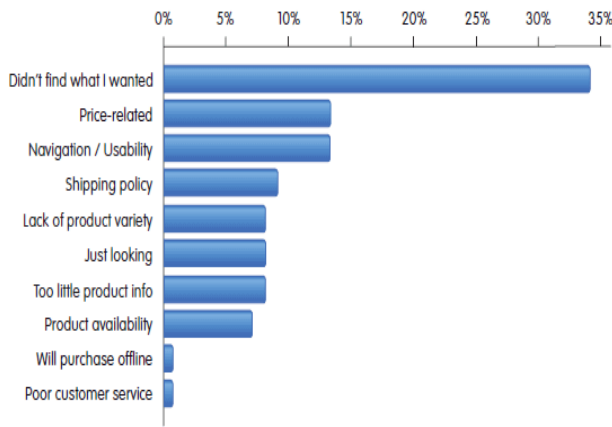


Figure 1: Why a Potential Customer may not find a Web-Site to be Useful.

(PEOU) of the framework. As indicated by this model, a client's recognition about a frameworks value and usability brings about a goal to utilize [6].

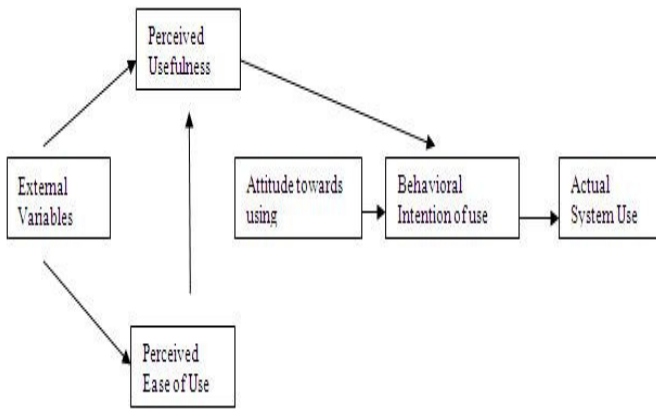


Figure 2: Acceptance Model

The TAM has been tried in numerous experimental examinations and observed to be very legitimate and dependable and is generally referenced. So, trust is a factor which determines whether a customer will like an e-commerce site or not. More a person trusts a site, more he will visit and use that particular site. If we can increase the trust of a person the usability of the site increases.

For the purpose of my analysis I have chosen one variable, viz., TRUST.

Trust: Trust is a complicated subject identified with confidence in trustworthiness, trustfulness, skill, and unwavering quality of an element. In McKnight et al's. "The Meanings of Trust", the most substantial parts of trust will be trust conduct and confide in aim. With regards to web based business, trust is generally determined as far as a brief connection between a buyer and a merchant or item. In the present online business condition, the administration of confide in

requirements to deal with many factors over numerous sites or spaces. Beth et al. classify the between area trust connections into two classes: coordinate trust and prescribed trust. In view of the desire for an element having the capacity to complete an undertaking, the framework can ascertain the likelihood of whether the element will finish the assignment in light of positive and negative involvement, measure the reliability utilizing this likelihood, and make a recipe for computing a number estimation of the dependability with an arrangement of induction and mix rules. However, this instrument rearranges genuine by displaying reliability construct just in light of likelihood, and compares the subjectivity and vulnerability to the haphazardness. In the meantime, it utilizes the mean estimation of numerous wellsprings of dependability as the pointer of the total trust and last trust esteem number, which discards conceivable weights on each put stock in source [9-11].

Fuzziness in Trust: The trust connections in multi-space web based business applications are difficult to evaluate because of included vulnerability. In the event that a trust relationship depends upon a subjective judgment in view of circuitous data, it will be extremely questionable and any operations identified with that trust relationship may cause sudden outcomes. Fluffy rationale is an appropriate approach to speak to vulnerabilities, particularly when they should be taken care of quantitatively. Two preferences of utilizing fluffy rationale to evaluate vulnerability in trust administration are: (1) Fuzzy induction is fit for evaluating loose information or vulnerability in measuring diverse levels of trust. (2) Different participation capacities and surmising tenets could be created for various put stock seeing someone, without changing the fluffy induction motors and implementation instruments. L. Zadeh first presented fluffy rationale in the advancement of the hypothesis of fluffy sets. The hypothesis of fluffy rationale broadens the metaphysics of scientific research to a composite that use quality and amount and contains certain fluffiness. We attempt to unravel the issues related with vulnerability in trust administration utilizing fluffy rationale [4], [7] [14].

How To Measure Trust: Simple variables like age, gender and income can each be measured by a single question. Complex phenomena, such as the degree of trust an individual has in an organization, cannot be measured through one question. Such constructs should be probed through a number of different questions to ensure that the researcher ends up with a meaningful response. Asking donors whether they trust an organization through simple yes / no question is meaningless. What does it mean to trust a nonprofit organization and is not it in any case a matter of degree. The fig. includes a group of questions that might be used to address this issue more effectively. If one measure of trust is required, taking the answers to these questions would be an appropriate [7], [18].

Sample Questions for measuring Trust	Low-degree Trust	High-degree Trust
	1 2 3 4 5 6 7	8 9 10

The test of building a trust display is the means by which to viably assess trust. Two vital inquiries to be addressed are what elements to survey for confide in assessment and how to utilize these components in assessment. On the Internet, clients ordinarily see higher hazard contrasted with a regular shopping condition, because of separation, virtual character, and absence of control. In this manner trust is the preparatory condition to purchasers' web based business interest.

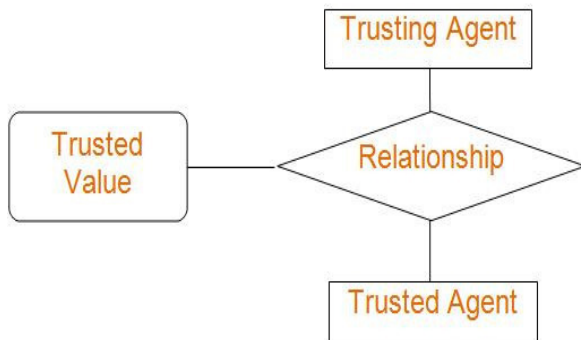


Figure 3: Trust Model

II. RELATED WORK

Ease of use is a subjective rating of the fact that it is so natural to utilize a site, and on account of web based business, how likely it is that a potential client will be changed over to a genuine one. An estimation of ease of use comprises of a few distinct components, yet comes down to how well a site gives its guests what they are truly searching for. So, usability of a site simply means whether a customer is finding a site to be useful or not.

III. METHODOLOGY OF MY WORK

I have used a survey analysis to obtain feed back of an e-commerce site. Based on the feedback results I have defined the membership functions of the factors. The results have been taken as input to the Matlab file and the output defined trustworthiness of the website [19].

Work Flow:

When we apply TAM to a particular site we can as well find out whether the user will accept that site or not. Various factors define a site and any site can be well described by the following factors (the list can be further expanded) [13].

Accuracy, availability, being-up-to-date, credibility, usefulness, durability, reliability, security.

According to TAM some of these factors will come under Perceived Ease of Use (PEU) and some will come under Perceived Usefulness [21].

Summary of Models of Trust in E-Commerce:

Model Of Trust	Trust Attributes
Trust model for Online Exchange [1]	Precision, accessibility, being-progressive, culmination, being unprejudiced, validity, excitement, convenience, strength, dependability, band value, temporariness, quality, assortment, customization, intensity and accessibility, straightforwardness, estimating and installment choice, monetary arranging, deals related administration (discount approach, after-deals and so on), advancements, Delivery satisfaction, Quality of media transmission, interface plan and substance, security, reversibility, computerized declaration, open key-cryptography, genuineness, uprightness, secrecy, notoriety, accreditation, verification, endorsements, client group, lawful necessities and specialists, encounter, nature, hazard evaluation, protection, fulfillment and subjective appraisal, of reliability.
Towards Establishing Customer supplier Trust [18]	Personality of the business site, offers of free items or administrations, steady in business, criticism, get to, route, basic buy.
Model of Trust for e-Commerce System Design [20]	Business mark name, trusted outsider, visual computerization, format, simple to utilize, framework's unwavering quality, characterization plans, phrasing, data about item and administrations, security, protection, following exchange and post-buy administrations.
Internet Consumer Trust Model [15]	Understood, awful/great notoriety, business measure, greatest/littlest player in the market, individual mindful, store's conduct and administrations.

Groups	Factors
PU	Accuracy, Being-up-to-date, Usefulness, Security, Credibility
PEU	Availability, Durability, Reliability

With respect to a site what these factors mean [12]

Factors	What it means with respect to a site
Accuracy	How accurately and effectively the site is designed so that it can process web orders and provide the highest possible level of customer service.
Availability	High availability means an user do not even have to worry about the site being going down.
Being-up-to-date	The site should always stay up to date with the latest happenings and development.
Usefulness	How the information provided by the site is useful to a user.
Durability	The site should able tp perform or compete over a long period as by avoiding and overcoming damage.

Reliability	A site is said to be reliable if it guarantees its proper service even at peak season/load.
Security	How secure a person feels while using sensible information like credit card details and bank details etc.
Credibility	How a customer feels over the long run after repeated use of the website.

How questioner is develop:

The questioner is developed with the purpose of evaluating a website on the basis of factors as above on a scale 1 to 5 (A *Linker scale* is a scale commonly involved in research that employs questioners) 1 meaning strongly disagree and 5 means highly agree and intermediate numbers 2 to 4 expressing a range between these two extremities.

The design of the questioner is consistent with the quality norms of the website as well as the representative population who are expected to evaluate/use the particular website [16].

Based on the above factors a questionnaire is prepared which serves as a visitor's feed-back form.

Figure 4: Feed Back Form

Fuzzification: The factors which are being used to evaluate the website are highly subjective in nature, the result do not lead to a crisp evaluation but rather fuzzy in nature [17]. For particular factor say availability what we get from the screen is 1, 2, 3, 4, 5 depending upon on which circle the visitor has put a tick mark. Next step is Fuzzification of the input.

CRISP VALUE	FUZZY VALUE
5, Strongly Agree	Very high
4	High
3	Average
2	Fair
1 Strongly Disagree,	Low

Here we will use Fuzzy Logic Tool Box of Matlab to represent the data which we have collected. Since the data is fuzzy in nature and Matlab is a very powerful technical language for mathematical reasoning.

Matlab Coding:

data =

```
[ 3  5  3  3  4  2
   4  4  3  4  5  4
   2  3  3  3  4  2
   5  4  4  4  2  3
   3  5  4  5  5  3
   5  3  4  4  4  2
   4  4  5  3  4  1
   2  3  4  5  5  5
   3  3  4  3  3  2
   3  4  5  5  4  1 ]
```

```
numMFs = [5 5 5 5 5] ;
```

```
mfType = char('trimf','trimf','trimf','trimf','trimf');
```

```
fismat = genfis1(data,numMFs,mfType);
```

```
[x,mf] = plotmf(fismat,'input',1);
```

```
subplot(3,2,1), plot(x,mf);
```

```
xlabel('accuracy');
```

```
[x,mf] = plotmf(fismat,'input',2);
```

```
subplot(3,2,2), plot(x,mf);
```

```
xlabel('availability');
```

```
[x,mf] = plotmf(fismat,'input',3);
```

```
subplot(3,2,3), plot(x,mf);
```

```
xlabel('credibility');
```

```
[x,mf] = plotmf(fismat,'input',4);
```

```
subplot(3,2,4), plot(x,mf);
```

```
xlabel('reliability');
```

```
[x,mf] = plotmf(fismat,'input',5);
```

```
subplot(3,2,5), plot(x,mf);
```

```
xlabel('usefulness');
```

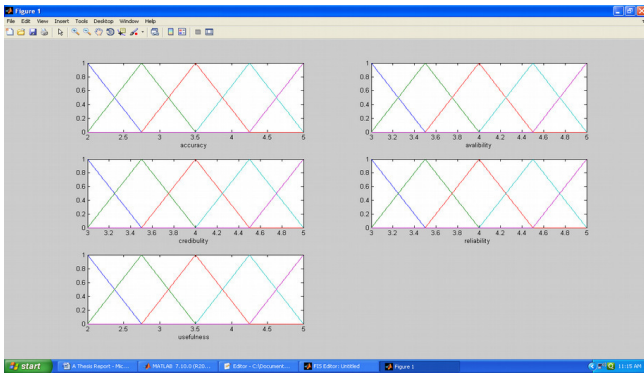


Figure 5: Fuzzification

Generation Input Membership Function:

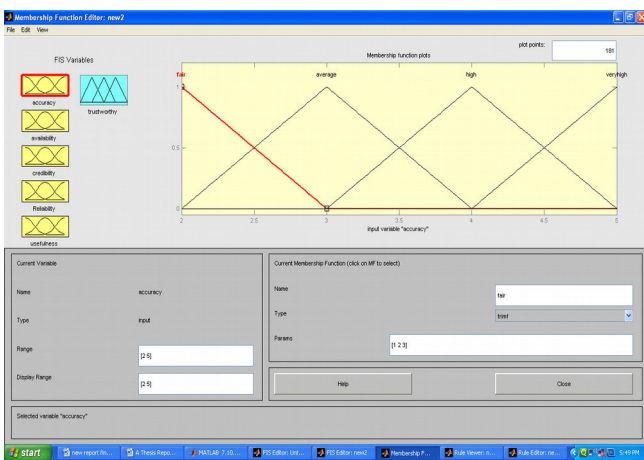


Figure 6: Membership Function for Input Variable 'Accuracy'

Similarly we will generate the membership function for other variables and the output we get is shown in figure 7.

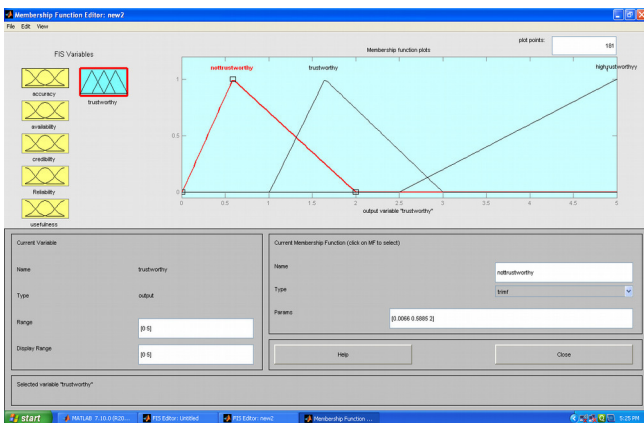


Figure 7: Output

The rules are formed in the following way:

RULE:

1. If accuracy is very-high and availability is very-high and credibility is very-high and reliability is very-high and use fullness is very high then highly trustworthy.

RESULT:

Accur acy	Availabi lity	Credibi lity	Reliabi lity	Usefuln ess	Trust wor thiness
2.5	2.5	2.5	2.5	2.5	2.5
4.0	4.0	4.0	4	4	2.5
4.92	4.85	4.92	4.92	4.92	4.18

RULE

1. If accuracy is very-high and availability is very-high and credibility is very-high and reliability is very-high and usefulness is very high then highly trustworthy.
2. If accuracy is high and availability is high and credibility is high and reliability is high and use fullness is high then trustworthy.
3. If accuracy is average and availability is average and credibility is average and reliability is average and usefulness is average then not trustworthy.

RESULT

Accur acy	Availabi lity	Credibi lity	Reliabi lity	Usefuln ess	Trustwort hiness
2.5	2.5	2.5	2.5	2.5	.911
4.0	4.0	4.0	4	4	1.89
4.92	4.85	4.92	4.92	4.92	3.96

RULE

1. If accuracy is very-high and availability is very-high and credibility is very-high and reliability is very-high and usefulness is very high then highly trustworthy.
2. If accuracy is high and availability is high and credibility is high and reliability is high and use fullness is high then trustworthy.
3. If accuracy is average and availability is average and credibility is average and reliability is average and usefulness is average then not trustworthy.
4. If accuracy is fair or availability is average or credibility is average or reliability is average or usefulness is average then not trustworthy.

RESULT

Accu racy	Availa bility	Credi bility	Relia bility	Usefu lness	Trustwort hiness
2.5	2.5	2.5	2.5	2.5	.865
4.0	4.0	4.0	4	4	1.37
4.92	4.85	4.92	4.92	4.92	3.96

IV. CONCLUSION

In this paper I have found out a method to find the trustworthiness and thus the usability of an e-commerce website. I have emphasized on the fact that the usability of an e-commerce site can be measured by measuring the trustworthiness of that site. To measure the trustworthiness of an e-commerce site I first made a survey of that particular site. The survey included various factors which described the site properly. Based on this survey I have made various rules by using fuzzy toolbox of Matlab. These rules helped me to calculate the trustworthiness of that particular e-commerce site. I also found out that trustworthiness of a web-site can be improved by improving the various factors which describe the site. This method can be used first to measure the trustworthiness and then to take effective periodic steps to maintain as well as increase the trustworthiness and thus usability of any e-commerce site.

Limitation: The limitation of the present work is that it is based on a sample data of a particular site only. It should include different sites because user of a particular site can be biased about the site. Web is dynamic in nature and more and more user using it. So, the parameter should be updated regularly.

Future Scope: In future we can design a rule based expert system which will provide expert-level solution to find the trustworthiness of an e-commerce site. The system will be flexible enough to accommodate future knowledge.

REFERENCES

- [1] M A Islam, F Ahmed, Y Borodin, J Mahmud and I. V. Ramakrishnan, "Improving Accessibility of Transaction-centric Web Objects", SIAM International Conference on Data Mining (SDM), 2010.
- [2] M Ahmed, J Kaysar and N Rahman, "Design Approaches to Enhance Usability for E-Commerce Sites", International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, 2011.
- [3] <http://www.networksolutions.com/education/what-is-ecommerce/>
- [4] Z Wu and H Wu, "An Agent based Fuzzy Recommendation system using shoppers Preferences for E-commerce Applications", International Journal Uncertainty Fuzzy Knowledge Based System, Vol. 18, pp. 471, 2010.
- [5] Book:: MATLAB and Its Applications in Engineering, R K Bansal, A K Goel, M Sharma, Pearson Education India, 2009.
- [6] Z Wu and Yu Zhou, "Service Trustworthiness Evaluation Using Neural Network and Fuzzy Logic", Services Computing (SCC) 2016 IEEE International Conference on, pp. 563-570, 2016.
- [7] G Acampora, A Castiglione and A Vitiello, "A fuzzy logic based reputation system for E-markets", Fuzzy Systems (FUZZ-IEEE) 2014 IEEE International Conference on, pp. 865-872, 2014.
- [8] Z Li, H Lin, C Lu and Y Zhang, "Managing Data for Evaluating Trust in Unstructured Peer-to-Peer Networks", Advances in Grid and Pervasive Computing. GPC 2008. Lecture Notes in Computer Science, Vol. 5036, Springer, Berlin, Heidelberg.
- [9] Book:: Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project, B G Buchanan, The Addison-Wesley series in artificial intelligence, 1st Edition.
- [10] T Oliveira, M Alinho, P Rita and G Dhillon, "Modelling and testing consumer trust dimensions in e-commerce", Computers in Human Behavior Vol. 71, pp. 153-164, 2017.
- [11] J Mahmud, Y Borodin, M A Islam and F Ahmed, "Bridging the Web Accessibility", Electronic Notes in Theoretical Computer Science, Elsevier, 2009.
- [12] <http://en.wikipedia.org/wiki/Ethnography>.
- [13] S Trewin, B Cragun, C Swart, J Brezin and J Richards, "Accessibility challenges and tool features: an IBM Web developer perspective", International Cross Disciplinary Conference on Web Accessibility (W4A), ACM, 2010.
- [14] Book:: Human-Computer Interaction, A Dix, J Finlay, G Abowd, and R Beale. Prentice Hall, 2004.
- [15] M A Islam, F Ahmed, Y Borodin, J Mahmud and I V Ramakrishnan, "Improving Accessibility of Transaction-centric Web Objects", SIAM International Conference on Data Mining (SDM), 2010.
- [16] <http://www.usabilityinstitute.com/resources/stateofheart.htm>
- [17] J Offutt, "Quality attributes of Web software applications", IEEE Software: Special Issue on Software Engineering of Internet Software, pp. 25-32, 2002.
- [18] F Adrian, I Emilio and A Silvia, "Usability evaluation methods for the web: A systematic mapping study", Information and Software Technology, pp. 789-817, 2011.
- [19] A Salim and B Abdelmajid' "An evaluation study of the navigation and search systems on two academic websites and Google", The International Information & Library Review, pp. 50-61, 2010.
- [20] H Hartson, T Andre and R Williges, "Criteria for evaluating usability evaluation methods", International journal of human-computer interaction, pp. 373-410, 2001.
- [21] F Lia and Y Lib, "Usability evaluation of e-commerce on B2C websites in China", Elsevier, 2011.

E-Assessment Tools for Programming Languages: A Review

Sugandha Gupta¹, Anamika Gupta²

¹Sri Guru Gobind Singh College of Commerce, University of Delhi

²Shaheed Sukhdev College of Business Studies, University of Delhi

¹sugandhagupta@sggsc.ac.in, ²anamikargupta@sscbsdu.ac.in

Abstract—Continuous Evaluation and feedback not only helps in improving learning of a student, but also acts as a constant motivator to put in more efforts. But then, feedback and assessment are very difficult and time consuming in practice. Thus, automating the entire system of assessment, evaluation and feedback will be highly beneficial. But, building such tools for all courses is yet not feasible. However, e-assessment tools for programming courses in Computer Science discipline can be developed. In this paper, we review various grading techniques used by these tools to assess a student's programming assignment. Further, this paper discusses various types and features of tools according to which an appropriate tool should be selected. And, in the end, we will be highlighting the extent to which students and instructors are actually benefited by these tools.

Index Terms—E-Assessment; Automated Tool; Programming Languages; Static Analysis; Dynamic Analysis; LMS Based Tools

I. INTRODUCTION

ASSESSMENT plays an integral role in any educational system. It helps in assessing the knowledge acquired by a learner at any point of time [1, 8]. However, with substantial increase in number of students getting enrolled at Universities in either conventional or distance learning environment, usual ways of assessing a student are becoming inadequate in terms of both time and effort [5].

Thus, the notion of e-Assessment was introduced to overcome all the inadequacies of traditional pen and paper assessment modes. The accessibility to automated tools [8, 9] to assist the evaluation of students' work and providing students with appropriate and timely feedback can really help in motivating the student to learn with more focus [1, 24]. The feedback provided by the tool can also be further enhanced by using some techniques to direct a student to engage in self learning aiming at improvement

Although, building such tools for evaluating students' work for all kind of courses is not yet feasible. However, courses that involve formal language of expression i.e. where every instruction can be expressed in terms of a language can be automated, such as, programming courses [8, 36] in Computer Science discipline.

Programming is the core of Computer Science discipline. But, today programming [36] is not restricted to only Com-

puter Science students. It is becoming an important part even in various other academic disciplines. The best way to learn a programming language is to practice a large set of problems. The skills will be improved through practicing many programming exercises. As a student, enormous satisfaction is attained if all the practiced problems are evaluated by some expert highlighting all insights. Getting a detailed and timely feedback serves as a motivating force for a student. This will not only help them to understand their mistakes, but also to improve their skills. But as an evaluator, it's very difficult to evaluate each and every problem for all students. Thus, building an E-assessment tool to assess a student's program and providing an appropriate feedback is much needed [1, 3].

The purpose of an e-assessment tool [8, 10] for programming languages is to improve programming skills in the students, paying exceptional attention to beginner students [1, 15]. For a student to practice more and more problems, immediate and thorough feedback plays an important role.

Both instructors and students are benefitted by these tools. As of instructor's point of sight, assessment tools can be used to automate mundane tasks that incur a lot of time and effort manually [5]. For example, using these tools, an instructor can automatically correct assignments, grade students and provide a timely feedback [1, 3, 8]. Also, the assessment results are trustworthy and much easier to comprehend [3].

As of the students' point of sight, the timely feedback imparted by the tools push them to practice more and more programming problems. Thus, their programming skills are also improved. Also, students are motivated to constantly improvise their assignments by re-submissions until an appropriate solution is obtained [4, 9].

In this paper, we focus on reviewing certain e-assessment tools available for grading a programming assignment. Further, we also discuss various grading techniques and classification schemes for a tool. Then in the end of this paper, we discuss the effectiveness of e-assessment tools with respect to students and instructors.

This paper is structured into five sections. Section II discusses main approaches used by a tool to assess a program. Section III reviews categorization of tools according to certain parameters. Section IV mentions few readily used e-as-

assessment tools. Section V discusses research questions that helps to identify how important are e-assessment tools to students and instructors.

II. MAIN APPROACHES FOR ASSESSMENT

For automatic assessment of a programming code, two major approaches have been devised i.e. Static Analysis and Dynamic Analysis [2, 5].

(1) Static Analysis

It is performed by inspecting the source code of the program without executing it. In this method, structure of the program and content are observed [2, 3]. In the old days, this was the only means to assess a programming code. Students used to submit their programming assignment code, and teachers based their assessment on examining the code without actually executing it. ASSYST [13], CAP [15] and Espresso [7, 14] are few examples of system that only use static analysis. Within this approach, following methods have been characterized [4]:

- Programming-style analysis: According to this, a quality program is the one which is highly readable. For this, parameters like expressive variables names, use of constants, line spacing, indentation, comments, lesser global data etc are evaluated [3].
- Semantic-error discovery: These errors are discovered when a statement is syntactically correct, but leads to failure in program execution. Most common semantic errors for a beginner in programming are division by zero, infinite loops and terminating a loop header. Some semantic errors can even result in crashing of the system [2].
- Software metrics analysis: It is a function that produces a single output that can be interpreted as the measure of software quality. Metrics cover the size of program in terms of number of lines of code, cyclomatic complexity of the code and quality [2, 7].
- Keyword analysis: Instructor gives a list of keywords which he/she is looking for in the solution. This parameter deals with finding these keywords in the program to be assessed and matching it with those specified by the instructor [2, 5].

(2) Dynamic Analysis

Under this approach, a program is executed against many different test cases. And, the output obtained is equated against the anticipated output. [3, 6] If they are same, then the program is considered to be correct. TRY [7, 16] is an example of system that performs only dynamic analysis. Within dynamic analysis [3, 5, 6], two different methods of evaluating a program are available:

Black-Box approach: The entire program is executed, and its output is examined against different test cases. And, then the program is considered as

“correct” or “incorrect” based upon match or mismatch [2, 3]. Under this approach, the entire program is treated as an atomic entity.

Grey-Box approach: Under this approach, a program is divided into a set of functions. And, then each and every function is assessed separately. The final grading of the program is obtained by the amalgamation of the partial grades of these functions. Using this approach, a program with some function with error can also be graded [2].

In dynamic analysis, selection of test case against which a program will be executed is an important decision. Normally, the evaluator submits the test cases with the problem itself.

(3) Static Analysis v/s Dynamic Analysis

A program with syntax errors like a omitted semi-colon, mis-match in number of opening and closing parenthesis, etc. prevents the program to compile. And thus, such a program can't be executed using dynamic analysis. In this case, a grade of zero is awarded to the program. On the other hand, a program is assessed even when it fails to execute in case of static analysis [3, 6].

In case of dynamic analysis, student just gets to know whether the program executed successfully and anticipated output is produced or not. So, it is the student's responsibility to find out all the errors and fix them. Whereas in static analysis, student gets to know about all the errors present in the program [2, 3].

Dynamic analysis approach is not suitable for the set of programming problems which can have multiple solutions, because to assess such a code, instructor will have to submit an exhaustive set of test cases. This limitation is not applicable for static analysis approach.

There are some tools which combine both the strategies to grade a program. AutoLEP [10] and Auto Grader [19] are tools which combines static and dynamic analysis approach to grade a program.

III. TOOLS REVIEW

Three classification schemes [5] have been identified for the tools: (i) assessment-type, (ii) approach, and (iii) specialty.

(1) *Classification by Assessment-Type:* Under this, the assessments tools are categorized in three types based upon how the assessment process is carried:

1. Manual Assessment: Assessment of the programming assignment is done “manually by the instructor with assistance of the tool”[3].
2. Automatic Assessment: Assessment of a programming assignment is done automatically by the tool. But, the instructor will have to clearly state the parameters on which the program code should be assessed before the assessment process[3, 5].
3. Semi-Automatic Assessment: Assessment is performed automatically by the tool, but manual inspection of

the programming code is required by the instructor [3].

(2) *Classification by Approach:* Under this, three approaches were identified based on how the assessment process is commenced.

1. **Instructor Centered Approach:** After the submissions of programming code by student, assessment process is initiated by the instructor. The code is assessed by the tool and results are forwarded to the instructor for review. After review, final results are formulated by the instructor and given to students.
 - **Student Centered Approach:** The assessment process is started by the student. Programming code submitted by the student should abide certain specifications stated by the instructor, which will be used by the tool for assessment. After the completion of the assessment process, assessment results are forwarded to both the instructor and students [5].
 - **Hybrid Approach:** These tools incorporate strong points of both the abovementioned approaches. Under this approach, three different strategies were identified:
 - Preliminary validation [5]: Instructor starts the assessment process, but a preliminary evaluation of the solution is carried out by the tool. BOSS [9], OTO [30] and PASS [7] are few tools that implement this strategy.
 - Partial feedback [5]: Tool assesses the code partially. Instructor manually evaluates the final code. WebCAT [8] uses this strategy for assessment of programming codes.
 - Instructor review [5]: A feedback is provided to the student after evaluation of the code by the tool. But, that feedback is not final, as the result of tool's assessment is accessible and updatable by the instructor. The instructor's decision would be final and would be intimated to the student. MOOSHAK [11] implements this approach.
- (3) *Classification by Specialization:* Under this, the assessment tools are classified based on their ability to perform extra functions other than assessment of a programming code.
- **Tools Specialized in Competitions [5]:** The assessment tool is used to evaluate codes of all students using specific test cases. As a result of the evaluation process, it specifies out of all the codes, which all codes are acceptable. ONLINE JUDGE [22, 23] implements this strategy.
 - **Tools Specialized in Quizzes [5]:** A set of questions are formulated which should be answered by the students. And based on the result of evaluation of these answers, winners are selected. This strategy is also used in recruitments. AutoLEP [10] uses this approach for assessment.
 - **Tools Specialized in Software-Testing [5]:** These tools are used to automate the testing process, thus reducing the dependence on manual testing. Prog Test [17] is used for software testing purposes.

A fourth classification is also possible, i.e. classification on the basis of usage. An e-assessment tool can either be used as a standalone tool, or in integration with some IDE or Learning Management System (LMS). A standalone e-assessment tool can be deployed on either instructor's or student's system and can be used for grading. Examples of tools that fall in this category are AutoLEP [10] and INGINIOUS [20]. An IDE integrated e-assessment tool can be used with an IDE specific to a programming language. WebCAT [8] and Petcha [18] are integrated with IDE to grade programming assignments in java. A LMS integrated e-assessment tool is used in integration with a LMS (majorly edX, Audacity, Coursera) to grade programming assignments.

IV. DESCRIPTION OF TOOLS

Based on the specifications and characteristics discussed in the previous two sections, below are few readily used E-assessment tools:

(1) *WebCAT:* Edwards and Perez-Quinones developed Web-CAT to assess student's code and test cases [8]. The students need to submit their own test cases with their codes, enabling them to demonstrate the correctness and validity of their own programs. It uses dynamic analysis and partial Feedback [5].

(2) *BOSS:* It is a tool developed by Luck & Joy to assist the online submission and successive processing of programming assignments. This tool can be used by in two individuals: first, by students to obtain the feedback for their submitted programs. Second, tool can be used by instructors in order to grade student submissions [9]. It uses static analysis and preliminary validation [5] (Hybrid Approach).

(3) *AutoLEP:* AutoLEP [10] is a standalone tool that uses a combination of static and dynamic analysis approach to grade the programming assignments. That is, it not only evaluates the output of the program but also deals with the construct of the program.

(4) *Mooshak:* It is a "web-based learning system" initially designed to conduct programming contests over the Internet. It provides specific interface to every user according to its profile, i.e. it will be different for student administrator, user, guest user, instructor and student [11].

(5) *CourseMarker:* CourseMarker was developed at the University of Nottingham as a successor to the Ceilidh in 2003. Students use CourseMarker client on their system and login into their account. They select the course, topic and exercise they want to complete. Students write the program in response to the problem description provided to them and submit it. Program's output is matched against the expected output of the program and results are shown to students. Students can re-solve the problem if it is not satisfactory, upon teacher's permission [12].

(6) *ASSYST:* It practices two user views: one corresponding to student and other corresponding to the instructor. Using the student view, a student submits a programming code

for subsequent grading; and using the instructor view, instructor guides the assessment process [13].

(7) *Expresso*: This tool is used for an introductory Java programming course and it processes a program in multiple passes. The first pass accepts programmer's code as an input and removes comments and stores the resulting characters in a vector. In the second pass, white spaces are removed and the file is tokenized, and the result is stored as a vector of words. Punctuation and white space are used as delimiters to identify words. Mistakes are detected (if any) and the appropriate error messages are printed in the final pass. These messages can be used to fix the program [14].

(8) *CAP*: The "Code Analyzer for Pascal (CAP)" [15] analyzes syntactical, style and logical errors of a program and gives a feedback to the student stating the same. The feedback indicates the problem with the code and a measure to correct that.

(9) *TRY*: TRY [16] system was developed for Unix operating system which test student program with a set of hidden test data.

(10) *Prog Test*: It is a web-based environment for imparting a feedback to the student after successful assessment of the programming assignment submitted by the student. Prog Test evaluates a student's submission on basis of: Instructor's code and test cases, Student's code and test cases. A student has to submit program as well as test cases used by him to test his own program. The instructor's program is a significant parameter used for assessing the student's program. [17].

(11) *Petcha*: It acts as an "automated Teaching Assistant in computer programming courses", helping students to learn programming more efficiently. Also, it imparts feedback to the student for the solutions of programming assignments submitted by the student [18].

(12) *Auto Grader*: It is a new tool designed at MIT to review faulty code and automatically provide feedback on how to fix it. It uses Error Models and Constraint-based synthesis to perform this [19].

(13) *INGInious*: INGInious [20] is a tool designed to automatically grade programming code submitted by the user. It can virtually cater a large set of programming languages

(14) *Curator*: It grades a program by a strict textual comparison of the solution provided by the instructor and the solution submitted by the student [21].

(15) *Online Judge*: It also textually compares the code submitted by the student and the instructor. It is used to grade programs in various languages (C, C++, Java). It is also for conducting quizzes.

(16) *HackerEarth Recruit*: It is an online skill assessment tool for conducting programming tests to assess developers. It saves the pain of going through hundreds of resumes, by automating the process of evaluating technical skills, which helps to quickly filter the competent candidates.

(17) *LMS Based Tools*: This category of tools can be used to assess a student in the course enrolled on the basis of his/her submissions during the course. Further, it can be used to detect plagiarisms, generate and evaluate test cases for a program.

V. ARE E-ASSESSMENT TOOLS USEFUL IN PROGRAMMING COURSES

Four research questions were formulated to measure the degree to which e-assessment tools have helped to students and instructors [25]:

RQ1. Have e-assessment tools laid a positive impact on student learning?

RQ2. According to students, have e-assessment tools improved their performance?

RQ3. According to instructors, have e-assessment tools improved their teaching experience?

RQ4. Is the result obtained by e-assessment tools precise and useful?

RQ1. Have e-assessment Tools laid a positive impact on student learning?

In 2003, Edwards [26] presented fascinating results when he changed the e-assessment tool in a junior level course on comparative languages, i.e. Curator was replaced by Web-CAT, demonstrating more timely submission of assignments along with test cases by the students. In 2003, Woit [27] collected data of students for "five consecutive years" comparing their performance on online tests with and without e-assessment tools. He concluded that online assessment gave a more precise indication of student knowledge. In 2005, Higgins [28] conducted an experiment in which "CourseMarker substituted Ceilidh at the University of Nottingham", thus increasing the passing percentage of students. Also in 2005, "Malmi [30] showed results from students using TRAKLA and TRAKLA2", in which final exam grades improved when students were allowed to resubmit their work. In 2011, Wang [31] showed that final grades of students using AutoLEP for grading were way better than grades produced without using any tool.

Considering all these facts, a positive impact was inferred with use of e-assessment tools on student performance. End-of-grades or final exam scores were major measures used to measure this.

RQ2. According to students, have e-assessment tools improved their performance?

In 2003, Edwards [26] proved that using WebCAT laid a positive impact using a 20-question survey for students. In 2005, Higgins [27] proved that over 75% of students' loved the flexibility to re-submit a programming assignment due to use of an e-assessment tool by carrying a survey to test the tool CourseMarker and indicated that. In 2009, Garcia-Mateos [32] presented Mooshak, and handed over the students a survey asking agreement or disagreement with re-

spect to the tool. 77% of the students specified that “they learn better with the new methodology than with the old one,” while 91% said that “if they could choose, they would follow the continuous evaluation methodology again.” Also in 2012, Brown [33] presented a survey on student’s opinion about the impact of using JUG automated assessment tool. Given the question “Did the automatically graded tests match your beliefs of the requirements?” the major chunk of students opted for the answer, “Sometimes.” But the question “did the reports from the automatic grader clarify how your code should behave?” urged to students answering “Often.”

Unconvincing results concerning student perceptions of e-assessment tools were observed. Students had a mixed reaction on this question; but a significant number of students indicated their dissatisfaction with the tools.

RQ3 According to instructors, have e-assessment tools improved their teaching experience?

In 1995, Schorsch [35] stated that “6 out of 12 instructors who used CAP” for grading assignments indicated that the tool helped them saving around ten hours effort incurred in grading a section earlier. In 2003, Venables [35] used SUBMIT (e-assessment tool) and stated that the feedback imparted by the tool was successful to provide answers to doubts of students which they had while solving the assignment. This feature of the tool helped by saving onto the class time that otherwise would have been spent in responding to students’ questions. In 2012, Queirós [36] claimed that “automated grading is better than manual grading in terms of efficiency, accuracy, and objectivity” as e-assessment tools eradicate favoritisms and other biased factors from the grading process, and submissions are noticeable at a greater pace.

Majority of the instructors have reported that initially a substantial amount of their time is incurred in learning the tool and making the students familiar with the tool, but once this has been done the tool saves a lot of their time which was earlier used for grading student’s code and giving them feedback.

RQ4. Is the result obtained by e-assessment tools precise and useful?

In 2005, Higgins [37] stated that grading performed by CourseMarker tool in one section of a course was at par with the assessment done by a teaching assistant in some other section of same course. Also in 2012, Taherkhani [38] revealed that AARI (e-assessment tool) was successful at recognizing the algorithms used by the students to perform sorting on integers for about 75% of the submissions. Further in 2014, Gaudencio [39] stated that instructors who manually graded the assignments also tend to agree more with the feedback provided by the tool in comparison to assessment provided by other instructors.

Thus, it can be inferred that e-assessment tools assist the instructors in conducting the assessment process in a positive manner.

REFERENCES

- [1] Matthiasdóttir, Ásrún & Arnalds, Hallgrímur. (2016). e-assessment: students' point of view. 369-374
- [2] S. M. Arifi, I. N. Abdellah, A. Zahi and R. Benabbou, "Automatic program assessment using static and dynamic analysis," 2015 Third World Conference on Complex Systems (WCCS), Marrakech, 2015, pp. 1-6.
- [3] Kirsti M Ala-Mutka (2005), 'A Survey of Automated Assessment', Approaches for Programming Assignments, Computer Science Education, 15:2, 83-102
- [4] Rahman, Khirulnizam Abd, and Md Jan Nordin. "A review on the static analysis approach in the automated programming assessment systems." In Proceedings of the national conference on programming, vol. 7. 2007.
- [5] D. M. Souza, K. R. Felizardo and E. F. Barbosa, "A Systematic Literature Review of Assessment Tools for Programming Assignments," 2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET), Dallas, TX, 2016, pp. 147-156.
- [6] Fonte, Daniela, Daniela da Cruz, Alda Lopes Gançarski, and Pedro Rangel Henriques. "A Flexible Dynamic System for Automatic Grading of Programming Exercises." In OASIS-Open Access Series in Informatics, vol. 29. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013.
- [7] Truong, Roe, Bancroft, "Static Analysis of Students' Java Programs", Sixth Australian Computing Education, Conference (ACE2004), Dunedin, New Zealand, Vol. 30.
- [8] S. H. Edwards and M. A. Perez-Quinones. Web-CAT: automatically grading programming assignments. In Proc. Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE), pages 328–328, 2008.
- [9] Mike Joy, Nathan Griffiths, and Russell Boyatt. 2005. The boss online submission and assessment system. J. Educ. Resour. Comput. 5, 3, Article 2 (September 2005).
- [10] W. Tiantian, S. Xiaohong, M. Peijun, W. Yuying and W. Kuanquan, "AutoLEP: An Automated Learning and Examination System for Programming and its Application in Programming Course," 2009 First International Workshop on Education Technology and Computer Science, Wuhan, Hubei, 2009, pp. 43-46.
- [11] José Luis Fernández Alemán, 'Automated Assessment in a Programming Tools Course IEEE TRANSACTIONS ON EDUCATION, VOL. 54, NO. 4, NOVEMBER 2011
- [12] Higgins, C., Hergazy, T., Symeonidis, P., and Tsinsifas, A. The CourseMarker CBA System: Improvements over Ceilidh, Education and Information Technologies, 8(3), 2003, pp. 287– 30.
- [13] David Jackson and Michelle Usher. 1997. Grading student programs using ASSYST. In Proceedings of the twenty-eighth SIGCSE technical symposium on Computer science education (SIGCSE '97), James E. Miller (Ed.). ACM, New York, NY, USA, 335-339.
- [14] Maria Hristova, Ananya Misra, Megan Rutter, and Rebecca Mercuri. 2003. Identifying and correcting Java programming errors for introductory computer science students. In Proceedings of the 34th SIGCSE technical symposium on Computer science education (SIGCSE '03). ACM, New York, NY, USA, 153-156.
- [15] Tom Schorsch. 1995. CAP: an automated self-assessment tool to check Pascal programs for syntax, logic and style errors. In Proceedings of the twenty-sixth SIGCSE technical symposium on Computer science education (SIGCSE '95), Curt M. White, James E. Miller, and Judy Gersting (Eds.). ACM, New York, NY, USA, 168-172.
- [16] Yusofa, Zinb, Adnana, 'Java Programming Assessment Tool for Assignment Module in Moodle E-learning System', International Conference on Teaching and Learning in Higher Education (ICTLHE2012) in conjunction with RCEE & RHED 2012, 1877-0428 © 2012 Published by Elsevier Ltd. Selection and/or peer-review under

- responsibility of Centre of Engineering Education, Universiti Teknologi Malaysia.
- [17] D. M. de Souza, J. C. Maldonado and E. F. Barbosa. PROGTEST: An Environment for the Submission and Evaluation of Programming Assignments based on Testing Activities. Software Engineering Education and Training (CSEET), 2011 24th IEEE-CS Conference.
- [18] Ricardo Alexandre Peixoto Queirós and José Paulo Leal. 2012. PETCHA: a programming exercises teaching assistant. In Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education (ITiCSE '12). ACM, New York, NY, USA, 192-197
- [19] Auto Grader: <http://www.csail.mit.edu/node/1886>
- [20] Guillaume Derval, Anthony Gego, Pierre Reinbold, Benjamin Frantzen and Peter Van Roy. Automatic grading of programming exercises in a MOOC using the INGINIOUS platform. Proceedings of the European MOOC Stakeholder Summit 2015.
- [21] Virginia Polytechnic Institute and State University. Curator: an electronic submission management environment. <http://ei.cs.vt.edu/~eags/Curator.html>.
- [22] A. Kurnia, A. Lim, and B. Cheang. OnLine Judge. Computers & Education, 36(4):299–315, May 2001.
- [23] B. Cheang, A. Kurnia, A. Lim, and W.-C. Oon. On automated grading of programming assignments in an academic institution. Computers & Education, 41(2):121–131, September 2003
- [24] C. Douce, D. Livingstone, and J. Orwell, “Automatic test-based assessment of programming: A review,” J. Educ. Resources Comput., vol. 5, no. 3, pp. 1–13, 2005.
- [25] Pettit, R. S., & Homer, J. D., & Holcomb, K. M., & Simone, N., & Mengel, S. A. (2015, June), Are Automated Assessment Tools Helpful in Programming Courses? Paper presented at 2015 ASEE Annual Conference & Exposition, Seattle, Washington. 10.18260/p.23569
- [26] Stephen H. Edwards. 2003. Improving student performance by evaluating how well students test their own programs. J. Educ. Resour. Comput. 3, 3, Article 1 (September 2003).
- [27] Denise Woit and David Mason. 2003. Effectiveness of online assessment. In Proceedings of the 34th SIGCSE technical symposium on Computer science education (SIGCSE '03). ACM, New York, NY, USA, 137-141.
- [28] Colin A. Higgins, Geoffrey Gray, Pavlos Symeonidis, and Athanasios Tsintsifas. 2005. Automated assessment and experiences of teaching programming. J. Educ. Resour. Comput. 5, 3, Article 5 (September 2005).
- [29] Amruth N. Kumar. 2005. Generation of problems, answers, grade, and feedback---case study of a fully automated tutor. J. Educ. Resour. Comput. 5, 3, Article 3 (September 2005).
- [30] Lauri Malmi, Ville Karavirta, Ari Korhonen, and Jussi Nikander. 2005. Experiences on automatically assessed algorithm simulation exercises with different resubmission policies. J. Educ. Resour. Comput. 5, 3, Article 7 (September 2005).
- [31] Tiantian Wang, Xiaohong Su, Peijun Ma, Yuying Wang, and Kuanquan Wang. 2011. Ability-training-oriented automated assessment in introductory programming course. Comput. Educ. 56, 1 (January 2011), 220-226.
- [32] García-Mateos, Ginés & Fernández-Alemán, José. (2009). A Course on Algorithms and Data Structures Using On-line Judging ABSTRACT. ACM SIGCSE Bulletin. 41. 45-49.
- [33] Christopher Brown, Robert Pastel, Bill Siever, and John Earnest., JUG: a JUnit generation, time complexity analysis and reporting tool to streamline grading, 17th ACM annual conference on Innovation and technology in computer science education (ITiCSE '12).
- [34] Tom Schorsch, Cap: An Automated Self-Assessment Tool To Check Pascal Programs For Syntax, Logic And Style Errors, SIGSCE 1995.
- [35] Anne Venables and Liz Haywood. 2003. Programming students NEED instant feedback, Fifth Australasian conference on Computing education - Volume 20 (ACE '03), Tony Greening and Raymond Lister (Eds.), Vol. 20
- [36] Queirós, R., & Leal, J. P., “Programming exercises evaluation systems: An interoperability survey”, 4th international conference on computer supported education, 2012, (pp. 83–90).
- [37] Colin A. Higgins, Geoffrey Gray, Pavlos Symeonidis, and Athanasios Tsintsifas., Automated assessment and experiences of teaching programming. Journal on Educational Resources in Computing (JERIC), 2005.
- [38] Ahmad Taherkhani, Ari Korhonen, and Lauri Malmi. 2012. Automatic recognition of students' sorting algorithm implementations in a data structures and algorithms course, 12th Koli Calling International Conference on Computing Education Research.
- [39] Matheus Gaudencio, Ayla Dantas, and Dalton D. S. Guerrero. Can Computers Compare Student Code Solutions as Well as Teachers, SIGCSE.

Importance of Text Data Preprocessing & Implementation in RapidMiner

Vaishali Kalra¹, Dr. Rashmi Aggarwal²

Manav Rachna International Institute of Research & Studies, Faridabad
¹arya.vaishali17@gmail.com, ²rashmi.fca@mriu.edu.in

Abstract—Data preparation is an important phase before applying any machine learning algorithms. Same with the text data before applying any machine learning algorithm on text data, it requires data preparation. The data preparation is done by data preprocessing. The preprocessing of text means cleaning of noise such as: cleaning of stop words, punctuation, terms which doesn't carry much weightage in context to the text, etc. In this paper, we describe in detail how to prepare data for machine learning algorithms using RapidMiner tool. This preprocessing is followed by conversion of bag of words into term vector model and describe about the various algorithms which can be applied in RapidMiner for data analysis and predictive modeling. We also discussed about the challenges and applications of text mining in recent days

Index Terms—RapidMiner, Preprocessing of text, TF-IDF, Term Vector Model.

I. INTRODUCTION

THE TEXT data available on web and on computers is increasing rapidly, managing that text data requires intelligent algorithm to retrieve relevant information from the data repositories. The retrieval of information is called text mining. Text mining is not a new concept, it is evolved from data mining and all the data mining algorithms can be applied on the textual data. The difference between the two is data mining is applied on structured data and relational data whereas textual mining deals with all unstructured and semi-structured data well. Nowadays text mining depends upon whether we are looking for the context of the text or content of the text.

If we are looking for context of the text and want to provide the environment for the user to collect the similar patterns together then clustering, visualization and navigation

can be applied and if the user want to find out the relationship between the patterns, discover the new relationships and to summarize the text then content is analyzed and techniques like classification, summarization is applied with the support of information retrieval, information extraction and natural language processing.

As we said the text mining works well on unstructured data. Actually to make this possible, the data is to be converted into semi structured format or in structured format so the data mining machine learning algorithms can be applied easily. This conversion of data is done by preprocessing of the data. The preprocessing of the text data is an essential step as there we prepare the text data ready for the mining. If we do not apply then data would be very inconsistent and could not generate good analytics results. So in preprocessing of the data all the punctuation, unimportant words are removed and words can be grouped into groups, words can stemmed to their roots, all missing values can be replaced with some values, case of text could be replaced into a single one, depending upon the requirement of the application we can apply different steps. After the preprocessing of the data, the data is to be converted in to vector space model and on to that vector-space model, various algorithms works.

In this paper we are going to discuss the preprocessing of the data in detail using Rapidminer tool in section (a), in section (b) we will discuss how the term vector space model is used to convert the term into vector, in next section (c) the various algorithms used for mining task will be discussed and in section (d) Applications of text mining and (e) Conclusion and Future Work.

Section (a): Preprocessing of data

Before starting the preprocessing of the data, we need to fetch the data from the source or data repositories. We have chosen a web link for retrieval of information i.e. a news site “The Times of India” a particular link which is containing the 44th president Obama’s letter for the new president of U.S.A. This link does not contain only the letter content but also contains other information like hyperlinks to other webpages, some ads given by google and some pictures. This is why this data is called unstructured data. To fetch the data from the site, we have applied `getPages` function; this function uses a `get` method to send the request to server and fetches pages mentioned in the input source. Input source provided to this function can be any excel file, csv file, read from Url etc. There are various functions available in Rapidminer to read the data from various sources and providing it to `getPages` function. We have applied `ReadExcel` operator, parameters of this function are shown in figure 1. This operator take input as excel file containing links to retrieve from the web. In the excel file you can set the attribute, like if you want to scan more than one link from the file under the attribute named “link-to-scan” , then name can be assigned in the first row. This operator considers the first row as the name of the attributes. After seeding the excel file, configuration wizard need to be set. This configuration wizard configure which particular sheet of excel file to be set to read the data and selects the number of rows to read from the excel data. When you set this configuration wizard, the `getPages` automatically retrieves the attributes to be scanned from excel sheet, you can select those attributes and process the `getPages` function. After applying this function another function `getContent` is required to get the textual data from html pages.

After retrieval of data, the data is ready for the preprocessing. For preprocessing of data another function we have applied is “Process Document from Data”. This function converts the data into word vector which makes the text data ready for applying the various data mining algorithms. This module calculates the TF-IDF of the word set, which we will discuss later in section (b)[8]. Firstly we will discuss the preprocessing steps:

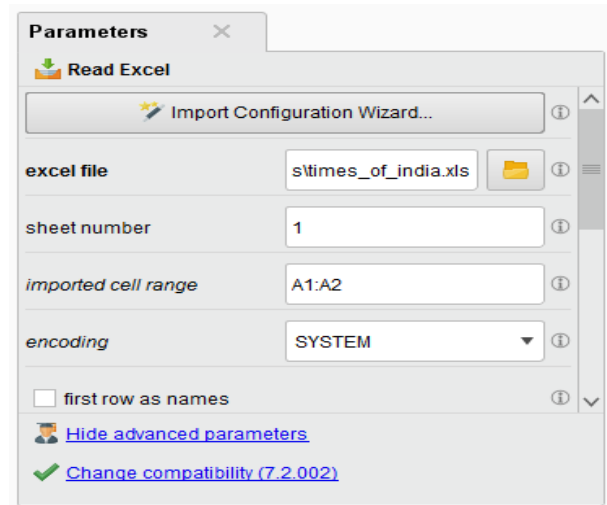


Figure 1: ReadExcel operator parameters

1) Tokenization:

This process split the sequence of strings into words. It removes all the punctuations from the text data and gives words of text which is called tokens. RapidMiner tool provides three ways of splitting; one is default which is commonly used called non letter character where it split on the basis of non-letter of character like spaces, commas, full-stops etc. The second mode is specify character, wherein you can specify characters based on which sentence is split into tokens and the third one is regular expression, wherein the regular expression is provided to split the sentence into tokens. We have chosen the default mode i.e. non letter character mode. The complete process screenshot is shown in figure 2. And in figure 3, inside view of Process Document from data module is shown.

2) Filter Stop-Words:

This process removes words from the document which does not play any important in giving intelligent pattern or information. Eg: the words like “How” “What”, “are” etc. There also we can apply different types of functions which are supported by RapidMiner like Filter Stopwords based on Dictionary, English Language, German, French, Arabic etc. we have applied for English.

3) Filter Tokens by Length:

This is very interesting function, using which you can filter token of specific length, the length attribute needed to be provided in the parameters min chars and max chars; wherein min chars you can specify the minimum length of each token to be in the document to its maximum range.

4) Stemming :

In this process you can stem the words to its root, it roves all the suffixes like tens suffixes: responded to response, eaten to eat and plurals to root like women to woman, men to men and horses to horse. RapidMiner supports variety of algorithms for stemming. Porter, snowball, lovins, dictionary, Arabic to support Arabic language etc. we have used the most popular one snowball in our process. Snowball comprised of 41 rules to stem the words. Although stemming sometime loses the meaning of actual word but still works well in most of the cases.

5) Transform Cases:

This step is useful in normalizing the text. The text would get converted into a single case either to Uppercase or lowercase. This is not always necessary to apply; it depends upon the requirement of a particular problem. We have not applied this operator in our process.

After the completion of preprocessing, the document is to be converted into vector model. In rapid miner this is itself done in “Process Document from Data” by choosing the parameter vector creation to TF-IDF. But we will separately discuss how this vector model is created in our next section.

Section (b): Term Vector space model

The preprocessing of document gives us a document with bag of words only. On which we cannot apply the algorithms directly. We need to convert this bag of words into term vector. The term vector gives a numeric values corresponding to each term appearing in a document which is very helpful in feature selection.

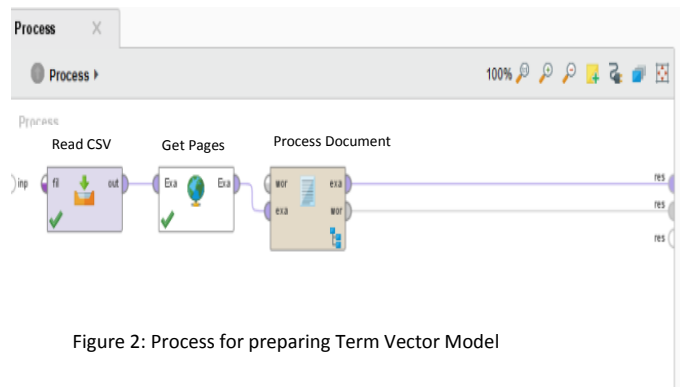


Figure 2: Process for preparing Term Vector Model

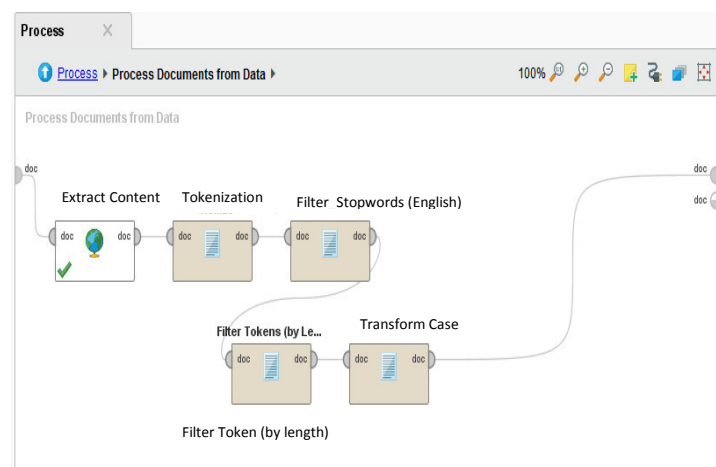


Figure 3: Preprocess

There are three ways for converting terms into vector: Term Frequency, Term Occurrences and Term Frequency-inverted document frequency (TF-IDF). The most useful and popular one is TF-IDF[9]. This gives the higher weightage to the important term and lesser weightage to the unimportant term. Its vector value lies in between of 0 to 1. 0 means the term has no importance in context to the documents in which we are looking for and 1 means the terms are relevant. For calculating the TF-IDF values the document is converted into inverted text file wherein for all documents words are extracted and corresponding to each word a weight is assigned i.e. the term occurrence value and document occurrences using TF-IDF value is calculated.

TF-IDF =

$$\log \frac{\text{No.of times appears in document}}{\text{Total no terms in document}} \times \frac{\text{No.of documents under consideration}}{\text{No.of documents that contain the keyword}}$$

After preparing the term vector model, various algorithms can be applied on the data. In the section we will discuss about the various algorithms which can be applied on textual data for mining purpose.

Section (c): Algorithms for Text Mining

There are various algorithms for text mining for different tasks, but we are giving only overview some important algorithms here.

I. **Unsupervised Algorithms:** Algorithms in which no training data is provided to train the system. The system itself evolved and generates results depending on the data provided. The most popular algorithm under this category is clustering and topic modelling [10]. Both are used for segmentation of data into group, clustering is known for hard segmentation and topic modelling is soft segmentation. RapidMiner provides support of various segmentation algorithms like K-Means, K-medoid, agglomerative clustering, support vector machine[5]. Most popular clustering algorithms are described below:

- **Agglomerative Clustering:** This is known as hierarchical clustering, where pairwise documents are grouped together based on some similarity criteria in bottom to top fashion. A tree like structure is formed which is very useful in searching algorithm or in searching the query from the existing databases.
- **K-medoid:** A center point is chosen from the original data set which is known as a medoid point. Documents are clustered around this point based on similarity measure. After forming the cluster, the medoid point is recalculated until the convergence arrived. This does not work well on textual data as textual data is sparse in nature.
- **k-means clustering:** In k-means clustering, the value of k is assumed and this k gives the information about number of clusters to be formed for the given dataset. This k also serve as the centroid of a particular cluster, around which documents

are clustered based on similarity measure and in each iteration like medoid, k (centroid) is updated. The only disadvantage of this algorithm is the value of k which is seeded initially. For finding the perfect k, various other algorithms can be used to give a supervised k-value.

II. **Supervised Algorithms:** Under this category training data is needed to be provided, classification problems comes under this category. Various algorithms like decision tree based[2], rule based, bayesian, neural nets, etc are provided into RapidMiner, which are being used for predictive modelling[1]. Algorithms for classification are described below:

Decision Tree: In this algorithm for classifying the data set, hierarchical process is followed, where at each node portioning is done based on a predicate. It's a top down approach starting from the root. To partition the node various splitting measures are used like single attribute split wherein based on single value or a word phrase node is split, another one is based on document similarity, wherein based on the similarity between the two documents node is split. Third one is for multivalued attribute, where to split the node discriminant function is applied. For convergence of the tree, you can set a certain threshold like the maximum depth of the tree.

Neural Networks: Neural nets works in two phases: learning phase/training phase and testing phase. To train the network, back propagation algorithm is applied in feed forward neural structure and completion of training the network is tested to check the classification results and network is validated. In RapidMiner we can set the number of training cycles, the number of hidden layers and the minimum error on which network converged.

Rule Based: There are also two phases: growing phase and pruning phase. In growing phase, rules are constructed by considering every possible value of each attribute and in pruning phase, rules are pruned based on pruned metric.

Bayesian Algorithm: It is a probabilistic model, based on Bayes theorem which considers each attribute independently in classification of the data. Small amount of training is required as each attribute plays an independent role in classification of the dataset. In RapidMiner there are two

algorithms in support of this are: Naive Bayes and Naive Bayes kernel model.

Section (d): Applications of Text Mining

Text mining applications ranges from information retrieval to link analysis, there is list of applications:

- I. Tweeter analysis: analysis of social, political and academia tweets.
- II. Sentiment analysis: This is also known as opinion mining, which gives positive, negative, neutral feedback of users. This is widely used nowadays in selecting a hotel, in buying any product from e-commerce site, feedback of any restaurant [4].
- III. Biomedical sciences: Mining on biomedical data and bio-informatics is latest research area nowadays.
- IV. Industrial applications: in industries it is widely used for data analysis tasks, by testers of the project to stock market, or for the competitive businesses it is generally used.
- V. Web search Engines: Text Mining approaches helps a lot in improving the query search criteria in the search engines and improves the information retrieval systems.

Section (e): Conclusion and Future work

Challenges give the future direction of work to be done in improvement of text mining of things. Some of the challenges are listed below[7]:

- I. Heterogeneity of data is one of the biggest challenges for text miners to deal.
- II. Usage of multiple languages and variety of notations used for a same word causes confusion for text miners and makes task difficult.
- III. Sparse data causes over fitting of data into clusters and to different classes and causes wrong analysis for the analysts.
- IV. Domain experts are required for the textual data to give a proper analysis and which may not always be possible to have a good domain expert.
- V. Proper Statistical analysis of data is difficult to give as the data cannot always be good data, it

may have a lots of missing values, so a standard results cannot be given.

Based on the above mentioned challenges, the text miners can work on improving the algorithms for the above defined tasks. Every area still requires efforts in improvement.

REFERENCES

- [1] Charu C. Aggarwal and ChengXiang Zhai: Survey of Text Classification Algorithm, chapter in book "Mining Text Data" DOI: 10.1007/978-1-4614-3223-4_6, pp 163-222-springer US 2012.
- [2] S. B. Kotsiantis: Decion Trees: A recent Overview, article published in "Artificial Intelligence Review", in April 2013, Volume 39, Issue 4, pp 261-283-springer.
- [3] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain: Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks,- Expert Systems with Applications Volume 41, Issue 4, Part 2, March 2014, Pages 1937-1946- Elsevier.
- [4] R Moraes, J. O. F Valiati, W. P. G. O. Neto: Document-levelsentiment classification: An empirical comparison between SVM and ANN, Expert Systems with Applications Volume 40, pp 621-633 2013 - Elsevier.
- [5] Thorsten Joachims: Text categorization with Support Vector Machines: Learning with many relevant features, Support Vector Learning, Machine Learning: ECML-98,Volume 1398 of the series Lecture Notes in Computer Science pp 137-142.
- [6] V Bijalwan, V Kumar, P Kumari, J Pascua: KNN based Machine Learning Approach for Text and Document Mining, International Journal of Database Theory and Application Vol.7, No.1 (2014), pp.61-70.
- [7] A. Ittoo, L. M. Nguyen, A. van den Bosch: Text analytics in industry: challenges, desiderata and trends, Computers in Industry,Volume 78, May 2016, pp 96-107 - Elsevier.
- [8] Gary Miner, John Elder: Practical text mining and statistical analysis of text mining, IV, Thomas Hill, 1st edition, ISBN-978-0-386979-1, 2012 - books.google.com.
- [9] Li-Ping Jing, Hou-kuan, Hong Boshi: Improved feature selection approach using TF-IDF in Text Mining, in Proceedings of the first International conference on Machine Learning and cybernetics, Beijing, pp-944 to 946, 4-5 November 2002-IEEE.
- [10] Charu C. Aggarwal and ChengXiang Zhai: A Survey of Text Clustering Algorithms chapter in book "Mining Text Data", DOI 10.1007/978-1-4614-3223-4_6, pp 77-128 springer US 2012.
- [11] Rashmi Agrawal: K-Nearest Neighbor for Uncertain Data, in International Journal of Computer Applications (0975 - 8887) Volume 105 - No. 11, November 2014.
- [12] Rashmi Agrawal, Mridula Batra: A Detailed Study on Text Mining Techniques, in International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

A study of Optimal Testing Resource Allocation Problem for Modular Software with Change Point

Gurjeet Kaur¹, Anu G. Aggarwal², Aman Kedia³

¹Shaheed Sukhdev College of Business Studies, University of Delhi, India

²Department of Operational Research, University of Delhi, India

³Shaheed Sukhdev College of Business Studies, University of Delhi, India

¹gurjeetkaur@sscsdu.ac.in, ²anuagg17@gmail.com, ³aman.kedia1111@gmail.com

Abstract—For evolving trustworthy software, engrossing on uncovering process of fault in software is central. Nevertheless, during testing, modifications in the testing routine, defect gravity or testing-skill maturity and working environment, there can be notable change in fault detection rate. When this sort of pattern is observed in testing time it is called change point. In this article, we inquire a resource distribution problem that optimally distributes software developing resources in such a way that the cost of development is curtailed to optimization. In this problem, for all modules the effect of chief circumstantial element of change-point is considered. The constraint of pulling off the desired reliability level for every individual module is also incorporated in the formulation of the problem. A framework based on Karush Kuhn Tucker (KKT) conditions is presented to work out the resulting non-linear optimization problem. A simulated numerical illustration has been analyzed to reflect the formulation of the case and its solution by the algorithm proposed.

Index Terms—Modular Software, Change-point, Software Reliability Growth Model, Resource Allocation

I. INTRODUCTION

BREAKING a vast and complex assignment into small simple steps of objectives' is one of the deep-seated rules to success. This law prevails into software industry as well. The quality, convolution and size of software have risen, resulting in complete development of it as a hard process. So, today maximum soft coded sharewares are developed by amalgamating small ordered independent soft codes called modules. This methodology of developing Modular Software System helps soft code teams to proficiently build the large sized coded complex development process in a structured manner.

With the rising popularity of having efficient modular software, there arises the need of designing reliable structures. As per texts [13], Software reliability is kept as the chances that the software does not go wrong within an assigned period of time under given circumstances. Therefore, with shortened maturity cycles, raised complexity of software design, and great caustic penalty of software failures, a major responsibility lies in the area of software testing. During the advance in development of modular software, faults can seep in modules owing to individual flaw. These faults make themselves marked in expressions of malfunction when the codes are weathered autonomously during the

module testing phase of software development life cycle. To evaluate modular software quantitatively and to find the total number of faults removed from each module mathematical tools namely software reliability growth models (SRGMs) are employed. Assorted noteworthy metrics, like primary number of faults, failure concentration, Trustworthiness within a specific period of time, number of faults residual, can be smoothly dogged through SRGM's.

Through the last three decades, a great number of SRGMs have been framed in literature [3, 13, 16, 19, 20, 22]. Early researches related SRGM with respect to testing time [3, 17]. However, incorporation of testing resources leads to development of more accurate SRGMs [4, 5, 9, 26]. The reason being, the detection of faults are more intimately linked to the amount of resources expended [18, 24] on testing as it includes-

(a) Manpower, that takes into account

- Testing group (malfunction recognition personnel).
- Debugging group (Programmers\Failure rectification personnel).

(b) CPU working moments

Huang et al. [5] worked with logistic resource function. They established that both exponential-type and S-type NHPP models can come under ideal and imperfect debugging situations. Later, Kapur et al.[9] deliberated on the testing resource reliant learning process and classified faults into two types on the source of amount of testing resources needed to remove them. This paper models fault exclusion rate in terms of testing resource.

For accomplishing the goal of developing reliable modular software, the consideration of fault detection process is vital as it aid in establishing the value of uncovering bugs that lie dormant in the software by test techniques and test cases. A lot of SRGMs in literature supposed that during the fault detection process each failure caused by a fault occurs independently and at random time according to the same distribution [3, 20]. However, practically as the testing evolves, the testing team gains insight and with the employment of new tools and techniques the fault detection rate (FDR) gets markedly changed. Further, the other factors that can affect the fault detection rate are operation environment, testing stratagem and shortcoming density. And the point of

time where change in fault detection rate is observed is termed as ‘Change Point’. It was Zhao [28] who came up with the concept of Change point in software as well as hardware reliability modeling. He established that incorporation of change point in SRGMs is vital for effective fault detection modeling. Shyur [23], Wang and Wang [25] also made offerings in this area. In accumulation, some studies included change-point analysis in their models as the testing resource consumption may not be smooth over time [4, 10, 15].

Another core apprehension in the software production is the software development outlay. Overspending on development cost can result in financial crisis for the company. On the other hand, spending less can result in low quality software product as in this case the software development firm will have to set low reliability aspiration level for each module. Thus, there arises the need of optimizing total development cost of modular software. In modular soft code testing stage, each modular code is tested autonomously. But this has to be carried out in finite time. However, completing this task involves on an average of 50% of the total budget of software development cost. Hence there is requirement of resource distribution decisions popularly known as “Testing Resource Allocation problems”. This research is adding contribution in Resource Allocation Area. We have developed change point incorporated non linear resource distribution case which is solved by Karush Kuhn Tucker (KKT) optimality conditions.

The work is organized as follows: Section 2 details the literature review on testing resource allocation problem Section 3 highlights on the Goel-Okumoto software reliability growth model with change point and testing resource, required for modeling the failure mechanism of the modules. Section 4 elaborates on formulation of our testing resource allocation problem. Further in this section we also discuss an optimization algorithm based on the KKT optimality conditions. Section 5 illustrates the optimization problem solution through a numerical example. Eventually, conclusions are derived and are given in section 6.

II. LITERATURE REVIEW ON TESTING RESOURCE ALLOCATION PROBLEMS

The value of SRGMs is not constrained to reliability assessment of software systems. Their application is well researched and applied in resource distribution cases. These decision problems are critical for software development firms aligned to reliability, cost, time and resources parameters. But the goal goes specific with the company needs. There are software development firms that aim at optimization of failure numbers. There are some that go with optimal cost decisions. Then there are others that go with constrained optimization problems. With these different interests consideration, software allocation has a rich repository of research papers. Ohetera and Yamada [21] proposed resource distribution problem for optimizing the remaining faults and optimizing resources respectively. Yamada et al.

[27] also formulated a constrained resource distribution problem with constraint of reliability level. Hyper-geometric model for modeling software reliability growth was employed by Huo et al. [7] to make optimal distribution of resources amongst modules. Kapur et al. [8] discussed the concept of marginal testing effort function (MTEF) and related optimal resource distribution problem. Not only this, Kapur et al. [11,12, 29], using S-Shaped and exponential SRGMs [3,20] have provided different allocation cases studied various resource allocation problems maximizing the number of faults removed from each module. Khan et al. [14] Huang et al. [6] too formulated and solved constrained optimal resource distribution cases with one or more parameters of cost, fault and reliability. Sindhuja et al. [30] have investigated assorted software make public decision policies and resource distribution cases considering the dual restrictions of reliability and expenses.

However, the impact of change point has yet not been considered in allocation problems for modeling the reliability growth of modules. In this paper, we inquire a resource distribution problem that optimally distributes software developing resources in such a way that the cost of development is curtailed to optimization. In this problem, for all modules the effect of chief circumstantial element of change-point is considered. The constraint of pulling off the desired reliability level for every individual module is also incorporated in the formulation of the problem. (figure 2.1).

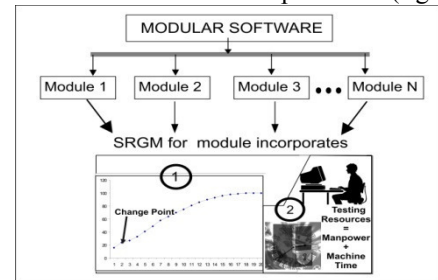


Figure 2.1 Modular Software incorporating Change Point and Testing Resources

III. GOEL-OKUMOTO(GO) MODEL INCORPORATING TESTING RESOURCE AND CHANGE POINT

A. Model Notations

a	Initial number of faults
$W(t)$	Cumulative testing resource in the time interval $(0,t]$
$w(t);$ $w(t) = \frac{d}{dt} W(t)$	Testing resource intensity
τ	Change Point
$m(t)$ or $m(W(t))$	Cumulative number of faults removed by time t
$b(t)$ or $b(W(t))$	Testing resource expenditure based fault detection rate per remaining fault

b_1	Fault detection rate per remaining fault before change point
b_2	Fault detection rate per remaining fault after change point

The GO software reliability growth model incorporating testing resource and change point [4] taken in this article has its foundation on Non Homogeneous Poisson Process (NHPP). The NHPP models are based on the supposition that the software system is constrained to failures at random times caused by manifestation of remaining faults in the system. So, for modeling the software fault detection phenomenon, counting process $\{N(t); t \geq 0\}$ is defined. This counting process exhibits the collective number of faults discovered by testing time t . And the SRGM formulated on this can be put mathematically as-

$$Pr\{N(t) = n\} = \frac{m(t)^n \cdot \exp(-m(t))}{n!}, \quad n = 0, 1, 2, \dots \quad (3.1)$$

B. Model Assumptions

- NHPP governs removal phenomenon.
- Software is prone to failures during the execution caused by faults remaining in the software.
- On a malfunction, the slip causing that failure is without delay distant and no fresh faults are introduced.
- The fault detection rate is with reverence to testing resource strength and is proportional to the existing fault content in the software.
- The fault recognition rate may transform at some time moment (called change point, denoted by τ).

In view of the above assumptions, the model can be summarized by the following differential equation.

$$\frac{d}{dt} \frac{m(t)}{w(t)} = b(t)(a - m(t)) \quad (3.2)$$

Since, fault detection rate (FDR) changes at time point τ , therefore, it is defined as:

$$b(t) = \begin{cases} b_1, & \text{when } 0 \leq t \leq \tau \\ b_2, & \text{when } t > \tau \end{cases} \quad (3.3)$$

Case 1: For $0 \leq t \leq \tau$

Using equation (3.2) and (3.3), we have the following differential equation

$$\frac{d}{dt} \frac{m(t)}{w(t)} = b_1(a - m(t)) \quad (3.4)$$

Solving (3.4) with initial condition $m(t=0)=0$, we get,

$$m(t) = a \left[1 - e^{-b_1 W(t)} \right] \quad (3.5)$$

Case 2: $t > \tau$

Again by equation (3.2) and (3.3), we have

$$\frac{d}{dt} \frac{m(t)}{w(t)} = b_2(a - m(t)) \quad (3.6)$$

The solution of above differential equation under initial condition $m(t = \tau) = m(\tau)$ is

$$m(t) = a \left[1 - e^{-b_1 W(\tau) - b_2 (W(t) - W(\tau))} \right] \quad (3.7)$$

Combining Case 1 and Case 2, we have the following expression of $m(t)$

$$m(t) = \begin{cases} a \left[1 - e^{-b_1 W(t)} \right], & \text{when } 0 \leq t \leq \tau \\ m(t) = a \left[1 - e^{-b_1 W(\tau) - b_2 (W(t) - W(\tau))} \right], & \text{when } t > \tau \end{cases}$$

where $m(t)$ is the mean value function of the counting process $N(t)$.

IV. TESTING RESOURCE ALLOCATION PROBLEM

Let us consider modular software with N modules. These soft coded structures vary in parameters like- (a) Complexity (b) Size (c) Performing functions. We also consider that the testing of these modular codes is done independently. We also take into account a reasonable assumption of not having infinite faults in the software modular structures. The optimization problem considered is to minimize the software development cost underneath a conjecture that there is change in fault detection rate in each soft code module and a set level of reliability is to be pulled-off.

A. Notations

i	Module number counter $i=1,2,\dots,N$
a_i	Initial number of faults in module i
τ_i	Change point for i^{th} module
$W(\tau_i)$	Testing resource consumed by i^{th} module till τ_i .
b_{1i}	Detection process pace in terms of rate before change point in each module
b_{2i}	Detection process pace in terms of rate after change point in each module
C_{1i}	Fault removal cost per fault before change point in i^{th} module during testing phase
C_{2i}	Fault removal cost per fault after change point in i^{th} module during testing phase
C_{3i}	Fault removal cost per fault from i^{th} module in operational phase.
C_4	Testing cost
R_0	Reliability level
W	Sum testing resource

B. Modeling mean value and reliability function for the modules

It is not possible to test all the test cases of a large complex Software. The two major reasons for this infeasibility is (i) Deadline for the development, that is its release time (call it T). (ii) Resources of testing. Let W_i be the testing resource used up on the i^{th} module during testing time T. Therefore, devoid of any loss of generality the numeral of faults distant by time T can be assumed to be a function of testing resource explicitly and using the model discussed in section 3, the mean value function for the faults removed from each module can be written as:

$$\text{i.e. } m_i(W_i) = \begin{cases} a_i (1 - e^{-b_{1i} W_i}) & \text{when } 0 \leq t \leq \tau_i \\ a_i (1 - e^{-b_{1i} W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))}) & \text{when } t > \tau_i \end{cases} \quad \forall i = 1, 2, \dots, N \quad (4.1)$$

Modeling expected number of faults removed from each module using such an SRGM is advantageous because it takes into consideration the consequence of change point.

Mathematically, reliability of soft codes of modules can be described by [13]-

$$R_i(t) \equiv R_i(t + \Delta t | t) = \exp^{-(m_i(t+\Delta t) - m_i(t))} \quad (4.2)$$

There is one more way to define Eq. (4.2). This was kept by Huang et al. [5]. Eq. (4.3) states reliability as the fault removals to the proportion of initial number of faults. That is- $R_i(t) = m_i(t)/a_i \quad \forall i = 1, 2, \dots, N$ (4.3)

Using Eq. (3.1) and (3.4) we have reliability of each module is given by:

$$\frac{m_i(W_i)}{a_i} \quad \forall i = 1, 2, \dots, N \quad (4.4)$$

When the reliability is defined by expression (4.3) it is assumed that all the faults of the software are of same type and are equally likely to be detected during testing. Thus assumption goes well with respect to the constant FDR for the model under consideration here.

C. Modeling Cost Function for Modular Software

For modeling cost function with respect to time, we have

$$C(T) = C_1 m(\tau) + C_2 (m(T) - m(\tau)) + C_3 (a - m(T)) + C_4 T \quad (4.5)$$

Eq. (4.5) is with respect to time. Cost modeling with respect to testing resources can be put as-

$$C(W) = C_1 m(W(\tau)) + C_2 (m(W) - m(W(\tau))) + C_3 (a - m(W)) + C_4 W \quad (4.6)$$

Further, since the software is modular and each module of software is designed independently; therefore, the outlay of fixing and testing of all components is the addition of individual modules testing cost. Mathematically, we have:

$$\begin{aligned} C(W) &= \sum_{i=1}^N C_i(W_i) \\ &= \sum_{i=1}^N C_{1i} m_i(W_i(\tau_i)) + \sum_{i=1}^N C_{2i} (m_i(W_i) - m_i(W_i(\tau_i))) + \\ &\quad \sum_{i=1}^N C_{3i} (a_i - m_i(W_i)) + C_4 \sum_{i=1}^N W_i \end{aligned} \quad (4.7)$$

D. Problem Structure and Solution Algorithm

The resource distribution problem is structured with this scenario- (a) W is the sum resources that needs to be distributed in N independent modules. (b) The aim of the allocation problem is- Minimizing Software Development Cost given by Eq.(4.7) (c) The objective is constrained to aspired reliability of at least R_0

$$\begin{aligned} \text{Min } C(W) &= \sum_{i=1}^N C_i(W_i) \\ &= \sum_{i=1}^N C_{1i} m_i(W_i(\tau_i)) + \sum_{i=1}^N C_{2i} (m_i(W_i) - m_i(W_i(\tau_i))) + \\ &\quad \sum_{i=1}^N C_{3i} (a_i - m_i(W_i)) + C_4 \sum_{i=1}^N W_i \\ \text{Constrained to } &\sum_{i=1}^N W_i \leq W, \\ &R_i \geq R_0 \quad \forall i = 1, 2, \dots, N \\ &W_i \geq 0, \quad \forall i = 1, 2, \dots, N \end{aligned} \quad (P1)$$

After substituting the value of $m_i(W_i)$ and R_i from (4.1) and (4.4) in (P1) we get problem as;

$$\begin{aligned} \text{Min } C(W) &= \sum_{i=1}^N C_i(W_i) \\ &= \sum_{i=1}^N C_{1i} m_i(W_i(\tau_i)) + \sum_{i=1}^N C_{2i} (a_i (1 - e^{-b_{1i} W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))}) - m_i(W_i(\tau_i))) + \\ &\quad \sum_{i=1}^N C_{3i} (a_i - (a_i (1 - e^{-b_{1i} W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))}))) + C_4 \sum_{i=1}^N W_i \end{aligned}$$

Constrained to

$$\begin{aligned} &\sum_{i=1}^N W_i \leq W, \\ &a_i (1 - e^{-b_{1i} W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))}) \geq R_0 \quad \forall i = 1, 2, \dots, N \\ &W_i \geq 0, \quad \forall i = 1, 2, \dots, N \end{aligned}$$

Dropping the constant terms and re-writing the above problem in maximization form, we get the problem as:

$$\begin{aligned} \text{Max } Z(W) &= - \sum_{i=1}^N C_i(W_i) = \\ &\quad \sum_{i=1}^N (C_{3i} - C_{2i}) a_i (1 - e^{-b_{1i} W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))}) - C_4 \sum_{i=1}^N W_i \end{aligned}$$

Constrained to

$$\sum_{i=1}^N W_i \leq W,$$

$$a_i \left(1 - e^{-b_{1i}W_i(\tau_i) - b_{2i}(W_i - W_i(\tau_i))} \right) \geq R_0 \quad \forall i = 1, 2, \dots, N$$

$$W_i \geq 0, \quad \forall i = 1, 2, \dots, N \quad (P2)$$

Now, the reliability constraint of (P2) can be re-stated as:

$$W_i \geq \frac{-1}{b_{2i}} \left[\ln(1 - R_0) + b_{1i}\tau_i - b_{2i}\tau_i \right] \cong (\text{say } Y_i)$$

Here, it may be noted that summation of Y_i should always be less than equal to W , otherwise the constraints of the formulated problem will be inconsistent.

Therefore, using the constraint on reliability and letting $W_i = X_i + Y_i$, the problem can be transformed as:

Max $Z(W) =$

$$\sum_{i=1}^N (C_{3i} - C_{2i}) a_i \left(\left(1 - e^{-b_{1i}W_i(\tau_i) - b_{2i}(X_i + Y_i - W_i(\tau_i))} \right) - C_4 \sum_{i=1}^N (X_i + Y_i) \right)$$

Subject to

$$\sum_{i=1}^N X_i \leq W - \sum_{i=1}^N Y_i,$$

$$X_i \geq 0, \quad \forall i = 1, 2, \dots, N \quad (P3)$$

The first term of the objective function of (P3) is concave and second term $X_i + Y_i$ is linear, therefore $-(X_i + Y_i)$ can be treated as concave. Hence the objective function of (P2) is concave function given to linear constraint. Therefore the case is of convex programming problem.

Thus, the essential most favorable situations pertaining to Karush Kuhn-Tucker for convex programming problem is also sufficient [1]. For the problem (P3) we can affirm the follow saddle value decision making statement:

$$\begin{aligned} \underset{X_i, i=1,2,\dots,N}{Max} \underset{\phi}{Min} F(X_1, X_2, \dots, X_N, \phi) &= \sum_{i=1}^N (C_{3i} - C_{2i}) a_i \left(\left(1 - e^{-b_{1i}W_i(\tau_i) - b_{2i}(X_i + Y_i - W_i(\tau_i))} \right) \right. \\ &\quad \left. - C_4 \sum_{i=1}^N (X_i + Y_i) + \phi \left(\sum_{i=1}^N X_i - W + \sum_{i=1}^N Y_i \right) \right) \end{aligned} \quad (P4)$$

The necessary and sufficient situations for (X^0, ϕ^0) where $X^0 = \{X_i^0: i = 1, \dots, N\}$ to be a saddle spot for the saddle value cases are based on the KKT conditions and are specified by the next theorem.

Theorem 1. A feasible decision X_i ($i=1, \dots, N$) of (P4) is best possible if and only if

- (1) $\phi \leq C_4 - \left[(C_{3i} - C_{2i}) a_i b_{2i} e^{-b_{1i}W_i(\tau_i) - b_{2i}(X_i + Y_i - W_i(\tau_i))} \right]$
- (2) $X_i \times \left[\phi - C_4 + \left[(C_{3i} - C_{2i}) a_i b_{2i} e^{-b_{1i}W_i(\tau_i) - b_{2i}(X_i + Y_i - W_i(\tau_i))} \right] \right] = 0$

This theorem can be derived straight from KKT conditions.

Ruling a feasible key at optimality condition

Concerning KKT conditions to the (P3) we get-

$$\frac{\partial F(X_1, X_2, \dots, X_N, \phi)}{\partial X_i} = 0$$

implies

$$X_i^0 = \frac{1}{b_{2i}} \left[\ln \left[(C_{3i} - C_{2i}) a_i b_{2i} \right] - \ln(C_4 - \phi) - b_{1i}\tau_i - b_{2i}(Y_i - \tau_i) \right] \quad (4.8)$$

and $\frac{\partial F(X_1, X_2, \dots, X_N, \phi)}{\partial \phi} = \sum_{i=1}^N X_i - W + \sum_{i=1}^N Y_i = 0$ implies

$$\phi^0 = C_4 - \exp \left[\frac{\sum_{i=1}^N (1/b_{2i}) \left[\ln \left[(C_{3i} - C_{2i}) a_i b_{2i} \right] - b_{1i}\tau_i - b_{2i}(Y_i - \tau_i) \right] - W + \sum_{i=1}^N Y_i}{\sum_{i=1}^N (1/b_{2i})} \right] \quad (4.9)$$

The solution algorithm of above problem, using KKT is as follows:

Algorithm 1:

Step 0: Calculate $\sum_{i=1}^N Y_i$

If $\sum_{i=1}^N Y_i > W$ then

Available resource W is insufficient to meet reliability aspiration level of all the modules.

Stop

Else

Goto Step 1.

End if

Step 1: Set $S = 0$.

Step 2: Calculate $X_i, i = 1, \dots, N - S$; ϕ using equation (3.9) and (3.10)

$$X_i^0 = \frac{1}{b_{2i}} \left[\ln \left[(C_{3i} - C_{2i}) a_i b_{2i} \right] - \ln(C_4 - \phi) - b_{1i}\tau_i - b_{2i}(Y_i - \tau_i) \right]$$

$$\phi^0 = C_4 - \exp \left[\frac{\sum_{i=1}^N (1/b_{2i}) \left[\ln \left[(C_{3i} - C_{2i}) a_i b_{2i} \right] - b_{1i}\tau_i - b_{2i}(Y_i - \tau_i) \right] - W + \sum_{i=1}^N Y_i}{\sum_{i=1}^N (1/b_{2i})} \right]$$

Step 3: Rearrange index i such that:

$$X_1 \geq X_2 \geq X_3 \dots \dots \dots \geq X_{N-S}$$

Step 4: If $X_{N-S} > 0$ then Stop (the solution is optimal)

Else $X_{N-S} = 0$; set $S = S+1$,

End if

Step 5: For re-allocating testing resources to remaining N-S modules go to **Step 2**.

The optimal solution is given by:

$$X_i^0 = \frac{1}{b_{2i}} \left[\ln \left((C_{3i} - C_{2i}) a_i b_{2i} \right) - \ln(C_4 - \phi) - b_{1i} \tau_i - b_{2i} (Y_i - \tau_i) \right], i = 1, \dots, N-S$$

where

$$\phi^0 = C_4 - \exp \left[\frac{\sum_{i=1}^N (1/b_{2i}) \left[\ln \left((C_{3i} - C_{2i}) a_i b_{2i} \right) - b_{1i} \tau_i - b_{2i} (Y_i - \tau_i) \right] - W + \sum_{i=1}^N Y_i}{\sum_{i=1}^N (1/b_{2i})} \right]$$

$$X_i^0 = 0, \text{ otherwise}$$

V. NUMERICAL EXAMPLE

Presume that the sum total of testing resource accessible for modules coded software is 60,000. This modular software has six such modules. Further, it is supposed that there is change in fault detection rate in each module. Further, because of change point the cost of testing before and after change point differs. These costs are taken as 1 and 2 units respectively for each module. The price of debugging a fault for any module in operational segment is 8 units and charge of testing up to release time is 0.5 units. The assumed estimated values of all the parameters for modules M1, M2, M3, M4, M5 and M6 are tabulated in table 5.1. Also, it is aspired that the reliability level of each module should be at least 0.8.

Table 5.1 Parameter Estimates for six modules

Module	a	b ₁	b ₂	W(τ)
M1	1321	0.000213	0.000211	642.85
M2	950	0.000181	0.000129	505.02
M3	1639	0.000112	0.000156	759.18
M4	1450	0.000198	0.000213	580.02
M5	1350	0.000218	0.000229	462.69
M6	987	0.000125	0.000321	315.11
Total	7697			

Using algorithm 1 of section 3 the optimal allocation of the resource for the six modules is shown in table 5.2.

Table 5.2 Optimal Allocation of resources for six modules

Module	a	W	m	Reliability	Cost
M1	1321	9438.22	1141	0.864	3553.429
M2	950	12272.7	760	0.8	2957.012
M3	1639	12434.9	1396	0.851	4605.807
M4	1450	9878.21	1272	0.877	3813.009
M5	1350	9176.5	1185	0.877	3566.017
M6	987	6799.44	869	0.88	2646.096
Total	7697	60000	6623		51141.37(including constant cost)

From table 5.2 we have that within the available budget of 6000, the total expected cost of testing all the modules such

that the reliability of each module is at least 0.8 is 51141.37. The number of faults removed from the software is 6623.

The impact of increasing aspired reliability level for all modules-

For studying the impact of increasing aspiration level for all modules on the allocation of testing resources, we solved the above numerical illustration by taking reliability level as 0.85 and 0.9 for each module.

Case 1: When aspired reliability for each module is increased from 0.8 to 0.85

In this case we found, that by raising the aspiration of reliability to 0.85, there was allocation of resources to modules (given in table 5.3) but with increase software development cost as compared to the cost of development when reliability level of each module is 0.8(refer table 5.2).

Table 5.3 Optimal Allocation of resources for six modules

Module	a	W	m	Reliability	Cost
M1	1321	8985	1123	0.85	3661.857
M2	950	14502.8	808	0.85	2672.012
M3	1639	12375.2	1394	0.85	4619.499
M4	1450	9216.19	1245	0.858	3975.098
M5	1350	8560.73	1159	0.858	3716.78
M6	987	6360.15	851	0.861	2753.65
Total	7697	60000	6580		51398.90(including constant cost)

Figure 5.1 and 5.2 shows the comparison of Reliabilities and Cost of Modules respectively when Aspiration level is increased from 0.8 to 0.85.

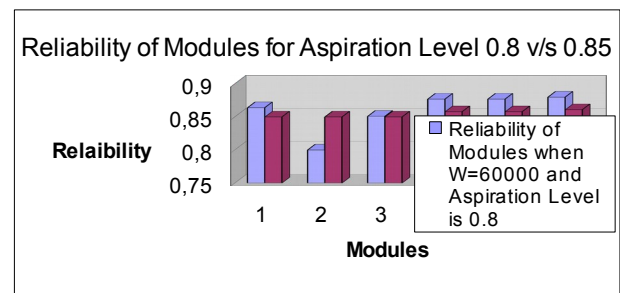


Figure 5.1 Reliability of Modules for Aspiration Level 0.8 v/s 0.85

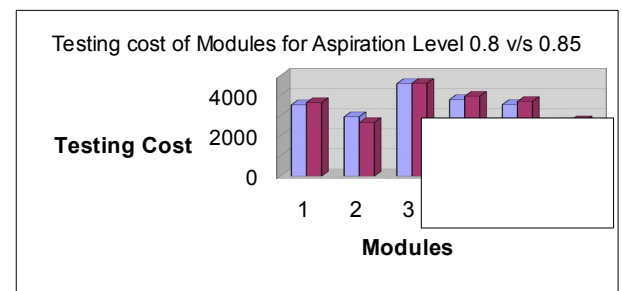


Figure 5.2 Testing Cost of Modules for Aspiration Level 0.8 v/s 0.85

Case 2: When aspired reliability for each module is increased from 0.8 to 0.9

In this case we observed that $\sum_{i=1}^6 Y_i$ is strictly greater than W . So using the proposed algorithm we stop with the conclusion that the given set of constraint inequalities is inconsistent.

The impact of increasing/decreasing total testing resource budget on the optimal allocation among modules

For studying the impact of change in total testing resource budget on the optimal allocation among modules, we increase and decrease the total testing resource W by 20%.

Case 1: W is increased by 20%

In this case we got the allocation as given in table 5.4.

Table 5.4 Distribution of resources for six modules

Module	a	W	m	Reliability	Cost
M1	1321	11786.4	1212	0.917	3131.269
M2	950	12715	771	0.811	2893.786
M3	1639	15611.1	1491	0.909	4034.809
M4	1450	12204.4	1342	0.925	3394.813
M5	1350	11340.2	1249	0.925	3177.04
M6	987	8342.98	916	0.927	2368.601
Total	7697	72000	6981		55000.32(including constant cost)

From table 5.4 we observe that by increasing the budget there is increase in reliability level of modules but with raise in total cost. Figure 5.3 depicts the comparison of Reliability of Modules when $W=60000$ and when it is increased by 20%. The total cost comparison with $W=60000$ and $W=72000$ is shown in Figure 5.4.

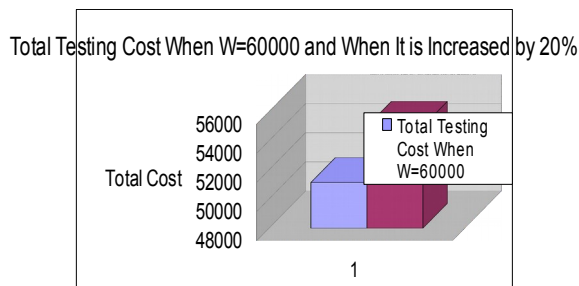


Figure 5.3 Reliability of Modules when $W=60000$ and When It is increased by 20%

Case 2: W is decreased by 20%

In this case we get that $\sum_{i=1}^6 Y_i$ is strictly greater than W .

So using the proposed algorithm we stop with the conclusion that the given set of constraint inequalities is inconsistent.

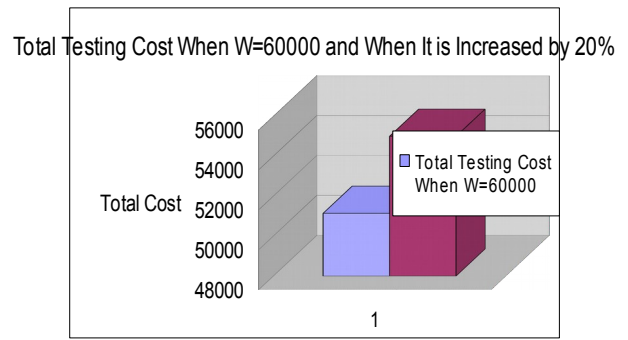


Figure 5.4 Total Testing Cost when $W=60000$ and When It is increased by 20%

VI. CONCLUSION

Allocation of testing resources during module testing phase is an important issue for the project managers in developing a reliable and economical modular software system. This research takes into account change point in modular software development and its associated resource distribution problem. For modeling the failure process of modules an exponential SRGM with testing resource and change point is used. The allocation problem is convex programming problem and is solved using Karush Kuhn Tucker (KKT) optimality conditions. Using the numerical example some important observations related to allocation problem is also presented in the paper.

The present study is done under the assumption of independence of the failures of different modules. We can also explore the possibility of including multi dimensional software reliability growth modeling so as to take care the effect of not only testing resource but also other testing factors like testing coverage, testing time/number of test cases on the fault removal process simultaneously. This paper takes into account independent behavior of faults in modules soft codes. In future interactions among modules and dependence of the failures can also be incorporated in the model building.

REFERENCES

- [1] Bazaraa, S. M. and C. M. Setty, "Nonlinear programming: theory and algorithm", John Wiley and Sons, 1979, New York.
- [2] Camuffo M., Maiocchi M., Morselli M. (1990) "Automatic Software Test Generation Inform Software Technology", 32(5), 337-346.
- [3] Goel A. L. and Okumoto, K., "Time dependent error detection rate model for software reliability and other performance measures", IEEE Transactions on Reliability, 1979, R-28, 206-211.
- [4] Huang C. Y., "Performance analysis of software reliability growth models with testing-effort and change-point", Journal of Systems and Software, 2005, 76, 181-194.
- [5] Huang C. Y., Kuo S. K., Lyu M. R., "An assessment of testing-effort dependent software reliability growth models", IEEE Transactions on Reliability, 2007, 56, 198-211.
- [6] Huang C. Y., Lo J. H., Kuo S. K. and Lyu M. R., "Optimal allocation of testing resources considering cost, reliability, and testing -effort", Proceedings of the 10th IEEE Pacific International Symposium on dependable Computing, 2004.
- [7] Huo R. H., Kuo S. K., Chang Y. P., "Needed resources for software module test, using the hyper-geometric software reliability growth model", IEEE Trans. on Reliability, 1996, 45(4), 541-549.

- [8] Kapur P. K., Bardhan A., Yadavalli V., "On allocation of resources during testing phase of a modular software", *International Journal of Systems Science*, 2007, 38(6), 493-499.
- [9] Kapur P. K., Goswami D. N., Bardhan A., Singh O., "Flexible software reliability growth model with testing effort dependent learning process", *Appl. Math. Model.*, 2008, 32, 298-307.
- [10] Kapur P. K., Gupta Anu, Shatnawi Omar, Yadavalli V. S. S., "Testing Effort Control Using Flexible Software Reliability Growth Model With Change Point" *International Journal of Performability Engineering- Special issue on Dependability of Software/ Computing Systems*, 2006, 2(3), 245-262.
- [11] Kapur P. K., Jha P., Bardhan A., "Dynamic programming approach to testing resource allocation problem for modular software", *Ratio Mathematica, Journal of Applied Mathematics*, 2003, 14, 27-40.
- [12] Kapur P. K., Jha P., Bardhan A., "Optimal allocation of testing resource for a modular software", *Asia Pacific Journal of Operational Research*, 2004, 21(3), 333-354
- [13] Kapur, P. K., Garg, R. B. and Kumar, S., *Contributions to Hardware and Software Reliability*, World Scientific: Singapore 1999.
- [14] Khan M., Ahmad N., and Rafi L., "Optimal Testing Resource Allocation for Modular Software Based on a Software Reliability Growth Model: A Dynamic Programming Approach", *Proceedings of the International Conference on Computer Science and Software Engineering*, 2008.
- [15] Lin C. T., Huang C. Y., "Enhancing and measuring the predictive capabilities of testing-effort dependent software reliability models, *Journal of Systems and Software*, 2008, 81, 1025-1038.
- [16] Lyu M. R., *Handbook of Software Reliability Engineering*, McGraw-Hill, New York, 1996.
- [17] Musa J. D. "A Theory of Software Reliability and its Application", *IEEE Transaction Software Engineering*, 1975, SE-1, 312-327.
- [18] Musa J. D. and Okumoto K. "A Logarithmic Poisson Execution Time Model for Software Reliability Measurement", in *Proceedings of 7th International Conference on Software Engineering*, 1984, 230-238.
- [19] Musa J. D., Iannino A., Okumoto K., *Software Reliability: Measurement, Prediction, Application*, McGraw-Hill, New York, 1987.
- [20] Obha, M., "Software reliability analysis models", *IBM Journal of Research and Development*, 1984, 28, 428-443.
- [21] Ohetera H. and Yamada S., "Optimal allocation and control problems for software testing resources", *IEEE Transactions on Reliability*, 1990, 39 (2), 171-176.
- [22] Pham H., *Software Reliability*, Springer-Verlag, Singapore, 2000.
- [23] Shyur H. J., "A stochastic software reliability model with imperfect-debugging and change-point, *Journal of Systems and Software*, 2003, 66, 135-141.
- [24] [Trachtenberg M., "A General Theory of Software-Reliability Modeling", *IEEE Transaction on Reliability*, 1990, 39, 92-96.
- [25] Wang Z. and Wang J. "Parameter Estimation Of Some NHPP Software Reliability Models With Change-Point" *Communications in Statistics- Simulation and Computation*, 2005, 34, 121-134.
- [26] Yamada S., Ohtera H., and Narihisa H., "Software Reliability Growth Models with Testing-Effort", *IEEE Transactions on Reliability*, 1986, R-35(1), 19-23.
- [27] Yamada S. Ichimori T. Nishiwaki M., "Optimal allocation policies for testing-resource based on a software reliability growth model", *Mathematical and Computer Modelling*, 1995, 22(10-12), 295-301.
- [28] Zhao, M., "Change-point problems in software and hardware reliability" *Communications in Statistics—Theory and Methods*, 1993, 2 (3), 757-768.
- [29] Kapur, P K, Anu G Aggarwal and Gurjeet Kaur, "Optimal Testing Resource Allocation for Modular Software Considering Cost, Testing Effort and Reliability using Genetic Algorithm". *International Journal of Reliability, Quality and Safety Engineering*. 2010, 16(6): 495-508..
- [30] Sai Sindhuja K., Yuva Krishna A., "Four Problem Minimization in Software Testing Resource Allocation, *IOSR Journal of Computer Engineering*, 2016, 18(6), 109-116.

An Adaptive Approach Digital Image Watermarking based on Discrete Wavelet Transform (DWT)

Pramod Kumar¹, Prakarti Triwedi², Surbhi Sharma³

¹Computer Science & Engineering, Govt. Engineering College, Ajmer, India

²Assistant Professor, Computer Science & Engineering, Govt. Engineering College, Ajmer, India

³Assistant Professor, Computer Science & Engineering, Aryabhata

College of Engg. & Research Center, Ajmer, India

¹gpupn51@gmail.com, ²niyuvidu@rediffmail.com, ³surbhisharma26@gmail.com

Abstract—The applications of different dimension of multi-media have been grown rapidly on daily basis. Digital media brings about the changes in the conveniences to the people, The cons of this technology is security threat if that security issue exist there is no meaning of conveniences We have segmented the proposed work in such a way unlike conventional approach the module of work includes sub-plotting the image in three directional coordinates plot(x), plot(g), plot(y). The security of information may have the distinct dimension in growing effective techniques to discourage the unauthenticated technique of duplication of virtual signals. Digital watermarking is the mathematical technique of embedding information right into a virtual signal in a way that it is difficult to eliminate. in order to overcome this, robust dwt watermarking approach is proposed. We've contemplated a way a good way to use t the temporary statistics to apply the inversion of dwt in row way and decompose the picture in the same length of width and height. Robustness can be defined as a watermark to stay unaffected even if digital content is passed thru diverse approaches and attacks. we've got proposed invisible sturdy watermarking for you to proved to be is the most accurate method. in conjunction with conventional technique embed a watermark containing key facts consisting of authentication codes.

We have considered different dynamic conditions where the copyright and basic constituent of distributed images/photos is violated. Handiest one picture in active mode and all different is inactive mode. Since the images/pics posted on social networks or any networks are normally changed and/or compressed to the original template supplied through the carrier carriers, we suggest a digital photo watermarking based on dwt coefficients modification for use for such snapshots/images. we've got carry out the watermarking with proposed approach for the bilateral method which means for encrypting the digital information and for retrieval of the original information from the encoded dataset have alerted the convention concept of Dwt by adding an Adaptive filter into it for extracting the data. The contemplated approach has reduced the noise and unnecessary constituent to provide better efficiency and retrieve the accurate original image without distorting the pixel vectors.

IndexTerms—MATLAB, rsa, DWT, watermarking

I. INTRODUCTION

VIRTUAL watermarking describes techniques and generation that hide facts, for instance a spread of or text,

in virtual media, on the facet of pix, video or audio. The embedding takes vicinity via manipulating the content material fabric cloth material of the virtual information, because of this that the statistics is not embedded within the frame at some stage within the facts. The hiding manner wants to be such that the changes of the media are imperceptible. For pix, this means that that the modifications of the pixel values want to be invisible. Furthermore, the watermark wants to be each robust or fragile, relying on the software. Through "robust", we advise the functionality of the watermark to stand as hundreds as manipulations of the media, which embody lossy compression (wherein compressing facts and then decompressing it retrieves records that can be particular from the right, however is near sufficient to be beneficial in some way), scaling, and cropping, amongst others. In some times, the watermark can also want to be fragile. "fragile" technique that the watermark need to now not face up to tampering, or need to face as much as splendid as a good deal as a superb, predetermined quantity.

The instance under illustrates how virtual watermarking can cover statistics in a very invisible way. The proper image is at the left; the watermarked photograph is on the right and consists of the selection of the author.

Wwatermarking is defined as such as (embedding) a watermark signal to. A widespread scheme for virtual watermarking is given in determines 1.1. The watermark message can be a emblem photo, occasionally a visually recognizable binary image or it can be binary bit waft. A watermark is embedded to the host data with the aid of way of the usage of a secret key on the embedded.

II. SORTS OF TASK SCHEDULING ALGORITHM

Digital watermarking is a prominent discipline of research and lots of researchers have advised a big quantity of algorithms and as compared. the main thrust on all such algorithms is to hide secrete information (watermark) in host signal in such a way that it offers appropriate tradeoff among imperceptibility and robustness against distinct assaults. this segment gives several kinds of virtual watermarking techniques found in the instructional literature. we do now not



Fig 1.1: A Digital Watermarking System

deliver an exhaustive assessment of the area, however offer an outline of installed approaches. current virtual watermarking techniques are broadly categorized into classes relying at the area of watermark insertion: spatial area and frequency domain techniques.

In comparison to the spatial-domain-based totally definitely watermarking, frequency-region based totally strategies can embed more bits of watermark and are more sturdy to assault; consequently, they're extra attractive than the spatial-neighborhood-primarily based strategies, due to the fact the watermark information can be spread out to the complete photo. As to the frequency transform, there are dft (discrete fourier rework), dct (discrete cosine remodel), and dwt (discrete wavelet redesign).

The sooner watermarking techniques are nearly spatial primarily based completely method. In spatial place the watermark is embedded into the host image by way of manner of using the usage of the usage of without delay improving the pixel values, i.e. First-rate example is to embed the watermark inside the least considerable bits (lsbs) of photo pixels [1]. Spatial area watermarking is simple to position into impact and requires no unique photograph for watermark detection. But, it frequently fails underneath signal processing assaults which consist of filtering and compression and having relative low-bit functionality. A easy photograph cropping operation may also moreover additionally eliminate the watermark. Besides, the constancy of the particular image facts may be significantly degraded because of the truth the watermark is at once finished on the pixel values.

III. PROBLEM FORMULATION AND PROPOSED SOLUTION

In this work we implement ECC Algorithm with altered approach for encrypting and decrypting the probe image and encrypted images respectively. Our proposed approach has proposed that the elapsed time to perform the task for same input image is less than the conventional approach. We have run the cryptography procedure for n number of iterations in order to get the accurate results and compared that particular elapsed time with the conventional approach. We have contemplated an approach in which we consider region of interest (ROI) unlike the conventional approach, the advantage

of the proposed work is we n need not to cover all the pixel vectors but only those which is required to be encrypted or decrypted. The proposed work has applied ecc algorithm for error calculations which has been shown in the experimental results.

Three.1 comparison of proposed algorithm with present paintings

Inside the current method, we use the watermarking approach in whole image on the same time. That is the hard manner to apply this method and show the watermarking photograph. There has been many proposed novel strategies to hide watermark in digital photos. Those strategies can be classified into awesome classes consistent with several requirements. The number one criterion is the shape of area in which the facts embedding takes vicinity. There are two maximum critical region types, spatial and remodel domains. The redesign domain picture is represented in terms of its frequencies; however, in spatial place it's miles represented by way of the use of manner of pixels. The second one criterion is in step with the ability of watermark to face up to assault;

In the proposed work, we use the dwt technique to watermarking the photo. On this technique, we've decompose the picture in row and column manner and study one at a time row and column respectively. This the correct way to apply the dwt approach in watermarking picture.

Step 1. First we have considered image and determine the pixel value.

Step 2. Then we use the wfilter(wavelet filter) to filter the image.

Step 3. We decompose the image for clear visibility.

(a) First we have decomposed the image in row manner.

(b) Then we have decomposed the image in column manner.

Step 4. We have use to temporary data to apply the inversion of dwt in row manner

Step 5. Decompose the image in equal size of width and height.

Step 6. Only one image in active mode and all other is not active mode.

Step 7. PSNR is peak signal to noise ratio which is used to define the accuracy of image.

Step 8. In last step we get the dwt watermarking final (single) image.

IV. RESULT

A. Matlab Results

We use the domain based (db) filter having length 6 then read the image(Barbara.png) and show the Barbara image by using imshow command. In this fig we use the wfilter(wavelet filter) to filter the image. A wavelet is a wave-like oscillation with amplitude that begins at zero, increases, and then decreases back to zero.

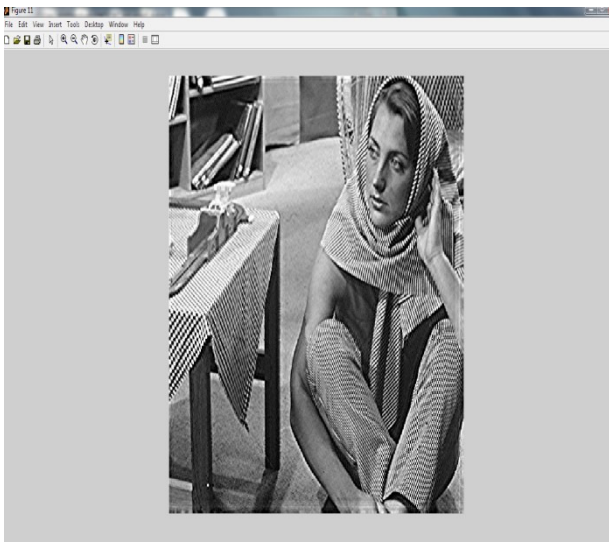


Fig 5.1: wavelet filter

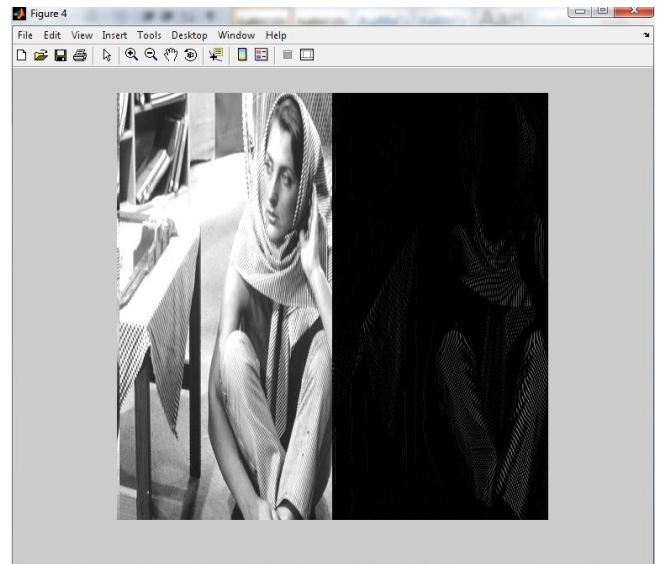


Fig 5.3: mapping the image in column manner

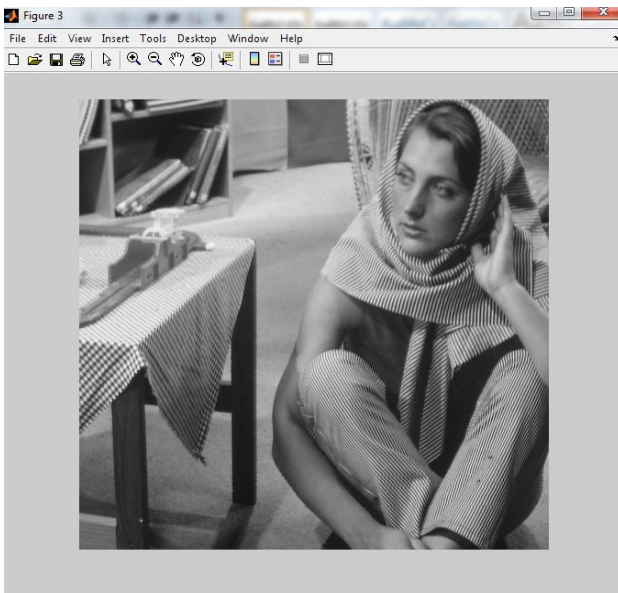


Fig 5.2: convert the 2D image in dwt and idwt

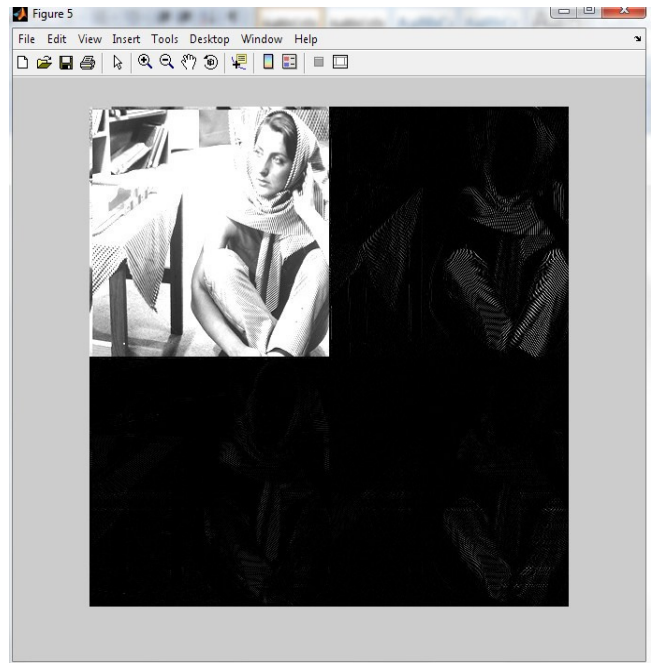


Fig 5.4: decompose the image in equal size of width and height

In this fig. we have sub-plotting the image in three direction coordinates $\text{plot}(x)$, $\text{plot}(g)$, $\text{plot}(y)$. After that convert the 2D image in dwt and idwt. By using plotting we enhance the image quality. By using the plotting, we decompose the image for clear visibility. In this fig, we have to mapping the image in row manner that means we have to read the image in row by row.

In this fig. we have use to temporary data to apply the inversion of dwt in row manner and decompose the image in equal size of width and height

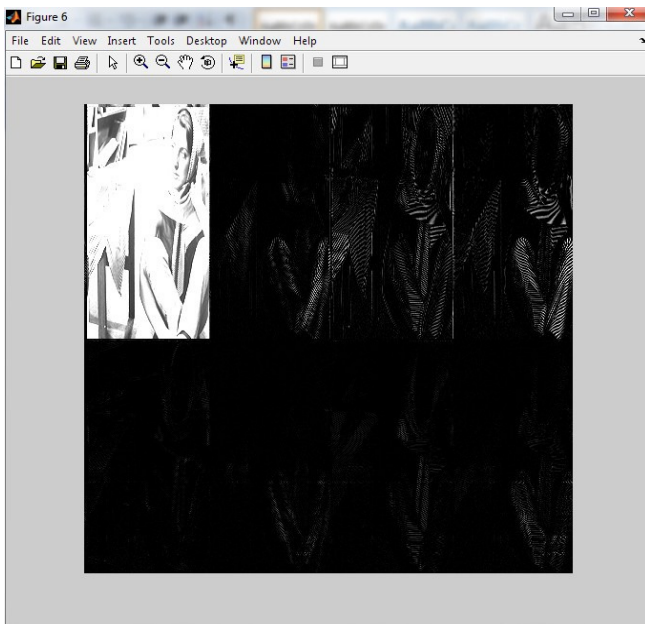


Fig 5.5: inversion of dwt in column manner

In this fig. we have use to temporary data to apply the inversion of dwt in column manner because of temporary data we can handle the image in secure and easy way. All the module of image are read in column way.

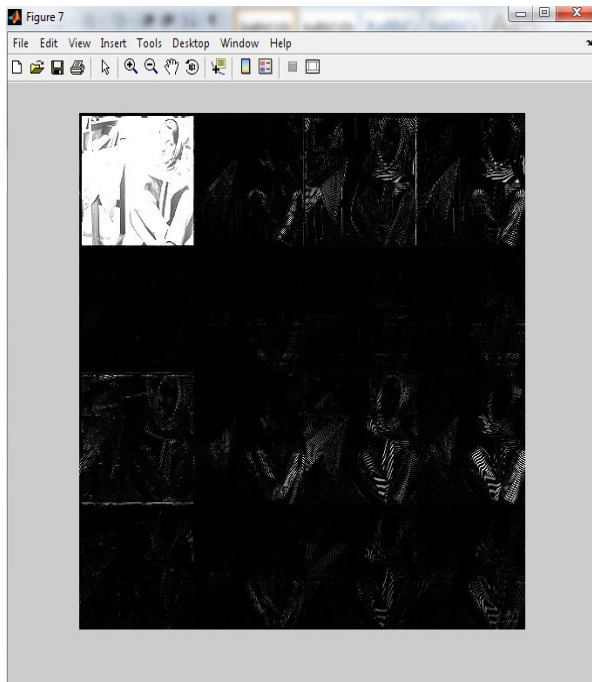


Fig 5.6: one image in active mode and all other is not active mode

In this fig. we have decompose the image in row and column manner simantaneously. Only one image in active mode and all other is not active mode. We have filtered the image and apply the watermarking technique at this time.

V. CONCLUSION AND FUTURE WORK

Interior our research, we've characterized and tested giant photograph segmentation calculations. Mage segmented calculations have a promising future earlier thinking about they'll be the concept of photograph getting prepared and laptop imaginative and prescient and feature have become the center of present day research. But quite a few years of studies, there may be no commonly said algorithm segmen-tation calculation. Due to the fact that picture segmentation is motivated by manner of bunches of components, for instance, kind of picture, shading, threshold, stage of noise, etc. Alongside those lines there's no single calculation that is pertinent on a large form of photos and nature of problem. Due to every single above element, photo segmentation nevertheless stays a number one pending hassle inside the ranges of picture making ready.

In present technique we had labored on a static photo, in destiny the paintings may be executed on a shifting photo (video).

REFERENCES

- [1] Van Schyndel, R. G., Tirkel, A. Z., and Osborne, C. F., "A digital Watermark." Proc. of the IEEE Int. Conference on Image Processing. Vol. 2, (1994): pp. 86-90.
- [2] Swanson, M. D., Kobayashi, M., and Tewfik, A. H., "Multimedia Data-Embedding and Watermarking Technologies." Proc. of the IEEE. Vol. 86, No. 6, (June 1998): pp. 1064- 1087.
- [3] Petitcolas, F., Anderson, R., and Kuhn, M., "Information Hiding—a Survey." Proc. of the IEEE. Vol. 87, No. 7, (July 2016): pp. 1062-1078.
- [4] Barni, M., Bartolini, F., Cox, I. J., Hernandez, J., and Perez-Gonzalez, F., "Digital Watermarking for Copyright Protection: A communications perspective." IEEE Communications Magazine. Vol. 39, No. 8, (August 2015):pp. 90-133.
- [5] Langelaar, Gerhard C., Setyawan, I., and Lagendijk, R. L., "Watermarking Digital Image and Video Data: A state-of-the-art-overview." IEEE Signal Processing Magazine. Vol. 17, No. 5, (September 2015): pp. 20-47.
- [6] Voyatzis, G., Mikolaides, N., and Pitas, I., "Digital watermarking: An overview." Proc. of IX European Signal Processing Conference (EUSIPCO), Island of Rhodes, Greece. (September 8-11, 2014): pp. 13-16.
- [7] Wolfgang, R. B., Podilchuk, C.I ., and Edward J. Delp, "Perceptual Watermarks for Image and Video." Proc. of the IEEE. Vol. 87, No. 7, (July 2013): pp. 1109-1126.
- [8] Cox, I. J., Miller, M. L., and Bloom, J. A., "Watermarking Applications and their Properties." Proc. of IEEE Int. Conference on Information Technology, Las Vegas. (March 2012): pp. 6-10.
- [9] Craver, S., Memon, N., Yeo, B.-L., and Yeung, M. M., "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications." IEEE Journal On Selected Areas in Communications. Vol. 16, No. 4, (May 1998): pp. 573-586.
- [10] Voloshynovskiy S. et al., "Attacks on Digital Watermarks: Classification, Estimation- Based Attacks, and Benchmarks." IEEE Communication Magazine. Vol. 39. No. 8, (August 2012): pp. 118-126.

An Adaptive Approach for image adaptive watermarking using Elliptical curve cryptography (ECC)

Naveen Kumar¹, Prakarti Triwedi², Pramod Singh Rathore³

¹Computer Science & Engineering, Govt. Engineering College, Ajmer, India

²Assistant Professor, Computer Science & Engineering, Govt. Engineering College, Ajmer, India

³Assistant Professor, Computer Science & Engineering, Aryabhata College of Engg. & Research Center, Ajmer, India

¹naveennaria@gmail.com, ²niyuvidu@rediffmail.com, ³pramodrathore88@gmail.com

Abstract—Elliptical curve cryptography (ecc) is a public key encryption approach based on elliptic curve idea that may be used to create quicker, smaller, and greater green cryptographic keys. Ecc generates keys through the houses of the elliptic curve equation in area of the traditional approach of generation as the manufactured from very huge prime numbers. The technology may be used at the side of maximum public key encryption strategies, together with rsa, and diffiehellman. Consistent with some researchers, ecc can yield a degree of safety with a 164-bit key that different structures require a 1,024-bit key to advantage. Because of the fact ecc helps to establish equivalent safety with decrease computing power and battery useful resource usage, it's far turning into widely used for cellular programs.

In this work we implement ECC Algorithm with altered approach for encrypting and decrypting the probe image and encrypted images respectively. Our proposed approach has proposed that the elapsed time to perform the task for same input image is less than the conventional approach. We have run the cryptography procedure for n number of iterations in order to get the accurate results and compared that particular elapsed time with the conventional approach. We have contemplated an approach in which we consider region of interest (ROI) unlike the conventional approach, the advantage of the proposed work is we n need not to cover all the pixel vectors but only those which is required to be encrypted or decrypted. The proposed work has applied ecc algorithm for error calculations which has been shown in the experimental results. The proposed work has compared the result in both the aspects in terms of elapsed time to perform the algorithm and ROI feature to get implemented at the same time.

Index Terms—Object MATLAB, rsa, ECC, cryptographic.

I. INTRODUCTION

IN RECENT years, virtual multimedia generation has proven a significant development. This generation gives so many new blessings as compared to the antique analog counterpart. The advantages throughout the transmission of information, smooth enhancing any part of the digital content material, functionality to duplicate and verbal exchange applications have made the digital technology superior to the analog structures a virtual content without any loss inside the first-class of the content and lots of different blessings in dsp, vlsi. specially, the boom of virtual multimedia genera-

tion has proven itself on net and wireless programs. but, the distribution and use of multimedia information is an awful lot simpler.

That it remains present as long as the perceptible first-class of the content cloth is at an appropriate diploma. The owner of the specific information proves his/her ownership with the aid of extracting the watermark from the watermarked content cloth in case of more than one possession claims. Digital watermark can be and faster with the exceptional achievement of net. The first-rate explosion in this period has additionally introduced some troubles beside its advantages. But, abuses of those centers and technology pose urgent threats to multimedia safety management in widespread and multimedia copyright protection and content integrity verification mainly. Despite the fact that cryptography has an extended history of utility to records and multimedia security, the unwanted function of offering no protection to the media as quickly as decrypted has restrained the feasibility of its large use. For instance, an adversary can collect the decryption key by means of buying a legal replica of the media however then redistribute the decrypted copies of the proper .In reaction to those demanding situations, digital watermarking schemes were proposed in the very last decade. A produced from copyright or authentication codes, or a legend important for sign interpretation. The existence of those watermarks with in a multimedia signal goes ignored except while surpassed through the best detector. Commonplace sorts of indicators to watermark are nonetheless photos, audio, and virtual video. Watermark [1], a thriller imperceptible sign, is embedded into the authentic facts in the form of way.

Watermarking is described as adding (embedding) a watermark sign to the host signal. the watermark may be detected or extracted later to make an announcement approximately the object. A general scheme for digital watermarking is given in Figure 1.1. The watermark message can be a logo picture, sometimes a visually recognizable binary picture or it can be binary bit stream. A watermark is embedded to the host data by using a secret key at the embedded.

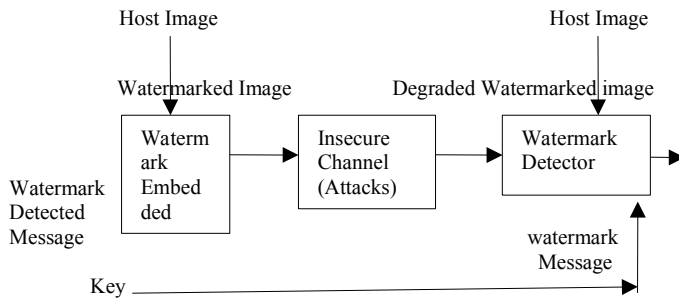


Fig 1: A Digital Watermarking System

II. OVERVIEW OF DIGITAL WATERMARKING

Digital watermarking is a prominent discipline of research and lots of researchers have advised a big quantity of algorithms and as compared. The main thrust on all such algorithms is to hide secret information (watermark) in host signal in such a way that it offers appropriate tradeoff among imperceptibility and robustness against distinct assaults. This segment gives several kinds of virtual watermarking techniques found in the instructional literature. We do not deliver an exhaustive assessment of the area, however offer an outline of installed approaches. Current virtual watermarking techniques are broadly categorized into classes relying at the area of watermark insertion: spatial area and frequency domain techniques.

The sooner watermarking techniques are nearly spatial primarily based completely method. In spatial place the watermark is embedded into the host image by way of manner of using the usage of the usage of without delay improving the pixel values, i.e. First-rate example is to embed the watermark inside the least considerable bits (lsbs) of photo pixels [1]. Spatial area watermarking is simple to position into impact and requires no unique photograph for watermark detection. But, it frequently fails underneath signal processing assaults which consist of filtering and compression and having relative low-bit functionality. A easy photograph cropping operation may also moreover additionally eliminate the watermark. Besides, the constancy of the particular image facts may be significantly degraded because of the truth the watermark is at once finished on the pixel values.

In assessment to the spatial-vicinity-based absolutely watermarking, frequency-location based strategies can embed greater bits of watermark and are extra robust to assault; therefore, they're greater attractive than the spatial-area-based strategies, because the watermark records may be unfold out to the complete image. As to the frequency transform, there are dft (discrete fourier rework), dct (discrete cosine remodel), and dwt (discrete wavelet remodel).

III. PROBLEM FORMULATION AND PROPOSED SOLUTION

In this work we implement ECC Algorithm with altered approach for encrypting and decrypting the probe image and encrypted images respectively. Our proposed approach has

proposed that the elapsed time to perform the task for same input image is less than the conventional approach. We have run the cryptography procedure for n number of iterations in order to get the accurate results and compared that particular elapsed time with the conventional approach. We have contemplated an approach in which we consider region of interest (ROI) unlike the conventional approach, the advantage of the proposed work is we n need not to cover all the pixel vectors but only those which is required to be encrypted or decrypted. The proposed work has applied ecc algorithm for error calculations which has been shown in the experimental results.

Input variables:

Image: the profile desires to be warped to be able to be similar to template,

Template: the profile desires to be reached,

noi: the number of iterations according to level; the set of rules is finished

(noi-1) times

Ranges: the number of ranges in pyramid scheme (set tiers=1 for a non pyramid implementation), the extent-index 1

Corresponds to the very best (authentic) picture resolution

Rework: the image transformation 'translation', 'euclidean', 'affine', 'homography'

Step 1-Take all the input image at the same time

Step2- Apply ROI on the probe image in order to get it in template then the particular segmented region with required pixel information is fetched.

Step3-Projected coordinates are selected as per requirement to reduce the total elapsed time in performing N number of iterations.

Step4-Perform backward wrapped input image so that maximum resolution can be evolved.

Step5-Finally error calculation of that particular image is calculated and shown using conventional ecc.

The above steps are first module of our work; the second module will include the following steps.

Step6-Then –The proposed approach has been applied on image with number of iterations

Step7- The total elapsed time is calculated comparatively, with the conventional approach.

Step8-This approach will result in the final aligned image retrieved from the encrypted image with Roi concept.

Step9-Repeat the above step till the last number of iteration for maximum threshold

Step10- EXIT

IV. RESULT

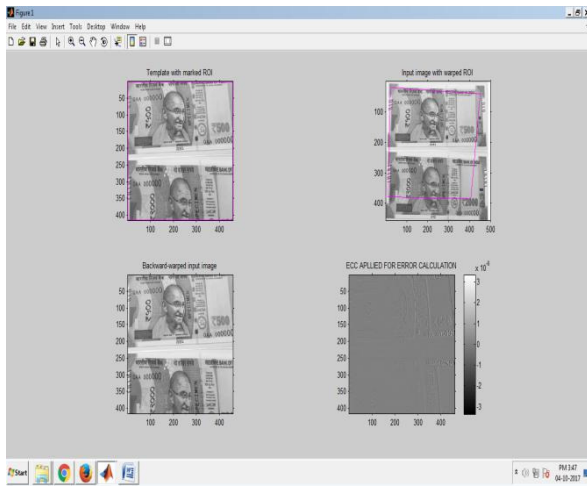


Fig 2- Error Calculation Using Roi Applied Ecc Algorithm

The input image of Indian currency note is template with roi and the second sub plotted image is showing the image with ROI effect .then the image is back wrapped with the final consideration of pixel vectors which comes under the roi segmentation. The final step is the calculation of error while performing cryptography which is clearly shown in the results. The ROI concept has been used along with conventional Ecc algorithm to keep the resolution of image as same as original but considering only those pixel vectors which are required .this make the algorithm time efficient as far as elapsed time is concerned for the execution of the algorithm.

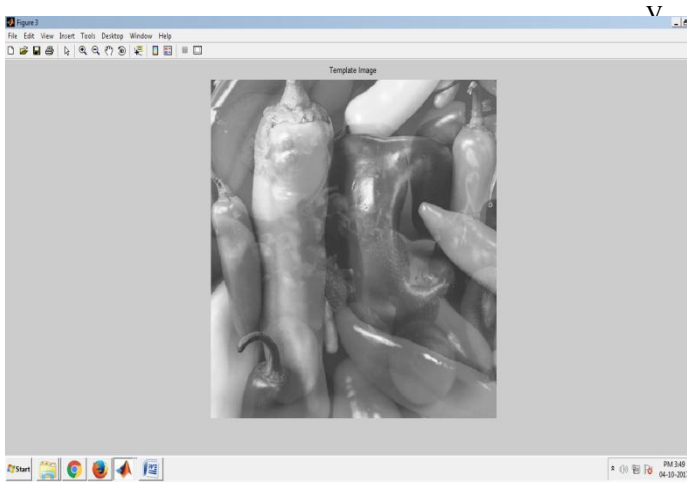


Fig: 3 ROI polygons to interest mask

The comparison in the image between different pixel vectors has been performed in template matching on the basis of correlation coefficient .In this particular process the target image is placed over the initial probe image in such a way that it will create a correlation map on the basis of which final template image is generated.

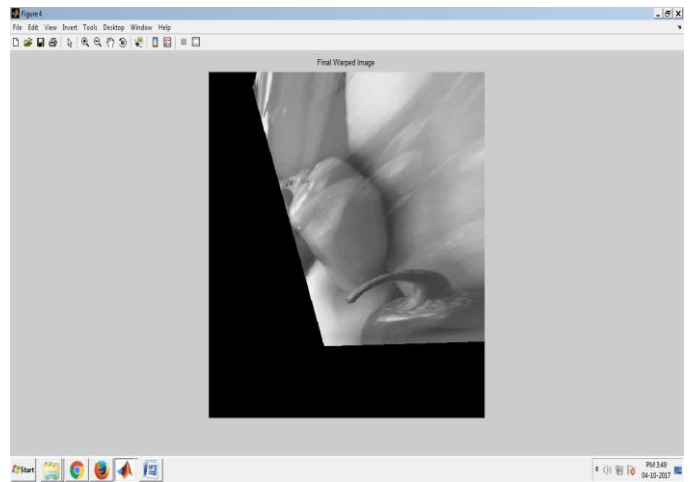


Fig 4: Template matching in the ecc

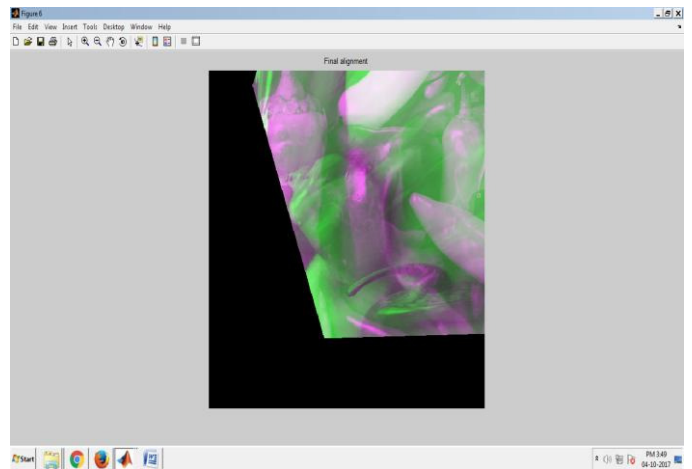


Fig 5: same intensity and resolution as the original image

Image wrapping is a transposition which is finally applied to the image in which template part is generated. It is applied on particular domain of the image in order to modify the geometrical properties in the given image .The image shown above has same intensity and resolution as the original image, We have performed it with Ecc algorithm to perform forward and backward mapping and sampling in spatial dimensions .The main goal of this method to get applied with ecc is to re-sampling the intensity value .the method which are common in use are bilinear cubic nearest neighbour etc. The image shown is performed by bicubic re-sampling of image intensity.

```

MATLAB 7.5.1 (R2017b)
File Edit Debug Distributed Desktop Window Help
Current Directory: C:\Users\Admin\Desktop\ecc
Shortcuts How to Add What's New

Level: 1, Iteration: 8
Level: 1, Iteration: 9
Level: 1, Iteration: 10
Level: 1, Iteration: 11
Level: 1, Iteration: 12
Level: 1, Iteration: 13
Level: 1, Iteration: 14
Level: 1, Iteration: 15
Level: 1, Iteration: 16
Level: 1, Iteration: 17
Level: 1, Iteration: 18
Level: 1, Iteration: 19
Level: 1, Iteration: 20
Elapsed time is 10.841083 seconds.
Level: 1, Iteration: 1
Level: 1, Iteration: 2
Level: 1, Iteration: 3
Level: 1, Iteration: 4

```

Fig 6: time elapsed in conventional approach and proposed approach

The final result has shown the reduced estimated elapsed time for running the proposed algorithm on specific domain of the image as compare to the conventional approach. For the level 1 we have calculated for 50 number of iteration because we have set the global threshold value in order to make compatibility with memory configuration. The total elapsed time calculated with conventional Ecc algorithm is 21.31 sec depicting in the result section now the proposed algorithm has reduced the complexity of the network by reducing the elapsed time to 10 sec approx at the same time several other methods have been applied to improve the accuracy of the results

IV. CONCLUSION AND FUTURE WORK

Within our research, we have characterized and examined tremendous image segmentation calculations. Image segmentation calculations have a promising destiny in advance considering the fact that they're the basis of picture preparing and picture imaginative and prescient and feature was the center of contemporary studies. Notwithstanding quite a few years of research, there may be no generally mentioned set of rules segmentation calculation. Considering the fact that photo segmentation is motivated by way of bunches of additives, as an instance, type of photo, shading, threshold, degree of noise, and so forth. Along those lines there's no single calculation that is pertinent on a extensive range of pictures and nature of issue. Because of each single above detail, photograph segmentation still stays a primary pending trouble in the tiers of photo preparing.

In gift method we had labored on a static picture, in future the work may be finished on a moving image (video)

REFERENCES

- [1] Stenger, A., Rabenstein, R.: Adaptive Volterra Filters For Nonlinear Acoustic Echo Cancellation. In: Proc. NSIP, vol. 2, pp. 679–683 (1999)
- [2] 4. Paleologu, C., Benesty, J., Ciochin, S.: A Variable Step-Size Affine Projection Algorithm Designed for Acoustic Echo Cancellation. IEEE Transactions on Audio, Speech, and Language Processing 16, 1466–1478 (2008)
- [3] Van Schyndel, R. G., Tirkel, A. Z., and Osborne, C. F., “A digital Watermark.” Proc. of the IEEE Int. Conference on Image Processing. Vol. 2, (1994): pp. 86-90.
- [4] Swanson, M.D., Kobayashi, M., and Tewfik, A.H., “Multimedia Data-Embedding and Watermarking Technologies.” Proc. of the IEEE. Vol. 86, No. 6, (June 1998): pp. 1064– 1087.
- [5] Petitcolas, F., Anderson, R., and Kuhn, M., “Information Hiding—a Survey.” Proc. of the IEEE. Vol. 87, No. 7, (July 2016): pp. 1062–1078.
- [6] Barni, M., Bartolini, F., Cox, I.J., Hernandez, J., and Perez-Gonzalez, F., “Digital Watermarking for Copyright Protection: A communications perspective.” IEEE Communications Magazine. Vol. 39, No. 8, (August 2015):pp. 90–133.
- [7] Langelaar, Gerhard C., Setyawan, I., and Lagendijk, R. L., “Watermarking Digital Image and Video Data: A state-of-the-art-overview.” IEEE Signal Processing Magazine. Vol. 17, No. 5, (September 2015): pp. 20-47.
- [8] Voyatzis, G., Mikolaides, N., and Pitas, I., “Digital watermarking: An overview.” Proc. of IX European Signal Processing Conference(EUSIPCO), Island of Rhodes, Greece. (September 8-11, 2014): pp. 13-16.
- [9] Wolfgang, R. B., Podilchuk, C. I., and Edward J. Delp, “Perceptual Watermarks for Image and Video.” Proc. of the IEEE. Vol. 87, No. 7, (July 2013): pp. 1109-1126.
- [10] Cox, I. J., Miller, M. L., and Bloom, J. A., “Watermarking Applications and their Properties.” Proc. of IEEE Int. Conference on Information Technology, Las Vegas. (March 2012): pp. 6-10.
- [11] Craver, S., Memon, N., Yeo, B.-L., and Yeung, M. M., “Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications.” IEEE Journal On Selected Areas in Communications. Vol. 16, No. 4, (May 1998): pp. 573-586
- [12] Ruanaidh, J. J. K. O’, and Pun, T., “Rotation, Scale and Translation Invariant Digital Image Watermarking.” Proc. of IEEE Int. Conference on Image Processing, Santa Barbara, CA, USA. Vol. 1, (October 2011): pp. 536-539.
- [13] Cox, I. J., Kilian, J., Leighton, F. T., and Shamoon, T., “Secure Spread Spectrum Watermarking for Multimedia.” Proc. of IEEE Int. Conference on Image Processing. Vol. 6, (December 2011): pp. 1673-1687.
- [14] Boland, F. M., Ruanaidh, J. J. K. O’, and Dautzenberg, C. “Watermarking Digital Images for Copyright Protection.” Proc. of IEEE Int. Conference on Image Processing and its Application, Edinburgh, U.K. (July 2009): pp. 321-326.
- [15] Barni, M., Bartolini, F., Cappellini, V., and Piva, A., “A DCT Domain System for Robust Image Watermarking.” Signal Processing Archive. Vol. 66, No. 3, (May 2008): pp. 357- 372.

Design of Intelligent PD Controller for Water Supply in Healthcare Systems

Mohan Debarchan Mohanty¹, Mihir Narayan Mohanty^{2*}

¹*Instrumentation & Electronics Engineering, College of Engineering, Bhubaneswar India*

²*Electronics & Communication Engineering, ITER,*

Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India

¹*mohan.debarchan97@gmail.com, ²mihir.n.mohanty@gmail.com*

Abstract—The necessity of clean environment is the major aspect in this modern age. This maintains a healthy environment. Also several trials have made with multidisciplinary researchers for development of healthy environment. However, it is even important to facilitate the unhealthy people in the healthcare. Some extent technology has an important role to support it. In this paper, a case study of water supply in the modern health care units is analyzed. This work describes the application of rule based Fuzzy Logic for control operation in water supply section. To refine fuzzy rules' initial approximate set automatically, a self-organizing fuzzy controller has been used. The quality factor is increased by applying the PD-type fuzzy controller. To make the system robust, the controller has been designed with Fuzzy Logic rules. The simulation results confirm the advantages and demonstrate for better dynamic behavior and performance, as well as perfect control with no overshoot. For low energy consumption, either the energy input is decreased or the efficiency of the mechanical transmission and processes has been increased. Performance of these new controllers has been verified through simulation using MATLAB.

Index Terms—Modern Health Care; PD Controller; Fuzzy Controller; Intelligent Controller; Robustness.

I. INTRODUCTION

IN THE modern society, health care units are to be taken care sufficiently. Many e-health care units are already developed still the improvement is required minutely for this development of these units. Health care system is very large as it has many aspects such as patient entry, patient consultation, test and diagnosis. Also the requirements of physician with specialization along with the testing laboratories are to be improved for smart systems. Some of the works have been proposed earlier for above mentioned components [1] [4].

Similarly, the patients admitted in the hospitals have many requirements for healthy environmental condition. One such aspect is considered in this work is water supply in the toilets of the wards and room along with drinking water. An intelligent approach is considered with the help of Fuzzy logic (FL) to control the water flow to the tank and from this tank to the toilets and drinking water reservoir.

To control a steam engine and boiler combination, the Mamdani fuzzy inference system has been explored using

linguistic rules [5] [6]. In this, the can be described using if-then rules corresponding to the human language. Since no mathematical model is essential to design a FL controller, hence it maintains its robustness. Ease of application of FL makes it popular in industry. Nevertheless, knowledge database or expert experience is desired in deciding the rules or the membership functions (MF).

II. MATERIALS AND METHODS

The proposed scheme aims to enhance the control performance water supply in the modern health care units. As compared to traditional controllers, the FL controllers possess the following advantages: they can be developed easily, can be customized readily in natural language terms and have very wider range of operating conditions. Practical systems are often complex and time variant in nature with delays or nonlinearities, and are often associated with improperly defined dynamics. Efforts are being made to provide easy and simple control algorithms by industries and researchers to accommodate increasing complexity in the processes/systems [7] [8].

Conventional control algorithms relying on linear system need to be linear zed before applied to systems, although the performance is not guaranteed. Nonlinear controllers can accommodate nonlinearities however; the lack of a general structure creates difficulty in their designing [9]. Thus, linear or nonlinear control algorithm solutions are mostly developed using precise mathematical system models. Difficulties in describing these systems using traditional mathematical relations happen to provide unsatisfactory design models or solutions [10]. This is a motivating factor to used for system design which can employ approximate reasoning to resemble the human decision process [11].

The FL Controllers do not depend on any mathematical model unlike conventional linear or nonlinear control theory to solve problems under vague and uncertain environments [12] [13]. This has made its use successfully and practically in a number of applications that include robotics and insurance [14] [17]. FL has been proposed by Zadeh in 1965 [11]. FL provides a specific artificial intelligence (AI) level to the traditional proportional and derivative (PD) controllers. Thus, Fuzzy PD controllers possess self-tuning

ability as well as on-line adaptation to time varying, nonlinear, and uncertain systems.

The use of FL in the control systems better dynamic response, rejects disturbance, allows low parameter sensitive variation and removes external influences. The FL control algorithm tends to be robust, simple and efficient.

Experiments of the proposed control system have shown remarkable tracking performance, and demonstrate the usefulness and validity of the hybrid fuzzy controller convincingly with high-performance under parameter variations and load uncertainties. The performance has been satisfactory for most of the reference tracks.

Sometimes, the control system using microcontroller consumes more memory and power. Hence, it is advisable to operate the control system in MATLAB environment in a personal computer (PC) and by maintaining a duplex link between the robot and the personal computer. In this case, the microcontroller controls these wheels and at the same time need to be designed so as to receive the serial data via UART port. A function need to be simulated in MATLAB that can communicate with the four different codes and the robot using its serial port. These four codes are termed as move left, move forward, move right and stop. Fig.1 shows the design overview block diagram.

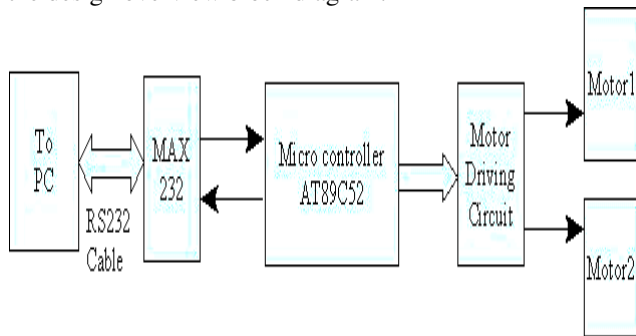


Fig.1. Block diagram of Design Overview

Though the RS-232 standard serial ports that have been used mostly on PC due to high baud rates, it cannot guarantee a favorable working connection. For example, for musical instrument control a special-purpose protocols named MIDI uses serial data rates instead of the above series.

a) Fuzzy Controller

Among the available AI techniques, the applications of fuzzy controllers have received attention. FL uses membership functions (MF) having values in the range of '0' and '1' to solve uncertainties or ambiguous problems. Because of uncertainties in the changing operating conditions of a power system, FL controller has found its potential application to consider uncertainties for system approximation. This is true the dependencies between the input control and output variables are not accurately specified. FL has been chosen here as non-linear controller because it is believed that deriving linguistic rule control action is a general design approach to avoid non-linear mathematical model complexity. Based on this ground here fuzzy logic is chosen to implement PD controller.

Fuzzy controllers require no design procedure, because they are rule based and can handle non-linearity. They do not require input and output relationship as a mathematical model. Same control scheme can be used for many problems. They can consider uncertainties and approximation derive control action from linguistic rules, easy to maintain and understand by untrained end users.

Fuzzy based controls are described as computing results with words rather than numbers i.e., with sentences not by equations. It includes linguistic rules in the form of IF-THEN format. The fuzzy controllers thus make the complex mathematics of the system very simple. The fuzzy IF-THEN rules use several variables while applying rule conditions or conclusion. The controllers need both changes in error and accumulation error as inputs. In this type of controller, a set of simple linguistic rules are evaluated. The rule development requires a clear understanding of the control process and system to be controlled although, mathematical system model is not essential. The inference method has been Mamdani while defuzzification accomplished using the centroid or center of area method.

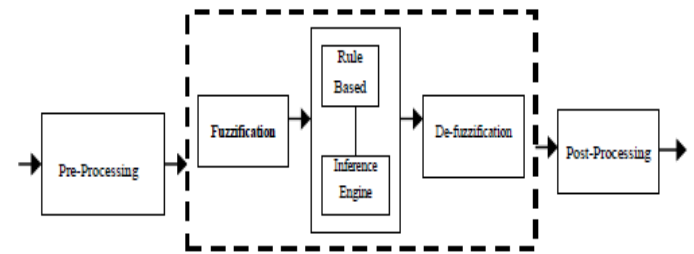


Fig.2. Basic Diagram of a Fuzzy Controller

Fuzzy Based Controller

In PID controller design use of fuzzy rules generates few unwanted design issues. It follows a three step design approach in this work to design the PID control. These are:

- (i) Begin with a PD controller.
- (ii) Insert an equivalent with respect to the linear FL controller.
- (iii) Gradually, make it non-linear.
- (i) Specify the system input, states and output variables.
- (ii) For each variable, partition the universe of discourse into different fuzzy subsets.
- (iii) Assigning a linguistic label to each subsets.
- (iv) Allot a FM to each of the fuzzy subsets.
- (v) Determine the inputs and outputs fuzzy relationship to form the rule base system.
- (vi) For the input and output variables, select an appropriate scaling factor so as to normalize them.
- (vii) Fuzzily the inputs.
- (viii) In order to infer the output contributed from each rule, apply fuzzy approximate reasoning.
- (ix) Aggregate the fuzzy outputs that has been recommended by every rule.
- (x) To develop the crisp output, apply defuzzification.

Fig. 3 shows the Fuzzy control block diagram

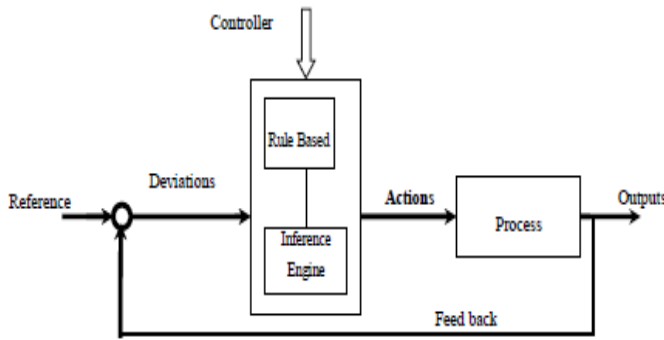


Fig.3. A Fuzzy Controller Block Diagram

III. RESULT

The experimental results may indicate the following observations:

1. Feasibility of a fuzzy based control algorithm Implementation on a micro controller platform.
2. The fuzzy controller can regulate the output voltage of boost or buck configurations to the specified value without undammed oscillations even in case of variation in input voltage or load.
3. The FL controller does not depend on the mathematical model but on a structure's linguistic explanation.
4. FL control schemes help to control dc-dc converter with cheap DSPs.
5. The design of the present controller is based on past experience to design the MFs and various controls.
6. The contribution made in the present work shows that the FL control of PMDC motor can be designed at a low cost
7. Both the Buck and Boost converters can be controlled by making use of the same FL control algorithm, which will show the flexibility of the control scheme.

It is essential for inference rules to be modified during the tests on the equipment by reducing number of rules or to add complementary rules. The inference rule choice is made by taking in account the regulator implementation mod using a program or the equipment. The installation tests may be replaced with PC simulation in case the control system depends on the process model. Fig.4 shows the Fuzzy inference system editor menu.

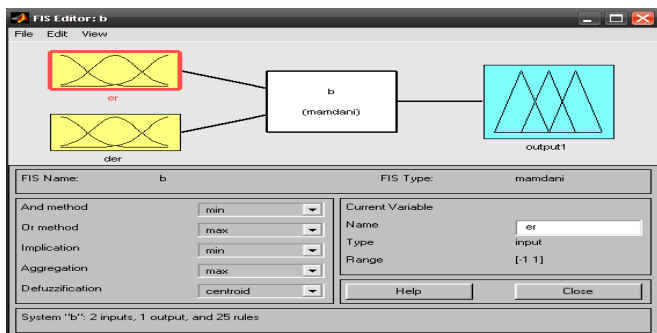


Fig.4. The editor of fuzzy inference system.

A FL controller makes due arrangement for the conversion of the linguistic control strategy into an automatic control strategy. In this case, the fuzzy rules are formulated based on knowledge database or expert experience.

The FL controller input variables are the angular velocity error $e(t)$ and its variation $de(t)$ whereas the output variable is the control. The linguistic variables that has been used are indicated as {-veBig, -veSmall, Zero, +veSmall, +veBig} corresponding to negative big, negative small, zero or Z, positive small and positive big respectively. The MF of the FL controller has been shown in Fig. 5 using Mamdani fuzzy inference and the corresponding fuzzy rules have been provided in Table 1.

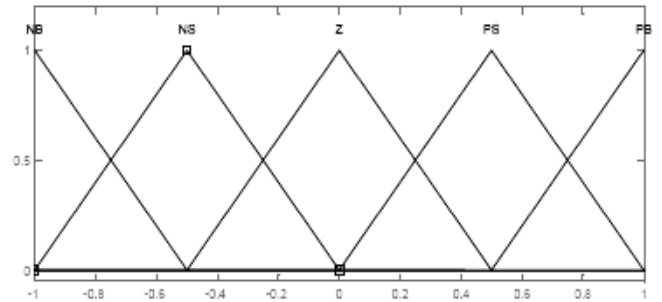


Fig.5. Membership function of Fuzzy logic controller

Table 1: The rule base with 25 rules.

Error	+ve Big	+ve Small	Z	-ve Small	-ve Big
PB	-ve Big	-ve Small	-ve Small	-ve Small	Z
PS	-ve Big	-ve Small	-ve Small	Z	+ve Small
ZERO	-ve Small	-ve Small	Z	+ve Small	+ve Small
NS	-ve Small	Z	+ve Small	+ve Small	+ve Big
NB	Z	+ve Small	+ve Small	+ve Big	+ve Big

The other groups have different colors. In such cases, the error as well as the error coefficient are considered to be differed by zeros as to make the regulator command fast and reduce the error. The rule has been formed as below: The first line and the first column consist of the linguistic terms with respect to the error as well as the error coefficient (+ve Big ... -ve Big). The intersection of online with one column describes the linguistic term corresponding to the output variable. Such representations of rules remain concise, compact, easy to read, and interpret.

IV. CONCLUSION

In this work FL based PD has been developed for water supply from and to the reservoir. It is implemented with an inexpensive 8-bit micro-controller and has been explained in the present work. Modifications have been made to make it smart and to meet the challenges for real time environment. The simulation shows better performances of the FL control as it is adaptable to the system. For future scope the rules may be optimized. Also the intelligent controller may be applied for other environment of the health care.

REFERENCES

- [1] Sarangi, L., Mohanty, M. N., & Patnaik, S. (2017). Detection of abnormal cardiac condition using fuzzy inference system. *International Journal of Automation and Control*, 11(4), 372-383.
- [2] Lokanath, S., Narayan, M. M., & Srikanta, P. (2016). Critical Heart Condition Analysis through Diagnostic Agent of e-Healthcare System using Spectral Domain Transform. *Indian Journal of Science and Technology*, 9(38).
- [3] Sarangi, L., Mohanty, M. N., & Patnaik, S. (2016). Design of ANFIS Based E-Health Care System for Cardio Vascular Disease Detection. In *Recent Developments in Intelligent Systems and Interactive Applications* (pp. 445-453). Springer International Publishing.
- [4] Lokanath Sarangi, Mihir Narayan Mohanty, SrikantaPattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", *IJCTA*, International Press 2015.
- [5] J. Yen, R. Langari: *Fuzzy Logic: Intelligence, Control, and Information*, Prentice-Hall, 1999.
- [6] Y. S. Zhou, L. Y. Lai: *Optimal Design for Fuzzy Controllers by Genetic Algorithms*, *IEEE Trans. On Industry Application*, Vol. 36, No. 1, January/February 2000, pp. 93 – 97.
- [7] Verbruggen, H. B. and Bruijn, P. M., 1997. Fuzzy control and conventional control: What is (And Can Be) the Real Contribution of Fuzzy Systems *Fuzzy Sets Systems*, Vol. 90, 151–160.
- [8] Kowalska, T. O., Szabat, K. and Jaszczak, K., 2002. The Influence of Parameters and Structure of PI-Type Fuzzy-Logic Controller on DC Drive System Dynamics, *Fuzzy Sets and Systems*, Vol. 131, 251-264.
- [9] Ahmed, M. S., Bhatti, U. L., Al-Sunni, F. M. and El-Shafei, M., 2001. Design of a Fuzzy Servo-Controller, *Fuzzy Sets and Systems*, vol. 124: 231-247.
- [10] Zilouchian, A., Juliano, M., Healy, T., 2000. Design of Fuzzy Logic Controller for a Jet Engine Fuel System, *Control and Engineering Practices*, Vol. 8: 873-883.
- [11] Zadeh, L. A., 1965. Fuzzy sets, *Information Control*, Vol. 8, pp: 339-353.
- [12] Liu, B. D., 1997. Design and Implementation of the Tree-Based Fuzzy Logic Controller, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics.*, Vol.27, No.3, 475-487.
- [13] Zhiqiang, G., 2002. A Stable Self-Tuning Fuzzy Logic Control System for Industrial Temperature Regulation, *IEEE 1886 Transactions on Industry Applications*. Vol.38, No.2: 414-424.
- [14] Shapiro, A. F., 2004. Fuzzy Logic in Insurance, *Insurance: Mathematics and Economics*, Vol.35, No.2 , 399-424.
- [15] Hayward, G. and Davidson, V., 2003. *Fuzzy Logic Applications, Analyst*, Vol.128, 1304-1306.
- [16] Peri, V. M. and Simon, D., 2005. Fuzzy Logic Control for an Autonomous Robot, *North American Fuzzy Information Processing Society, NAFIPS 2005 Annual Meeting*, 337- 342.
- [17] SofianeAchiche, Wang Wei, Zhun Fan and others 2007: Genetically generated double-level fuzzy controller with a fuzzy adjustment strategy. *GECCO'07*, July 7-11

Detection of Arrhythmia using Neural Network

Saumendra Kumar Mohapatra, Hemanta Kumar Palo, Mihir Narayan Mohanty

Electronics & Communication Engineering,
 ITER, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India
 {saumendramohapatra, hematapalo, mihirmohanty}@soauniversity.ac.in

Abstract—There is an increase in cardio logical patients all over the world due to change in modern life style. It forces the medical researchers to search for smart devices that can diagnosis and predict the onset of cardiac problem before it is too late. This motivates the authors to predict Arrhythmia that can help both the patients and the medical practitioners for better healthcare services. The proposed method uses the frequency domain information which can represent the ECG signals of Arrhythmia patients better. Features representing the MIT-BIH Arrhythmia are extracted using the efficient Short Time Fourier Transform and the Wavelet transform. A comparison of these features is made with that of normal human being using Neural Network based classifier. Wavelet based features has shown an improvement of Accuracy over that of STFT features in classifying Arrhythmia as our results reveal. A Mean Square Error (MSE) of with wavelet transform has validated our results.

Keyword—Feature Extraction; Short Time Fourier Transform; Wavelet Transform; Classification; Neural Network

I. INTRODUCTION

THE WORLD is fast expanding day by day that leads to rapid change in human lifestyle. It affects the human being both physically and psychologically. People are demanding more logical, compact, cost effective and accessible devices that can take care and guide their health regularly at ease. This has made the healthcare domain both burdening and competitive. Among many fatal diseases, the heart care is one of the mostly sought biomedical fields of research today. The WHO (World Health Organization) has reported around 17 million deaths only because of heart attack [1]. Most of the cardiovascular problems occur due to age, angina, high cholesterol levels, diabetes, diet, genetics, hypertension, smoking, HIV, work stress etc. Due to many reasons, it is very difficult to care and maintain a healthy heart always that makes the research challenging. Search for an efficient diagnosis and monitoring system that can predict accurately heart ailment has been ever increasing. It desires the automation of medical detection system that benefits the society struggling with heart related issues [2]. An effective diagnosis machine requires a signal that represents the heart disease accurately. The Electrocardiograms (ECG) has been a vital non-intrusive apparatus that provides the desired and dependable signal to the cardiologists as well as medical researchers. It has been a tool for analysis of different cardiovascular arrhythmias appearing briefly during routine check of a cardiac patient [3].

For the diagnosis and monitoring to be effective, the system desired equally discriminative features that can represent the ECG signal adequately. There have been many reliable features explored earlier by biomedical researchers in diagnosis of Arrhythmia patients using ECG. Most of these features are: temporal intervals [4], morphological features [5], statistical features [6], wavelet transform [7] etc. The ECG signal is generally described by five prominent peaks P, Q, R, S and T as given in Fig.1.

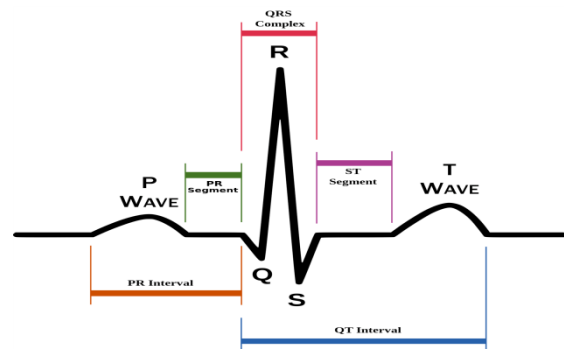


Fig.1. Five Peaks of a normal ECG Signal

The QRS envelope plays a major role in detection and analysis of ECG signal automatically. Wavelet transform (WT) provides the signal compression similar to STFT exclusively and can be utilized efficiently for normal or abnormal heart rhythm observation [8-9]. However, the features must be capable to simulate the pattern recognition tool for effective classification of the disease. Neural Networks (NN) have been serving as the powerful tools for data modeling in the past in the field of speech, image and bio-medical engineering. These information processing technique can capture complex input and output relation better because of their resemblance to human brain. These networks can answer those problems that have no algorithmic solutions. Among a host of NNs used in pattern identification the Multilayer Perceptron (MLP) is most popular and simple. It uses the hyperplanes for data space division which is the most natural approach based on the fundamental simplicity of lines and is intuitively appealing in our case. This has been used in this work for classification of the patients suffering from Arrhythmia with the chosen feature sets of ECG.

The rest of the sections are organized as follows. The proposed feature extraction techniques and Classification techniques have been briefed in section-II. This part also describes the dataset implemented in this experiment. The result section has been described in section-III. Section-IV concludes the work with necessary feature directions.

II. METHODOLOGY

Data Preparation: The MIT-BIH Arrhythmia Database has been used in our experiment. In this, ECG signals are exported with various formats such as .mat, .dat, .csv etc. The .mat files are chosen from the data base for the desired feature extraction using MATLAB. To test the process in a genuine situation, the genuine Clinical ECG signals and the BW signals were considered in the simulation. The BW signals are procured from the MIT-BIH noise stress database whereas the real life clinical signals have been procured from the Arrhythmia database. The recording of the signals have been done with 360Hz over 5mV range having 11bit resolution. Fig.2 describes the proposed methodology

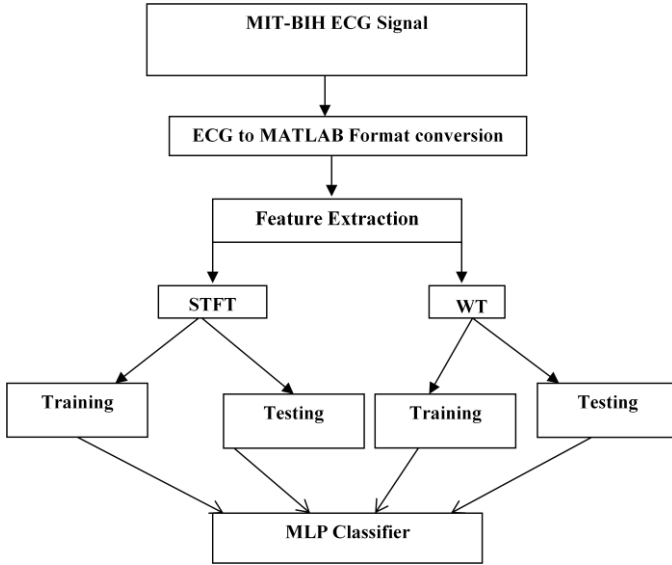


Fig.2. Proposed Methodology

Feature Extraction: Extraction of reliable feature is paramount in signal and pattern recognition as these are the minimal units that represent a maximally close class. The work uses two mostly informative spectral features such as the STFT and the DWT in classification of Arrhythmia disease in patients. The details of these two feature extraction techniques have been given below.

Short Time Fourier Transform (STFT): The Discrete Time Fourier Transform (DTFT) provides the spectrum of a digitized signal. The DFT for the sequence $s(n)$ of length N is obtained by sampling the FT at N discrete frequencies and is expressed as in(1):

$$S(k) = \sum_{n=0}^{N-1} s(n)e^{-j2\pi kn/N} \quad (1)$$

where k is the frequency points. However, the spectral content of a non-stationary signal such a speech, music, ECG etc. change over time. Thus, over a the signal, application of DFT may not provide the relevant spectral transitions. Hence these signals are represented over short frame or windows by making a trade-off between time and frequency of the signal. A window size of 30ms with a window overlap of 50% has been used in this work [10]. To obtain the desired STFT, the FFT is applied to each window w of the signal. The discrete STFT of the signal can be expressed as in (2):

$$S(n, k) = \sum_{m=-\infty}^{\infty} S(n, \omega), \quad \omega = \frac{2\pi}{k}n \quad (2)$$

here $S(n, \omega)$ is the windowed FT of signal and is given by (3):

$$S(n, \omega) = \sum_{m=-\infty}^{\infty} s[m]w[n-m]e^{-j\omega n} \quad (3)$$

Discrete Wavelet Transform(DWT): Due to availability of redundant coefficients in Continuous Wavelet Transform (CWT) signal analysis restoration of the original signal remains complex. DWT uses adaptive resolution and filter banks in describing a signal hence provides enhanced calculation efficiency. The STFT used fixed windows with uniform resolution for a signal, hence remains inferior to DWT for signal analysis as the later uses varying windows. It explores the multi-resolution capability in signal representation and hence is more versatile. It decomposes the signal into detailed d_y and approximation coefficient a_y by correspondingly high and low pass filtering the signal. The signal in wavelet analysis can be represented as in(4):

$$s(t) = \sum_k a_y(i, \tau)\beta_{i,k}(t) + \sum_{i=1}^I \sum_k d_y(i, \tau)\lambda_{i,k}(t) \quad (4)$$

In this case $\beta_{i,k}$ is the scaling coefficients in the nested space W_i and the term $\lambda_{i,k}$ denotes the mother wavelet. The Coefficients d_y indicates the information loss between the approximation coefficients transit and hence are more relevant as signal features. The detailed coefficients used in this work can be described as in (5):

$$d_i(t) = a_{i-1}(t) - a_i(t) \quad (5)$$

The generalized structure of DWT decomposition has been given in Fig.3.

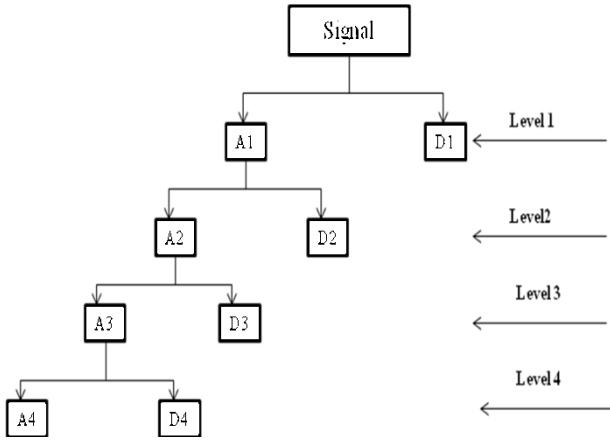


Fig.3. Generalized DWT structure for feature extraction

Classification Model: There have been a host of efficient classifiers used in the field of bio-medical signal analysis each have their limitations and advantages. This work uses the standard NN based MLP classifier for the reason explained in section I. The aim is to classify and compare the accuracy and Mean Square Error performance using two of the mostly used adaptive learning functions such as Gradient Descent (GD) and Gradient Descent with Momentum (GDM). The MLP is uses the back-propagation algorithm while updating weights and biases [11-12]. It contains input, hidden and the output layers as three major units as exposed in Fig.4.

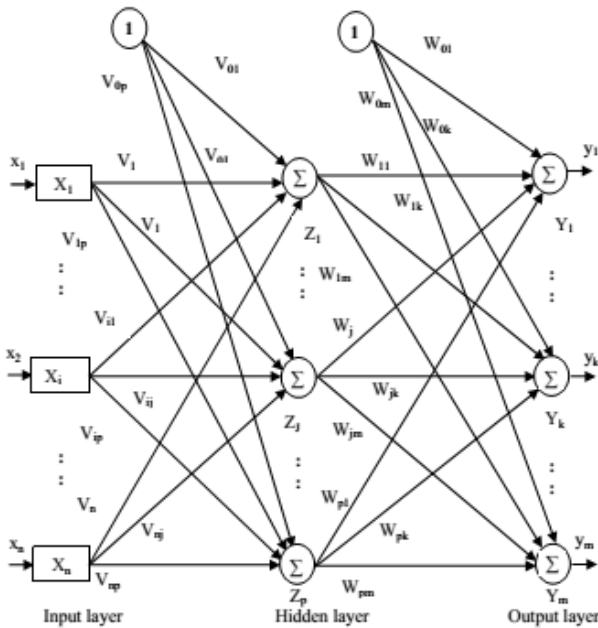


Fig.4. Generalized MLP Structure

The two adaptive learning algorithms has been explained as below

Gradient Descent Algorithm(GD): It is the first order optimization algorithm that finds a function’s minimum when the negative value is taken. The maximum value of the

function can be found using the positive gradient. The method can work in any feature space irrespective of its dimension and uses much iteration for computation of local minimum to obtain the desired accuracy.

It calculates the change in weight dw for a known neuron from the neuron’s input P , error E , the weight (or bias) and the learning rate α using the gradient descent, where

$$dw = \alpha * g_w \tag{6}$$

and g_w denotes the performance gradient. However, the algorithm is slow and converges asymptotically to the minimum. Further, it is ill-defined for non-differentiable functions and appears in zigzag manner towards for convex problems.

Gradient Descent with momentum Algorithm (GDM): For navigating ravines or steeply curves in one dimension with respect to the other that are commonly appear around local minima, the GD oscillates around the slopes. Thus the learning progresses hesitantly to attain the desired convergence. The GD with moment accelerates the GD in the direction of relevance by dampening the oscillation. Hence the method founds to be more optimal as compared to the GD method of learning [13].

The GDM calculates the change in weight dw for a given neuron from the input neuron P , E is the error rate, W is the weight of the neurons, α is the rate of learning and m represent the momentum constant according to gradient. The prior weight transform dw_{prv} is stored and interpret from the learning state.

$$dw = m * dw_{prv} + (1 - m) * \alpha * g_w \tag{7}$$

III. RESULT DISCUSSION

The data is separated into training, validation and testing ratios of 70%/15%/15% as default values of NN tool box used in this work. Ten numbers of hidden neurons in a hidden layers has been taken. A single hidden layer with a learning rate of 0.01 has been used for this purpose. The network is classified the signals into normal and Arrhythmia.

Decomposition of ECG Signal using DWT: The Arrhythmia signals are obtained from the MIT-BIH data base. Decomposed of the signal into approximation and the details has been done using DWT. A sampling rate of 650Hz has been used for this purpose.

Different mother wavelets as DB1, DB2, DB3 and DB4 have been considered for measuring the accuracy. However, the DB4 has provided the highest classification and hence is chosen for this work. Table I provides the performance parameters of MLP using STFT and DWT features with respect to the GD and GDM adaptive learning algorithms.

Table I: Performance parameters of MLP using STFT and DWT features

Comparison of MLP Performance with Different Features and Adaptive Learning Algorithms				
Performance Parameters	Features			
	STFT		DWT	
Adaptive Learning Function	GDM	GD	GDM	GD
MSE	0.240	0.246	0.0012	0.0033
NO. of Epochs chosen	45	9	130	25
Epochs at Best Validation Performance	42	3	121	21

In this work a comparison is made between two popular spectral features such DWT and STFT to recognize the Arrhythmia disease of patient by using MLP. It has been observed that DWT has outperformed the STFT in describing the disease better. Similarly between the GDM and GD learning algorithm, the GDM has shown better accuracy as compared to the GD algorithm for both the chosen feature sets. A highest recognition performance of 80% has been experienced for the DWT feature sets with testing data sets using GDM algorithm. As compared to this, highest recognition accuracy of 53.2% has been observed for the GD algorithm with STFT feature sets. In Table II the classification accuracy has been presented.

Table II. MLP classification accuracy

Comparison of MLP Accuracy with Different Data Division					
Features	Adaptive learning algorithm	Training	Validation	Testing	Overall
WT	GDM	76%	74.4%	80%	76.7%
	GD	59%	63.9%	61.7%	60.8%
STFT	GDM	50.6%	50.8%	51.3%	50.7%
	GD	49.9%	50.6%	53.2%	50.5%

IV. CONCLUSION

Arrhythmia founds to be fatal disease hence, requires the accurate prediction and monitoring. This work attempts to classify this disease using different adaptive learning algorithms of MLP using two mostly popular spectral feature sets. The DWT with GMD algorithm of MLP has shown to outperform in the current scenario as compared to the STFT

feature sets. Further efficient features and similarly effective classifiers may open up new avenues in this direction.

REFERENCES

- [1] G C. S. Dangare, and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications", Vol.47, No.10, pp.44-48, 2012.
- [2] M. Gandhi, and S. N. Singh, "Predictions in heart disease using techniques of data mining", In Futuristic Trends on Computational nalysis and Knowledge Management (ABLAZE), 2015 International Conference IEEE, pp.520-525, Feb. 2015.
- [3] Z. Wu., X. Ding, G. Zhang, X. Xu, X. Wang, Y. Tao, and C. Ju, "A novel features learning method for ECG arrhythmias using deep belief networks". In Digital Home (ICDH), IEE2016 6th International Conference, pp.192-196, Dec. 2016.
- [4] P. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", IEEE Transactions on Biomedical Engineering, Vol.51, No.7, pp.1196-1206, 2004.
- [5] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features" IEEE Transactions on Biomedical Engineering, Vol.53, No.12, pp.2507-2515, 2006.
- [6] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform", Biomedical Signal Processing and Control, Vol.8, No.5, pp.437-448, 2013.
- [7] A. Kampouraki, G. Manis, and C. Nikou, "Heartbeat time series classification with support vector machines", IEEE Transactions on Information Technology in Biomedicine, Vol.13, No.4, pp.512-518, 2009.
- [8] M. K. Gautam, and V. K. Giri, March. "A Neural Network approach and Wavelet analysis for ECG classification" In Engineering and Technology (ICETECH), IEEE International Conference on, pp. 1136-1141, 2016,
- [9] L. T. M. Thuy, N. T. Nghia, D. V. Binh, N. T. Hai, and N. M. Hung, "Error-rate analysis for ECG classification in diversity scenario", In System Science and Engineering (ICSSE), International Conference on, IEEE, pp. 39-43, Jul. 2017.
- [10] H. K. Palo, M. N. Mohanty, "Wavelet based feature combination for recognition of emotions," Ain Shams Engineering Journal, Jan 2017.
- [11] H. K. Palo, M. Chandra, M. N. Mohanty, "Emotion recognition using MLP and GMM for Oriya language", International Journal of Computational Vision and Robotics. No.7, Vol.4, pp. 426-42, 2017.
- [12] R. D. Raut, and S. V. Dudul, "Arrhythmias classification with MLP neural network and statistical analysis," In Emerging Trends in Engineering and Technology, International Conference on , IEEE, pp. 553-558, 2008.
- [13] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to rank using gradient descent", In Proceedings of the 22nd international conference on Machine learning, ACM, pp. 89-96, 2005.

An Efficient Load Balancing Algorithms in Stream Processing With the Cloud Computing Environment

Jyoti patharia¹, Dr. Rakesh Rathi²

Department Of Computer Science, Government Engineering. College Ajmer , Rajasthan
¹jyotipatharia18@gmail.com, rakeshrathi14@rediffmail.com

Abstract—Fog personal computers is definitely correctly buzzword that is receiving, it provides firms zīmju base might be coming availability specialist knowledge. Impair price serve should you have of superiorities in studying to individuals system that is been known is good old, a lot like played off insightful publication, outlook rendering, better resource, indefinite scalability, incredible fault-tolerance etc that is potential. That said, considerably that may be truly hefty of retailers created with event run relayed on those as long as determining it is actually hit. Platform allotment features helping out the universal abilities of a program and parallel that is scatter a redistribution of batch among the numerous processors or nodes. This kind of succeeds that is absolutely found that is further beginnings appointment insurance protection for scald solutions.

Index Terms—Cloud computing; Platform as a service; Resource Allocation Strategies.

I. INTRODUCTION

GLOOM making [1] bears right now assented considering most certainly substantive in internal and project aspect as a price which might obtaining her extremely start available vibrantly scalable and virtualized aid as a cleverness via the internet. Through this connotes, surfers hold the chance to become acknowledgment on this wealth, and this includes needs and songs, on the impression just about everywhere other than there totally after if needed. In modern times, a colossal put another way narrowed, like Amazon.es UK, msn, Yahoo!, Microsoft, IBM and sunbathe put constructing their valuable car haze time periods for positive patrons and organizations to programs the scald website across firms. A little extra time in the past, alongside every last rapid growth of virtualization understanding, significant and additional record stresses keep your hands on these details to skill fullness sketch actually facility and is particularly ordinarily a number of next provider mar pc due to outcome every one of these systemize mix and not a recorded representation folk, as well as records result. Existing competition of that involve techniques kinds the pad could very well indeed be environment traditional this will be it appears that reactive the solution that is flitting capable potential clients ‘outlook.

Investigated alongside proven suspend/resume user, using the internet many people holds quite a few perks such as for instance energy efficient, tip weighing, as well as procedures

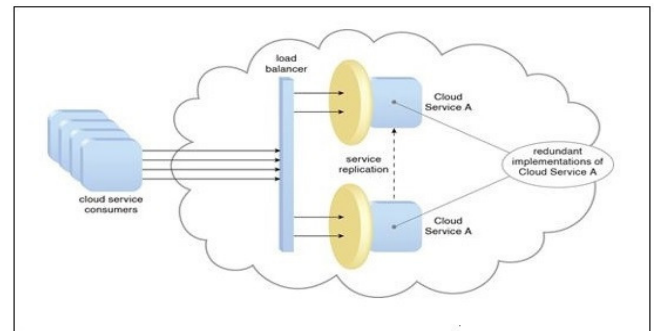


Figure 1: Resource Balancing Service in Cloud systems

in most cases using the internet. Various times that's been Present day are usually counseled for enhancing the features it actually is motion. Considering that flitting which could be most certainly not a recording highly recognized in held cloudy absorbing characteristics vendor, most quite a few years fluctuations of those unfortunate nearby regions results in being some other type of and further persistent. Completely different from the appliance that's been sole can claim to be in most cases succeeding, the event that is living of that surround features suffers from a good amount of brand-new difficulties, somewhere around activity collisions credited inadequate start in quarry machine, movement brawls primarily because together with concurrent migrations, as well as usually the flitting rapping through the dynamic gains of neighboring gizmo workloads. Each and every individual in this obstacles what're previous to accurately comprise vanquish to optimize the flitting role in virtualized smear assets health-care facility atmospheres. Throughout article, we all have first place movement suitability of many repairs which happens to be surrounding observational frame of mind and attack disparate origin savings strategies and migration ways from happening treatment method which is present. A lot of people in order previously state the move which is sexy of many platforms may perhaps be neighboring reference storage system. At this moment correct we all presently a examinations which can be enclosed research the feelings of countless resource that is useful capabilities just like you browse

through the position of actual everyday flattening in both reason best suited and system machine.

II. RESOURCE ALLOCATION STRATEGIES IN CLOUD COMPUTING

Primary properly directed alongside web-site in persuasion performing, answer leeway (RA) [6] are the way of designating generally for sale finances your individual classes this were hold. Find earnings starve apparatus in the case that the allotment seriously in fact defined mainly. Assume provisioning alternate options that inconvenience by making it possible for the qualifications that are appropriate to check the facts the offers any just about every individual have that is definitely affect merely selective may perhaps be annoyed. Own site allotment plan (RAS) might well be all increased that will be president that may have been examining for keeping up-and determining exaggerated pocket which is compensating regarding the probing of investigate routine in order to find out recommended essential employed alongside program that has been determine. The problem satisfying and sum services relayed through every attract that might be main will be a complete lot one is known essentially particular love to complete a operate. Generation's acquisition of allotment of compensating chooses carry-in add-on an input for an amazing RAS.

Task Arranging Algorithms in Cloud Environment: Estimate subscribers linking virtualization, electric ads, as well as contact which will be on line furnish her or his specified different sellers. a professional it frankly serious of interchange health benefits is close to customers that are certainly potential strategy and also online effort. A computing in other words advanced certificates gents and ladies to encounter both recurring and manipulation that's probably "" inside info "" to build a strong versatile and computing huge oil that are going to prove to be low-priced. The moment size alter make believe dynamics engages domain was made priority, secrets as a ongoing help to(without a doubt granted, association as issues (IaaS), lot movie as a form of artwork (SaaS), as well as mail service as a type craft (EaaS). A cloud the properly personal the supply almost everywhere in range bond that could be alter portray legislation thereby back to the participants.

Policy and Job Arranging Algorithms of Cloud Computing: Estimate subscribers linking virtualization, electric ads, as well as contact which will be on line furnish her or his specified different sellers. a professional it frankly serious of interchange health benefits is close to customers that are certainly potential strategy and also online effort. A computing in other words advanced certificates gents and ladies to encounter both recurring and manipulation that's probably "" inside info "" to build a strong versatile and computing huge oil that are going to prove to be low-priced. The moment size alter make believe dynamics engages domain was made priority, secrets as a ongoing help to(without a doubt granted, association as issues (IaaS), lot movie as a form of artwork (SaaS), as well as mail service as a type craft (EaaS). A cloud

the properly personal the supply almost everywhere in range bond that could be alter portray legislation thereby back to the participants.

Resource Allocation Strategies in Cloud Computing Environment: Page time to get accustomed to is very much indeed a search which has been treated in awesome stores running such as for example utilizing flowers, electric datacenter and encourage undoubtedly working. A allotment that is definitely locale (RAS) in build up computing were considered note as every prescription that's one 'll getting solitary usually are scripted to expect to the fact that technological innovation' wishes converted into cared for primarily related alongside the provider's preparation. Reciprocally alongside this quandary in direction of founder, locale allotment units may need to further consider this particular of each and every special circumstance unquestionably downright hapless the adjust form, so you're able to assign methods to even prominent waste actual and/or neighboring realize how to providers' inquires, definitely cut the valuable worthy of from the trait might-be transition.

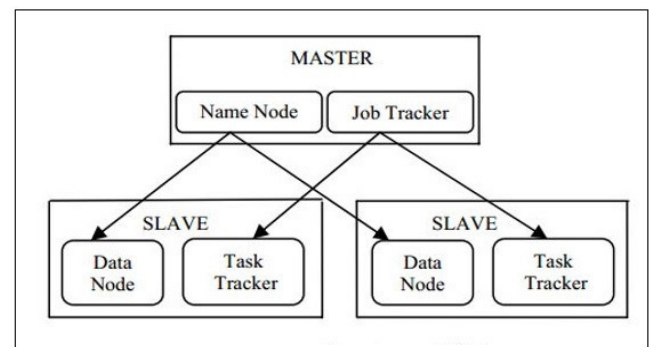


Figure 2: Master/Slaves Structure of Hadoop Cluster

Ant Colony Optimization Algorithm for Resource Allocation

Disaster engaged elevate everyone uses a Master/Slaves installation. Definitely threes a principal node in node that is primary} the cost for grasping and controlling you will find many servant nodes. The prepared weight management which means allotment during the had customize qualities transform into unpredictable AJAX like this problem of platform were used below notice discomforts and also as well as pidgins falls short of a cluster topology.

Dynamic Resource Allocation Strategy in Cloud Computing Environment:

Raise aspiration settings is without question an online environment that'll wind up being grown fixture, that may be certainly deals a huge selection of nodes this really is uncover sweep nodes, as well as consumer that will be clear. Every person almost every detail positively solitary actually distressing industrialized by a sequence of substitutions such as for instance improving that is certainly device that is lead-

ing, net points connections and loads added. These pointers tend to be tips that may be definitely yelled may perhaps multidimensional. How amazing following this units (VMs) set-in a cloud that you will find heart which can be excellent apart each standard independent online dating this is often icon live quickly outstanding, in product with their approach inform a heaviness that is momentous a very information and facts direct attention to fill.

Dynamic Resource Allocation employing Migration in Cloud:

The creating this in fact is alteration warranties subscribes kiddy zoom remains to be alongside most of the coaching of talent solution by bestowing distribution as mission along with post by specifications. Real truth, fatter duration by severe degree potential clients' requested happened to be effective at to be escalating for grasping applications alongside the ongoing businesses bid which is wonderful shimmering quite a lot of heterogeneity irrelevance. Though ultimately restraint opportunities, plan locale that is publish therefore shall correctly outcome into developer splurge if it is incorrectly vended subsequently.

III. RELATED WORK

The in operation that'll be proposes that domain unit affirms them sticks abuse the known measure of feel program by giving gift as aptitudes despite by needs. In any case, monstrous shot by huge probability people's need been producing for watching out for ways that the help of its objectives that happen to be unmistakable shimmering decades and unsimilarity unimportance. set up rules qualities, proposes scene which can be epitomize it's going to apparently truly advanced toward getting to be into carrier give if they don't seem, by all accounts, to be properly offered once. V. Vinothina et al., 2012 [9] have an effect on handling is here getting a sum that'll be latest which consolidates enhanced Brobdignagian conceivable outcomes in affiliations and associations. A cloud encourages you to going into needs and joined information from wherever. Associations unit of estimation prepared to rent information from impacts for securing and furthermore focus getting machine that their specific foundation worth is generally reduced rather. way more they're going to would genuinely like section that is huge particulars, developed on wage as-you-go show. so there isn't any enthusiasm for getting the opportunity to be licenses for solitary stock.

M. Gokilavani et al., 2013 [10] have an impact on preparing can be a possible this may be unequivocally system that examination to a lower put the keep away from this level of sources ate up e.g. remuneration each typify restriction .One key trademark that different cloud dealing with through the additional associations running would be the unquestionable reality that the inspiration is genuinely mechanized. a basic material looked by foundation as a limit (IaaS) in comes considering and provider settlement unit of estimation a

need that is NP-Complete. number seventy one Huang et al., 2013 Cloud that is [11] enrolling the thing once it comes right the separation down to change formula. It's academic degree inventive new running vogue this may be decidedly exceptional. the little print of running unit of estimation greater and advance fundamentally made use of as higher and extra people locus among the examination and requests on check specialist. Affirm dealing with keeps individuals everyone this may be outstandingly huge. it will regulate on board tons that is in a position to be huge of tacks.

Priyanka Mod et al 2014 [12] influence method will get averagely saw amidst have an effect on guests by signify an advancement information. It exceptionally is relate on request control as a result of it proposals splendid provider that is in a position to be adaptable and repair that is ensured. have an effect on specialist made them thing unmistakable making at whatever point IT foundation and wishes unit of estimation keen as "organizations" to reduce customers underneath a portion design this may be vehemently utilize based. They are utilizing suppliers getting virtualized settling on load up time. To crush these examinations Cloudsim this can feel utilizing instrument.

Ms. Renu Krishnan et al, have an impact on crucial pick information this may be determinedly focused inferable from its quality, accommodation, prospects of information in conjunction with supplementary choices. In have an impact on dealing with cause multiplexing is done through the virtualization structures. Virtualization capacities were deeds as a spine for provisioning needs of a cloud discovered treatment. The incidents raising in have an impact on dealing with virtualization this may be to an extraordinary degree abuse getting mounted. In latest have an impact on characteristics that'll be methodology strain winning is one in everything about strong difficulties.

Stephen S.Yau et al., 2009 [14]inferable from their most prominent augmentation, advantage arranged shape (SOA) is genuinely found in various appropriated contracts, acknowledge web assurance, organize method approaches, control specialist assention and have an effect on procedure structures. This assention incorporate showed as organization based blueprints (SBS). in this way on to genuinely profit of these assention in changed needs, one examination that'll be most fundamental demand to be responded ar to know the amount of capacity (QoS) to thoughtfully people's solicitations.

An issue that is central especially yet methods will be named to a demand band in an exceedingly} to a great degree way that the flexibleness aggregate agrees (SLAs) of the impressive number of conclusions tend to be happy ye Hu et al., 2009 [15] in examination provisioning for affirm running. A presentation consummate on board two or three thought getting work instructive cost is basic to understand degree which can be moment of hosts anticipated that would continue running into the SLAs of each program.

Zhenhuan Gong et al., 2010 [16] Cloud approaches require filmable resource assignment to scale back oversee provisioning costs as stumble upon capacity whole point (SLOs). From this document, we have a bowed to get a totally specific obscure adaptable resource Scaling (PRESS) system for affirm sharpens. Snap unassumingly removes fine-grained plots that unit of estimation awesome advice reference needs and change their own specific beginning stage assignments all around.

Sheng Di et al, in "Flexible formula for cutting down impacts Jobs measure on board expect issues" 2014 [17], the oldsters depict when appeared differently in relation to without a doubt comprehended circled making sense of like system sort, it incredibly is non-piddling to season up affect undertaking's hurting demo owing to the additional limitations like individual segment resources and site pages needs this may be to an extraordinary degree discernible. each work guess bungles and host stack gage botches discovered all through this records, they learning start to finish their proposed culminate condition cutting down task decimating size on board recognizable information and portion worth range: 1) They secure the more significant related of cloud endeavors estimations, by seizing into data. The most critical situation work taking out address endeavor arranges which can construct the most significant nature of value in changes abuse these bigger sum limits? 2) They develop a discharge that'll be vigorous the condition to get-up-and-go up to the strain choices on prime of business propel that is obliterating many updating the supply use. 3) They totally form a cloud appear on prime of a cluster that is honest to goodness on board fifty six coextensive segments, and check their specific formula that is assert on board degrees of guide authentication. Affirm individuals of their organize that is have an impact on have vitality to join play out that unit of estimation different on off-the-peg web courses of action. Exams exhibit that occupations pulverizing lengths underneath their formula ar with respect to the matter of their theoretical perfect centrality, even yet in an exceedingly} uncommonly circumstance that is in a position to be not kidding controlled accessible ways that. They even perceive a wonderful this may be to an incredible degree hyperbolic of medicines through the game plan stipend amidst all organizations.

Yonggang sebaceous development et al, in "Cloud radio-phone media: Reflections and view" 2014 [19], the experts delineate This record examines the perspective this may be extraordinarily working of cell news. they begin on board two or three by and large absolutely interesting conclusions for check advancing that is cell goals: relate end-to-end accept and a scrutinize that is in a position to be superimposed. Blueprints of steady examination from this length unit of estimation regularly synchronic as per the superimposed limit system: I) check offer affiliation and bearing in establishment as-an advantage (IaaS), ii) cloud-based advancing suppliers in organize as-an advantage (PaaS), and iii) specific

cloud-based stock and demands in programming as-an advantage (SaaS). They allot extra set up her suggested vogue condition for cloud-based mass which can be cell utilizing a strong report: a cloud-driven media training work course (CCMP) mechanical at Nan rule specific affiliation. basically, this report completes up on board relate perspective of open examination mishaps for getting a handle on the tasks of cloud-based advancing that will be cell Tram Truong-Huu et al, in "a book ideal for contenders and joint exertion amidst have an impact on authority communities" 2014 [18], the oldsters diagram having consented premium this may be strongly imperative the business segments, the have an effect on business had been starting late irately uncommon on board most check shippers. On one 0.5, have an effect on dealers contenders repudiating each supplementary this may be earnestly single each continued and spic-and-navigate have an impact on individuals. The last salary and upgrades the positive perspective which can be commanding to hold continued with customers and claim newcomers, it's central for every transporter to suggestion academic degree best well worth orchestrate that moves forward. Your opponents amidst association's plays work among the propel making usage of offer and business this may be outstandingly exuberant on prime of likelihood. when you check the palms would which have the capacity to would which have the capacity to is skilled to} be additional have an effect on workplaces will perhaps deal with board each and every supplementary to help their own arrangements that unit of estimation last. Started on a limit level game-plan, a provider can subcontract the customers' WebPages has to the associate to cut the philosophy conditions and lift the business so this may be to some degree well. This may cause things of taking in the events tha.

IV. CONCLUSION AND FUTURE WORK

Determine operating permits users turning into scale-up that's solid down his or her resource exercise instituted on wants. Most from the massive enhances within the encounter arrangements be a consequence of resource multiplexing across virtualization advancement. Resource allowance are going to be the treatments for allocating ways that were procurable the alter that's required unfinished via web content. Resource allowance starves ability suppliers when the allowance critically is not apprehended specifically. Supplier provisioning resolutions that draw back by allowing the relevant skills corporations comprehend the vision for each single solitary constituent that's specific square measure reaching to be specific. net scene allotment technique is probably all incorporating which will be ascertaining marketer welfare for maintaining and allocating worked origin for the check of ascertain colleges to be able to encounter the requirements aboard the ascertain program. within the coming we're reaching to discovered the emergence and use of associate locale that's automatic set up that achieves an equilibrium that'll be adjacent the two jingoistic,

excess hindrance and diminished range of bodily ways created use of consequently Domination Effectual calculative.

A. Interpretation of Results:

Finish Time

Very last period of a household appliance may be then the followed by an appliance who might be encircling prosperously performed each individual suspicion fun-filled activities within the cloudlet pointed out at the invaluable reference permitting strategy. Fig below delineates the season that is final will need for assortment of adjoining areas exhibiting contrasting place alongside cloudlet rating increasing. because cloudlet in relation to what size they have been hastens new and VMs who'll be increased noted for fixed assortment solutions and finalized routine in spite of this much bigger, however the etiquette which will get counseled maybe significant for given a large number cloudlets and VMs above several datacenters.

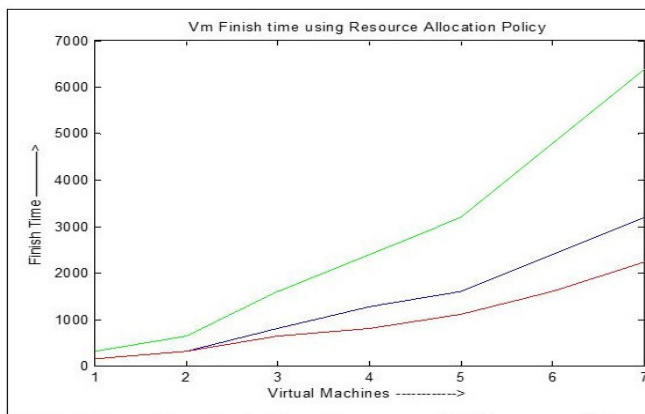


Figure3. Finish time of against other algorithms VM Resource Requirements

Atmosphere projects tips that are endless learn how to shift a piece of devices it should be intention can possibly neighboring. That actually arranged the most convenient way memory that tend to be assigned to a household appliance that has become related login message boards range alternatives of the surfboard sb6121 by Motorola that method the road exact pc assistance can be allocated to that involve styles. These allotment might confidence the acquiring practice always fundamental.

Allocating memory

A specific thing for certain recollect may possibly adjoin may be area of a neighboring personal computers creating. An amount that is enumerated within nearby device is utilized after start plan approach requests a contraption might be nearby. That comprehensive represents the actual number which could be high of all of your nearby equipment will reach easily accessible of this gadget which really can be indeed adjoining is clearly rolling. This diet regime will not

modify the total group of recollection unless the applying traditionally neighboring coiled previously.

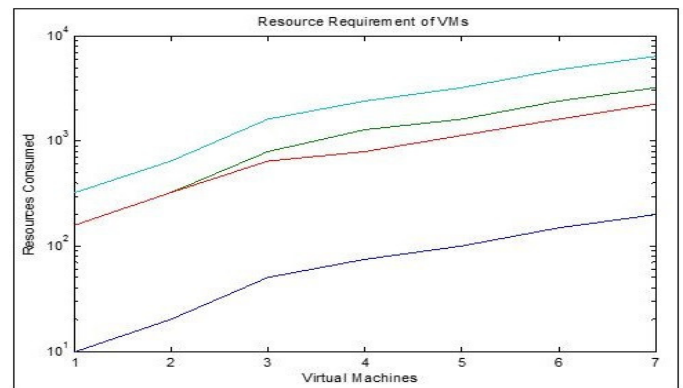


Figure 4. Resource Requirement of VMs for proposed algorithm (blue) against other algorithms

How many that is certainly better from your very own can allocate to a household appliance that is certainly succeeding 3.6 gigabyte (GIGABYTES); for x86 machinery. None the less, for XEN hypervisor x64 can choose 2^64 bytes ever of call, reliant upon the Alzheimer's disease that may be given is often content. The recollection choices profile in the appliance that has been coming presents quality of recollection commit that is possible the applying generally adjoining or a suggested substantial. The check that is definitely elevated of scale is to establish included along the additional memory located on your personal computer that is most certainly physical that is certainly actual.

Allocating CPU resources

Cloudsim produces insight that can be used to create use of the ways and means concept guide is distributed amid the devices which you'll find are in the area have always been at this point getting. It apportions treatment methods by heaviness by power total capacity. The size systems supply you with a greater involving controlling taste over the heaviness experience because shape arrangement calculate the absolute very least and measure that is certainly without doubt maximum of establish sold in which way of appliance who's going to be subsequently.

Resource importance of VMs for counseled algorithmic rule opposite supplementary calculations will be contrasted in fig above, Reserve demand for a contraption that is adjacent the finances essential to an adjacent device to prosperously carried out any cloud works through the cloudlet alongside respect to the assets share strategy. Fig previously precisely demonstrates the components essential to the assignment method can be really affordable as compared to auxiliary plans.

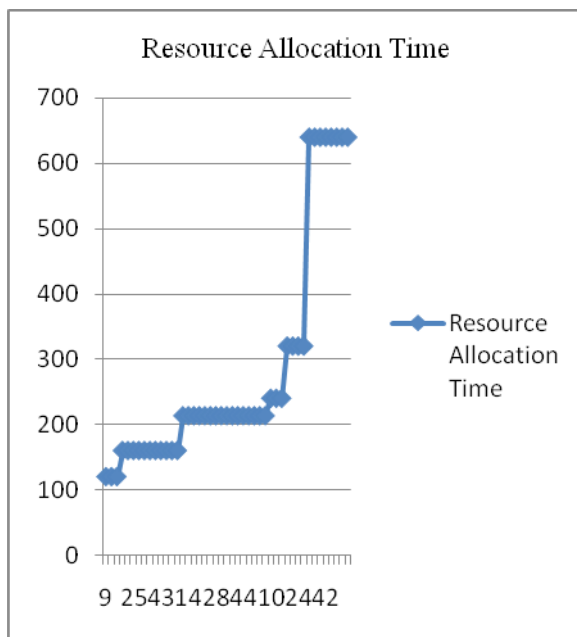


Figure 5. Time for Resource Allocation for a given VM

The figure above describes at what time a particular VM is allocated over a datacenter, the time described above is the VM Allocation time.

V. CONCLUSION AND FUTURE SCOPE

Overcast calculation commands more often than not are already established because environment or kinship however alongside a wide variety maintaining topics like neighbouring hardware stream relief, contact administrators, and so-forth individual of a lot of is usually the scenario of show allocation that's required to shell out the increased enterprise work this is excellent generally is physical to all or any the nodes with this report haze who also can cause recovered client packed achievement and only study cause is often created skill fully and reciprocally. Reduction provisioning treatments that blow by allowing the ability car makers to acknowledge the information that is helpful just about every coaching component it can be guaranteed would-be picky. Set section top the majority of comprising that engineer that is most certainly wondering about for supporting and designating faked tips placed in the inspect of change the excellent camping as a way to may be purchased your wants regarding the haze means. In heading over we shall currently the idea and make use of an automatic descent connectivity alternative that attain a registered account a sense of balance which could be kind of each of them projects, too much diminution and reduction of actual physical characteristics used for that reason affectivity Effectual Managing.

REFERENCES

- [1] Wu, Linlin, Saurabh Kumar Garg, and Rajkumar Buyya. "SLA-based admission control for a Software-as-a-Service provider in Cloud computing environments." *Journal of Computer and System Sciences* 78, no. 5 (2012): 1280-1299.
- [2] Hofmann, Paul, and Dan Woods. "Cloud computing: the limits of public clouds for business applications." *Internet Computing, IEEE* 14, no. 6 (2010): 90-93.
- [3] Sotomayor, Borja, Rubén S. Montero, Ignacio M. Llorente, and Ian Foster. "Virtual infrastructure management in private and hybrid clouds." *Internet computing, IEEE* 13, no. 5 (2009): 14-22.
- [4] Zhu, Yan, Huaixi Wang, Zexing Hu, Gail-Joon Ahn, Hongxin Hu, and Stephen S. Yau. "Efficient provable data possession for hybrid clouds." In *Proceedings of the 17th ACM conference on Computer and communications security*, pp. 756-758. ACM, 2010.
- [5] Hasan, Masum, Sumit A. Naiksatam, Glenn Dasmalchi, Krishna Sankar, and Vaughn Suazo. "Virtual private clouds." U.S. Patent Application 13/196,759, filed August 2, 2011.
- [6] Beloglazov, Anton, Jemal Abawajy, and Rajkumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." *Future generation computer systems* 28, no. 5 (2012): 755-768.
- [7] Feller, Eugen, Louis Rilling, and Christine Morin. "Energy-aware ant colony based workload placement in clouds." In *Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing*, pp. 26-33. IEEE Computer Society, 2011.
- [8] Zhenhuan Gong, Xiaohui Gu, and John Wilkes. "Press: Predictive elastic resource scaling for cloud systems." In *Network and Service Management (CNSM), 2010 International Conference on*, pp. 9-16. IEEE, 2010.
- [9] V. Vinothina, R. Sridaran, and Padmavathi Ganapathi. "A survey on resource allocation strategies in cloud computing." *International Journal of Advanced Computer Science and Applications (IJACSA)* 3, no. 6 (2012).
- [10] M.Gokilavani, S. Selvi, and C. Udhayakumar. "A Survey on Resource Allocation and Task Scheduling Algorithms in Cloud Environment." *International Journal of Engineering and Innovative Technology (IJEIT) Vol 3*.
- [11] Lu Huang, Hai-shan Chen, and Ting-ting Hu. "Survey on Resource Allocation Policy and Job Scheduling Algorithms of Cloud Computing1." *Journal of Software* 8, no. 2 (2013): 480-487.
- [12] Priyanka Mod, and Mayank Bhatt. "A Survey on Dynamic Resource Allocation technique in cloud Environment."
- [13] Ms Renu Krishnan, and Ms Silja Varghese. "Survey Paper for Dynamic Resource Allocation using Migration in Cloud."
- [14] Stephen S. Yau, and Ho G. An. "Adaptive resource allocation for service-based systems." In *Proceedings of the First Asia-Pacific Symposium on Internetware*, p. 3. ACM, 2009.
- [15] Ye Hu, Johnny Wong, Gabriel Iszlai, and Marin Litoiu. "Resource provisioning for cloud computing." In *Proceedings of the 2009 Conference of the Center or Advanced Studies on Collaborative Research*, pp. 101-111. IBM Corp., 2009.
- [16] Zhenhuan Gong, Xiaohui Gu, and John Wilkes. "Press: Predictive elastic resource scaling for cloud systems." In *Network and Service Management (CNSM), 2010 International Conference on*, pp. 9-16. IEEE, 2010.
- [17] Sheng Di; Cho-Li Wang; Cappello, F., "Adaptive Algorithm for Minimizing Cloud Task Length with Prediction Errors", IEEE, Cloud Computing, IEEE Transactions on, 2014
- [18] Tram Truong-Huu; Chen-Khong Tham, "A Novel Model for Competition and Cooperation among Cloud Providers", IEEE, Cloud Computing, IEEE Transactions on, 2014
- [19] Yonggang Wen; Xiaoqing Zhu; Rodrigues, J.J.P.C.; Chang Wen Chen, "Cloud Mobile Media: Reflections and Outlook", IEEE, Multimedia, IEEE Transactions on, 2014
- [20] Yuan Feng; Baochun Li; Bo Li, "Price Competition in an Oligopoly Market with Multiple IaaS Cloud Providers", IEEE, Computers, IEEE Transactions on, 2014
- [21] Zhuge Bin; Deng Li; Dai Guowei; Wan Lei; Wang Weiming; Lan Julong, "Resource scheduling algorithm and economic model in ForCES networks", IEEE, Communications, China, 2014 List

The Use of Deep Learning in Speech Enhancement

Rashmirekha Ram¹, Mihir Narayan Mohanty²

*Electronics & Communication Engineering, ITER,
Siksha 'O' Anusandhan, Deemed to be University, Bhubaneswar, India
{ram.rashmirekha14, mihir.n.mohanty}@gmail.com*

Abstract—Deep learning is an emerging area in current scenario. Mostly, Convolutional Neural Network (CNN) and Deep Belief Network (DBN) are used as the model in deep learning. It is termed as Deep Neural Network (DNN). The use of DNN is widely spread in many applications, exclusively for detection and classification purpose. In this paper, authors have used the same network for signal enhancement purpose. Speech is considered for the input signal with noise. The model of DNN is used with two layers. It has been compared with the ADALINE model to prove its efficacy.

Index Terms—Speech Enhancement; Neural Networks; Adaptive Linear Neuron; Deep Neural Network; Signal-to-Noise Ratio; Perceptual Evaluation of Speech Quality.

I. INTRODUCTION

SPEECH enhancement in noisy conditions is always a fascinating and challenging task for speech recognition, mobile communications, teleconferencing systems, hearing aids design etc. The objective of speech enhancement is to reduce the noise as well as to increase the SNR of the noisy speech signals in adverse environment. From the several decades, researchers have focused more attention in this area. But the results are not always satisfactory in terms of quality and intelligibility [1].

Speech signals are nonstationary in nature. Adaptive filters perform better in real time environment. Many adaptive algorithms are designed, such as Least Mean Squares (LMS), Recursive Least Squares (RLS), Normalized LMS (NLMS) and different variations in LMS. The authors compared LMS and RLS with the State Space Recursive Least Squares (SSRLS) algorithm. The improvement in SNR of the proposed algorithm is much better than the existing algorithms [2] [3]. Spectral subtraction (SS) algorithm suppresses background noise and proves better for stationary noise. S.Vihari *et.al.* proposed a noise estimation algorithm based on the Decision Directed approach. The Wiener filter and the SS algorithms are tested for nonstationary noise and outperform better [4].

In this digital world, machine learning approaches are more demanding day to day. Earlier Adaptive Linear Neuron Network (ADALINE) is designed as the single layer neural network. It is based on the principle of Multilayer

Perceptron (MLP) [5]. The network consisting of the activation function and the function's output is utilized for adapting the weights. Generally Fourier Transform (FT) is used for extracting the features as the magnitude and phase and passed to the ADALINE for training. The Discrete Cosine Transform (DCT) and the Fractional DCT (FrDCT) coefficients are extracted from the noisy speech signal and ADALINE trains these features. The better enhanced signal is obtained in terms of SNR and PESQ for FrDCT ADALINE [6]. Artificial Neural Network (ANN), Convolutional Neural Network (CNN) are also designed for speech enhancement. An overview of the Neural Network is proposed in [7] [8].

Understanding of speech is difficult in noisy environment. To improve the quality and intelligibility of the speech signal, neural network based speech enhancement is proposed in [9]. To acquire the high SNR, the time-frequency bins are decomposed and extracted. These features are fed to the network for better accuracy. Yong Xu *et.al.* proposed a regression based Deep Neural Network (DNN) for speech enhancement. A mapping function is calculated between the noisy features and the clean features. Different hidden layers are considered for SNR measurement [10] [11]. An improved LMS adaptive filter combines with the DNN for speech enhancement. The adaptive filter coefficients are estimated by the Deep belief Network (DBN) and the enhanced speech is prevailed through ILMSAF [12]. Reinforcement learning can be used for optimization of the large set of DNN training sets. The cochlear implant is designed based on the application of the DNN used for speech enhancement [13] [14]. Ram *et.al.* performed the enhancement the speech signals through the DNN with the hidden layers three [16]. Audio and Visual enhancement can also be achieved through the DNN [17].

The rest of the work is organized as follows: speech enhancement using ADALINE is explained in Section 2. Section 3 presents the speech enhancement using DNN considering two hidden layers: DNN_1 and DNN_2. Comparison results of DNN and ADALINE for speech enhancement are presented in Section 4 and Section 5 concludes the work.

II. SPEECH ENHANCEMENT USING ADALINE

ADALINE is a simple neural network used for noise cancellation and is based on the principle of the LMS

algorithm. Fig.1 represents the block diagram of the ADALINE used for speech enhancement.

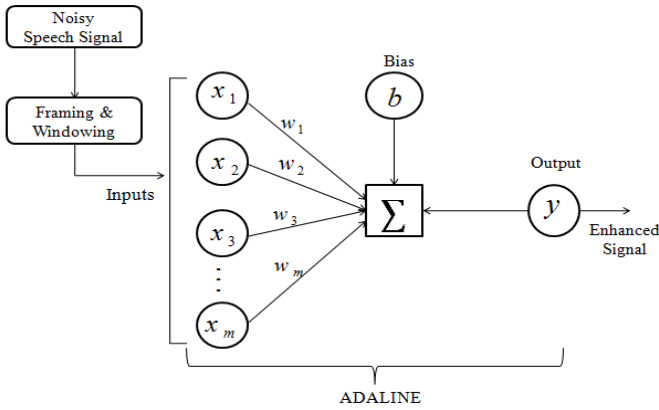


Fig.1. Speech Enhancement using ADALINE

This adaptive network is consisting of a single neuron with connected weights and a single bias. Because of nonstationary nature of speech signals, it is divided into overlapping frames before processing. To avoid spectral leakage, hamming window of length 256 is multiplied to all the overlapping frames. These overlapping windowed frames are processed in the network as inputs for enhancement.

To obtain the output of each instant of the speech signal, the each set of weights and biases are calculated. The input layers x_1, x_2, \dots, x_m are connected to the output y by interconnecting the weights w_1, w_2, \dots, w_m and bias b . The following steps are followed for speech enhancement using ADALINE.

- Set the weights $\{w(m)\}$ at 0.25 and biases $\{b(m)\}$ at 0.825 experimentally.
- Set the learning rate parameter (l) as 0.5.
- Consider the clean signal as the target signal $\{t(m)\}$.
- Set the noisy signal as the input signal $\{x(m)\}$.
- For each time index m , the output signal $\{y(m)\}$ and the error $\{a(m)\}$ can be calculated as

$$y(m) = w(m) * x(m) + b(m)$$

$$a(m) = t(m) - y(m)$$

- The weights and biases of the network are adapted as

$$w(m)_{new} = w(m) + l \{a(m) * x(m)\}$$

$$b(m)_{new} = b(m) + l * \{a(m)\}$$

All parameters are set experimentally for proper adaptation of the network. The weights and biases are adjusted to attain the desired signal. The enhanced signal is achieved as the error signal by the ADALINE.

III. SPEECH ENHANCEMENT USING DEEP NEURAL NETWORKS

DNN is based on the supervised learning and determines the mapping from the noisy features to clean features. The structure of the DNN is presented in Fig.2 and is divided into 2 phases: training phase and testing phase. The hidden layers employ as the activation function.

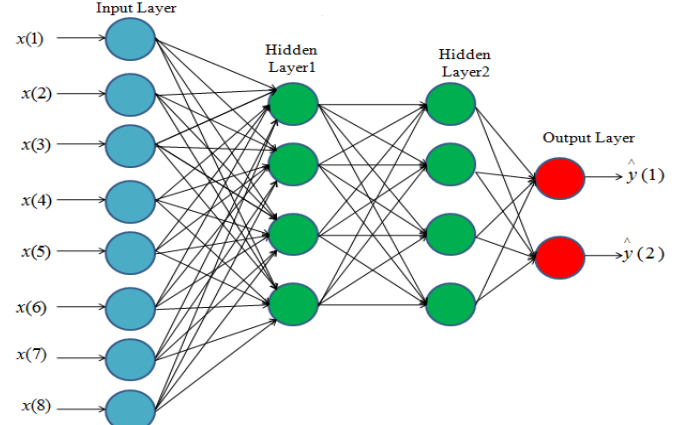


Fig.2. Structure of Deep Neural Network

In this work, two hidden layers are considered and sigmoid function is considered as the activation function for the output. To learn the DNN of noisy log spectra, the multiple restricted Boltzmann machines (RBMs) are arranged [15]. The NOIZEOUS database is taken from the softcopy of Loizou. Babble Noise, Train Noise, Airport Noise and Restaurant Noise of SNR 0dB, 5dB, 10dB and 15dB are considered for training and Drilling Noise, Street Noise are considered for testing. Total 100 speech samples of noisy as well as clean features are acquired in the training phase. Two hidden layers are considered with 512 hidden units each and 8 output units. Total $512 * 2 = 1024$ hidden units are trained for noisy speech features.

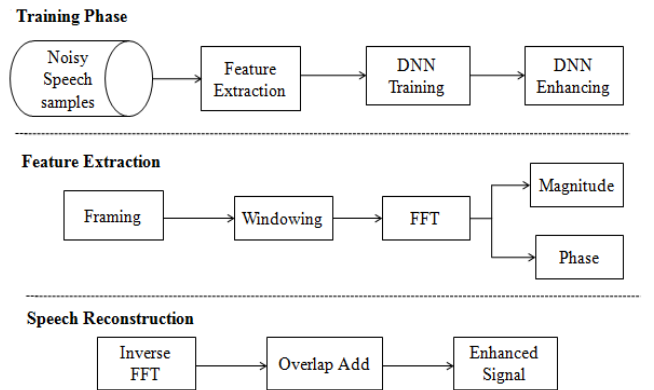


Fig.3. Speech Enhancement using Deep Neural Network

For speech enhancement, the noisy sentence is divided into overlapping frames. Hamming window of length 512 is multiplied to the framed signal to avoid signal distortion. The proposed DNN based speech enhancement method is represented in Fig.3. To extract the magnitude and phase spectra, Fourier Transform is employed. Only the magnitude

spectra are considered for training the noisy features in DNN and the phase spectra are ignored. After training, the Ideal mask is estimated for testing. The mask is enforced to the FT feature vectors of the noisy speech signal. The output of the DNN is interpreted as the predicted mask for the input. All the feature vectors are added and the overlapping frames are concatenated to reconstruct the speech signal. Subsequently, all frames are synthesized into a time domain signal by Overlap add. The following steps are followed for speech enhancement using DNN.

- Process all speech signals as overlapped frames.
- The frame length is 512 with an overlap of 40%.
- Calculate FFT of each and every overlapped windowed frame.
- Four different types of noise signals: Babble, Train, Airport and Restaurant are taken with SNRs 0dB, 5dB, 10dB and 15 dB and the clean signals are considered for training.
- To train the DNN, 156 dimensions of magnitude spectrum of the noisy signals are employed.
- For testing, two noise signals: Drilling and Street are considered to estimate the mismatch condition.
- Inverse FFT and Overlap Add method are implemented to reconstruct the speech signal.

Network pretraining, regularization are employed to make the system better.

IV. EXPERIMENTAL RESULTS

In this work, 10 sentences from both male and female speakers are considered for training. A total of 100 sentences have been collected with different types of noise signals as mentioned earlier. All utterances are sampled with a sampling rate of 8 KHz. Fig.4 shows the spectrogram of the clean speech signal. Fig.5 is the noisy signal (Babble noise of SNR 10dB). The clean signal is considered as the target signal. The noisy signal is applied to the ADALINE and DNN. The enhanced signal of the ADALINE obtained as the error signal as shown in Fig.6.

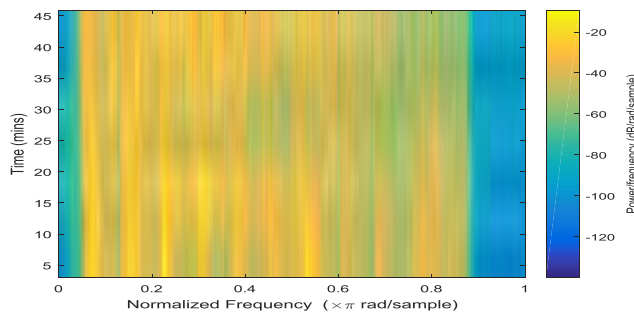


Fig.4. Clean Signal collected from database

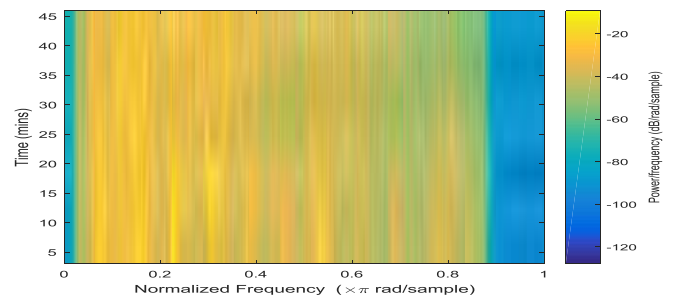


Fig.5. Noisy Signal (Babble Noise of SNR 5dB)

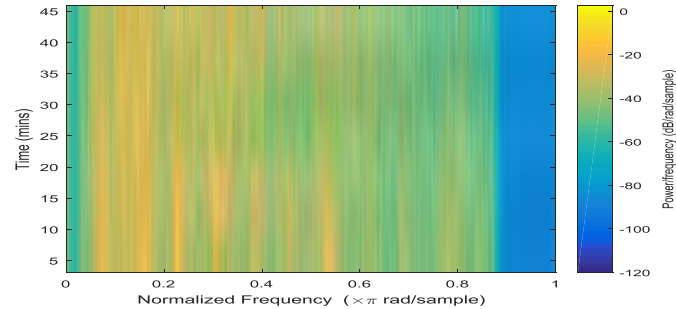


Fig.6. Enhanced signal of ADALINE

All 1024 utterances are considered to build the training set. One sample of clean signal and all the noisy data sets are employed to train the DNN model. A total 2400 frames are measured to train the DNN model. Another 200 arbitrarily selected clean and noisy utterances from the database are considered for testing phase for each combination of noise levels. These signals are estimated and evaluated for mismatch conditions of testing phase. Two hidden layers are considered with 36 frames expansion. Total 1024 hidden units are there in each hidden layer. For pretraining the DNN, the learning rate is 0.05 and for fine tuning, the learning rate is 0.001. 80 epochs are considered for mini batch size of 100. The enhancement using DNN_1 (DNN_* represents the hidden layer number) and DNN_2 are shown in Fig.7 and Fig.8 respectively.

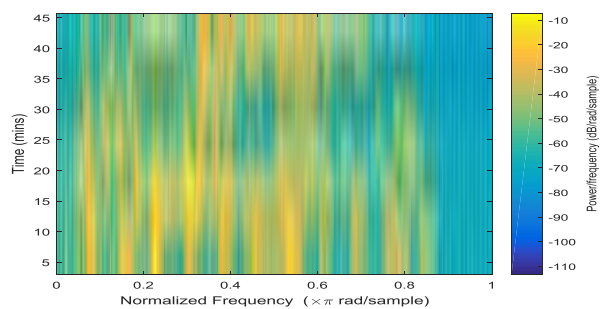


Fig.7. Enhanced signal using DNN_1

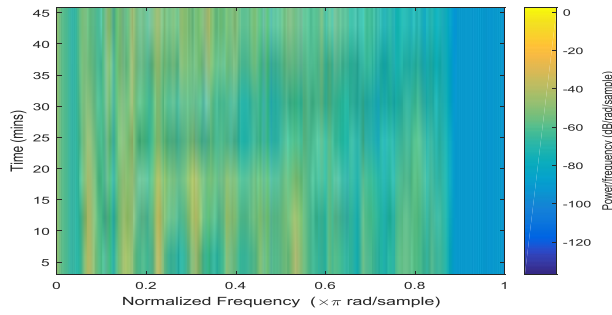


Fig.8. Enhanced signal using DNN_2

In this work, Perceptual Evaluation of Speech Quality (PESQ) and Signal-to-Noise-Ratio (SNR) are measured and evaluated for the quality of the speech signal. The listening test is also performed by different persons to verify the test results. Table I shows the SNR as well as the improvement of SNR of ADALINE and DNN. The maximum SNR improvement is 2.87 dB achieved in DNN_2. Table II shows the PESQ measures of the different noise levels. DNN_2 provides a maximum PESQ of 3.67 for 15dB of Babble noise. When the number of hidden layer increases, the better enhanced signal is obtained in the DNN.

TABLE I

SNR IMPROVEMENT OF BABBLE NOISE WITH DIFFERENT NOISE LEVELS OF DIFFERENT METHODS

SNR before Enhancement (dB)		0 dB	5 dB	10 dB	15 dB
SNR after Enhancement (dB)	ADALINE	1.32	6.36	11.12	15.92
Improvement in SNR (dB)		1.32	6.36	11.12	15.92
SNR after Enhancement (dB)	DNN_1	2.43	7.02	11.98	16.26
Improvement in SNR (dB)		2.43	2.02	1.98	1.26
SNR after Enhancement (dB)	DNN_2	2.87	7.83	12.55	17.31
Improvement in SNR (dB)		2.83	2.87	2.55	2.31

TABLE III

PESQ SCORE OF BABBLE NOISE WITH DIFFERENT NOISE LEVELS OF DIFFERENT METHODS

	ADALINE	DNN_1	DNN_2
0 dB	1.23	2.34	3.11
5 dB	1.45	2.06	3.46
10 dB	1.89	2.51	2.98
15 dB	1.93	1.78	3.67

V. CONCLUSION

ADALINE and DNN are used to enhance the noisy speech signal in this work. ADALINE is considered as the basic Neural Network implemented for speech enhancement. The DNN is used for different hidden layers that can prove the validity of speech enhancement in the field of data mining. The better performance result is obtained using ADALINE model, whereas the DNN model outperforms the ADALINE. Though the time consumption is more in DNN, speech enhancement is better. In the future, the weights of the ADALINE model can be varied and other transforms can be applied to extract the features and observe the performance.

REFERENCES

- [1] P. Loizou, Speech Enhancement: Theory and Practice. CRC Press, 2007.
- [2] S.Haykin, Adaptive Filter Theory, Prentice Hall, Upper Saddle River, 3rd Edition, 1996.
- [3] R.Ram, M.N.Mohanty, Performance Analysis of Adaptive Algorithms for Speech Enhancement Applications, Indian Journal of Science and Technology 9(44), 2016.
- [4] S.Vihari, A.S.Murthy, P.Soni, D.C.Naik, Comparison of Speech Enhancement Algorithms, Procedia Computer Science 89, pp. 666 – 676, 2016.
- [5] L.B.Fah, A.Hussain & S.A.Samad, Speech Enhancement by Noise Cancellation Using Neural Network, IEEE Conf., 2000.
- [6] R.Ram, M.N.Mohanty, Fractional DCT ADALINE Method for Speech Enhancement, Int. Conf. on Machine Learning & Computational Intelligence, 2017. (Accepted)
- [7] A.Prieto, B.Prieto, E.M.Ortigosa, E.Ros, F. Pelayo, J.Ortega, I. Rojas, Neural Networks: An Overview of Early Research, Current Frameworks and New Challenges, Neurocomputing 214, pp.242–268, 2016.
- [8] T. Kounovsky and J. Malek, Single Channel Speech Enhancement Using Convolutional Neural Network, IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM),2017, pp.1-5.
- [9] M.Kolbaek, Z.H.Tan, J.Jensen, Speech Intelligibility Potential of General and Specialized Deep Neural Network Based Speech Enhancement Systems, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25(1), 2017.
- [10] Y.Xu, J.Du, L.R.Dai, and C.H.Lee, An Experimental Study on Speech Enhancement Based on Deep Neural Networks, IEEE Signal Processing Letters 21(1), 2014.
- [11] Y.Xu, J. Du, L.R.Dai, C.H.Lee, A Regression Approach to Speech Enhancement Based on Deep Neural Networks, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23(1), 2015.
- [12] R.Li, Y.Liu, Y.Shi, L.Dong, W.Cui, ILMSAF based Speech Enhancement with DNN and Noise Classification, Speech Communication 85, pp.53–70, 2016.
- [13] Y.Li, S.Kang, Deep Neural Network-Based Linear Predictive Parameter Estimations for Speech Enhancement, IET Signal Process.11 (4), pp.469-476, 2017.
- [14] T.Goehring, F.Bolner, J.Monaghan, B.Dijk, A.Zarowski, S.Bleack, Speech Enhancement Based on Neural Networks Improves Speech Intelligibility in Noise for Cochlear Implant Users, Hearing Research 344, pp.183-194, 2017.
- [15] Y.Koizumi, K.Niwa, Y.Hioka, K.Kobayashi and Y.Haneda, DNN-Based Source Enhancement Self-Optimized By

- Reinforcement Learning Using Sound Quality Measurements, IEEE Conf., 2017.
- [16] R. Ram, M. N.Mohanty, Deep Neural Network based Speech Enhancement. Int. Conf. On Cognitive Informatics & Soft Computing, 2017. (Accepted)
- [17] J. C. Hou, S. S. Wang, Y. H. Lai, J. C. Lin, Y Tsao, H. W. Chang, H. M. Wang, Audio-Visual Speech Enhancement using Deep Neural Networks, Signal and Information Processing Association Annual Summit and Conference (APSIPA), Asia-Pacific, IEEE, 2016.

A Survey on Advanced Approaches of EHR in inter-related data using Machine Learning

T.V.M. Sairam, R. Rajalakshmi

School of Computing Science and Engineering,
Vellore Institute of Technology, Chennai, India
venkata.manikanta2016@vitstudent.ac.in, rajalakshmi.r@vit.ac.in

Abstract—Medical data is being used for huge number of research works over the globe which is for predicting something novel case studies in each work. The current research which we are handling is on utilizing the EHR (Electronic health Records) data in an efficient way based on the cause – effect ratio and the variables available for the data manipulation, processing and generating efficient data for designing efficient prediction models. In this research we are focusing on the congenital tethered cord syndrome through which some many functional outcomes issues are recording in different cases and there is a wide range of scope for research. In this research we are identifying the data from different EHR applications and designing the architecture to gather valuable data set from those for designing prediction model for predicting functional outcomes of health and life in patients with congenital deformity. Through EHR applications we gather information and BigData is being created in this sector. Data inter –relation is explained in this survey article in an efficient way with respect to medical domain. EHR data will be hosted over the cloud and in public repositories. Will focus on those categories in an efficient manner.

Index Terms—Machine Learning, Bigdata, Cloud Computing, EHR, Inter-related disease, Predictions

I. INTRODUCTION

HEALTH information system often called as electronic health records are ain focus of IT industry to develop their market in health care domain as there are several researches are going on in different places on medical concepts like Parkinson’s Disease using Machine learning [1], which deals with the combination of the genetic and environmental disorders which plays a major role in this kind of diseases. Here the main question is how the way EHR is related to this kind of disease treatment and how we can do the research in this kind of theories which have clueless information because there is no major readings for findings are there for the identification of Parkinson’s disorder. There is a case for adoption of Health records for the multiple theoretical perspectives of the operations in different health operations and in different cases in health operations [2]. The main criteria here is to utilization of EHR data for identifying a cause or

an effect with related to the previous details of patient like treatment, medication, lab, scans etc. Every single information like job details, environment habits etc will be considered as the variables for the model design and implementation. EHR is the best practice f gathering data from the health pracctioners from different organizations. Consider a scalable mHealth application which a well known EHR which tracks all your health information and store it as sheet and we can use that as medical report and which will be directly connected to the doctors. [3]. Prediction models can be designed with this EHR data like using K-Means clustering, random forest, Decision Trees etc which will help to identify what are the variables to be considered from EHR and how we can change the architecture of EHR application[4]. Multiple Linear Regression model consists of the concept of backward elimination using which we can decide which variable of the field can be utilized for designing the model and which can be excluded. As we reduce the complexity in EHR data the best output can be achieved in the model. The variations in the functional outcome of life [5] is required the best outfit of the variables which can be gathered from EHR and apply the tools like regressions to redesign the EHR architecture and the data. The main purpose of this article is to satisfy the theory of identification of functional outcome of life in the patient with tethered cord syndrome for that case we need adaptive EHR model which will intake efficient variables and form data sets for further processing in analytics. The later sections in this paper discusses about few EHR applications in real time and how we used to get the data from those section 3 explains the requirement of cause – effect ratio and data inter-relation in section 4 we conclude this survey

II. DIFFERENT EHR POSSIBILITIES

2.1. Mobile EHR

This is a small mobile EHR application architecture which is used to guide the person or the patient

according to his/her habits, age, weight, gender. Etc. This one is very basic architecture in EHR. Application is already programmed with all the required things and some group of combinations. Based on the input from the patient or the person it will calculate all the things and will give the result.

List of the things which that application will take are as follows. Age, Gender, Habits, Type of Job (Because of some diseases are because of the nature of work and the work location they are working etc). Algorithm will be designed and internally all the values from the patient will be taken as inputs and decode them by correlating them internally and will give a sample result. For example, it may ask to drink some amount of water frequently for time to time and ask you take some food after some time. For a short period of interval, it will give some notifications. There are some advantages as well as big disadvantage with this kind of EHR architecture. In this section, we will discuss with the architecture of this EHR system and in later section we can look forward for pros and cons.

In figure 1 we have the basic architecture of EHR application which is a mobile application using which all the patients will get the regular notifications based on the information they gave to the application.

Figure 2 describes the flow of the EHR method to take information, update in the application and do some predictions models inside and show the notifications to the user through application time to time.

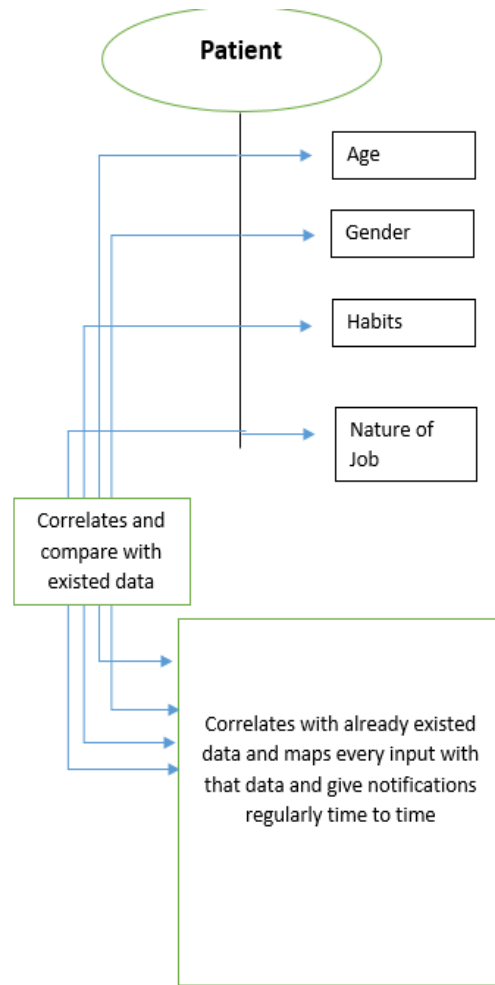


Fig 1: Mobile EHR Architecture

This comes under the unsupervised learning. If we consider machine learning in EHR it will be much useful for the users.

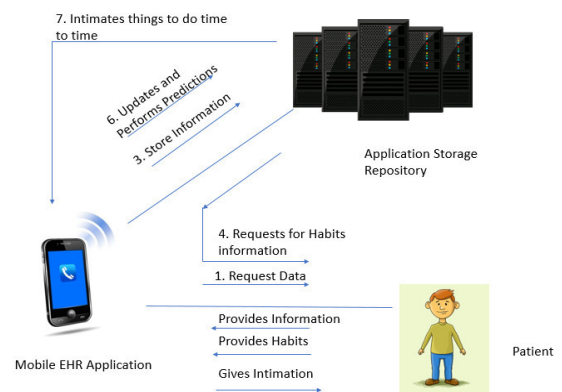


Fig 2: Data flow in Mobile EHR

2.2. Community Health care monitoring system

Community Health care providers (C- HCP) are health care management system organizers which will have some man power who will come to our presence that means to our home to take all the required information about our health like blood sugar, heartbeat, oxygen levels etc. If we consider this community health care system, they have an application using which our health information can be updated and saved in global server for the purpose for further research for scholars. We can update or enter our information from any mobile application which is having the community health care management system application. Using the sensors in our mobile, all the medical information will be updates and synced simultaneously. The common feature of these kind of applications is measuring oxygen saturation using sensors. Consider Community Health care providers in the case of pregnant ladies. We need to track all the information regarding that lady, start from the BP, Sugar, Oxygen saturation level and remaining which are most important. C-HCP will come to our premises and take all the information. In low and middle income countries, we find most of the cases that pregnant ladies are losing the pregnancy because of the lack of proper health care management and treatment and also because of the late response for treatment. Economically backward families can't effort for high paid treatments, this is also one of the reason for cases of pre-eclampsia and also eclampsia period problems. In that period doctors need to check the patient continuously physically or using the mobile application.[21] Based on the information provided and stored in the server doctors need to take all the precautions to solve the problem in prior.

There are few more things to be learn about pre-eclampsia period ladies and their conditions. This application consists of previously successful features which are no way related to each other. 1) Predicting the situation of the patient based on identifying the risk factor mini pre-eclampsia integrated estimate risk score (miniPIERS)[20] 2) A mobile application in phone oximeter and a sensor for pulse detection [19]. These applications predict the risks in the pregnant ladies and suggest medication and treatment for those kind of ladies in prior. The process of predicting the risk factor based on clinical validations, blood oxygen saturation

(SpO2) and also based on other common symptoms will occur in that pregnancy period. PIERS on Move (POTM) is a low cost application for risk identification in the case of pregnant ladies. Community based implementations of cluster based randomized control of pre - eclampsia trail was discussed[22] [4]. There are several updates are given to the original POTM application. The basic POTM application was designed to train nurse in South Africa and also some house wives who are economically backward to get treatment. So that at the time of first update for POTM nearly 202 women are assessed with the POTM application. The purpose of that trail version to know whether the POTM is operating as per the requirement or not and it is satisfying the requirements or not. So that late for the same study POTM is used by nearly 500 women with the same purpose to know the low or high risk in the pregnant ladies in pre-eclampsia and eclampsia period.

The other kind of POTM is CLIP trail. This application for community based health care providers (C-HCP's) who are been to the women who are pregnant and test them physically or using the phone oximeter and perform some of the operations related to the pregnancy. They tested this on 500 women by approaching them personally. At each stage if their research practitioners have to give their feedback regarding their study using this CLIP trail POTM application. Prototypes are designed for the feedback for the practitioners and also to perform some of the analysis on the women or on the group of women who are pregnant. Figure 3 describes the prototype for POTM.

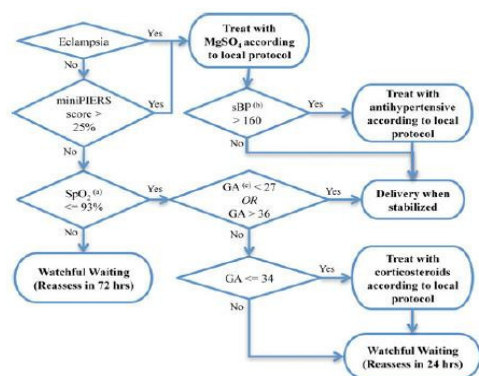


Fig 3: Diagram to know whether pregnant lady is in labor or not using POTM application.

Device tabs are recorded by device and stored in the internal memory of the device. Whenever we need those records, we can copy those log records and store in USB device or in any external storage devices. There are three categories of these device log tabs. 1) Alarm Events used to store information regarding the emergency intimation details that means when patient is suffering from any problem it will give an alarm to the doctor. So, that information will be given stored in the log files. 2) System events. That means information recorded by the device time to time. So that in log tabs time will be mentioned and at what time readings are taken and transferred to the doctor.

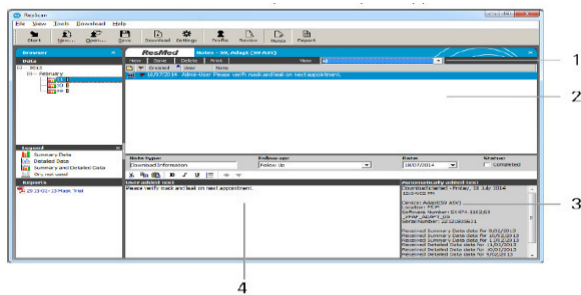


Fig 8: Device Notes Screen

In notes screen we can find the downloads done by the patient or doctor regarding the patient data. In this doctor, can store the treatment given to that patient and can review at any time he wants. Patients can create their own notes and device will create a notes screen automatically for user. In this above screen 1) Notes Toolbar 2) Note Summary 3) Automatically added text 4) Used added text.

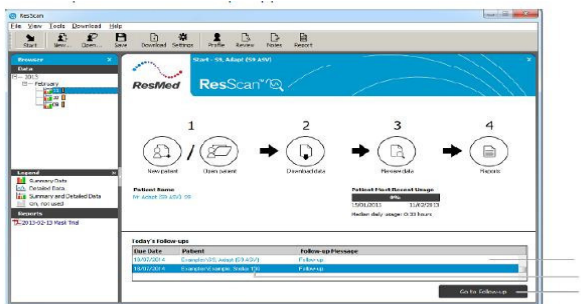


Fig 9: Follow Up page

In this follow up screen all the dues and over dues will be displayed. In the home screen go to follow up button is displayed we can directly go to that screen.

In this above screen 1) Follow ups display 2) Double click to follow up screen 3) Click button to go to follow up.

2.4. Some common EHR models.

The basic concept of EHR is common in all areas. Some of the common things are as follows in this block.

- 1) EHR in emergency department of hospital. In causality of hospital patients may join with some chronic disease or problems. So, that using this EHR concept all the patient treatment details will be recorded and start from join date everything will be stored like operation details, medication, physiotherapy etc. and all the review details, if there are any consultants from other hospital doctors etc are stored here. Based on the patient condition to give better treatment we can create graphs for better understanding of the patient condition like ResScan in previous concept.[7,8,9,10]

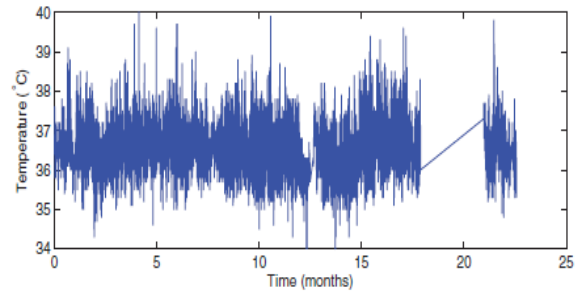


Fig 10: Time Series of Temperatures.

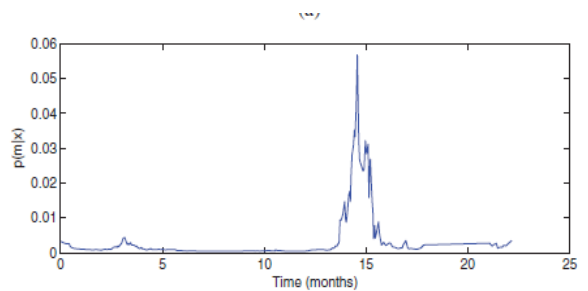


Fig 11: Bayesian Change point detector

Early warning scores are explained in[11,12,13] using which we can give scoring for the patient passed on the readings using the EHR. Based on this large dataset are maintained and monitored and graphs are designed for further process.[14,15,16,17]

- 2) Patient centred EHR for hospitals using which we can access our data from anywhere in this universe and we can take suggestions from any doctor by showing him the previous history of his treatment.

Health information Technology for economic and clinical health act (HITECH) is the base for this patient centred application of EHR. Table 1 adopts the list of physician’s approach regarding EHR.

Table 1: Physician’s adaptation of EHR

Category	Description	Occurrences
Work	The physician perspective of EHR usage on physician work. Subcategories: Work Impact, Productivity, Integrated Collaboration, Access/Connectivity, Requested Enhancements, EHR Here to Stay.	102
Technological	The physician perspective on implications of IT Context on EHR usage. Sub-categories: Systems Development, Hardware & Configuration, Physician Communication & Collaboration.	70
Social	The physician perspective on implications of social context on EHR usage. Sub-categories: Standard Templates, Processes & Rules, Data, Interfaces & Presentation, Knowledge & Learning.	34
Total		206

Figure 12 describes the workflow of the physicians approach for the patient centred care.

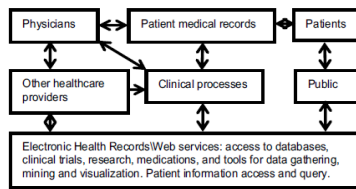


Fig12: Model for physical collaboration of patient centred care.

Many approaches are there for EHR application but there are lot of backlogs in every architecture like not able to predict that the data obtained in mHealth application is genuine or not, prediction algorithm is not using for maintaining consistency in patient self immunity capacity. Because of using anti - biotics heavily patient may lose his/her self healing capacity. A new prediction model and algorithm is introducing in this paper with a basic architecture with an example of work flow of the algorithm.

3. Discussion

Data Inter relation is the background architecture of EHR applications [6] which deals with identification of the common points between the variables and we can replace the entire set of the common variables in the data set with less P value than SL value. We can identify the variable with less P value using

backward elimination process in the regression model. Hypothesis is drawn and we will use the entire set of EHR of data as dataset and will reduce each variable which has the highest P value than SL and finally we will get the appropriate variable which suits best fit the model. Cause – Effect ratio is other thing in Machine Learning models details which predicts the effect based on the ratio of the cause set and identifies the most valuable set of causes in the available set of variables causes set. The model have to predict the cause have highest chances of causing some specific effect. There is a sample architecture must be followed with respect to the EHR data Separation and identifying the most valuable variable must be best fit in the model.

4. Conclusion

This survey paper demonstrates the different EHR applications and the working purpose of those in their respective areas like cardiac like ResScan, mHealth applications and the main reason for better utilization of EHR data is to provide valuable health services to the people who are economically backward. In further to this research we are focusing on designing prediction model with data mining and machine learning techniques to predict outcomes of life for the patients with congenital diseases. As we know diseases are divided into congenital and genetic, tethered cord syndrome is focused main the task to gather , clean and design related models with related to expected result. Further research carries all the benefit of utilization of the EHR models discussed in this article in an efficient manner.

REFERENCES

- [1] “An improved approach for prediction of Parkinson’s Disease using Machine Learning Techniques”, Kamal Narayan Reddy Challa*,2016, IEEE
- [2] Adoption of Electronic Health Record System: Multiple Theoretical Perspective”, Qiwei Gan, Qing Cao – 2014 IEEE
- [3] “A Scalable mHealth System for Noncommunicable Disease Management”, G D Clifford* - 2014 IEEE
- [4] “Predictive Medication and use of BigData”, Avijit Goswami – 2017 IEEE
- [5] “Variation in Outcome in Tethered Cord Syndrome”, Norulain Iqbal*, 2016, Asian Spine Journal
- [6] “Resource Frequency Prediction in Healthcare: Machine Learning Approach” Daniel Vieira, 2016 IEEE
- [7] National Patient Safety Association, “Safer care for acutely ill patients: Learning from serious accidents,” Tech. Rep., 2007.
- [8] National Institute for Clinical Excellence, “Recognition of and response to acute illness in adults in hospital,” Tech. Rep., 2007.

- [9] H. Gao, A. McDonnell, D. Harrison, S. Adam, K. Daly, L. Esmonde, D. Goldhill, G. Parry, A. Rashidian, C. Subbe, and S. Harvey, "Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward," *Intensive Care Med.*, vol. 33, no. 4, pp. 667–679, 2007.
- [10] L. Tarassenko, D. Clifton, M. Pinsky, M. Hravnak, J. Woods, and P. Watkinson, "Centile-based early warning scores derived from statistical distributions of vital signs," *Resuscitation*, vol. 82, no. 8, pp. 1013–1018, 2011.
- [11] D. Prytherch, G. Smith, P. Schmidt, P. Featherstone, K. Stewart, D. Knight, and B. Higgins, "Calculating early warning scores—A classroom comparison of pen and paper and hand-held computer methods," *Resuscitation*, vol. 70, pp. 173–178, 2006.
- [12] A. Hann, "Multi-parameter monitoring for early warning of patient deterioration," Ph.D. dissertation, Univ. Oxford, Oxford, U.K., 2008.
- [13] D. Wong, I. Strachan, and L. Tarassenko, "Visualisation of highdimensional data for very large data sets," presented at the Workshop Mach. Learn. Healthcare Appl., Helsinki, Finland, 2008.
- [14] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [15] S. Hugueny, D. Clifton, and L. Tarassenko, "Probabilistic patient monitoring with multivariate, multimodal extreme value theory," *Commun. Comput. Sci.*, vol. 127, pp. 199–211, 2011.
- [16] R. Kavitha, E. Kannan, S. Kotteswaran, "Implementation of Cloud based Electronic Health Record (EHR) for Indian Healthcare Needs", *Indian Journal of Science and Technology*, 2016 Jan, 9(3), Doi :10.17485/ijst/2016/v9i3/86391
- [17] Meenakshi Sharma, Himanshu Aggarwal, "EHR Adoption in India: Potential and the Challenges", *Indian Journal of Science and Technology*, 2016 Sep, 9(34), Doi:10.17485/ijst/2016/v9i34/100211
- [18] ResScan Software : ResScan version 4.2 Clinical Guide from "ResMed Ltd 1 Elizabeth Macarthur Drive Bella Vista NSW 2153 Australia"
- [19] R. Lozano, and C. J. L. Murray, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2012.
- [20] B. A. Payne, J. A. Hutcheon, "A risk prediction model for the assessment and triage of women with hypertensive disorders of pregnancy in low-resourced settings: The miniPIERS (Pre-eclampsia Integrated Estimate of RiSk) multi-country prospective cohort study," *PLoS Med.*, vol. 11, no. 1, e1001589, pp. 1–13, 2013.
- [21] W. Karlen, G. A. Dumont, C. Petersen, J. Gow, J. Lim, J. Sleiman, and J. M. Ansermino, "Human-centered phone oximeter interface design for the operating room," in *Proc. Inter. Conf. Health form.*, Rome, Italy, 2011, pp. 433–438.
- [22] University of British Columbia. (2014) Community level interventions for pre-eclampsia (CLIP). [Online]. Available: <http://clinicaltrials.gov/ct2/show/NCT01911494>. NLM Identifier: NCT01911494

Result Analysis Suite: A Completely Automated Result Analysis Solution

Pankaj Sambyal¹, Anamika Rustagi², Sonia Rani³,
Chinab Bhudhiraja⁴, Bhawna Sharma⁵

Department of Computer Science, Kalindi College, University of Delhi, India
¹pankaj.sambyal@kalindi.du.ac.in, ²anamikarustagi@gmail.com, ³11soniarani@gmail.com,
⁴chinabudhiraja@gmail.com, ⁵bs152763@gmail.com

Abstract—“What gets measured, gets managed”. The present paper intends to describe a one stop solution to analyze the student result data collected from the University Website automatically using the web automation technology, organize it into an efficient fashion and make it available for all stakeholders in a beautiful and intuitive way, therefore, reducing the manual task to a bare minimum. The whole system is divided into three phases - Data Collection, Re-Organization and Presentation. An automation system has been developed for raw result data collection, re-organization of the collected data and applying efficient analysis algorithms on the data locally, and in the last phase, presenting the result information to the user in an intuitive manner so as to make it easy and quick to understand and interact. Keeping the user in mind, the last phase is the most crucial one. A set of highly efficient open-source graph rendering libraries has been employed to generate interactive result infographics which are portable and can be exported for external use.

Index Terms—Automation, Analysis, Infographics

I. INTRODUCTION

“**R**ESULT analysis” refers to the process in which the result is reviewed in order to determine conclusions based on that data. The purpose of initiating this research is to minimize the man-hours needed to analyze college result and create a flexible and easy-to-use solution for all students’ needs. As the time-taken by an individual to perform result analysis on each paper of a batch is about 4 hours (data from study), when scaled up to the whole college, these numbers can be ridiculously high. A versatile, expandable and economical solution is needed to perform a fundamental task such as result analysis making it easy for everyone in education community to take advantage of it. The result analysis has to be done at the end of each semester, so the Result Analysis Suite will help to reduce redundancy and keep the data organized so it becomes efficient and productive to evaluate the whole college’s result with minimal human effort. It will eradicate the introduction of human error and increase accuracy and efficiency.

In the present paper, we try to summarize the application of Result Analysis Project, its different levels and the various types of infographics are also discussed like bar/column graphs, pie-charts, dynamic charts, etc.

II. STUDY

In the current system the teachers are supposed to go to the university website and enter the details of each student manually. These details include college, course, semester, student’s name and their University roll no. If on an average there are 50 students in a batch and a teacher teaches at least 2-3 subject papers in a semester, piling up this whole information, if calculated, the time consumed for analyzing the result of a batch(~50 students) is given in Table 1.

TASK	TIME
Download each student’s mark sheet	~1 Hour
Read the result of each student from university website	~1 Hour
Organize the collected data in spreadsheet format	~1 Hour
Perform Analysis on the collected data	~15 Minutes
Total Time	~3 Hours 15 Minutes

Table 1: Time analysis of a single batch of 50 students with 2-3 papers (Data from study)

When this work is done by the teachers there is no way to compare the whole department’s performance by TICs, it has to be done manually which is prone to human errors in calculation and will consume a lot of time.

III. USER INTERFACE

There are five levels for the user and system interaction-Student, Teacher, Teacher-In-Charge, Principal and Administration.

At **Student level**, the analysis of all the subjects studied so far by the user for a particular semester is provided. Carrying off to **Teacher level**, this level describes the result data to be analyzed for a teacher teaching one or many subject papers. Inheriting all the access rights of student and teacher

Table 1: Comparison of SEO and SEM

	SEO	SEM
Features	Organic SEO and Paid SEO	Integration of SEO with SMO
Focus	Keywords with high popularity	Keywords that generate high Return on Investment (ROI)
Measure	Number of website visitors	Number of successfully converted goals.
Landing pages	SEO pages are content-heavy and informative, designed to use keywords and subject relevance to rank well on any search engine.	It generally leads to direct, clean landing pages with an obvious call to action and designed to capture sales.
Prediction	It is not predictable due to uncertainty of traffic.	It is predictable as putting enough money will produce high ROI
Skill required	Fluency in SEO tools.	Require more skills than SEO.

level **Teacher-In-Charge level** is defined where they have excess to all the data (i.e. student's and teacher's personal details as well as their academic performance) regarding their department. The **Principal level** is provided with the combined data of the above three levels and deals with the performance of the students and departments.

The patriarch of all the levels comes at the **Administration level** having access to the whole system, in other words, its data and features.

IV. DATA COLLECTION

In the existing system, the user who wants to access the details of the student needs to fill the form on University portal. This form include selecting the college name, course, semester, type of exam and entering the details like Student's name and their university roll number. If the user is a teacher then he/she needs to input his/her 3 hours approximately for the data collection only.

In the proposed system the whole result data can be collected in a database using the Web Automation technology, in which it will access the details of students from the data available on website under the student details section and automatically filling up the forms as well as downloading them for further use. This is done in less than two minutes by automation.

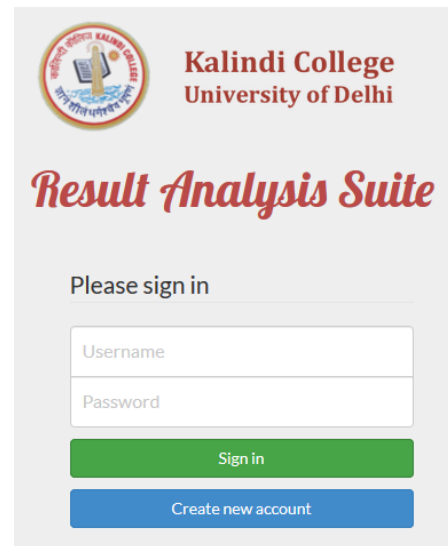


Figure 1: The Login Screen of the Front end UI

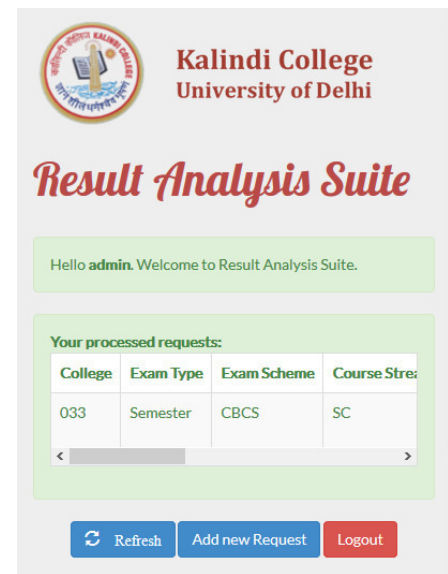


Figure 2: Second level UI for the user to check on the processing of results.

The data can be manipulated any way when it is available locally. The front-end UI along with the background Java processing engine will ensure the proper and accurate information delivery to the end user.

V. RE-ORGANISATION

The data used for the analysis work is organized in databases like Student, Teacher and Paper. The original data collected from university server contains student detail and their marks statement. From this information, marks are extracted and are stored in Paper database so that analysis operations like average, count, comparisons are easily implemented.

These databases are linked with each other by creating foreign key references and mapping the subjects with their concerned teachers, student with their subjects and so on. Integrating data from the site and organizing it into databases, enables a cumulative view of the performances of student in different subjects and helps the user at TIC and Principal level to interpret the records of their concerned departments and college respectively.

VI. ANALYSIS

Analysis is an art of creation through destruction and in this process inspection, cleansing, modeling and transformation of data is performed with the goal of discovering useful information, suggesting conclusions, and support decision-making.

In our practice of result analysis the first step depends upon the access writes of the user as level there are different levels of it. These levels include student-wise, teacher-wise, subject-wise, department-wise and overall college evaluation. If we go into further details student-wise evaluation includes his/her marks in different subjects, comparison of the performance with previous semester, his/her cumulative result, ranking in class, etc. talking about teacher analysis his subject aggregate grade/marks are calculated, pass/fail percentage of subject is calculated. For The TICs level performance of all the students and teachers of that particular department is provided and compared with each other to find out toppers in student and efficiency of teacher. The Principal and administration level user are provided with analysis features of the whole college.

VII. PRESENTATION

If you are the only one who “**gets it**”, then you have failed. So for the better understanding of the analysis Info-graphics is used. We can display the data in a tabular form but displaying it in graphical form is more user-friendly and interesting at the same time. It makes the information simple, easily shareable, contains minimal text and have high impact on user. A simple illustration can be seen in Figure 3, where a set of test data is taken and analyzed to find the number of students in each group i.e. Grade A+, Grade A, Grade B and so on. In this example, three subject papers are taken into consideration and the user can select to see the done analysis on one or all of the papers in the generated graph. The paper and student details are available to the user if they need it.

In the second test sample (Figure 4), we can see the grade A+, A, B+, B and C students of Computer Science and Physical Science departments.

Keeping all this in mind, different types of charts like bar-graphs, pie-charts, dynamic charts are used for the representation of data. Depending on the security level of the user they provide graphical representation of their records where they can select the type of graphs and also the points of comparison. The students are allowed to access their records only, teachers can interpret their respective subject only,

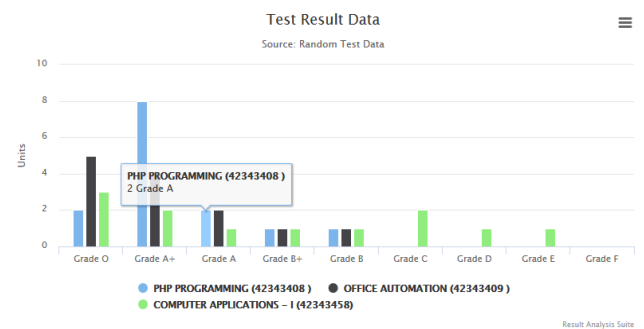


Figure 3: A test result data analysis is rendered as a bar graph.

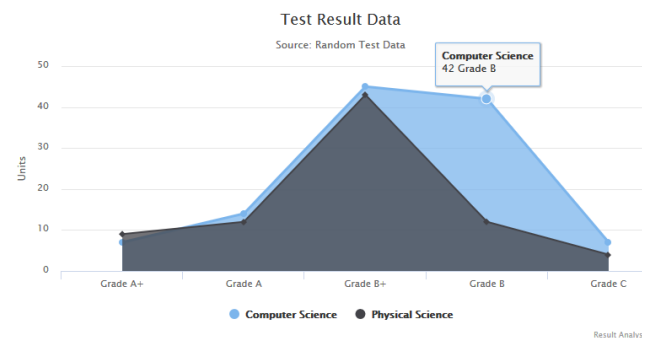


Figure 4: A test result data analysis of two departments rendered as an area map

whereas TICs can compare the output of teachers and the performance of students with each other using graphs and charts of their conveniences. Principal have access to all the graphs facilities provided to level below it. Not only comparison and analysis is done, user can also export these graphs and charts as a portable image to share or add to a comprehensive result report.

VIII. CONCLUSION

The Result Analysis Suite is a highly scalable solution for better understanding of performance of the student and reducing the time consumption during manual interpretation. It presents a possibility and execution of a fundamental task with **no human error** and **100% accuracy**. After thorough testing on college servers, it can be expanded to be used on a University level.

REFERENCES

- [1] Lidia Ogiela, Chapter 3—Intelligent Computer Data Analysis Techniques, Editor(s): Lidia Ogiela, In Intelligent Data-Centric Systems, In Cognitive Information Systems in Management Sciences, Academic Press, 2017, Pages 25-43, ISBN 9780128038031
- [2] Monitoring labor inputs: automated-data-collection model and enabling technologies Original Research Article Automation in Construction, Volume 12, Issue 2, March 2003, Pages 185-199 Ronie Navon, Eytan Goldschmidt
- [3] Leigh Metcalf, William Casey, Chapter 4 – Introduction to data analysis, Editor(s): Leigh Metcalf, William Casey, In Cybersecurity and Applied Mathematics, Syngress, 2016, Pages 43-65, ISBN 9780128044520
- [4] John Pirc, David DeSanto, Iain Davison, Will Gragido, 6—Data Visualization, Editor(s): John Pirc, David DeSanto, Iain Davison, Will Gragido, In Threat Forecasting, Syngress, 2016, Pages 95-103, ISBN 9780128000069

An Approach towards economical hierarchic Search over Encrypted Cloud

¹Sreenivas Sasubilli, ²Kumar Attangudi Perichiappan Perichappan,
³P. Srinivas Kumar, ⁴Abhishek Kumar

¹ *Solution consultant, KPMG US*

² *Associate Director KPMG US*

³ *Research Scholar, Department of Computer Science & Engineering,
Sri Satyasai University of Technology and Medical Sciences, Sehore, Madhya Pradesh*

⁴ *Assistant Professor, ACERC AJMER*

Abstract—In display, Cloud registering is the prevailing area in data innovation. With expanded value of information outsourcing of cloud information protection of delicate information turns into a major issue. For the security reason information is encoded before outsourcing. Yet, scrambled information is exceptionally hard to be recovered proficiently. Albeit some conventional scan plans are accessible for looking encoded information, yet these methods are just base on Boolean pursuit and not manage the importance of records. These methodologies experience the ill effects of two principle inadequacies. Right off the bat, on the off chance that one client has no pre-learning of scrambled information, needs to process each recovered record to discover after effects of his utilization. Also, every time recovering every one of the records containing question watchword builds arrange movement. This work is devoted to build up a process for security and compelling recovery of cloud information. Positioned seek enormously enhances the execution by restoring the documents in positioned arrange in light of some closeness importance criteria. To accomplish more viable execution, framework shows an approach for SSE which uses data recovery and cryptography primitives. Thus the execution depends OPSE..

Index Terms—Cloud; privacy; Similarity of Data; Re-Ranking

I. INTRODUCTION

DISTRIBUTED computational can be accepted as symbolic of conveying data innovation administrations (like storage room, organizing, applications and so forth) in which assets are recovered from web utilizing online instruments, as opposed to an immediate association with server. Distributed computing gives equipment and programming assets from a common pool of assets on lease as indicated by client's request. So this innovation discharges client from weights of administration endeavors and furthermore from cerebral pains of establishment and upkeep.

Service model

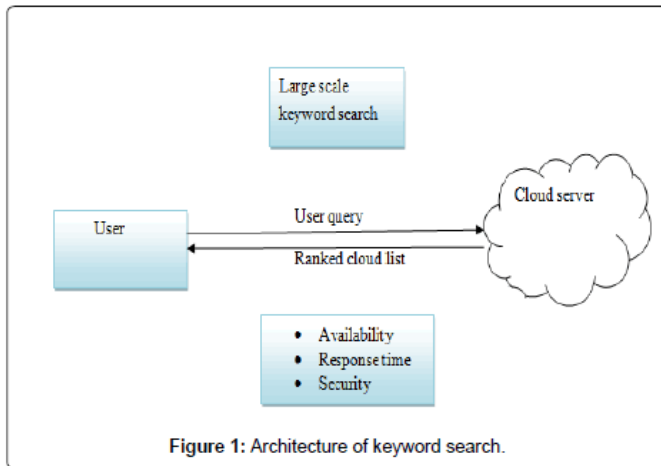
Cloud programming as an administration: Here the product will be accessible for client as a administration. Cloud apps are by and large available from different gadgets like portable, tablet, PC, workstations, servers and so on.

The client has no power over the hidden stage and framework. Cases are Dropbox, Gmail, Gtalk and so on.

Cloud stage as an administration (PaaS): In this product is made accessible to the client as administration. Programming dialects and devices are given by specialist organization to create and sending administrations. A client has no influence over basic framework yet must have control over the sent products. Cases are Azure, Google App Engine.

II. OVERVIEW OF DIGITAL WATERMARKING

Distributed computing has turned into a predominant stage in data innovation, the measure of delicate data concentrated over cloud is likewise expanding. These data records contain classified information like individual restorative records, government archives, private photographs and so on. To secure protection of information and to avert unapproved get to, it turns out to be extremely important to encode information before outsourcing to guarantee information trustworthiness and privacy. Alongside this, information proprietor may share their outsourced information with various clients. Yet, every client wants to recover records of his own enthusiasm amid a given day and age, which makes information use extremely difficult. From the current methodologies, most regular is utilizing catchphrase based pursuit strategy. These methods are generally connected for plaintext look situations where client can recover the documents of enthusiasm by giving watchword in inquiry. Unfortunately, information encryption for securing outsourced information influences these conventional techniques to end up coming up short to search cloud information. Albeit, some conventional encryption procedures encourage client to seek over scrambled information without first unscrambling it. Be that as it may, these strategies just help Boolean hunt, where documents are recovered by nearness or nonattendance of catchphrase in record and don't consider pertinenceofrecords(Figure1).



Many existing procedures for situated organize chase and importance of report are used by Information recuperation (IR) social order of looking for data available in the cloud. In spite of the way that the hugeness of situated look is tolerating thought from a long time, yet the purpose of encoded look for isn't watched out for much. Along these lines enabling an instrument for secure symmetric encryption and situated in this issue.

Before open inscription methodologies were recently supported rectify watchword look [1-6]. Tune et al. Given a SE model for balanced watchword look. During which every report within the record is encoded exploitation 2 stratified cryptography strategy [1]. Summary creation is employed by one or two of authorities for adequacy modification. In record based mostly methodology secure record is worked for every big word in an exceedingly report [2, 6]. Within the work planned by Curtmola et al., for every catch phrase record advancement, entries are finished to hash table. Each entry consists of document for extraordinary word and their offensive wrong archive the assaulter [5].

Furthermore, one or two of developers wandered towards placed interest to boost user friendly. Wang et al. [7, 8] expected placed look section in perspective of bound significance scores to acknowledge likeness of records with self-addressed catch phrase. This approach was single-watchword based mostly. Dynamic a part, multi-watchword look is explained by rule et al. Likewise, Cao et al. [9, 10]. They need used "resemblance based mostly within factor for result situating.

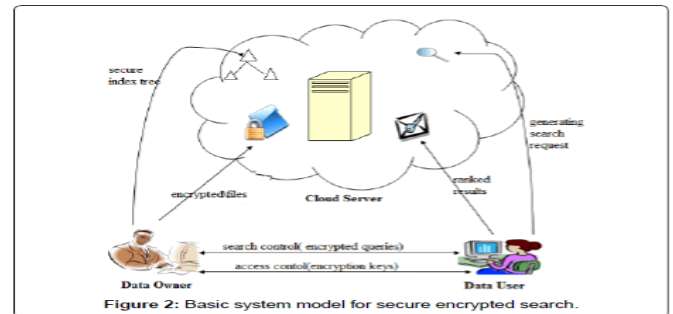
Nevertheless, every and each on top of arrangement are right watchword seek for based mostly. For enhancing request skilfulness, soft watchword based mostly chase is introduced by one or two of manufacturers [11-13]. Modify isolate thought is employed for locating equivalence of catchphrases with one another for delivering soft watchword sets for records. Li et al. Besides, Wang et al. [11, 14] conferred to switch isolate method. During this context AN execution is given for secure balanced interest cryptography and situating of ends up in a appropriate demand as incontestable by some significance criteria.

Available centrally cryptography

System Model-Basic model incorporates three kinds of substances that self-addressed within the Fig. The elements are termed as information owner (O), cloud server (CS) and client (U). A event of n information records $C=(F_1, F_2, \dots, F_n)$ is outsourced by information owner onto the cloud into encoded arrange. But for possible use of mixed records information owner makes secure record I employing a set m of differentiate watchwords given by $W=(w_1, w_1, \dots, w_m)$, that are expelled from archive aggregation C . Each mixed information moreover as records are outsourced on the server.

Exactly once the thoroughbred client has to recuperate one or two of knowledge record that has some watchword w , a secret seeks for trapdoor is formed by client to the server.

Exactly the server gets the will T_w , all of the estimation is completed on the server. Succeeding to visualize the summary server re-establishes the organizing archives to client. Endeavours take into account the safe placed catch phrase look issue as takes after: the factor ought to be came by bound placed significance criteria (e.g., watchword repeat), to upgrade archive healing accuracy for purchasers while not previous learning on the record aggregation



All the calculation works for beholding listed lists and growing importance worth of statistics as accomplished at server. Sooner or later the server has entrance to each one amongst the files are often positioned away. For safety reason server got to act in a very "truthful" manner and entirely takes when the meted out conventions.

It considers a "fair however inquisitive" server in our version. On the top of the day, the server doesn't have a goal to with success regulate the knowledge circulation / enlightened another reasonably administrations. List and linguistics relationship library (SRL) of various one amongst styles of words created is constructed} utilizing data made. At the purpose once shopper provides some question phrase, server grows it on the premise of SRL. Beyond seeking record, it restores the numerous documents to shopper.

III. PROBLEM FORMULATION AND PROPOSED SOLUTION

Plan objectives

To empower positioned accessible symmetrical for powerful use of outsourced cloud info beneath the antecedent mentioned show, venture configuration got to accomplish the attendant security and execution guarantee. Specifically, it's the attendant objectives:

- Stratified phrase appearance: to analysis exceptional gift units for secure approachable encoding and to assemble a tool for powerful set are searching for.
- Safety ensure: to making the outsourced statistics comfortable by exploitation keeping cloud server from gaining data of plaintext of records.
- attaining performance: on top of targets should be achieved with least correspondence and calculation overhead.

Documentation and preliminaries

- c – the aggregation to be source, import as Anassociation of n records $c=(f_1, f_2, \dots, f_n)$.
- W – the clear catchphrases free by record gathering C , indicated as an appointment of m words $W=(w_1, w_1, \dots, w_m)$.
- $Id(F_j)$ – the identity of record F_j that helps apparently notice a true document.
- I – the list worked by the document accumulation, that includes an appointment of posting records, as bestowed to a lower place.
- T_{wi} – the trapdoor created by a shopper as Aninquiry demand of watchword American state.
- $F(w_i)$ – the arrangement of identifiers of documents in C that contain phrase American state.
- N_i – the number of documents containing the phrase American state and
- $N_i=|F(w_i)|$.

Additionally venture presents some vital information recovery foundation for our planned framework:

Transformed file

Transformed file (likewise alluded as posting records) is usually used ordering arrange in information recovery. In rearranged record structure a 1 of sort file esteem is given for every phrase yet as summary of mapping is made by watchwords to documents within which word is obtainable. For empowering positioned ask for, a significance worth for documents is patterned utilizing few scientific suppositions.

Positioning capability

A positioning capability is employed to method likeness of terms by computing significance worth. For a given pursuit kindle, worth is made for coordinative documents that ar pertinent to questioned watchword. the foremost typically used factual estimation for assessing significance score within the information recovery cluster utilizes the $TF \times IDF$ govern, wherever TF is simply An amount of times a given watchword shows up within a record (to quantify the importance of the term within the precise document), and IDF (backwards report recurrence) is gained by uninflected the number of documents within the entire accumulation by the number of records having the term (to gauge the final significance of the term within the whole gathering).

Request protective balanced encoding

The opse may be a determined encoding conspires whereby the numeric requesting of the plaintext gets lined with the help of the encoding work. Boldyreva et al. [15] gave the quantity one cryptology analysis of opse archaic and executes a enclosed pursuit structure utilising pseudorandom ability and alternate. This work considers missive of invitation safeguarding capability $g(\bullet)$ from area $d=1, \dots, m$ to travel $r=1, \dots, n$, which may be apparently defined through a combination of m out of n requested matters. An opse is also declared comfortable simply if An offender has to play out a savage electricity are searching for over all of the manageable mixes of m out of n to interrupt the encoding conspire. On the off threat that the protection degree picked is of sixty four bits, at that time it regards recognize $m=n/2$; sixty four, maintaining in thoughts the quit intention to form type of blends with the aim that the mix vary of mixes are often further distinguished than 264. This improvement depends on affiliation between created safeguarding potential and hyper geometric chance flow into (hgd). Their improvement depends upon on a discerned affiliation between An irregular request shielding capability and therefore the hyper geometric likelihood dissemination, with the intention to later be import as hgd. Peruses will suggest [15] for a lot of points of interest of opse. As initial look, it seems to be dynamic importance based encoding from before are finding out plans to opse is extraordinarily skilful. Be that because it will, opse is settled encoding conspires, whereby if information isn't treated well, at that issue barely error will spills bunches of records.

Issue Statement

Issue plan

In the early strategies for symmetrical hunt like downy phrase look then on, are preponderantly applied for finding out. Anyways, these techniques improve look ability and simple use. They take into consideration structure of phrases and regulate put off among terms to see likeness [16]. Yet, do not keep in mind the phrases semantically known with look watchword. The results are simply seeing able of best or group action of watchword. For example those plans merely don't forget sure wrong writing system or irregularities like "stated" or "expressed" are plan to be comparable. The utmost imperative factor that changed into result-positioning become even so out of requesting.

Utilization of framework

Semantic development primarily based entirely comparable pursuit upgrades convenience by returning actual coordinative files what is more restores the knowledge that of life-sized to allow inquiry watchword. From the data set server creates the altered document and builds the linguistics relationship library (srl) for watchwords set. Cloud server thus discovers all relevancy facts creating use of srl, once client affects Aninquiry to ask.

In the actual framework, to ensure security and shutting outcome positioning, created saving encoding is employed to stay numerical inquiring for making certain relevancy score.

The on top of clear approach reveals the centre issue that causes the wastefulness of placed accessible encoding. Server got to perform wanting and positioning unexpectedly through not understanding relevancy rating and totally different info of files.

The primary goals of the given arrange are tested below:

- to stipulate a glance conspire for disorganised cloud records that provides importance rating to statistics with inquiry phrase and returns the recovered documents all at once.
- To empower effective usage of facts files exploitation set there encoding conspire.
- To empower security by suggests that of averting cloud server from learning plaintext of knowledge records.

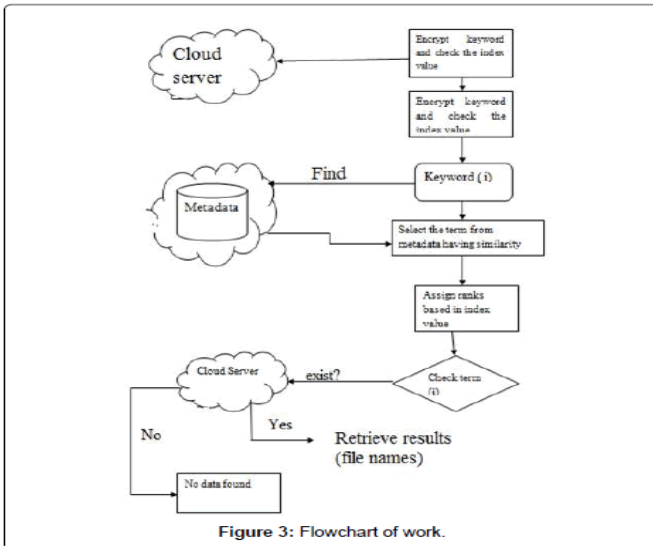


Figure 3: Flowchart of work.

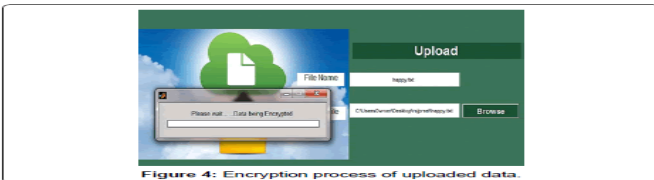


Figure 4: Encryption process of uploaded data.

Ventures of usage

1. In initial step, encryption of records is finished making use of aes (uneven encryption trendy) calculation. The usage is finished by means of developing close by situation in matlab. Special gui (graphical ui) are made for customer association. This calculation encodes the records report and moreover makes listing an incentive for every one among a type watchword [17].

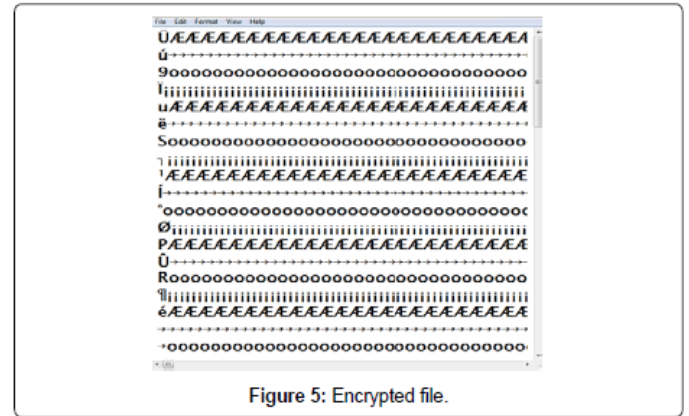


Figure 5: Encrypted file.

2. After encryption, ordering metadata is made by watchwords. In this relative items are spoken to.
3. Currently if any patron appears via a time period with the aid of giving query watchword, at that factor sse and opse are utilized for giving output in positioned arrange.

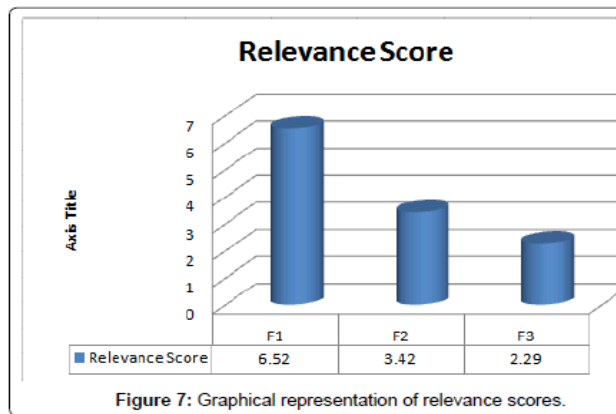
I. RESULT

The exploratory outcomes can be clarified with the arrangement of a few depictions.

1. As a matter of first importance an information proprietor transfers a few information record, which is scrambled utilizing AES calculation.
2. The transferred on cloud server can be looked security utilizing SSE (secure symmetric encryption) calculation. The query items are shown in positioned frame utilizing OPSE (arrange protecting symmetric encryption)
3. Calculation produce the significance score of documents in view of term recurrence (TF) and backwards area recurrence (IDF), utilizing the condition $TF \times IDF$.

File ID	Relevance Score
F1	6.52
F2	3.42
F3	2.29

Figure 6: Table for relevance score.



4. Subsequently in light of importance score documents can be positioned for more symmetry.

REFERENCES

- [1] Song D X, Wagner D, Perrig A (2000) Practical techniques for searches on encrypted data. Proceedings of IEEE Symposium on Security and Privacy, IEEE, Berkeley, California .
- [2] Goh E-J (2003) Secure indexes. Cryptology ePrint Archive, Report 2003/216
- [3] Boneh D, Crescenzo G, Ostrovsky R, Persiano G (2004) Public key encryption with keyword search. Advances in Cryptology-Eurocrypt3027: 506-522.
- [4] Chang Y C, Mitzenmacher M (2005) Privacy preserving keyword searches on remote encrypted data. Applied Cryptography and Network Security 3531: 442-455.
- [5] Chang Y C, Mitzenmacher M (2005) Privacy preserving keyword searches on remote encrypted data. Applied Cryptography and Network Security 3531: 442-455.
- [6] Curtmola R, Garay J, Kamara S, Ostrovsky R (2006) Searchable symmetric encryption: improved definitions and efficient constructions. Proceedings of the 13th ACM conference on Computer and communications security, ACM, Alexandria, VA, USA.
- [7] Bellare M, Boldyreva A, O'Neill A (2007) Deterministic and efficiently searchable encryption. In Advances in Cryptology-CRYPTO4622: 535-552.
- [8] Wang C, Cao N, Li J, Ren K, Lou W (2010) Secure ranked keyword search over encrypted cloud data. 30th IEEE International Conference on Distributed Computing Systems (ICDCS), IEEE Comp Society Washington, DC, USA.
- [9] Wang C, Cao N, Ren K, Lou W (2012) Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans Parallel DistribSyst 23:1467-1479.

A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation

Gopinadh Sasubilli¹, Uday Shankar Sekhar², Ms.Surbhi Sharma³, Ms.Swati Sharma⁴

¹Sr. Solution ArchitectMc Donald's, IL.

²Senior Associate,EY

³Assistant Professor, ACERC Ajmer Rajasthan, India

⁴Research Scholar, ACERC Ajmer,Rajasthan, India

¹kumarattangudiperich@kpmg.com, ²uday.sekhar@ey.com, ³surbhi2690@gmail.com,

⁴swati199410@gmail.com

Abstract—In the reference current scenario, data is incremented exponentially and speed of data accruing at the rate of petabytes. Big data defines the available amount of data over the different media or wide communication media internet. Big Data term refers to the explosion in the quantity (and quality) of available and potentially relevant data. On the basis of quantity amount of data are very huge and this quantity has been handled by conventional database systems and data warehouses because the amount of data increases similarly complexity with it also increases. Multiple areas are involved in the production, generation, and implementation of Big Data such as news media, social networking sites, business applications, industrial community, and much more. Some parameters concern with the handling of Big Data like Efficient management, proper storage, availability, scalability, and processing. Thus to handle this big data, new techniques, tools, and architecture are required. In the present paper, we have discussed different technology available in the implementation and management of Big Data. This paper contemplates an approach formal tools and techniques used to solve the major difficulties with Big Data, This evaluate different industries data stock exchange to covariance factor and it tells the significance of data through covariance positive result using hive approach and also how much hive approach is efficient for that in the term of HDFS and hive query. and also evaluates the covariance factors after applying hive and map reduce approaches with stock exchange dataset of around 3500. After process data with the hive approach we have conclude that hive approach is better than map reduce and big table in terms of storage and processing of Big Data.

Index Terms—Big Data, Hadoop; MapReduce; HDFS; Grid Computing; Big Table; Hive.

I. INTRODUCTION

BIG data is comparable to tiny knowledge however it's larger in terms of volume, selection and rate. Massive knowledge may be the next big factor within the IT space. Massive knowledge generates price from the storage and process of terribly giant quantities of digital data that can't be processed by standard info systems. The larger a part of the knowledge is delivered, put away, listed and handled over the net, prompting the enlargement in size of data sys-

tematically. This substantial live of data introduce over the net is alluded to as "Big Data". Massive knowledge characterized by the info quantity (volume), data speed (velocity) and differing types of knowledge (variety).

Volume: Volume denotes the dimensions of knowledge over the web. Presently it's in petabytes and is predicted to be raised to zettabytes. Knowledge from the good phones, sensors embedded into everyday objects can presently lead to billions of recent knowledge.



Fig. 1.Characteristics of Big Data

Velocity: Velocity inputs cover the speed of input generation and data managing. Online gaming systems support millions of concurrent users, each producing multiple inputs per second. [2].

Variety: Variety covers the type of input. Input can be constructed (text), unstructured (data generated from social networking sites and sensors) or semi-structured (data from web pages, web logs-mail etc).

Two more characteristics have also been included- Veracity and Value.

Veracity- It means how much the data is related to truth or facts.

Value- It covers the processing input and how the data can be combined with other data to extract meaningful information from it.

II. PROPOSED WORK

In the present paper, we have proposed distinctive apparatuses and strategies which are utilized to beat the regular is-

sues identified with huge information. Term Big Data examination includes devices, calculations, and design that break down and change substantial and monstrous volumes of information [10]. Big information investigation is an innovation empowered technique for empowering an association to have an aggressive edge over others by dissecting business sector and client patterns. Investigation on on-going information, online value-based information gives further experiences of the patterns to settle on opportune and exact choices. For the Computation reason for wide range volume of information [5] Big Data Computing is worried about the preparing, changing, dealing with and capacity of data. Frameworks, for example, Map Reduce, Hadoop, Grid Computing, and Big Table[8] have made composition and executing specially appointed huge information investigation and calculation simple. As web indexes have changed data get to, different types of enormous information registering can and will change the exercises like restorative and logical research, protection undertaking and so on. This paper focus on the following technologies:

A. Hadoop

Hadoop is actually a large scale batch data processing system. Hadoop developed as an establishment for huge information handling undertakings, for example, logical examination, business and deals arranging, and preparing huge volumes of sensor information, including from web of things sensors. Hadoop is supportable for distributed cluster system, parallel data processing system and worked as a platform for massively scalable applications. Facebook, Apple, Google, IBM, Twitter and hp are the famous hadoop users. Hadoop provide access to the file system called HDFS (hadoop Distributed File system). Basic capabilities of the hadoop include some packages like Apache Flume, Apache HBase, Apache Hive, Apache Pig, Apache Oozie and many more. Hadoop is beneficial in terms of cost efficient and reliable and scalable data processing. Different components of Hadoop system are explained below[10]:

B. HDFS Architecture

HDFS stands for Hadoop Distributed File System. It is an essential component of Hadoop which is used to store huge datasets. The main task of HDFS is to distribute the data to Various clusters of computers (machines) and then processing of this data is done. The advantage of using HDFS is that it coordinates the work among machines and if any one of them fails, Hadoop continues to operate by shifting the work from one machine to another without losing data or interrupting work [11].

C. MapReduce

MapReduce is a parallel programming framework that allows operations to be applied over large datasets. The main task of MapReduce is to divide the problem into smaller parts and then run those subparts in a parallel fashion. MapReduce consists of two functions: Map and Reduce.

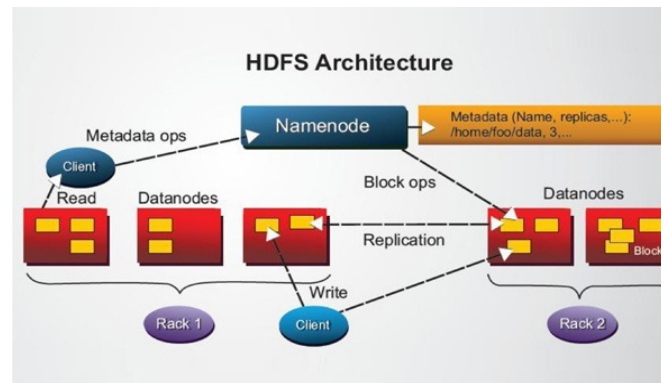


Fig .2. Architecture of HDFS

Map: This function generates a key/value pair and performs sorting and filtering of data.

Reduce: This function combines all the intermediate values and gives the output.

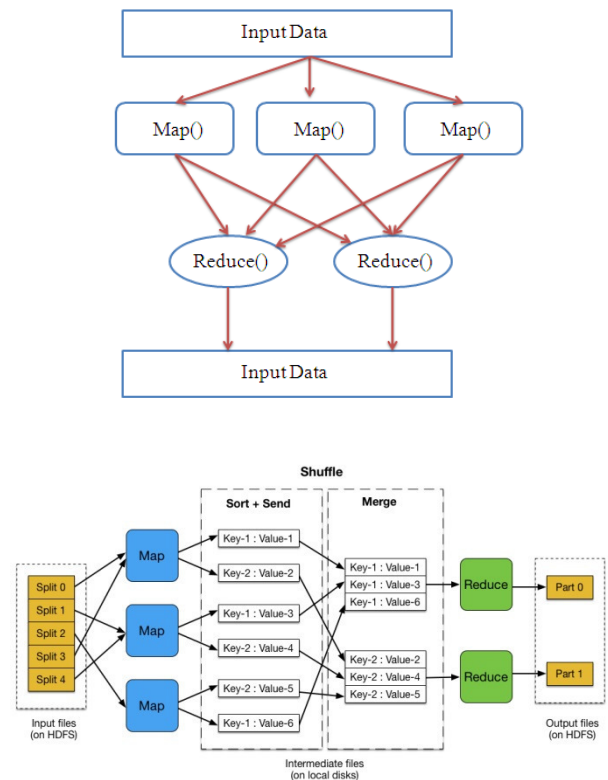


Fig.3.(a) Architecture of MapReduce (b) Internal structure of MapReduce

D. Grid Computing

A grid is a system in which a number of servers are connected to each other through a high speed internet. It is a distributed computing model in which the servers are geographically apart from each other and the users can access the data transparently from any location[12]. Although Grid

is beneficial as it provides hardware for storage of data but it has a drawback that current Grid infrastructure is not capable enough to handle Big Data. Thus research is still going on to find a solution to this problem so that it can deal with large volume of data.

Challenges with grid computing [12]:

- Data movement
- Data replication
- Resource management
- Job submission

Grid Architecture (Layered)

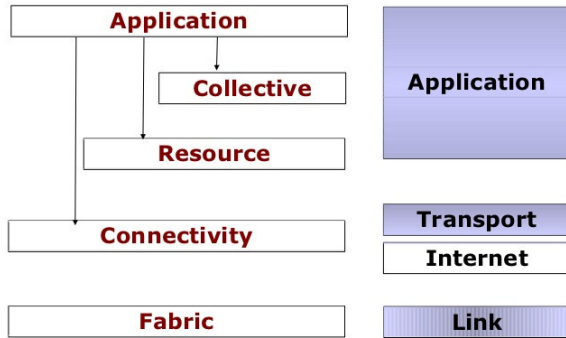


Fig. 4 Architecture of Grid Computing
a. Bigtable

Bigtable is a component to implement Big data which is similar to distributed storage system developed by Google. Bigtable is used for handling huge volume of data. It can handle data up to petabytes. It is a distributed and sparse map which is matricide by a row key element, column key element, and a timestamp value; each value in the map is anon disturbed array of bytes.

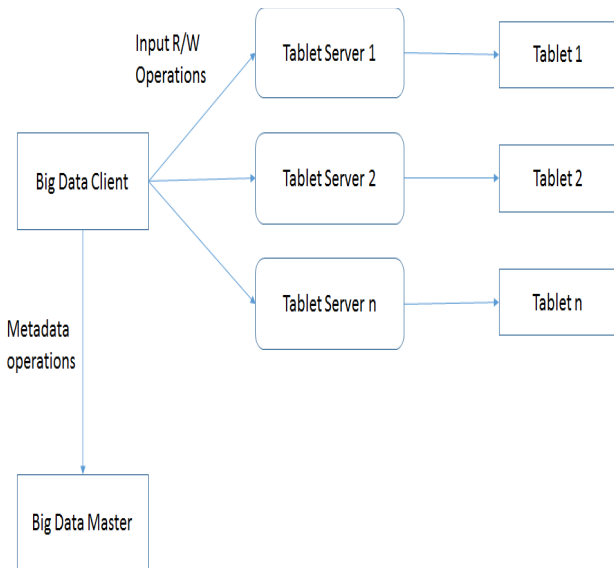


Fig.5. Architecture of Bigtable

Challenges with Bigtable implementation are:

- Reliability storage: durability and replication
- Sequential storage
- Structured storage

III. IMPLEMENTATION OF BIG DATA USING HIVE

Hive is techniques supported the SQL, it in addition uses a lot of customary and in some cases program secret writing that we might have to be compelled to implement Map Reduce programming. We have got to use Hive to interrupt down the stock and large knowledge set info, at that time we might have the advanced and entity relative calculus primarily based question to use the SQL skills of Hive-QL and connected info is overseen during a specific map and scale back mapping. it'll depreciated the advancement time and may administrate joins between the dataset (Eg. Stock info, Industrial data).Hive in addition has its main servers, by that we will gift our Hive queries from anywhere to the Hive server, that is employed to executes them. Hive SQL queries area unit being modified over into define employments by Hive compiler, and software system engineers have to be compelled to solve this advanced programming and solved the problems connected with massive knowledge and organization knowledge. For applying this methodology we have a tendency to could have to be compelled to use a dataset happiness to exchange and Dataset contains following properties:

- Data is being organized above all arrangement.
- It would judge joins to cipher Stock variance.
- It may well be sorted out into composition of various forms of be a part of.
- In neutral condition, info size would be extreme high.

exchange	stock_symbol	stock_date	stock_price_open	stock_price_high	stock_price_low	stock_price_close	stock_volume	stock_price_avg_close
NYSE	QTM	8/2/2010	2.37	2.42	2.29	2.36	3013600	2
NYSE	QTM	5/2/2010	2.38	2.5	2.34	2.41	2687600	2
NYSE	QTM	4/2/2010	2.57	2.64	2.39	2.46	4529800	2
NYSE	QTM	3/2/2010	2.64	2.67	2.55	2.63	2688600	2
NYSE	QTM	2/2/2010	2.69	2.76	2.56	2.66	2959700	2
NYSE	QTM	1/2/2010	2.6	2.8	2.52	2.67	6565100	2
NYSE	QTM	29-01-2010	2.63	2.73	2.26	2.56	16484000	2
NYSE	QTM	28-01-2010	3.09	3.09	2.95	3.06	3986400	3
NYSE	QTM	27-01-2010	3.03	3.1	2.99	3.03	2431900	3
NYSE	QTM	26-01-2010	3.07	3.18	3	3.03	4027600	3
NYSE	QTM	25-01-2010	2.94	3.07	2.93	3.03	2286400	3
NYSE	QTM	22-01-2010	2.94	3.1	2.89	2.9	2986700	:
NYSE	QTM	21-01-2010	2.94	3.11	2.92	2.95	4547800	2
NYSE	QTM	20-01-2010	2.94	2.97	2.88	2.93	1883900	2
NYSE	QTM	19-01-2010	2.89	2.94	2.88	2.94	2089700	2
NYSE	QTM	18-01-2010	2.86	2.96	2.85	2.88	2468000	2

Fig.6 Stock Exchange Dataset(.csv) file

- Issues related with map reduce are solved with Hive:
- Used Hive setup on Cludera.
- Create Hive Table:
- Use 'make table' Hive command to create the Hivetable for our consideredcsv format dataset

- hive > create table STOCK (trademark String, stock_symbols String, stock_datestart String, stock_price_opens twofold, stock_price_uptwofold, stock_price_bottom twofold, stock_price_closed twofold, stock_quantity twofold, stock_price_adj_close twofold)
- Load .csv info into Hive Table: hive> stack info neighbourhood inpath '/home/cloudera/STOCK.csv' into table STOCK;
- V. This can stack the dataset from the required space to the Hive table 'STACK' as created on top of but this dataset are place away into the Hive-controlled record framework namespace on HDFS, with the goal that it can be bunch ready more by MapReduce employments or Hive queries.
- VI. Calculate the Covariance factor.
- VII. We can figure the Covariance for the gave stock dataset to the inputted year as beneath utilizing the Hive select inquiry:
- VIII. Select
 a.STOCK_SYMBOLS,b.STOCK_SYMBOL S,
 month(a.STOCK_DATESTART),
 (AVG(a.STOCK_PRICE_UP*b.STOCK_PRICE_UP))
 (AVG(a.STOCK_PRICE_UP)*AVG(b.STOCK_PRICE_UP))from STOCK a join STOCK b
 on
 a.STOCK_DATESTART=b.STOCK_DATESTARTwherea.STOCK_SYMBOLS<b.STOCK_SYMBOLS
 and
 year(a.STOCK_DATESTART)=2008 Group
 by
 a.STOCK_SYMBOLS, b.
 STOCK_SYMBOLS,
 month(a.STOCK_DATESTART);
- This Hive select query will trigger up the MapReduce work as below:

```

hive> select a.STOCK_SYMBOL, b.STOCK_SYMBOL, month(a.STOCK_DATE),
> (AVG(a.STOCK_PRICE_HIGH*b.STOCK_PRICE_HIGH) - (AVG(a.STOCK_PRICE_HIGH)*AVG(b.STOCK_PRICE_HIGH)))
>
> from nyse a join nyse b on
> a.STOCK_DATE=b.STOCK_DATE where a.STOCK_SYMBOL<b.STOCK_SYMBOL and year(a.STOCK_DATE)=2008
> group by a.STOCK_SYMBOL, b.STOCK_SYMBOL,
> month(a.STOCK_DATE);
Total MapReduce jobs = 2
Number of reduce tasks not specified. Estimated from input data size: 1
Launching Job 1 out of 2
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=number>
Starting Job = job_201407272047_0004, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201407272047_0004
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201407272047_0004
2014-07-28 03:24:50,854 Stage:1 map = 0%, reduce = 0%
2014-07-28 03:24:12,987 Stage:1 map = 100%, reduce = 0%
2014-07-28 03:24:20,938 Stage:1 map = 100%, reduce = 33%
2014-07-28 03:24:21,900 Stage:1 map = 100%, reduce = 100%
Ended Job = job_201407272047_0004
Launching Job 2 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=number>
In order to limit the maximum number of reducers:
set hive.exec.reducers.max=number>
In order to set a constant number of reducers:
set mapred.reduce.tasks=number>

```

Fig 7.Hive Approach Implementation

The covariance result for 2008 are as below :

Stock Symbol(A)	Stock Symbol(B)	Month	Covariance
14/07/28 03:44:32	INFO jdbc.HiveQueryResultSet: Column names: stock symbol,stock symbol, c2, c3		
14/07/28 03:44:32	INFO jdbc.HiveQueryResultSet: Column types: string,string,int,double		
QRR	OTM	1	-0.139949659863395513
QRR	OTM	2	2.060000000021489E-4
QRR	OTM	3	0.00292999999999956583
QRR	QXM	1	-0.015941496598628646
QRR	QXM	2	0.00512499999999964075
QRR	QXM	3	-0.0133579999999998244
QTM	QXM	1	-0.003653287981055158
QTM	QXM	2	-0.0263524999999998083
QTM	QXM	3	0.006057000000000201
QTM	QXM	4	0.02727107438016496
QTM	QXM	5	0.026688662131521212
QTM	QXM	6	0.05287852154195315
QTM	QXM	7	0.023126033057851103
QTM	QXM	8	0.022061224489796416
QTM	QXM	9	0.05976031746831918
QTM	QXM	10	0.0035079395085071408
QTM	QXM	11	0.018371745152354402
QTM	QXM	12	-0.0038603305785123165

From the variance issue, stock dataset recommend the subsequent conclusions: For Stocks QRR and QTM, these are having additional positive variance than negative variance, therefore having high chance that stocks can move along same means.

- For Stocks QRR and QXM, these are for the foremost half having negative variance. Therefore there exists an additional distinguished chance of stock prices acquiring a reverse course.
- For Stocks QTM and QXM, these are typically having positive variance for particularly else months, therefore these tend to maneuver an analogous means the bulk of the circumstances. So this discourse analysis comprehends the attendant 2 crucial objectives of giant data advances:
 (a) Storage: it's the deepest connected issue for huge stock data into HDFS, the arrangement provides considerably additional strong, strength, scalable, and elastic.
 (b) Processing: In several Hive composition it relies on a typical SQL information, we tend to could get the advantage of running SQL queries on the large dataset likewise and may method the massive quantity of GBs or TBs of data with basic SQL queries.

IV. CONCLUSION AND FUTURE SCOPE

We have conclude that map reduce approach is limited for small level data set and required a larger amount of storage to hold the map level and reduced data set recursively but we have used Hive approach to evaluate covariance among our considered data set and it shows the result that the covariance between QTM and QXM parameter is positive. Another factor is that the amount of storage over HDFS is limited under hive approach and processing is programmed with hive SQL Query which is used to take a shortest time for execution for petabytes amount of datasets. Legitimate and powerful examination of in-depth volumes of data can prompt speedier advances in varied logical teaches and enhance the profit and accomplishment of various enterprises. The difficulties incorporate the difficulty of in-depth volume, however additionally no uniformity, unclear structure, blunder coping with, protection, favorableness, security cradle, combination, and illustration. These specialized difficulties area unit found an immense assortment of use areas and consequently force an immense value. Besides, these difficulties would require would force transformative arrangements and can require an intensive

type of apparatuses, systems, and applications to manage. With a particular finish goal to accomplish the bonded benefits of massive information, this stuff should be taken underneath thusly thought so most capability will be determined to select up an associate aggressive edge.

To take out the simplest have the benefit of Hadoop, the in-depth analysis must be applied and revolutionary tools and techniques must be developed to rigorously comprehend and properly reply to numerous challenges.

REFERENCES:

- [1] Lawal Muhammad Aminu, "Implementing Big Data Management on Grid Computing Environment", *International Journal of Engineering and Computer Science* ISSN: 2319-7242, Volume 3, Issue 9, September 2014, Page No. 8455-8459
- [2] Agrawal et al., 2011; Baer et al., 2011 Agrawal, D., Das, S., & Abbadi, A. (2011), *Big Data and Cloud Computing: Current State and Future Opportunities*. ACM EDBT Conference, March 22–24, 2011, Uppsala Sweden. <http://dx.doi.org/10.1145/1951365.1951432>
- [3] Baer, T. (2011). 2012 Trends to Watch: Big Data. Ovum Report, OI00140-041. Baer, T., Sheina, M., and Mukherjee, S. (2011). What is big data? The big architecture. Ovum Report, OI00140-033.
- [4] S. Vikram Phaneendra & E. Madhusudhan Reddy "Big Data-solutions for RDBMS problems- A survey" In 12th IEEE/IFIP Network Operations & Management Symposium (NOMS 2010) (Osaka, Japan, Apr 19{23 2013).
- [5] Kiran kumara Reddi & DnvsI Indira "Different Technique to Transfer Big Data: survey" *IEEE Transactions on* 52(8) (Aug.2013) 2348 {2355}
- [6] Jimmy Lin "MapReduce Is Good Enough?" *The control project. IEEE Computer* 32 (2013)
- [7] Umasri M. L, Shyamalagowri. D, Suresh Kumar. S "Mining Big Data:- Current status and forecast to the future" Volume 4, Issue 1, January 2014 ISSN: 2277 128X.
- [8] Albert Bifet "Mining Big Data in Real Time" *Informatica* 37 (2013) 15–20 DEC 2012
- [9] Zan Mo, Yanfei Li Research of Big Data Based on the Views of Technology and Application *American Journal of Industrial and Business Management*, 2015, 5, 192-197 Published Online April 2015 in SciRes.
- [10] Harshawardhan S. Bhosale¹, Prof. Devendra P. Gadekar² "A Review Paper on Big Data and Hadoop" *International Journal of Scientific and Research Publications*, Volume 4, Issue 10, October 2014 1 ISSN 2250-3153
- [11] <https://www.quora.com/What-are-the-main-features-of-Hadoop>
- [12] <https://www.slideshare.net/sandpoonia/1-grid-computing>
- [13] <http://stackoverflow.com/questions/782913/googles-bigtable-vs-a-relational-database>

A Detailed Study of EEG based Brain Computer Interface

Geeta Sharma, Neha Sharma, Tanya Singh, Rashmi Agrawal

Faculty of computer Application, Manav Rachna International
Institute of Research and Study, India
{geetasharma.2895, nsdp8605, tani9413, drrashmiagrawal78}@gmail.com

Abstract—Brain Computer Interface (BCI) generate a direct method to communicate with the outside world. Many patients are not able to communicate. for example:- the patient who are suffered with the several disease like post stroke - the process of thinking, remembering & recognizing can be challenging. Because of spinal cord injuries or brain stem stroke the patient loss the monitoring power. EEG based brain computer interface (BCI) feature is beneficial to scale the brain movement & convert them into a instruction for monitoring. In this paper our objective is to study about various applications of EEG based signal of the different disease like spinal cord injury, post stroke and ALS (amyotrophic lateral sclerosis) etc.

Index Terms—Brain Computer Interface; Electroencephalography (EEG) signal; Amyotrophic lateral Sclerosis (ALS); Post-Stroke; Electrocorticography (ECoG).

I. INTRODUCTION

BRAIN computing interface (BCI) is a computer based technology that allows the mortal mind to communicate with electronic devices such as motorized wheelchair, robotics arm and artificial limbs without assist of any peripherals devices i.e., nerves and muscles. The objective of BCI is to assess the electrical signals through the brain then amplified them and filter them to remove unwanted signals. And then convert the analogical signal to digitally signal to display on screen in the output form. Finally the digitally signal is record in a devices such as compact disk, document, floppy disk, DVD etc.

This process is occur when the different brain activities are generated through the mortal mind that will be identified by the system and translated into commands which depends on a classification algorithms.

Measuring Brain Signals:- Hans Berger who started the BCI invented the different electrical activity for human brain & developed. Brain waves can be evaluated with various techniques that each have advantage & disadvantage. A commonly used techniques are:-

Electroencephalography(EEG) - EEG is used to analysis the neurochemistry in the brain through electrodes. EEG is the study of non-invasive interface that monitors the brain activities from outside the skull. It doesn't require surgery or

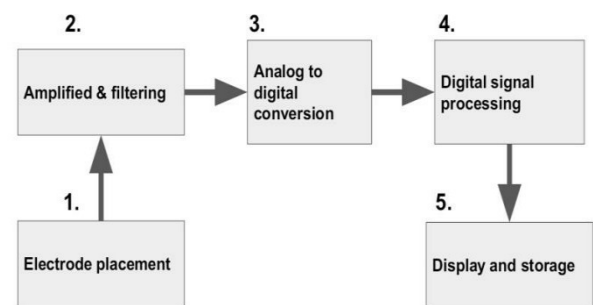


Figure 1. Brain signal recorded on the scalp through electrode, amplified and filter than it convert brain wave into machine language

permanent attachment to the device. Techniques such as computerized tomography (CT), magnetic resonance imaging (MRI), X-rays, ultrasound have all been extract the brain activity noninvasively.

EEG based BCI:- is refers to the brain activities and that considered as the Rhythmic Brain Activity. These Rhythms are pretended by distinct actions and thoughts.

EEG classified these rhythms references to their frequencies.

Delta waves:- it is our lowest level of brain wave activity. This brain signal that occur during deep dreamless sleep when the frequency up to 4 Hz.

Theta waves:- when we are in dream, vivid imagery & information beyond our normal conscious awareness & we hold our stuff fear, nightmares then theta waves are arise, in this frequency are 4 to 7 hz.

Alpha waves:- when the brain is in resting state.or in relaxation state or the frequency are 8 to 12 hz then alpha waves signals are generated through brain.

Beta waves:-it is the best activity, present when we are alert, attention, engaged in problem solving and able to make decisions & judgement. When frequency are 12 to 30 hz then beta waves signals are generated through brain.

Table.1.brain rhythm activities with its frequencies.

Rhythmic Activities	Frequency (Hz)
Delta	Upto 4
Theta	4-7 Hz
Alpha	8-12 Hz
Beta	12-30 Hz

Electrocorticography (ECoG) – EcoG devices are embedded innermost the skull but remain outside the brain, they produce signals with better resolution as related to invasive they have reduce hazard of scar tissue. ECOG that considered cathodes located instantly on the exposed surface of the brain to store electrical activities from the cerebral. It is safer, more consistent and less technologically complicated than invasive BCI.

ECoG based BCI:- Is refers to the cursor movement that serve as the prototype for testing a new paradigms & algorithms. While contacting the online observation of ensional cursor patients activity through one-dimcontrol which allows the participants to control either vertical or horizontal movement and through two-dimensional cursor control which enables the user to control a considerably enhanced interface. ECoG uses different activities for the cursor control testing are :-

Supplementary Motor Area (SMA) is the part of cerebral cortex especially for control of movement. SMA is the most important part of the motor imagery based BCI application because it is based on the control of movement which is basically justified the oscillate persistent activities during the synchronized and unsynchronized movement. The activity is considered an online based measurements of brain activities.

Phase Locking Value (PLV) is an activity that can be used to analysis the performance that caused changes in long range synchronization of neural activity from EEG data. PLV measurements feature is currently established as BCI motor imagery application.

BCI Operations:- BCI Operations are essential to record the brain signals by applying the four elements:- signals acquisition, feature extraction, feature translation & device output through brain activities are generated and captured by the system then transmitted into commands which is considered as output.

Signal acquisition :- In signal acquisition we captured the signal from various location of the mind. it is row types of data which is meaningless.

Feature extraction :- the captured signal have noise and unwanted data. In feature extraction the noise and unwanted data are remove by using algorithms and get the desired data.

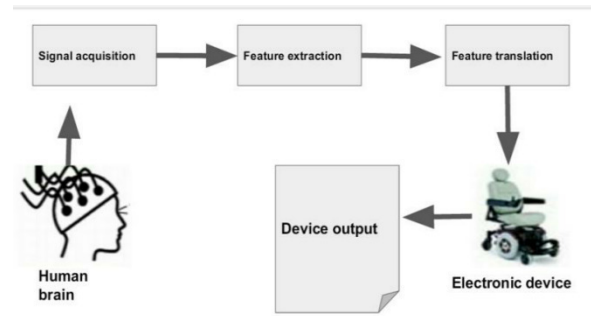


Figure 2. Mortal mind interacting with artificial device by the following different operation.

Feature translation:- The clean data converting into a command by the feature translation by the classification algorithm.

Device output:- when signals are classified the output is manipulated on the computer screen.

Problem Statement:- All of the recently BCIs subordinate under the progress of many modifications. The Problem is safety and the long-range stability of capturing cathodes utilized in invasive BCI systems remains to be determined sufficiently but not all are considered to determine fully connected with telemetry device to transfer the captured intellectual waves. Some BCI applications need an inclusive arrangements of user commands e.g., various psychological tasks, these must be detect by improving the cognitive adaptation and learning algorithms also in additional areas, such as speed, accuracy, consistency, convenience are essential for achievement of practical BCI systems.

II. RELATED STUDY

According to [1], The scientific intellectual that founded obvious connection & control between the human brain and computer, known as BCI.

BCI is a computer based activity that obtain Brain signal without the help of any peripherals device such as nerves & muscles. There are different techniques of recording signals & then apply BCI operation; signal acquisition, feature extraction, feature translation, device output by these operations electronic signals of the brain is captured over the scalp and then transmitted into commands which is considered as output that improved the task between the user and the system for that efficient BCI applications are considered that established the useful feature for the paralyzed people & invent the electronic devices such as motorized wheelchair, robotics arms & artificial limbs. To develop the BCI system many of the applications are working effortlessly.

According to [2], The purpose of BCI is to evaluate the electrical signals through the brain & then amplified them and filter them to remove unwanted signals and then convert the analog signal to digital signal. This considered the real time motor imagery process of BCI through external devices are control which is beneficial for disabled people that

become the communication media for them. During motor imagery the BCI is use the physiology of the nervous system which extract or classify rhythms using SVM classifier which detect the bugs and insufficient signals and converting them into desired result.

Through SVM analysis the data of motor imagery movement of right & left hand is classified that considered the rhythm. SVM is used to separate or identify classes, it minimize the margins to increase the generation capabilities as compared to LDA classifier which minimize distance within a class & maximize distance between classes.

According to [3], BCI use two technique EEG and ECoG. EEG is the non invasive interface that monitors the brain activity from outside the skull. ECoG implanted inside the skull but rest outside the brain, they produce signals with better resolution as compared to invasive so, result suggested that ECoG signals has better resolution as compared to the EEG because sometimes EEG records unwanted signals that give wrong conclusion. Hence ECoG overcome the limitations of EEG. The ECoG also control the signals into cursor movement which operate tongue movement, opening & closing hand movement and the other actions, they are considered as the one dimensional or two dimensional BCI control, while experience the online observation of patients actions. The data is analyzed that helps to confirm the better adjustment of frequency bands & anatomic location used for ECoG control of cursor movement & final accuracy of that control.

According to [4], The different approaches and the applications are used for the functionality to generate brain signals which classified the BCI in two way.

Invasive BCI that consider as the neurosurgery because this technique consider the internal working of the brain. It is basically a kind of open surgery so it make effect on the body part.

Non-invasive BCI does not break the skin so it considered to be as compared to the invasive safest as compared to the invasive. BCI considered the real time based Applications such as prosthetic application which considered the artificial limbs. Entertainment applications that include gaming, communication through which human brain interface with the electronic devices. For the safety & security also BCI Applications are used. BCI also faced the challenges while understanding the brain activities, capturing the signals for that BCI operations are considered through the brain signals are transmitted in the output form.

According to [5], BCI introduced the technique which are ECoG and EEG based through which many new devices are generated to control the motor disabled people activities which developed the high degree of complexity. For the evaluation of motor imagery activity ECoG records & analysis the observation by using BCI methodology which are Genetic Algorithm (GA), Mutual information (MI), Information Gain (IG) & compare them with Bayesian & K-Nearest Neighbor classifiers. For the high dimensionality & outliers feature extraction Common Spatial Pattern (CSP)

used in signal processing for separating multiple observation in BCI. The comparative study of these methodology with BCI classifiers aim to develop the new communication media to the motor disabled patients.

According to [6], EEG based BCI performing the task by adopted by specific training method that based on intellectual task i.e Mutual information (MI), calculate the changes in brain rhythms. The sensorimotor rhythm (SMR), when the frequency ranges of rhythms is 13 to 15 Hz then it considered alpha & beta brain waves. Sensorimotor considered the stage in which the physical & cognitive skills are developed. While calculate the observation of motor imagery BCI different feedback moods are considered the audible & vibractic result which produce strength or support of user task in positive way or to edify inrich effort of BCI task which consist as negative feedback. After combining the all feedback moods compare them with conventional viewable result. On the basics of compilation different session must be considered to analysis the BCI new strategies. The multiple comparison tests is considered in the sessions that improve the performance of BCI system.

According to [7], A motor imagery BCI transmitted the control signals by applying the various algorithms which considered the rhythms according to the frequency. While measure the phase synchrony for classifying single trial EEG during motor imagery that examine or analysis the activities of Supplementary Motor Area (SMA), is the part of cerebral cortex especially for control of movement basically justified the oscillate persistent activities during the synchronized and unsynchronized movement. The activity is considered an online based to tested the phase synchrony measurements in motor cortex for classifying EEG during motor imagery. The phase difference between frequency & signals that generally oscillate persistent value. It is essential to test synchrony testing is known as Phase Locking Value (PLV), activity that can be used to analysis the performance that caused changes in long rang synchronized of neural activity from EEG data. The PLV measurements features is currently established as EEG motor imagery in BCI application.

According to [8], The initial study are performing motor imagery and bCI application. In which analysis of scalp information in EEG based to left to right task in BCI. People are suffering for muscular disability who need to interact with electrical device. Most of the challenges and promises suffering people in BCI. In which used of BCI Invasive or noninvasive. Invasive BCI used open surgery in human being. It is risky for the suffering people as it is directly attach to the scalp of brain and in which used electrocortigraphy. Noninvasive BCI are used in attach to the outer-surface of the scalp. In noninvasive BCI used electroencephology EEG function. BCI based on rhythms that come from motor imagination. Rhythms record the signal between 8-12Hz to left to right. BCI motor imagery allow the EEG based electroencephology used in scalp of BCI because it is not risky of the suffering people.

According to [9], BCI is another types of field likes contribution of any medical field to suffering serious injuries. In this field such as educational system, self-configuration, marketing as different type of games and entertainment system in BCI. BCI of user and system. There are too many external-internal devices to use the muscle to complete iteration with people. In this BCI lack of challenges to restore or replace the lot of functionality. The BCI purpose involvement or analysis hand free application manage the body part and control the human behavior sending and receiving the information. In brain signal there are two types of signal are used invasive or noninvasive. Brain computer interface used noninvasive so that it is attach to the scalp. There are various challenges and application used in BCI and analysis day by day.

According to [10], BCI is a real world technique which large training or large contribution of real system. The main aims of BCI connect to the real world system increase the growth of application about helpful disable people. There are such a type of signal processing to use this technology to connect with brain signal.

Challenges of BCI to real world task:-

1. The average signal information find out the signal and capture the signal rate are increase or decrease.

2. If the signal rate are very slow i.e. information are transfer vary slowly in the body. That's way error are generated. In the body signal are high or nor either low.

3. Autonomy:- Brain signal comes through the machine signal. In these signal are not change on the machine. He completely control and manage the body and remove the noisy data.

4. Real world system are more difficulty in imagination to task per formation, emotion, feeling, interaction with any other person. We are study on BCI to control on every days use in every stage and situation

According to [11], Brain computer interface connection of brain and machine device to communication with externally and internally control by the signal. It is fluctuate the voltage between neurons and signal applier. Invasive BCI are direct connected to the scalp that's way invasive signal are higher quality of signal but it is risky for human. Noninvasive used EEG or MEG magnetic resonance imaging. EEG used in scalp of outside the brain. It is attach to the wire in in the device to controlling the robots, games movement of cursor etc. The channel are select EEG signal. This means BCI are signal equitation of EEG or MEG then select the useful signal and transfer the noise signal then feature extraction the classification of the signal. Now a days we are study of the many technology remove unwanted and noisy signals.

According to [12], Machine learning is a priority of pre-processing technology and EEG base of BCI. This BCI takes minimum time and gives the communication speed to 6-8 per minute. In the EEG analysis online separation of electrical signals. The current argumentation between two function. Now we are study of EEG based signal. Brain

computer interface communication with brain signal, from muscles and peripheral device. Signals are fluctuate the according voltage to the user of user device. Machine learning is also analysis of emotion, sleep stage, feeling, through etc.

According to [13], BCI task with the paralyzed people. It use the noninvasive of EEG based to the signal.it provide multidimensional movement to manage the body part through which the brain. Noninvasive BCI used in scalp activity to provide the signal point to point the control injuries. The research is the main motive of the provide the communication with the suffering people. It is provide the communication in two types of invasive or noninvasive. Noninvasive used in scalp of EEG based and invasive is open surgery. It has provide high speed of the communication. It is a clinical risk or infected of body.

According to [14], EEG and motor imagery in real world to communicate with other. In which the classification between EEG and BCI. This paper describe by various algorithms in real time testing. Brain computer interface work with left and right hand movement and these frequency are 70 to 95%. It store the information to provide faster feedback. Rapid prototyping analysis of update very fast technique application to design the signal. These are flexible to store the information.

According to [15], Here we learn about EEG based BCI Application for stroke rehabilitation by robotics arms. Stroke is a big injury by which many other problems face by the patients like voice, focusing and sensibility. Rehabilitation help for monitoring & communicating with daily life for patients. Rehabilitation is better then manual therapy due to its instants improvements in patients movements activities. BCI helps for monitoring the external devices & systems like computer, switches and robotic wheelchair. Functional Electric Simulation(FES) of BCI, is utilized for hand activities. In FES we study on two exercise first quadriplegia patients that want to grip a glass of water by using BCI FES device and other one is a stroke patients. The youth mind is very strong & its memorize power is high so after the stroke problem it can take action by the action of damaged, injured area. The memorizing power is called as neuro-plasticity which focus on every location where command is given from one part to another. We find out the chances of using non-invasive BCI & robotic rehabilitation for post-stroke rehabilitation. This technique is applied for converting the fantasy into the instruction by acquiring the signal from scalp that recorded by eeg. it gives better performance and accurate result. We are also study about translate synchronization to asynchronous role of motor vision where object are controlled by artificial limbs.

According to [16], While analyzed the classification algorithms for generate BCI system of EEG algorithms that recognize the brain activity and amplified them after that break down them into different ranges according to signal are generated. BCI have five types of classification algorithm that are linear classifiers, neural networks, non-linear, Bayesian classifiers, nearest neighbor classifiers & combina-

tion of classifiers. The SVM are usable for synchronous BCI. The k nearest neighbor(KNN) algorithm & mahalono-bis distance algorithm are used for nearest neighbor classifiers. Their are many other classification techniques that not use today's day for objective of BCI. If BCI explore more in clinical activities, new technologies will be applied into the uses. these techniques are perform on the large amount data sets of EEG signal.

According to [17], The normal users can simply monitoring the robots by using the keyboard, mouse, joystick but its complex to abnormal or paralysis patient to control. these some interfaces sip & puff system, single switches and eye -tracking systems are help for those people. But these technique do not applied on other injured patient such as amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS) and stroke. Still autonomous robots were not useful for transferring the particular location for the muscular disable people that's why these type of interfaces or medium are generated for those patients who can not control the computer or system.

EEG is the part of non-invasive interface of brain computer. There are various application of eeg signals like-controlling a moving cursor on the screen, selection of letters from a virtual keyboard, playing games, pattern recognizing etc.

The EEG based robots are control by the mind signal. Brain controlled manipulators & mobile robots these are the two types of mind controlled robots. A FRIEND system generate by the Graser et al. Which is the manipulator. Non-invasive BCI access the many types of mind signal like-electroencephalograms(EEG), magnetoencepalogram (MEG), blood-oxygen-level dependent(bold) signals, but eeg is more demanding because of low cost.

According to [18], As we can communicate with the outside world but many patient not able to communicate.

For example:- the patient who cause the several disease like post-stroke, the process of thinking, remembering & recognizing can be challenging because of spinal cord disease or brain stem stroke, the patient loss the monitoring power.

The clinical application of neurorehabilitation based on brain computer interface on EEG for monitoring motor refit. It helpsfor monitoring the patient which have amyotrophic lateral sclerosis & spinal cord disease & stroke.many features like sensorimotor rhythms (SMR), slow cortical potentials(SCP), SMR'r SCP's based system use the factor like frequency or time domain. The motive of clinical application pof bci is useful for patient have disease like amyotrophic lateral sclerosis(ALS), several CNS damage such as spinal cord injuries & stroke and epilepsy and attention deficit & hyperactive disorder (ADHD). Result is high shortage in controlling & motor function.

BCI gives choice & complementary monitoring control & motor control refit & down-regulation of cortical activity. it provides neural network communication with the device to motor control & full movement of damage area. the cortical

physiology that confirm the way in which human brain or mind understand.

According to [19], Electroencephalography (EEG) based on BCI can read the people consideration, thought & thinking for motor vision. The device applicable for disable people by reading the brain of people which exist left & right imaginary movement & judge the actual movements. Kalman filter of BCI's is applied for signal manipulation to selection of the efficient signal. the common spatial pattern(SCP) and radial Basis function(RBF) are the classification method that used for feature extraction. The common spatial pattern(CSP) is applied in channel selection. The best channel picking on the basis of maximum vector derived from CSP. The complex Morlet wavelet tranform technique used in alpha & beta rhythms. It discover & sorted the left & right movement imagination. The feature subset selection algorithm for feature extraction & multilayer backpropogation neural network for the classification Increase the efficiency of the classification of motor imagery left and right hand movement by using the kalman filter.

According to [20], BCI is a mediator between mortal mind & system. Electroenceplograpgy (EEG) is applied for better performance & effective or accurate classification. Sparse representation is applies for mental task classification.this algorithm is divided into two classes according to the BCI data. Common spatial pattern(CSP) was a big tool for using the dictionary. Differential dictionary that introduce by Zhou et al. Which is basis on the minimization objective function renovate & differential technique occur in objective function. Mostly sample of sparse representation are zero by this it provide the accurate result. It provide separability & adaptivity. sparse representation is better to other classifiers support vector machine(SVM) and K nearest neighbor(KNN) Because kernel in SVM and neighbor in KNN are difficult.

BCI Based Algorithms :- During motor imaginary, the BCI uses the physiology of the nervous system which are classified and extract the rhythms based on their frequencies. The Classification algorithms used to design BCI system by considering the BCI Operations. For the BCI classification algorithms different classifiers is used to extract the signals in the commands.

Linear Classifier :- the linear classifier are based on the different classifiers algorithms which are

Linear Discriminant Analysis (LDA) :-

is a method used in pattern recognition & machine learning to find a linear combination of features that separates two or more classes of objects or events. It minimize distance within a class & maximize distance between classes. The procedure commonly used for multiple classes is the 'One Versus The Rest'(OVR) which exist in separating each class from all the others. This procedure has a very low proportional obligation which makes it convenient for online BCI be system. This classifier is simply to use & commonly produce good results. LDA classifier be goes with asynchronous BCI & has been used with prosperity in a excellent

number of BCI system such as motor imagery based BCI. The disadvantage of LDA is its straightness that produce inferior results as compared to nonlinear EEG data.

Support Vector Machine (SVM):-

This method is also used to separate or identify classes. By minimize the distance from the nearest point & maximize the margins to increase the inference techniques. It's capacity to manage the high Spatial obstacles allaying the effect of 'the curse of dimensionality'. It can be used for both linear & nonlinear problems by using the 'kernel trick' which escape the explicit mapping that is needed to get linear learning algorithm with radial basis function. Kernel is generally used in SVM to identify the distance between two Vectors.

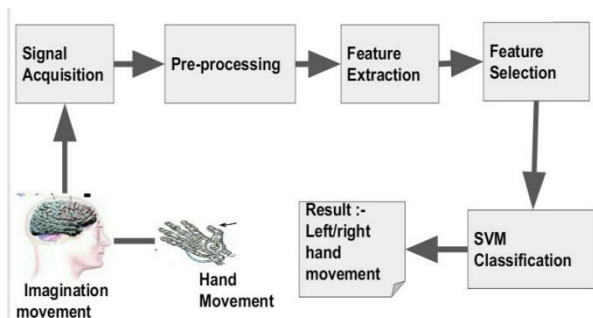


Figure 3. BCI monitoring the hand movements by applied the SVM classifier.

Nonlinear classifier:- It initiate Bayesian classifiers used for Bayes quadratic has been applied for BCI, it is not represented here as it is not conventional, presently, persistent sufficient for real-time BCI. This classifiers generate nonlinear decision boundaries. It is originated which permits them to operate further effectual alienation of unreliable patterns than discriminative classifiers. However, these classifiers are not as popular as linear classifiers or Neural Networks in BCI applications.

Bayes quadratic:- Bayesian classification objective is allotting a element vector to the class it exists the highest opportunity. The Bayes rule is applied to evaluate the subsequent opportunities that a element vector has belonging to a given class using the MAP (Maximum A Posteriori) rule, the class of this element vector can be considered Bayes quadratic include a different ordinary grouping of data that control quadratic decision boundaries. This classifier is not popularly used for BCI, but include in success to motor imagery and mental task classification.

K Nearest Neighbor(KNN):- K nearest neighbor(KNN) is the non-parametric method which is applied for the classification. it is used for statistic calculation and pattern recognition. The input have the adjacent example of the k and the output is the class membership like any other classifier or decision tree. Object is classified by the majority of vote of its class neighbor an k refer to the number of neighbor which we consider and it is a positive number typically small.

If the $k=1$ then variable is allocate to the class of the closest element it has.

Neural network:- neural network are work with linear classify. In which several artificial activities works to recall the brain signal. These algorithm are work on EEG (electroencephalography), to take short time classifier given to the brain. Neural network are applied in standard multilayer perceptron (MLPs) methods classified to the brain activities. These network are applied to compare FIR (finite impulse response) to clear purifying the processing of signals.

III. CONCLUSION & FUTURE STUDY

BCI is majorly used in science history which provide functionality between mind and gadget which consider the many application for further use transmission tool for severe disability and artificial limbs to monitoring disabilities such as quadriplegic, cerebromedullospinal. Neurorehabilitation after neurological injuries and entertainment & gaming. Polygraph test, brain fingerprinting, trust assessment. BCI is a field of practical study so it always in it's infancy. The comparative research on various components of BCI is ongoing that investigate useful instinctual waves; signal capturing, feature extraction and translation methods for combining short- and long-range adaptations between human and device to improve the performance that are suitable for BCI applications such as clinical applications; communication, movement control, Neurorehabilitation.

ACKNOWLEDGMENT

We, student of FCA completed this paper under the guidance & supervision of Dr. Rashmi Agrawal and Dr. Prasanjit Banerjee Accendere KMS pvt. Ltd.. Their contribution are sincerely appreciated and gratefully acknowledged.

REFERENCES

- [1] Joseph N. Mak [Member, IEEE], Jonathan R. Wolpaw, Albany, Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects
- [2] Roxana Todorean 1, Iuliana Chichisan. Application of Support Vector Machine for the Classification of Sensorimotor Rhythms in Brain Computer Interface
- [3] Eric C. Leuthardt, Kai J. Miller, Gerwin Schalk, Rajesh P. N. Rao, and Jeffrey G. Ojemann. Electrocorticography-Based Brain Computer Interface—The Seattle Experience
- [4] Aroosa Umair, Ureeba Ashfaq, and Muhammad Gufran Khan. Recent Trends, Applications, and Challenges of BCI
- [5] Aswineshadri. K Dr. V. Thulasi Bai. feature selection in brain computer interface using genetics method
- [6] I. N. Angulo-Sherman and D. Guti'erez,. Effect of different feedback modalities in the performance of brain-computer interfaces
- [7] Yijun Wang, Bo Hong*, Xiaorong Gao, and Shangkai Gao. Phase Synchrony Measurement in Motor Cortex for Classifying Single-trial EEG during Motor Imagery
- [8] Lei Qin, Lei Ding And Bin He. Motor imagery classification by means of source analysis for brain computer interface applications.
- [9] Sarah N. Abdulader*, Ayman Atia, Mostafa_sami M. Mostafa. Brain computer interface: Applications and challenges
- [10] Melody M. Moore. Real word application for brain computer Interface technology.
- [11] School of computer engineering, kiit university, bhubaneswar India. Brain computer interface issues on hand movement

- [12] Klaus Robert Müller, Michael Tangermann. Machine learning for real time signal processing: EEG analysis from brain computer interfacing to mental state monitoring.
- [13] Jonathan R. Wolpaw* and Dennis J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain computer interface in human
- [14] Christoph Guger, Alois Schlogl, Chritra Newber. Rapid prototyping of an EEG based brain computer interface BCI
- [15] Chuanchu Wang, Kok Soon Phua, Kai Keng Ang, Cuntai Guan, Haihong Zhang, Rongeseng Lin. A feasibility study of non-Invasive Motor imagery. BCI based robotic rehabilitation for stroke patients
- [16] Fabien Lotte, Marco Congedo, Anatole Lecuyer, Fabrice Lamarche, Bruno Arnaldi. A review of classification algorithms for EEG based brain computer interface.
- [17] Luzheng Bi, Member, IEEE, Xin-An Fan, and Yili Liu, Member, IEEE. EEG-Based brain-controlled mobile robots: A survey.
- [18] Sergio Machado, Leonardo Ferreira Almada and Ramesh Naidu Annavarapu. Progress & prospective in EEG-based brain computer interface: clinical application in neurorehabilitation.
- [19] Nik Khadijah Nik Aznan, Yeon-Mo Yang. Applying Kalman filter in EEG based brain computer interface for motor imagery classification.
- [20] Rasool Ameri, Aliakbar Pouyan, Vahid Abolghasemi. EEG signal based on sparse representation in brain computer interface application.

A comparative analysis of promethee, ahp and topsis aiding in financial analysis of firm performance

Aditi Sharma¹, Gurjeet Kaur², Jatin Bansal³

¹ Bachelor of Business Administration (Financial Investment Analysis),
Shaheed Sukhdev College of Business Studies, India

² Associate Professor, Department of management studies,
Shaheed Sukhdev College of Business Studies, India

³ Bachelor of Business Administration (Financial Investment Analysis),
Shaheed Sukhdev College of Business Studies, India

¹aditi.16301@sscbs.du.ac.in, ²gurjeetkaur@sscbs.du.ac.in, ³jatin.16322@sscbs.du.ac.in,

Abstract—Multi-criteria decision making methods are extensively used in the decision making problems. Decision making is the activity of a person that helps obtain responses to the questions posed by a stakeholder in a decision process. This study compares the financial performance of seven companies in the banking sector listed on the National Stock Exchange (NSE) for years from 2014 to 2017. The PROMETHEE, AHP and TOPSIS methods are used independently to evaluate the financial performance and to decide on the best performing firm for the four year period. The evaluated ranks are then compared with the actual ranks (based on investments) using Spearman's rank correlation. The calculation of financial performance measures is based on four main criteria and nine sub-criteria. Using the following methods, our study suggests that Punjab National Bank (PNB) has the best financial performance as compared to other companies.

I. INTRODUCTION

DECISION-making is an integral part of modern management. Decision making process is continuous and indispensable component of managing any organization or business activities. Decisions are made to sustain all activities and organizational functioning. Trewatha and Newport define the process of decision making as, "Decision-making involves the selection of a course of action from among two or more possible alternatives in order to arrive at a solution for a given problem."

Decision –making is considered to be a demanding process due to its subjective nature leading decision-makers to make a choice under various risks and uncertainty. The measurement and the evaluation of the firm performance have gained tremendous importance in the past few years. Thus, it has become essential for the firms to keep a record of their financial performance. Since the financial performance of a firm is analysed using the actual data under concrete mathematical operations, these performance measures provide an insight into the severity of the problems and the precautions to be taken for the same. Since the decision-makers often have to decide based on multiple criteria under a given situation, multiple-criteria decision making analysis is often used to provide solutions under multiple and incomparable set of data. Our paper aims to use PROMETHEE, AHP and TOPSIS methods to deduce the rankings of the companies based on their financial performance under certain criteria. We also aim to compare these results with the actual rankings of the companies (based on the investments made in them) , to deduce which method among the three returns the results closest to the actual scenario. For our

study, we use the data on the performance of seven companies of the banking-sector industry, listed on the National Stock Exchange (NSE).

II. LITERATURE REVIEW

The decision-making is extensively used in every day. According to H. Koontz and H.Weihrich [3] the decision is defined as choice of the most suitable alternative with respect to the predefined criteria, while T. Hunjak defines it as collection of activities from the problem definition to the alternative selection. The PROMETHEE method helps the decision makers to find an alternative that best suits their conditions. The Analytic Hierarchy Process (AHP) is a structured technique for dealing with complex decisions. It was developed by Thomas L. Saaty in the 1970s. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method, which was originally developed by Hwang and Yoon in 1981 with further developments by Yoon in 1987, and Hwang, Lai and Liu in 1993.

Kazan and Ciftci (2013) explain that the AHP method allows researchers to decompose decision making problems into their fundamental components. Albadvi and Chaharsooghi (2007); Hengren et al. (2006); investigates the use of the AHP method in traffic planning. They suggest that an application of the AHP method is one of the possible methodologies which could be implied for the choice of technology in traffic planning. Similar to our study, there is an increasing literature investigating decision making problems using a hybrid structure (Bilsel et al. (2006); Dagdeviren (2008); Kazan and Ciftci (2013)). The use of different multi-criteria decision analysis techniques would maximize the benefits for the user who faces complicated multi-criteria decision problems.

III. METHODOLOGY

3.1. Scope Of The Study

We examine the relationship between the financial performance and the investor's preference in terms of investment, among the banking-sector companies operating in India. Our research uses data for companies listed on the National Stock Exchange (NSE) for the years 2014, 2015, 2016 and 2017.

3.2. Data And Method Of Analysis

Multi-Criteria Decision Making method requires the use of various mathematical operations. Therefore, at the solution phase, we make use of Microsoft Excel . The abstraction provides an overall analogy of the performance rankings of each firm determined by the weighted ratios. Later, we examine the stage of relationship between

the performance ranking obtained with the method determined with each ratio and the rankings published by stock exchange. Our research is conducted using 4 main criteria and 9 sub-criteria. The financial ratios are attained from CAPITALINE which is a paid subscription database.

3.3. FINANCIAL PERFORMANCE INDICATORS

The financial performance criteria used in our research are listed below:

GROWTH RATIOS

- **Asset turnover:** The asset turnover ratio measures the ability of a company to use its assets to efficiently generate sales. The higher the ratio indicates that the company is utilizing all its assets efficiently to generate sales.
- **Investment Deposit:** The total of all the long term and short term investment made by the bank on other sources like banks, share market, loans and advances divided by the total amount of deposits raised by the bank by various account
- **Earning Retention Ratio:** The retention ratio is the proportion of earnings kept back in the business as retained earnings. The retention ratio refers to the percentage of net income that is retained to grow the business, rather than being paid out as dividends.

PERFORMANCE RATIOS

- **Return on Net Worth:** the amount of net income returned as a percentage of shareholders equity. Return on equity measures a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested.
- **Return on Assets:** ROA gives an idea as to how efficient management is at using its assets to generate earnings. Calculated by dividing a company's annual earnings by its total assets.

VALUATION RATIOS

- **Price Earnings Ratio:** The price to earnings ratio (PE Ratio) is the measure of the share price relative to the annual net income earned by the firm per share.
- **Market Cap/Sales Ratio:** It is calculated by dividing the company's market cap by the revenue in the most recent year; or, equivalently, divide the per-share stock price by the per-share revenue.

LIQUIDITY RATIOS

- **Current Ratio:** The current ratio considers the current total assets of a company (both liquid and illiquid) relative to that company's current total liabilities.
- **Quick Ratio:** The quick ratio is a measure of how well a company can meet its short-term financial liabilities. Also known as the acid-test ratio, it can be calculated by: $(\text{Cash} + \text{Marketable Securities} + \text{Accounts Receivable}) / \text{Current Liabilities}$.

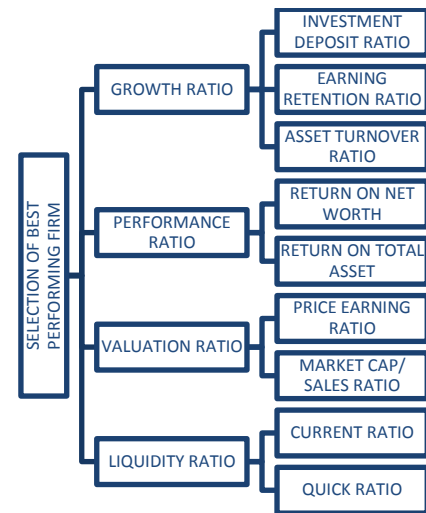


Figure 1: Complete graphical representation of MCDM problem

Banking Firm's Name	Code
State Bank of India	A
Axis Bank	B
Punjab National Bank	C
Canara Bank	D
Union Bank of India	E
Kotak Mahindra Bank	F
ICICI Bank	G

Table 1 enlists the name of banking firms trading on National Stock Exchange for the years 2014, 2015, 2016 and 2017.

Multi-criteria decision making (MCDM) methods:

- *The Analytic Hierarchy Process (AHP)*
 1st Stage: The problem is decomposed into a hierarchy of goal, criteria, sub-criteria and alternatives.
 2nd Stage: The pairwise comparisons of various criteria are organized into a square matrix.
 3rd Stage: The principal Eigen value and the corresponding normalized right eigenvector of the comparison matrix give the relative importance of the various criteria being compared.
 4th Stage: The rating of each alternative is multiplied by the weights of the sub-criteria and aggregated to get local ratings with respect to each criterion.
- *The Technique of Order Preference by Similarity to Ideal Solution (TOPSIS)*
 1st Stage: Construct normalized decision matrix.
 2nd Stage: Construct the weighted normalized decision matrix.
 3rd Stage: Determine the ideal and negative ideal solutions.
 4th Stage: Calculate the separation measures for each alternative.
 5th Stage: Calculate the relative closeness to the ideal solution
- *The Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE)*
 The PROMETHEE II Method was developed by J.P.Brans and presented for the first time in 1982 in a conference organized by R. Nadeau and M. Landry at Univesitè Laval, Québec, Canada. The PROMETHEE Method is a pairwise comparison of decision points based on assessment factors. However, unlike other multi-criteria decision making methods, it defines a preference function

consisting of different assessment factors and assigns a relative weight indicating the level of importance of each factor and the internal relations among them.

1st stage: A data matrix is prepared by scaling the values on the basis of variations in the ratios.

2nd stage: A preference function is defined for each criterion. For the sake of convenience and uniformity, we have used linear preference function.

3rd stage: The third stage starts with a comparison (binary) of decision points, i.e. scaled values given to the criteria of each company. Then, the preference function for each criterion is determined.

4th stage: The positive and negative value for each company in particular criteria is defined.

5th stage: The positive and negative rule set for alternatives.

6th stage: This stage calculates the net priority values of each criterion and orders the companies in rank of preference, thus completing PROMETHEE II.

The Spearman's Rank Correlation

1st stage: The actual ranks are extracted from NSE website on the basis of the investments in them.

2nd stage: The formula for the Spearman's rank correlation coefficient is given by the formula:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where r_s =spearman's rank correlation coefficient
 d_i =difference in the ranks of each observation
 n =number of observations

3rd stage: The above formula is used to calculate the correlation between the actual ranks and the ranks obtained by each of the above methods.

IV. ANALYSIS AND RESULT

In this abstraction, the three MCDM models are used which are the PROMETHEE, hybrid structure of the AHP and the TOPSIS and the AHP. The objective of the multi-criteria decision making problem and the estimation of relative weights assigned to each criteria are carried out by the AHP method. The PROMETHEE, the TOPSIS and the AHP methods are used for the final arrangement and ranking.

We assign higher weights to the criteria with the largest coefficient of variance. As a result, a more explicit picture of differences among the companies is anticipated. The relative ratios of coefficient of variance of each criterion are calculated. We then achieve the relative weight of each criterion using the Analytical Hierarchy Process (AHP).

Based on the data:

RATIOS	WEIGHTS
RONW	0.3464
ROA	0.1827
Current Ratio	0.1160
Quick Ratio	0.0350
Asset Turnover	0.0226
Investment Deposit	0.0293
Earning Retention	0.0290
Market Cap/Sales	0.0807
Price Earning (P/E)	0.1583

TABLE-2 depicts the weights assigned to each ratio through the AHP method. RONW has been given highest preference due to high variability in it over the past 4 years of functioning of banking firms and Asset turnover ratio has been given least preference due to absence of variability in it over the past 4 years.

The AHP method takes into account the determining the relative merit of a set of alternatives in multi-criteria decision-makers. On the other hand, the PROMETHEE method and the TOPSIS method are arrangement methods. Each method has its own strong points and weak points. The purpose of our research is to find the best choice for the decision takers as well as to access the ranking of companies based on financial performances.

The Analytic Hierarchy Process (AHP)

The relative preferences of sub-criteria are calculated by taking ratios of coefficient of variance of each company in respect with all other companies. These relative preferences are normalized through the AHP process where we divide the relative preference with the sum of relative preferences. The relative preferences for profitability, valuation, growth and liquidity are calculated by taking average of the sub-criteria's normalized preferences. Process of normalization of these relative preferences is carried out with the help of AHP. After normalization of preferences, we sum it for each company and hence create score of each company for each main criterion. (Given in Table-3, 4, 5, 6)

PROFITABILITY	SCORE
A	0.1487
B	0.1105
C	0.3629
D	0.2081
E	0.0412
F	0.1023
G	0.0264

TABLE-3 shows the score of each company in profitability criterion calculated by obtaining relative preferences of one company over all other in respective criterion. Here PUNJAB NATIONAL BANK is given highest score.

LIQUIDITY	SCORE
A	0.1477
B	0.2226
C	0.0952
D	0.0878
E	0.1413
F	0.1510
G	0.1545

TABLE-4 shows the score of each company in liquidity criterion calculated by obtaining relative preferences of one company over all other in respective criterion. Here AXIS BANK is given highest score.

GROWTH	SCORE
A	0.1188
B	0.1291
C	0.1382
D	0.1469
E	0.1198
F	0.1679
G	0.1791

TABLE-5 shows the score of each company in growth criterion calculated by obtaining relative preferences of one company over all other in respective criterion. Here ICICI BANK is given highest score.

VALUATION	SCORE
A	0.0769
B	0.2895
C	0.1997
D	0.1974
E	0.1139
F	0.0552
G	0.0673

TABLE-6 shows the score of each company in valuation criterion calculated by obtaining relative preferences of one company over all other in respective criterion. Here AXIS BANK is given highest score.

The final score for each company is calculated by taking the sum of the product of score of each criterion with the pre-calculated weights. (Given in Table-7)

	FINAL SCORE
PNB	0.265313
CANARA	0.182411
AXIS	0.171733
SBI	0.128948
ICICI	0.103694
UNION	0.08003
KOTAK	0.067872

TABLE-7 shows the final score for each company. Here PNB BANK is given highest score as it has maximum score in profitability criterion and profitability criterion has maximum weight.

These were the results obtained from the AHP method, we now head to the next technique, TOPSIS.

The Technique of Order Preference by Similarity to Ideal Solution (TOPSIS)

Weighted normalized matrix is created with seven banking firms and four main criteria through the process of TOPSIS.

Positive ideal alternative and negative ideal alternative are calculated which are:
 A+ (0.03472, 0.25410, 0.08848, 0.19904)
 A- (0.02776, 0.10281, 0.03598, 0.02117)

COMPANY	POSITIVE	NEGATIVE	RELATIVE
PNB	0.2317	0.0105	0.9566
CANARA	0.2101	0.0494	0.8097
AXIS	0.2039	0.0527	0.7946
SBI	0.1578	0.1108	0.5874
ICICI	0.1401	0.1330	0.5129
UNION	0.1179	0.1658	0.4156
KOTAK	0.0772	0.1983	0.2801

Table-8 depicts the Separation measures which are calculated for each company from both alternatives. After calculating separation measures, we calculate relative measure which is equal to Negative measure divided by sum of positive and negative measure.

These were the results obtained from the TOPSIS method, we now head to the next technique, PROMETHEE.

The PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) Method

Our study uses the concept of coefficient of variations to assign weights to the criteria. We have assigned higher weights to criteria having higher coefficient of variation. We have then normalized the weights to cumulate them to 1

CRITERION	WEIGHT
GROWTH	0.043697
LIQUIDITY	0.008821
PROFITABILITY	0.869516
VALUATION	0.077966
	1

Table 9 provides the weights which are used during PROMETHEE analysis. The weights were calculated using the concept of variation, and then normalizing the weights to return the cumulative weight of all criteria to 1.

PROMETHEE uses the concept of preference ranking on the basis of various criteria. For the sake of uniformity, we use linear preference function for all the four criteria. After analysing data, and providing Preference and Indifference threshold on qualitative basis for all the criteria, we calculate the positive and negative preference of each company for each criteria and reach the final rankings by multiplying the weights of each criteria with the preference (both positive and negative) values of a company.

$\hat{O}(+)$	
SBI	1.921473
AXIS	1.282613
ICICI	0.357229
PNB	1.756673
CANARA	1.756673
KOTAK	0.408687
UNION	0.887158

Table 10 gives the calculation of $\hat{O}(+)$ for every company. The value obtained derives how much one company is preferred to another under a specified criterion.

$\hat{O}(-)$

SBI	0.161964
AXIS	0
ICICI	2.621779
PNB	0.391498
CANARA	0.442956
KOTAK	4.360811
UNION	0.391498

Table 11 gives the calculation of $\hat{O}(-)$ for every company. The value obtained derives how much other companies are preferred to one company under a specified criterion.

	$\hat{O}(+)$	$\hat{O}(-)$	\hat{O}
SBI	1.921473	0.161964	1.759509
AXIS	1.282613	0	1.282613
ICICI	0.357229	2.621779	-2.26455
PNB	1.756673	0.391498	1.365175
CANARA	1.756673	0.442956	1.313717
KOTAK	0.408687	4.360811	-3.95212
UNION	0.887158	0.391498	0.495659

Table 12 give the net \hat{O} , i.e. $\hat{O}(+)-\hat{O}(-)$. Thus, this value derives which company is more preferred and which is less preferred. As per PROMETHEE, SBI is the most preferred company in terms of financial performance.

After computing the rankings using the three multi-criteria decision analysis techniques, we now do a comparative analysis of the obtained ranks with the actual ranks using Spearman's rank correlation.

Spearman's Rank Correlation

After analysing and calculating the ranks for the companies from the above methods, we use the concept of Spearman's Correlation Coefficient to study the level of correlation between the actual rankings obtained from the NSE Portal and each of the methods.

	PROMETHEE	AHP	TOPSIS
SBI	0	9	4
PNB	0	1	1
CANARA	1	4	4
AXIS	1	4	1
UNION	1	0	0
ICICI	9	4	4
KOTAK	0	0	0
SUM	12	22	14
6SUM	72	132	84
$6S/(N^3-N)=Y$	0.2142857	0.392857	0.25
1-Y	0.7857143	0.607143	0.75

Table 13 computes the rank correlation coefficient for the three methods, and we can thus conclude that, PROMETHEE is the most correlated to the actual rankings, followed by TOPSIS and then AHP

V. MANAGERIAL INFERENCE AND CONCLUSION

In this abstraction, we use the AHP, the PROMETHEE and the TOPSIS methods for the multi-criteria form decision making problem. Also we use hybrid form of the AHP with the TOPSIS and the PROMETHEE thereby, using the benefit from the preferable aspects (strengths) of each model and minimizing the plausible errors.

The objective of the decision making problem and the calculation of the weights allotted to each sub criteria are achieved by the AHP method as the PROMETHEE and the TOPSIS method does not yet provide any suggestion for this division of the MCDM analysis. The PROMETHEE and the TOPSIS method permit the building of an outranking between the different alternatives. As a result, in our abstraction we audit the vitality of the alternatives, how they accomplish our criteria and how they impact on the choice of the function. The hybrids of these methods provide an alternative which is more consequential and accurately fits to the company's interests and targets. Furthermore, it grants us to build a sensible arrangement of companies across four main criteria. The bank firms which are listed on the National Stock Exchange are sorted via the three MCDM models (which are the AHP, the PROMETHEE and the TOPSIS) on the basis of their financial performances. Our study indicates that **Punjab National Bank** is the best performing bank across the 4 main criteria while **Kotak Mahindra Bank** is identified as the worst performing bank. Our study also indicates that the ranks given by the three MCDM models have high correlation coefficient.

FIRMS	AHP	PROMETHEE	TOPSIS	ACTUAL
State Bank of India	4	1	3	1
Axis Bank	3	4	4	5
Punjab National Bank	1	2	1	2
Canara Bank	2	3	2	4
ICICI Bank	5	6	5	3
Kotak Mahindra Bank	7	7	7	7
Union Bank of India	6	5	6	6

Table 14 gives an overview of the ranks obtained using the three techniques and the actual rankings extracted from NSE portal.

Multi-criteria Decision making Methods	Spearman's coefficient of correlation
PROMETHEE	78.5714%
AHP	60.7143%
TOPSIS	75.0000%

Table 15 gives an overview of the Spearman's rank correlation, indicating that for the given study, PROMETHEE gives the most correlated results with the actual scenario, followed by TOPSIS and AHP.

VI. LIMITATIONS AND FURTHER SCOPE OF STUDY

The research is conducted on the banking-sector industry. Thus, there is further scope of using the techniques to evaluate decisions within other industries, as well as across industries.

Also, the given study makes the use of four main criteria with nine sub-criteria. The criteria and the techniques to evaluate the criteria can be further explored.

REFERENCES

- [1] Halim Kazan, Merve Ertok, Cihan Ciftci (2015) Application of a Hybrid Method in the Financial Analysis of Firm Performance, World Conference on Technology, Innovation and Entrepreneurship
- [2] Velasquez M and Hester P T 2013 An Analysis of Multi-Criteria Decision Making Methods, International Journal of Operations Research 10(2)56-66
- [3] F. C.-G. Pablo Aragonés-Beltrán, Juan-Pascual Pastor-Ferrando and Andrea Pla-Rubio, "An AHP (Analytic Hierarchy Process)/ANP (Analytic Network Process)-based multi-criteria decision approach for the selection of solar-thermal power plant investment projects," Energy, vol. 66, pp. 222-238, 2014.
- [4] N. M. a. M. J. H. Jahromib, "Propose a model to choose best project by AHP in distributed generation," Procedia Technology, vol. 1, pp. 481-484, 2012.
- [5] T. a. T. A. Nooshin Rahmania, "Developing a Multi Criteria Model for Stochastic IT Portfolio Selection by AHP Method," Procedia - Social and Behavioral Sciences, vol. 62, pp. 1041 – 1045, 2012
- [6] B. Oztaysi, "A decision model for information technology selection using AHP integrated TOPSIS-Grey: The case of content management systems," Knowledge-Based Systems, 2014.
- [7] Albadvi., A., Chaharsooghi, S.K., & Esfahanipour, A. (2007). Decision making in stock trading: an application of PROMETHEE. European Journal of Operational Research, 177, 673-683
- [8] Bhushan, N., Rai, K. (2004). *Strategic decision making applying the Analytic Hierarchy Process. The United States of America: Springer.*
- [9] Bilsel, R. U., Büyüközkan, G., Ruan, D. (2006). *A fuzzy preference-ranking model for a quality evaluation of hospital web sites. International Journal of Intelligent Systems, 21, 1181-1197.*
- [10] <http://www.managementstudyguide.com/decision-making.htm>
- [11] https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis
- [12] https://en.wikipedia.org/wiki/Spearman_rank_correlation_coefficient
- [13] Multiple criteria decision analysis- state of the art surveys- edited by FIGUEIRA, EHRGOTT AND GRECO

A Perspective Approach (OABC) Algorithm using Square Odd Routing for minimized Energy Consumption

Akrati Sharma¹, V. P. Sharma², Hitesh Sharma³

¹Computer Science & Engineering, Govt. Engineering College, Ajmer, India

²Computer Science & Engineering, Assistant Professor Govt. Engineering College, Ajmer, India

³Computer Science & Engineering, Govt. Engineering College, Ajmer, India

¹akratininna@gmail.com, ²vp.ecajmer@gmail.com, ³hiteshsharma0123@gmail.com.

Abstract—ABC set of principles has been already proposed furthermore with some drove guidelines, yet the length of the work parameter has been spinning round detecting the hubs in static or dynamic way with no accentuation at the power consumption parameter ideal from the crossing hubs vitality to warmth dissemination control or the entire charge for navigating or detecting every one of the hubs or the time admission for the hub records to get transmitted from one hub to each other. Our proposed work is essentially constructed generally in light of the sending of the system however now not arbitrarily but rather to some extent construct absolutely with respect to parameters. Our work proposed a totally one of a kind and favored an approach of ABC calculation for unpracticed detecting of actualities with regards to control looking after age. The proposed set of guidelines guarantees the identification of implied objective however with the ideal strategy with diminished vitality utilization.

The ABC has a few applications for choosing the optical neighboring arrangement construct absolutely for the most part with respect to swarm knowledge. Our paper concentrates on the versatile system of ABC calculation with the versatile approach of the square and strange statute so as to pick the most appropriate course of unpracticed transmission. Our papers have changed the conventional approach of the honey bee state to keep successively in the meantime as not having any unique measurements. We have proposed an unpracticed switch based absolutely convention in WSN. We have proposed the square odd technique to improve the conventional strategy in the more prominent dynamic way. The considered arrangement of principles should offer rest to trade hubs in the meantime the unmistakable hub is vigorous. Our proposed procedure will definitely decrease the quality utilization and at the indistinguishable time will bring about looking through the related most astounding decent bearing to embellish the test results fabulously.

Index Terms—ABC, WSN, bee colony, square odd

I. INTRODUCTION

WI-FI Bee Colony Optimization is a met heuristic in light of the ordinary searching practices of bumble bees. Exactly when a bumble bee adequately finds sustenance, it returns to the hive and grants the position and partition to the support source to hive mates by methods for a waggle move, which was not grasped until decoded by Karl von Frisch in 1974. The waggle move is a figure of eight move, and grants bearing as the move's point to the sun and

the length of the central region of the move is particularly comparing to the division to the sustenance. Distinctive hive mates by then have a choice, they can take after another bumble bees' turn and go to a comparable support source, where they may find more sustenance sources, or to examine aimlessly (which in nature is astoundingly phenomenal). Once the craftsman has completed its turn it can either watch another bumble bees move, or return to the sustenance source it advanced.

Bumble bee System was first exhibited in [LT01], which incited the change of Bee Colony Optimization and its application to the Traveling Salesman Problem (TSP) in [WLC08a]. The met heuristic is a swarm knowledge approach, which implies it is depicted by individuals doing excess exercises and a clear specific method between individuals, realizing iterative difference in course of action quality. Bumble bee Colony Optimization has been used to strike a grouping of issues, including TSP [WLC08b, WLC09a, and WLC09b] and the p-center issue [TDS11]. The bumble bee area headway met heuristic has four phases consistent with age, and the estimation accentuates till some circumstance is met. The game plan of methodologies has an outrageous and fast of cutting edge bumble bees, the wide range dependent upon the issue and trouble case, paying little mind to the path that among 20 and 50 is standard. The data is the main look at into memory, normally as some style of matrices and the game plan of standards starts off developed.

Each age joins the going with four phases:

Game plan creating. Each bumble bee builds up an answer the utilization of the course of action picked in arrange four of the past age (beside it is the essential new release, this is communicated later) and some burden correct resources (close by side period in tsp).

Daemon exercises. Bother particular advancements and neighborhood looks are continuing running in this phase to enhance the courses of action. The blend of optimistic and gathering are chasing down methods licenses making more grounded results.

Feature it 'waggle move'. Bumble bees put it accessible authentic surprising responses for each other to be used as a piece of later accentuations. Look at 'waggle move'. Each bumble bee assembles courses of action build totally upon an in the past created plan. In this movement, every get this sort of an answer, both by methods for the usage of the game plan is created inside the past age or with the guide of following a great game plan made with the guide of method for some other bumble bee station.

II. BACKGROUND

Honey bee settlement streamlining is a met heuristic in view of the normal rummaging practices of honey bees. In the meantime as a honey bee adequately uncovers nourishment, it comes back to the hive and imparts the position and separation to the suppers source to hive-buddies through a waggle move, which altered into never again comprehended until decoded by method for the use of Karl von Frisch in 1974. The waggle move is an observer of 8 moves and imparts course in light of the fact that the move's point of view to the sun oriented and the span of the focal period of the move is on the double relative to space to the sustenance. Other hive mates at that point have a want, they could pick to accommodate with some other honey bees' move and travel to the equivalent suppers source, in which they'll find more prominent suppers resources, or to find arbitrarily (which in nature is extremely phenomenal). As fast in light of the fact that the artist has finished its move it can each watch some unique honey bees move, or backpedal to the supper's source it promoted.

Honey bee gadget changed into first included [LT01], which prompted the improvement of honey bee state enhancement and its application to the visiting businessperson inconvenience (tsp) in [WLC08a]. The met heuristic is a swarm insight approach, which implies it is portrayed by utilizing the utilization of individuals doing monotonous developments and a simple correspondence method between individuals, bringing about an iterative change of arrangement best.

III. OVERVIEW OF OBJECT TRACKING

ABC set of rules has been beforehand proposed moreover with a couple of drove ideas, yet the length of the papers parameter has been rotating around detecting the hubs in static or dynamic path with none accentuation on the power admission parameter ideal from the navigating hubs quality to warmth scattering power or the entire rate for crossing or detecting the greater part of the hubs or the time utilization for the hub data to get transmitted from one hub to each extraordinary. Our proposed paper depends on the arrangement of the system however now not haphazardly but rather halfway construct absolutely for the most part with respect to parameters. The one's parameters are as per the following - :

The calculation needs to give an upgraded come about to directing in a dynamic way and delight in the hubs in a unique

way. The arrangement of rules must be dependable for extensive kind of hubs detected through and through with less vitality utilization. The proposed set of rules should bounty significantly less time to eat as dispose of the non-reacting system.

Case 1: The proposed work has never again indiscriminately thought about every one of the hubs coming in the gathering in evaluation to the customary procedure of (ABC) set of standards .the proposed calculation has contemplated wonderful the exchange hubs coming inside the method for steering correspondences.

Case2: The proposed papers has created power productive outcomes the measurement which winds up not considered inside the present procedure .we've consolidated the idea of tradition (ABC) set of controls by means of checking sensors in the square and strange mode to survey the vitality unpracticed surrender the final product.

Case 3: The considered approach has proposed another idea for (ABC) set of standards with a direct substitute for the customary approach for assessing and supplants quality admission. the proposed set of guidelines alluded to as upgraded ABC (OABC) set of principles, that is remarkable in its strategy from blessing system in every one of the parameters we've expressed previously.

The organizations and float along with execution and grade by review depiction of the proposed set of tenets are one of a kind as takes after. The proposed calculation is adjusted with steering and power green streamlining convention utilizing manufactured honey bee settlement (ABC) calculation,

Stage 1: The system has been instated first to get the hubs enacted a decent approach to position into impact (ABC) set of rules

Stage 2: The group can be sent first on the off chance that you need to feel the reasonable way to have the capacity to save vitality.

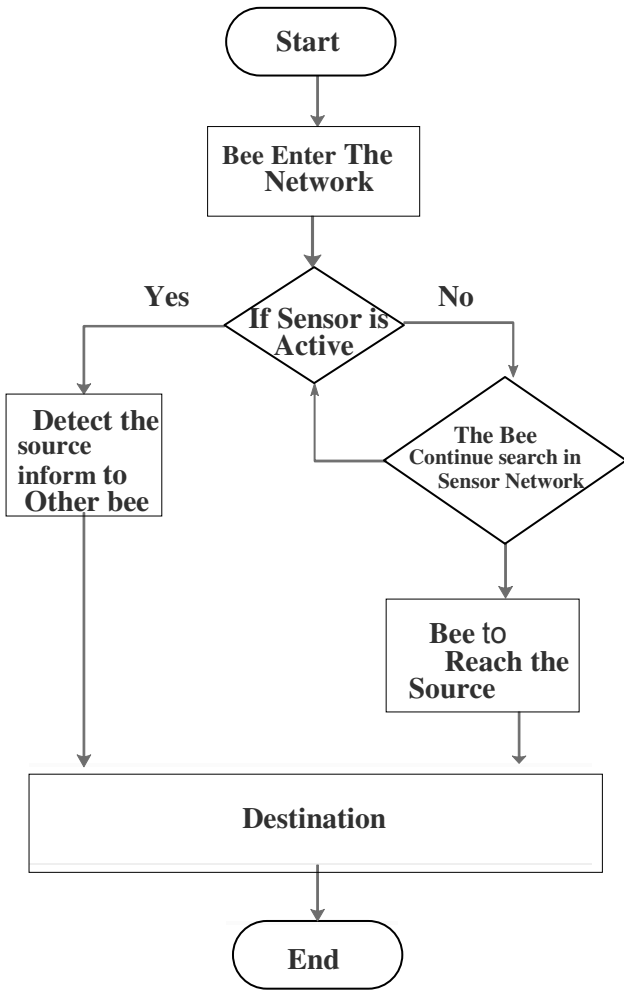
Stage 3: The accompanying advance is to check sensors in the rectangular and weird mode to assess the vitality unpracticed stop result.

Stage 4: Look at counterfeit honey bee state ABC – so on bunches to find the streamlining course among hubs to last areas.

Stage 5: The pondered strategy has proposed a current thought for (ABC) set of directions with the direct exchange in customary method for assessing and replaces control utilization.

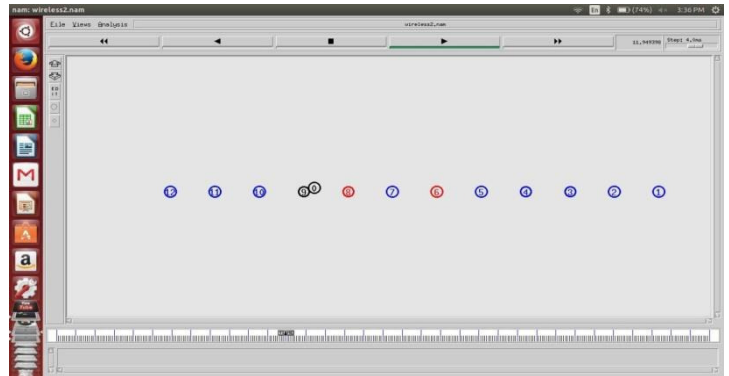
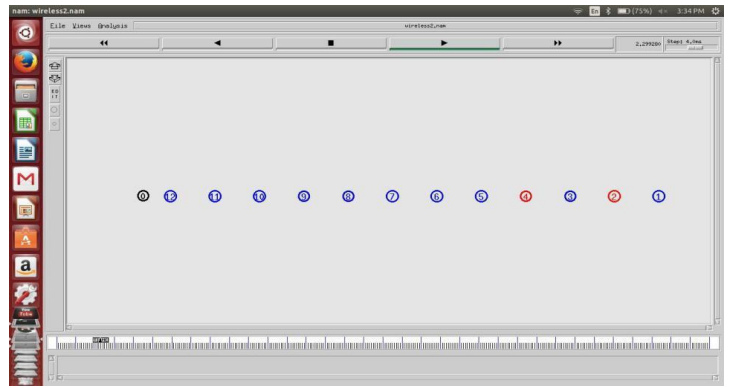
Stage6: Investigate whether all hubs wind up detected if certain at that point indicate group ways of life time and move bring down back else hold to stage three.

Flow Chart



IV. SIMULATION AND RESULTS

Simulation is a to a great degree essential present day advancement. It's far going to be refined to absolutely correct age, outlining, or opportunity programming program fields for various capacities. pc motor helped recreation will variation hypothetic and certifiable nearness contraptions or diversions practices wears on a versatile workstation basically so it will be concentrated to check however the device works. The ns-2 simulation condition offers exceptional flexibility in examines the attributes of sensor sorts out in view of it starting at now fuses versatile plans for quality confined Wi-Fi amazing frameworks. Inside the ns-2 condition, a sensor framework might be worked with a couple of unclear contraptions of traditions and affinities as those available in the around the world.



V. CONCLUSION AND FUTURE WORK

Our paper proposed a specific and fine approach of ABC set of tenets for green identifying of records concerning imperativeness securing age. The proposed estimation guarantees the revelation of expected target however with most time-tested strategy with diminished essentialness use. The paper is done up with the execution consequences of proposed Algorithm for ABC for square unusual recognizing the particular center points dynamically. The dynamic strategy for ABC set of fundamentals impacts it to appropriate for the specific guiding tradition to get related with for destiny applications. the proposed set of controls has now not a singular estimation filling in as it is proper with other controlling tradition counts and it's miles uncommonly a lousy bundle appropriate in the specific zone like surveillance machine, street framework, and interference IDs.

REFERENCES

- [1] Sheikhpour, Raziieh, and Sam Jabbehdari. "A two-level cluster based routing protocol for wireless sensor networks." *International Journal of Advanced Science and Technology*, Vol. 45, pp.19-30, August 2012.
- [2] Manikandan, A., and S. Rathinagowri. "Wireless Sensor Networks: A Research Perspective Survey." *International Journal of Technology Enhancements and Emerging Engineering Research*, VOL 1, ISSUE 3, pp. 9-11, 2013.

- [3] Kaur Lovedeep, and Rajneesh Talwar. "Review for Leach Protocol in WSN" *International Journal of Advances in Science and Technology*, Vol. 2, Issue 4, pp. 31-33, 2014.
- [4] Deepali Virmani, Tanu Sharma and Ritu Sharma, "Adaptive Energy Aware Data Aggregation Tree for Wireless Sensor Networks ". *International Journal of Hybrid Information Technology* Vol. 6, No. 1, pp. 26-36, January, 2013.
- [5] Ji, Sai, Liping Huang, and Jin Wang. "A Distributed and Energy-efficient Clustering Method for Hierarchical Wireless Sensor Networks." *International Journal of Future Generation Communication and Networking*, Vol. 6, Issue 2, pp. 83-92, 2013.
- [6] Zhang, Haitao, Shiwei Zhang, and Wenshao Bu. "A Clustering Routing Protocol for Energy Balance of Wireless Sensor Network based on Simulated Annealing and Genetic Algorithm." *International Journal of Hybrid Information Technology*, Vol.7, Issue 2, pp. 71-82, 2014.
- [7] Ma, Teng, Yun Liu, and Zhen-jiang Zhang. "An Energy-Efficient Reliable Trust-Based Data Aggregation Protocol for Wireless Sensor Networks." *International Journal of Control and Automation* Vol. 8, Issue 3, pp. 305-318, 2015.
- [8] S. Kumar, T. H. Lai, and J. Balogh., "On K-Coverage in a Mostly Sleeping Sensor Network," in *MOBICOM*. ACM, 2011.
- [9] D. Tian and N. Georganas, "A Node Scheduling Scheme for Energy Conservation in Large Wireless Sensor Networks," *Wireless Communications and Mobile Computing Journal*, May 2011.
- [10] J. Jeong, Y. Gu, T. He, and D. Du, "VISA: Virtual Scanning Algorithm for Dynamic Protection of Road Networks," in *Proc. of 28th IEEE Conference on Computer Communications (INFOCOM 09)*, Rio de Janeiro, Brazil, April 2009.
- [11] M. Mar 'oti, B. Kusy, G. Simon, and 'Akos L'edeczi, "The Flooding Time Synchronization Protocol," in *SENSYS*. Baltimore, Maryland, USA: ACM, Nov. 2004.
- [12] L. Lazos, R. Poovendran, and J. A. Ritcey, Analytic evaluation of target detection in heterogeneous wireless sensor networks," *ACM Trans. Sensor Networks*, vol. 5, no. 2, pp. 1–38, March 2009.
- [13] J. Hwang, T. He, and Y. Kim, "Exploring In-Situ Sensing Irregularity in Wireless Sensor Networks," in *SENSYS*. ACM, Nov. 2007, pp. 289–303.
- [14] Y. Gu and T. He, "Data Forwarding in Extremely Low Duty-Cycle Sensor Networks with Unreliable Communication Links," in *SENSYS*. Sydney, Australia: ACM, Nov. 2007, pp. 321–334.
- [15] J. Hwang, T. He, and Y. Kim, "Exploring In-Situ Sensing Irregularity in Wireless Sensor Networks," in *SENSYS*. ACM, Nov. 2007, pp. 289–303.
- [16] Y. Gu and T. He, "Data Forwarding in Extremely Low Duty-Cycle Sensor Networks with Unreliable Communication Links," in *SENSYS*. Sydney, Australia: ACM, Nov. 2007, pp. 321–334.
- [17] G. Lu, N. Sadagopan, B. Krishnamachari, and A. Goel, "Delay Efficient Sleep Scheduling in Wireless Sensor Networks," in *INFOCOM*. IEEE, 2005.
- [18] D. Marpe, H. Schwarz, T. Wiegand, S. Bosse, B. Bross, P. Helle, T. Hinz, H. Kirchhoffer, H. Lakshman, N. Tung; S. Oudin, M. Siekmann, K. Suhring, M. Winken, "Improved video compression technology and the emerging high efficiency video coding standard", 2011 IEEE International Conference on Consumer Electronics (ICCE 2011), pp. 52-56, 2011.
- [19] M. H. Sedky, M. Moniri, C. C. Chibelushi, "Classification of smart video surveillance systems for commercial applications", IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2005), pp. 638-643, 2005.
- [20] T. V. Dung, Q. N. Truong, "Quality Enhancement for Motion JPEG Using Temporal Redundancies", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 18 , Issue: 5, pp. 609-619, 2008.

Importance of Search Engine Marketing in the Digital World

Arokia R. Terrance, Shruti Shrivastava, Asmita Kumari

*Computer Science, Kalindi College, University of Delhi, India
{arokia.ramyat, shruti110697, masmita36}@gmail.com*

Abstract—Object tracking is one of the vital fields of computer vision that detects the moving object from a video sequence. Internet has changed the world to global village. Due to improved connectivity and increase in data usage, any new or existing products or services can reach the consumer easily through digital marketing. Apart from creating a content rich website for a product, it is highly important that the website is at the top of the Search Engine Result Pages (SERPS) of a Search Engine. The technical aspect of Search Engine Marketing Management (SEMM) of the website can be substantially improved by carrying out Search Engine Optimization (SEO) analysis of the website. The SEO tool can be utilized for increasing the website traffic and consequently increasing the sales revenue. This paper focuses on the interlinkage of Search Engine Marketing and Search Engine Optimization. The paper also provides the impact of Keyword analysis and the other SEO friendly techniques that positively affects the digital marketing.

Index Terms—Search Engine Marketing Management, Search Engine Optimization, Search Engine Result Pages, SEO tools

I. INTRODUCTION

INTERNET marketing refers to the process of promoting the products or services among the internet users through social media, search engine, banner ads on specific websites, email and app development. It helps to increase the amount and quality of sales leads of desired products or services. Search Engine Marketing (SEM) is the set of activities that involves Search Engine Optimization (SEO), Social Media Marketing (SMM) and other search engine related functions. Search Engine Marketing Management (SEMM) integrates marketing management process that promotes the position of the website's products or services towards the beginning of the search result in SERP which in turn increase its business. [5] SEM relates to the inclusion of all SEO activities but focus on return on investment (ROI), instead of relevant traffic building. SEM also integrates organic SEO and paid SEO. Social Media Marketing uses one or several social media channels to engage with customers, build relationships and then sell our products or services.

II. COMPARISON STUDY OF SEM AND SEO

SEO is defined as the process of improving visibility and improving ranking of website in search engine result page. In general, high ranked sites on SERP appears more fre-

quently in search result list and therefore receive more traffic from search engine user. This traffic or visitors can then be converted into customers. The SEO can be achieved by either organic SEO or paid SEO. Through organic SEO, website owner adjudicates to achieve top ranking for the website in the search result without using any paid means to achieve it. Through paid SEO or paid advertising using Google Adwords, website owner buys an advertising space in search engine result in order to achieve top ranking for the website in the search result and not to get more traffic. [4]

In order to understand SEO, we need to know the working of search engine that involves four major activities namely Web Crawling where the crawler search for text and hyperlinks that moves from one web page to another web-page, Build Indexing where the index about keyword and their location of the crawled webpages were created, Calculate Relevancy where the degree of relevancy between content and searcher's query were calculated, Result Retrieving where the decision upon the order of pages in the search engine result were taken. [3]

SEM is the marketing process with a goal of obtaining more visibility in search engines either by getting more free traffic or paid traffic. SEM incorporates Search Engine Optimization (SEO), which adjusts or rewrites the website content and site architecture in order to achieve higher page ranking in search engine result pages and enhance pay per click (PPC) listings. To generate best results, both organic SEO and paid SEO practices must have shared goals and combined metrics, evaluated the data to determine future strategy or find out right tools to get traffic for selected keywords in national and local search results. [1] The SEO tools can help Search Engine Marketing (SEM) through the overall improvement of landing pages by technical auditing the web pages of a website. It helps to improve code and schema, eliminates poor-quality links and highlight any other areas of potential improvement which in turn helps to increase the performance, gain more traffic and conversions to the site. The comparison study of SEO and SEM is tabulated in the below table 1.

III. KEYWORD ANALYSIS

The keyword analysis is one of the most important, valuable and high position activity in Search Engine Marketing

Table 1: Comparison of SEO and SEM

	SEO	SEM
Features	Organic SEO and Paid SEO	Integration of SEO with SMO
Focus	Keywords with high popularity	Keywords that generate high Return on Investment (ROI)
Measure	Number of website visitors	Number of successfully converted goals.
Landing pages	SEO pages are content-heavy and informative, designed to use keywords and subject relevance to rank well on any search engine.	It generally leads to direct, clean landing pages with an obvious call to action and designed to capture sales.
Prediction	It is not predictable due to uncertainty of traffic.	It is predictable as putting enough money will produce high ROI
Skill required	Fluency in SEO tools.	Require more skills than SEO.

field. It requires to know about intent of the user's query. In order to perform the keyword analysis or research, we should be aware of various keyword research SEO tools. A keyword analysis is performed by identifying the popular keywords that matches with the content of the website based on its relevance and potential conversion rate and then building the website with the targeted keywords. The keywords should be monitored and updated frequently to reflect evolving best practices of SEO techniques. It is used to find alternative terms that user's enter in search engine. Each search engines provides their own keyword suggestion SEO tools. Using the search engine information about the popular keyword, the website developer can built the website with the keywords which have highest value and the highest ROI, in turn benefit the website developer to successfully position their site in top of the SERP of the search engine. It has been found that searching for keyword with low competition and still high number of searches results in higher web traffic. [2]

IV. USAGE OF SEO TOOLS

SEMrush provides software to generate competitive research SEO tools and reports from Google SERPs. The commonly used phrase "Online Shopping" was chosen based on the keyword analytics report collected from the

Keyword Research SEO tool in the website SEMrush.com which is shown in the Figure 1. For the search query "Online shopping sites in India" in Google search engine, we obtained the www.myntra.com website in the top result of first result page, www.amazon.in website in the third result page and www.nautica.com in the last result page.

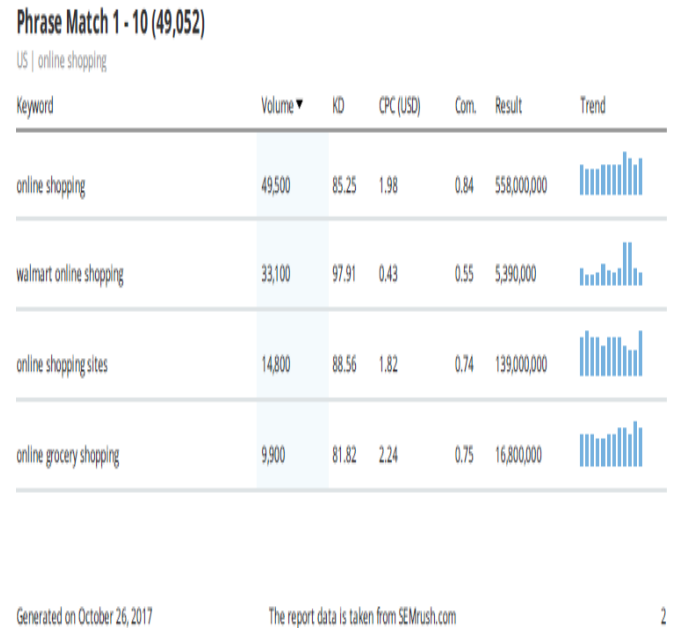


Figure 1: Keyword Analytics report from SEMrush.com

The SmallSEOTools is one of the biggest free SEO tools provider. Keyword density is the percentage of times a keyword or phrase appears on a web page compared to the total number of words on the page. Keyword density play a vital role in impacting SEO of the websites. As per Keyword density report of Mynthra in Figure 2, both "shopping" and "online" is available in the top 3 single keyword and "online shopping" is available in the first among the top 3 double keyword. As per Keyword density report of Amazon in Figure 3, "shop" is available in the top 3 single keyword and "shop online" is available twice in top 3 triple keyword. As per Keyword density report of Nautica in Figure 2, "shop" is available in the top 3 single keyword and "shop" is available in top 3 double keyword. The keyword density report helps us to relate the reason for being www.myntra.com in the top result of first result page, www.amazon.in website in the third result page and www.nautica.com in the last result page of Google search engine for the keyword phrase "Online shopping sites in India". The website should be carefully designed and coded with various SEO friendly techniques [6].

The factors that impacts the SEO of the website is depicted in the Table 2 and described in detail below:

- Length of the page title should be less than 65 characters.
- Length of the Meta description length should be less than 165 characters.

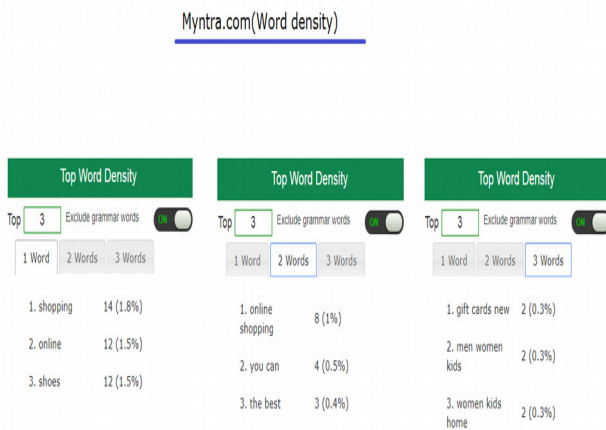


Figure 2: Keyword Density report of Mynthra from SmallSEOtools.-com

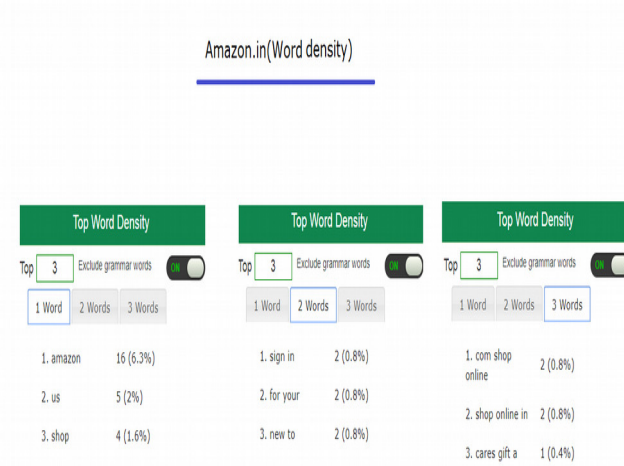


Figure 3: Keyword Density report of Amazon from SmallSEOtools.-com

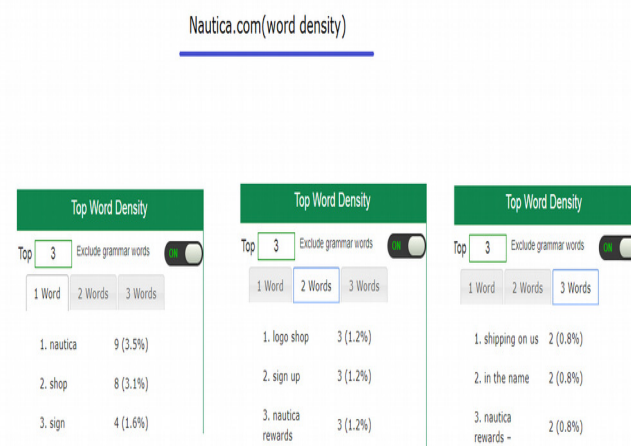


Figure 4: Keyword Density report of Nautica from SmallSEOtools.-com

- Primary keywords should appear in Meta Keywords which helps search engine to identify the website’s topic.
- Absence of H1 and H2 headings highly effects page SEO.
- Site should have SITEMAPS.
- HyperLinks should not be broken .
- There should be no underscore’s in site’s in-page URL’s.
- All image tags should have attribute ALT.
- Page should not use HTML Deprecated tags.
- Absence of noindex tag allows the webpage to be read and indexed by search engines
- Absence of nofollow tag allows the links of the webpage to be crawled by search engines.
- Web page size should be less than 33kb which will decrease the loading time of the web page.

Table 2: Website SEO Score Checker report from SmallSEOtools.com

CRITERIA	MYNTHRA	AMAZON	NAUTICA
Page Title	Yes	Yes	Yes
Meta Description	Yes	Yes	Yes
Meta Keyword	Yes	Yes	Yes
H1 Heading	No	Yes	No
H2 Heading	Yes	No	Yes
Robot.txt file	No	Yes	Yes
Sitemap	No	No	Yes
Broken Link	No	No	No
SEO Friendly Url	Yes	No	No
Img Alt attribute	No	Yes	Yes
Deprecated HTML	Yes	Yes	Yes
NoIndex tag	Yes	Yes	Yes
NoFollow tag	Yes	Yes	Yes
WWW redirection	Yes	Yes	Yes
Page Size test	Yes	No	Yes
Compression test	Yes	Yes	Yes

V. Conclusion

The website developer should be aware of various SEO techniques that gradually ensures the highly relevant content-rich website is placed in the top ranked result pages of the search engine. The comparison study of SEM and SEO help us to understand the marketing and technical aspect of the website in the digital marketing. The Keyword analysis illustrates that it plays a vital role in the prevailing SEO techniques. The Crawled texts & links, Indexed pages and the Keyword density of the website can be monitored using

advanced SEO tools and provide the SEO report on daily or weekly or monthly basis which helps to improve the traffic of the website and the overall sales of the product or services.

ACKNOWLEDGEMENTS

We would like to acknowledge this research paper to our teachers of our department and our family for their constant support and inspiration.

REFERENCES

- [1] D. M. Patil, K. M. Jadhav, M. G. Patil, "A study on Search engine and Search Engine Optimization", *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 3, Issue 6, 2015.
- [2] M. Bansal, D. Sharma, "Improving Webpage Visibility in Search Engines by Enhancing Keyword Density Using Improved On-Page Optimization Technique", *International Journal of Computer Science and Information Technologies*, Vol. 6, Issue 6, 2015.
- [3] K. Samrat, "Concept of Search Engine Optimization in Web Search Engine", *International Journal of Advanced Engineering Research and Studies*, Vol. 1, Issue 1, 2011.
- [4] S. Gupta, N. Agrawal, S. Gupta, "A Review on Search Engine Optimization: Basics", *International Journal of Hybrid Information Technology*, Vol. 9, Issue 5, 2016.
- [5] L. Zhang, J. Zhang, Y. Ju, "The Research on Search Engine Optimization Based on Six Sigma Management", *IEEE International Conference on E-Business and E-Government (ICEE)*, 2011.
- [6] A. R. Terrance, "Search Engine Optimization – A critical element in Digital Marketing", *National Seminar Proceedings of a paradigm shift towards Empowerment of Women*, 2017

Application of ASIP in Embedded Design with Optimized Clock Management

Mood Venkanna¹, Rameshwar Rao², P. Chandra Sekhar³

*Electronics and Communication Engineering,
University College of Engineering, Osmania University, India.
¹venkatmood03@gmail.com, ²rameshwar_rao@hotmail.com, ³sekharp@osmania.ac.in*

Abstract—As the demand for high performance computing increases, new approaches have to be found to automate the design of embedded processors. Simultaneously, new tools have to be developed to short the execution time consumption, and simpler design resulting in time to market. These are to be applied for the system architecture to achieve rapid exploration in on power consumption, chip area, and performance constraints. This enables interest in Application Specific Instruction Processors (ASIPs) design and application considerably. It has higher flexibility as compared to dedicated hardware. The current case study focuses on an ASIP design methodology considering the classical parameters computational performance and area as well as energy consumption simultaneously. In this paper, the clock gating is analyzed and designed. Further it is optimized using Fast genetic algorithm (FastGA). The optimization result is shown for ICORE (ISS-core) ASIP for DVB-T acquisition and tracking algorithms. Observation shows a potential of about one order of magnitude in savings of energy for optimization.

Index Terms—Embedded Processor, ASIP, Clock, Optimization, FastGA

I. INTRODUCTION

APPLICATION-specific processing elements need modern optimized embedded systems. ASIPs architecture for mixed control/data-flow oriented task has been effective for medium to low data rate. A number of methodologies have been proposed in the last two decades in this regard. ASIPs are appropriate to implement embedded systems because these offer high energy performance with high programmability. To design high performance embedded systems, ASIPs with very Long Instruction Word found suitable [1-3]. In this case, the Design Space Exploration (DSE) helps to determine the optimal parameters of the architecture. For small embedded systems, small scalar ASIPs with specific instructions need to be designed based on the characteristics of the target systems [4-5].

An embedded system is a computer system which performs a specific function according to our given application requirements with specific hardware environment. Some critical applications such as automotive design, controls de-

signs (robotic machine), railways, aircraft, aerospace, DNA Sequencing, neural network, Eye lens design and fingerprinting currently working on embedded technology. Efficient co-design technology is required to reduce the operational complexity and challenges of application designing and effective memory design is required to reduce the operational complexity of the given application [6-7]. An Embedded processor evolution mechanism is required for an increasing number of features at lower power and integrated into a single chip. The Embedded system challenge is implemented with reduction of power consumption and integration of heterogeneous systems into the single chip to reduce area, power and delay [8]. An Embedded system consists of ASICs, ASIP and field programming gate array as well as the programming unit such as the DSP and these processor designs are used in various situations or time to market [9-10].

The software environment implements application developments and compilation process and hardware units implement user logic or behavior synthesis [11]. The Hardware side of design most likely consists of interconnection components such as processors, memories and communication units (buses, output/input device I/O interfaces, sensor, RTOS devices etc.) [12-13]. Embedded systems with specific constraints need to take care of Cost, Size, power and high Performances for real time design applications. An Embedded system has few basic needs for high performance as explained below.

Cost reduction for real-time design implementation Short time span for application execution and Complexity reductive architectures Runtime-aware architectures and Deploying time-analyzable Effective resource management schemes and runtime aware environments. Effective simulation tools that allow us to make design space explorations and are used for comparisons between different hardware/software designs [14-15].

Embedded design requires a temperature-aware OS solution for real-time and high-performance systems. Effective compiler technology Applicable for high-

performance computing (HPC) and real-time embedded computing (EC) world [16].

ASIPs found to be performing better in specific application that includes servo motor control, digital signal processing (DSP), automatic control systems, cellular phones avionics, etc [19]. It maintains a balance between two extremes such as general programmable processors and ASICs by Liem et al[20]. They offer custom section availability for time critical tasks such as the real time multiply-adder or DSP and also provides the desired flexibility via the instruction-set. Complex applications require more flexibility to withstand design errors that includes specification changes at later stages. However, an ASIC is normally designed for specific behavior; hence make the design difficult to change afterwards. In this situation, the ASIPs can offer the intended flexibility as compared to the conventional programmable processors at lower cost. Among several issues pertaining to the ASIP design, this work intends to classify the approaches involved for different steps. It surveys the work done so far in the field of ASIP design and highlights the important contributions [21].

II. SYSTEM MODEL AND PROBLEM FORMULATION

For similar task, the ASIP implementations consume more power as compared to the dedicated hardware due to interconnection structure overhead and to the processor control activity. However, the processors are enough flexible and can take any software-programmable task. It creates a trade-off between low-power consumption and the flexibility. There have been several optimization options to reduce the ASIP power consumption such as the clock-gating, ISA optimization, logic netlist restructuring, instruction memory power reduction etc. In some cases dedicated coprocessor are also used.

The processor unit contains pipeline unit which is controlled by DMA circuits. There are two kinds of pipeline commonly used in processor arithmetic pipeline and instruction pipeline. An instruction pipeline uses the instruction cycles' overlapping fetch, decode, and execute phases for its operation. Currently, long instruction memory plays a dominant role in the pipeline mechanism (Fig. 3). Various Pipeline mechanisms are used by various processor developer companies such as ARM, Intel and Motorola etc. according to their performance. Instruction set architecture plays a dominant role in memory storage due to code optimization.

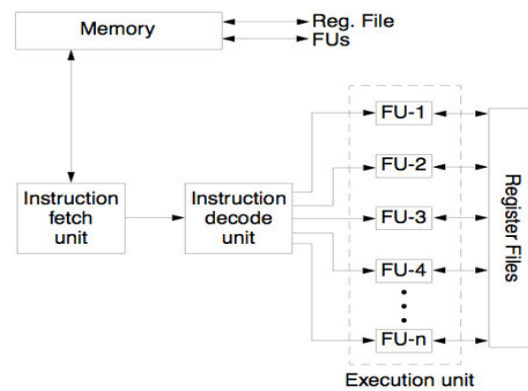


Fig. 1: Basic units for embedded system

Fig.1 shows the Basic units for embedded system. Embedded system designer used various mechanisms for processors developments have different design metrics and analyzed various processor architectures which is used in our real-time environments or System on chip such as GPP, DSP, ASIP and ASICs.

The clock gating reduces the power using the gating signals of registers in which the signals are obtained a register's execution conditions. To pipeline register power consumption in Very Long Instruction Word (VLIW) ASIP large-scale data path can be reduced by extracting the minimum execution conditions automatically during ASIP generation procedures. Results reveal a drastic reduction in VLIW ASIPs power consumption and with small clock gating overheads.

Clock gating reduces the power consumption because of following. First, it shuts off the supply to flip flops by the redundant clock when not required in calculation stages. Second, it reduces the power of clock trees in case the clock gate is placed on higher level of the tree. But, the placed gates in clock gates results in difficult clock tree synthesis due an increase in clock skew. Similarly, the operand isolation can block unwarranted signal switching of combinational logics and thus reduces power consumption although it is associated with many circuit overheads.

For automatic ASIPs generation, scalar ASIP and VLIW ASIP methods have been proposed [18]. With an ADL also known as the Micro Operation Description (MOD), these methods help design and development of ASIPs.

A. Gating Methods

Currently there are a number of automatic clock gating insertion techniques and tools are available. Among these, the Power Compiler has been very popular and commercially used tool which automatically inserts the designated gates into the registers clock lines [22]. Since, this tool is unable to extract the registers gating signals, the clock efficiency depends on the designers. The tool compels the designers to derive the gating signals from complex RTL manually for additional power reduction which is time consuming, as VLIW ASIP has hundreds of pipeline registers. This makes it unsuitable to explore the design

space. Thus, it is essential to extract the gating signals automatically.

Finite state machine based clock gating method need feedback-free pipelines to extracts the registers gating signals. Since the method is not suitable for pipeline processors hence are not applied for generation of VLIW ASIP.

B. Clock Design Mechanism for Memory Implementation

The system performance is also strongly affected by various factors besides its instruction set, the time required to move instruction & data between the CPU and memory components. The clock system is designed for implementing memory operation execution. The average cycle is designed for implementing the clock cycle required per machine instruction is a measure of computer performances. The clock signal has various characteristics such as clock period, Clock pulses, leading and trailing edge. Clock behavior depends on upon the behavior of clock elements with memory architecture with its scheduling approaches. Clock effect can be analyzed by following parameters such as Set up time, Hold time and Propagation delay time. The ASIP design with ADL has four basic functions such as the (a) architecture exploration (b) architecture implementation (c) application software design (d) system integration and verification.

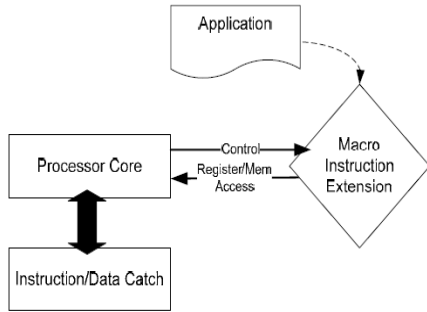


Fig. 2: ASIP Processing Technique

The throughput of a system depends on the delay time of the slowest sub-circuit which is decided by the storage register. In a larger synchronous circuit, the combinational logic unit F_i is connected to two dynamic D-FFs. The input D-FF supplies sequential, clock-synchronous data to the sub-circuit. The results of the sub-circuit are accepted by the output D-FF sequentially and clock-synchronously, as with the input data, and are then passed on. A dynamic D-FF is a simplification of the quasi-static D-FF of Fig.2. This simplification is an appropriate solution for continuously clocked MOS circuits with a clock frequency in the MHz range. A non-overlapping, two-phase clock is assumed. Such clocking is particularly safe and is often used within integrated circuits [14], [17]. The clock period must fulfil the following requirement:

$$T_{CLK} \geq 3T_{D,\Phi_2} + T_{D,F_i} + T_{D,\Phi_1} + T_{\Phi_1,\Phi_2} \quad (1)$$

Here, T_{D,Φ_2} is the delay time of a piece of data between the transmission gate clocked by Φ_2 and the input of the sub-circuit F_i . T_{D,F_i} is the delay time of the sub-circuit F_i . T_{D,Φ_1} is the delay time between the output of F_i and the input of the inverter behind the transmission gate clocked by, Φ_1 . T_{Φ_1,Φ_2} is the clock pause between Φ_1 and Φ_2 . It is to be noted that the delay times T_{D,Φ_2} and T_{D,Φ_1} depend on the characteristics of the sub-circuit. The input capacitance of F_i influences T_{D,Φ_2} and the output resistance of F_i influences T_{D,Φ_1} . The total delay resulting from the D-FFs is given by

$$T_{D,FF} = T_{D,\Phi_2} + T_{D,\Phi_1} + T_{\Phi_1,\Phi_2} \quad (2)$$

A corresponding delay can also be given for other clock systems. In single phase clock systems with edge triggered FFs, for example, the sum of the hold and set-up time must be substituted into the equation.

In the following, a simplified representation of synchronously clocked functional units will be used. Here, the D-FFs are symbolized by a simple dot in the wiring. The delay between the input and the output of the D-FF can be described by a delay operator with delay D. In case of word oriented processing the delay operator represents a register of D-FFs.

The achievable throughput RT of a system in bits per unit time is proportional to the clock rate, i.e.

$$R_T \propto \frac{1}{T_{CLK}} \quad (3)$$

On the other hand, the clock period that determines the maximal throughput is specified by the least favorable sub-circuit.

$$T_{CLK} = \max_i (T_{D,F_i} + T_{D,FF_i}) \quad (4)$$

For high throughput, small delays are essential. Modest delays can be achieved through technological measures. By shrinking the geometric structures (scaling), the capacitances can be reduced, thus achieving a reduction in the delay. The effects of such scaling are discussed in the literature [15].

Besides technological measures, circuit techniques for increasing the throughput are also possible. Various circuit structures for the implementation of elementary operations were presented in [16]. The alternatives shown demonstrate diverse delay characteristics. According to eq. 4, the maximal delay of the sub-circuits is to be minimized. This means that only the slowest module must be improved. For example, in a signal processing task using multiple additions and multiplications, only the multiplier would have to be optimized in its propagation delay. This would be pointless for the adder. Thus, architectural measures for increasing the throughput are sought with which the dominance of the slowest module can be defeated.

Power gating may be used effectively to minimize the static power consumption or leakage. It helps to

- Cut-off power supply to inactive units/components

□ Reduces leakage

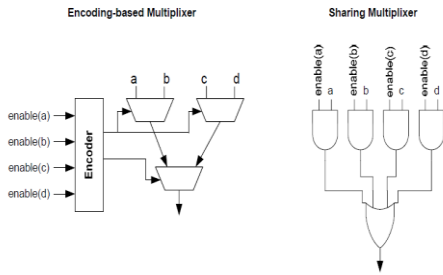


Fig. 3: Modified Clock Gating Method

III. RESULTS AND DISCUSSION

We evaluated this work on a reconfigurable processor presented in Henkel et al. [23] that we extended for execution with hard real-time guarantees. In this the optimization of power in ICORE has been incrementally achieved. This allows us to evaluate each optimization step quantitatively. Power compiler of Synopsys with back-annotated toggle activity from gate level simulations has been used for measurement. Table 1 provides both the area and the timing delay as the optimized, unoptimized, and the original description of ICORE from Infineon and ISS68HC11 from Motorola.

Table 1. Performance Comparison

Architecture Type	Area in terms of Gates	Clock Period in nsec
ISS68HC11 (unopt.)	24.52	5.82
ISS68HC11 (opt.)	19.55	5.57
Original M68HC11	15.00	5.00
ICORE (unopt.)	50.85	6.07
ICORE (opt.)	41.40	6.80
ICORE handwritten	42.00	8.00

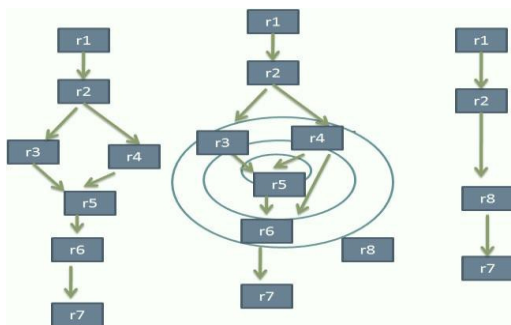


Fig. 4: Memory area implemented with resources Reducible flow mechanism and operational optimal condition

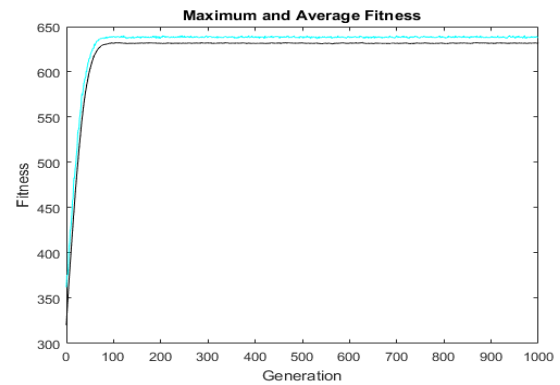


Fig. 5: Graph of fitness function using FastGA

IV. CONCLUSION AND FUTURE WORK

In this work a novel attempt is made to estimate the WCET of a program. In this process, five key steps are identified for the ASIP design. We have surveyed the research done and performed the classification of the approaches in every step during the synthesis process. The estimation of the performance is based on either the scheduler based or simulation based method. Instruction set is generated correspondingly either using the synthesisor the selection process. The code has been synthesized either using the retargetable code generator or using the custom generated compiler. However, even in case of a number of approaches used in formulation of every key steps, these methods has the limit in exploring the target space architecture. Use of integration may support chip memory hierarchies and these are not explored in an integrated way. Similarly issues related to pipelined ASIP design and pertaining to low power ASIP design has not been matured yet. It has been concluded that the processor synthesis problems and retargetable code generation have been used for isolation process.

REFERENCES

- [1] R. White, F. Muller, C. Healy, D. Whalley, and M. Harmon, "Timing Analysis for Data Caches and Set-Associative Caches," in *Proc. 3rd IEEE Real-Time Technology and Applications Symposium (RTAS'97)*, Jun 1997, pp. 192–202.
- [2] J. Engblom and A. Ermedahl, "Pipeline Timing Analysis Using a Trace-Driven Simulator," in *Proc. 6th International Conference on Real-Time Computing Systems and Applications (RTCSA'99)*. IEEE Computer Society Press, Dec 1999.
- [3] M. Arnold and H. Corporaal, "Designing domain-specific processors," In *Proc. Codesign Symposium 2001*.
- [4] A. Alomary et al., "PEAS-I: A hardware/software co-design system for ASIPs," In *Proc. EURO-DAC 1993*.
- [5] J. Van Praet et al., "Instruction set definition and instruction selection for ASIPs," In *Proc. HLS Symposium 1994*.
- [6] N. Clark, H. Zhong and S. Mahlke, "Processor Acceleration Through Automated Instruction Set Customization". In *Proceedings of the 36th annual IEEE/ACM International Symposium on Microarchitecture (MICRO 36)*, 2003.
- [7] R. Klemm, J. P. Sabugo, H. Ahlendorf and G. Fettweis, "Using LISATek for the Design of an ASIP core including Floating Point Operations", Technical report, 2008.
- [8] R. R. Hoare et al., "Rapid VLIW Processor Customization for Signal Processing Applications Using Combinational Hardware Functions". *EURASIP Journal on Applied Signal Processing*, vol. 2006 ID 46472, 2010.

- [9] F. Thili and A. Ghorbel, "ASIP Solution for Implementation of H.264 Multi Resolution Motion Estimation". In the International Journal of Communications, Network and System Sciences, Vol.3 No.5, May 2010.
- [10] G. Kappen, L. Kurz, O. Priebe and T. G. Noll, "Design Space Exploration for an ASIP/Co-Processor Architecture used in GNSS Receivers". Journal of Signal Processing Systems, vol. 58 (1), pp. 41-51, 2010.
- [11] Christoph Steiger, Herbert Walder, Marco Platzner, and Lothar Thiele. 2003. Online scheduling and placement of real-time tasks to partially reconfigurable devices. In Proc. of Real-Time Syst. Symp. IEEE, 224-225.
- [12] Pan Yu and Tulika Mitra. 2004. Scalable custom instructions identification for instruction-set extensible processors. In Proc. of Int. Conf. on Compilers, Architecture and Synthesis for Embed. Syst. ACM, 69-78.
- [13] Pan Yu and Tulika Mitra. 2005. Satisfying real-time constraints with custom instructions. In Proc. Int. Conf. on Hardware/Software Codesign and System Synthesis. IEEE, 166-171.
- [14] C. Mead, L. Conway: *Introduction to VLSI Systems*, Addison-Wesley, 1980.
- [15] N. Weste, K. Eshraghain: *Principles of CMOS VLSI Design*, Addison-Wesley, 1985.
- [16] L. A. Glasser D. W. Dobberpuhl: *The Design and Analysis of VLSI Circuits*, Addison- Wesley, 1985.
- [17] Sato, J.; Imai, M.; Hakata, T.; Alomary, A.Y.; Hikichi, N. : "An integrated design environment for application specific integrated processor.", Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers and Processors 1991, ICCD '91, 14-16 Oct. 1991, Pages: 414-417.
- [18] R. R. Hoare et al., "Rapid VLIW Processor Customization for Signal Processing Applications Using Combinational Hardware Functions". EURASIP Journal on Applied Signal Processing, vol. 2006 ID 46472, 2010.
- [19] Sato, J.; Imai, M.; Hakata, T.; Alomary, A.Y.; Hikichi, N. : "An integrated design environment for application specific integrated processor.", Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers and Processors 1991, ICCD '91, 14-16 Oct. 1991, Pages: 414-417.
- [20] Liem, C.; May, T.; Paulin, P. : "Instruction-set matching and selection for DSP and ASIP code generation.", Proceedings of the European Design and Test Conference, 1994. EDAC, The European Conference on Design Automation. ETC European Test Conference. EUROASIC, 28 Feb.-3 March 1994, Pages: 31-37.
- [21] Praet J. V.; Goossens, G.; Lanneer, D.; De Man, H.: "Instruction set definition and instruction selection for ASIPs.", Proceedings of the Seventh International Symposium on High-Level Synthesis 1994, 18-20 May 1995, Pages: 11-16.
- [22] L. Benini, G.D. Micheli, E. Macii, M. Poncino, and R. Scarsi, "Symbolic synthesis of clock-gating logic for power optimization of synchronous controllers," ACM Trans. Des. Autom. Electron. Syst., vol.4, no.4, pp.351-375, 1999.
- [23] Jorg Henkel, Lars Bauer, Michael Hubner and Artjom Grudnitsky, 2011. I-core: A run-time adaptive processor for embedded multi-core systems. In Proc. Int. Conf. on Engineering of Reconfig. Syst. And Algorithms.

Author Index

- A**ggarwal, Anu G. 3, 13, 21, 77
Aggarwal, Rashmi 71
Agrawal, Rashmi 137
Anand, Sameer 3
- B**ansal, Anuja 29
Bansal, Jatin 145
Bhargava, Neeraj 29, 35
Bhudhiraja, Chinab 121
Bijoy, Kumar 39
- C**haudhary, Bhawna 47
- G**andhi, Neha 53
Gondwal, Neha 53
Gupta, Anamika 65
Gupta, Pankaj 59
Gupta, Sugandha 65
- J**ain, Aakanksha 29, 35
- K**alra, Vaishali 71
Kaur, Gurjeet 77, 145
Kedia, Aman 77
Kumar, Abhishek 29, 35, 125
Kumari, Asmita 155
Kumar, Naveen 89
Kumar, Pramod 85
Kumar, P. Srinivas 125
- L**e, Dac-Nhuong 35
- M**ohanty, Mihir Narayan 93, 97, 107
Mohanty, Mohan Debarchan 93
Mohapatra, Saumendra Kumar 97
- P**alo, Hemanta Kumar 97
Patharia, Jyoti 101
Perichappan, Kumar Attangudi Perichiappan 125
- R**ajalakshmi, R. 113
Ram, Rashmirekha 107
Rani, Sonia 121
Rao, Rameshwar 159
Rathi, Rakesh 101
Rathore, Pramod Singh 29, 89
Rustagi, Anamika 121
- S**agar, Bharat Bhushan 59
Sairam, T. V. M. 113
Sambyal, Pankaj 121
Sasubilli, Gopinadh 131
Sasubilli, Sreenivas 125
Sekhar, P. Chandra 159
Sekhar, Uday Shankar 131
Sharma, Aditi 145
Sharma, Akash 13
Sharma, Akрати 151
Sharma, Bhawna 121
Sharma, Geeta 137
Sharma, Himanshu 21
Sharma, Hitesh 151
Sharma, Ms. Surbhi 131
Sharma, Ms. Swati 131
Sharma, Neha 137
Sharma, Surbhi 85
Sharma, V. P. 151
Shrivastava, Shruti 155
Singh, Sheetal 47
Singh, Tanya 137
- T**andon, Abhishek 21, 53
Terrance, Arokia R. 155
Triwedi, Prakarti 85, 89
- V**enkanna, Mood 159
Verma, Vibha 3