ISSN : 2165-4069(Online)
ISSN : 2165-4050(Print)

# IJARAI

International Journal of
Advanced Research in Artificial Intelligence

Volume 5 Issue 3

www.ijarai.thesai.org

A Publication of
The Science and Information Organization

# Editorial Preface

*From the Desk of Managing Editor…*

Artificial Intelligence is hardly a new idea. Human likenesses, with the ability to act as human, dates back to Geek mythology with Pygmalion's ivory statue or the bronze robot of Hephaestus. However, with innovations in the technological world, AI is undergoing a renaissance that is giving way to new channels of creativity.

The study and pursuit of creating artificial intelligence is more than designing a system that can beat grand masters at chess or win endless rounds of Jeopardy!. Instead, the journey of discovery has more real-life applications than could be expected. While it may seem like it is out of a science fiction novel, work in the field of AI can be used to perfect face recognition software or be used to design a fully functioning neural network.

At the International Journal of Advanced Research in Artificial Intelligence, we strive to disseminate proposals for new ways of looking at problems related to AI. This includes being able to provide demonstrations of effectiveness in this field. We also look for papers that have real-life applications complete with descriptions of scenarios, solutions, and in-depth evaluations of the techniques being utilized.

Our mission is to be one of the most respected publications in the field and engage in the ubiquitous spread of knowledge with effectiveness to a wide audience. It is why all of articles are open access and available view at any time.

IJARAI strives to include articles of both research and innovative applications of AI from all over the world. It is our goal to bring together researchers, professors, and students to share ideas, problems, and solution relating to artificial intelligence and application with its convergence strategies. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations.

We hope that this journal will inspire and educate. For those who may be enticed to submit papers, thank you for sharing your wisdom.

**Editor-in-Chief**
**IJARAI**
**Volume 5 Issue 3 March 2016**
**ISSN: 2165-4069(Online)**
**ISSN: 2165-4050(Print)**

# Editorial Board

# Reviewer Board Members

# CONTENTS

# Creation of a Remote Sensing Portal for Practical Use Dedicated to Local Goverments in Kyushu, Japan

Kohei Arai [1]
[1] Graduate School of Science and Engineering
Saga University
Saga City, Japan

Masaya Nakashima [1]
[1] Department of Information Science
Saga University
Saga City, Japan

***Abstract*—Remote sensing portal site for practical uses which is dedicated to local governments is created. Key components of the site are (1) links to data providers, (2) links to the data analysis software tools, (3) examples of actual uses of the satellite remote sensing data in particular for local governments. Users' demands for remote sensing satellite data are investigated for the local governments situated in Kyushu, Japan. According to the users' demands, the remote sensing portal site is created with the aforementioned key components. For the examples of remote sensing data applications, creation of land use maps, disaster mitigations, forest maps, vegetation index map for evaluation of vitality of agricultural fields and forests, etc. are taken into account. In particular for forest map creation, it is created with free open source software: FOSS of classifiers together with open data API derived training samples applied to Landsat-8 OLI data. On the other hand, volcanic eruption is featured for disaster relief with 3D representation by using open data derived DEM data. In accordance with the users' evaluation reports, it is found that the proposed portal site is useful.**

***Keywords—Remote Sensing; Satellite data; Land use map creation; Disaster prevention; Normalized Difference Vegetation Index: NDVI; Forest map; Volacanic eruption; 3D representation with Digital Elevation Model: DEM; Open data API***

## I. INTRODUCTION

There are many remote sensing portals in the world. Most of them are dedicated to scientific research based data driven portal sites. The most famous and useful site among them is Geoss portal which allows access to the data servers through the portal. On the other hand, most of space agencies create their own remote sensing satellite data servers. Meanwhile, Free Open Source Software: FOSS of data analysis software tools is also available through the different software providers. It seems that there is no portal site which allows remote sensing satellite data access, FOSS of software access, examples of data analysis and instructions of data analysis procedure with government providing open data in particular for practical use of the remote sensing satellite data dedicated to local governments. For instance, federal and state governments provide open data together with open data API. By using such government providing open data, practical use of remote sensing satellite data can be done easily.

There are many previously proposed methods for information and image retrievals in particular for remote sensing satellite data [1]-[20]. This paper is intended to provide the aforementioned purpose of remote sensing portal site for practical uses which is dedicated to local governments is created. Key components of the site are (1) links to data providers, (2) links to the data analysis software tools, (3) examples of actual uses of the satellite remote sensing data in particular for local governments. Users' demands for remote sensing satellite data are investigated for the local governments situated in Kyushu, Japan. According to the users' demands, the remote sensing portal site is created with the aforementioned key components. For the examples of remote sensing data applications, creation of land use maps, disaster mitigations, forest maps, vegetation index map for evaluation of vitality of agricultural fields and forests, etc. are taken into account. In particular for forest map creation, it is created with FOSS of classifiers together with open data API derived training samples applied to Landsat-8 OLI data. On the other hand, volcanic eruption is featured for disaster relief with 3D representation by using open data derived DEM data. In accordance with the users' evaluation reports, it is found that the proposed portal site is useful.

This paper is organized as (1) investigation of users' demands on practical uses of remote sensing satellite data in particular for the local governments which are situated in Kyushu, Japan, (2) design of the proposed remote sensing portal site, (3) examples of practical uses of remote sensing satellite data utilizing open data which are provided by the government, and (4) concluding remarks together with some discussions.

## II. PROPOSED PORTAL SITE FOR PRACTICAL USES OF REMOTE SENSING SATELLITE DATA

### A. Hearing of the Users' Demands on Practical Uses of Remote Sensing Satellite Data from the Local Governments in Kyushu, Japan

Investigation is conducted for clarifying users' demands on practical uses of remote sensing satellite data from the local governments in Kyushu, Japan. The followings are users' demands,

(1)Agricultural and forestry applications
Rice crop quality map
High quality of tealeaves map
Total nitrogen content in agricultural fields
Agricultural productivity map
Forest inventory map
Forest type and age estimations
Bamboo forest map
(2)Water resources and quality monitoring

Run-off water resource map
Water resource management
(3)Disaster
Disaster mitigation
Hazard map
Volcanic monitoring
Tsunami prediction
Illegal disposal findings
(4)Atmospheric environment
Air pollution map
Solar irradiance estimation
(5)Ocean monitoring
Ocean monitoring
Renewable resources monitoring
Nutrient rich water map
Red tide
River and coastal area monitoring and planning

The largest users' demands are agricultural and forest monitoring followed by disaster monitoring. The reason for this is the fact that the largest industry in Kyushu is agriculture and forest resources. Ocean monitoring and water resources as well as atmospheric environment are followed by. The Kyushu is surrounded by the ocean. Therefore, fishery is a major industry in Kyushu. In Kyushu, there are so many active volcanoes. It is so frequently that typhoon hit Kyushu. This is because the disaster related users' demands are dominated. Other than these, there is strong demand on solar energy monitoring for solar power plantation of electricity provides. Also, air pollution comes from the Asian continent. Therefore, air pollutions including PM2.5 is major concern for Kyushu.

### *B. Design Concept*

There are four major key components for the proposed portal site. Those are as follows,

*1) Links to the major remote sensing satellite data providers*
*2) Links to the major sites for data analysis software providers*
*3) Links to the major sites for open data which are applicable to remote sensing data analysis*
*4) Examples of practical uses of remote sensing satellite data.*

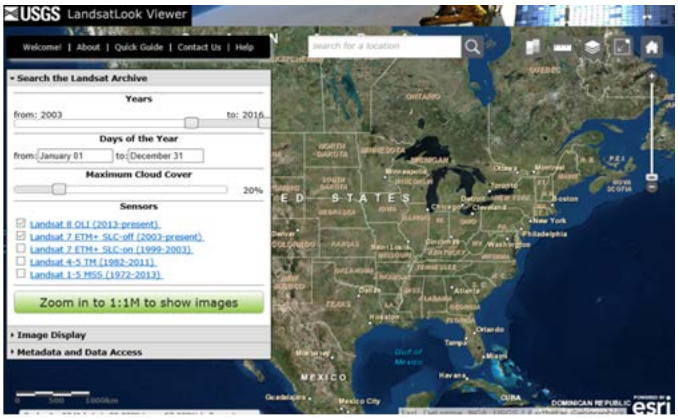
Namely, the proposed portal site is intended to provide the aforementioned four major links and information, data providers, software providers, open data providers, and examples of practical uses of the remote sensing satellite data for local governments.

### *C. The Links to the Remote Sensing Satellite Data Providers*

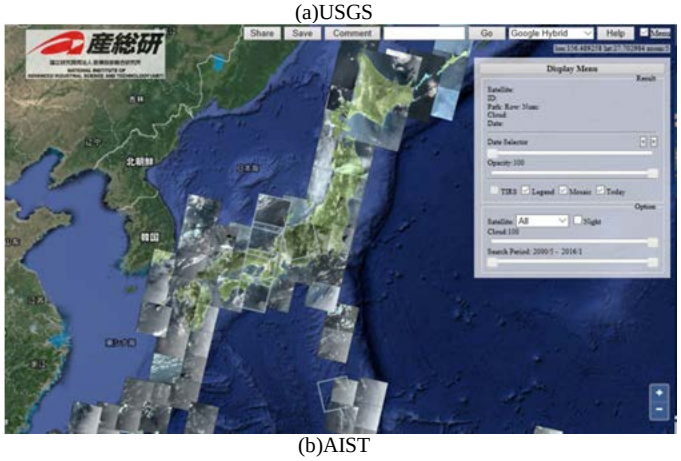
There are many remote sensing satellite data providers. Local governments need solar reflectance wavelength region of the surface reflectance channels of data with a high spatial resolution (higher than 30 m) with free of charge (downloadable from their sites freely). Therefore, Landsat ETM+, TM, OLI, ASTER/VNIR, ASTER/SWIR are candidates. Such these users' requirements are matched to the following sites,

*1) Unite State Geological Survey: USGS*[1]
*2) National Institute of Advanced Industrial Science and Technology: AIST*[2]
*3) Libra*[3]
*4) Reverb | ECHO*[4]
*5) JAXA G-Portal*[5]

Screen shot images of the data providers are shown in Fig.1. Fig.1 (a) shows USGS site followed by AIST of Landsat viewer in Fig.1 (b). In particular for AIST site, there is a comprehensive map utilizing retrieval site as shown in Fig.1 (c). Fig.1 (d) shows Libra site provided by Libra development seed organization. NASA/EOSDIS provides the Reverb | ECHO as shown in Fig.1 (e). Meanwhile, Fig.1 (f) shows JAXA G-Portal which allows remote sensing satellite data by application fields and by mission instruments.

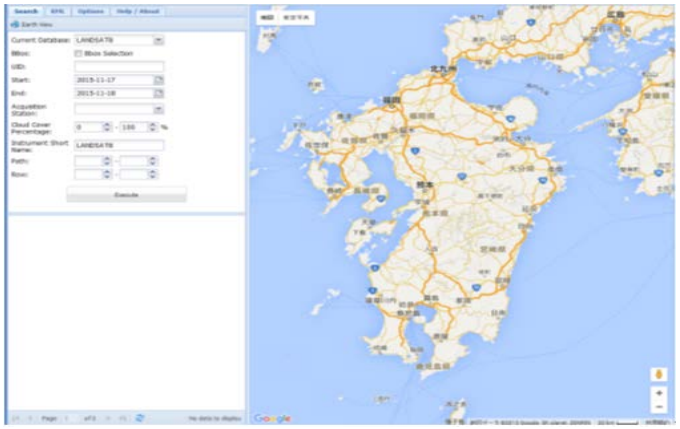


(a)USGS

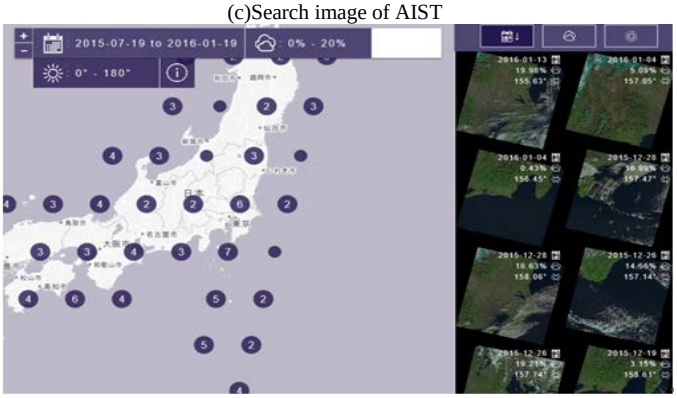


(b)AIST

[1] http://landsatlook.usgs.gov/
[2] http://landbrowser.geogrid.org/landbrowser/index.html
[3] http://libra.developmentseed.org/
[4] http://reverb.echo.nasa.gov/reverb/#utf8=%E2%9C%93&spatial_map=satellite&spatial_type=rectangle
[5] https://www.gportal.jaxa.jp/gp/date-and-area.html?F1108489815937N0QOGR=_



(c)Search image of AIST



(d)Libra

(e)NASA/EOSDIS

(f)JAXA G-Portal

Fig. 1. Examples of the remote sensing satellite data retrieval sites

## D. The Links to the Open Data Providers

Open data[6] is available and is useful for remote sensing satellite data analysis. For instance, training samples for land uses of the open data is useful for land use map creations. Open data initiative of Japan[7] is launched in 2012. The home page of the open data site in Japan is as shown in Fig.2 (a). Through the home page, the open data provided by ministries and local governments are accessible as shown in Fig.2 (b).


(a)Home page


(b)List of the available open data

Fig. 2. Open data initiative in Japan (Home page and the list of available data of the open data)


(a)Local governmental open data




(b)Geographical spatial information of open data

Fig. 3. Screen shots of local governmental open data and geographical statial information

[6] https://en.wikipedia.org/wiki/Open_data
[7] http://www.data.go.jp/?lang=english

A portion of available open data provided by the local governments in Japan is shown in Fig.3 (a) while the home page of the clearing house of geographical spatial information is shown in Fig.3 (b). Links to the local government open data sites of URLs are available from the site of Fig.3 (a) while JMP2.0 based metadata search is available for the clearing house.

### *E. The Links to the Software Providers*

Most of local governments prefer Free Open Source Software: FOSS of analysis software tools rather than commercially available software. MultiSpec[8] is one of those of FOSS. Meanwhile, RSP [9] is sophisticated image analysis software dedicated to the remote sensing satellite data analysis which is provided by the Aoyama construction company limited. On the other hand, QGIS[10] is sophisticated software which is developed by QGIS development team. One of the specific features of QGIS is available programming languages, C++, Python[11], Qt[12]. Therefore, it is relatively easy to utilize the software with the uses' developed software. Furthermore, QGIS is cross platform FOSS which allows refers, edit, and analyze the imagery data on GIS (Geographical Information System).

Common functionalities of these software tools are as follows,

*1) File manipulations including format conversion, image portion extraction, pan-sharpening, color composite, etc.*

*2) Geometric corrections including Affine transformation, pseudo Affine transformation, etc.*

*3) Filtering processing which includes mask processing, median filter, edge extraction, etc.*

*4) Image operation processing including add, subtract, multiply, division, NDVI calculation, etc.*

*5) Analysis including correlation analysis, principal component analysis, etc.*

*6) Color information manipulations which include enhancing, binarization, histogram manipulations, pseudo color representation, multi-level slicing, etc.*

*7) Image display which includes enlargement, shrinking, etc.*

*8) Image classification including maximum likelihood classification, clustering, etc.*

*9) Geographical analysis including Digital Elevation Model: DEM representation, DEM editing, slope elevation and azimuth angle calculation, etc.*

### *F. Examples of Practical Uses of Remote Sensing Satellite Data*

Major concerns of local governments have to be referred through the proposed portal site. In the site, (1) NDVI calculation for forest vitality monitoring, (2) forest inventory map creation which includes forest type classification, (3) sediment disaster due to volcanic eruptions, etc. are referred.

*1) NDVI calculation for forest vitality monitoring*

NDVI is expressed as follows,

$$NDVI=(NIR-Red)/(NIR+Red) \quad (1)$$

where NIR, Red denote leaf surface reflectance at Near Infrared wavelength region (more than 700nm) and that at red color wavelength (around 600nm), respectively. NDVI represents vitality of the tree and or the forest. In Kyushu, most of prefectural local governments concern forest vitality. NDVI has positive correlation to total nitrogen content in the leaves and has negative correlation to fiber content in the leaves. Total nitrogen content is highly correlated to tree or forest vitality while fiber content is highly correlated to age of the leaves. Therefore, total vitality of tree or forest can be estimated with NDVI. By using the correlations, quality of agricultural products, in particular, tea trees vitality and quality of tealeaves can be estimated.

Fig.4 shows an example of the NDVI estimated with Terra/ASTER/VNIR data which are acquired in the winter seasons. In the winter season, tea trees are patients for the next coming sprinter season maintaining their vitality. In such season (approximately five months), a fine condition of remote sensing satellite data can be acquired.Fig.5 (a) to (e) shows NDVI images in 2011 to 2015 which are estimated with the VNIR image of Ureshino district, Saga, Kyushu, Japan in where many tea farm areas are situated. Thus quality of new fresh tealeaves can be estimated. These NDVI images can be created with image operations and manipulations using RSP, QGIS and the other remote sensing imagery data processing software.

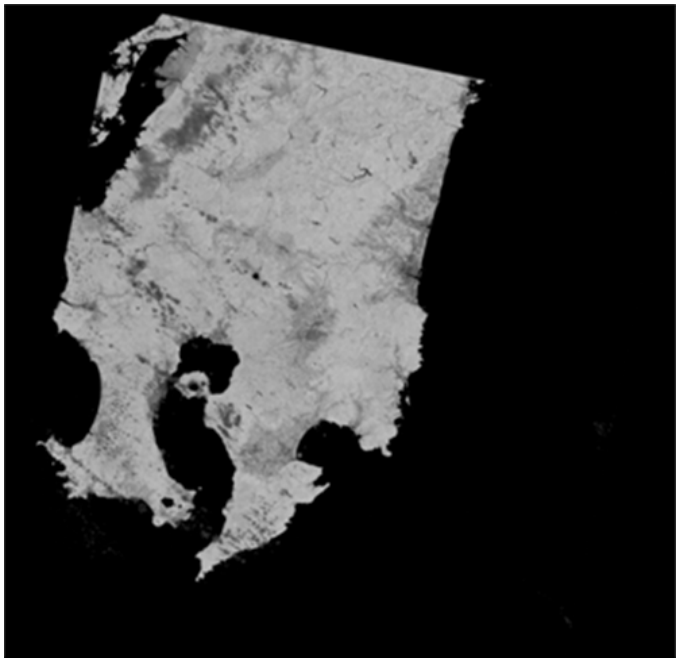


Fig. 4. Example of NDVI image of Southern Kyushu, Japan

[8] http://itcweb.cc.affrc.go.jp/affrit/sidab-info/satellite/multispec
[9] http://rs.aoyaman.com/soft/item.html
[10] http://www.qgis.org/ja/site/forusers/download.html
[11] https://www.python.org/
[12] https://ja.wikipedia.org/wiki/Qt

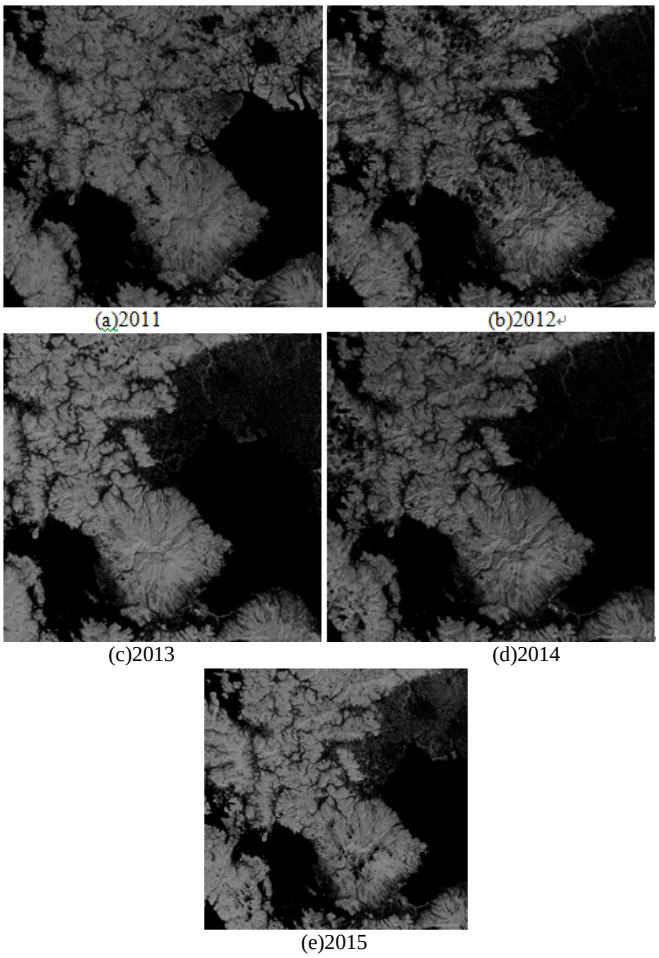



Fig. 5. NDVI images of Ureshino city Saga Japan in the winter seasons of 2011 to 2015

*2) Forest inventory map creation which includes forest type classification*

Landsat-8/OLI imagery data which is acquired on May 2 2015 can be retrieved and download through AIST site. One of the examples of false color representation of image is shown in Fig.6 (a). Meanwhile, classified image is easily created with RSP software (class#1: Ocean, class#2: River water, class#3: vegetated areas, class#4: Bare soil, class#5: Urbanized areas) as shown in Fig.6 (b). It is possible to extract training samples for classification through referring Open Data of previously classified image with Open Data API.

Forest type classification is also available with previously created open data of forest map (Fig.7 (a)) for forest inventory provided by prefectural local government which is downloaded through Open Data portal with Open Data API. Example of the classified result is shown in Fig.7 (b) with the legends as shown in Fig.7 (c). By using DEM data provided by geological survey of Japan through Open Data portal, bird view image of classification result can be created easily as shown in Fig.7 (d). Therefore, classified results can be represented from different aspects with QGIS software.

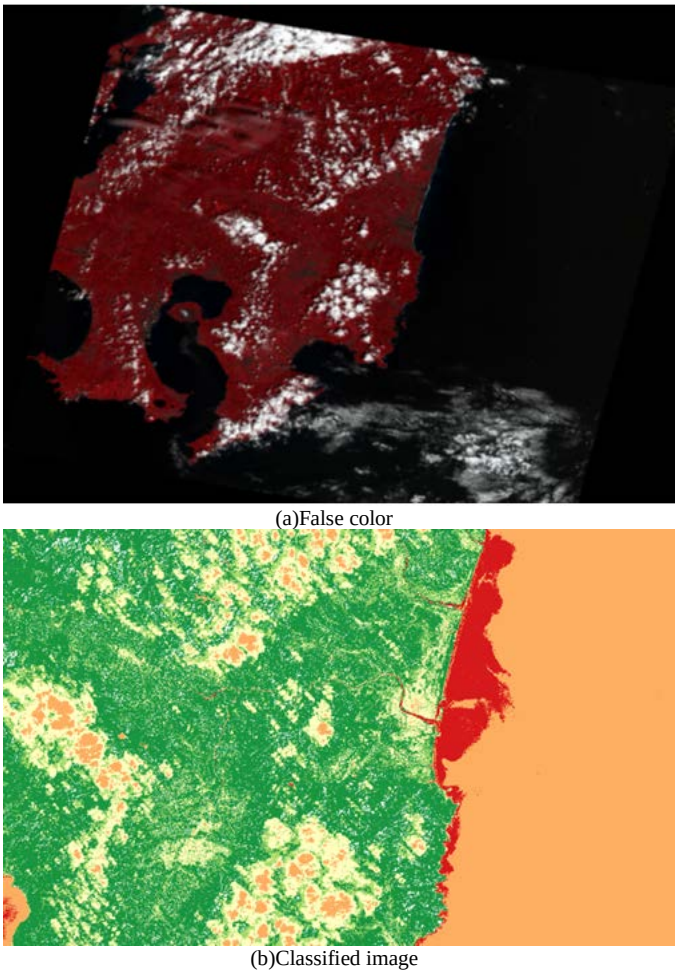
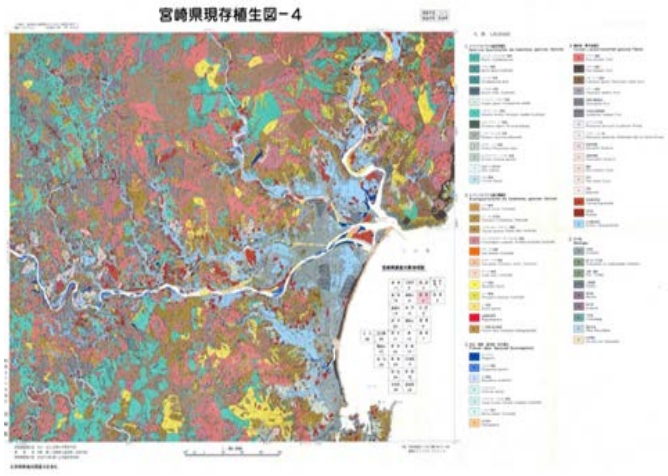
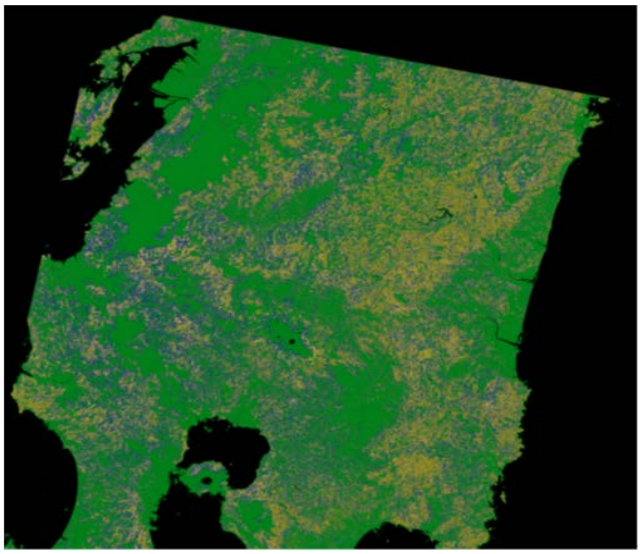


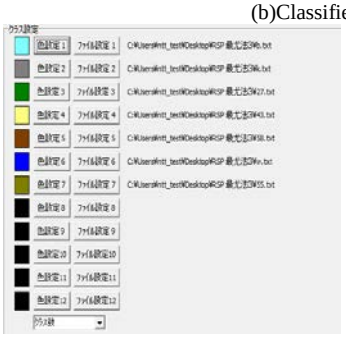

Fig. 6. Example of downloaded image of Landsat-8/OLI and classified image



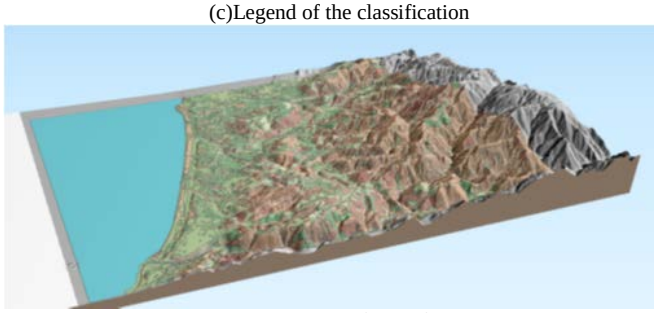

(a)Forest Map



(b)Classified image



(c)Legend of the classification

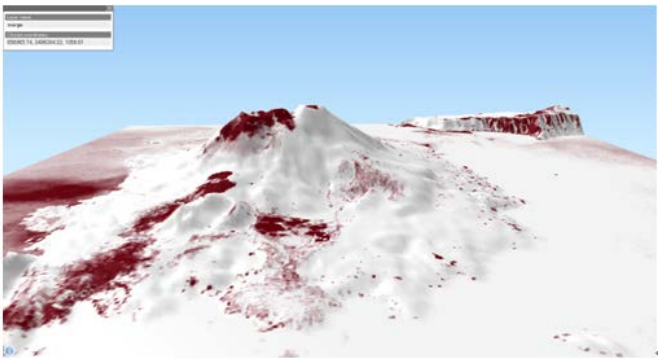


(d)3D representation of classified image

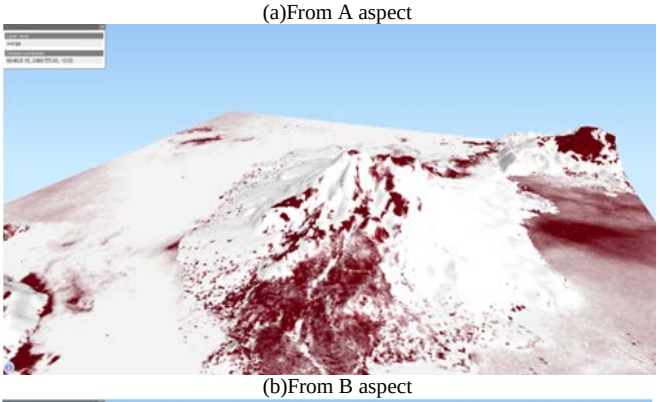
Fig. 7. Forest inventory application of remote sensing satellite data

*3) Sediment disaster due to volcanic eruptions*

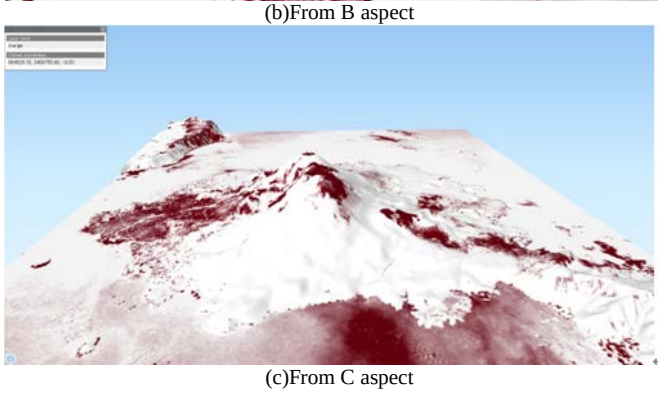
Land slide (sediment disaster) due to volcanic eruption can be detected by extracting land cover changes from vegetated areas to non-vegetated areas (bare soil) form the two different remote sensing satellite imagery data which are acquired before and after the eruption. Active volcano of Sakurajima Mountain, Kagoshima, Kyushu, Japan is erupted during August 9 and September 10 2015. On August 15, caution level is raised from 3 to 4 and then that is dropped from 4 to 3 on September 1. By comparing two Landsat-8/OLI images which are acquired on August 9 and September 10, land slide areas are detected. Using QGIS, land slide areas which are colored in red can be detected and represented as shown in Fig.8 (a) to (c) from the different aspects. Through comparison between analyzed land slide areas and local government provided land slide data (truth data), it is found that they show a good coincident.



(a)From A aspect



(b)From B aspect



(c)From C aspect

Fig. 8. Land slide and/or sediment disaster areas (red color portion) of Sakurajima Mountain in August 2015

## III. CONCLUSION

Remote sensing portal site for practical uses which is dedicated to local governments is created. Key components of the site are (1) links to data providers, (2) links to the data analysis software tools, (3) examples of actual uses of the satellite remote sensing data in particular for local governments. Users' demands for remote sensing satellite data are investigated for the local governments situated in Kyushu, Japan. According to the users' demands, the remote sensing portal site is created with the aforementioned key components. For the examples of remote sensing data applications, creation of land use maps, disaster mitigations, forest maps, vegetation index map for evaluation of vitality of agricultural fields and forests, etc. are taken into account. In particular for forest map creation, it is created with free open source software: FOSS of classifiers together with open data API derived training samples applied to Landsat-8 OLI data. On the other hand, volcanic eruption is featured for disaster relief with 3D representation by using open data derived DEM data. In

accordance with the users' evaluation reports, it is found that the proposed portal site is useful.

Further investigation is required for increasing application examples in particular for matching to the prefectural local government need.

## ACKNOWLEDGMENT

The author would like to thank all the participants to this project (Kyushu Aeronautical and Space Development Promotion Committee Members) and Mr.Masanori Sakashita of Graduate School of Science and Engineering of Saga University in Japan for his effort to conduct users' demand investigations and simulation studies.

## REFERENCES

[1] Shin-ichi Sobue, Kohei Arai, Osamu Ochiai, Inter-Operable Protocol for Earth Observation Satellite Data Retrievals, Journal of Information Processing Society of Japan, Vol.39, No.3, 222-228, Mar.1998.

[2] Kohei Arai, Manabu Arakawa, Hirofumi Etoh, Fuzzy Search of Earth Observation Data Retrievals Using Physical and Spatial features Based on Fuzzy Theory, Journal of Japan Society of Photogrammetry and Remote Sensing,, Vol.38, No.4, pp.17-25, Aug.1999.

[3] Hirofumi Etoh, Takahiro Yamamoto, Kohei Arai, Indexing Method for Image Database Retrievals by Means of Spatial Features, Journal of Japan Society of Photogrammetry and Remote Sensing,, Vol.39, No.3, pp.14-20,(2000).

[4] Kohei Arai, Bu Kenkyo, Image portion retrievals in large scale imagery data by using online clustering taking into account pusuite algorithm based Reinforcement learning and competitive learning, Journal of Image Electronics Society of Japan, 39, 3, 301-309, 2010

[5] Kohei Arai, Visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals, Journal of Visualization, DOI:10.1007/s, 12650-011-0118-6, 2011.

[6] K.Arai, C.Rahmad, Wavelet based image retrieval method, International Journal of Advanced Computer Science and Applications, 3, 4, 6-11, 2012.

[7] K.Arai, DP matching based image retrieval method with wavelet Multi Resolution Analysis: MRA which is robust against magnification of image size, International Journal of Research and Review on Computer Science, 3, 4, 1738-1743, 2012.

[8] K.Arai, Free Open Source Software: FOSS based GIS for spatial retrievals of appropriate locations for ocean energy utilizing electric power generation plants, International Journal of Advanced Computer Science and Applications, 3, 9, 95-99, 2012.

[9] K.Arai, Visualization of link structure and URL retrievals utilization of interval structure of URLs based on brunch and bound algorithms, International Journal of Advanced Research in Artificial Intelligence, 1, 8, 12-16, 2012.

[10] Kohei Arai, Method for image portion retrieval and display for comparatively large scale of imagery data onto relatively small size of screen which is suitable to block coding of image data compression, International Journal of Advanced Computer Science and Applications, 4, 2, 218-222, 2013.

[11] Kohei Arai, Cahya Rahmad, Content based image retrieval by using multi-layer centroid contour distance, International Journal of Advanced Research in Artificial Intelligence, 2, 3, 16-20, 2013.

[12] Kohei Arai, Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 32038, 2013.

[13] Kohei Arai, Image retrieval and classification method based on Euclidian distance between normalized features including wavelet descriptor, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 19-25, 2013.

[14] Kohei Arai, Numerical representation of web sites of remote sensing satellite data providers and its application to knowledge based information retrievals with natural language, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 26-31, 2013.

[15] K.Arai, Indra Nugraha Abudullar, Hiiroshi Okumura, Image retrieval based on color, shape and texture for ornamental leaf with medicinal functionality, International journal of Image, Graphics and Signal Processing, Vol.6, No.7, June 2014

[16] Cahya Rahmad, K.Arai, Comparison contour extraction based on layered structure and Fourier descriptor on image retrieval, International Journal of Advanced Research on Artificial Intelligence, 4, 12, 71-74, 2015.

[17] S.Sobue and K.Arai, Metadata Definition and Retrieval of Earth Observation Satellite Data, Proceedings of the IEEE Meatadata Conference, (1997)

[18] Kohei Arai, Open GIS with spatial and temporal retrievals as well as assimilation functionality, Proceedings of the Asia Pacific Advanced Network Natural Resource Workshop, Utilization of Earthy Observation Satellite-Digital Asia Special Session 1,p8, 2003.

[19] Kohei Arai, Yuji Yamada, Image retrieval method based on hue information and wavelet description based shape information as well as texture information of the objects extracted with dyadic wavelet transformation, Proceedings of the 11th Asian Symposium on Visualization, ASV-11-08-10, 1-8, 2011.

[20] K.Arai, C.Rahmad, Wavelet based image retrievals, Proceedings of the 260th conference in Saga of Image and Electronics Engineering Society of Japan, 243-247, 2012.

## AUTHORS PROFILE

Kohei Aarai He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission-A of ICSU/COSPAR since 2008. He wrote 33 books and published 510 journal papers. He is Editor-in-Chief of International Journal of Advanced Computer Science and Applications as well as I ternational Journal of Intelligent Systsems and Applications. http://teagis.ip.is.saga-u.ac.jp/

# Implement Fuzzy Logic to Optimize Electronic Business Success

Fahim Akhter
Department of Management Information Systems
College of Business Administration
King Saud University, Saudi Arabia

***Abstract*—Customers are realizing the importance and benefits of shopping online such as convenience, comparison, product research, larger selection, and lower prices. The dynamic nature of e-commerce evokes online businesses to make alterations in their business processes and decisions making to satisfy customers' needs. Online businesses are adopting Business Intelligence (BI) tools and systems with the collaboration of fuzzy logic system to forecast the future of the e-commerce. With the aid of BI, businesses have more possibilities to choose types and structures of required information to serve customers. The fuzzy logic system and BI capabilities would allow both customers and vendors to make right decisions about online shopping. Many experts believe that trust and security are critical risk factors for the embracement of e-commerce. Online trust may be influenced by factors such as usability, familiarity and conducting business with unknown parties. This paper discusses fuzzy logic and BI approach to gauge the level of trust and security in online transactions. The paper further addresses the issues and concerns related to the equilibrium of trust, security, and usability in online shopping.**

***Keywords*—Business Intelligence; Fuzzy Logic; Electronic Business; Trust; Security; Usability***

## I. INTRODUCTION

E-commerce is a widely accepted way of doing business, and within a relatively short time, its services have risen to become a core element of the Internet. The Census Bureau of the Department of Commerce declared that the estimate of U.S. retail e-commerce sales for the fourth quarter of 2015 was $89.1 billion, an increase of 2.1 percent (±0. 9%) since the third quarter of 2015 [1]. E-commerce sales in the USA are projected to reach $482 billion by 2018 [2], accounting for approximately 9% retail sales within the country. The number of digital buyers reached 171 million in 2015 and continues to increase, with the total number of digital buyers projected to surpass 190 million by 2018. Credit and debit cards (73%) are the payment method of choice for USA online shoppers with digital payments (16%) increasing in popularity [2]. The growth of e-commerce is not alone in the US, but the sign points towards continued growth globally. To this end, we aimed at identifying major elements that back the acceptance of e-commerce among the users. This study established the elements that contribute to the growth of users' trust, leading them to complete online retail sales in the United States transactions.

Consumers perceive security threats from different perspectives, for example, whether the web server is owned and operated by a legitimate company, how one can verify that the web pages do not contain malicious code, or how one can ensure that the Web server will not distribute the information to a third party. Similarly, it is important to be able to confirm that the information sent back and forth between the server and the user's browser has not been altered. These concerns illustrate the types of a security issue that can arise in e-commerce transactions. The importance of security implementation is reflected in the policies and actions of the company. Consumers analyze the security policies of vendors mainly through the company's statements on their homepage. These statements normally describe the terms and conditions of the vendor's security policies. For example, through the introduction of security features such as the presence of a secure socket layer, encryption, password and third party security seals. Some companies explain these security policies to alleviate consumers concern while some make these issues hard to find and difficult to understand.

The effectiveness of communicating a commitment to consumers by the use of the latest technology such as third party verification programs, encryption methods, and data protection is important for online companies to appreciate. Previous studies [3] have found no negative correlation between the presence of security statements and the perceived risk of a site. Research also revealed that security statements of Web based companies do show a positive correlation with an increased likelihood that consumers' will purchase from those companies. However, it is not clear during the decision making process, under what conditions the consumer considers security features more important than 'pleasure' features such as convenience, ease of use, and ease of navigation.

An expert has argued that the growth of online e-commerce is attributed not only to ease of use, but also to the fact that they provide reputation and feedback systems that help marketplace platforms create trust [4]. Security features are complex to understand by consumers, but, provided they are properly handled, consumers will confidently conduct their purchase online. Thus, the second hypothesis suggests that the security indicators are the leading factors in the customer's decision to purchase online when compared to design features.

## II. LITERATURE REVIEW

An expert has argued that the operational Business Intelligence supports an analysis and control of business processes by an integration of information systems from a technical and business perspective [5]. BI enables the web-

based business to make informed business decisions and thus can be the source of competitive advantages. This advantage is necessary to improve the timeliness and quality of information to serve consumers 'needs. The access to current and accurate information could eliminate the vagueness within a business process. Due to the vagueness and ambiguous information, consumers' find the decision-making process hard and painful. BI would enhance coordination among business process and enable the business to respond timely to consumers' concern and expectations.

Fuzzy logic provides a means for coping with the ambiguity and vagueness that are often present in B2C commerce [4]. Indeed, it was reported that qualitative and fuzzy type information is commonly used in the trust evaluation process [6]. MATLAB was used because of the built-in support that assisted in understanding the intrinsic relationships between the driving parameters and their effects on the degree of B2C transactions in e-commerce. In conclusion, this study has provided a deeper insight into the factors affecting consumer perception of B2C commerce. Nowadays, consumers have many online alternatives to explore and make a sensible and safe purchase decision. They could find the same items offered by different online retailers with different price options in a matter of a couple of clicks. Consumers buying decision could be influenced by different factors such as trustworthiness, brand, reputation, familiarity, third party seal, security and privacy, fulfillment, presentation and many more. Consumers have to analyze and compare these factors to make a final decision of pursuing online transactions. The purpose of research is to uncover hidden relationships between the critical factors and their effect on the human decision process.

The study [7] provides a review of the literature on the trust issue mainly considering organizational level relationships. The review highlights trust evaluation being a multi-layer, multi-criteria and often a context-dependent process which commonly uses various sources of evidence to judge the trustworthiness of trustees. It has been mentioned [8] that trust has been the most important factor for consumers to do business with each other Contrarily, the trustworthiness of a system provides assurance for consumers to choose a particular e-commerce platform in the first place. They have proposed a fuzzy hybrid multi-criteria analysis approach to measure the trustworthiness of e-commerce systems.

It's extremely critical to acknowledge and understand the trust issues that are associated with the websites which hold the customer back from shopping online [9]. They believe that the consumers' concerns have to be acknowledged in respect of security, usability and trust parameters that leads to the development of the fuzzy system. The researchers has [10] presented and discussed a fuzzy logic based reputation system that assessed the consumers and vendors through the exploitation of a fuzzy trust model which takes into account a set of metrics and a fuzzy based reputation aggregation taking into account credibility concept to discriminate false trust values.

## III. METHODOLOGY

This study adopts a fuzzy logic approach and utilizes a mathematical research toolset known as Matlab fuzzy logic toolbox® to achieve its objectives. The rationale for choosing the fuzzy logic approach is based on the underlying reasoning process behind B2C transactions, which is based on human decision-making. Though many factors influence the decision process of B2C transactions, the perception of an influencing feature is more important than the actual level of the feature itself. For example, if the perceived security level is higher than its actual implementation, then it will contribute positively to the level of B2C outcome. There may be cases where the inverse is true as well, but for such cases, a high level of persuasion will be needed to alter the perception level.

The concept of a linguistic variable is paramount to fuzzy logic, where values of a linguistic variable are expressed as words rather than numerical values. For example, the statement 'e-commerce successfully implies that the linguistic variable *e-commerce* takes the linguistic value *successful*. The term *rule* in Fuzzy logic, which is the most commonly used types of knowledge representation, can be defined as an IF-THEN structure that relates has given information or facts in the IF part to some action in the THEN part [11]. A rule provides some description of how to solve a problem. The rules are relatively easy to create and understand. Any rule consists of two parts: the IF part, called the *antecedent* (premise or condition) and the THEN part called the consequent (conclusion or action). Furthermore, a rule can have multiple antecedents joined by the keywords AND (conjunction), OR (disjunction) or a combination of both.

Fuzzy reasoning includes two distinct parts: (1) evaluating the rule *antecedent* and (2) implication or applying the result to the *consequent*. In the classical expert system, if the rule antecedent is true, then the consequent is also true. In fuzzy systems, where the antecedent is a fuzzy statement, all rules fire to some extent, or in other words, they fire partially. This can also be understood as that if the premise is true to some degree, then the conclusion will also be true to the same degree. It is required to examine the various hedges of this set, which will automatically create some additional subset of the trust indicators. In order to get a complete picture of the fuzzy expert system, an inference diagram can give a detailed explanation of the processes involved.

It should be noted that the initial input(s) are a crisp set of numbers. These values are converted from a numerical level to a linguistic level. Following that, the fuzzy rules are applied and fuzzy inference engine is executed. This will result in a given B2C level as varying degree of membership of fuzzy subsets of the B2C superset. The last step is the defuzzification process, which provides a numeric value for the likelihood of the B2C transaction, such as how is this system useful to the consumer and what benefits can the B2C vendor expect from utilizing such a tool? For the consumer who is unaware or unable to reach a sound buying decision this tool will assist him/her in understanding the parameters that could influence or

ascertain the strength and weaknesses of the B2C site. Similarly, the B2C vendor can use this tool to discover the critical factors on which consumer bases a B2C transaction irrespective of the product/services offered. Consequently, a more realistic picture can be drawn of the factors influencing a consumer's B2C decision.

## IV. Discussion & Analysis

The raw data from the survey has been entered into the worksheet and no interpretation is provided. In the security worksheet, a linguistic input has been assigned a numeric value such as 2, 1 or 0 to calculate the accumulated security level. Accumulated security level is the total sum of the four numeric values and maximum is the percentage of the sum (x/8 multiply by 100). This security level is then calculated as a percentage of a maximum value (usually 8) and from that, a linguistic security level is drawn. It was decided to express the security level as one of three linguistic values, namely low, moderate, and high. The percentage of maximum is evenly distributed to establish a linguistic security level such as low (0-33), moderate (34-66), and high (67-100). The Familiarity worksheet followed the same pattern as used in the security worksheet. The Usability worksheet also followed the same pattern except the linguistic design level was labeled as poor (0-33), moderate (34-66), and good (67-100). In the linguistic data worksheet, trust data have been entered and aligned with security, familiarity, and usability levels by expressing them in terms of linguistic parameters as explained earlier. A Visual Basic language was used to extract and organized data. Trust and other categories are labeled as very low, low, moderate, high, very high and fair, moderate and highly respectively. In the trust rules worksheet, a maximum of twenty-seven unique rules could be identified from respondents input by sorting the security, familiarity and usability columns in alphabetic order. The rules describing the basis for a given trust level were based on degrees of security, familiarity, and usability. These degrees were formulated in terms of their linguistic variables such as low, moderate and high. The degree for usability level was expressed in terms of poor, moderate and good. Similarly, the degree for a trust level was ranging from very low to very high, in five distinct fuzzy sets.

These rules were derived from the survey data after a thorough organization and analysis represent the users' views of the Trust level of a given website based on the given factors. A rule can be extracted such as if (security = high) and (familiarity = low) and (usability = moderate) then (trust = moderate trust). In order to fully understand the contributions from various factors contributing to the Trust level, it is required that we examine the contribution of each factor separately. Figure 1 shows the contribution to Trust of a given website originating from the Security. Therefore, the contribution from Familiarity and Usability has been kept constant at three levels, namely: low, moderate and high corresponding to numeric values for Familiarity and Usability of (1 - 4 and 7). The figure 8 shows that Trust level is monotonically increasing for increasing the perceived security of a website for any given level of Familiarity and Usability (F&U). However, when both F&U is "High" (numeric value of 7) the Trust level is at its maximum for maximum Security. The three curves have one common feature that they exhibit a "staircase shaped" curvature. It is interesting to note that for "low" and "moderate" levels of F&U the developed Trust is almost identical up to a Security level of about 5. Then there is a sharp change on the Trust level between "low" and "moderate", and the perceived Trust for "moderate" F&U is approaching that of "high". A general observation is that the Trust is positively related to Security for any given value of Familiarity and Usability. This observation is also plausible to the human mind. One feature that is disclosed from this figure is that for "high" levels of Security the Trust difference is less significant for "moderate" and "high" levels of F&U. This result could not be anticipated from the outset.

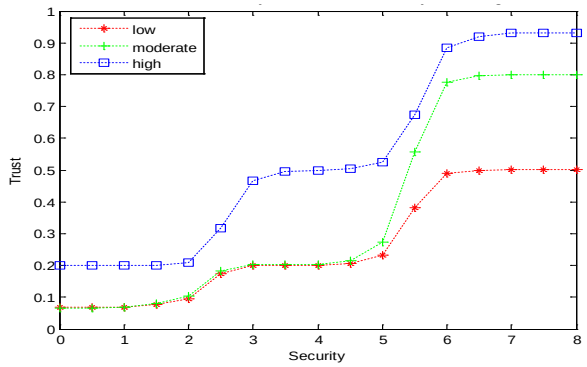



Fig. 1. Trust versus Security for constant familiarity and design

The study has attempted to visualize the Trust level as a continuous function of its input parameters. It should be noted that since the contribution of Familiarity and Usability is identical, it suffices to view Trust as a function of Security and one other factor say Familiarity. The figure 3 attempts to portray variation of Trust as encapsulated in the rules for the Trust.

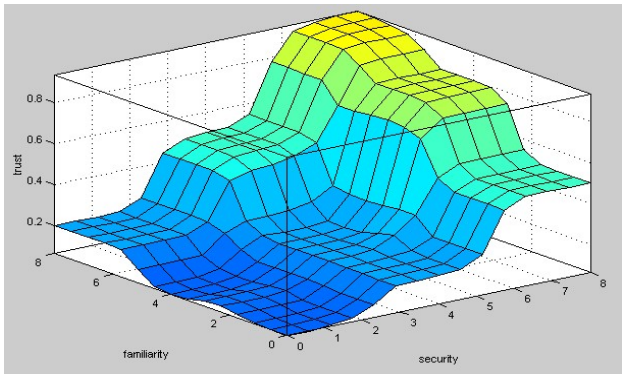



Fig. 2. The trust level is positively related to Levels of Security and Familiarity

The figure 2 shows that Trust level is positively related to Levels of Security and Familiarity. That is when Familiarity and Security are low, then Trust is also low. Furthermore, the Trust is at its maximum when the both Security and Familiarity are maximized. Since the low Familiarity Trust is increasing in "steps" with increasing Security, attaining its maximum at a level of about 0.5. Looking at figure 3 from its topmost point the gradient perpendicular to Security is less than that which is perpendicular to Familiarity axis. This suggests that lowering the Security level has a greater detrimental effect on Trust than

that attained when decreasing Familiarity levels of similar magnitude. The highest gradient for the Trust is when Familiarity is "moderate" and Security is "moderate" to "high". This suggests that when people are somewhat familiar with a website then a small increase in security levels from between moderate to high security will boost their trust in a significant way. Looking at figure 5 diagonally for (low, low) to (high, high) levels of Security and Familiarity one observes three plateaus where the last one is around 0.925 and remains at that level even when the input factors are increased further. This result is somehow unexpected and may be due to the fuzzy nature of the expert system where a "Trust" or "Truth" level of 100 % is unrealistic. The contributions of Familiarity and Usability levels on Trust are similar and, in fact, identical. This can also be deduced from the figure 3, showing complete mirror symmetry of the Trust mapping. One interesting point to note however is that for maximum Familiarity and Usability the Trust level is never higher than 0.5. This plateau is also reached fairly rapidly with high gradients from both sides of the input variables.

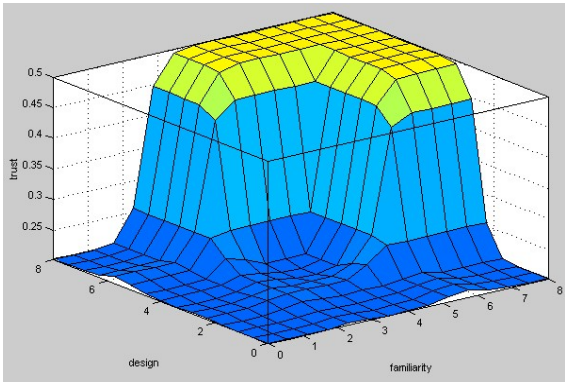



Fig. 3. Mapping of Trust

Consumers generally perceive online shopping as a risky option due to presents of numerous risks. Lack of trust is one of the major risk factors for consumers for not purchasing from online. Security and privacy issues have become an another leading risk factors major for consumers as well business due the growth of the Internet and the complex nature of e-commerce. Strong encryption and authentication are useful in terms of helping to build the trustworthiness of transactions over the Internet but are not yet sufficient for customers to use e-commerce with confidence. Many consumers do not trust an available security option for online transactions and neither trust e-commerce systems, nor believe they will be able to evaluate or control their information when providing them. The issue of trust is very important, especially when the success of e-commerce relies on the profitable exchange of business and consumer relationship.

Fuzzy logic based decision support system, containing BI tools, supports business and organizational decision-making relationship. A properly designed system is an interactive software-based system intended to help consumers' to gauge the level of risk factors in a given website. The proposed system in this study could compile useful information from raw data, documents, personal knowledge, and from business models to create rules to identify the potential risk factors in respect of conducting online transactions. Online business required a current and precision information about the consumers' background, needs and perception of the business process. In this respect, the fuzzy logic approach is capable of collaborating with BI to collect and process the consumers' information. This partnership will assist in creating the rules for making right decisions.

The proposed system is also beneficial to web-based businesses. It allows the businesses to reach out the consumers' possibly by BI. The recommendations and suggestions of the proposed systems would assist the management to address the needs of the customers, ability to act on market drivers, optimization of business and customer retention.

## REFERENCES

[1] U.S. Census Bureau News. Quarterly Retail E-Commerce Sales 4th Quarter 2015. Available at: https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf (Accessed: 2 March 2016).

[2] PFS web. eCommerce Summary. Available at: http://www.pfsweb.com/pdf/global-ecommerce-book/2016-Global-eCommerce-Book-USA.pdf (Accessed: 5 March 2016).

[3] Fahim Akhter and Wafi Albalawi, "Online Trust and Young Entrepreneurs: A case Study of Saudi Arabia", International Journal of Theoretical and Applied Information technology, E-ISSN 1817-3195 / ISSN 1992-8645, 2016.

[4] S. Tadelis, "The Economics of Reputation and Feedback Systems in E-Commerce Marketplaces," in IEEE Internet Computing, vol. 20, no. 1, pp. 12-19, Jan.-Feb. 2016

[5] T. Hänel and C. Felden, "An Empirical Investigation of Operational Business Intelligence Perspectives to Support an Analysis and Control of Business Processes," System Sciences, 2015 48th Hawaii International Conference on, Kauai, HI, 2015, pp. 4722-4731.

[6] Z. M. Husseini, M. H. Fazel Zarandi and S. M. Moattar Husseini, "Trust evaluation for buyer-supplier relationship concerning fuzzy approach," Fuzzy Information Processing Society held jointly with 2015 5th World Conference on Soft Computing, 2015 Annual Conference of the North American, Redmond, WA, 2015, pp. 1-6.

[7] Husseini, Z.M.; Fazel Zarandi, M.H.; Moattar Husseini, S.M., "Trust evaluation for buyer-supplier relationship concerning fuzzy approach," in Fuzzy Information Processing Society, 2015 Annual Conference of the North American, vol., no., pp.1-6, 17-19 Aug. 2015

[8] Zhengping Wu; Lifeng Wang, "Trustworthiness Measurement of E-commerce Systems Using Fuzzy Hybrid Multi-criteria Analysis," in Trustcom/BigDataSE/ISPA, 2015 IEEE, vol.1, no., pp.668-675, 20-22 Aug. 2015

[9] Kaur, B.; Madan, S., "A fuzzy expert system to evaluate customer's trust in B2C E-Commerce websites," in Computing for Sustainable Global Development (INDIACom), 2014 International Conference on, vol., no., pp.394-399, 5-7 March 2014

[10] Fahim Akhter and Zakaria Maamar. "Evaluation of Fuzzy Models to Support Online-Trust Assessment." Comparison-Shopping Services and Agent Designs. IGI Global, 90-100, 2009

[11] Zadeh, L. Outline of a new approach to the analysis of the complex system and decision processes. IEEE Transactions on System, Man, and Cybernetics, SMC-3 (1), 1974. 28-44.

# Applying Swarm Optimization Techniques to Calculate Execution Time for Software Modules

Nagy Ramadan Darwish
Department of Computer and Information Sciences, Institute of Statistical Studies and Research, Cairo University, Cairo, Egypt

Ahmed A. Mohamed
Department of Information System, Higher Technological Institute, 10th of Ramadan City, Egypt

Bassem S. M. Zohdy
Department of Business Technology, Canadian International College, Cairo, Egypt

***Abstract*—This research aims to calculate the execution time for software modules, using Particle Swarm Optimization (PSO) and Parallel Particle Swarm Optimization (PPSO), in order to calculate the proper time. A comparison is made between MATLAB Code without Algorithm (MCWA), PSO and PPSO to figure out the time produced when executing any software module. The proposed algorithms which include the PPSO increase the speed of executing the algorithm itself, in order to achieve quick results. This research introduces the proposed architecture to calculate execution time and uses MATLAB to implement MCWA, PSO and PPSO. The results show that PPSO algorithm is more efficient in speed and time compared to MCWA and PSO algorithm for calculating the execution time.**

***Keywords*—Particle Swarm Optimization; Parallel Particle Swarm Optimization; MATLAB Code without Algorithm***

## I. INTRODUCTION

Testing is a crucial phase that is performed during software development. It is a primary technique which is used to gain consumer confidence in the software. It is conducted by executing the program developed with test inputs and comparing the observed output with the expected one [1, 2]. Testing is the writing and applying all software tests to ensure the confidence in the operation of the program. Testing is the phase of development that is carried out after the main coding efforts [3].

Execution time is the time during which software is running. Calculating execution time is very important in many fields, such as; medical system, army system, and airlines system … etc., Where any time delay in these software misfortunes may occur.

This research selects primary studies that published between 2005 and 2015, while searching many electronic databases in order to determine if similar work had already been performed, and locates potentially relevant studies.

C. Mao [4] proposed a search-based test data generation solution for software structural testing using particle swarm optimization (PSO) technique. A. Windisch, S. Wappler and j. Wegener [5] applied a particle swarm algorithm for evolutionary structural testing. R. Ding and H. Dong [6] proposed a hybrid particle swarm genetic algorithm to apply in software testing using case automate generations. A. S. Andreou, K. A. Economides and A. A. Sofokleous [7] proposed an enhanced testing framework that combines data flow graphs with genetic algorithms (GA) to generate optimum test cases. P. Palangpour, G.K. Venayagamoorthy, and S.C. Smith [8] presented a pipelined architecture for hardware particle swarm optimization (PSO) implementation to achieve much faster execution times than possible in software. A. Mansoor [9] developed AI technique based on genetic algorithm for the optimization of software test data. M. Syafrullah and N. Salim [10] proposed a new approach based on particle swarm optimization techniques in order to improve the accuracy of term extraction results. J. H. Andrews, T. Menzies, and F.Li [11] described a system that is based on a genetic algorithm (GA) to find parameters for randomized unit testing that optimize test coverage. J. CHANG and J. Pan [12] presented a Parallel Particle Swarm Optimization (PPSO) algorithm and a three communication strategies. D. Arora, A. S. Baghel [1] presented a method that uses genetic algorithm and particle swarm optimization for optimizing software testing by finding the most error prone paths in the program.

The method used here in this research is to calculate execution time for software modules, this could be achieved by using the Particle Swarm Optimization (PSO) techniques, as the each population in each iteration search for best execution time through particles in this population, and finally compare the best solution to produce the best execution time, also the use of parallel particle swarm optimization helps to run the populations and particles in distributed processing systems to help find the best solution in parallel, then selecting the best execution time. It is very crucial phase for any software to determine its quality and ability to meet requirements, which could be achieved through test this software, testing as a phase of software engineering process, literally takes about 40~50% of the development efforts in software houses [13].

It is noteworthy that life critical software could use more efforts and resources, if it is not tested perfectly, the software may cause dangerous consequences as timetable delays, cost overrun. Also software community aims to deliver high quality software to customers, to ensure that the software will run perfect with no delays in execution time, as this is the aim of this research is to calculate the execution time [14, 15, 16]. Also the proposed algorithm in this research is done automatically through a testing tool that produces the results of execution time, also trials have been done and the results of sample code is depicted below in implementation part.

The paper is organized as follows: in part II an introduction and brief description of PSO algorithm, in part III brief description of the two types of PPSO techniques, in part IV description of the proposed PSO and PPSO algorithms, followed by the implementation in part V, and then analysis and results in part VI, the conclusion and future work in part VII.

## II. PARTICLE SWARM OPTIMIZATION

Modelling of swarms was initially proposed by Kennedy to simulate the social behaviour of fish and birds, the optimization algorithm was presented as an optimization technique in 1995 by Kennedy and Eberhart, PSO has particles which represent candidate solutions of the problem, each particle searches for optimal solution in the search space, each particle or candidate solution has a position and velocity. A particle updates its velocity and position based on its inertia, own experience and gained knowledge from other particles in the swarm, aiming to find the optimal solution of the problem [17].

The particles update its position and velocity according to the following Equation:

$$v_1^{k+1} = wv_1^k + c_1 rand_2 \times (pbest_i - s_i^k) + (gbest_i - s_i^k) \quad (1)$$

Where:

$v_i^{k+1}$ = Velocity of agent i at iteration k,
w = Weighting function,
cj = Weighting factor,
rand = Random number between 0 and 1,
$S_i^k$ = Current position of agent iteration k,
pbesti = Pbest of agent i,
gbesti = gbest of the group.

The weighting function used in Equation 1:

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \quad (2)$$

Where:

Wmax= Initial weight,
Wmin= Final weight,
itermax= Maximum iteration number,
Iter = Current iteration number.

According to more than ninety modifications are applied to original PSO [17, 18].

## III. PARALLEL PARTICLE SWARM OPTIMIZATION

PSO is optimized to be implemented on distributed systems, the iterations and particles within each iterations of the PPSO are independent of each other, so results could be parallel analysed. PPSO could be divided into two types [19, 20, 21]:

*1) Synchronous Parallel Particle Swarm Optimization*

PSO parallel implementation is to simply evaluate the particles (solutions), or in other words the execution time produced within each iteration in parallel, without changing the overall logic of the algorithm itself. In this implementation, all particles within design iteration are sent to the parallel computing environment, and the algorithm waits for all the analyses to complete before moving to the next iteration. This implementation is referred to as a synchronous implementation. This method is used in this research [22, 23].

*2) Asynchronous Parallel Particle Swarm Optimization Algorithm*

Considering an asynchronous algorithm means that particles (solutions) or as mentioned before execution time produced in the next iteration are analysed before the current design iteration is completed. The goal is to have no idle processors as one move from one iteration to the next. [22]

The key to implementing an asynchronous parallel PSO algorithm is to separate the update actions associated with each point and those associated with the swarm as a whole. These update actions include updating the inertia value and the swarm and point histories. For the synchronous algorithm, all the update actions are performed at the end of each design iteration. For the asynchronous algorithm, researchers want to perform point update actions after each point is analysed and the swarm updates actions at the end of each design iteration. The parts of the algorithm that need to be considered when looking at the update actions are the velocity vector, and the dynamic reduction of the inertia value. [22]

The velocity vector is the centre point of any PSO algorithm. For each design point, the velocity vector is updated using the following dynamic properties for that point: the previous velocity vector; the current position vector; and the best position found so far. In addition, the updated inertia value and the best position for the swarm as a whole are also required. To do the velocity update in an asynchronous fashion, researchers need to update the position vector and the best position found so far for each design point directly after evaluating that point. For the best position in the swarm, researchers have two choices:

Use the best position in the current iteration, or use the best position found so far. To keep the best position for the swarm current when moving to the next design iteration, before the current iteration is completed, it is necessary to use the best position found so far rather than the best position in the current iteration. This setup allows the algorithm to update all required dynamic properties of the velocity vector directly after evaluating each design point, except for the inertia value. The inertia value is the only iteration level update required to compute the velocity vector and is updated at the end of each design iteration. The craziness operator is the only other iteration level update and is also performed at the end of each design iteration. [23]

The asynchronous algorithm is thus very similar to the synchronous algorithm, except that researchers update as much information as possible after each design point is analysed. The inertia is only applied when design iteration is completed. Of course, this could result in some points of the next design iteration being analysed before the inertia operator is applied for that design iteration. However, the influence on the overall performance of the algorithm seems to be negligible [24].

## IV. PROPOSED PSO & PPSO ALGORITHMS

The proposed architecture is based on PSO and PPSO algorithms to calculate execution time depicted below in figure 1. Additionally, in order to evaluate the execution time for software module, a proposed PPSO (Parallel PSO) algorithm to calculate execution time also introduced and the recommended execution time strategy is determined for implementing this PPSO algorithm.

- The PSO Algorithm to Calculate Execution Time can be listed in following steps:

*A. Initialize the population with N Particles. And Set iterations counter I = 0.*

*B. Apply Fitness function: Calculating the fitness value by calculating the percentage of this particle will share in minimizing the total processing time to find the optimal solution.*

*C. Compare the calculated fitness value of each particle with its (lbest). If current value is better than (gbest), then reset (gbest) to the current index in particle array. Select the best particle as (gbest).*

*D. Calculated fitness value among the neighboured particles in the network achieved so far in the iteration.*

*E. Update each Particle Velocity and position according to Eq. (1).*

$$v_1^{k+1} = wv_1^k + c_1 rand_2 \times (pbest_i - s_i^k) + (gbest_i - s_i^k) \quad (1)$$

Where I= 0, 1, 2… M-1

To prevent the velocity from becoming too large, researchers set a maximum value to limit the range velocity as

$$-VMAX \leq V \leq VMAX$$

*F. Cost function assigns the highest fitness value in the iteration and which has a current position (xi).*

t = t +1.

*G. End While.*

*H. Stop.*

So these recursive steps continue until reaching the termination condition, and the termination condition achieved when the cost function finishes the execution, and finds the optimal time and solution.

In PPSO, computation time of PSO can be reduced with the parallel structure. Parallel processing aims at producing the same results achievable using multiple processors with the goal of reducing the run time. The same steps described in PSO will be applied, but in step (a) PPSO will define how many group of processors needed for the cost function to be executed, because it can be designed to be 2n sets.

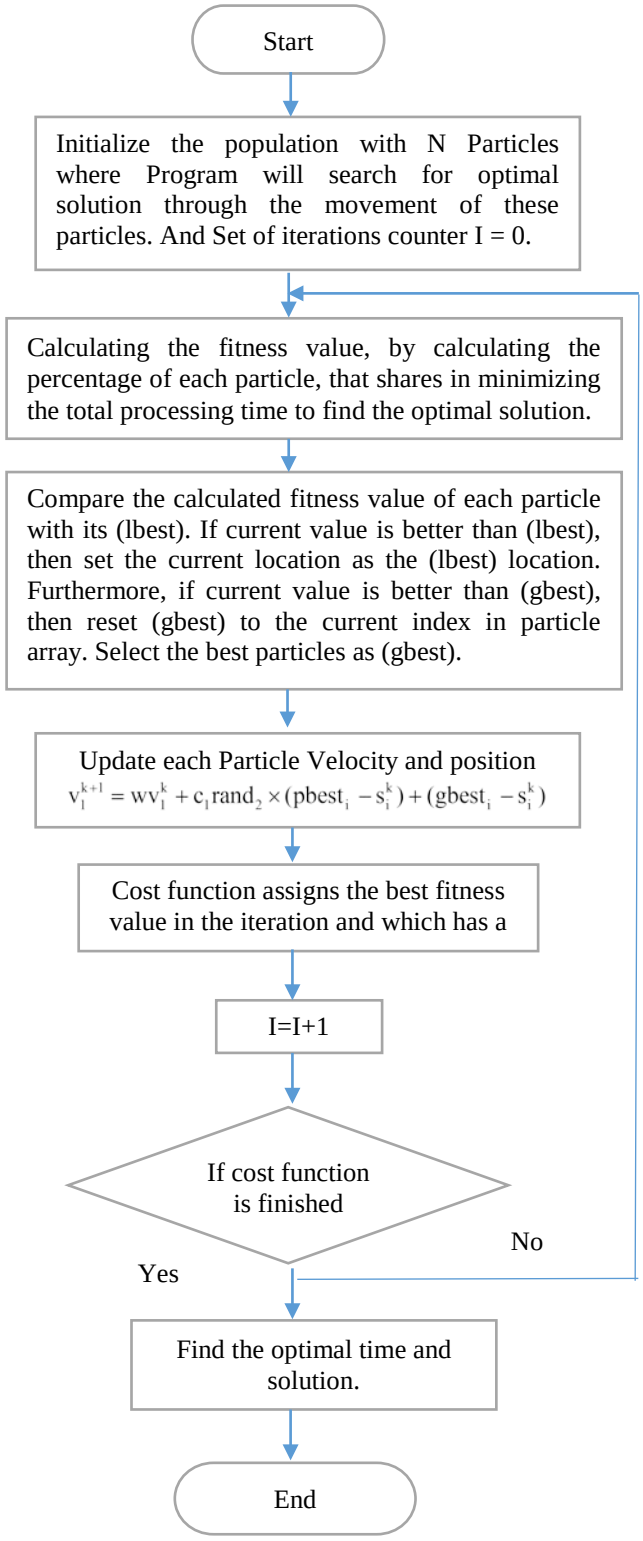



Fig. 1. Proposed PSO Based Algorithm to Calculate Execution Time

The performance of the Parallel PSO can be evaluated using Amdahl's Law Eq. [25].

Speedup (Sp) = $1/f_s + f_p/p$

Where:

fs= serial fraction of code

fp= parallel fraction of code

P= number of processors

Suppose serial fraction of code (0.5), parallel fraction of code (0.5) and number of processors (2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024).

If P=2: Sp= 1 / (0.5+0.25) = 1.33

TABLE I. PPSO ALGORITHM

| P | SP | Elapsed Time |
|---|---|---|
| 2 | 1.33 | 0.02 |
| 4 | 1.60 | 0.017 |
| 8 | 1.79 | 0.012 |
| 16 | 1.89 | 0.008 |
| 32 | 1.96 | 0.005 |
| 64 | 2.00 | 0.003 |
| 128 | 2.00 | 0.002 |
| 256 | 2.00 | 0.001 |
| 512 | 2.00 | 0.001 |
| 1024 | 2.00 | 0.0003 |

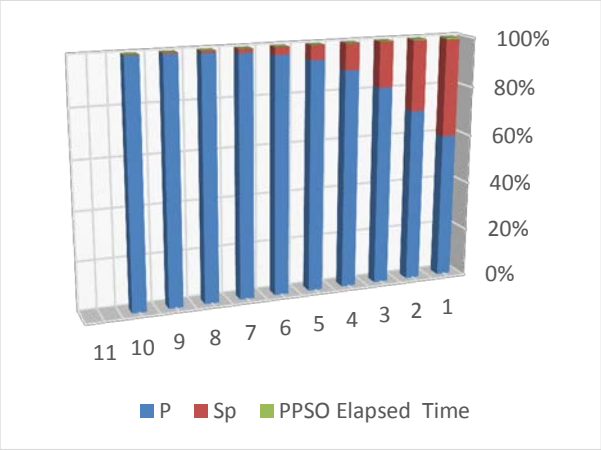



Fig. 2. PPSO Algorithm

This figure shows that the increase of number of processors, increase in speed results and decrease in elapsed time.

## V. IMPLEMENTATION OF THE PROPOSED ALGORITHM

This section introduces implementation of MCWA, PSO and PPSO where cost function (CF) is:

```
CF: function z=Sphere(x)
    z=sum (x. ^2);
   end
```

This paper implementation proposed cost function uses MCWA, PSO and PPSO. The first implementation uses MCWA where results are shown below in table 2:

TABLE II. RESULTS OF MCWA

| Iteration | CF | Elapsed Time(ET) |
|---|---|---|
| 1 | 1 | 16.16 |
| 2 | 4 | 97.15 |
| 3 | 9 | 109.26 |
| 4 | 16 | 118.05 |
| 5 | 25 | 126.38 |
| 6 | 36 | 133.59 |
| 7 | 49 | 140.11 |
| 8 | 64 | 147.25 |
| 9 | 81 | 157.59 |
| 10 | 100 | 169.34 |

In table 2 each test case (iteration) to optimize the cost function but elapsed time increased and shown below in figure3:

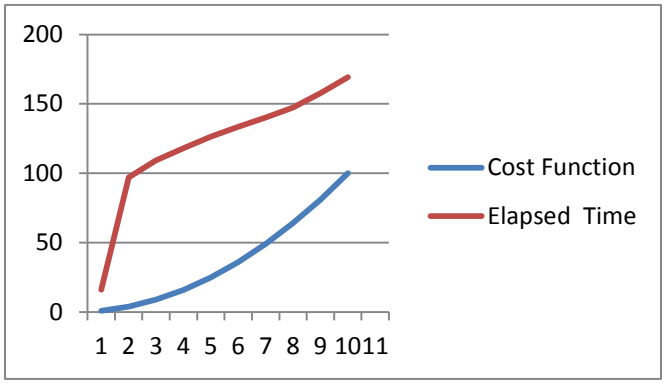



Fig. 3. Releationship between CF and ET in MCWA

In figure 3 shown relationships between CF and ET where found positive relationship between them.

The second implementation uses PSO where results are shown below in table 3:

TABLE III. RESULTS OF PSO

| Iteration | CF | ET |
|---|---|---|
| 1 | 58.73 | 0.04 |
| 2 | 24.48 | 0.07 |
| 3 | 15.00 | 0.10 |
| 4 | 5.85 | 0.13 |
| 5 | 5.85 | 0.16 |
| 6 | 5.85 | 0.19 |
| 7 | 5.34 | 0.21 |
| 8 | 2.73 | 0.24 |
| 9 | 2.73 | 0.27 |
| 10 | 2.73 | 0.30 |

In table 3 each test case (iteration) to optimize the cost function but elapsed time decreased compared with MCWA and shown in figure 4.

In figure 4 shown relationships between CF and ET where found inverse relationship between them.

The third implementation uses PPSO where results are shown in table 4.

In table 4 each test case (iteration) to optimize the cost function but elapsed time decreased compared with MCWA, PSO and shown in figure 5.

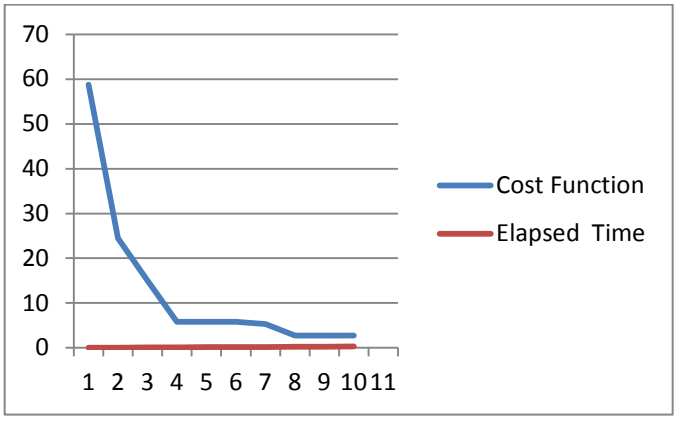



Fig. 4. Releationship between CF and ET in PSO

TABLE IV. RESULTS OF PPSO

| x | P | Sp | CF | ET |
|---|---|---|---|---|
| 1 | 2 | 1.33 | 58.73 | 0.02 |
| 2 | 4 | 1.60 | 24.48 | 0.017 |
| 3 | 8 | 1.79 | 15.00 | 0.012 |
| 4 | 16 | 1.89 | 5.85 | 0.008 |
| 5 | 32 | 1.96 | 5.85 | 0.005 |
| 6 | 64 | 2.00 | 5.85 | 0.003 |
| 7 | 128 | 2.00 | 5.34 | 0.002 |
| 8 | 256 | 2.00 | 2.73 | 0.001 |
| 9 | 512 | 2.00 | 2.73 | 0.001 |
| 10 | 1024 | 2.00 | 2.73 | 0.0003 |

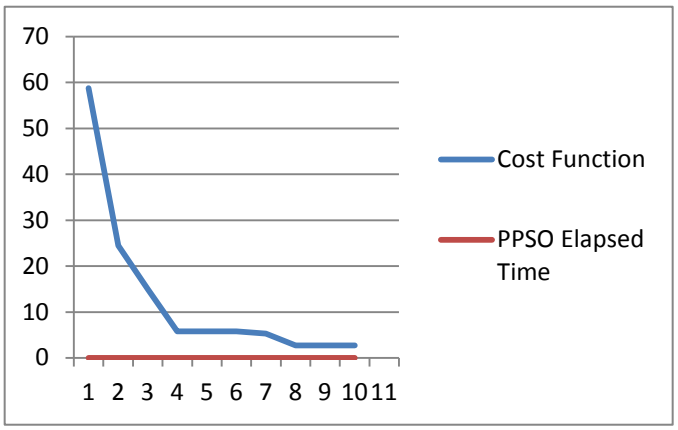



Fig. 5. Releationship between CF and ET in PPSO

In figure 5 shown relationships between CF and ET where found inverse relationship between them.

This paper introduces compared between PSO and PPSO where shown in figure 6.

In figure 7 shown inverse relationships between SP, PSO and PPSO Whenever an increase in speed occur where decreased in PSO and also more decreased in PPSO.

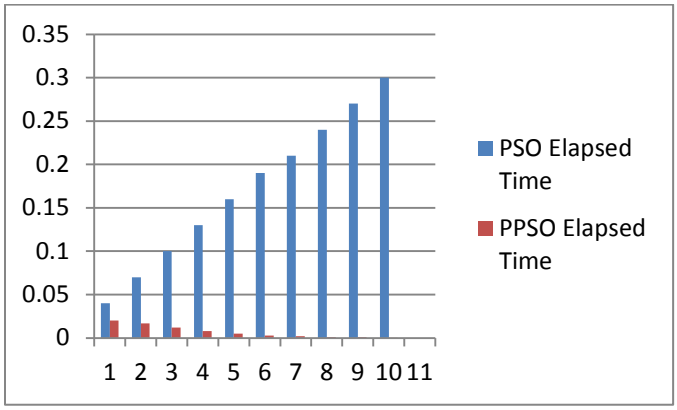



Fig. 6. Releationship between PSO and PPSO

In figure 6 shown relationships between PSO and PPSO where elapsed time in PPSO decreased compared with PSO. In figure 7 shown relationships between SP, PSO and PPSO.

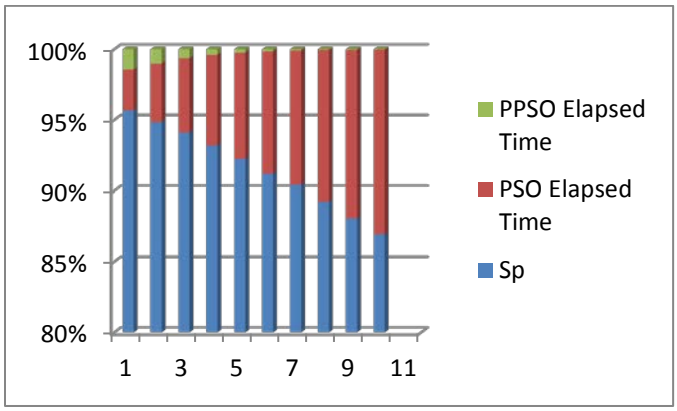



Fig. 7. Releationship between SP, PSO and PPSO

## VI. ANALYSIS AND RESULTS

This paper presented the results of PSO and PPSO. In PSO, the results show inverse relationship between CF & ET, although the optimizing, the cost function elapsed time decreased compared with MCWA. In PPSO found that there is an inverse relationship between CF & ET although optimizing the cost function, but elapsed time decreased compared with MCWA and PSO. This paper shows the relationship between PSO and PPSO where elapsed time in PPSO decreased compared with PSO and shows inverse relationship between SP, PSO and PPSO Whenever an increase in speed occur it decreased in PSO and also time decreased more in PPSO.

## VII. CONCLUSION AND FUTURE WORK

PSO is relatively recent heuristic approach; it is similar to PPSO in a way that they both are population based evolutionary algorithms.

The research presents the application of PSO and PPSO. The proposed research described the basic concepts of PSO and PPSO, calculating execution time for software modules for using PSO and PPSO and how they are useful in finding the optimal solution to the problem. Comparative study is done between both the algorithm where PPSO can be useful, and showing how PPSO overcome the drawback of PSO. This paper shows that PPSO algorithm is more efficient in speed and time compared with MCWA and PSO algorithm to calculate the execution time.

In future, the researchers aim to apply the proposed approach using ante colony optimization (ACO) and cat optimization (CO) and compare the results with other results produced from other evaluation techniques of swarm optimization algorithms. In addition, the researchers aim to enhance the proposed approach by examining more hybrid techniques to calculate execution time for software modules.

## REFERENCES

[1] D.Arora, A.Baghel "Application of Genetic Algorithm and Particle Swarm Optimization in Software Testing "Journal of Computer Engineering , Volume 17, PP 75-78 ,2015.

[2] James H. Andrews, Tim Menzies, Felix C.H. Li, "Genetic Algorithms for Randomized Unit Testing", IEEE Transactions on Software Engineering, vol.37, no. 1, pp. 80-94, January/February 2011.

[3] Nagy Ramadan Darwish "Towards An Approach For Evaluating The Implementation Of Extreme Programming Practices" International Journal of Intelligent Computing and Information Science, Volume 13, PP 55-67,2013.

[4] C. Mao, "Generating Test Data for Software Structural Testing Based on Particle Swarm Optimization," Arabian Journal for Science and Engineering, vol. 39, pp. 4593-4607, 2014.

[5] A. Windisch, S. Wappler and j. Wegener, "Applying Particle Swarm Optimization to Software Testing," in Proc. 9th Conf. Genetic and evolutionary computation, 2007, pp. 1121-1128.

[6] R. Ding and H. Dong "Automatic Generation of Software Test Data Based on Hybrid Particle Swarm Genetic Algorithm," in Electrical & Electronics Engineering (EEESYM), 2012, pp. 670 - 673.

[7] A. S. Andreou, K. A. Economides and A. A. Sofokleous "An Automatic Software Test-Data Generation Scheme Based on Data Flow Criteria and Genetic Algorithms," in 7th Int. Conf. Computer and Information Technology, 2007.

[8] P. Palangpour, G.K. Venayagamoorthy, and S.C. Smith, "Particle Swarm Optimization: A Hardware Implementation," in Proc. Int. Conf. Computer Design, 2009.

[9] A. Mansoor, "Automated Software Test Data Optimization Using Artificial Intelligence," International Journal of Information and Communication Trends, vol. 1, pp. 1-60, 2014.

[10] M. Syafrullah and N. Salim, "Improving Term Extraction Using Particle Swarm Optimization Techniques," Journal of Computing, vol. 2, 2010.

[11] J. H. Andrews, T. Menzies, and F.Li, "Genetic Algorithms for Randomized Unit Testing," IEEE Transactions on Software Engineering, vol. 37, no. 1, 2011.

[12] J. CHANG and j. Pan, "A Parallel Particle Swarm Optimization Algorithm with Communication Strategies," Journal of Information Science and Engineering, vol. 21, pp. 809-818, 2005.

[13] Jovanovic and Irena,"Software TestingMethods and Techniques," May 26, 2008.

[14] Saswat Anand, Edmund Burke, Tsong Y. Chen, et al., "An Orchestrated Survey on Automated Software Test Case Generation" Journal of Systems and, Software, 2013.

[15] "Addressing Software Testing Costs, Complexity and Challenges: Sogeti UK's Annual TestExpo Survey", Survey Report, Sogeti Inc., 2013.

[16] Myers, Glenford J., IBM Systems Research Institute, Lecturer in Computer Science, Polytechnic Institute of New York, "The Art of Software Testing", Copyright 1979. by John Wiley & Sons, Inc.

[17] Aly, Walid; Yousif, Basheer; Zohdy, Bassem "A Deoxyribonucleic Acid Compression Algorithm Using Auto-Regression and Swarm Intelligence". Journal of Computer Science, 690-698, 9(6), 2013.

[18] Syafrullah M., Salim N., 2010, "Improving Term Extraction Using Particle Swarm Optimization Techniques", Journal of Computing, Vol. 2, Issue2, pages 116-120, 2010.

[19] Chang, J.F., Chu, S.C., Roddick, J.F., Pan, J.S., "A parallel particle swarm optimization algorithm with communication strategies", Journal of information science and engineering. 21, 809–818, 2005.

[20] Shu-Chuan, Chu et al. "Parallel Particle Swarm Optimization Algorithms With Adaptive Simulated Annealing", Springer, 31, 2006.

[21] Venter G, Sobieszczanki-Sobieksi J. A parallel particle swarm optimization algorithm accelerated by asynchronous evaluations. Journal of Aerospace Computing, Information, and Communication. 6th World Congresses of Structural and Multidisciplinary Optimization, 2005

[22] Şaban Gülcü, and Halife Kodaz, "A novel parallel multi-swarm algorithm based on comprehensive learning particle swarm optimization," Engineering Applications of Artificial Intelligence, vol. 45, pp. 33–45, 2015.

[23] R. V. Kulkarni and G. K. Venayagamoorthy, "Particle swarm optimization in wireless-sensor networks: a brief survey," IEEE Transactions on Systems, Man and Cybernetics Part C, vol. 41, no. 2, pp. 262–267, 2011.

[24] C. A. Voglis , K. E. Parsopoulos , I. E. Lagaris, "Particle swarm optimization with deliberate loss of information", Soft Computing - A Fusion of Foundations, Methodologies and Applications, v.16 n.8, p.1373-1392, August 2012.

[25] Herb Sutter, "The Free Lunch is Over: A Fundamental Turn Toward Concurrency in Software", Dr. Dobb's Journal, 30(3), March 2005.

# Improving Performance of Free Space Optics Link Using Array of Receivers in Terrible Weather Conditions of Plain and Hilly Areas

Amit Gupta
Professor, Department of ECE
Chandigarh University,
Gharuan, India

Shaina
Research Scholar, Department of ECE
Chandigarh University,
Gharuan, India

Surbhi Bakshi
Associate Professor, Department of ECE
Chandigarh University,
Gharuan, India

Mandeep Chaudhary
School of Engineering & Digital arts
University of Kent
Canterbury,UK

***Abstract*—Free-space optical (FSO) communication is a cost effective and high data rate access technique, which has been proving itself a best alternative to radio frequency technology. FSO link provides high bandwidth solution to the last mile access bottleneck. However, for terrestrial communication systems, the performance of these links is severely degraded from atmospheric loss mainly due to fog, rain and snow. So, a continuous availability of the link is always a concern. This paper investigates the dreadful weather effects such as rain, fog, snow, and other losses on the transmission performance of FSO systems. The technique of using an array of receivers for improving the performance of FSO links is explored in this paper. It involves the deployment of multiple photo detectors at the receiver end to mitigate effects of various weather conditions. The performance of the proposed system is evaluated in terms of bit error rate, received signal power, Q- factor and height of eye diagram. The influence of various weather conditions of plain and hilly areas are taken into consideration and results are compared with conventional FSO links.**

***Keywords—Free space optics (FSO) communication; Array of photo detectors; Bit Error Rate (BER); Eye diagram; Quality factor (Q factor); Bad weather effects***

## I. INTRODUCTION

The optical wireless communication (OWC) systems have attracted a lot of interest of the users because they can solve the last mile problem in urban environments. OWC, also recognized as free space optical (FSO) communication, has emerged as a commercially feasible alternative to radio frequency (RF) and millimetre waves wireless communication for reliable and rapid deployment of data and voice networks. FSO communication using high bandwidth transmission links has enormous potential to serve for requirements of high data rate transmissions. License free bandwidth, high carrier frequency (range 20 THz- 375Thz), easy deployment, appreciable security of data, avoiding electromagnetic pollution, low power consumption enables FSO links to provide high data rates communications[1]. Its various advantages over existing Radio Frequency (RF) technologies like wider bandwidth that can support a large number of users without any delay or interference in communication, have increased its demand in the market.

Despite of being on the list of most desirable technologies of the next generation, its deployment is highly dependent on atmospheric variations thus related to its reliability and availability issues. Fog, snow and clouds scatter or absorb the optical signal, which causes transmission errors. Maintaining a clear line of sight (LOS) between transmitter and receiver is also one of the major challenges in establishing optical wireless links in the free space [6]. The LOS is disturbed due to atmospheric influences like fog, rain, snow, dust, sleet, clouds or temporary physical obstructions like birds and airplanes. Various researchers have come up with the results that optical attenuation can reach up to 128dB/km in heavy fog and snow conditions in different areas [7]. The scattering, absorption and refraction of light signals reduce the link capacity and availability in different weather conditions.

To lessen these effects, techniques like using multiple transmit lasers and multiple receive apertures can be applied [8-9]. The performance of FSO links in the presence of atmospheric turbulence had been analyzed using spatial diversity [10-11]. To calculate the error rate performance, outage probability and diversity gain for multi-input multi-output FSO links, the combined effect of atmospheric turbulence and misalignment was also considered [12-13]. Then the effect of weather conditions was taken into account using array of receivers [14-16].

But the effect of weather conditions of hilly areas like heavy fog, wet snow, dry snow etc was not discussed by researchers in previous literature. The consideration of these parameters cannot be ignored while installation of FSO link especially in hilly areas. In this paper a comprehensive analysis of FSO link in all weather conditions has been performed using one of the most important approaches of array of photo detectors to reduce the effects of attenuation on received signal. At an ideal case, the only cause of signal attenuation is distance

of transmission. So, the additional losses we have taken into account are due to weather conditions only and other losses are considered to be 0 dB/km. The study of bit error rate, height of eye diagram, Q factor and maximum received power is taken into account for studying the performance of FSO link.

The paper is organized as: After the introduction of FSO systems, system analysis of proposed FSO model is provided and various parameters that affect the quality of the signal in the link are discussed in section II. In Section III, the results obtained after simulations are evaluated for both hilly and plain areas using array of receivers. Finally, conclusions are drawn in Section IV.

## II. System Analysis

A synoptic diagram of the considered system model is depicted in Figure 1 below. The block diagram shows the three key function elements of FSO system that are the transmitter, the atmospheric channel and the receiver. The transmitter which is used for converting electrical signal into optical signal consists of a modulator, a laser driver, a laser and a power meter.

The modulator used in the link converts the information into the desired signal and controls the amplitude of an optical signal. Laser driver provides the power to the laser for its proper functioning and helps to prevent aging and other environmental effects of laser. The range of the link for evaluating the performance is chosen as 500m and the transmission wavelength chosen for the working of laser is 1550 nm which is the 3rd optical window of wireless transmission. It is chosen to work on this wavelength as the functioning of FSO link is more robust and safe for human eye at this value [2-3].

The information signal is transmitted over FSO channel where it undergoes attenuation and power loss as a result of absorption, scattering and turbulence. At the receiver end, the signal is amplified and detected by an array of receivers which improve the overall efficiency and accuracy of the system. The filter and regenerator are used to preserve the wave shape of the signal. Power meter and BER analyser are used to measure the parameters of received signal. The data rates up to 100 Gb/s can be achieved using FSO technology [4-5].

Attenuation present in the atmosphere of the system can affect its performance. Atmospheric attenuation and geometric losses constitute all attenuation. It is the effect of particles present in the air for e.g. haze, rain, fog, snow etc. These particles can stay a longer time in the atmosphere. So, attenuation values depend upon the visibility level at that time. To reduce these effects, a system is proposed that can work properly under these conditions. The value of parameters on which the system is operating is mentioned in table 1.

The total attenuation of wireless medium communication system can be estimated [17] as:

$$\alpha = \alpha\text{fog}_{\gamma} + \alpha\text{snow}_{\gamma} + \alpha\text{rain}_{\gamma} + \alpha\text{scattering}_{\gamma}, \text{dB/km} \quad (1)$$

where, α=attenuation and

γ=is transmission wavelength in μm

The main attenuation factor for optical wireless links is fog, but the attenuation caused by an effect of rain cannot be ignored, especially in environments where rain is more frequent than fog.

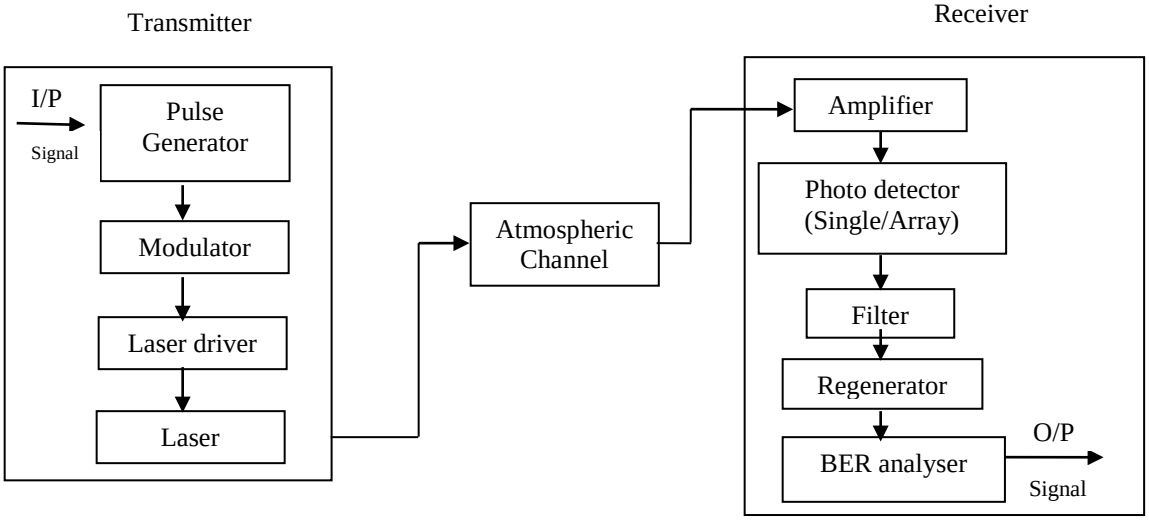



Fig. 1. Block diagram of FSO link

TABLE I. PROPOSED OPERATING PARAMETERS FOR FREE SPACE OPTICAL COMMUNICATION SYSTEMS

| Operating parameters | Value |
|---|---|
| Signal transmitted power (plain areas) | 1mW |
| Signal transmitted power (hilly areas) | 20 W |
| Attenuation (plain areas) | 0-43 dB |
| Attenuation (hilly areas) | 110-128 dB |
| No. of photo detectors used | 8 |
| Range of Link | 500 m |
| Operating signal wavelength, | 1550 nm |
| Transmitter lens diameter | 100 cm |
| Receiver aperture diameter | 50 cm |

Let R be the rain rate in mm/h, the specific attenuation of optical wireless link is given by [18]:

$$\alpha_{rain}=1.076\ R^{0.67}\ \text{dB/km} \quad (2)$$

If S is the snow rate in mm/h then specific attenuation in dB/km is given by [19] as:

$$\alpha_{dry\ snow} = aS^{b}\ \text{dB/km} \quad (3)$$

If λ is the wavelength, the parameters a and b for dry snow is given as the following:

a =5.42×10−4λ+5.495876, b =1.38

The specific attenuation in the case of wet snow can be expressed as the following formula [20]:

$$\alpha_{wet\ snow} = h\ S^{g}\ \text{dB/km} \quad (4)$$

The parameters h and g for wet snow are as, h =1.023×10−4λ+3.7855466, g =0.72

The amount of received power is proportional to the amount of power transmitted and the area of the collection aperture but inversely proportional to the square of the beam divergence and the square of the link range. It is also inversely proportional to the exponential of the product of atmospheric attenuation coefficient times the link range [21-22].

$$P_{received} = P_{received} * \frac{d_2^2}{[d1+(D*R)]^2} * 10^{(-a*r/10)} \quad (5)$$

Where, P = power,

d1 = transmit aperture diameter (m),

d2 = receive aperture diameter (m),

D = beam divergence (mrad), R = range (km),

a = atmospheric attenuation factor (dB/km).

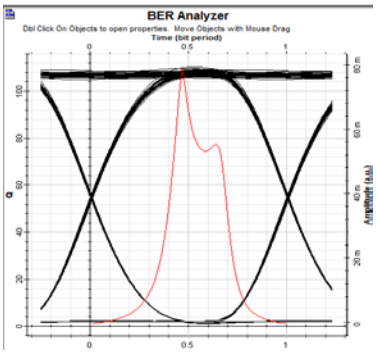
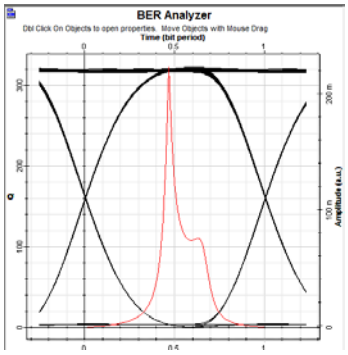
Also, the bit error rate can also be expressed in terms of signal to noise ratio (SNR) as:

$$BER = \frac{2}{\pi.SNR}.\exp\left(-\frac{SNR}{8}\right) \quad (6)$$

## III. RESULTS AND PERFORMANCE EVALUATION

The model has been investigated to show the weather effects on the transmission and overall performance on free space communication in hilly and plain areas by using single photo detector and an array of photodetectors. FSO system with link range 500m operating at a wavelength of 1550 nm is considered such that it can show useful results over a wide range of weather conditions. The values of attenuation effecting information signal considered in table 2 are taken from [6].

TABLE II. COMPARISON OF OUTPUTS OF BER ANALYSER IN PLAIN AREAS

| Attenuation | Weather conditions | Output using single receiver | Output using array of receivers |
|---|---|---|---|
| **0.43 dB** | **Clear air** |  |  |

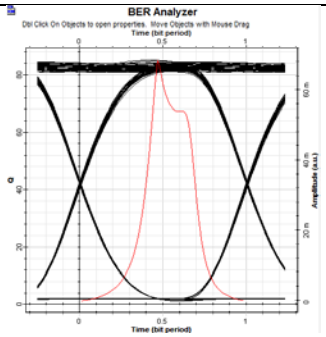
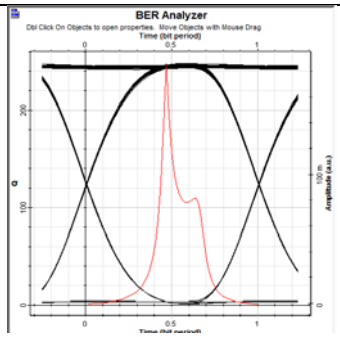
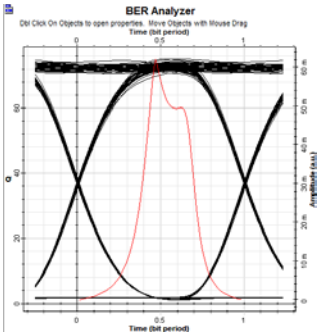
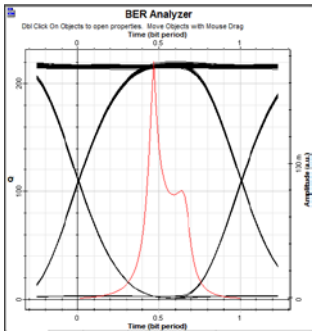
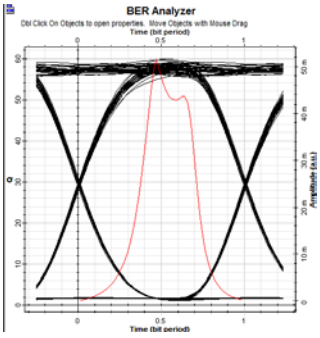
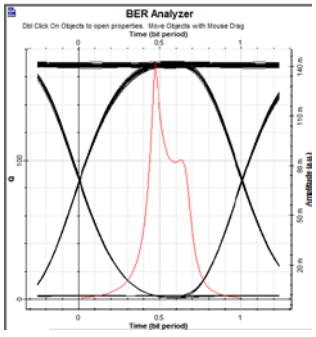
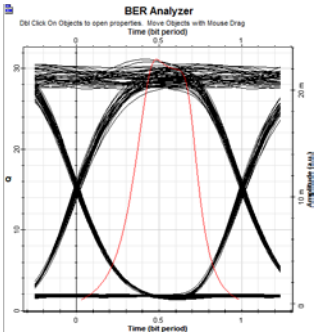
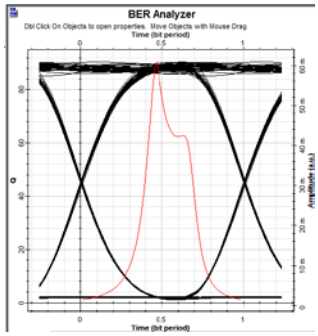
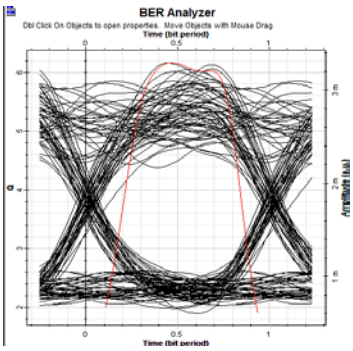
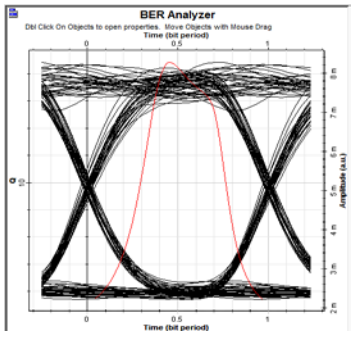
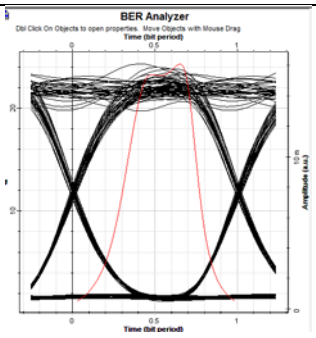
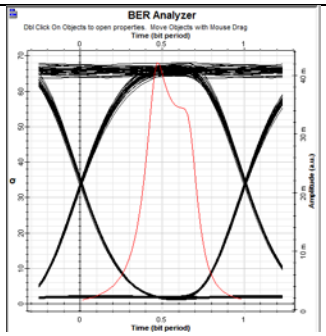
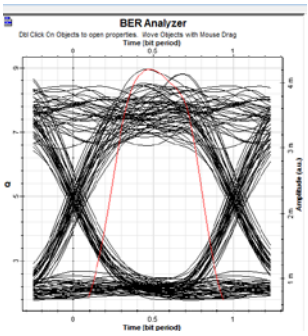
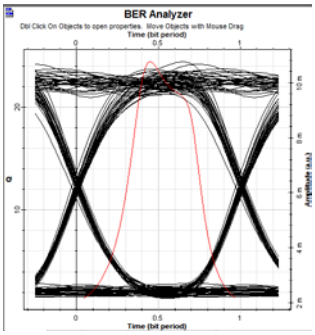
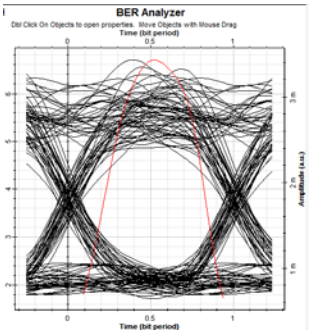
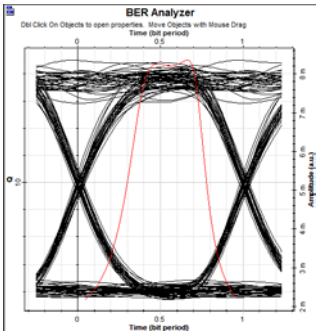
| | | | |
|---|---|---|---|
| **4.2 dB** | **Haze** |  |  |
| **5.8 dB** | **Moderate rain** |  |  |
| **9.2 dB** | **Heavy rain** |  |  |
| **20 dB** | **Light fog** |  |  |
| **42.2 dB** | **Moderate fog** |  |  |

Table 2 shows the output of BER analyser for FSO link located in plain areas. The comparison is showing the improvement in the output when more than one receivers are inserted in the link. It can be seen than width and height of eye have increased and curve of Q factor has become sharp after using array of receivers at different values of attenuation. Figures in the table have proved that received signal power decreases with increasing atmospheric attenuation for in the presence of bad weather effects. But using an array of photo detectors has presented the highest received signal power compared to a single receiver. From the above discussion it is clear that array of receivers is giving better results in these weather conditions also. So, this technique can be further implied on the weather conditions of hilly areas too.

TABLE III. COMPARISON OF OUTPUTS OF BER ANALYSER IN HILLY AREAS

| Attenuation | Weather conditions | Output using single receiver | Output using array of receivers |
|---|---|---|---|
| **110 dB** | **Wet snow** |  |  |
| **125 dB** | **Heavy fog** |  |  |
| **128 dB** | **Dry snow** |  |  |

Similarly, Table 3 is the comparison of the output of BER analyzer in hilly areas. The values of attenuation of the signal effecting communication considered in table 3 in taken from [7]. The value of attenuation in hilly areas is very large as compared to plain areas.

So using an array of receivers in hilly areas gives very effective results by improving values of simulation parameters far better than optimum values.

The figures show that even in high attenuation conditions FSO systems can be deployed reliably by slightly modifying the conventional FSO systems.

The system operation characteristics have been plotted under varying weather conditions using different simulation parameters. A brief comparison is made to show the improvement in all simulation parameters using more than one receiver for receiving the information signal.

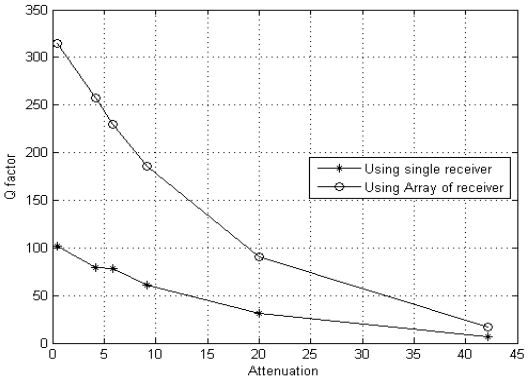



Fig. 2. Comparison of Q factor in weather conditions of plain areas using single and array of Photo detectors

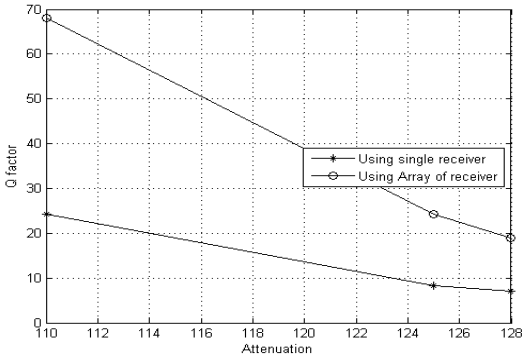



Fig. 3. Comparison of Q factor in weather conditions of hilly areas using single and array of Photo detectors

The graphs in figure 2 and 3 show the comparison of various values of Q factor of system model, as a result of using single receiver and array of receivers at different weather conditions of plain areas and hilly areas. The curve in figure 2 shows that, at transmitter power of 1mW there is a large difference in output signals of both the cases. Below 20 dB attenuation the technique employed is improving the quality of the received signal in huge ratio. After 20 dB the results start decreasing linearly with increasing attenuation and for the values above 40 dB there are not many variations in the results.

In figure 3 also, the results have assured that the Q factor of the system increases with increasing number of photo detectors in the link. It is observed that the quality of a signal received using an array of receivers is much better than the quality of the signal using single detector under different attenuation conditions. There is a significant difference in the value of Q factor at 128 dB, thus improving the reliability of the communication in bad weather conditions.

Figures 4 and 5 have demonstrated that the width of eye increases after the use of array of receivers in the link in both plain as well as hilly areas because the resultant signal is chosen such that it have maximum signal to noise ratio thus the opening of eye is more as compared to signal detected with single photo detector.

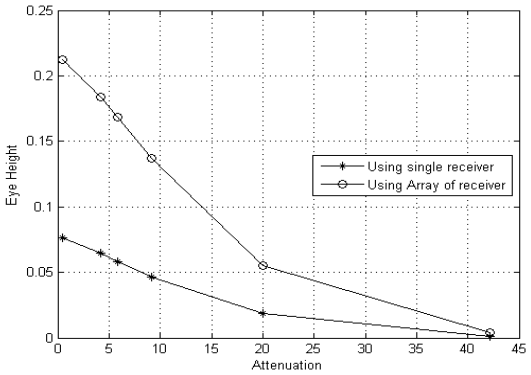



Fig. 4. Comparison of Eye height in weather conditions of plain areas using single and array of Photo detectors

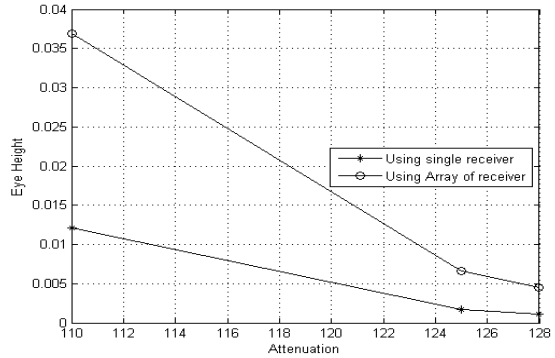



Fig. 5. Comparison of eye height in weather conditions of hilly areas using single and array of Photo detectors

In figure 4, there is a significant difference in the values before 20 dB attenuation but after 20 dB it is almost same. But figure 5 illustrates that in hilly areas using this technique is quite helpful in removing the noise and jitter from the signal.

At 128 dB, the value of eye height in the conventional system is 0.0012 which is improved to 0.0036 using array of receivers.

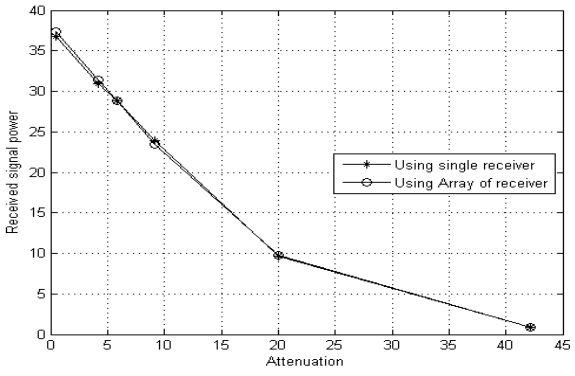



Fig. 6. Comparison of received signal power in weather conditions of plain areas using single and array of Photo detectors

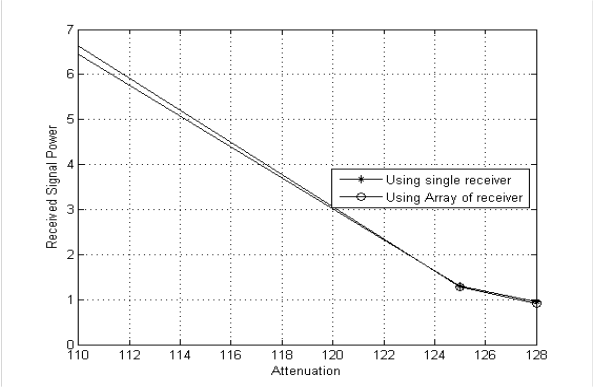



Fig. 7. Comparison of received signal power in weather conditions of hilly areas using single and array of Photo detectors

Figures 6 and 7 are indicating that with an increase in a number of photo detectors at the receiver end, there is a slight increase in received signal power because the multiple numbers of detectors are detecting the same signal independently. The final signal obtained is the maximum value of all detected signals.

Finally, a comprehensive comparison has been presented in an organized manner among the performances of links having single photo detector and multiple photo detectors in tabular form. Table 4 is illustrating the values of simulation parameters analyzed in plain areas with transmission power one mW. It can be clearly visualized that in the case of an array of photo detectors, there is a significance decrease in bit error rate. Also the values of Q factor, the height of eye and received signal power increases with the introduction of more than one photo detectors in the link.

TABLE IV. EVALUATION OF FSO LINK IN PLAIN AREAS WITH POWER 1 MW

| | | Max. Q factor | Min. BER | Eye height | Received signal power(mW) |
|---|---|---|---|---|---|
| **Single photo detector** | Clear air (0.43 | 102.192 | 0 | 0.076034 | 36.7936 |
| | Haze (4.2dB) | 79.0142 | 0 | 0.0647861 | 30.911234 |
| | Moderate rain (5.8dB) | 78.2872 | 0 | 0.0583843 | 28.845105 |
| | Heavy rain (9.2dB) | 60.5079 | 0 | 0.0461851 | 23.91591 |
| | Light fog (20 dB) | 30.7312 | 7.34165e-208 | 0.0184382 | 9.69035 |
| | Moderate fog (42.2 dB) | 6.93705 | 1.83037e-12 | 0.00107621 | 0.90622 |
| **Array of photo detectors** | Clear air (0.43 | 314.806 | 0 | 0.21236 | 37.388885 |
| | Haze (4.2dB) | 257.198 | 0 | 0.183567 | 31.362605 |
| | Moderate rain (5.8dB) | 229.584 | 0 | 0.168624 | 28.845105 |
| | Heavy rain (9.2dB) | 186.242 | 0 | 0.136811 | 23.480854 |
| | Light fog (20 dB) | 90.6124 | 0 | 0.055339 | 9.8212256 |
| | Moderate fog (42.2 dB) | 17.0842 | 8.87462e-66 | 0.00437094 | 0.920583 |

TABLE V. EVALUATION OF FSO LINK IN HILLY AREAS WITH POWER 20W

| | | Max. Q factor | Min. BER | Eye height | Received signal power(mW) |
|---|---|---|---|---|---|
| **Single photo detector** | Wet snow (110 dB) | 24.2834 | 1.12641e-130 | 0.0121424 | 6.4458854 |
| | Heavy fog (125 dB) | 8.19081 | 1.13342e-16 | 0.0017453 | 1.3038745 |
| | Dry snow (128 dB) | 7.04754 | 8.40774e-13 | 0.0011057 | 0.9438791 |
| **Array of photo detectors** | Wet snow (110 dB) | 68.0498 | 0 | 0.0369363 | 6.6352527 |
| | Heavy fog (125 dB) | 24.1941 | 2.00845e-104 | 0.0065796 | 1.2833537 |
| | Dry snow (128 dB) | 19.0077 | 8.40774e-13 | 0.0044685 | 0.9144441 |

The performance of link in hilly areas for different cases is summarized in Table 5. It shows the values of various parameters in hilly areas, where attenuation is very high as compared to plain areas. So, the power used for transmission of signal in hilly areas is 20W. The quality of received signal is improving as the values of simulation parameters are attaining optimum values.

It can be clearly visualized from the above results that as the number of photo detectors are increased at the receiver end; it is possible to obtain performance that may not be possible by using other techniques like increasing the transmitter power or aperture averaging.

## IV. CONCLUSION

In this paper, the performance of the free space optical communication systems under the effects of bad weather conditions especially for heavy rain, fog, dry and wet snow has been analyzed. The performance of link is investigated for these conditions and is further improved by the technique of using an array of receivers. The results reveal that use of array of receivers is advantageous over that of a single receiver in FSO link for detecting the signal more accurately as the quality factor of received signal is improved by approximately 28% in all the cases under consideration. With further research and development, communication in FSO system can also be enhanced at higher data rate over a longer link range under all weather conditions and atmospheric turbulences to enhance the usage of free space optics technology.

## REFERENCES

[1] F. Nadeem, V. Kvicera, M. S. Awan, E. Leitgeb, S. S. Muhammad, G. Kandus, "Weather Effects on Hybrid FSO/RF Communication Link," IEEE J. On selected areas in Comm., Vol.27, No. 9, pp. 1687-1696, Dec. 2009

[2] H.H. Refai, J.J. Sluss Jr., H.H. Refai, "Comparitive study of performance of analog fiber optic links versus free- space optical links" J. of Optical Engineering, Vol. 45, No.2, Feb. 2006

[3] S. Saini, A. Gupta," Investigation to find optimal modulation format for low power inter-satellite optical wireless communication (LP-IsOWC)", Eleventh International conference on Wireless and Optical Communication Networks, pp. 1-4, 2014

[4] V. Sarup, A. Gupta" Performance analysis of an ultra high capacity 1 Tbps DWDM-RoF system for very narrow channel spacing", Eleventh International conference on Wireless and Optical Communication Networks, pp. 1-5, 2014

[5] N. Cvijetic, Q. Dayou, Y. Jianjun, H. Yue-Kai, W. Ting, "100 Gb/s per channel free space optical transmission with coherent detection and MIMO processing,"in Proc. ECOC, pp.1-2, 2009

[6] B. He, R. Schober," Bit-Interleaved Coded Modulation for Hybrid RF/FSO Systems", IEEE Trans. on Comm., Vol. 57, No. 12,pp. 3753-3763, Dec. 2009

[7] A.N.Z. Rashed, M.M.E. El-Halawany, "Transmission characteristics evaluation under bad weather conditions in optical wireless links with different optical transmission windows", Wireless Personal Communications, Springer, Vol. 71, No. 2, pp. 1577-1595, July 2013

[8] A. K. Majumdar and J. C. Riclkin, "Free-space laser communications," in Principles and Advances. New York, NY, USA: Springer-Verlag, 2007.

[9] M. M. Ibrahim and A. M. Ibrahim, "Performance analysis of optical receivers with space diversity reception," IEE Proc. Commun., vol. 143, no. 6, pp. 369–372, Dec. 1996.

[10] S. Saini, A. Gupta," Modeling and Performance Analysis of DWDM Based 100 Gbps Low Power Inter-satellite Optical Wireless Communication (LP-IsOWC) System", SOP transactions on signal processing, Vol. 2, No. 1, 2015

[11] M. Khalighi, N. Schwartz, N. Aitamer, and S. Bourennane, "Fading reduction by aperture averaging and spatial diversity in optical wireless systems," J. Opt. Commun. Netw., vol. 1, no. 6, pp. 580–593, Nov. 2009.

[12] H. G. Sandalidis, T. A. Tsiftsis, G. K. Karagiannidis, and M. Uysal, "BER performance of FSO links over strong atmospheric turbulence channels with pointing errors," IEEE Commun. Lett., vol. 12, no. 1, pp. 44–46, Jan. 2008.

[13] A. A. Farid and S. Hranilovic, "Diversity gain and outage probability for MIMO free-space optical links with misalignment," IEEE Trans. Commun., vol. 60, no. 2, pp. 479–487, Feb. 2012.

[14] P. Kaur, V. K. Jain, and S. Kar," Performance Analysis of FSO Array Receivers in Presence of Atmospheric Turbulence", IEEE Photonics Technology Letter, Vol. 26, No. 12, pp. 1165-1168, June, 2014

[15] A. Gupta , R S Kaler, H Singh, "Investigation of OBS assembly technique based on various scheduling techniques for maximizing throughput", Optik-International Journal for Light and Electron Optics, 124(9), pp. 840–844, 2013

[16] A Gupta, RS Kaler, H Singh," An inimitable scheduling technique for optical burst switched networks", Optik-International Journal for Light and Electron Optics, vol. 124, no. 8, pp.689-692, 2013

[17] A. K.Majumdar& J. C. Ricklin, "Free space laser communications, principles and advantages", Berlin: Springer, 2008

[18] S. Muhammad, B. Flecker, E. Leitgeb, & M. Gebhart, "Characterization of fog attenuation in terrestrial free space optical links" Journal of Optical Engineering, vol. 46, no. 4, pp.1–9, 2007

[19] Akiba, M., Ogawa, K., Wakamori, K., Kodate, K., & Ito, S, "Measurement and simulation of the effect of snowfall on free space optical propagation", Journal of Applied Optics, Vol. 47, No. 31, pp. 5736– 5743, 2008

[20] Betti, S., Carrozzo, V., & Duca, E. "Over stratospheric altitude optical free space links: System performance evaluation transparent optical networks". In ICTON, International Conference on Transparent Optical Networks, July 2007

[21] S. Bloom, E. Korevaar, J. Schuster and H. Willebrand, "Understanding the performance of free-space optics", Journal of Optical Networking, Vol. 2, No. 6, June 2003.

[22] M. Jeong, J. Lee, S. Kim, S. Namgung, J. Lee, M. Cho, S. Huh, Y. Ahn, J. Cho, and J. Lee, "Weather-Insensitive Optical Free-Space Communication Using Gain-Saturated Optical Fiber Amplifiers", IEEE Journal of Lightwave Technology, Vol. 23, No. 12, 2005.

[23] A. E. N. A Mohamed, H. A. Sharshar, A. N. Z. Rashed & A. El-Nabawy, " Integrated service quality enhancement of wireless optical communication systems for long haul transmission distances" Canadian Journal on Electrical and Electronics Engineering, vol. 2, no. 12, pp.-557–570, 2011

[24] A Gupta, H Singh, J Kumar," A novel approach to reduce packet loss in OBS networks" International Journal of Computer Applications, vol. 58, no. 3, 2012

[25] N. Vinayak, A gupta," Comparative analysis of WDM system using Cascaded amplifiers in Optical wireless channel over a distance of 10000 km", SOP Transactions on Signal Processing, Vol. 1, No. 1, pp.25-32, 2014

# Packrat Parsing: A Literature Review

Manish M. Goswami
Research Scholar,
Dept. of Computer Science and Engineering,
G.H.Raisoni College of Engineering,
Nagpur, India

Dr. M.M. Raghuwanshi,
Professor,
Department of Computer Technology,
YCCE,
Nagpur, India

Dr. Latesh Malik,
Professor,
Dept. of CSE,
G.H.Raisoni College of Engg.,
Nagpur, India

***Abstract*—Packrat parsing is recently introduced technique based upon expression grammar. This parsing approach uses memoization and ensures a guarantee of linear parse time by avoiding redundant function calls by using memoization. This paper studies the progress made in packrat parsing till date and discusses the approaches to tackle this parsing process efficiently. In addition to this, other issues such as left recursion, error reporting also seems to be associated with this type of parsing approach and discussed here the efforts attempted by researchers to address this issue. This paper, therefore, presents a state of the art review of packrat parsing so that researchers can use this for further development of technology in an efficient manner.**

***Keywords*—Parsing Expression Grammar; Packrat Parsing; Memoization; Backtracking**

## I. INTRODUCTION

Parsing consists of two processes: lexical analysis and parsing. The job of the lexical analysis is to break down the input text (string) into smaller parts, called tokens. The lexical analyzer then sends these tokens to the parser in sequence. During parsing, the parser takes the help of a grammar to decide whether to accept the input string or reject .i.e. whether it is a subset of the accepting language or not. A set of grammar rules or productions is used to define the language of grammar. Each production can then, in turn, compose of several different alternative productions. These productions guide the parser throughout the parsing to determine whether to accept the input string or to reject it. Top-down parsing is a parsing strategy that attempts left-to-right leftmost derivation (LL) for the input string. This can be achieved with prediction, backtracking or a combination of the two. LL(k) top-down parser makes its decisions based on lookahead, where the parser attempts to ”look ahead” k number of symbols of the input string. A top-down parser that uses backtracking instead evaluates each production and its choices in turn; if a choice/production fails the parser backtracks on the input string and evaluates the next choice/production, if the choice/production succeeds the parser merely continues. The bottom-up parsing is a parsing method that instead attempts to perform a left-to-right rightmost derivation (LR) in reverse of the input string. Shift-reduce parsing is widely used bottom-up parsing technique. A shift-reduce parser uses two different actions during parsing: shift and reduce. A shift action takes a number of symbols from the input string and places them on a stack. The reduce action reduces the symbols on the stack based on finding a matching grammar production for the symbols. The decisions regarding whether to shift or reduce are done based on lookahead. Several different parsing techniques have been developed over the years, both for parsing ambiguous and unambiguous grammars. One of the latest is packrat parsing [5]. Packrat parsing is based upon a top-down recursive descent parsing approach with memoization that guarantees linear parse time. Memoization employed in the packrat parsing eliminates disadvantage of conventional top-down backtracking algorithms which suffer from exponential parsing time in the worst case. This exponential runtime is due to performing redundant evaluations caused by backtracking. Packrat parsers avoid this by storing all of the evaluated results to be used for future backtracking eliminating redundant computations. This storing technique is called memoization which ensures guaranteed linear parsing time for packrat parsers. The memory consumption for conventional parse algorithms is linear to the size of the maximum recursion depth occurring during the parse. In the worst case it can be the same as the size of the input string. In packrat parsing, the memory consumption for a packrat parser is linearly proportional to the size of the input string. Packrat parsing is based upon parsing expression grammars (PEGs) which have the property of always producing unambiguous grammars. It has been proven that all LL(k) and LR(k) grammars can be rewritten into a PEG[7]. Thus, packrat parsing is able to parse all context-free grammars. In fact, it can even parse some grammars that are non-context-free [7].

Another characteristic of packrat parsing is that it is scannerless i.e. a separate lexical analyzer is not needed. In packrat parsers, they are both integrated into the same tool, as opposed to the Lex[10]/Yacc[9] approach where Lex is used for the lexical analysis and Yacc for parsing phase of the compiler.The founding work for packrat parsing was carried out in 1970 by A. Birman et. al.[4]. Birman introduced a schema called the TMG recognition schema (TS). Birman’s work was later refined by A. Aho and J. Ullman et. al.[2], and renamed into generalized top-down parsing language (GTDPL). This was the first top-down parsing algorithm that was deterministic and used backtracking. Due to deterministic nature of resulting grammar they discovered that the parsing results could be saved in a table to avoid redundant computations. However, this approach was never put into practice, due to the limited amount of main memory in computers at that time [10,14]. Another characteristic of GTDPL is that it can express any LL(k) and LR(k) language, and even some non-context-free languages[2,3]. Rest of the paper consists of introduction to Parsing Expression grammar

along with discussion on its properties followed by the work carried out by researchers in this area. Finally the paper focuses upon the open problems in packrat parsing and concluded with future work.

## II. PARSING EXPRESSION GRAMMAR

As an extension to GTDPL and TS, Bryan Ford introduced PEGs [7]. CFGs (which were introduced mainly for usage with natural language [7]) may be ambiguous and thereby either (1) produce multiple parse tree's, which is not necessary due to only one is needed, and (2) produce a heuristically chosen one, which might not even be correct[8]. However, one of the characteristics of PEGs is that they are by definition unambiguous and thereby provides a good match for machine-oriented languages (since programming languages supposed to be deterministic). It is also shown that PEGs, similar to GTDPL and TS, can express all LL(k) and LR(k) languages, and that they can be parsed in linear time with the memoization technique [7].

### A. Definitions and Operators

PEGs, as defined in [7], are a set of productions of the form A <- e where A is a nonterminal and e is a parsing expression. The parsing expressions denote how a string can be parsed. By matching a parsing expression e with a string s, e indicates success or failure. In case of a success, the matching part of s is consumed. If a failure is returned, s is matched against the next parsing expression. Together, all productions form the accepting language of the grammar. The following operators are available in PEG productions: [7,14]

**Ordered choice:** $e_1/.../e_n$, expression e1,...,en is evaluated in this order, to the text ahead, until one of them succeeds and possibly consumes some text. If one of the expressions succeeded, indicate success. Otherwise indicate failure and input is not consumed.

**Sequence:** $e_1,...,e_n$, expressions $e_1,...,e_n$, is evaluated in this order, to consume consecutive portions of the text ahead, as long as they succeed. If all succeeded, success is indicated. Otherwise indicate failure and input is not consumed.

**And predicate:** &e, if expression e matches the text ahead; indicate success otherwise indicate failure. Text is not consumed.

**Not predicate:**!e, if expression e matches the text ahead, failure is indicated; otherwise, indicate success. Do not consume any text.

**One or more:** e+, expression e is repeatedly applied to match the text ahead, as long as it succeeds. Matched Text is consumed if any and success is indicated if there is at least one match. Otherwise failure is indicated.

**Zero or more:** e*, As long as expression e matches text ahead it is applied repeatedly and consumed the matched text (if any). Always report success.

**Zero or one:** e?, if expression e matches the text ahead, consume it. Always report success.

**Character class:** [s], character ahead is consumed if it appears in the sting s and success is indicated. Otherwise failure is indicated.

**Character range:** [c1- c2], if the character ahead is one from the range c1 through c2, consume it and indicate success. Otherwise indicate failure.

**String:**'s', if the text ahead is the string s, consume it and success is indicated. Otherwise failure is indicated.

**Any character:** (dot), if there is a character ahead, consume it and indicate success. Otherwise (that is, at the end of input) indicate failure.

### B. Ambiguity

The unambiguousness of a PEG comes from the ordered choice property. The choices in CFGs are symmetric, i.e., the choices need not be checked in any specific order. However, the choices in a PEGs are asymmetric, i.e., the ordering of the choices determines in which order they are tested. For a PEG the first expression that matches are always chosen. This means that a production such as A<-a/aa is perfectly valid and unambiguous. However, it only accepts the language {a} on the contrary, a CFG production A->a|aa is ambiguous but accepts the language {a,aa}.

A traditional example that is hard to express with the use of CFGs is the dangling else problem. Consider the following if-statement:

*if cond then if cond then statement else statement*

This statement can be matched in the following two ways:

if cond then (if cond then statement else statement)
if cond then (if cond then statement) else statement

If the intended matching is the former of the two (in fact, this is how it is done in the programming language C[8]) then the following PEG production is sufficient:

*Stmt<- 'if' Cond 'then' Stmt 'else' Stmt*

*/'if' Cond 'then' Stmt /...*

Note: Matching the outermost if with the else-clause is believed not to be possible with a PEG. However, no source to either prove or contradict this statement was found.

Discovering if a CFG production is ambiguous is sometimes a non-trivial task. Similarly, choosing the ordering of two expressions in a PEG production without affecting the accepting language is not always straightforward [7].

### C. Left Recursion

Left recursion is when a grammar production refers to itself as its left-most element, either directly or indirectly. Similar to the conventional LL(k) parsing methods, left recursion proves to be an issue for PEGs, and therefore a problem also for packrat parsing [10,7]. Consider the following alteration of the production:

A<-Aa / a

For a CFG, this modification is not a problem. However, for a PEG, parsing of nonterminal A requires testing that A matches, which requires testing that A matches etc, producing

an infinite recursion. However, it was early discovered that a left-recursive production can always be rewritten into an equivalent right-recursive production [1], and thus making it manageable for a packrat parser. However, if there is indirect left recursion involved, the rewriting process may become fairly complex.

*D. Syntatic Predicates*

PEGs allow the use of the syntactic predicates! and &. Consider the following grammar production:

A <- !B C

Every time this production is invoked it needs to establish if the input string matches a B and if it does, signal a failure. If there is no match the original input string is compared with the nonterminal C. This ability to "look ahead" an arbitrary amount of characters combined with the selective backtracking gives packrat parsers unlimited lookahead [10, 6, 7, 13].

*E. Memoization*

The introduction of memoization was treated as a machine learning method in 1968 by D. Michie et. al. [11]. By storing calculated results, the machine "learned" it. The next time it was asked for the same result, the machine merely"remembered" it by looking up the previously stored result. The storage mechanism used was a stack. This makes the look-up process become linear. However, insertions of results are constant; they are merely pushed on top of the stack. In packrat parsing, the results are instead stored in a matrix or similar data structure that provides constant time look-ups (when the location of the result is already known) and insertions [10]. For every encountered production this matrix is consulted; if the production has already occurred once the result is thereby already in the matrix and merely needs to be returned; if not, the production is evaluated and the result is both inserted into the matrix and returned. Conventional recursive descent parsers that use backtracking may experience exponential parsing time in the worst case. This is due to redundant calculations of previously computed results caused by backtracking. However, memoization avoids this problem due to the fact that the result only needs to be evaluated once. This gives packrat parsing a linear parsing time in relation to the length of the input string (given that the access and insertion operations in the matrix are done in constant time). Let us look at the following trivial PEG, taken from [10]:

Additive <- Multitive '+' Additive / Multitive

Multitive <- Primary '*' Multitive / Primary

Primary <- '(' Additive ')' / Decimal

Decimal <- [0-9]

With this grammar and the input string 2*(3+4) the following memoization matrix can be produced:

The columns correspond to each position of the input string; the rows correspond to each of the parsing procedures. To make it clear that the rows are in fact procedures they have been given a prefix 'p'. Each cell contains either a number that represents how much of the input string that have been consumed by a previous call to the procedure, or the cell

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| pAdditive | 7 | -1 | 5 | 3 | -1 | 1 | -1 | -1 |
| pMultitive | 7 | -1 | 5 | 1 | -1 | 1 | -1 | -1 |
| pPrimary | 1 | -1 | 5 | 1 | -1 | 1 | -1 | -1 |
| pDecimal | 1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 |
| input | '2' | '*' | '(' | '3' | '+' | '4' | ')' | $ |

A Matrix Containing the Parsing Results of the Input String 2*(3+4) contains a '-1' which indicates a failed evaluation. For instance, if backtracking occurs at input position four (where the number '3' is present) and procedure pAdditive is called, the parser first checks if a previous computation is stored in the storage matrix. In this case the number 3 is stored and thereby the parser immediately knows that it can advance three steps on the input string and end up at the seventh character of the input string. If the stored value is '-1' the parser knows that a previous call resulted in a failed parse and can thus avoid continuing with the procedure call and instead return a failure response to the calling function. This illustrates how redundant computations and thereby also potential exponential parsing times are avoided with the help of memoization. Calculating the whole matrix in Table1 would be unnecessary since many of the cells are not needed. The idea behind packrat parsing is not to evaluate all of the cells in the parsing matrix, only the results that are needed [10]. This effectively reduces the amount of memory space required during parsing.

*F. Scannerless*

Conventional parsing methods are usually divided into two phases: the lexical analysis phase and the parsing phase. The tokenization phase is called lexer, lexical analyzer, tokenizer or scanner. The lexical analysis splits the input string into tokens which hopefully corresponds to the permitted terminals of the grammar. This lexical analysis is important for conventional parsers due to their inability to refer to nonterminals for lookahead decisions [6]. Thus, the parser treats the tokens acquired from the lexical analysis as if they were terminals.

Packrat parsers can on the other hand be scannerless, which means that it requires no lexical analysis. When a scannerless packrat parser evaluates different alternatives, it can rely on already evaluated results. This effectively makes a packrat parser able to use both terminals and nonterminals during lookahead [10, 6]. Large parts of the code base of programs may consist of white spaces, comments and other irrelevant information that is not needed for the semantic analysis. A lexical analyzer can effectively disregard such information by simply opting not to create any tokens for them, thus no specific productions for white spaces and/or comments need to be included in the grammar specification of the parser. For a packrat parser that does not use a lexical analyzer, however, this is not the case. A packrat parser that uses no lexical analyzer, the white spaces and comments need to be incorporated into the productions of the grammar.

As previously mentioned, the conventional parsers treat the created tokens given by the lexical analyzer as if they were terminals, and between each token the lexical analyzer disregards any white spaces or comments. To achieve this

effect for a packrat parser, a production to manage white space or comments can be created and used after each terminal symbol of the grammar. For instance, the grammar for recognizing arithmetic expression altered in the following way to be able to correctly handle white spaces inside an arithmetic expression:

Additive<- Multitive '+' Spaces Additive / Multitive
Multitive<-Primary '*' Spaces Multitive / Primary
Primary <- '(' Spaces Additive ')' Spaces /Decimal
Decimal <- [0-9] Spaces
Spaces <- (' ' / ' \t' / ' \n' / ' \r')*

This grammar is now able to parse an arithmetic expression such as: 2 * (3+ 4).

## III. Literature Survey

PEGs are a recently introduced technique for describing grammars by Ford in [5] with implementation of the packrat parser. Theory is based upon strong foundations. Ford [18] showed how PRGs can be reduced to TDPLs long back in the 1970s. It was shown by Roman[24] that primitive recursive-descent parser with limited backtracking alongwith integrated lexing can be used for parsing Java 1.5 where requirement is of moderate performance. PEG is not good as a language specification tool as shown in [25]. The characteristic of a specification is that what it specifies is clearly to be seen. But this is, unfortunately, not valid for PEG. Further it gives reasonable performance when C grammar is slightly modified and also in [16] he studied that classical properties like FIRST and FOLLOW where he demonstrated those can be redefined for PEG and can be obtained even for a large grammar. FIRST and FOLLOW are used to define conditions for choice and iteration that are similar to the classical LL(1) conditions, although they have a different structure and semantics. This is different from classical properties like FIRST and FOLLOW where letters are terminal expressions, which may mean sets of letters, or strings. Checking these conditions gives an idea of useful information like the absence of reprocessing or language hiding which is helpful in locating places that need further examination. The properties FIRST and FOLLOW are kind of upper bounds, and conditions using them are sufficient, but not necessary which may results in false warnings In [17] a virtual parsing machine approach is proposed for implementing PEG which is can be applied to pattern matching. Each PEG is converted directly into its equivalent corresponding program Virtual parsing machine then using scripting language excutes the translated program. Creation and composition of new programs are done on fly.

In [7] Robert grimm parsing technique made practical for object-oriented languages. This parser generator employs simpler grammar specifications. Error reporting is also made easy by this parser generator and shown to be better performing parsers through aggressive optimizations.

In [19] cut operator was introduced to parsing expression grammars (PEGs) and when applied to PEG on which packrat parsing is based. Disadvantage is largely addressed with this approach. Concept of cut operator was borrowed from Prolog [6]. It introduces degree of controlling backtracking. An efficient packrat parser can be developed avoiding unnecessary space for memorization by inserting cut operators into a PEG grammar at appropriate places. To show effectiveness and usefulness of cut operators, a packrat parser generator called Yapp was implemented and used. It accepts Parsing Expression Grammar marked with cut operator. The experimental evaluations showed that the packrat parsers generated using grammars with cut operators inserted can parse Java programs and subset of XML files in mostly constant space, unlike conventional packrat parsers. In [12] automatic insertion of cut operators was proposed that achieves the same effect. In these methods, a statistical analysis is made of a PEG grammar by parser generator in order to find the places where the parser generator can insert cut operators without changing the meaning of the grammar and cut operators are inserted at these identified points. Definite clause grammar rules and memoing can be a possible combination for implementation of packrat parser as shown in [20] .Further it points out that packrat parsing may degrade its performance over plain recursive descent with backtracking, but memoing the recognizers of just one or two nonterminals can sometimes give reasonable performance.

Warth [21] tweaked memoization approach used by packrat parser because of which left-recursion even indirectly or mutually was supported. But some experiments were conducted out to show that this is not the case for typical uses of left recursion. In [8] Coq formalization of the theory of PEGs is proposed and with this as a foundation a formally verified parser interpreter for PEGs, TRX is developed. This gives rise to writing a PEG, together with its semantic actions, in Coq and then a parser can be extracted from it a parser with total correctness guarantees. This ensures that the parser will terminate on all kind of inputs and produces output as a parsing results correct with respect to the semantics of PEGs.

In [27] concept of elastic sliding window is used and it is based upon the observation of worst longest backtrack length. Particularly author noted that if a window in the form of small memorization table slides and covers the longest backtrack then redundant calls are avoided since the storage is sufficient enough to store all the results. Practically, it is difficult to get the longest backtrack before parsing as it is runtime entity. Here window is approximated from empirical investigation and if needed may be expanded during parsing.

[28] introduces derivative parsing with memorized approach algorithm for recognition of PEG. Main problem in this algorithm since derivative parsing attempts all possible parses concurrently is to identify which constructs exactly in the current parse tree can match against or consume the current character. This problem is solved by using concept of a backtracking generation (or generation) as a means to take into account for backtracking choices in the process of parsing. Execution of the algorithm is found to be in worst case quadratic time and cubic space. However, it is stressed in this paper that due to the limited amount of backtracking and recursion in grammars when put in practical use and input, practical performance may be nearer to linear time and constant space and requires experimental validation for the same which is in progress. Table II summarizes the comparative study of major packrat parser generators.

TABLE I. COMPARISON OF VARIOUS PACKRAT PARSER GENERATORS

| Name of Parser | Language used for implementation | Working Principle | Memory Utilization | Execution Time |
|---|---|---|---|---|
| YAPP | Object oriented Language Java | Optimized Packrat Parser with CUT operator | Memory requirement has been cut down as cut operator reduces redundant calls | Moderate amount of parsing time required |
| RATS | Object oriented Language Java | Packrat Parser with some aggressive optimization | Less memory space as some aggressive optimizations used | Moderate amount of parsing time required |
| Mouse | Object oriented Language Java | Straightforward Recursive descent Parser implemented using PEG with no memorization support | Least amount of storage used | Amount of parsing time is highest among all the parsers as repeated backtracking is not avoided |
| PAPPY | Functional Programming Language Huskell | Basic Pakrat Parser with memoization | Use of significant storage space | Moderate amount of parsing time required |
| Nez | Object oriented Language Java | Packrat Parser implemented with elastic sliding window concept | Significant and least amount of storage is used among all packrat parser which use memoization | Moderate amount of parsing time required |

## IV. OPEN PROBLEMS IN PACKRAT PARSING

### A. Memoization

One of the drawbacks with using packrat parsing is the additional memory consumption when compared with conventional parsing techniques. A tabular approach where the whole m*n matrix is evaluated would require Θ(mn) space, where m is the amount of nonterminals and n is the size of the input string. However, for a packrat parser, only the required cells in the matrix are evaluated, and these cells are directly related to the input string and not the nonterminals of the grammar [22]. In other words, adding more productions to the grammar may not necessarily increase the storage consumption while increasing the size of the input string will always increase the memory consumption. This makes the required size of the memoization matrix for a packrat parser be proportional only to the size of the input string, thus O(n). Even if the space consumption is upper-bounded by the input string and can therefore be written as O(n) there is a ”hidden constant multiple” of n [22].This is because there can be more than n elements in the produced memoization matrix. Conventional LL(k) and LR(k) parsing algorithms only require storage space proportional to the maximum recursion depth that occurs for the given input. This causes these conventional algorithms to have the same worst case memory requirement as a packrat parser. However, a packrat parser is also lower bounded by n and this worst case behavior for LL(k) and LR(k) parsers rarely occurs [10, 22]. In fact, the maximum recursion depth is usually significantly smaller than the size of the input string [22].

### B. Maintaining States

A parser for the programming languages C and C++ requires that the parser is able to maintain a global state. The reason is the nature of typedef’s for the two languages. The parser needs to be able to distinguish whether the input is a typedef symbol, an object, a function or an enum constant, and change the global state accordingly if the meaning of a specific token changes. For instance, the following C code requires this feature: [7]

```
T(*b)[4]
```

By only looking at this snippet, the parser has no way of knowing whether T refers to a function call or a typedef name, it is context-sensitive. If it is a function call, the snippet corresponds to accessing the fifth element of the resulting call to function T with the pointer b as input parameter. If T instead is a typedef, the snippet corresponds to b pointing to an array consisting of four elements of type T.

C or C++, however, can still be parsed with a packrat parser that changes its state whenever a variable changes its type [8]. This is because of the requirement that the type of a variable needs to be declared before its usage and therefore no parsing information prior to a definition of a variable is lost. This way, a separate symbol table can be constructed during the parse which keeps track of the type for different tokens. However, for the general case of context-sensitive grammars, packrat parsers may experience exponential parsing time and memory consumption. This is because during parsing a packrat parser assumes that an already evaluated cell of the result matrix is the correct result, and that this value will not have to change. But if a state change occurs the result matrix may have to be re-evaluated to ensure a correct result during backtracking. This can potentially break the guaranteed linear time characteristic due to cells being evaluated multiple times [22].

### C. Left recursion:-

Left recursion is an issue in PEG and solution is proposed for the same in [11] by Wrath et al. But this approach fails for some PEGs as shown by Tratt[37] .The solution works for a safe subset of left-recursive PEGs with this approach. By extending this algorithm where allowing left-recursive rules with definite right-recursion to work as expected. In order to parse right-recursive PEGs safely, a number of subtle issues need to be addressed, and the set of right-recursive PEGs safely parseable is less than might originally have been hoped for. Next step obviously is to extend the solutions presented in this paper to tackle with indirect left and indirect right recursion. But this may be quite challenging and may impose further restrictions on valid PEGs. Therefore, this gives rise to a open problem: are PEGs really safe for left-recursion?

### D. Error Reporting:-

One important property of parser is to provide good syntax error support. For example, if user enters invalid expression

and to recover from it, it is necessary for parser needs to know if it is parsing an array index or, say, an assignment. It is preferable that the parser should resynchronize by skipping ahead to a token. In the later case, it should skip to a; token since top-down parsers maintains a rule invocation stack and is able to report things like invalid expression in array index. Ambiguous context poses a problem before packrat parser since they are always speculating. In practice, recovery from syntax errors cannot be possible because they cannot detect errors until they have seen the entire input.

### *E. Specification Tool:-*

PEG is looked at as a advanced tool for describing syntax and considered to be better than CFGs and regular expressions. The reason behind this is cited as grammar is unambiguous. But though it is an unambiguous specification of a parser, the language specified by it is whatever that parser happens to accept. But the language we want is easily seen? "Specification" means its meaning must be clear to a human reader. "Prefix capture" in PEG is not immediately visible which the main pitfall is. In addition to left recursion, detection of prefix capture is inevitable for any usable parser generator for PEG. To make sure the grammar defines the language required are there any other conditions that must be detected? This gives rise to think in terms of Parsing Expressions. This raises an argument here that BNF make it to understand easier because it better reflects the working of human mind, or because we are using it since long time?

## V. Conclusion and Future Work

In this paper, compressive review is taken about the packrat parsing introduced by Ford in 2002. It presents the state of the art about the development and research about packrat parsing based upon parsing expression Grammar. Specifically this type of parsing is presented by using memoization technique but subsequently it is this memoization which is shown to be hindrance from applying this parsing technique though it guarantee linear time of execution by avoiding redundant calls. Therefore authors focused on this problem so that wide use of this technique is feasible. In addition to this other issues such as left recursion, error reporting also seems to be associated with this type of parsing approach and discussed here about the initiatives made by researchers to address this issue. Future work may be to apply these techniques to wide variety of grammars so that authenticities of these techniques are to be endorsed.

## References

[1] Alfred V. Aho, Ravi Sethi, and Je_rey D. Ullman. Compilers: Principles, Techniques, and Tools. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1986.

[2] Alfred V. Aho and Je_rey D. Ullman. The Theory of Parsing, Translation, and Compiling.Upper Saddle River, NJ, USA, 1972. Prentice-Hall, Inc.

[3] A. Birman and J. D. Ullman. Parsing Algorithms with Backtrack. Journal of Information and Control, 23(1):1{34, 1973.

[4] Alexander Birman. The TMG Recognition Schema. PhD thesis, Princeton University,Department of Electrical Engineering, Princeton, NJ, USA, 1970.

[5] B. Ford. Packrat Parsing: Simple, Powerful, Lazy, Linear Time - Functional Pearl. In Proceedings of the ACM SIGPLAN International Conference on Functional Programming,ICFP, pages 36{47, New York, NY, USA, 2002. ACM.

[6] Bryan Ford. Packrat Parsing: A Practical Linear-time Algorithm with Backtracking. Master'sthesis, Massachusetts Institute of Technology, Department of Electrical Engeneering and Computer Science, Cambridge, MA, USA, 2002.57

[7] Bryan Ford. Parsing Expression Grammars: A Recognition-based Syntactic Foundation.In Proceedings of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Program-ming Languages, POPL '04, pages 111{122, New York, NY, USA, 2004. ACM.

[8] Robert Grimm. Better Extensibility Through Modular Syntax. In Proceedings of the 2006 ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI '06, pages 38{51, New York, NY, USA, 2006. ACM.

[9] Stephen C Johnson. Yacc: Yet Another Compiler-Compiler. Bell Laboratories Murray Hill, NJ, 1975.

[10] M. E. Lesk and E. Schmidt. UNIX Vol. II. chapter Lex; a Lexical Analyzer Generator,pages 375{387. W. B. Saunders Company, Philadelphia, PA, USA, 1990.

[11] D. Michie. Memo Functions and Machine Learning. Journal of Nature, 218:19{22, 1968.

[12] Kota Mizushima, Atusi Maeda, and Yoshinori Yamaguchi. Packrat Parsers Can Handle Practical Grammars in Mostly Constant Space. In Proceedings of the 9th ACM SIGPLAN-SIGSOFT Workshop on Program Analysis for Software Tools and Engineering, PASTE '10, pages 29{36, New York, NY, USA, 2010. ACM.

[13] Terence J. Parr and Russell W. Quong. Adding Semantic and Syntactic Predicates To LL(k): Pred-LL(k). In Proceedings of the 5th International Conference on Compiler Con-struction, CC '94, pages 263{277, London, UK, UK, 1994. Springer-Verlag.

[14] R. R. Redziejowski. Parsing Expression Grammar as a Primitive Recursive-Descent Parser with Backtracking. Journal of Fundamenta Informaticae, 79(3-4):513{524, 2007.

[15] R. Redziejowski. Some aspects of parsing expression grammar. In Fundamenta Informaticae 85, 1-4, pages 441–454, 2008.

[16] R. Redziejowski. Applying classical concepts to parsing expression grammar. In Fundamenta Informaticae 93, 1-3, pages 325–336, 2009.

[17] S. Medeiros and R. Lerusalimschy : A Parsing Machine for PEGs. In Proc. PEPM, ACM (January 2009) 105-110.

[18] R. Grimm. Better extensibility through modular syntax. In Proceedings of the ACM SIGPLAN 2006 Conference on Programming Language Design and Implementation, pages 19–28, 2006.

[19] K. Mizushima, A. Maeda, and Y. Yamaguchi. Improvement technique of memory efficiency of packrat parsing. In IPSJ Transaction on Programming Vol.49 No. SIG 1(PRO 35) (in Japanese), pages 117–126, 2008.

[20] R. Becket and Z. Somogyi. Dcgs + memoing = packrat parsing but is it worth it? In Practical Aspects of Declarative Languages, January 2008.

[21] Warth, A., Douglass, J., Millstein, T.: Packrat parsers can support left recursion. In: Proc. PEPM, ACM (January 2008) 103-110.

[22] Warth, A., Piumarta, I.: OMeta: an object-oriented language for pattern matching. In: Proc. Dynamic Languages Symposium, ACM (2007) 11-19.

[23] R. Redziejowski. Mouse: from parsing expressions to a practical parser. In Concurrency Specification and Programming Workshop,September 2009.

[24] R. Redziejowski. Parsing Expression Grammar for Java 1.5.

[25] http://www.romanredz.se/papers/PEG.Java.1.5.txt.

[26] Adam Koprowski and Henri Binsztok. TRX: A formally veri_ed parser interpreter. In Proceedings of the 19th European Symposium on Programming (ESOP '10), volume 6012 of Lecture Notes in Computer Science, pages 345-365, 2010.

[27] Kuramitsu, K.: Packrat Parsing with Elastic Sliding Window, Journal of Infomration Processing, Vol. 23, No. 4, pp. 505–512 (online), DOI: http://doi.org/10.2197/ipsjjip.23.505 (2015).

[28] AaronMoss: Derivatives of Parsing Expression Grammars. CoRR abs/1405.4841 (2014)

# Micro-Blog Emotion Classification Method Research Based on Cross-Media Features

Qiang Chen, Jiangfan Feng
College of computer science and technology
Chongqing University of Posts and Telecommunications
Chongqing, China

***Abstract*—Although the sentiment analysis of tweet has caused more and more attention in recent years, most existing methods mainly analyze the text information. Because of the fuzziness of emotion expression, users are more likely to use mixed ways, such as words and image, to express their feelings. This paper proposes a classification method of tweet emotion based on fusion feature, which combines the textual feature and the image feature effectively. Due to the sparse data and the high degree of the redundancy of the classification feature, we adopt the canonical correlation analysis to reduce dimensions of data expressed by the text emotional feature and image feature. The dimension reduction of data can maximally retains the relevance of characteristics of the text and the emotional image on the high-level semantic and utilize the support vector machine (SVM) to train and test the feature fusion data set. The results of data experiment on Sina tweet show that the algorithm can obtain better classification effect than the single feature selection methods.**

***Keywords—tweet sentiment classification; CCA; Text emotional; Image emotional***

## I. INTRODUCTION

Recently, people witness a rapid development of social networks, such as twitter, facebook, sina tweet and tencent tweet. For example, Twitter, according to reports, the registered users of twitter had outnumbered 500000000 until July 1, 2012. Tweet attracts more than 500000000 users and have about 100000000 messages everyday. Because of the increasing number of tweet users, tweet gains attention and wide space of development. Mining the emotional value of tweet has been extensively studied and applied in many fields, such as business advertisement, social network analysis, public opinion monitoring, cause analysis of accident. At present, the widespread practical technology of sentiment analysis is divided into two categories: (1) the type of adopting emotional dictionary [1]with the help of dictionary counts the amount of positive and negative emotional words in the text and then analyze the emotional polarity of the text according to the difference of positive and negative emotional words; (2)other method utilized machine learning [2]by labeling the training corpus and test corpus and use support vector machine (SVM), "maximum entropy", and KNN classifier to classify emotions. Wang et al. [3] construct a analysis system of Twitter sentiment, which can analyze the emotional tendency of the comments about the presidential election in real time; Agarwal et al. [4] characterized the polarity and part-of-speech of the words to investigate emotion classification towards the text of tweet based on the kernel tree model and have obtained certain outcomes; Jiang [5] et al. adopt the approaches of relevant and irrelevant to the topic classify emotional polarity, and it can be divided into positive affection and negative affection. Zhiming Liu [6] et al. study three kinds of machine learning algorithms, three feature selection methods and three calculation methods of feature weight for micro blog emotion classification research, but this method fail to consider the impact of emoticon on the emotional polarity of the whole micro blog. Lixing Xie et al. [7] propose muti-strategy sentiment analysis of micro blog based on hierarchical structure which has certain improvement in classification results compared with rules of emoticons, but this method ignores the characteristics of the polarity in the micro blog text.

A new trend of micro blog message is the increase of the visual content, just as users sometimes sent status words with pictures. This is very common, especially mobile phone users, it is more convenient for them to express the mood by photographs, rather than lengthy words. Fig. 1., for example, contains two posts. The one on the right is positive, on the left is negative.

Fig. 1. Microblog messages with images. Left: It is so lucky, beautiful fireworks, Liuyang fireworks awesome .right: People reflect the air hot eyes throat uncomfortable

Obviously, the message of emotion can be expressed more clearly by images rather than words. It shows that the image is meaningful for micro blog emotion classification. To understand the transfer of visual emotion, Borth et al. [8] put forward visual emotional ontology contains emotional detector library and their methods are chiefly concentrated in the image analysis, but the accuracy of image emotional analysis relies on the accuracy of image semantic labels and machine learning. Weining Wang et al. choose line direction histogram to describe the "dynamic" and "static" types of emotional images and completed emotion classification based on the line [9]by studying the relationship of the image line and emotion; When Dai researched the component of the HSV color and the texture parameters in the gray level co-occurrence matrix , he

discovered the effects of texture on five kinds of emotion [10];H W. Yoo proposed feature extraction methods combining the feature of color and texture as the core technology, thus set up a general framework of image retrieval based on emotion [11]. T.Hayashi et al., firstly, segmented an image into L*L. Then, take the average color of each image block as the color features of the image. Finally complete the mapping of image bottom-layer feature and the emotional keyword by neural network [12]; Weiwei Lu adopt CSIFT generating emotional visual words, and combine with based on global HSV color histogram ,forming graphical semantic expression types of muti-feature for the image semantics [13]; Xia Mao et al. consider the link between the fluctuations of graphics and emotional reaction and the 1 / f wave theory, then obtain the relationship between image features and the emotion information using power spectrum characteristics of the image [14]; Yali Fu investigate the unique characteristics of the wood's shape and extract the image texture feature of the wood, such as directivity, roughness, strength and contrast, completing the extraction of color features under the L * a * b color space and implementing the classification of images of wood between the "gorgeous" and the "natural" emotion [15].

The existing classification of microblog-oriented emotion primarily consider the emotion of the text. Due to the increasing number of with microblog users who sometimes express their feelings by images, so we consider the emotional characteristics of images and text comprehensively. The textual features mainly include the characteristics of emoticons, emotional dictionary and cyber language; The emotional features of images include the feature of color, texture and shape. We classify emotions of microblog with SVM. In the paper, the main contributions are: (1) consider the characteristics of textural and graphical emotion more comprehensively and classify emotions of Chinese microblog accurately; (2) select feature with Canonical correlation analysis and after dimension reduction characteristics it can maximumly keep semantic emotional correlation between the original text and image characteristics matrix. The proposed algorithm reduce the redundancy and improve the operation efficiency and accuracy.

The rest of this paper is organized as follows: In Section 2, 3 we extract the feature of the textual and image emotion. In Section 4, we reduce the dimension of the original characteristics with CCA. In Section 5,The experimental results and analysis of the new approach has been given. Finally, we summarize the main results of the paper.

## II. Text Emotional Characteristics in Tweet

We summarize the related researches and extract several emotional characteristics of micro blog text with its unique characteristics.

### A. The characteristic of emotional dictionary

HowNet (called HowNet in Chinese), established by Zhendong Dong and Qiang Dong, is one of commonsense knowledge bases of the describing object represented by the concept of English and Chinese [16].In Hownet, each concept and what it describes is the content of a record and each word is explained correspond to a number of concepts，the concept is a kind of word semantic description, and, the concept is called meanings, each concept explaned by several meanings original. HowNet the analysis set of emotional words in Chinese contains 3730 positive evaluation words, 3116 negative evaluation words, 836 positive emotional words and 1254 negative emotional words, view words, degree level etc.six parts. National Taiwan University Sentiment Dictionary (NTUSD), an emotional dictionary, which is organized and published by Taiwan University has traditional Chinese and simplified Chinese with 2810 positive words and 8276 negative words [17]. The paper select the simplified Chinese version as the emotional dictionary of feature extraction.

### B. The features of emoticon

Sina microblog platform provides some default emoticons, "emoticons" in crawl down in the text is in the form of being parentheses. For example "  " is the expression of the corresponding text "[happy]."A message may contain multiple emoticons.

TABLE I. Error Rate and Percentage in the Given Areas

| | number | Content |
|---|---|---|
| **Sina weibo expression** | Positive expression34 | For example "  " corresponds to [happy],accord to" [(.*?)]" regular expression |
| | Negative expressions 32 | For example "  ",[hum]) ,  ( ,[cry]), |

### C. Network language features

With the rapid development of Internet, Internet in the process of communication also generates enormous novel online language network language. We collected 16 positive emotional network words, 24 negative emotional network language. Those words are shown in Table 2.

TABLE II. The Part of Network Emotion Words

| Positive network words | Negative network words |
|---|---|
| counterattack,Very good very powerful, GeiLi | miracle, Ceezy, too delicate to bear a blow, a tear-inducingmisery |

## III. Image Emotional Features of Micro Blog

Low-lever features such as color, texture and shape can express rich emotional information, different color, texture and shape can arouse people's different associations and emotional reaction. However, not every a low-level features are our concerns and needs, the image low-level feature selection has the vital effect on the [18] high-level affective semantic expression of the image.

### A. Color feature

Color is the visual feature of object surface, which is the basic element of the content of the image, and is one of the main perceptual features of human recognition. It can be said that in all the visual features of the image, The color is the

most emotional features. Generally, The obvious color was able to attract people's attention and make people have a certain subjective feelings.

Color is represented by color space and HSV color space conform to human visual and psychological feelings, and the color feature don't be affected by illumination and observation angles, also HSV color space quantization results can also be in line with the color feature smaller dimension of visual feature. So the color feature of the image is represented by the HSV space model is appropriate when the semantic image emotion is classified. In this paper, we use the 64 bit histogram method based on the HSV color space to represent the color feature of the image. According to the visual discrimination ability, the tone H, saturation S and brightness V were divided into 16, 4 and 1 respectively. Specific quantitative formula is shown below:

$$H=15\begin{cases}0 & if \quad h\in(345,15]\\ 1 & if \quad h\in(15,25]\\ 2 & if \quad h\in(25,45]\\ 3 & if \quad h\in(45,55]\\ 4 & if \quad h\in(55,80]\\ 5 & if \quad h\in(80,108]\\ 6 & if \quad h\in(108,140]\\ 7 & if \quad h\in(140,165]\\ 8 & if \quad h\in(165,190]\\ 9 & if \quad h\in(190,220]\\ 10 & if \quad h\in(220,255]\\ 11 & if \quad h\in(255,275]\\ 12 & if \quad h\in(275,290]\\ 13 & if \quad h\in(290,316]\\ 14 & if \quad h\in(316,330]\\ 15 & if \quad h\in(330,345]\end{cases} \quad S=\begin{cases}0 & if \quad s\in(0,0.15]\\ 1 & if \quad s\in(0.15,0.4]\\ 2 & if \quad s\in(0.4,0.75]\\ 3 & if \quad h\in(0.75,1]\end{cases} \quad V=\begin{cases}0 & if \quad v\in(0,0.15]\\ 1 & if \quad v\in(0.15,0.4]\\ 2 & if \quad v\in(0.4,0.75]\\ 3 & if \quad v\in(0.75,1]\end{cases} \tag{1}$$

Can be seen, in the three component, the human visual system sensitivity of V, S, H is increased in turn, so according to the H, S, V quantitative series, for three characteristic component , calculation of weights combination can get one dimensional feature vector L can be obtained, it can expressed by the following equation:

$$L=16H+4S+V \tag{2}$$

### B. Texture feature

Texture reflects the homogeneous phenomenon of visual features which existing in image, it usually performs irregular in Local but regular in whole, as the clouds, distant lakes. The coarseness, concave-convex and other characteristics of texture can evoke psychological reflection and emotional perception. Material determines the organizational structure of the object surface, so that objects of different materials would create a different psychological feeling.

Tamura texture features is proposed based on the basis of human visual characteristics of texture perception Psychology Research [19] , it divided into six components correspond to six properties of Psychology perspective texture feature, which are coarseness ,contrast, directionality, linelikeness , regularity and roughness. Since the first three components have an intuitive visual sense ,and can directly engender psychological changes and emotional reactions. Therefore, this paper is to extract the coarseness, the contrast and the directionality of Tamura texture to represent the texture feature of images .The article [20]describes the calculation method of the three texture features.

### C. Analyze of Image features in different corpus

These images contain a wealth of features and other semantic information [24], and if color feature, texture feature can distinguish micro-blogging, for the above assumptions, the paper was verified by experiment and found that the image features, texture features in different micro Bo has the ability to distinguish between apparent emotion class, micro-blogging emotion classification has a good effect.

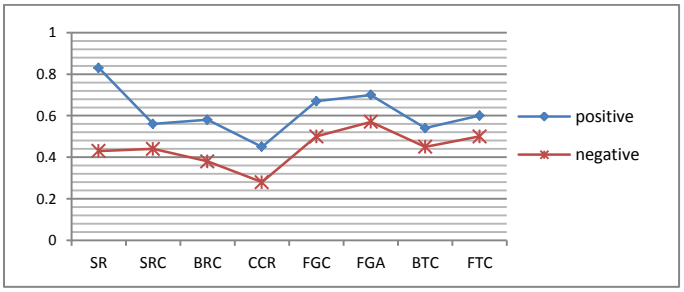



Fig. 2. Image interpretations. We demonstrate how each visual feature distributes over each category of images by the proposed model. The visual features include saturation (SR), saturation contrast (SRC), bright contrast (BRC), cool color ratio (CCR), figure-ground color difference (FGC), figure-ground area difference (FGA), background texture complexity (BTC), and foreground texture complexity (FTC)

Fig.1demonstrates how each visual feature distributes over different emotions the proposed model. For example, in the Happiness category, images tend to have high saturation and bright high contrast, which both bring out a sense of peace and joy. On the contrary, images in Sadness category tend to have lower saturation and saturation contrast, which both convey a sense of dullness and obscurity. Sad images also have low texture complexity, which gives a feeling of pithiness and coherence. The distribution during features value of two types of micro-Bo corpus is significantly different on the color and texture , these two features have a clear distinction. Then use the two features to classify with good results.

## IV. Micro-Blog Feature Reduction and Fusion

Each feature of micro-blog in different degree reflects the partial information of the researched question, but features redundancy will increase the amount of calculation and increase the complexity of the research problem. Therefore , I hope through quantitative analysis, using less feature subset to express more information, feature selection method is based on the purpose. In this paper, in order to fully exploit the advantages of each feature selection method, a new feature fusion algorithm with CCA, through a combination of two types selection methods to obtain effective integration features. The classification process is as follows:

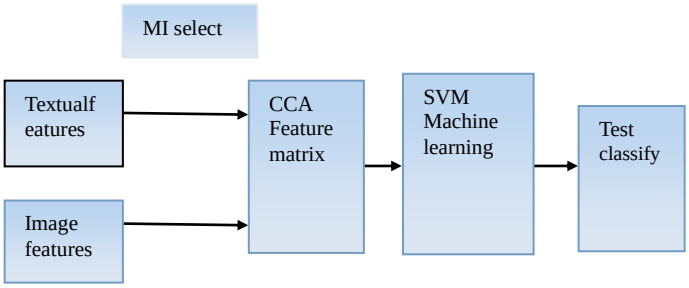



Fig. 3. classification process

Feature selection methods can be divided into two categories: supervised and unsupervised feature selection methods. Supervised feature selection commonly include: Document Frequency, Information Gain, Chi-Square Statistic, Mutual Information and other methods [21]. IG, CHI and MI is important to measure the degree of correlation of the feature item, but IG is for the category as a whole to consider the importance of a feature item. The DF methods use thresholds to select characteristics which are representative and strong distinguish ability, that use class discrimination lever to measure the importance of the features. These four methods can be used to get have a major impact on the classification of features from different levels, but its drawback is that the calculation of the metric associated with the Corpus categories marked.

The article first with DF method to select the class distinction between good features, then CCA method [22]to reduce redundancy between features and information as soon as possible to preserve the original features information , after feature dimensions ,text and image feature emotional semantic correlation is maintained .CCA always is used in cross-media retrieval [23].

Different types of multimedia data can be co-expressed similar feelings semantics, such as the "smog" image and text(The fog is too terrible 😂) data. In the statistical sense, "terrible smog" there is a correlation between the corresponding image data and text data, this section uses canonical correlation analysis, $X_{(n\times p)}$ showing concentrated extract image data from micro-blog visual emotion feature matrix, $Y_{(n\times q)}$ showing data from text concentrated extract of emotional text feature matrix.So the definition of the correlation of two variables X and Y between the fields as follows: Field variable with n samples, p-variable denoted $X_{(n\times p)}$, additional samples n, q-variable denoted $Y_{(n\times q)}$, in order to maximize to extract the main features of the correlation between X and Y as a criterion, extract from a combination of variable X in L, extracted from a combination of variable Y in M, as follows:

$$X_{(n\times p)} \xrightarrow{W_{X(p\times m)}} L_{(n\times m)};$$
$$Y_{(n\times q)} \xrightarrow{W_{Y(q\times m)}} M_{(n\times m)} (m<p, m<q) \quad (3)$$

Where $W_X$ and $W_Y$ are the feature vector space, also know as the canonical variables. According to equation (1) , the relevant variables with more field variables X and Y is a combination of less variables between L and M interrelated, through the distribution of values to determine the form of space-related distribution of X and Y. Instead, the value $W_X$, $W_Y$ determined corresponding variables importance. So the question boils down to how to get the canonical variables, the correlation coefficient is defined under p = r(L,M), constraint in equation (3), so that the correlation coefficient is optimized.

$$\rho = r(L,M) = \frac{W_X^T C_{XY} W_Y}{\sqrt{W_X^T C_{XX} W_X W_Y^T C_{YY} W_Y}} \quad (4)$$

$$v(L) = L^T L = W_X^T X^T X W_X = 1;$$
$$v(M) = M^T M = W_Y^T Y^T Y W_Y = 1; \quad (5)$$

In formula (2), $C_{XY}$ represent covariance matrix constructed by $X_{(n\times p)}$ and $Y_{(n\times q)}$. Combined with formula (2) and (3) using the Lagrange multiplier method can be obtained $C_{XY} C_{YY}^{-1} C_{YX} W_X = \lambda^2 C_{XX} W_X$, it will transform the optimization problem to characteristic root problem $Ax = \lambda Bx$. And further according to the formula (1) to give the smallest variables combination of $L_{(n\times m)}$ and $M_{(n\times m)}$, to maximize reveal correlation between $X_{(n\times p)}$ and $Y_{(n\times q)}$.

## V. EXPERIMENT

This article is aim to do the positive and negative classification research about related text of most discussed topic on Sina micro-blog. And the data about film and television, people's life and products have been collected. Each field chooses 2000 micro-blog comments as linguistic data. And labels will be added artificially. The results are listed in Table I below:

TABLE III. THE STATISTICAL RESULTS OF CORPUS

| Topic | sample | positive | negative | neutral | sum |
|---|---|---|---|---|---|
| **Film** | The Secret of Grave Robber | 600 | 464 | 936 | 2000 |
| **Society** | smog | 236 | 1250 | 514 | 2000 |
| **Product** | iPhone | 1124 | 320 | 556 | 2000 |
| **Total** | | 1960 | 2034 | 2006 | 6000 |

### A. Validation method

The text makes accuracy, recall rate, value of F and macro-averaging as the evaluation parameters. Suppose $m_{righti}$ to be the correct amount of micro-blog text classfied to ci. $m_{wrongi}$ is the amount of Micro blog belonging to other kinds of Micro blog wrongly classified to ci. $m_{all}$ is the accurate amount of Micro blog this belonging sort ci.

accuracy of sort ci:

$$Precision_i = \frac{m_{righti}}{m_{righti} + m_{wrongi}} \times 100\% \quad (6)$$

recall rate of sort ci:

$$\text{recall}_i = \frac{m_{righti}}{m_{alli}} \times 100\% \quad (7)$$

Value of F of sort ci:

$$Fi = \frac{2 \times precision_i \times recall_i}{precison_i + recall_i} \times 100\% \quad (8)$$

### B. The result and analysis of experiment

For each dataset of microblog, 70% of dataset is randomly selected as training set respectively. The other 30% is used as test set. Performance assessment adopts the percentage of the rightly classficated microblog in data sets, i.e,classification accuracy. 20 experiments will do for each of the two algorithms in order to write down the classification accuracy and the average accuracy of each experiment. Matlab2011a toolkit must be used to realize simulation in the system of Windows7.

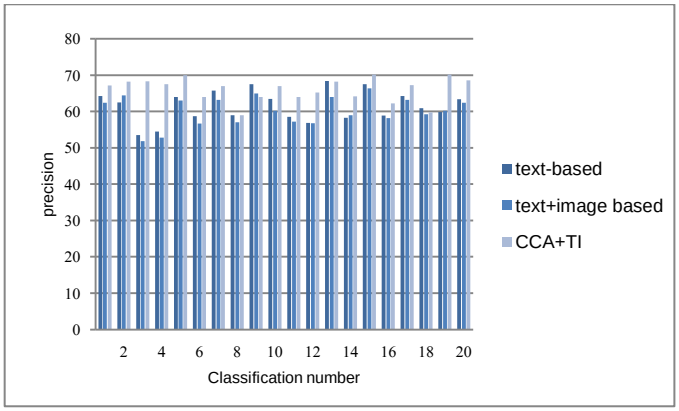



Fig. 4. Precision contrast figure

Above, the horizontal axis represents classification number, a total of 20 times, vertical axis represents classification accuracy. Three methods of classification method respectively represent the text, the text - image method, CCA + TI classification method. Analysis of the results can be seen that the average accuracy between the text method and text- image classification algorithm is similar, although the introduction of image emotional characteristics, but the feature attribute redundancy and feature dimension increasing, make accuracy rate does not improve. By using CCA method to reduce the emotional characteristics dimension, getting maximum correlation characteristic matrix, and then classified. By comparing the accuracy can be improved by 4%. Illustrate after application of the method, statistical correlation can be maintained after dimension reduction between the text emotional characteristics and image emotional characteristics. Emotional characteristics further reduce redundancy, with improved accuracy.

## VI. Conclusion

To solve the sparse problem of Microblog -Text emotional characteristics ， we propose a novel approach for Microblog sentiment analysis based on CCA cross-media model (CBM). Previous researches always focus on the text emotion , neglecting the effect of images emotion with growth of image in the message. Considering more and more people express their feelings through images in Microblog, thus we take images into account in our model. There are three advantages of our method. First, the sentiment of the messages is analyzed by combining the images and texts. Second, this model gives a unified representation of texts and images for cross-media sentiment analysis with CCA method. And the Finally, we use Logistic Regression to relax conditional independence assumption. Experiment results illustrate the effectiveness of our model, with classification accuracy 4% higher than the text-based method.

## Acknowledgment

The work is funded by the Natural Science Foundation Project of Chongqing cstc2014 jcyjA00027.The author authors are also grateful to cstc.

## References

[1] LunWei Ku， TungHo Wu， LiYing Lee， et al. Construction of an evaluation corpus for opinion extraction［C］ ‖ NTCIR － 5 Japan. 2005: 513-520

[2] DasguptaS， NgV. Mine the easy classify the hard: S semi-supervised approach to automatic sentiment: classification［ C］ ‖ ACL ’09: 701 － 709

[3] Wang Hao， Can Dogan， Abe Kazemzadeh. A system for real-time Twitter sentiment analysis of 2012 U. S. presidential election cycle［ C］ ‖Proceedings of the ACL 2012 SystemDemonstrations，2012: 115 － 120

[4] Apoorv Agarwal，Xie Boyi，Ilia Vovsha. Sentiment analysis of Twitter data ［ C ］ ‖Proceedings of the Workshop on Languages inSocial Media，2011:30-38

[5] Jiang Long ， Yu Mo ， Zhou Ming. Targetdependent Twitter sentiment classification［C］ ‖Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:Human Language Technologies，2011:151-160

[6] LIU Zhiming, LIU Lu. Empirical study of sentiment classification for Chinese microblog based on machine learning. Computer Engineering and Applications, 2012, 48（1）: 1-4.

[7] Xing Lixing， Zhou Ming，Sun Maosong. Hierarchical Structure Based Hybrid Approach to Sentiment Analysis of Chinese Micro Blog and Its feature Extraction［J］. Journal of Chinese information processing.， 2012，26( 1) : 73 － 83

[8] D. Borth, R. Ji. Large-scale visual sentiment ontology and detectors using adjective noun pairs. Proc. of the 21st ACM international conference on Multimedia, 2013, 223-232.

[9] WANG Weining, YU Yinglin, ZHANG Jianchao. Image Emotional Semantic Classification based on Line Direction Histogram [J].

Computer Engineering, 2005, 31(11): 7-9.

[10] Yoo H W, Jung S H, Jang D S, et al. Extraction of major object features using VQ clustering for content-based image retrieval[J]. Pattern Recognition, 2002, 35(5): 1115-1126.

[11] Hayashi T, Hagiwara M. An image retrieval system to estimate impression words from images using a neural network[C]// Computational Cybernetics and Simulation, 1997 IEEE International Conference on. IEEE, 1997, 1: 150-155.

[12] Lv Weiwei. Research on Image Classification Algorithm based on emotion [D]. Beijing Jiaotong University, 2012.

[13] 17 MAO Xia, DING Yu-kuan, Muta Itsuya. Analysis of Affective Characteristics and Evaluation on Harmonious Feeling of Image [J]. ACTA ELECTRONICASINICA, 2009, (z1): 1923-1927.

[14] Fu Yali , Guo Na, Zhang Jiawei . Image kansei feature classification based on color and texture features [J].Journal of Zhengzhou university of light industry (natural science), 2009, 23(6): 118-121.

[15] Lu Quan, Ding Heng. Review of research on image retrieval based on emotion [J]. Information theory and Practice, 2013, 36(2): 119-124.

[16] Dong Zhendong, Dong Qiang. [DB/0L]. HowNet is introduced[2013-7-20]. http://www. kcenagc. com.

[17] Chen Xiaodong .A Thesis Submitted in Full Fulfillment of the Requirements for the Degree of the Master of Engineering:[D].HuBei province:Huazhong University of Science and Technology ,2012.

[18] Liu Zengrong ,Yu Xueli ,Li Zhi . Idengtification of Image Emotional Semantic based on Feature Fusion [J]. Journal of Taiyuan University of Technology, 2012, 43(5): 553-557.

[19] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception[J]. Systems, Man and Cybernetics, IEEE Transactions on, 1978, 8(6): 460-473.

[20] YU Jin. Content based visual information retrieval[M]. Science Press, 2003.

[21] Yang YM ,Pedersen JO . A Comparative Study on Feature Selection in Text Categorization[C]/ / Proceedings of the ICML,1997:412-420

[22] H.Hotelling. Relations between Two Sets of Variates .Biometrika , PP. 321 一 377 , 1936.

[23] Wu F, Zhang H, Zhuang Y. Learning Semantic Correlations for Cross-Media Retrieval [A]. Proceeding of 2006 IEEE International Conference on Image Processing [C], 1465-1468, Atlanta, 2006.

[24] Yang, Y., Jia, J., Zhang, S., et al.: How Do Your Friends on Social Media Disclose Your Emotions. In: Proc. AAAI, 14, pp. 1–7 (2014)

# The True Nature of Numbers is that they are a Group Associated with the Painlevé Property

## Mathematical life structures description hypothesis and notation

Yoshimi SHIMOKAWA
Kiiro Book Store LLC.
Tokyo, Japan

***Abstract*—The true nature of numbers is that they are a group associated with the moving Painlevé property. In the past, humans considered numbers to be individual entities. The two-point selective ability of living beings can be considered to have established today's mathematical logic. We can consider mathematical logic to have been formed by the minimum condition of a consistent group (There is no discrepancies group) resulting from a light spectrum being transmitted to or reflected on a three-dimensional closed manifold. In other words, examining the characteristics of the light spectrum enables an understanding of the characteristics of human perception, and an understanding of the characteristics of the numbers perceived by human perception. Based on the understanding of these characteristics, humans think about perceiving mathematical logic. What was unfortunate in the past about the logical constructions that started with Gödel, is that they sought for the concept of numbers amid mathematical logic. Cellular automatons have advanced this approach by physicalizing it, and Conway revealed discrepancies by creating actual machines [12].The decimal system, in itself, is simply a notation tool and merely a notation. Owing to the structure of base conversion, it is convenient to use decimals in two-dimensional notations. However, decimals cannot be used in three-dimensional notations; instead, circular numbers are generated. This can be considered the result of using "0," although the concept of writing out a three-dimensional notation consistently should be created separately. With no concept of null (א) in the " Set theory and the continuum hypothesis"[5], that the perception of the human body is simply maximizing the margin of classification boundaries in a SVM problem [3]. How to consider maximizing the margin of classification boundaries in perceiving the problem could be thought of as the theme for the theory of numbers. Current, by "The International System of Units (SI)" has been defined, A unit got association that the unified. By using the unit, it will be elucidated also algorithms of the molecular structure in the organism. We need a third algorithm. It is the algorithm of the molecular structure organism, that is"the number of recognition" and "the number of physical"Is an algorithm to unify. In other words, temporary has been described structure in this paper, a common algorithm that is "the number of physical" and"the number of recognition". Once again, it was discussed the number of nature, for that algorithm. In the future, to make a key operation using this algorithm. I wish to apply to display notation the development of new devices.**

***Keywords—cognitive science; pattern recognition; scale invariance; Painlevé Property***

## I. INTRODUCTION

### Preconditions and Assumptions

If we were to assume that the two-point selective ability has established the present day mathematical logic (the foundation for classification boundaries), it is necessary to understand its characteristics. First, I considered Conway's "puffer train" as a two-point classification boundary figure structure. It has a main body portion that moves independently at half the "speed of light" without interference from fragments, and two "spaceships" that attain two-point discrimination. For considering a nontrivial self-renewal process, a binary method with two-point discrimination provides the required confidence level. When introducing the physical settings on a computer, this is the resulting appearance because the notations on the display are limited to two dimensions.

Next, It is considered why organisms form clod-like continuous shapes from chaotic locations. It is considered examples (of such growth) in nature as well as of organisms that grow in the above mentioned way in nature. Then focused on portions of organisms that remain motionless as objects instead of as growth, on the opposite axis of discrimination. Through the induction of epitope light, plants ultimately produce fruits that can be counted—one, two, three, and so on. Fruits are considered complete after separation from its main body. In other words, the perception of "separation" in the boundary problem is thought of for organisms as being conducted by epitope induction, some other condition (algorithms) for displays. Then examined the boundary problem for this object formation and numbers, and then created 3D figures illustrating the method by which those shapes can be transformed from "Concept 1," which is explained below. By using a three-dimensional multi-connected closed manifold, "the difficulty of classifications" is interpreted along with the state of changes.

### Question the continuum hypothesis initial setting

The claim of this research questions the initial settings of the continuum hypothesis. My claim is that numbers are "a simple concept arising from the human awareness of two-point discrimination, and its symmetrical transformation patterns" and that the concepts of the natural numbers 1, 2, 3, 4, 5, 6, 7, 8, and 9 are an expression of how this two-point

discrimination is moved. To avoid confusion, the concepts of natural numbers will be referred to below as "Concept 1," "Concept 2," "Concept 3," and so on. Cantor and Gödel focused on cardinality as well as ordinal numbers, but my approach goes a step further, taking the view that cardinality is perceived by the human body, which is created by antigenic epitope induction.

Scope of Analysis and Method

The analysis covers three areas.

1) *Organism model analysis*
2) *Mechanical model analysis*
3) *Conceptual model analysis*

In other words

1) *classification boundary epitope induction*
2) *classification boundary algorithm*
3) *classification boundary cognition*

In other words

*1) Analysis of the organism exponentiation conditions save deformation*

*2) Analysis of the ends of the puffer train and the two spaceships described in The Recursive Universe*

*3) Analysis of tori (three-dimensional multi-connected closed manifolds) that represent the transformation of Concept 1*

To analyze the binary numerical machine minimum model for Conway's puffer train described in The Recursive Universe (Item 2)[12], Be inferred the property of life (in other words, the independence that is characteristic of life), from a different perspective by making a comparison to the conditions of a closed manifold. To analyze the organism model (Item 1), Be inferred numerical formulas from epitope induction, examining "copying'' numbers and "control or stop" numbers that can be linked to the binary system. In other words, cell division was interpreted as light induction using copying, and the transformations of Concept 1 (Item 3) were analyzed from the state of perception in which understanding changes when position changes. In other words, the manifold of Concept 3 was displayed.

The minimum figure of perception for considering the logic of numbers is the binary machine minimum model described in The Recursive Universe (Item 2) as the logical theoretical framework. It is considered co-dimensional problems that generate discrepancies in mathematical logic to be inseparable from the structure of perception. It analyzed the structure of the location of visual perception using an organism model, and viewed it along with its constituent equations. Then surmised other examples of this kind of equation being used for other organisms. Then, cognitive recognizes the "concept 1", also, how to recognize a mistake as "concept 2" and "concept 3", I shown in the figure. It was constructed a simple figure with the whether 3D can be deformed in any way from the "Concept 1". Using this 3D multi-linked closed manifold, changes were observed regarding "The difficulty of classification" at the same time.

Comparative analysis targets three, respectively

1) *physical status quo analysis*
2) *physical state compression model analysis*
3) *cognitive status analysis That is the meaning.*

## II. Organism Model Analysis

### A. Number of Visual Perception Iris Structures

First, I used mathematical logic to examine the two-point discrimination structure of two eyeballs. It is considered discrimination to be the margin maximization of a classification boundary in an SVM, and searched for its discrepancies. Vision (i.e., light-gathering) done by organic perception is structured with the structure of the iris (the higher-order equations for "3," "5," and "6"). However, when viewed through a vector analysis of a three-dimensional space, this structure resembles a Seifert surface structure with topology classification E6. It is a 52-dimensional real Lie group composed of four automorphic mappings over four rings. It is a space given by a closed curve C, meaning that it can be created without S and L touching. This is the base of perceptual cardinality. Seifert surfaces are composed of areas with four boundaries and linking numbers. From there, strokes are added to two locations and by folding over to the inner side, three rings overlap on the inside[7]. This is a knot structure called a (p, q)-torus knot, and in two dimensions, it is also a 3 + 1 threshold discrimination structure. Spaces that resemble the horizontal figure eight in two dimensions are two-threshold discriminations, but they can be made consistent (no contact between discrimination spaces S and L) when they have a butterfly stroke structure in three dimensions. In other words, they can be considered a 1-threshold discrimination in three dimensions. This space is a simply connected 3-manifold floating in a temporary 3-sheeted covering space at the center of a simply connected 3 dimensional closed manifold. By using these four homeomorphisms, adjustments are perceived, and far-point perceptions can be surmised.

Fig. 1. Seifert space, in four of areas, have been constructed. Even when it is overlapped with the complexity, even when just break up into three regions and one of the areas, Linking number of them does not change

From the opposite perspective, even for the same two-point discrimination, a two-point discrimination of an included area is structured from the overlap of three portions. In other words, vision is defined in terms of a simply connected 3-manifold and two-object discrimination, but these can be considered as an E3 multi connected three-dimensional closed manifold. Doing so secures a spot in only one region, enabling a stable and consistent discrimination in the interior.

Organisms perceive the world using the "simply connected three-dimensional closed manifold in a temporary 3-sheeted covering space" as the source of reflection. In other words, it is possible that conditions omitted from this intricate structure are not perceived. To put it differently, the perception used by organisms is one for which this structure is consistent.

From the discussion above, the human visual perception can be considered to be able to discriminate 9 regions (3 points, 3 locations) using the continuum of concentrated light.

### *B. Number of Epitope Inductions Composing Plants and Animals*

Strictly speaking, a two-point discrimination can be considered two special locations in a three-point discrimination of a convex function. Although there is an origin serving as a harness that moves the two locations, in the case of the structure of the spectrum and iris, it indicates the light-gathering act from the outside.

Pursuing this discrepancy reveals the cardinality of numbers and the discrepant portion of the decimal system. It is considered the origin of perception by organisms (which begins with the two-point discrimination) to be Concept 2 and Concept 3, and the final stage of perception by organisms to be Concept 5. The concept currently called "3" can be broken down into "2 + 1," with the "1" portion of "2 + 1" being an adhering item. In other words, It is considered generating Concept 4 to have a common harness. Furthermore, Concept 3 has self-homeomorphism, and Concept 6 is the result of moving unchanged to a higher order and becoming three-dimensional. It is considered the (n + 1) pattern that moves to a higher level and has an external origin acting as a harness to be Concept 7. If considered in terms of the algebraic method, it is what becomes the characteristic function.

Among the Painlevé equations, the third-order equations are the same as the equations representing the structure of the iris, which is the source of visual information. These Painlevé property have four singularities: a moving pole, infinity point, point 0, and point 1. Among these four singularities, the moving pole is the source of the relativity theory and the infinity point is the source of the theory of cosmological inflation. The shape and characteristics of the universe are explained using the characteristics of light. For the dispersion point/separation point directly before point 0/point 1, on the other hand, Dr. Barbara McClintock used the characteristics of light to explain the shapes and characteristics of terrestrial plants [1]. In terms of separation/dispersion, the location affected by the real array is the antigenic determinant (epitope). In other words, the iris perceives both the far point and the dispersion/separation point.

When explained in terms of a garlic that is a pentagon and okra, antigenic determinants that react to light correspond structurally to the origin of the Hasse diagram, and can be considered radicals that manipulate genetic information through the action of light. In terms of signal logic, they correspond to a harness. The types with two axes passing through them may be describable by second-order equations, and types with three axes by third-order equations.

Now It will discuss the shapes of plants. A normal garlic bulb (Photo 1-1) is formed by a Concept 5 epitope, while a garlic bulb with a shifted focal point (Photo 1-2) is formed by a "5 + 9" epitope. An okra pod is formed in an inverted manner by an epitope of Concept 5, but the internal structure of the pod is formed by a seed using 14 epitopes in the inverted cavity. The cross-section of a normal garlic bulb (Photo 1-3) has a mathematical logic in which a complex-plane curve circles while spinning toward the center, with the starting and ending points aligning. In a garlic bulb with a shifted focal point, the epitope formation of Concept 5 becomes shifted such that the end point of the first "lap" does not overlap, causing the $H_2O$ ions (a constituent of the formation) to reverse its polarity and make two "laps" with "5 × 2" – 1. In other words, Concept 1 results in this case because Concept 5 of a normal garlic bulb (which is the "one before" in a quinary calculation) is interpreted as Concept 1 in the epitope. The second "lap" of the shape formation can be considered a covering space in the numerical calculation.

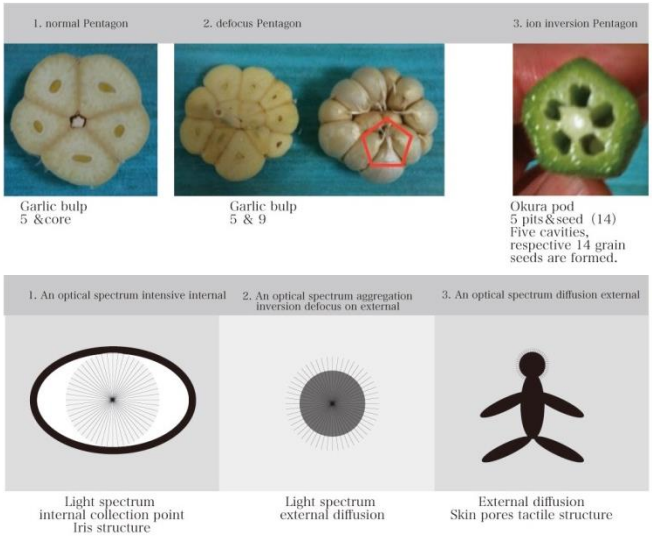



photo1：1. normal Pentagon 2. defocus Pentagon 3. ion inversion Pentagon

Fig. 2. Of visual and tactile, perception condensing reversal

Each of the pores along the human skin has individual epitopes, and is sealed by pentagon-shaped molecules. From there, two "tails" extend from each pore, leading to the foliation structures under the skin. For the eyeballs previously mentioned, visual acuity is determined by two-point discrimination with two eyeballs, but the discrimination with human skin may differ according to the day/night lighting and temperature differences. In other words, temperature differences may be used for the detection of the maximization of the margin of classification boundaries in an SVM problem with cardinality[3]. The light structure from the spherical surface used by vision can be considered a Gaussian mapping on the skin in the case of touch, and used in touch perception. Skin cells are replaced in 28-day cycles (peaking after 14 days). Since the human body is also an aggregate formed using the epitopes of Concept 5, it can be inferred that the skin surface (sealed with the epitopes of Concept 5) uses the same identification number as okra (Concept 14) for replacements.

### C. Reaction-diffusion Path Estimation and Number of Equations

The panda's coat of fur is a special example of predictable transmission. The panda has a diet of bamboo grass with strengthened epitope paths, and has a coat of fur identified by a distinctive two-color pattern. Juvenile pandas that do not eat bamboo grass do not have the distinctive two-color identification pattern. Among humans, the birthmark known as the Mongolian spot that is prevalent among those of Mongolian descent may be due to an accumulation of B12 cobalamin (which contains a cobalt molecule at its center) at the final portion during the final stage of ecological formation from epitope transmission. The birthmark disappears when the half-life of cobalt (3 to 5 years) has elapsed. The mathematical model of the transmission path of these epitopes simulates a signal network of a group of cells.Pattern of these plants and animals are thought to possible written in mock function, which is a Exponentiation conditions save deformation.

Some examples of epitope inductions with reversed ions are as follows. Among plants, there is the example of lotus petals, which are formed as 16 "$2^{2^2}$" petals with the third-order equation "$n^{n^n}$." Epitope induction No. 17 mays occur in the hollow stalk created from negative ions. When the shape of the lotus root is "$n^{n-1}$" "$3^{3-1}$", it is created by the second-order equation for "9," which may be because the polarity of the ions in the underwater portion of the lotus root is reversed. Another example is the second molar tooth, which has the same structure as the lotus. The visible portion is composed of four mountain shapes, and the ion-inverted root portion has three prongs. The first-order equation (n + 1) that represents Concept 17, can be considered one of the convergence epitopes in the human body. Object programs that represent computer units (binary) can all be the result of the epitopes having a heteroclitic-loop reaction-diffusion path.

## III. Computer Display Analysis: Numbers of Ends of the Puffer Train and Two Spaceships

The ends of the puffer train and the two spaceships described in The Recursive Universe are paired with each other, and are related in three dimensions like points of contact in a Seifert surface. A non-cyclical region with a pulsar and the leading edge of the train are at the top and bottom of both edges, respectively. This arrangement can be considered to be merely the result of planarization of the left/right orientation of molecular structures existing in higher-dimensional worlds, which itself is a polytope of compressed notation.

Ultimately, it is the moving objects that are "fixed," and the puffer train has a bottom and front. In other words, the puffer train has directionality, and the existence of this directionality is considered the true nature of movement. Having directionality implies that symmetry is spontaneously broken. This spontaneous breaking of symmetry is the "movement" that gives a region its independence from the background, or what could be considered "life-like."

The relationship between the main body portion that moves independently at half the speed of light without interference from fragments and the two spaceships that attain two-point discrimination can be expressed by a tau-function relationship represented by three objects aligned in a straight line. When considered as points in a 5-dimensional projected space, they become consistent. The packaged structure is seen in molecules such as collagen. When considered only in terms of mathematical logic, it is the result of a triangular pyramid calculation formula consistently describing self-homeomorphism. For the six tau functions[9], the three dependent variables of the symmetrical forms correspond to the three f variables.

In other words, the fixation of objects is a temporary fastening supported by directionality in at least two directions. It is balance that appears to be stopped, and this is the true nature of "stoppage." To put it differently, the true nature of a three-dimensional closed manifold is supported by equations of at least the third order, and it is a balanced, closed object bearing the potential for movement. The appearance varies according to the dimensions in existence, and it is the portion thought to have a covering space in quantum physics. It is considered interpret this fixed balance as Concept 1.

Both ends of Conway's puffer train and the two spaceships described in The Recursive Universe are a pseudo-open space with a barrier and are closed by two dimensions. However, by making their length massive, they become a simulation of infinity. What this signifies is a pseudo-open model enclosed by straight lines, but this could be a length calculated as a curve. This could also be a pseudo-model for space that humans once perceived to be opening or continually expanding.

New solutions can also be obtained for polyhedral objects by using curves on the surfaces of the Goldberg solid body. The higher-order equations "3," "5," and "6" are used for these solid bodies. When these facts are linked objectively, the logical shapes that form nature can be considered to be dependent on curves. The recognition that straight lines are special cases of curves may further advance the potential of science. "3" can be considered the minimum figure of self-homeomorphism, "5" the minimum figure of the condition of stopping with bridging [8][2], and "6" the minimum figure of dimension-raising using two self-homeomorphisms. Conway created a two-dimensional model of "life" that moves autonomously and has self-homeomorphism. In other words, it can be said that the figure was created which becomes the fixed balance of Concept 1 that is independent from outer space.

However, the Seifert structure of the eyeball represents a 52-dimensional real Lie group E6 (composed of four geometric automorphic mappings over four rings) where it is built infinitely externalized internally [4] [6]. In other words, it can logically be considered a structure able to perceive a far-away point arbitrarily close to a straight line.

## IV. Analysis of Tori (Three-Dimensional Multi Connected Closed Manifolds) that Represent the Transformation of Concept 1

### A. Perception Analysis of Concept 1 to Concept 9

Concept 1

signifies a simply connected three-dimensional closed manifold and a multi connected three-dimensional closed manifold.

Concept 2

signifies the action of maximizing the margin of classification boundaries of two-point discrimination.

Concept 3

signifies the minimum number of automorphic constituents; a polytope towards two dimensions.

Concept 4

signifies center 0 at the time of nonlocal stable opening/closing.

Concept 5

signifies a simply connected higher-dimensional open manifold; a polytope towards three dimensions. For third-order Painlevé equations, "1" and "5" are processed as the same simple group, where "R5/2" and "R1/2" contained in a sphere of radius R are the same covering space of a cyclic group.

Concept 6

signifies 2n + 0; a diploid of a perfect number with the minimum self-homeomorphism in three dimensions (Concept 3).

Concept 7

signifies 2n + 1.

Concept 8

signifies $n^3$ + 0; "4" in three dimensions, along with its covering space and origin 0.

Concept 9

signifies $n^{n-1}$ + 0; "$n^n$" in three dimensions, along with its covering space and external origin 0.

The transformation of Concept 1 is represented by tori (three-dimensional multi connected closed manifolds).

The compact theory (Kaluza-Klein Theory) of the co-dimensional portion that encloses the independent Concept 1 described contains a discrepancy [11]. The reason for this discrepancy is a question of how the locations interpreted were changed and the interpretation of locations is related to the problem of perception. Accordingly, It is considered in terms of the number of material structures that support perception as well as the sense of sight and touch, and the number of epitope inductions using characteristics of light that are indispensable for perceiving organisms. When re-examining what numbers are in terms of these locations covering spaces, It is considered the problem along with its initial settings—living things that perceive numbers. Human perception, in particular, is composed of what resembles a covering space of higher-order equations. The homeomorphism of the mathematical structure of the iris may be reflected in the target and create visual perception.

From the observations above, the true nature of numbers can be considered a creation of the rationality that describes "life." In macro terms, everything can be understood from the shining or reflection of the light spectrum—from the molecular structure of minerals to the trajectories of celestial objects.

Humans have moved from the age of paper and writing instruments used to make compressed notations in two dimensions, to the age of recording transmitted manuscripts enabling uncompressed notation. The age of notation using light has begun. Writing on the two dimensions of paper is a notation method that requires the compression of co-dimensions, and the decimal system is designed mainly to handle this application. Protists are completed in five decimal world,However, achievement that won the longitudinal and movable creatures, Have adopted the decimal system, which was to have a symmetry in quinary. However, notations that do not compress co-dimensions are needed when drawing non-orientable organisms such as sea urchins, oriented manmade structures, the shifting of orientations that occurs in outer space, or the structure of viruses. While computational models already exist, there are no comprehensive concepts or operation keys. Therefore, new connotations for these new notations are required. As the mathematical logic and three-dimensional computations are both ready, keys used to write extra dimensions are required for use as "procedure words."When re-examining the nature of numbers in terms of these types of locations (involving folding co-dimension) covering spaces, I considered the initial settings—living things that perceive numbers. The only reason for doing so was that the locations in which number discrepancies appear are special intermediate-dimensional planes (i.e., co-dimensions), where the locations in which the settings of words, perception, and tools are vague or poorly set. My discussion has been advanced by using "procedure words" instead of "sequences/quantities" as the words of mathematics, defining perception as "two-point discrimination" and using the decimal system as the current tool of notation.

I have derived the problem of the divided perception of a set known in a group from the structure of human perception, which is based on two-point discrimination, by approaching the problem using a model involving simple movements in mathematical logic.

## V. Discussion

### A. Control Numbers

Control numbers are locations having extra dimensions, in which the characteristics of light are noticeable, or locations, where shifts can be created with their reflection angles. "5" is a condition that consistently (There is no discrepancies) closes a boundary as a co-dimensional polytope in a regular polyhedron in three dimensions. Since, in general, it is not

possible to solve fifth-order or higher algebraic equations by the Exponentiation root, fixed form transformations using Exponentiation symmetry are not possible. In other words, this means that the path for epitope induction is not definite. Even in the evolved organism of the human body, the fingers, toes, and body as a whole constitute objects that have stopped by dividing into five branches. In other words, the creation of "undividability" can be considered a condition for controlling or stopping convergence. Therefore, Concept 5 is the number that functions as a controller. If observing based on the characteristics of an individual, primitive organism such as sea urchins are independent individuals that have their entire structure formed by a simple equation for "5." Sea urchins have been shaped to withstand the conditions of the highly acidic primitive sea, (in other words, conditions where the state of ions are different that of today) and can be considered to have a shape that is a reaction to the numbers at the rear. The acidic sea can be considered the neighborhood(affect around, and distributed the separation point) condition for creating organisms with acute-angle shapes. In terms of numerical formulas, these organisms may be patterns of manifolds with the origin of epitope inductions (which mammals have inside their bodies) that moved to the outside and have been induced by the 17 types of metals (halogen metals) present in the composition of the sea.

### B. Number of Inductions

Contrary to the control numbers, morphogenesis up to immediately prior to the five-dimensional stoppage number depends on the "substitution group." As can be seen from the example of the lotus discussed above, even in a case where the appearance expresses such numbers as "9" for the hollow stalk, with "16" petals and "1" lotus Exponentiation for the flower, it may possess conformity in the group overall simply by virtue of the fact that the root to be taken differs due to ion inversion and swinging back by the substitution group operation. This can be rephrased as it being engaged in monodromy preserving deformation. In other words, the number of inductions is equipped with a Painlevé Property.

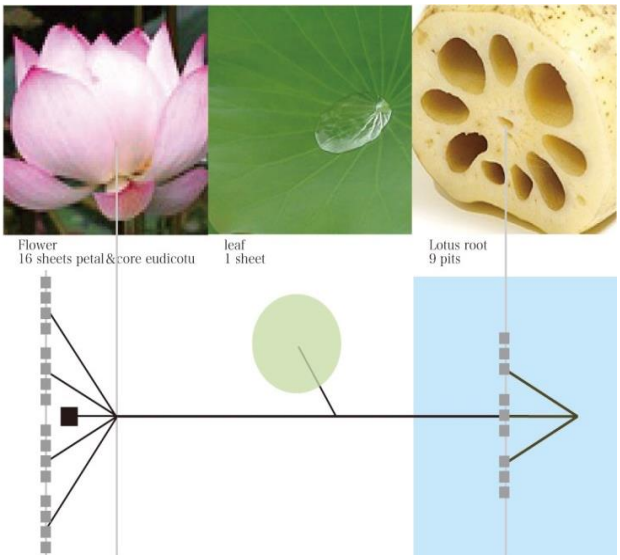



Fig. 3. Ion, plus or minus reversal and transformation It was on the basis of the Euler pentagonal theorem

Moreover, if one considers a case in which animals and plants are established as solids to be "Concept 1," this includes an agglomeration (DiscretelySubset)of the discrete parts of number R.

Epitope induction generally consists of two-part division and three-part composition, and is thought to move in accordance with the knot theory. Induction lines that can be calculated using Gröbner bases are thought to exist on the knots or crossings. The number of twists is thought to be related to speed.

### C. Perception of Numbers

Initially, the idea for this research arose from wondering if an attempt could be made to prove non-independence using the continuum hypothesis. As my hypothesis, It is considered what the result would be if the perceived numbers "1," "2," and "3" were not independent—in other words, did not have the characteristic of being a sequence. The answer is chaos. However, when It is considered how "1," "2," and "3" could be created as concepts when considering chaos as a single set, I also considered the case of perceiving Concept 1 as Concept 2 in a three-dimensional space, and Concept 2 as Concept 3. As pursued these facts, It noticed that when the concept human perceive as Concept 1 takes on a different appearance, it becomes Concept 2 and then Concept 3. Fig.4 shows the transformation and regeneration of a doughnut-shaped torus of a three-dimensional open manifold normally perceived as "1." Concept 5 viewed from above in three dimensions can appear as "3" or "2" when the representation is compressed into two dimension and viewed from the side. Ultimately, it can be transformed into a state enabling the calculation of its homeomorphism to a simply connected three-dimensional closed manifold. While the numbers discussed here are mere concepts, in practice, they can also be considered characteristics of the light spectrum composing living things.

It is necessary to establish boundary classifications and domain theory that use two-point discrimination cognition, the basis for classification boundaries, and that are in line with reality. It is proper to think of boundary classification as pattern recognition of an n-dimension polytype. The way of approaching the question in the continuum hypothesis was simply regrettable, and it is interesting as training in the method for thinking. Adopting Dr. J H. Conway's expressions of haploid and diploid for doing structural pattern recognition for the cardinality of the number are thought be in line with reality and appropriate[3][4]. This is because structural pattern recognition uses a Hasse diagram and preserves the consistency as a group with an orientation. To phrase it differently, it serves as a method for coming and going through an intermediate dimension consistently(There is no discrepancies ) .It is considered as characteristic of the optical spectrum.

## VI. Conclusion

### A. The number of coming and going Topology

Organisms may use these characteristics to construct organic structures. For example, the external appearance of an individual tangerine can be viewed as a simply connected

three-dimensional closed manifold, but the internal structure can be considered a transformation of a doughnut-shaped torus of a simply connected three-dimensional closed manifold. The same applies to bulbs of garlic.

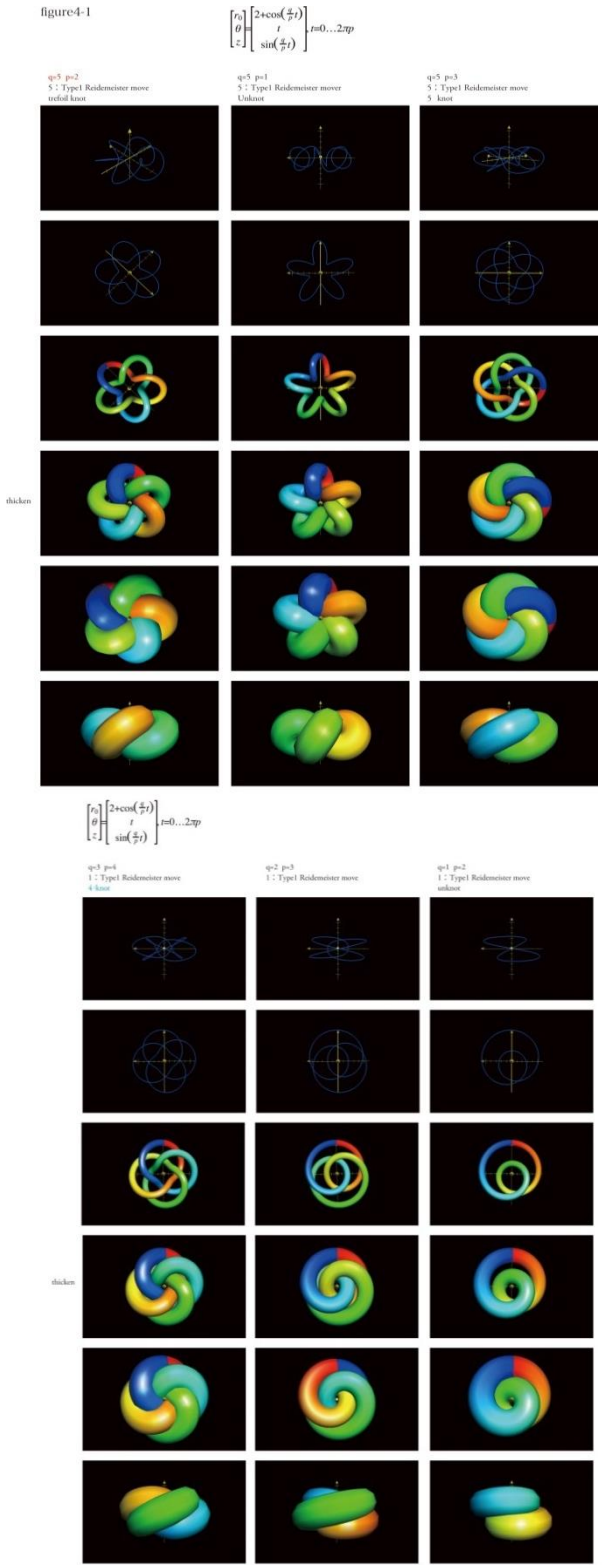



Fig. 4. The transformation and regeneration of a doughnut-shaped torus of a three-dimensional open manifolds normally perceived as "1"

The "separation" boundary problem may be perceived using substances that function as the smallest grinders during the cell division of an organism. These substances may be related to the 17 types of metals (halogen metals) subjected to epitope induction, have dielectric polarization properties and they may be able to transform into shapes such as Kepler manifolds. It is considered that they form cells while in motion. In other words, they could be considered to have the dielectric polarization characteristics of the 17 types of metals, which is the result of the reactions to epitope inductions.The initial setting of the continuum hypothesis has been examined only by the concept, but some ambiguity remains therein. It is necessary to examine the definition of the number before proving the consistency of the number, and in particular to examine why it has reached the point where it is recognized in that manner.

That decimal is, in a compressed description of the calculation in a two-dimensional, slightly deviated from the phenomenon of a three-dimensional space of reality.

Described above

1) *biological model*
2) *machine model*
3) *conceptual model*

Respectively

1) *4 or more dimensions model*
2) *two-dimensional model*
3) *3-dimensional model*

That is the meaning.

These extra dimensions (intermediate dimension), how to traffic to no contradiction, is a Exponentiation symmetry can be calculated the coating space.

Numbers can be considered to be products of light perception by organisms, and their true nature is that they move subtly depending on the dimensions of their notation. In general, they are a group with the Painlevé property.

In other words, the types of equations described here exist to describe independent organisms, making folds and lowering dimensions for clean convergence when induction has completed. To put it differently, the description of life starts from the symmetry of cell division.When it is assumed that two-point discrimination cognition has established current mathematical logic, it was thought that some condition or other with epitope induction is carrying out the recognition of the boundary problems thereof, but it is precisely this "some condition or other" that may be a Painlevé property that has Exponentiation symmetry.

## REFERENCES

[1] Barbara McClintock "The Significance of Responses of the Genome to Challenge" , Science 1984 vol226 November Copyright©1984 by the Nobel Foundation

[2] Bruce Alberts, Julian Lewis, Martin Raff, "Molecular Biology of the Cell" , Newton Press Co.,Ltd. 2010 ISBN-10: 4315518670

[3] Christopher Bishop "Pattern Recognition and Machine Learning" , Maruzen Co.,Ltd. 2009.

[4] J.H.Conway D.A.Smith "On uaternions and ctonions" , A K Peters,Lte 2003 ISBN 1-56881-134-9

[5] J.P.Cohen " Continuum hypothesis" , TokyoTosho Co.,Ltd. 1972 ISBN 4-489-00339-0

[6] Katsusada Morita " On uatrenions and ctonions and dirac equation" , Nippon Hyoron Sha Co.,Ltd 2011 ISBN 978-4-535-78676-9

[7] Kenji Fukaya " Electromagnetic fields and the vector analysis" , Iwanami Shoten Publishers 2006 ISBN 4-00-006883-0

[8] Linus Pauling "A Molecular Theory of General Anesthesia" , Science 7 July 1961 vol 134 November 3471

[9] Masatoshi Noumi "Painleve equation" , Asakura Publishing Co.,Ltd. 2000 ISBN 978-4-254-11554-3

[10] Tosaka, Nobuyoshi " Solution and application of differential equation" , University of Tokyo Press 2010 ISBN 978-4-13-062913-3

[11] Yukio Matsumoto "Basis of the manifold" , University of Tokyo Press 1988 ISBN 4-13-062103-3

[12] William Poundstone" The Recursive Universe" , Nippon Hyoron Sha Co.,Ltd , 1990

# Application of Vague Analytical Hierarchy Process to Prioritize the Challenges Facing Public Transportation in Dar Es Salaam City-Tanzania

Erick P. Massami
Unit for Applied Mathematics
Dar es Salaam Institute of Technology
Dar es Salaam, Tanzania

Benitha M. Myamba
Department of Logistics and Transport Studies
National Institute of Transport
Dar es Salaam, Tanzania

***Abstract*—Transportation is a key to the economy and social welfare; it makes mobility more accessible and enhances the social and economic interactions. On the other hand, the increase of urban population, pollution and other negative impacts has directly affected the existing transportation system in Dar es Salaam City - Tanzania. As the transportation challenges cannot be overcome simultaneously due to the scarcity of financial resources, a decision support tool is needed to prioritize these challenges. In this study, a composite model of Vague Set Theory (VST) and Analytical Hierarchy Process (AHP) is applied to appraise the challenges. The Vague Analytical Hierarchy Process (VAHP) uses opinions of experts collected from a survey questionnaire. The computational results reveal the ranking in descending order of the urban transportation challenges as poor traffic management, inadequacy of proper public transit service and inadequacy of road transport infrastructure. The results also depict that the VAHP model is a useful decision support tool for transport planners, transport policy makers and other industry stakeholders.**

***Keywords—Analytical Hierarchy Process; Vague Set; Urban Transportation; Transportation Challenge; Decision making***

## I. INTRODUCTION

The social-economic activities in any nation involve the movement of people and freight from one place to another. Transportation, as a major logistical element, plays a crucial role in this respect. For instance, the non-availability of products and/or services at the right time and right place can lead to negative consequences such as lost sales and customer dissatisfaction. Therefore efficient and effective transportation system is crucial for ensuring social-economic development [1].

Nonetheless, Dar es Salaam, the seaborne gateway in international trade for Tanzania and most of the East and Central African countries has been facing tremendous challenges in urban public transportation resulting from growth in travel demand and vehicular population. These challenges contribute to the malfunctioning of the city's public transport system. Due to resource constraints, it is impossible to tackle the causes of these challenges simultaneously. In addition, various actors may prioritize these challenges differently. Thus, it is essential to apply multi-criteria decision making technique i.e. Vague Analytic Hierarchy Process (VAHP) to optimally prioritize the transportation challenges in the city – Dar es Salaam. In particular, the Analytic Hierarchy Process (AHP) engineered by Saaty [2] is the most popular instrument in decision making for prioritizing alternatives. The AHP decision making approach combines deductive approach and systems approach of solving problems into one integrated logical framework and this makes it that much more effective in priority setting [3]. As it is hard for a decision maker to assess a factor/sub-factor based on a single number on a proposed scale of Judgement Number of Superiority, the application of the AHP alone could lead to a biased decision. Researchers and/or practitioners have suggested some approaches to overcome this challenge. In this study, we hybridize the AHP with Vague Soft Sets. The assessor's score on a particular factor/sub-factor based on the Vague Set Theory (VST) is given as an interval of crisp values matching with a particular linguistic term. Thus, the assessment of factors/sub-factors becomes much easier and unbiased.

This paper is structured as follows: Section II presents related studies; an overview of vague sets is presented in Section III; modeling procedure based on Vague Analytical Hierarchy Process is developed in Section IV; Section V gives application of the Vague Analytical Hierarchy Process to rank the challenges facing urban transportation in Dar es Salaam city - Tanzania. Lastly, conclusions are given in SectionVI.

## II. RELATED STUDIES

The challenges facing public transportation in cities have been attributed by continuous growth in urban population, private vehicle ownership, ineffective and inefficient traffic management system and the ineffectiveness of public transport services which are the causes of traffic congestion with direct consequences on social and economic activities [4]. More specifically, the immediate effects of ineffective transportation systems are the rising cost of logistic activities and business services [4]. Arasan [5] describes effective urban transportation factors as adequate road networks, traffic management systems and reliable public transport services.

Harriet *et al*. [1] argue that the existing road transport infrastructure capacity in most cities in the developing economies has reached critical level and is unable to meet the huge demand from the increasing number of vehicles. Miller [6] stipulates that underinvestment in transport infrastructure

has negative effects on logistic systems and the entire social and economic activities. Indeed, the rapid growth of China and other Asian countries is stimulated by more sophisticated infrastructure to support transportation and logistics [7].

Effective public transport service is essential for ensuring effective public transportation in urban areas. Xue *et al*. [8] present a study on urban road transportation strategy focusing on the mitigation of GHG emissions and public health damage, taking Xiamen city as a case study. Geng *et al*. [9] analyse the cost effectiveness and environmental benefits of various vehicles taking Shenyang –China as a case study. He *et al*. [10] estimate the energy consumption and $CO_2$ emissions from China's urban passenger transportation sector up to year 2030. Fan *et al*. [11] formulates a model that minimizes hazmat risk and transportation cost subject to road closure constraints. Fatima and Kumar [12] examine the impact of a new public bus transit system in the city of Bardoli, Gujarat, India. Siedler [13] investigates under what conditions the Bus Rapid Transit (BRT) system can be regarded as the best solution to meet the challenges facing the Oslo metropolitan area.

Studies on effective traffic management are proposed by some researchers and practitioners. Malecki *et al*. [14] investigate the utilization of mobile devices to support traffic management system. Saleh *et al*. [15] propose a mechanism for vehicle routing based on the availability of updated traffic information. Schreffler *et al*. [16] suggest that effective traffic management can improve utilization of road networks at much lower cost than constructing new and expanding existing ones.

The challenges facing the road transportation sector are investigated by some researchers and practitioners. Msigwa [17] points out that the challenges of public transportation in fast-growing cities of Tanzania (e.g. Dar es Salaam) are vehicular growth, inadequacy of parking space, high frequency of accidents, transport infrastructure, and environmental and noise pollution.

The objective of this study is to prioritize the challenges confronting road based public transportation in the Dar es Salaam city-Tanzania. The proposed procedure is based on the hybrid model composed of the most popular decision making tool for ranking and/or selecting the alternatives (i.e. Analytical Hierarchy Process) and Soft Set Theory. We should note that the decision making based on the AHP alone can be biased due to the difficulty of assessing influential factors on the proposed scale of natural numbers. This biasness can be reduced or eliminated by the hybridization of the AHP with Vague Set Theory (VST). When using the VST, the assessor's score on a particular factor/sub-factor is given as an interval of values corresponding to a specific linguistic term.

## III. Overview of Vague Sets

Let $S = \{a_1, a_2, \dots, a_m\}$ be the Universe. A vague set over *S* is characterized by truth-membership function $t_v$ and a false membership function $f_v$, $t_v: S \to [0,1], f_v: S \to [0,1]$, where $t_v(a_i)$ is a lower bound (LB) for the membership degree of $a_i$ derived from the evidence in favour of $a_i$, $f_v(a_i)$ is a lower bound (LB) on the negation of $a_i$ derived from the evidence against $a_i$, and $t_v(a_i) + f_v(a_i) \leq 1$.The membership degree of $a_i$ in the vague set belongs to the interval $[t_v(a_i), 1 - f_v(a_i)] \in [0,1]$. The vague value $[t_v(a_i), 1 - f_v(a_i)]$ indicates that the exact membership degree $\mu_v(a_i) \in [t_v(a_i), 1 - f_v(a_i)]$, $t_v(a_i) + f_v(a_i) \leq 1$.

## IV. Modelling Procedure Based on Vague Analytical Hierarchy Process

We develop the Vague AHP model which consists of the following sequential steps:

**Step 1**: Establish Vague Judgement Number of Importance (i.e. Vague Assessment Scale). In this study, we adopt a scale suggested by Massami [18]. This scale has 0 as the lower bound and 1 as the upper bound. The scale has five (5) intervals each with the width of 0.2 units. The relation between linguistic operator of importance and Vague Judgement Number of Importance is listed in Table I.

**Step 2**: Establish Vague JudgementTable of Importance (VJTI) of the criteria/sub-criteria for an objective under consideration. This is as defined in Table II.

TABLE I. The Relation Between Linguistic Term and Vague Judgement Number

| **Vague Judgement Number** | [0.0, 0.2] | [0.2, 0.4] | [0.4, 0.6] | [0.6, 0.8] | [0.8, 1.0] |
|---|---|---|---|---|---|
| **Importance** | Very Low | Low | Moderate | High | Very High |

TABLE II. Vague Judgement Table of Importance for the Criteria or Sub-Criteria

| **Factor ($F_m$)** | **$P_1$** | **$P_2$** | ... | **$P_N$** | $\frac{1}{32}\sum_{n=1}^{N}(a_{mn} + a_{mn}')$ |
|---|---|---|---|---|---|
| **$F_1$** | $[a_{11}, a_{11}']$ | $[a_{12}, a_{12}']$ | ... | $[a_{1N}, a_{1N}']$ | $\frac{1}{32}\sum_{n=1}^{N}(a_{1n} + a_{1n}')$ |
| **$F_2$** | $[a_{21}, a_{21}']$ | $[a_{22}, a_{22}']$ | ... | $[a_{2N}, a_{2N}']$ | $\frac{1}{32}\sum_{n=1}^{N}(a_{2n} + a_{2n}')$ |
| ⋮ | ⋮ | ⋮ | ... | ⋮ | ⋮ |
| **$F_M$** | $[a_{M1}, a_{M1}']$ | $[a_{M2}, a_{M2}']$ | ... | $[a_{MN}, a_{MN}']$ | $\frac{1}{32}\sum_{n=1}^{N}(a_{Mn} + a_{Mn}')$ |

**Note**: $P_n$: Assessor n; $\frac{1}{32}\sum_{n=1}^{N}(a_{mn} + a_{mn}')$: Arithmetic mean of crisp assessment values for factor $F_m, m \in \{1,2, \dots M\}$

**Step 3**: Construct a Crisp Judgment Matrix of Importance (CJMI), C, whose elements are found by using the following mathematical relation.

$$c_{mn} = \frac{\frac{1}{32}\sum_{n=1}^{N}(a_{mn} + a_{mn}')}{\frac{1}{32}\sum_{n=1}^{N}(a_{kn} + a_{kn}')}, \qquad \forall m, k \in \{1,2, \dots M\} \quad (1)$$

Where $[a_{mn}, a_{mn}']$ is a soft assessment value of factor $F_m$ by assessor *n*.

Thus, matrix C is given by

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ c_{M1} & c_{M2} & \dots & c_{MN} \end{bmatrix}_{M \times N} \quad (2)$$

The entry $c_{mn}$ denotes the number that estimates the relative importance of factor $F_m$ when it is compared with factor $F_n$ for an objective under consideration (i.e. a level with respect to the upper level). The entry $c_{mn}$ must satisfy the following:

*a)* $c_{mn} > 0$

*b)* $c_{mn} = \frac{1}{c_{nm}}$ (i.e. the matrix is a reciprocal one).

*c)* If $c_{mn} = c_{mk}.c_{kn}, \ \forall m, n, k$ then the weights are consistent (i.e. transitive).

*d)* If $m = n, \ c_{mm} = c_{nn} = 1$. Thus, $c_m$ and $c_n$ are of equal importance with respect to the objective in question.

**Step 4**: Find a weight vector i.e. priority vector $W$ of order $P$ (i.e. matrix size) satisfying the condition $AW = \lambda W$, the matrix W is called an *eigenvector* and $\lambda$ is an *eigenvalue*. If the Judgement Matrix is completely consistent, its maximum eigenvalue is equal to the dimension of the matrix i.e. $\lambda_{max} = P$. For matrices composed of elements made of human judgements, the condition $c_{mn} = c_{mk}.c_{kn}$ hardly hold as human judgments are inconsistent to a greater or lesser degree. Consequently, $AW = \lambda_{max} W$ and $\lambda_{max} \geq P$. $(\lambda_{max} - P)$ indicates the inconsistency of the judgements. If $\lambda_{max} - P = 0$ then the judgments are completely consistent.

We define the priority vector W as

$$W = [W_{F_m}], \qquad m \in \{1,2, \dots M\}$$

Where,

$$W_{F_m} = \frac{\sqrt[P]{(\prod_{n=1}^{N} c_{mn})}}{\sum_{m=1}^{M} \sqrt[P]{(\prod_{n=1}^{N} c_{mn})}}, m \in \{1,2, \dots M\} \quad (3)$$

$$\sum_{\forall m} \left( \frac{\sqrt[P]{(\prod_{n=1}^{N} c_{mn})}}{\sum_{m=1}^{M} \sqrt[P]{(\prod_{n=1}^{N} c_{mn})}} \right)$$

and $\sum_{m=1}^{M} W_{F_m} = 1$

**Step 5**: Find the weighted sum matrix ($C_W$)

$$C_W = \left[ \sum_{m=1}^{M} W_{F_m} c_{nm} \right], \qquad n \in \{1,2, \dots N\} \quad (4)$$

**Step 6**: Find the maximum eigenvalue ($\lambda_{max}$)

The maximum characteristic root of the equation $AW = \lambda W$ is given by

$$\lambda_{max} = \frac{1}{P} \sum_{n=1}^{N} \left( \frac{\sum_{m=1}^{M} W_{F_m} c_{nm}}{W_{F_n}} \right), n \in \{1,2, \dots N = M\} \quad (5)$$

**Step 7**: Calculate a Consistency Index (CI)

The Consistency Index (CI) is given by

$$CI = \frac{\lambda_{max} - P}{P - 1} \quad (6)$$

**Step 8**: Compute a Consistency Ratio (CR)

A consistency ratio (CR) helps to check the consistency of the judgements and is calculated by dividing the Consistency Index (CI) by the Mean Random Consistency Index ($R_P$). We use Table III for the mean random consistency index [19].

Thus,

$$CR = \frac{CI}{R_P} \quad (7)$$

If CR = 0 then we have a completely consistency case.

If $CR < 0.1$ then we have a satisfying consistency case (i.e. the judgement matrix is considered to satisfy the unanimity).

If $CR > 0.1$ then we have a non-satisfying consistency case (i.e. the judgements are too inconsistent to be reliable). Thus, a re-examination of the pairwise judgements is recommended until a CR less than or equal to 0.1 is achieved.

TABLE III. MEAN RANDOM CONSISTENCY INDEX FOR MATRIX DIMENSION *P*

| **P** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|
| $\mathbf{R_P}$ | 0.00 | 0.00 | 0.52 | 0.89 | 1.12 | 1.26 | 1.36 |
| **P** | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| $\mathbf{R_P}$ | 1.41 | 1.46 | 1.49 | 1.52 | 1.54 | 1.56 | 1.58 |

**Step 9**: Compute Overall Priority of the 2nd Level Factor (e.g. sub- transportation challenge)

The local priorities of elements of different levels are aggregated to obtain final priorities of the sub-factors as follows:

$$FP_{F_{mn}} = W_{F_m}.W_{F_{mn}} \quad (8)$$

$FP_{F_{mn}}$: The Final Priority of the 2nd Level Factor ($F_{mn}$);
$W_{F_{mn}}$: Local Priority of $F_{mn}$ with respect to $F_m$;
$W_{F_m}$: Local Priority of $F_m$ with respect to the goal.

## V. RANKING OF CHALLENGES FORURBAN TRANSPORTATION IN DAR ES SALAAM

Through reviewing the literature and consulting local experts in the transport industry we identify three main challenges and sub-challenges that have a direct bearing impact on urban transportation in Dar es Salaam City. These prime challenges and sub-challenges are shown in Table IV. The selection of the challenges is also supported by Anin *et al* [20] who argue that an effective transportation system needs to consist of the following optimally combined elements:

transport infrastructure, public transit system and traffic management system.

TABLE IV. TRANSPORTATION CHALLENGES IN DAR ES SALAAM CITY

| Public Transportation Challenge ($F_m$) | Sub-Public Transportation Challenge ($F_{mn}$) |
|---|---|
| Inadequacy of road transport infrastructure ($F_1$) | Inadequacy of road networks with lay-bys ($F_{11}$) |
| | Inadequacy of parking lot and terminals ($F_{12}$) |
| | Inadequacy of traffic lights and road sign equipment ($F_{13}$) |
| Inadequacy of proper public transit service ($F_2$) | Prevalence of vehicle emissions and noise pollution ($F_{21}$) |
| | Poor ethical status of drivers and conductors ($F_{22}$) |
| | Occurrence of traffic accidents ($F_{23}$) |
| | Inadequacy of road transport service ($F_{24}$) |
| Inefficient traffic management system ($F_3$) | Poor management of traffic lights and other signals ($F_{31}$) |
| | Poor management of drivers and pedestrians' indiscipline on roads ($F_{32}$) |
| | Poor management of vehicle breakdowns and road accidents ($F_{33}$) |

### A. Priority Vector Determination

The study uses data from sixteen (16) experts who were supplied with survey questionnaires. This team consists of traffic police officers, practitioners and academia in the road transport subsector. The experts' data is used to compose the Vague Judgement Table of Importance (VJTI) from which the Crisp Judgement Matrix (CJM) is deduced.

*1) Assessment of Dar es Salaam City Transportation System*

The interviewed experts assessed the transportation system i.e. first level transportation challenges to give the vague judgmentTable V.

We apply equation (1) to get the elements of the crisp judgement matrix obtained by pair-wise comparison of the transportation challenges leading to ineffective urban transportation system as represented in Table VI.

TABLE V. VAGUE JUDGEMENT VALUES FROM THE ASSESSMENT OF FIRST LEVEL TRANSPORTATION CHALLENGES

| Assessor | $F_1$ | $F_2$ | $F_3$ |
|---|---|---|---|
| $P_1$ | [0.8, 1.0] | [0.8, 1.0] | [0.6, 0.8] |
| $P_2$ | [0.8, 1.0] | [0.6, 0.8] | [0.6, 0.8] |
| $P_3$ | [0.2, 0.4] | [0.4, 0.6] | [0.2, 0.4] |
| $P_4$ | [0.4, 0.6] | [0.2, 0.4] | [0.2, 0.4] |
| $P_5$ | [0.8, 1.0] | [0.6, 0.8] | [0.4, 0.6] |
| $P_6$ | [0.0, 0.2] | [0.2, 0.4] | [0.0, 0.2] |
| $P_7$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] |
| $P_8$ | [0.4, 0.6] | [0.6, 0.8] | [0.6, 0.8] |
| $P_9$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] |
| $P_{10}$ | [0.4, 0.6] | [0.2, 0.4] | [0.6, 0.8] |
| $P_{11}$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] |
| $P_{12}$ | [0.4, 0.6] | [0.2, 0.4] | [0.4, 0.6] |
| $P_{13}$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] |
| $P_{14}$ | [0.2, 0.4] | [0.4, 0.6] | [0.6, 0.8] |
| $P_{15}$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] |
| $P_{16}$ | [0.2, 0.4] | [0.2, 0.4] | [0.0, 0.2] |
| $\frac{1}{32}\sum_{m=1}^{16}(a_{mn} + a_{mn}')$ | $\frac{41}{80}$ | $\frac{1}{2}$ | $\frac{39}{80}$ |

TABLE VI. CRISP JUDGEMENT MATRIX OF IMPORTANCE DUE TO INEFFECTIVE URBAN TRANSPORTATION SYSTEM

| Factor ($F_m$) | $F_1$ | $F_2$ | $F_3$ | Priority Vector (W $F_m$) | Rank |
|---|---|---|---|---|---|
| $F_1$ | 1 | $\frac{40}{41}$ | $\frac{39}{41}$ | 0.3251 | 3 |
| $F_2$ | $\frac{41}{40}$ | 1 | $\frac{39}{40}$ | 0.3332 | 2 |
| $F_3$ | $\frac{41}{39}$ | $\frac{40}{39}$ | 1 | 0.3417 | 1 |

$$c_{11} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')} = 1; c_{12} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')} = \frac{40}{41} \quad ;$$
$$c_{13} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')} = \frac{39}{41}$$

$$c_{21} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = \frac{41}{40}; \; c_{22} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = 1 \quad ;$$
$$c_{23} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = \frac{39}{40}$$

$$c_{31} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = \frac{41}{39}; c_{32} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = \frac{40}{39} \quad ;$$
$$c_{33} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = 1$$

We apply equation (3) to get the following elements of the weight vector.

$$W_{F_1} = 0.3251, \; W_{F_2} = 0.3332 \, , \; W_{F_3} = 0.3417$$

Obviously, the matrix as represented in Table VII is completely consistent i.e.

$$c_{mn} = c_{mk}.c_{kn}, \qquad \forall m, n, k \in \{1, 2, 3\}. \lambda_{max} = \frac{1}{3}\sum_{n=1}^{3}\left(\frac{\sum_{m=1}^{3} W_{F_m} c_{nm}}{W_{F_n}}\right)$$

$$\lambda_{max} = \frac{1}{3}\left[\frac{\sum_{m=1}^{3}(W_{F_m} c_{1m})}{W_{F_1}} + \frac{\sum_{m=1}^{3}(W_{F_m} c_{2m})}{W_{F_2}} + \frac{\sum_{m=1}^{3}(W_{F_m} c_{3m})}{W_{F_3}}\right] = 3.0000$$

P = 3 i.e. the matrix dimension

$$CI = \frac{\lambda_{max} - P}{P - 1} = \frac{3.0000 - 3}{3 - 1} = 0$$

$$P = 3,\ R_P = 0.52$$

$CR = \frac{CI}{R_3} = \frac{0}{0.52} = 0$ which means the crisp judgement matrix of importance is completely consistent.

The priority vector in Table VI reveals that poor traffic management is ranked the first urban transportation challenge of the three first level challenges. This result is largely contributed by poor management of drivers and pedestrians' indiscipline on roads (see the weight vector in Table XII). The second urban transportation challenge at the first level is inadequacy of proper public transit service which is aggravated by poor ethical status of drivers and conductors and inadequacy of mass transit service during peak hours (see Table X). The last first level challenge is inadequacy of road transport infrastructure which is largely contributed by the inadequacy of parking lot and terminals (see Table VIII). Nonetheless, the three challenges are approximately of equal significance i.e. 34.17%, 33.32% and 32.51%. Thus, the overcoming process of the three challenges should be carried simultaneously.

*2) Assessment of Transportation Challenges Leading to Inadequacy of Road Transport Infrastructure*

The experts assessed the factors contributing to inadequacy of road transport infrastructure to give the soft judgement Table VII.

TABLE VII. VAGUE JUDGEMENT TABLE MADE FROM THE ASSESSMENT OF INADEQUACY OF ROAD TRANSPORT INFRASTRUCTURE

| Assessor | $F_{11}$ | $F_{12}$ | $F_{13}$ |
|---|---|---|---|
| $P_1$ | [0.8, 1.0] | [0.8, 1.0] | [0.6, 0.8] |
| $P_2$ | [0.8, 1.0] | [0.8, 1.0] | [0.6, 0.8] |
| $P_3$ | [0.0, 0.2] | [0.0, 0.2] | [0.2, 0.4] |
| $P_4$ | [0.2, 0.4] | [0.2, 0.4] | [0.4, 0.6] |
| $P_5$ | [0.8, 1.0] | [0.4, 0.6] | [0.2, 0.4] |
| $P_6$ | [0.2, 0.4] | [0.0, 0.2] | [0.2, 0.4] |
| $P_7$ | [0.2, 0.4] | [0.0, 0.2] | [0.4, 0.6] |
| $P_8$ | [0.4, 0.6] | [0.2, 0.4] | [0.2, 0.4] |
| $P_9$ | [0.4, 0.6] | [0.0, 0.2] | [0.8, 1.0] |
| $P_{10}$ | [0.4, 0.6] | [0.0, 0.2] | [0.4, 0.6] |
| $P_{11}$ | [0.4, 0.6] | [0.0, 0.2] | [0.4, 0.6] |
| $P_{12}$ | [0.4, 0.6] | [0.4, 0.6] | [0.2, 0.4] |
| $P_{13}$ | [0.4, 0.6] | [0.6, 0.8] | [0.6, 0.8] |
| $P_{14}$ | [0.2, 0.4] | [0.2, 0.4] | [0.2, 0.4] |
| $P_{15}$ | [0.4, 0.6] | [0.6, 0.8] | [0.6, 0.8] |
| $P_{16}$ | [0.2, 0.4] | [0.2, 0.4] | [0.2, 0.4] |
| $\frac{1}{32}\sum_{m=1}^{16}(a_{mn} + a_{mn}')$ | $\frac{39}{80}$ | $\frac{3}{8}$ | $\frac{73}{160}$ |

We use equation (1) to get the elements of the crisp judgement matrix represented in Table VIII.

$$c_{11} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = 1;\ c_{12} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{10}{13}\ ;$$

$$c_{13} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{73}{78}$$

$$c_{21} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = \frac{13}{10};\ c_{22} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = 1\ ;$$

$$c_{23} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = \frac{73}{60}$$

$$c_{31} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = \frac{78}{73};\ c_{32} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = \frac{60}{73}\ ;$$

$$c_{33} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = 1$$

We apply equation (3) to get the following elements of the weight vector whose set is

$$\{W_{F_{11}} = 0.2969,\quad W_{F_{12}} = 0.3859,\ W_{F_{13}} = 0.3172\}$$

Since $c_{mn} = c_{mk}.c_{kn},\ \forall m, n, k \in \{1, 2, 3\}$, the matrix as represented in Table VIII is completely consistent. Thus, $\lambda_{max} = 3$, CI = 0, $R_3 = 0.52$, CR = 0.

From Table VIII, the first challenge associated with inadequacy of road transport infrastructure is shortage of supply of parking lot and terminals to meet the current demand for the facilities (38.59%). Both the Tanzania Roads Agency (TANROADS) and the Municipalities should increase their budget to finance these facilities. The second challenge is the shortage of traffic lights and road sign equipment (31.72%). This calls for the TANROADS and Municipalities to invest more in traffic lights and road sign equipment to be placed at relevant locations. The last challenge is the shortage of road networks with lay-bys (29.69%). Thus, the TANROADS and Municipalities should increase the supply of road networks to meet the current and projected demand.

TABLE VIII. CRISP JUDGEMENT MATRIX DUE TO INADEQUACY OF ROAD TRANSPORT INFRASTRUCTURE

| Factor ($F_{mn}$) | $F_{11}$ | $F_{12}$ | $F_{13}$ | Priority Vector | Rank |
|---|---|---|---|---|---|
| $F_{11}$ | 1 | $\frac{10}{13}$ | $\frac{73}{78}$ | 0.2969 | 3 |
| $F_{12}$ | $\frac{13}{10}$ | 1 | $\frac{73}{60}$ | 0.3859 | 1 |
| $F_{13}$ | $\frac{78}{73}$ | $\frac{60}{73}$ | 1 | 0.3172 | 2 |

*3) Assessment of Transportation Challenges Leading to Public Transit Service*

We use the experts' assessments on the transportation challenges associated with the inadequacy of public transit service to give the soft judgement in Table IX.

We apply equation (1) to get the elements of the crisp judgement matrix represented in Table X.

$$c_{11} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = 1; c_{12} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{17}{23} \quad ;$$
$$c_{13} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{41}{46}$$

$$c_{21} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = \frac{23}{17}; c_{22} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = 1 \quad ;$$
$$c_{23} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = \frac{41}{34}$$

$$c_{31} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = \frac{46}{41}; c_{32} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = \frac{34}{41} \quad ;$$
$$c_{33} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = 1$$

TABLE IX. VAGUE JUDGEMENT VALUES DUE TO THE INADEQUACY OF PUBLIC TRANSIT SERVICE

| Assessor | $F_{21}$ | $F_{22}$ | $F_{23}$ | $F_{24}$ |
|---|---|---|---|---|
| $P_1$ | [0.6, 0.8] | [0.8, 1.0] | [0.6, 0.8] | [0.8, 1.0] |
| $P_2$ | [0.8, 1.0] | [0.8, 1.0] | [0.6, 0.8] | [0.6, 0.8] |
| $P_3$ | [0.6, 0.8] | [0.2, 0.4] | [0.4, 0.6] | [0.4, 0.6] |
| $P_4$ | [0.0, 0.2] | [0.2, 0.4] | [0.4, 0.6] | [0.4, 0.6] |
| $P_5$ | [0.6, 0.8] | [0.4, 0.6] | [0.2, 0.4] | [0.4, 0.6] |
| $P_6$ | [0.4, 0.6] | [0.0, 0.2] | [0.6, 0.8] | [0.0, 0.2] |
| $P_7$ | [0.8, 1.0] | [0.2, 0.4] | [0.4, 0.6] | [0.4, 0.6] |
| $P_8$ | [0.4, 0.6] | [0.6, 0.8] | [0.4, 0.6] | [0.6, 0.8] |
| $P_9$ | [0.6, 0.8] | [0.2, 0.4] | [0.6, 0.8] | [0.4, 0.6] |
| $P_{10}$ | [0.0, 0.2] | [0.2, 0.4] | [0.2, 0.4] | [0.2, 0.4] |
| $P_{11}$ | [0.4, 0.6] | [0.4, 0.6] | [0.2, 0.4] | [0.2, 0.4] |
| $P_{12}$ | [0.4, 0.6] | [0.0, 0.2] | [0.4, 0.6] | [0.2, 0.4] |
| $P_{13}$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] | [0.2, 0.4] |
| $P_{14}$ | [0.8, 1.0] | [0.2, 0.4] | [0.4, 0.6] | [0.2, 0.4] |
| $P_{15}$ | [0.4, 0.6] | [0.4, 0.6] | [0.4, 0.6] | [0.2, 0.4] |
| $P_{16}$ | [0.4, 0.6] | [0.2, 0.4] | [0.4, 0.6] | [0.2, 0.4] |
| $\frac{1}{32}\sum_{m=1}^{16}(a_{mn} + a_{mn}')$ | $\frac{23}{40}$ | $\frac{17}{40}$ | $\frac{41}{80}$ | $\frac{7}{16}$ |

$$c_{41} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')} = \frac{46}{35}; c_{42} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')} = \frac{34}{35} \quad ;$$
$$c_{43} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')} = \frac{41}{34}$$

$$c_{14} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{35}{46}; c_{24} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')} = \frac{35}{34} \quad ;$$
$$c_{34} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')} = \frac{35}{41}$$

$$c_{44} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m4} + a_{m4}')} = 1$$

We apply equation (3) to get the following elements of the weight vector whose set is

$$\{W_{F_{21}} = 0.2088, W_{F_{22}} = 0.2825, W_{F_{23}} = 0.2343, W_{F_{24}} = 0.2744\}$$

Table X reveals the following. The first influential factor is poor ethical status of drivers and conductors (28.25%). This necessitates routine training in interpersonal skills and customer service to this group of stakeholders. The training could be offered by the Tanzania traffic policy department, the National Institute of Transport (NIT) or any training firm with expertise in the transport industry. The second factor is inadequacy of road transport service during peak hours (27.44%). As there is no barrier for the new entrants in this particular market i.e. for most routes of the city, the door is open for potential investors i.e. transport operators to provide public transit services. The third factor is the occurrence of traffic accidents (23.43%). As such, all three aspects contributing to road traffic accidents i.e. human element, vehicle element and road environment should be improved by the relevant parties. The last factor is the prevalence of vehicle emissions and noise pollution (20.88%). Consequently, all stakeholders including the Surface and Marine Transport Regulatory Authority (SUMATRA – Tanzania), vehicle owners and operators should take deliberate efforts to overcome the challenge.

TABLE X. CRISP JUDGEMENT MATRIX DUE TO THE INADEQUACY OF PUBLIC TRANSIT SERVICE

| Factor ($F_{mn}$) | $F_{21}$ | $F_{22}$ | $F_{23}$ | $F_{24}$ | Priority Vector | Rank |
|---|---|---|---|---|---|---|
| $F_{21}$ | 1 | $\frac{17}{23}$ | $\frac{41}{46}$ | $\frac{35}{46}$ | 0.2088 | 4 |
| $F_{22}$ | $\frac{23}{17}$ | 1 | $\frac{41}{34}$ | $\frac{35}{34}$ | 0.2825 | 1 |
| $F_{23}$ | $\frac{46}{41}$ | $\frac{34}{41}$ | 1 | $\frac{35}{41}$ | 0.2343 | 3 |
| $F_{24}$ | $\frac{46}{35}$ | $\frac{34}{35}$ | $\frac{41}{35}$ | 1 | 0.2744 | 2 |

Since $c_{mn} = c_{mk}.c_{kn}$, $\forall m, n, k \in \{1, 2, 3, 4\}$, the matrix above is completely consistent. Thus, $\lambda_{max} = 4$, CI = 0, $R_4 = 0.89$, CR = 0.

*4) Assessment of Transportation Challenges Related to InefficientTraffic Management System*

The experts assessed the factors contributing to inefficient traffic management system to give the vague judgement as represented in Table XI.

We apply equation (1) to get the elements of the crisp judgement matrix represented in Table XII.

$$c_{11} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = 1; c_{12} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2} + a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{31}{36} \quad ;$$
$$c_{13} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3} + a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m1} + a_{m1}')} = \frac{17}{18}$$

$$c_{21} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = \frac{36}{31}; c_{22} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = 1 \quad ;$$
$$c_{23} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')} = \frac{34}{31}$$

TABLE XI. VAGUE JUDGEMENT TABLE DUE TO INEFFICIENT TRAFFIC MANAGEMENT SYSTEM

| **Assessor** | $\mathbf{F_{31}}$ | $\mathbf{F_{32}}$ | $\mathbf{F_{33}}$ |
|---|---|---|---|
| $\mathbf{P_1}$ | [0.6, 0.8] | [0.6, 0.8] | [0.4, 0.6] |
| $\mathbf{P_2}$ | [0.6, 0.8] | [0.8, 1.0] | [0.6, 0.8] |
| $\mathbf{P_3}$ | [0.2, 0.4] | [0.0, 0.2] | [0.4, 0.6] |
| $\mathbf{P_4}$ | [0.2, 0.4] | [0.2, 0.4] | [0.2, 0.4] |
| $\mathbf{P_5}$ | [0.6, 0.8] | [0.4, 0.6] | [0.0, 0.2] |
| $\mathbf{P_6}$ | [0.2, 0.4] | [0.0, 0.2] | [0.0, 0.2] |
| $\mathbf{P_7}$ | [0.0, 0.2] | [0.2, 0.4] | [0.2, 0.4] |
| $\mathbf{P_8}$ | [0.4, 0.6] | [0.2, 0.4] | [0.4, 0.6] |
| $\mathbf{P_9}$ | [0.4, 0.6] | [0.0, 0.2] | [0.4, 0.6] |
| $\mathbf{P_{10}}$ | [0.2, 0.4] | [0.2, 0.4] | [0.4, 0.6] |
| $\mathbf{P_{11}}$ | [0.4, 0.6] | [0.2, 0.4] | [0.4, 0.6] |
| $\mathbf{P_{12}}$ | [0.2, 0.4] | [0.4, 0.6] | [0.6, 0.8] |
| $\mathbf{P_{13}}$ | [0.6, 0.8] | [0.2, 0.4] | [0.4, 0.6] |
| $\mathbf{P_{14}}$ | [0.2, 0.4] | [0.8, 1.0] | [0.2, 0.4] |
| $\mathbf{P_{15}}$ | [0.6, 0.8] | [0.2, 0.4] | [0.4, 0.6] |
| $\mathbf{P_{16}}$ | [0.2, 0.4] | [0.2, 0.4] | [0.2, 0.4] |
| $\frac{1}{32}\sum_{m=1}^{16}(a_{mn} + a_{mn}')$ | $\frac{9}{20}$ | $\frac{31}{80}$ | $\frac{17}{40}$ |

TABLE XII. CRISP JUDGEMENT MATRIX DUE TO THE INEFFICIENCY OF TRAFFIC MANAGEMENT SYSTEM

| **Factor ($\mathbf{F_{mn}}$)** | $\mathbf{F_{31}}$ | $\mathbf{F_{32}}$ | $\mathbf{F_{33}}$ | **Priority Vector** | **Rank** |
|---|---|---|---|---|---|
| $\mathbf{F_{31}}$ | 1 | $\frac{31}{36}$ | $\frac{17}{18}$ | 0.3106 | 3 |
| $\mathbf{F_{32}}$ | $\frac{36}{31}$ | 1 | $\frac{34}{31}$ | 0.3606 | 1 |
| $\mathbf{F_{33}}$ | $\frac{18}{17}$ | $\frac{31}{34}$ | 1 | 0.3288 | 2 |

Since $c_{mn} = c_{mk}.c_{kn},\ \forall m,n,k \in \{1,2,3\}$ , the matrix above is completely consistent. Thus, $\lambda_{max} = 3$, CI = 0, $R_3 = 0.52$, CR = 0.

$$c_{31} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m1}+a_{m1}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = \frac{18}{17}; c_{32} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m2}+a_{m2}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = \frac{31}{34} \quad ;$$
$$c_{33} = \frac{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')}{\frac{1}{32}\sum_{m=1}^{16}(a_{m3}+a_{m3}')} = 1$$

We apply equation (3) to get the following elements of the weight vector whose set is

$$\{W_{F_{31}} = 0.3106, W_{F_{32}} = 0.3606, W_{F_{33}} = 0.3288\}$$

According to experts' views, the first influential factor on inefficient traffic management system in Dar es Salaam city is poor management of drivers and pedestrians' indiscipline (36.06%). This challenge reveals that the Tanzania traffic police department is alleged to provide training to the public on road signs and safety awareness and enforce properly road traffic regulations.

The second factor is poor management of vehicle breakdowns and road accidents (32.88%). Thus, all human, vehicle and road environment aspects need to be improved to curb this problem. The last challenge is poor management of traffic lights and other signals (31.06%). This challenge exists when traffic lights and road signs are not maintained properly. The Tanzania Roads Agency (TANROADS) should carry out routine check and keep abreast of technology (technologies) which can enhance the optimisation of traffic flows.

### *B. Ranking of the Challenges for Urban Public Transpotationt in Dar esSalaam City*

We apply equation (9) to determine the overall priority of the transportation challenges as represented in Table XIII.

TABLE XIII. MATRIX OF PRIORITIZATION OF TRANSPORTATION CHALLENGES IN DAR ES SALAAM CITY

| **Factor ($\boldsymbol{F_m}$)** | **Local weight ($\boldsymbol{W_{F_m}}$)** | **Sub-factor ($\boldsymbol{F_{mn}}$)** | **Local weight ($\boldsymbol{W_{F_{mn}}}$)** | **Overall Priority $\boldsymbol{W_{F_m}.W_{F_{mn}}}$** | **Rank** |
|---|---|---|---|---|---|
| $\mathbf{F_1}$ | 0.3251 | $F_{11}$ | 0.2969 | $W_{F_1}.W_{F_{11}} = 0.0965$ | 6 |
| | | $F_{12}$ | 0.3859 | $W_{F_1}.W_{F_{12}} = 0.1255$ | 1 |
| | | $F_{13}$ | 0.3172 | $W_{F_1}.W_{F_{13}} = 0.1031$ | 5 |
| $\mathbf{F_2}$ | 0.3332 | $F_{21}$ | 0.2088 | $W_{F_2}.W_{F_{21}} = 0.0696$ | 10 |
| | | $F_{22}$ | 0.2825 | $W_{F_2}.W_{F_{22}} = 0.0941$ | 7 |
| | | $F_{23}$ | 0.2343 | $W_{F_2}.W_{F_{23}} = 0.0781$ | 9 |
| | | $F_{24}$ | 0.2744 | $W_{F_2}.W_{F_{24}} = 0.0914$ | 8 |
| $\mathbf{F_3}$ | 0.3417 | $F_{31}$ | 0.3106 | $W_{F_3}.W_{F_{31}} = 0.1061$ | 4 |
| | | $F_{32}$ | 0.3606 | $W_{F_3}.W_{F_{32}} = 0.1232$ | 2 |
| | | $F_{33}$ | 0.3288 | $W_{F_3}.W_{F_{33}} = 0.1124$ | 3 |

Table XIII reveals the overall prioritization of the urban public transportation sub-challenges as follows: Inadequacy of

parking lot and terminals; poor management of drivers and pedestrians' indiscipline; poor management of vehicle breakdowns and road accidents; poor management of traffic lights and signals; inadequacy of traffic lights and road sign equipment; inadequacy of road networks with lay-bys; poor ethical status of drivers and conductors; inadequacy of road transport service; occurrence of traffic accidents; and prevalence of vehicle emissions and noise pollution.

As there is limited financial resources attached to the management issues and investment of new facilities to overcome the proposed challenges, the government can use the ranking of the challenges for prioritization purposes. More specifically, the proposed VAHP model can be used as an investment appraisal tool for selecting the best alternative/project in other industries.

## VI. CONCLUSIONS

Rapid urbanization in Dar es Salaam city – Tanzania has led to the increased demand for public transport services which in turn necessitates improved transport infrastructure, traffic management and public transit services. Limited financial resources make these transportation challenges to be prioritized and allocate funds accordingly. In this paper, we formulate a multi-criteria decision making problem and appraise the transportation challenges using the Vague Analytical Hierarchy Process (VAHP) Model. The results of ranking of the challenges in descending order are poor traffic management, inadequacy of proper public transit service and inadequacy of road transport infrastructure. The findings also reveal that the hybridization of the AHP with the Vague Sets, improve the consistency of the Crisp Judgement Matrices (CJM). Moreover, the computational results validate the VAHP model which can be applied in other research fields as a decision support tool. Our future research direction is to model a hybrid decision support tool composed of Vague Sets, Soft Sets and Rough Sets to assess rural public transportation challenges in Tanzania.

## REFERENCES

[1] T. Harriet, K. Poku, & E.K. Anin, "Logistics Inefficiencies of Urban Transportation System in Ghana," .*International Journal of Humanities and Social Science*, 3(5): 308-314, 2003.

[2] T.L. Saaty, "*The analytic hierarchy process: planning, priority setting and resource allocation*," McGraw-Hill: New York, 1980.

[3] S. Yedla, & R.M. Shrestha, "*Application of Analytic Hierarchy Process to Prioritize Urban Transport Options- Comparative Analysis of Group Aggregation Method,"*Indira Gandhi Institute of Development Research, Mumbai, 2007.

[4] World Bank, "*Urban Transportation,"*Online available at www.web.worldbank.org (Accessed August 09, 2015), 2011.

[5] T.V. Arasan, "Urban Transportation systems planning," Unpublished.

[6] S.Miller, "*The challenges of logistics performance*," Available at http://EzineArticles.com/?expert=sammiller (Accessed August, 2012), 2012.

[7] A.M. Rodrigues, D.J. Bowersox, & R.J. Calantone, "Estimation of global and natural logistics expenditure: 2002 data update"*Journal of Business Logistics*, 26(2): 1 – 15, 2005.

[8] X. Xue, Y. Ren, S. Cui, J. Lin, W.Huang, & J. Zhou, "Integrated analysis of GHGs and public health damage mitigation for developing urban road transportation strategies,"*Transportation Research Part D: Transport and Environment*, 35: 84-103, 2015.

[9] Y. Geng, Z. Ma, B. Xue, W. Ren, Z. Liu, & T. Fujita, "Co-benefit evaluation for urban public transportation-sector-a case of Shenyang, China"*Journal of Cleaner Production*, 58: 82-91, 2013.

[10] D. He, H. Liu, K. He, F. Meng, Y. Jiang, M. Wang, J. Zhou, P. Calthorpe, J. Guo, Z. Yao, & Q. Wang, "Energy use of, and CO2 emissions from China's urban passenger transportation sector-carbon mitigation scenarios upon the transportation mode choices," *Transportation Research Part A: Policy and Practice*, 53: 53-67, 2013.

[11] T. Fan, W. Chiang, & R. Russel, "Modeling urban hazmat transportation with road closure consideration,"*Transportation Research Part D: Transport and Environment*, 35: 104-115, 2015.

[12] E. Fatima, & R. Kumar, "Introduction of public bus transit in Indian cities,"*International Journal of Sustainable Built Environment*, 3(1): 27-34, 2014.

[13] C.E. Siedler, "Can Bus Rapid Transit Be a Sustainable Means of Public Transport in Fast Growing Cities? Empirical Evidence in the case of Oslo," *Transportation Research Procedia*, 1(1): 109-120, 2014.

[14] K. Malecki, S. Iwan, & K. Kijewska, "Influence of Intelligent Transportation Systems on Reduction of the Environmental Negative Impact of Urban Freight Transport Based on Szczecin Example,"*Procedia-Social and Behavioral Sciences*, 151: 215-229, 2014.

[15] Y. Saleh, A. Tofigh, & A. Zahra, "Transportation Routing in Urban Environments Using Updated Traffic Information Provided Through Vehicular Communications,"*Journal of Transportation Systems Engineering and Information Technology*, 14(5): 23-36, 2014.

[16] E. Schreffler, D. Gopalakrishna, D. Vary, D. Friedenfeld, B. Kuhn, C. Dusza, R. Klein & A. Rosas, "Integrating demand management into the transportation planning process: A desk refere3nce,"*Jounal of Institute of Transportation Engineers*, 82(1): 38-41, 2012.

[17] R.E. Msigwa, "Challenges Facing Urban Transportation in Tanzania. *Mathematical Theory and Modelling*," 3(5): 18-26, 2013.

[18] E.P. Massami, "Risk Assessment of Port Competitiveness Based on Vague Soft Sets,"*Transport Policy*, In Press.

[19] H. Shang, (ed.), "Actuarial Science: Theory and Methodology,"Beijing: Higher Education Press, 2006.

[20] E.K. Anin, J. Annan, & F.A. Otchere, "Evaluating the Role of Mass Transit and its Effect on Fuel Efficiency in the Kumasi Metropolis, Ghana,"*International Journal of Business and Social Research*, 3(3): 107-116, 2013.