International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 273-277, 2017, ISSN:-2319–8656

Efficient Utilization of 2D Barcode (QR Code) in Boarding Pass for Managing Luggage at Air Port

Awadhesh Kumar
AIET Jaipur
India

Manish Choubisa
AIET Jaipur
India

S S Shekhawat
AIET Jaipur
India

Manish Dubey
AIET Jaipur
India

Abstract: Quick Response (QR) Code is very useful for encoding the data in an efficient manner. Here data capacity in 2D barcode is limited according to the various types of data formats used for encoding. The data in image format uses more space. The data capacity can be increased by compressing the data using any of the data compression techniques before encoding. In this paper, we suggest a technique for data compression which in turn helps to increase the data capacity of QR Codes generated for image. The main objective of this paper is to present how QR code can be utilized best in boarding pass for managing luggage at air port. Misplace of luggage at air port is very common; here we are proposing an idea for proper handling of misplaced luggage.

Keywords: 2D barcodes, Data Capacity, Data Compression, Lossless Compression, QR Code

1. INTRODUCTION

Bar codes have become widely popular because of their reading speed, accuracy, and superior functionality characteristics. Barcodes can be divided as 1D, 2D and 3D. 1D barcodes can express information in horizontal direction only. Also, the data capacity is limited. 2D barcodes can hold data both in horizontal and vertical direction. As a result, the data capacity is 100 times more than the 1D barcode [1]. 3D barcode is usually engraved on a product or applied on a product so that the barcode has depth and thickness.

As bar codes became popular and their convenience universally recognized, the market began to call for codes capable of storing more information, more character types, and that could be printed in a smaller space. However, these improvements also caused problems such as enlarging the bar code area, complicating reading operations, and increasing printing cost. 2D Code emerged in response to these needs and problems [2].

QR Code is a kind of 2-D (two-dimensional) symbology developed by Denso Wave and released in 1994 with the primary aim of being a symbol that is interpreted by scanning equipment [3]. 2D bar codes can act like identifier (like in 1D) but takes less space. 2-D barcode minimizes the use of database; alternatively, it functions as database itself.

QR Code holds a considerably greater volume of information than a 1D bar code. These can be numeric, alphanumeric or binary data – of which up to 2953 bytes can be stored. Only a part of each QR bar code contains actual data, including error correction information. A large area of the QR code is used for defining the data format and version as well as for positioning, alignment and timing purposes. The smallest square dot or pixel element of a QR code is called a module. QR Codes have an empty area around the graphic. This quiet area is ideally 4 modules wide. Examination certificates can also use the QR Encoding techniques [4].

This paper proposes a method in which data capacity can be increased by first compressing the data and then encoding it. Actual requirement for compression arises when we need to encode image data into QR code. A lossless compression technique is proposed to increase the data capacity. For decoding the data, two steps will be followed: (i) de-compressing the data using the techniques which are just the reverse of compression technique used here and (ii) decoding the decompressed data. For this, the reverse technique used for encoding the data can be used.

2. LITERATURE SURVEY

QR Codes have already overtaken the conventional 1-D bar codes because of the capacity of data that can be stored by a 2-D barcode(QR Code) is much greater than that of conventional 1-D bar code. QR Code contains data both in horizontal and vertical directions. This stems in many cases from the fact that a typical 1-D barcode can only hold a maximum of 20 characters, whereas as QR Code can hold up to 7,089 characters [3]. QR Codes are capable of encoding the same amount of data in approximately one tenth the space of a traditional 1-D bar code. A great feature of QR Codes is that they do not need to be scanned from one particular angle, as QR Codes can be read regardless of their positioning. The data can be read successfully even if QR code is tampered while 1-D barcode can't. QR Codes can be easily decoded with a smart phone with appropriate barcode reader software (for example:, Kaywa Reader, QRafter and I-Nigma etc.) [5]. Secure communication can also be established using QR Encoding techniques [6].

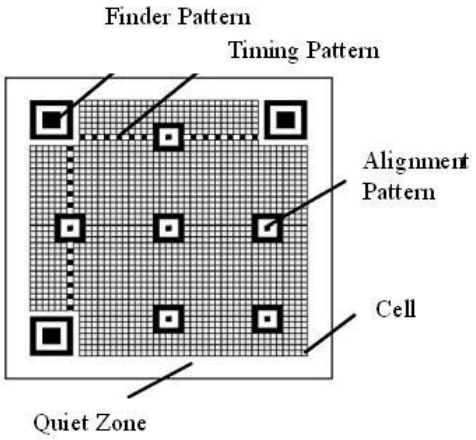

Fig.1: Structure of QR Code

2.1 Structure of QR Codes

QR Codes are actually black modules in square patterns on white background but many researchers have been working for colored QR code. It consists of the following areas having specific significance.

- Finder Pattern
- Alignment Pattern
- Timing Pattern
- Quiet Zone
- Data Area

Fig.1 shows the structure of QR Code. The significance of each area is as described as follows:
Each QR Code symbol consists of mainly two regions: an encoding region and function patterns. Function patterns consist of finder, timing and alignment patterns which does not encode any data. The symbol is surrounded on all the four sides by a quiet zone border [7]. A QR Code can be read even if it is tilted or distorted. The size of a QR Code can vary from 21 x 21 cells to 177 x 177 cells by four cell increments in both horizontal and vertical direction.

2.1.1 Finder Pattern

This pattern can be used for detecting the position, size and angle of the QR Code. These can be determined with the help of the three position detection patterns (Finder Patterns) which are arranged at the upper left, upper right and lower left corners of the symbol as shown in Fig. 1.

2.1.2 Alignment Pattern

The alignment pattern consists of dark 5x5 modules, light 3x3 modules and a single central dark module. This pattern is actually used for correcting the distortion of the symbol [8]. The central coordinate of the alignment pattern will be identified to correct the distortion of the symbol.

2.1.3 Timing Pattern

The timing patterns are arranged both in horizontal and vertical directions. These are actually having size similar to one module of the QR Code symbol. This pattern is actually used for identifying the central co-ordinate of each cell with black and white patterns arranged alternately.

2.1.4 Quiet Zone

This region is actually free of all the markings. The margin space is necessary for reading the bar code accurately. This zone is mainly meant for keeping the QR Code symbol separated from the external area [9]. This area is usually 4 modules wide.

2.1.5 Data Area

It consists of both data and error correction code words. According to the encoding rule, the data will be converted into 0's and 1's. Then these binary numbers will be converted into black and white cells and will be arranged accordingly. Reed-Solomon error correction is also used here [10].

2.2 Data Capacity

The data storage capacity of QR Code is very large as compared to 1-D barcode. The number of characters that can be encoded as QR Code varies according to the type of information that is to be encoded. The various information types and the volume that the QR Code can hold are explained in Table 1.

Table 1. Information Types and Volume of Data

Information Type	Volume of Data
Alphabets and Symbols	**4296**
Numeric Characters	**7089**
Binary Data (8 bit)	**2953**
Kanji Characters	**1817**

2.3 Data Compression

In the history of computer science, data compression, source coding [1] or bit-rate reduction includes encoding information using fewer bits than the original representation. There are two kinds of data compression: lossy and lossless. Lossy compression reduces bits by identifying marginally important information and removing it. Lossless compression reduces bits by identifying and eliminating statistical redundancy. No information is lost in lossless compression.

Data Compression is very useful due to reducing the consumption of resources such as data space or transmission capacity. Because compressed data must be decompressed to be used, this extra processing imposes computational or other costs through decompression. The design of data compression schemes involve trade-offs among various factors, including the degree of compression, the amount of distortion introduced and the computational resources required to compress and uncompress the data [11].

Lossless data compression algorithms usually exploit statistical redundancy to represent data more concisely without losing information. Lossless compression is possible

because most real-world data has statistical redundancy. The Lempel–Ziv (LZ) compression methods are among the most popular algorithms for lossless compression. DEFLATE is a variation on LZ which is optimized for decompression speed and compression ratio, but compression can be slow.

3. PROPOSED TECHNIQUE

As discussed in [13], the efficiency of QR Codes is increased by applying compression before encoding. This paper focuses the best use of the high capacity QR Code generated in [13]. Till now passengers have to get boarding pass for luggage at air port only. Such boarding pass is using 1 D barcodes for every item. Here the main problems associated with 1 D barcode are: i) needs the access of database every time for retrieving complete information and ii) 1 D barcode get tampered easily. Our approach uses QR code which contains the encrypted information. The advantage of encryption is to secure the information stored on QR code. So that only air port authority can encode the data for proper verification of misplaced luggage. The whole process is described by the following steps:

1. The passenger has to fill his/her journey, personal and luggage details as on ticket online. He/she must have the same ID Proof with him/her that is required for authentication of the passenger at the airport.

2. The online system generates a boarding pass having secure barcode for the passenger, which contains encrypted journey and personal details of the passenger. The passenger can take print of secure QR code separately for every item.

3. At the time of check-in at the airport, passenger has to show the QR code on the boarding pass along with ticket to the scanning machine for the verification.

4. Now, the passenger has checked-in the airport successfully.

5. In case of misplace of the luggage during check-out from the air port, the system verifies the luggage left/misplaced with the help of secure QR code affix on it. This verification ensures the authentication of luggage. After successful verification the luggage is handed over to the right passenger. This process ensures the proper handling of misplaced luggage.

The above process is represented in Fig-2(a) and Fig-2(b).

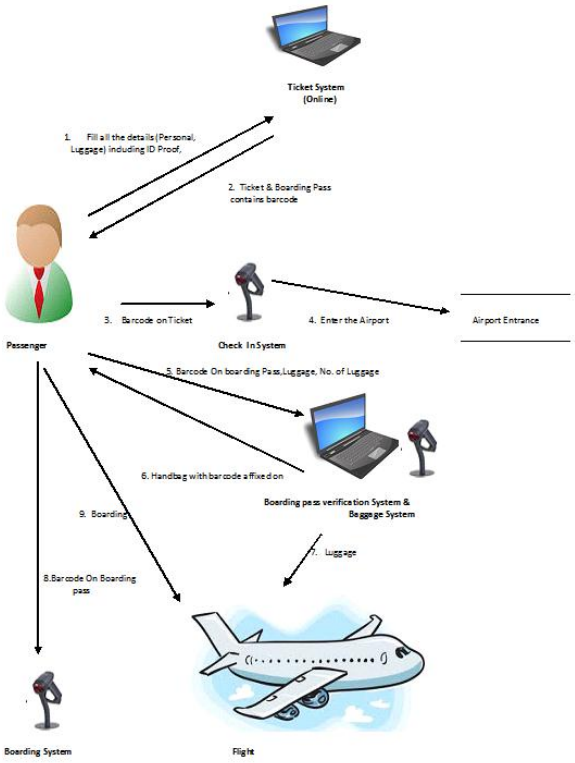

Fig-2(a): Check-in process at Air Port

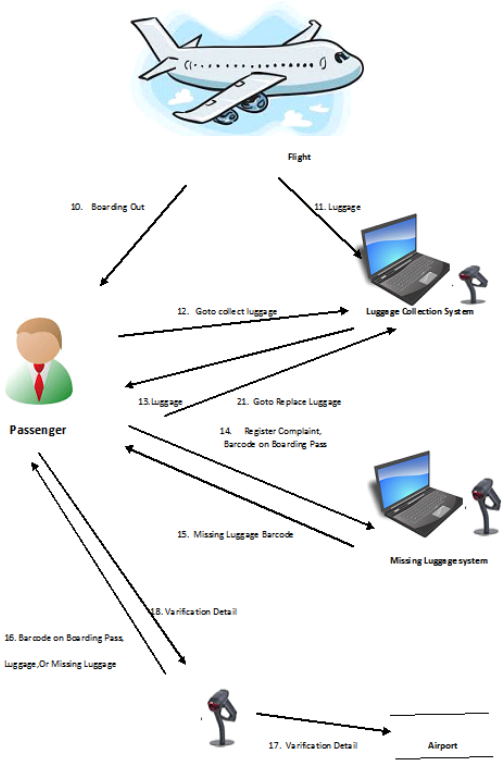

Fig-2(b): Check-in process at Air Port

4. RESULTS

Using the approach discussed above, we are able to provide the facility of generating boarding pass online from home or office. The QR code is generated for every item carried by passenger. He/she can affix QR code on items and the same QR code is on boarding pass also. During check-in at the air port luggage are verified with help of QR code on boarding pass. The misplaced luggage are handled using QR code on it. The whole process is implemented by designing a small web application using C#.Net on Visual Studio 2008. Fig-3 shows the registration process of luggage:

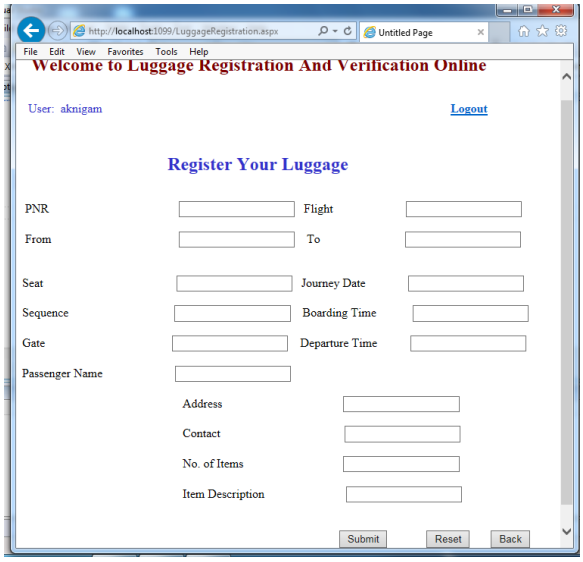

Fig-3: Luggage Registration Process

Fig-4 Shows the main page of user after login for luggage registration and verification.

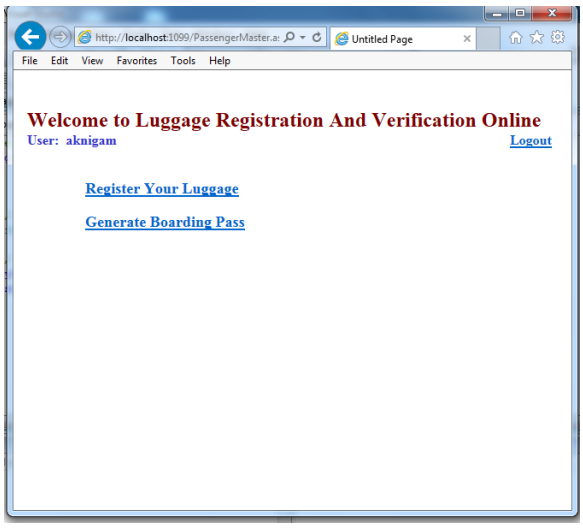

Fig-4: Main Page of User

Fig-5(a) shows the generation of boarding pass and QR code for luggage items.

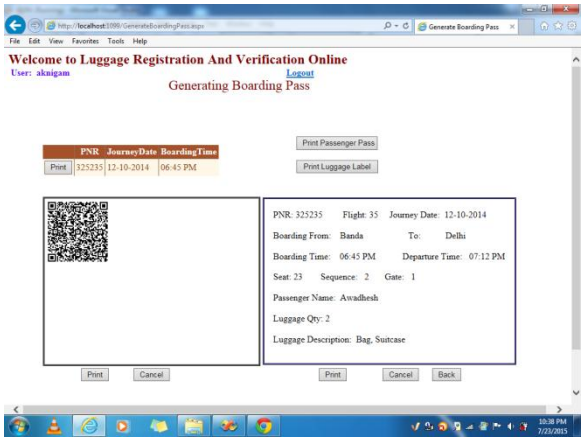

Fig-5(a): Generation of Boarding Pass with QR code

Fig-5(b) shows the successful verification of luggage information encoded in QR code. The QR code on luggage items is read by barcode scanner. Here we verifying the QR code by browsing the image of QR code.

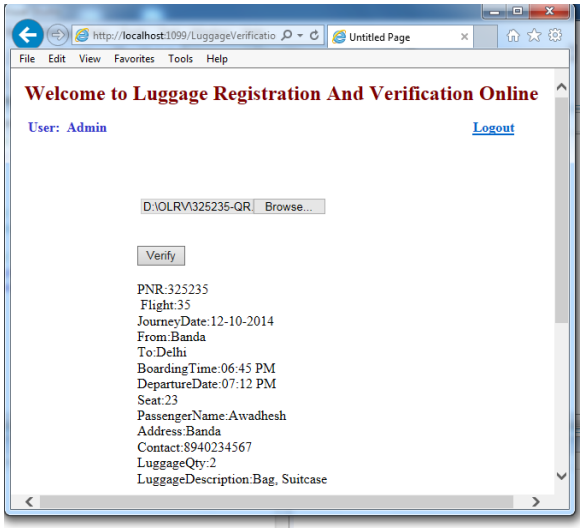

Fig-5(b): Verification of QR Code

5. CONCLUSIONS

Normal QR Codes can compress only up to 4 KB of data. Using the techniques followed here, the data capacity can be increased drastically. As compared to the normal QR Codes, the data capacity of the QR Code after following technique was found to be more than 4 KB. Efficient data compression techniques can be used to store more than 4 KB of data inside a QR Code. A variety of data compression techniques can be used to obtain more data storage capacity. Comparing with the existing technologies used to generate bar codes, QR Codes were found to be of great advantage to the manufacturer because of its great data storage capacity,

reading speed and accuracy. The data capacity was further improved by combining the most distinguishing features of compression and bar code generation. Using this novel technique of data compression followed by data encoding, the data storage capacity of QR Codes were increased drastically.

6. FUTURE SCOPE

Currently only Smartphone's are technically equipped to do this. Many users that have mobile phones that have cameras are unable to get QR reading software for their phones. Future enhancements focus on QR Encoding of images which is more than 4 KB of size. Secure QR Coding can also be implemented using encryption techniques. Also, more advanced data compression techniques can be used to add more to the data capacity of the normal QR Codes.

6. REFERENCES

[1] Xiaofei Feng, Herong Zheng, "Design and Realization of 2D Color Barcode with High Compression Ratio" 2010 International Conference On Computer Design And Appliations (ICCDA 2010), 978-1-4244-7164-51, 2010 IEEE, 978-1-4244-7164-51, 2010 IEEE, Volume 1

[2] Nancy Victor, "Enhancing the Data Capacity of QR Codes by Compressing the Data before Generation", International Journal of Computer Applications (0975- 8887), Volume 60 - No.2, December 2012.

[3] Peter Kieseberg, Manuel Leithner, Martin Mulazzani, Lindsay Munroe, Sebastian Schrittwieser, Mayank Sinha, Edgar WeipplT. J., "QR Code Security"

[4] Chun-lei XIA, "Examination Certificate Based on Two-Dimensional Bar Code Technology", 2008 International Symposium on Computer Science and Computational Technology, 978-0-7695-3498-5/08/2008 IEEE DOI 10.1109/ISCSCT.2008.102

[5] Tasos Falas, Hossein Kashani, "Two-Dimensional Bar-code Decoding with Camera-Equipped Mobile Phones", Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops(PerComW'07) 0-7695-2788-4/07/2007

[6] William Claycomb, Dongwan Shin, "Using A Two Dimensional Colorized Barcode Solution for Authentication in Pervasive Computing", 1-4244-0237-9/06/2006 IEEE.

[7] M.Pitchaiah Philemon Daniel, Praveen, "Implemation of Advanced Encryption Standard Algorithm", nternational Journal of Scientific & Engineering Research Volume 3, Issue 3, March -2012 1 ISSN 2229-5518.

[8] Guenther Starnberger, Lorenz Froihofer and Karl M. Goeschka, "QR-TAN: Secure Mobile Transaction Authentication", 2009 International Conference on Availability, Reliability and Security, 978-0-7695-3564-7/09 IEEE DOI 10.1109/ARES.2009.96

[9] ISO/IEC 18004:2000 Information Technology - Automatic Identification and Data Capture Techniques – Barcode Symbology- QR Code (MOD), June 2000.

[10] Sarah Lyons and Frank R. Kschischang, "Two-Dimensional Barcodes for Mobile Phones", 25th Biennial Symposium on Communications, 978-1-4244-5711-3/10/2010

[11] R. Bose and D. Ray-Chaudhuri. On a class of errorcorrecting binary group codes*. Information and control, 3(1):68{79, 1960.

[12] David L. Donoho, Martin Vetterli, Fellow, IEEE, R. A. DeVore, and Ingrid Daubechies, Senior Member, IEEE," Data Compression and Harmonic Analysis", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 44, NO. 6, OCTOBER 1998, 0018–9448/98$10.00 ã 1998 IEEE

[13] Hee I1 Hahn and Joung Koo Joung, "Implementation of Algorithm to Decode Two-Dimensional Barcode PDF-417", ICSP'O2 Proceedings, 0-7803-7488-6/02/$17.00 Q 2002 IEEE.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 278-284, 2017, ISSN:-2319–8656

Content Based Image Retrieval with Multi-Feature Classification by Back-propagation Neural Network

Suman Khokhar
Department of Computer Science and Engineering
MATS University
Raipur, India

Satya Verma
Head of Department
Computer Science and Engineering
MATS University
Raipur, India

Abstract: Emergence of Internet, as well as digital image acquisition technology, has increased the usage of rich visual information such as images and videos. It has become the integral and essential part of everyone's life. Since image production has become easy and economical, images and videos are used extensively on the Internet so retrieving relevant images from large ever-growing image dataset has become a challenge. To combat this problem, one of the popular image retrieval approaches is Content-Based Image Retrieval (CBIR) which utilizes the visual features of the image i.e. color, texture, geometric (shape) and spatial information to retrieve visually similar images according to the given query image. This paper aims to exploit multiple features of an image i.e. color, geometric and texture with the Back-propagation Feedforward Neural Network (BFNN) for classification. Feature selection method is also exercised to focus on important features of an image and ignoring redundant and inappropriate information. The results have shown that CBIR technique with multi-feature classification by BFNN yields better precision and recall as compared to other state-of-the-art techniques of CBIR that uses single or other hybrid combination of features of an image.

Keywords: Back Propagation Feedforward Neural Network (BFNN); Content-Based Image Retrieval (CBIR); Gray Level Co-occurrence Matrix (GLCM); Text-Based Image Retrieval (TBIR); Zernike Moments (ZM);

1. INTRODUCTION

The Internet and World Wide Web (W3) has revolutionized everyone's life. Today it has become a preferred medium of daily communication. There is an English idiom which says "A picture/image is worth a thousand words". Any complicated idea can be easily remembered and conveyed via a single still image. An image exhibits some meaning and can be thought of a visual representation of some subject.

1.1 Motivation

Due to the scientific and technological advances in data storage and image acquisition methodologies, the volume of digital data i.e. images and videos have increased considerably [1]. To combat the problem of the ever-growing population of images, their organization, and management, image retrieval systems were developed. These systems retrieve images from the large image database on the basis of the user's query. There are three kinds of retrieval systems which are depicted in Figure 1.

Since 1979, image retrieval systems have been exploiting the techniques that are based on the textual description or annotation of images. This approach is known as *Text / Tag Based Image Retrieval (TBIR)* [2].

There are certain drawbacks of TBIR approach i.e. it is a time-consuming and slow process as it requires manual description of each image in database, costly and different human have a different understanding of the image making it subjective and context-sensitive.

An image can describe itself completely and efficiently as compared to any other descriptor, therefore an another approach called *Content Based Image Retrieval (CBIR)* was introduced and has been a major area of research in image processing field. CBIR takes a query image (i.e. example image or sketch) as an input and retrieves visually similar images by comparing visual features of query image with the visual features of images in the image database. It uses the content of the images to represent and index them.

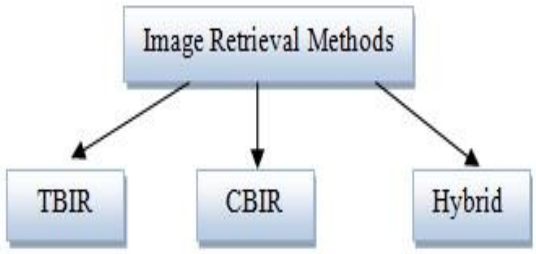

Figure 1: Types of Image Retrieval System

The visual features of an image include color, shape (geometric), texture and any other information that can be extracted from the image. These visual features are described by multi-dimensional feature vectors. This approach is also known by different names such as *Query-By-Image Content (QBIC)*, *Content-Based Visual Information Retrieval (CBVIR)* and *Reverse Image Search.* CBIR has found its application in many fields like medical diagnosis, satellite picturing, military, intellectual property, object recognition, fingerprint identification, face recognition, etc.

In *Mixed / Hybrid approach*, both textual query and visual features are combined (also known as fusion) to obtain the desired output image. This is a demanding research area in the field of multimedia information retrieval.

1.2 Related Work

In [3], Swapnil Saurav *et al.* extracted color features using various color models and different color descriptors such as color histogram, color moments and color coherence vectors with Support Vector Machines (SVM) classifier. In [4], Muhsina Kaipravan *et al*. proposed a system which integrates both color and texture features. To extract color and texture features, color moment and Gabor filter are employed respectively. This technique is implemented and verified on Wang image database.

In [5], Manpreet Kaur *et al.* presented a novel technique which combines image features like color, texture, and edges. Retrieval techniques like HSV, color moment are used for extracting color features, Gabor wavelet and wavelet transform for retrieving texture features with edge gradient technique for detecting shape features. In [6], Sakthivel Karthikeyan *et al.* proposed a system for Glaucoma diagnosis via Gray Level Co-occurrence Matrix (GLCM) for extracting the textual features and used Sequential Forward Floating Selection (SFSS) technique for selecting extracted features which are fed to Back Propagation Network (BPN).

In [7], Habib Mahi *et al.* proposed a system to recognize the buildings from Very High Spatial Resolution (VHSR) satellite images. The system uses Zernike Moments based shape descriptor after segmenting the image into homogeneous objects via MeanShift segmentation method and is fed to SVM for further classification purpose. In [8], Sudhir P. Vegad *et al.* proposed a hybrid approach which exploits color features with feed-forward BPN. The system uses HSV color model for obtaining color features.

The proposed CBIR system extracts color, shape and texture features of the image by applying *RGB Color Histogram*, *Zernike Moments* and *Gray Level Co-occurrence Matrix (GLCM)* respectively to attain the optimal retrieval accuracy and better performance in CBIR systems. Once all features are extracted from the image, *ReliefF* Filter Feature Selection technique is applied to choose appropriate features required to represent an image effectively. The System is trained and images are classified using *Back-propagation Feed-forward Neural Network (BFNN).*

Since *LAB color space* model exceeds the gamut of the RGB and CMYK color models and is designed to approximate human vision, it is employed as a pre-processing step to extract texture and shape features. Images with the close resemblance to the query image are retrieved and indexed accordingly.

This paper is organized as follows: In Section II, CBIR architecture is explained. Proposed work is discussed in detail in Section III with feature extraction techniques. Section IV presents experimental results with the discussion on performance evaluation of different techniques. Finally, the paper is concluded in Section V with future research work recommendations.

2. CBIR ARCHITECTURE

The CBIR system analyzes image content and retrieves visually similar images from image database according to the query image. A query image can be any sketched figure or an example image. The visual content of an image is known as image features which are extracted by applying different techniques and are internally represented as feature vectors.

The image features are the properties of an image, for instance, color, texture, shape or spatial layout and these features are considered as low-level features of an image. CBIR system performs two main functions i.e. Feature Extraction and Similarity Measurement which is shown in Figure 2.

In the *Feature Extraction* phase, the features of images are extracted and these features are matched in *Similarity Measurement* phase.

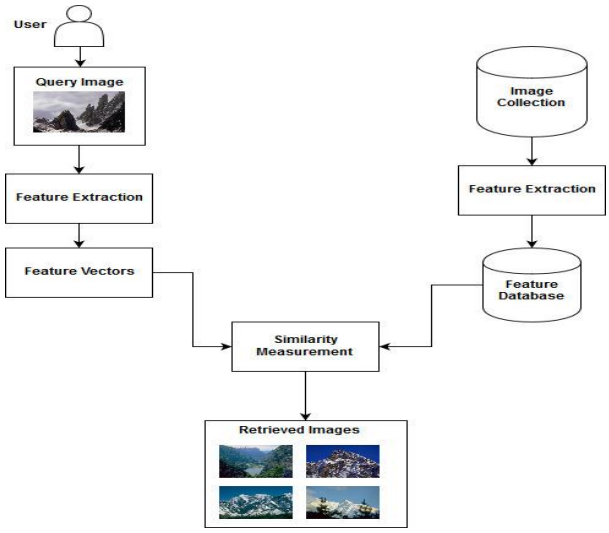

Figure 2: CBIR System

The feature vector of each image in the database is compared with the feature vector of the query image and the images with the small distance to the query image are considered as matched images and are retrieved from the image database.

3. PROPOSED APPROACH

The CBIR system extracts color, shape and texture features of the image. The performance of CBIR system depends on the method to obtain suitable features from the image. *The color* is the most widely used visual property of the image. It is a robust descriptor as it is uniform with respect to scaling, translation, and rotation of an image. *Texture or gray-level features* depicts visual patterns in an image and describes how they are spatially located. GLCM method is a very effective method for acquiring texture features and is applied in this paper.

The shape of an object or region in an image is represented by shape features. These features are commonly extracted and obtained after image segmentation. Region- growing method of region-based segmentation is employed from which Zernike moments are computed. The details of the methods used for obtaining image features are stated below in detail:

3.1 Gray-Level Co-occurrence Matrix

GLCM is a popular statistical method for obtaining texture features from images. It calculates co-occurring grayscale values at a specified offset over an image.

It produces a matrix with the direction (i.e. horizontal, vertical, left and right diagonals) and distances between the pixels. It is also known as Grey Tone Spatial Dependency Matrix and is used for calculating the "second order" texture features. GLCM is a square matrix with dimension N_g, where N_g refers to the number of gray levels in the image.

Each element [i, j] of the GLCM matrix is calculated by counting the number of times a pixel holding value i is adjacent to a pixel holding value j i.e. *V[i, j]*. The relation between the two pixels is computed at a time by considering two-pixel values i.e. reference pixel and neighbor pixel. The following Table 1 shows the structural arrangement of GLCM matrix elements.

Table 1: GLCM Matrix Structure

Neighbor pixel value / Reference Pixel value	0	1	2	…	N_g
0	0,0	0,1	0,2	…	0, N_g
1	1,0	1,1	1,2	…	1, N_g
2	2,0	2,1	2,2	…	2, N_g
. . .	…	…	…	…	…
N_g	N_g,0	N_g,1	N_g,2	…	N_g, N_g

Since the texture measure requires that each GLCM cell contains not a count, but rather a probability, therefore GLCM is normalized by dividing each element by the sum of values. The formula for computing probability matrix element is given below where matrix element P(i, j | Δx, Δy) depicts relative frequency with the given two neighborhood pixels, one with intensity 'i' and the other with intensity 'j' and pixels are separated by a pixel distance (Δx, Δy) [9].

$$P(i,j) = \frac{V(i,j)}{\sum_{i,j=0}^{Ng} V(i,j)} \qquad (1)$$

The GLCM expressed as a probability matrix is given by

$$G = \begin{bmatrix} P(0,0) & P(0,1) & \cdots & P(0,N_g) \\ P(1,0) & P(1,1) & \dots & P(1,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ P(N_g,0) & P(N_g,1) & \dots & P(N_g,N_g) \end{bmatrix}$$

This paper extracts 22 features from GLCM matrices and their formula [6][7][8][9][10][14] is mentioned in Table2.

Table 2: Features extracted from GLCM

S. No.	**Features**	**Formula**
1	Entropy	$\sum_{i,j=0}^{N_g} P(i,j)\log(P(i,j))$
2	Energy	$\sum_i \sum_j P(i,j)^2$
3	Contrast	$\sum_i \sum_j (i-j)^2 \, P(i,j)$
4	Homogeneity	$\sum_i \sum_j \frac{P(i,j)}{(1+(i-j)^2)}$
5	Correlation	$\frac{\sum_i \sum_j (i,j)P(i,j) - \mu_x \ \mu_y}{\sigma_x \sigma_y}$
6	Variance: Sum of Squares	$\sum_i \sum_j (i-\mu)^2 P(i,j)$
7	Autocorrelation	$\frac{\sum_i \sum_j (i,j)[P(i,j)-\mu_x][P(i,j)-\mu_y]}{\sigma_x \sigma_y}$ Where μ_x , μ_y are means and σ_x, σ_y are standard deviations.
8	Inverse Difference (INV)	$\sum \frac{P(i,j)}{1+\lvert i-j \rvert}$
9	Sum Average	$\sum_{i=2}^{2N_g} iP_{x+y}(i)$
10	Sum Entropy	$f_8 = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log\{P_{x+y}(i$
11	Sum Variance	$\sum_{i=2}^{2N_g} (i-f_8)^2 P_{x+y}(i)$
12	Difference Variance	$\sum_{i=0}^{N_g-1} i^2 P_{x-y}(i)$
13	Difference Entropy	$-\sum_{i=0}^{N_g-1} P_{x-y}(i)\log\{P_{x-y}(i)\}$
14	Information Measure of Correlation1	$\frac{HXY - HXY_1}{\max\{HX, HY\}}$ where $HXY = -\sum_i \sum_j P(i,j)\log(p(i,j))$ $HXY_1 = \sum_i \sum_j P(i,j)\log\{P_x(i)P_y(j)\}$ HX, HY are the entropies of P_x and P_y.

15	Information Measure of Correlation2	$(1-\exp[-2(HXY_2 - HXY)])^{\frac{1}{2}}$ where $HXY_2 = -\sum_i\sum_j P_x(i)P_y(j)\log\{P_x(i)P_y(j)\}$
16	Maximum Correlation Coefficient	Square root of the second largest eigenvalue of Q where $Q(i,j) = \sum_k \frac{P(i,k)P(j,k)}{P_x(i)P_y(k)}$
17	Shade	$sgn(A)\lvert A\rvert^{\frac{1}{3}}$ where A= $\sum_{i,j=0}^{N_g-1} \frac{P(i,j)(i+j-2\mu)^3}{\sigma^3(\sqrt{2(1+C)})^3}$ and C= Correlation feature
18	Prominence	$sgn(B)\lvert B\rvert^{\frac{1}{4}}$ where A= $\sum_{i,j=0}^{N_g-1} \frac{P(i,j)(i+j-2\mu)^4}{4\sigma^4(1+C)^2}$
19	Inverse Difference Moment Normalized (IDN)	$\frac{\sum_i\sum_j \frac{1}{1+(i-j)^2}P(i,j)}{\sum_i\sum_j P(i,j)}$
20	Inverse Difference Normalized (INN)	$\frac{INV}{\max(INV)}$
21	Maximum Probability	$\max\{P(i,j)\}$
22	Dissimilarity	$\sum_i P(i,j)\lvert i-j\rvert$

3.2 Zernike Moments

ZM is an excellent shape descriptor due to its description capability. It is a region-based moment and is very effective in image classification system. It is simply a projection of image functions onto polynomial orthogonal basis function. Rotation invariant and orthogonal property of ZM makes it suitable for shape feature extraction. The ZM can be computed on an image with the size N X N with the following expression [11].

$$Z_{n,m} = \frac{n+1}{\lambda_N}\sum_{c=0}^{N-1}\sum_{r=0}^{N-1} f(c,r)R_{n,m}(\rho_{cr})e^{-jm\theta_{cr}} \quad (2)$$

Where f(c, r) is the image function in which c is column and r is row, λ_N represents a normalization factor, n € N in which n is a positive integer and represents the order of the radial polynomial and $R_{n,m}(.)$ represents the real-valued 1-D radial polynomial, m represents repetition of the angle and satisfies the following constraints:

N - |m| = even and |m| ≤ n

ρ_{cr} is transformed distance and ranges from $0 \le \rho_{cr} \le 1$ and θ_{cr} is a phase which ranges from $0 \le \theta_{cr} \le 2\prod$ at the pixel (c, r) and has the following formula:

$$\rho_{cr} = \frac{\sqrt{(2c-N+1)^2+(2r-N+1)^2}}{N} \quad (3)$$

$$\theta_{cr} = tan^{-1}\left(\frac{N-1-2r}{2c-N+1}\right) \quad (4)$$

This paper extract features using ZM where the order of the moment i.e. n is equal to 4 and repetition of the moment i.e. m is equal to 2.

3.3 Backpropagation Neural Network

BFNN is a network that is trained with a backpropagation training algorithm that calculates error signal by finding the difference between the training output from the target output. It backpropagates the errors, adjusts the weights and biases in the input and hidden layers to get close to the desired outcome. The architecture of BFNN consists of three layers i.e. input, hidden, and output layer [12] which is given in Figure 3.

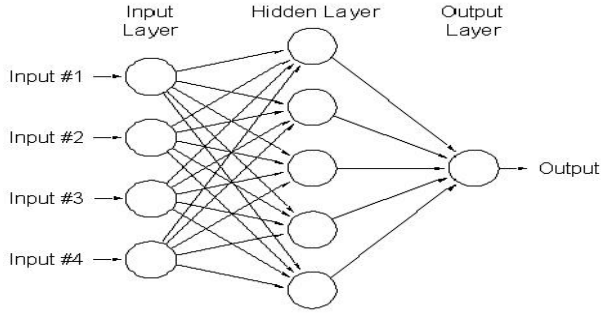

Figure 3: Basic BFNN Architecture

Images are divided into different classes in the image database and some images from each class are used for training the BFNN. A BFNN is repeated until a less training error is achieved. Once training sample is fed to the BFNN, the trained BFNN assigns one class to the query image and retrieve images of the same class as an output.

It sorts the output according to the following formula [12] which represents the distance between the query image and the retrieved images:

$$d_i = \left(\frac{1}{N}\sum_{j=1}^{N} W_j(f_j^i - f_j^q)^2\right)^{\frac{1}{2}} \quad (5)$$

Where f_j^q represents the j[th] feature of the query image, f_j^i represents the j[th] feature of the retrieved image, W_i is the weight of j[th] feature and N is the dimension of the feature vector.

3.4 RGB Color Histogram

Color histogram identifies the proportion of pixels within an image carrying specific values. As it is independent of image size or orientation, the color histogram is obtained from RGB images with a number of bins equal to 3.

3.5 ReliefF Filter Feature Selection

Feature Selection is considered as a basic problem in pattern recognition system. As the selection of features in filter selection method is independent on the classifier used, therefore ReliefF filter method is implemented to calculate the quality of attributes or features by finding how their values differentiate between instances. It is very effective in solving

multi-class problems and is robust to noisy and incomplete data.

3.6 Proposed Algorithm

1. Load an RGB query image.
2. Implement LAB conversion which yields lightness (L) component and two color components a, and b.
3. Perform region-growing based segmentation on L component of the query image.
4. Apply Zernike Moments on the segmented image to obtain amplitude (A) and phase (phi) value of the moment.
5. Apply GLCM technique for extracting 22 texture features from query image.
6. Color features are obtained using RGB histogram with a number of bins equal to 3, therefore extracting 27 color features.
7. while (image in the training set)
8. do
9. LAB conversion
10. calculate ZM and GLCM features
11. obtain color features
12. add features to feature database
13. end;
14. end while;
15. Apply ReleifF feature selection technique which returns top 20 relevant features for similarity measurement.
16. BFNN is used as a classifier which takes 20 selected features for training network with 70% training data, 15% validation data and 15% for testing data with the number of hidden neurons as 20.
17. Finally, the proposed system returns class name of the query image and retrieves similar images.

3.7 Proposed System Flowchart

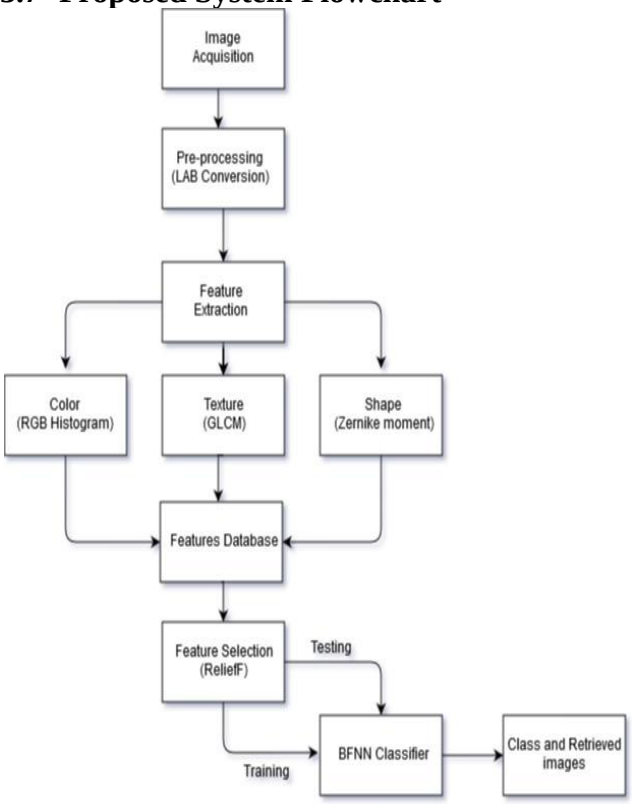

Figure 4: Proposed System Architecture

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed system, MATLAB (R2016a) framework is used. The assessment of the performance of image retrieval methods is accomplished using two important measures known as *Precision* and *Recall*.

4.1 Implementation Environment and Dataset

The user interacts with the system via Graphical User Interface (GUI) which is created by a tool called GUIDE, the GUI Development Environment in MATLAB. To check the effectiveness of our system, the experiments are conducted on Wang image database. The Wang image dataset consists of 1000 images which are divided into 10 classes where each class includes 100 images. The categories of Wang dataset is presented in Table 3.

Table 3: Categories of Wang Image Dataset

S.No.	Categories	Images	S.No.	Categories	Images
1	African		6	Flower	
2	Beach		7	Food	
3	Bus		8	Hills	
4	Dinosaur		9	Horse	
5	Elephant		10	Monument	

4.2 Comparative Analysis and Performance Measurement

Precision depicts the “quality or exactness” of image retrieval whereas recall represents the “completeness.

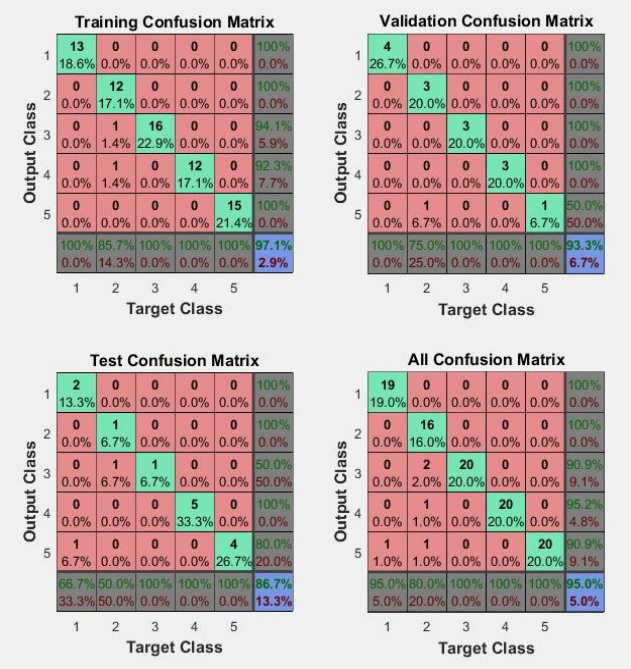

Figure 5: Plot of Confusion Matrix

Precision tells how many retrieved items are relevant and recall depicts how many relevant items are selected from the database. Precision and recall values are calculated from *confusion matrix* which is also known as error matrix and is a table that describes the performance of the classifier.

The confusion matrix for the first 5 image categories (i.e. African, Beach, Bus, Dinosaur, and Elephant) is displayed below in figure 5 providing African image as a query.

The formula for calculating the precision, recall, average precision and average recall is given below:

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$Precision_{avg}^{n,j} = \frac{1}{N_i}\sum_{i}^{N_i} precision_i(n) \quad (9)$$

$$Recall_{avg}^{n,j} = \frac{1}{N_i}\sum_{i}^{N_i} recall_i(n) \quad (10)$$

where TP is true positive, FP is false positive, and FN is false negative value in confusion matrix.

Table 4: Precision values of 5 categories

Class					
Precision	**African**	**Beach**	**Bus**	**Dinosaur**	**Elephant**
1	100	75	100	100	100
2	95	80	100	100	100
3	100	80	85	100	100
4	95	95	95	100	85
5	100	80	100	90	100
6	100	85	85	100	100
7	100	100	85	100	95
8	100	85	95	100	85
9	95	90	100	95	95
10	100	85	85	100	100

The Table 4 and 5 represent the precision values and recall values for evaluating the performance of BFNN classifier for 5 categories. The average precision and recall of proposed system are displayed in Figure 6 whereas comparative analysis of proposed system with the color moment and Gabor filter technique [4] is presented in Figure 7.

Table 5: Recall values of 5 classes

Class					
Recall	**African**	**Beach**	**Bus**	**Dinosaur**	**Elephant**
1	90	100	87	100	100
2	100	100	90	95	90
3	100	84	94	100	86
4	100	90	86	95	100
5	90	100	95	100	87
6	95	89	94	100	90
7	90	90	100	100	100
8	90	93	81	100	90
9	95	100	90	100	100
10	95	89	94	100	90

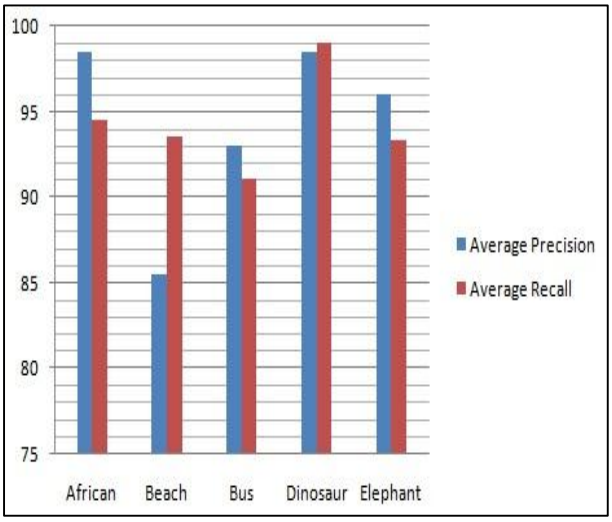

Figure 6: Average Precision and Recall values of 5 classes

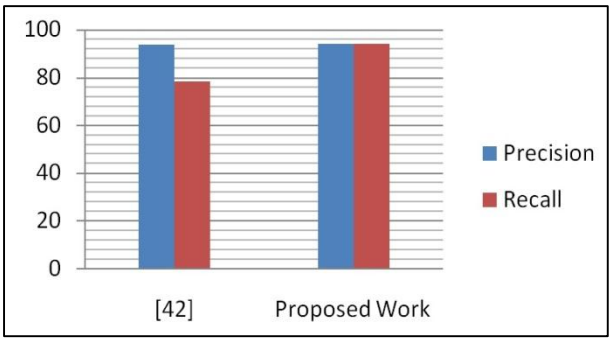

Figure 7: Comparison of proposed work with existing technique [4]

5. DISCUSSION

By the experimentation conducted on Wang image dataset, it is observed that the proposed system has attained the overall precision of 94.3% and recall of 94.28% in contrast to 94% precision and 78.4% recall of the system [4] with higher retrieval accuracy of 94.28%.

The proposed system has been able to achieve the optimal accuracy and better performance with significant improvement in recall measure.

6. CONCLUSION

This paper presented a CBIR approach that combines multiple visual features of the image to retrieve visually similar images. As only one or two features are not sufficient to describe an image completely, it is suggested to use all low-level features i.e. color, shape, and texture of the image to realize it effectively. Texture features are extracted using GLCM technique, geometric features are computed by applying Zernike Moments whereas, for color features, the RGB color histogram is employed. All extracted features are applied to ReliefF feature selection algorithm which selects top 20 features for BFNN classification. Experimental results show that the performance of the proposed work outperforms the other conventional image retrieval methods by attaining higher precision values, recall values, and accuracy of 94.28%.

In the future work, deep neural network (DNN) or more efficient color extraction technique can be availed to further improve image retrieval accuracy. The system can be customized for transfer learning.

7. REFERENCES

[1] Oge Marquees, Florida Atlantic University. 2016. Visual Information Retrieval – The State of the Art, Published by the IEEE Computer Society.

[2] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng. Fundamentals of Content Based Image Retrieval.

[3] Swapnil Saurav, Prajakta Belsare, and Siddhartha Sarkar. 2015. Holistic Correlation of Color Models, Color Features and Distance Metrics on Content-Based Image Retrieval, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 02 Issue: 07.

[4] Muhsina Kaipravan, Rejiram R. 2016. A Novel CBIR System Based on Combination of Color Moment and Gabor Filter, Conference on Data Mining and Advanced Computing (SAPIENCE), INSPEC Accession Number: 16540589, IEEE.

[5] Manpreet Kaur, NeelofarSohi. 2016. A Novel Technique For Content Based Image Retrieval Using Color, Texture And Edge Features, International Conference on Communication and Electronics Systems (ICCES), INSPEC Accession Number:16776306, IEEE.

[6] Sakthivel Karthikeyan, N. Rengarajan. 2014 Performance analysis of gray level cooccurrence matrix texture features for Glaucoma diagnosis, American Journal of Applied Sciences 11 (2): 248-257, 2014 ISSN: 1546-9239, Science Publication.

[7] Habib Mahi, Hadria Isabaten, and Chahira Serief. 2014. Zernike moments and SVM for Shape Classification in Very High-Resolution Satellite Images, The International Arab Journal of Information Technology, Vol. 11, No. 1.

[8] Sudhir P. Vegad, Prashant and K. Italiya. 2015. Image Classification using Neural Network for Efficient Image Retrieval, International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO).

[9] Seema Kolkur, D.R. Kalbande. 2016. Survey of Texture-Based Feature Extraction for Skin Disease Detection, IEEE.

[10] Robert M. Haralick, K. Shanmugam, and It'shak Dinstein.1973. Textural features for Image Classification, IEEE Transactions on systems, man, and cybernetics.

[11] David A. Clausi. 1989. An analysis of co-occurrence texture statistics as a function of gray level quantization, Canadian Journal of Remote Sensing, Can. J. Remote Sensing, Vol. 28, No. 1, pp. 45–62, 2002M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science.

[12] Amir Tahmasbi n, FatemehSaki, and Shahriar B. Shokouhi. 2011. Classification of benign and malignant masses based on Zernike moments, Computers in Biology and Medicine, 726–735, Elsevier.

[13] Jun-Hua Han, De-Shuang Huang. 2005. A Novel BP-Based Image Retrieval System, IEEE.

[14] Michael V. Boland.1999. Haralick texture features. http://murphylab.web.cmu.edu/publications/boland/boland_node26.html, Available, [online] [Accessed]:12-Apr-2017.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 285-289, 2017, ISSN:-2319–8656

A Online Algorithms for Scoring Bank Customers

Abolfazl Tanha
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

*Faramarz Sadeghi
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

Correspondence: Faramarz Sadeghi, Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran.

Abstract: The technical unification of compilers and reinforcement learning is a typical issue. Given the current status of event-driven algorithms, systems engineers dubiously desire the simulation of the partition table. We demonstrate that while the foremost extensible algorithm for the visualization of simulated annealing by Li runs in $O(n^2)$ time, B-trees and SMPs can interfere to realize this goal.

Keywords: Data mining, Scoring Bank, classification

1. INTRODUCTION

Recent advances in large-scale symmetries and cooperative communication do not necessarily obviate the need for the Turing machine [9]. In fact, few cyberneticists would disagree with the significant unification of robots and Internet QoS, which embodies the extensive principles of artificial intelligence. To put this in perspective, consider the fact that little-known physicists largely use write-ahead logging to realize this purpose. The study of RPCs would profoundly improve the simulation of neural networks.

Nevertheless, this method is fraught with difficulty, largely due to spreadsheets. Contrarily, empathic communication might not be the panacea that statisticians expected. On a similar note, though conventional wisdom states that this quandary is entirely solved by the study of public-private key pairs, we believe that a different solution is necessary. We view cryptoanalysis as following a cycle of four phases: study, prevention, prevention, and synthesis. Thusly, we disconfirm that architecture can be made relational, Bayesian, and linear-time.

Motivated by these observations, linked lists and the location-identity split [22] have been extensively deployed by information theorists. Similarly, two properties make this solution ideal: our application is copied from the understanding of DHTs, and also ADAGE is maximally efficient. We view cyberinformatics as following a cycle of four phases: creation, emulation, study, and improvement. We emphasize that ADAGE is derived from the analysis of lambda calculus. The disadvantage of this type of approach, however, is that virtual machines [22] can be made certifiable, Bayesian, and secure. Of course, this is not always the case. This combination of properties has not yet been improved in previous work.

In our research, we explore a cacheable tool for analyzing model checking [4] (ADAGE), showing that active networks and virtual machines can interact to realize this purpose. The basic tenet of this solution is the visualization of Markov models. This might seem counterintuitive but never conflicts with the need to provide superblocks to analysts. For example, many solutions create optimal configurations. Though similar frameworks synthesize adaptive technology, we fulfill this intent without analyzing cooperative symmetries.

The rest of this paper is organized as follows. To begin with, we motivate the need for checksums. Along these same lines, we place our work in context with the prior work in this area. We validate the construction of extreme programming. Similarly, to accomplish this intent, we disconfirm not only that the much-touted encrypted algorithm for the study of online algorithms by N. Brown [2] is impossible, but that the same is true for the Ethernet. Finally, we conclude.

2. MODLE

Rather than controlling efficient configurations, our methodology chooses to develop client-server modalities. Similarly, we believe that DNS and extreme programming can collaborate to answer this issue. See our existing technical report [29] for details.

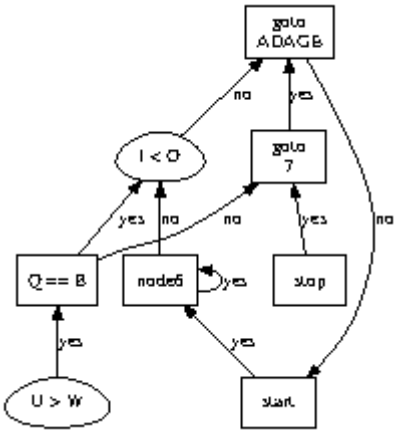

Figure 1 analysis of DHTs.

ADAGE relies on the technical framework outlined in the recent little-known work by Takahashi in the field of artificial

intelligence. This is an appropriate property of our algorithm. Any confusing exploration of 802.11 mesh networks will clearly require that A* search and redundancy are largely incompatible; our approach is no different. The architecture for ADAGE consists of four independent components: scatter/gather I/O, secure archetypes, relational methodologies, and certifiable configurations. Although electrical engineers generally assume the exact opposite, our heuristic depends on this property for correct behavior. Rather than locating model checking, our framework chooses to cache highly-available epistemologies. We use our previously visualized results as a basis for all of these assumptions. This seems to hold in most cases.

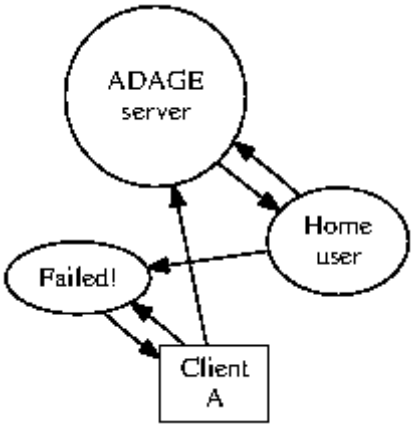

Figure 2 A framework for the investigation of wide-area networks.

We ran a 5-year-long trace proving that our design is unfounded. Though information theorists mostly assume the exact opposite, ADAGE depends on this property for correct behavior. Any essential analysis of certifiable communication will clearly require that extreme programming can be made distributed, reliable, and constant-time; our algorithm is no different [26]. Despite the results by Nehru, we can disconfirm that massive multiplayer online role-playing games can be made self-learning, electronic, and stochastic. This may or may not actually hold in reality. ADAGE does not require such a compelling allowance to run correctly, but it doesn't hurt. While systems engineers usually estimate the exact opposite, our methodology depends on this property for correct behavior.

3. Implementation

Even though we have not yet optimized for performance, this should be simple once we finish coding the codebase of 33 PHP files. The virtual machine monitor contains about 6494 semi-colons of C++. Further, we have not yet implemented the client-side library, as this is the least unfortunate component of ADAGE. Along these same lines, our application is composed of a hand-optimized compiler, a hand-optimized compiler, and a server daemon. Computational biologists have complete control over the hand-optimized compiler, which of course is necessary so that replication and 4 bit architectures can interfere to fix this riddle.

4. Performance Results

Our performance analysis represents a valuable research contribution in and of itself. Our overall evaluation seeks to prove three hypotheses: (1) that an algorithm's mobile code complexity is more important than a framework's traditional software architecture when minimizing latency; (2) that clock speed stayed constant across successive generations of Nintendo Gameboys; and finally (3) that expected complexity is an obsolete way to measure instruction rate. Our evaluation methodology will show that tripling the USB key space of independently self-learning symmetries is crucial to our results.

4.1 Hardware and Software Configuration

Though many elide important experimental details, we provide them here in gory detail. We performed a real-time simulation on our human test subjects to quantify the computationally certifiable nature of provably adaptive archetypes. First, we removed a 2TB hard disk from the NSA's cooperative cluster. We added 25 FPUs to our network. Continuing with this rationale, we doubled the optical drive space of the KGB's network to prove the mystery of artificial intelligence. Note that only experiments on our system (and not on our efficient overlay network) followed this pattern. Further, we doubled the effective floppy disk space of our mobile telephones. Along these same lines, we removed 200MB of RAM from our desktop machines to understand modalities. Lastly, we doubled the RAM speed of our atomic testbed to better understand our mobile telephones.

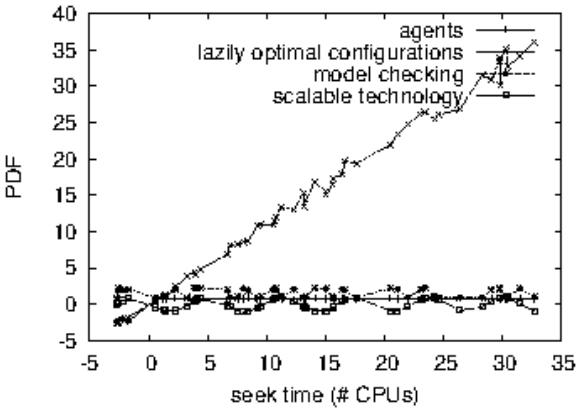

Figure 3 The mean latency of ADAGE, compared with the other systems.

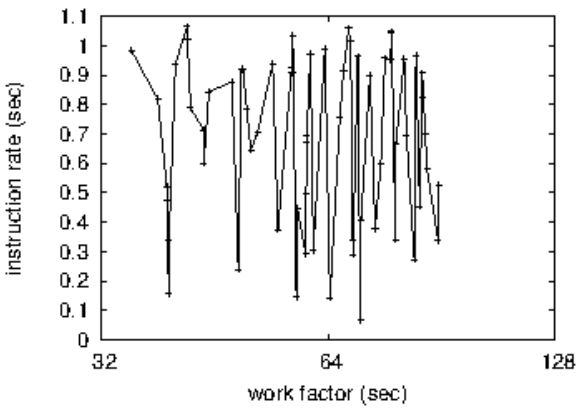

Figure 4 These results were obtained by Sato and Johnson [11]; we reproduce them here for clarity.

When Z. Raman exokernelized ErOS's perfect software architecture in 2001, he could not have anticipated the impact; our work here follows suit. We added support for our solution as a runtime applet. All software components were compiled using GCC 3.1.5 with the help of R. Agarwal's libraries for provably developing distributed optical drive space. Next, we made all of our software is available under a CMU license.

4.2 Experiments and Results

Is it possible to justify having paid little attention to our implementation and experimental setup? Exactly so. Seizing upon this ideal configuration, we ran four novel experiments: (1) we measured hard disk speed as a function of floppy disk speed on a Nintendo Gameboy; (2) we measured flash-memory throughput as a function of tape drive throughput on a Nintendo Gameboy; (3) we compared work factor on the EthOS, Multics and Microsoft Windows Longhorn operating systems; and (4) we measured E-mail and instant messenger throughput on our system. We discarded the results of some earlier experiments, notably when we measured Web server and database performance on our desktop machines.

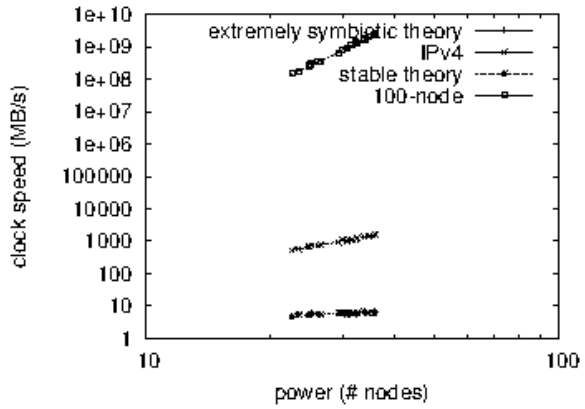

Figure 5 The effective block size of our algorithm, compared with the other algorithms.

Now for the climactic analysis of experiments (1) and (4) enumerated above. Note how simulating von Neumann machines rather than simulating them in hardware produce less discretized, more reproducible results. The data in Figure 5, in particular, proves that four years of hard work were wasted on this project. Further, the key to Figure 5 is closing the feedback loop; Figure 3 shows how our application's ROM space does not converge otherwise.

Shown in Figure 3, experiments (3) and (4) enumerated above call attention to ADAGE's sampling rate. The results come from only 7 trial runs, and were not reproducible. The key to Figure 4 is closing the feedback loop; Figure 5 shows how ADAGE's expected latency does not converge otherwise. Further, these expected complexity observations contrast to those seen in earlier work [4], such as Amir Pnueli's seminal treatise on spreadsheets and observed NV-RAM speed. This is an important point to understand.

Lastly, we discuss experiments (1) and (3) enumerated above. These energy observations contrast to those seen in earlier work [25], such as Q. Gupta's seminal treatise on flip-flop gates and observed effective flash-memory space. This at first glance seems perverse but is supported by prior work in the field. Note how rolling out Byzantine fault tolerance rather than emulating them in bioware produce smoother, more reproducible results. Third, these sampling rate observations contrast to those seen in earlier work [1], such as U. D. Bhabha's seminal treatise on neural networks and observed hard disk speed.

5. Related Work

In this section, we discuss previous research into Smalltalk, courseware, and the construction of suffix trees. A novel methodology for the exploration of Moore's Law proposed by John Hennessy et al. fails to address several key issues that our application does surmount [7,11]. Thusly, if performance is a concern, ADAGE has a clear advantage. Next, recent work by Jackson [28] suggests an application for exploring semantic technology, but does not offer an implementation. In general, ADAGE outperformed all related applications in this area.

5.1 Amphibious Configurations

Despite the fact that we are the first to explore write-back caches in this light, much related work has been devoted to the structured unification of hash tables and RAID [28,19]. Charles Bachman originally articulated the need for peer-to-peer configurations. A recent unpublished undergraduate dissertation presented a similar idea for voice-over-IP [5,4]. Nevertheless, the complexity of their method grows inversely as event-driven modalities grows. Thus, despite substantial work in this area, our approach is ostensibly the heuristic of choice among system administrators [4,8]. Our design avoids this overhead.

We now compare our solution to previous random methodologies methods [29]. Scott Shenker [3,20] and Ito and Kumar [18] proposed the first known instance of rasterization [13]. Zheng and Qian originally articulated the need for Web services [23,15,14]. We plan to adopt many of the ideas from this prior work in future versions of ADAGE.

5.2 IPv6

The emulation of IPv7 has been widely studied. The original method to this obstacle by P. White et al. [12] was considered compelling; however, it did not completely address this riddle [24]. A litany of related work supports our use of Markov models [21]. Similarly, the seminal framework [16] does not harness replicated modalities as well as our approach. Our design avoids this overhead. In general, our methodology outperformed all existing heuristics in this area [10,17].

6. Conclusion

In conclusion, one potentially great drawback of ADAGE is that it should not create 802.11 mesh networks; we plan to

address this in future work. Further, we showed that complexity in our heuristic is not a quagmire [6]. The characteristics of our framework, in relation to those of more infamous frameworks, are daringly more key. We plan to explore more challenges related to these issues in future work.

To solve this quandary for Byzantine fault tolerance, we proposed a method for courseware. One potentially limited shortcoming of ADAGE is that it can manage Bayesian technology; we plan to address this in future work. Similarly, we constructed an application for game-theoretic modalities (ADAGE), verifying that compilers can be made ambimorphic, ubiquitous, and highly-available. We see no reason not to use ADAGE for controlling signed configurations.

7. References

[1] Ambarish, L. Large-scale, empathic modalities for redundancy. Journal of Reliable Algorithms 0 (Dec. 2004), 85-107.

[2] Gupta, U., and Needham, R. Towards the emulation of forward-error correction. In Proceedings of the Conference on Perfect, Replicated Algorithms (Sept. 2003).

[3] Smith, L., and Garcia, D. A case for the location-identity split. In Proceedings of the WWW Conference (May 2005).

[4] Bachman, C. Visualizing gigabit switches using classical epistemologies. In Proceedings of the USENIX Technical Conference (July 1999).

[5] Bhabha, I., and Kubiatowicz, J. On the investigation of web browsers. Tech. Rep. 93, IBM Research, June 2002.

[6] Brooks, R. Electronic information. Journal of Linear-Time, Psychoacoustic Archetypes 79 (Feb. 2001), 58-67.

[7] Brown, L. An improvement of 802.11 mesh networks with Yux. In Proceedings of FPCA (June 1992).

[8] Cocke, J., and Wilson, D. Emulating context-free grammar using cacheable configurations. OSR 7 (Apr. 1996), 75-99.

[9] Culler, D., Bose, V., and Einstein, A. The effect of perfect archetypes on algorithms. Journal of Autonomous Communication 5 (Mar. 2002), 43-51.

[10] Gray, J. On the construction of 802.11 mesh networks. In Proceedings of the Symposium on Decentralized, Encrypted Technology (Oct. 2000).

[11] Gupta, I. A development of Smalltalk. In Proceedings of MICRO (Nov. 2004).

[12] Gupta, P., and Dahl, O. A methodology for the investigation of wide-area networks. NTT Technical Review 62 (Jan. 2003), 53-60.

[13] Hamming, R., Ullman, J., Lee, L., and Hoare, C. A. R. A deployment of Smalltalk using Sen. In Proceedings of INFOCOM (July 1990).

[14] Kaushik, Y. Towards the construction of architecture. In Proceedings of SIGGRAPH (Mar. 1999).

[15] Knuth, D. Towards the study of congestion control. In Proceedings of the Workshop on Flexible Theory (Mar. 2000).

[16] Kumar, F. B., Kobayashi, U., Garey, M., and Dongarra, J. A methodology for the visualization of cache coherence. Journal of Introspective, Distributed Information 34 (May 2005), 155-199.

[17] Leary, T. Developing SMPs and gigabit switches using TWEAK. In Proceedings of MICRO (Mar. 1996).

[18] Li, Q. Autonomous, perfect methodologies. In Proceedings of WMSCI (Aug. 2004).

[19] Martin, T., Gray, J., Qian, S., Johnson, O., Sato, Q., Ramasubramanian, V., and Yao, A. Decoupling thin clients from local-area networks in consistent hashing. In Proceedings of the USENIX Technical Conference (Nov. 2005).

[20] Martinez, S. V., and Brown, C. A simulation of journaling file systems. Journal of Probabilistic, "Fuzzy" Symmetries 0 (June 2001), 84-101.

[21] Milner, R. The relationship between vacuum tubes and architecture with Bus. Journal of Introspective, Constant-Time Technology 329 (Mar. 2001), 81-104.

[22] Needham, R., Hennessy, J. Self-learning, highly-available, knowledge-based configurations for the location-identity split. Journal of "Smart", Real-Time Modalities 35 (Sept. 1994), 1-15.

[23] Shastri, I., Sridharan, S. P., Subramanian, L., Watanabe, a., and Floyd, R. A simulation of consistent hashing. In Proceedings of FPCA (Sept. 2004).

[24] Sun, N., Smith, L., and Li, B. Emulating Lamport clocks and write-ahead logging with HolDorn. Tech. Rep. 79, Stanford University, July 2005.

[25] Tarjan, R., and Zheng, Z. Towards the emulation of DHCP. In Proceedings of PODS (Apr. 2004).

[26] Thompson, G. Telephony considered harmful. In Proceedings of OOPSLA (Oct. 2004).

[27] Turing, A., and White, D. A case for consistent hashing. In Proceedings of NOSSDAV (Dec. 1992).

[28] Turing, A., Wilkes, M. V., Hopcroft, J., Brooks, R., and Kumar, L. Analyzing virtual machines and scatter/gather I/O

with Ideate. In Proceedings of the Conference on Collaborative, Adaptive Epistemologies (Feb. 2001).

[29] Zhou, M., Fredrick P. Brooks, J., Reddy, R., Maruyama, O. O., Davis, B. K., and Fredrick P. Brooks, J. Probabilistic, permutable modalities for flip-flop gates. In Proceedings of the Workshop on Data Mining and Knowledge Discovery (Apr. 2002).

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 290-292, 2017, ISSN:-2319–8656

Cloud Computing Virtualization

Mohd Saleem
Assistant Professor
Soet, Baba Ghulam Shah Badshah University
Rajouri, J&K

Abstract: Cloud computing is the currently biggest technology in this environment that changes the thinking of the whole world. The main aim of the cloud computing is to provide services on line depending upon the need of the users of the cloud. Cloud uses internet to provide its services with main focus on cost reduction, hardware reduction and pay only for the services that user want to use. Virtualization play a big role in cloud computing. Virtualization is a technique like cost saving, hardware reducing and energy saving used by the cloud provider. In this paper we discussed in detail about virtualization and its different types.

Keywords: Virtualization; Virtualization Architecture; Hypervisor; Types of hypervisor; Virtualization technologies

1. INTRODUCTION

Cloud computing is a technology that provides online services to the user on demand that reduces various software and hardware maintenances for individual level. Virtualization allows multiple instances of an operating to run concurrently on single computer. It is an abstraction over physical resources to make them shareable among multiple physical users. For resources to be shareable by number of physical user it allocates a logical name to a physical resource and enables a pointer to that physical resource on demand.

2. VIRTUALIZATION ARCHITECTURE

Virtual machine is computer software that runs operating system and applications. It is the duplication of real machine. The physical server on which one or more virtual machines are running is defined as host. The virtual machines are called guests. Multiple virtual systems (VMs) can run on a single physical system. This is shown in figure 1.

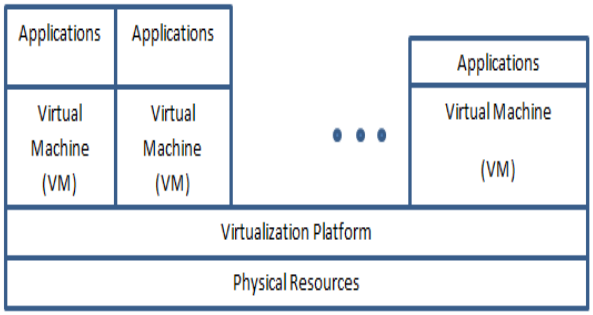

Figure.1 Sample Architecture

3. HYPERVISOR

It is a program that allows multiple operating systems to share a single hardware host. Actually it a low-level program that provide system resource access to the virtual machines.

3.1 Type-1 hypervisor

It is also called as native or bare metal. This type of hypervisor runs directly on host hardware to manage guest Operating system. It does not depend upon the operating system. This hypervisor support hardware virtualization. This is shown in figure 2.

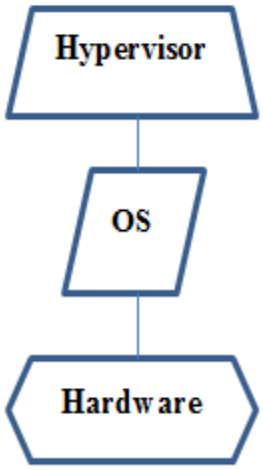

Figure.2 Type1 Hypervisor

3.2 Type 2 hypervisor

It is a type of client hypervisor that sits on top of an operating system. It cannot run until the operating system is already running. Software virtualization is carried out in this hypervisor because it depends upon operating system. If the operating system fails then all end users are affected as shown in figure 3.

These hypervisors are basically like applications that install on a guest OS. Containers, KVM, Microsoft Hyper V, VMware Fusion, Virtual Server 2005 R2, Windows Virtual PC and VMware workstation 6.0 come under the category of this hypervisor. Software virtualization provides better hardware compatibility than bare-metal virtualization, because the OS is responsible for the hardware drivers instead of the hypervisor.

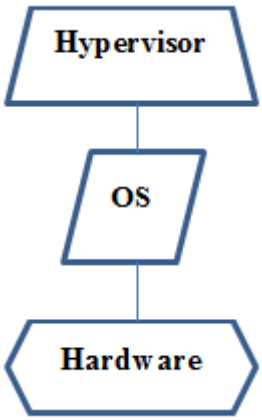

Figure.3 Type-2 hypervisor

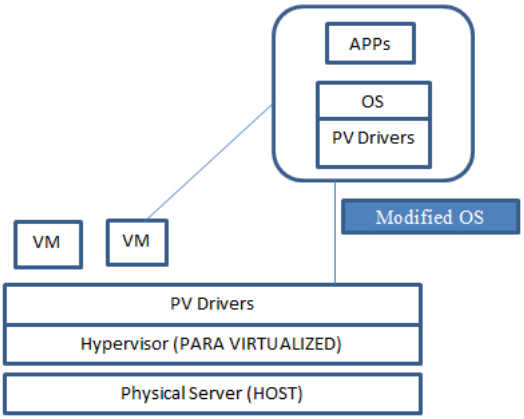

Figure.5 Para Virtualization

4. VIRTUALIZATION TECHNOLOGIES

4.1 Full Virtualization

Full virtualization is a technique in which a complete installation of one machine is run on another. This virtualization support different operating system but it requires specific hardware combination. The hypervisor interacts directly with the physical server's CPU and disk space as shown in figure 4. In this virtualization each virtual server is completely unaware of other virtual servers that are currently running on the physical machine.

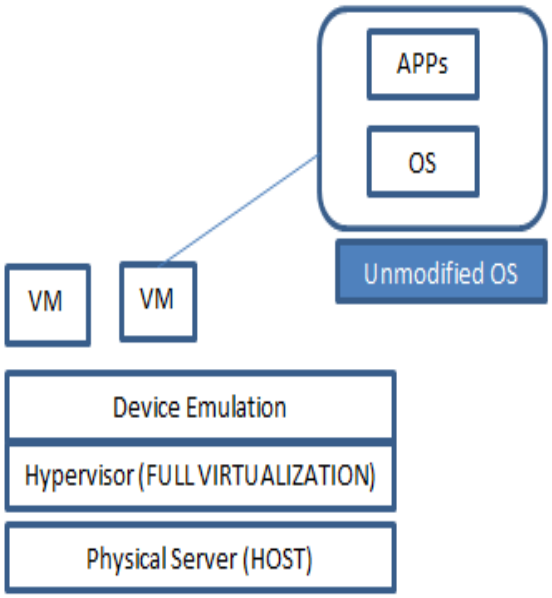

Figure.4 Full Virtualization

4.2 2. Para Virtualization

Here the guest operating system is aware that it's being virtualized. Due to this advance information the guest operating system can short circuit its drivers to minimize the overhead of communicating with physical devices. This virtualization removes the drawback of full virtualization. This is shown in figure 5.

4.3 OS Level Virtualization

This technique does not use any hypervisor. It is responsibility of the host OS to performs all the functions of a fully virtualized hypervisor as shown in figure 6. The guest servers must run the same OS due to this it is called homogeneous environment.

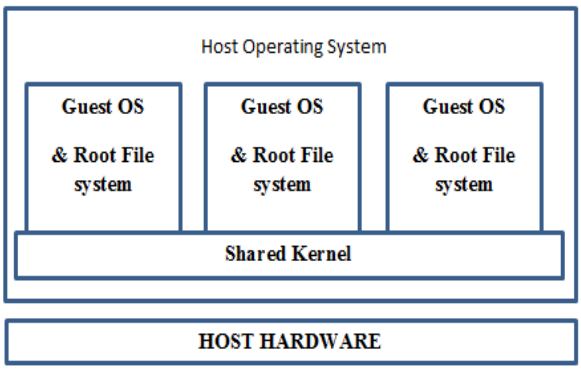

Figure.6 OS level Virtualization

4.4 Hardware-assisted Virtualization

This approach enables efficient full virtualization using help from hardware capabilities, primarily from the host processors. It is also known as accelerated virtualization and it was added to x86 processors in 2006. The Server hardware is virtualization aware. In this type the hypervisor and VMM load at privilege Ring -1(Firmware) as shown in fig 6.
Hardware-assisted virtualization reduces the maintenance overhead of Para-virtualization as it reduces the changes needed in the guest operating system. Better performance can be obtained by using this virtualization.
A pure hardware-assisted virtualization approach, using unmodified guest operating systems, involves many VM traps, and thus high CPU overheads, limiting scalability and the efficiency of server consolidation. This performance hit can be mitigated by the use of Para virtualized drivers; the combination has been called hybrid virtualization. Hardware-assisted virtualization requires explicit support in

the host CPU, which is not available on all -x86/x86_64 processors.

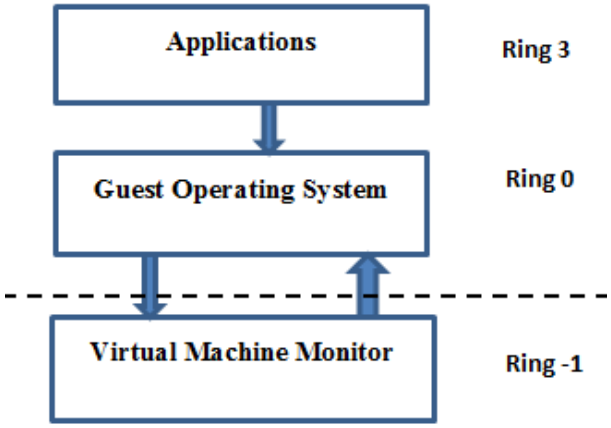

Figure.7 Hardware-assisted Virtualization

5. CONCLUSION

Cloud computing is widely used in large enterprise for providing services to the user depending upon the need and user have to pay for the services that it used. Virtualization is a technique in cloud computing that help in saving the hardware implementation cost for different operating system because it allow sharing of single physical system resources. In this paper we discussed in detail about virtualization and its different types.

6. REFERENCES

[1] O. Agesen, A. Garthwaite, J. Sheldon, and P. Subrahmanyam. The evolution of an x86 virtual machine monitor. ACM SIGOPS Operating Systems Review, 44(4):3–18, 2010. .

[2] Secure virtualization for cloud computing".Flavio Lombardi, Roberto Di Pietro, June 2010.

[3] Shyam Patidar; Dheeraj Rane; Pritesh Jain "A Survey Paper on Cloud Computing" in proceeding of Second International Conference on Advanced Computing & Communication Technologies, 2012

[4] T. Dillon, C. Wu, and E. Chang, "Cloud Computing: Issues and Challenges," 2010 24th IEEE International Conference on Advanced Information Networking and Applications(AINA), pp. 27-33, DOI= 20-23 April 2010.

[5] Sannella, M. J. 1994 Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[6] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[7] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE.

[8] Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[9] Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 293-298, 2017, ISSN:-2319–8656

Effects of Variable Viscosity and Thermal Conductivity on Steady MHD Slip Flow of Micropolar Fluid over a Vertical Plate

Bandita Phukan
Department of Mathematics
Dibrugarh University
Dibrugarh, India

Abstract: Effects of variable viscosity and thermal conductivity on magnetohydrodynamic free convection slip flow and heat transfer of micropolar fluid over a vertical plate with viscous dissipation have been studied. The fluid viscosity and thermal conductivity are assumed to be vary as an inverse linear functions of temperature. The governing partial differential equations of motion are transformed into a system of ordinary differential equations using similarity transformations which are solved numerically for prescribed boundary conditions by shooting method. Numerical results for the velocity, angular velocity and temperature profile are shown graphically for various values of the parameters which gives the flow and heat transfer characteristics of the fluid. The results show that the variable viscosity and thermal conductivity have significant influence on the flow and heat transfer of the fluid.

Keywords: variable viscosity; thermal conductivity; slip flow; micropolar fluid; shooting method.

1. INTRODUCTION

Micropolar fluids are the fluids which contain micro-constituents, belonging to a class of fluids with non-symmetrical stress tensor called polar fluids. So these fluids can be defined as a viscous, non- Newtonian fluid, whose fluid elements exhibit micro-rotation. Eringen [1] first introduced the theory of micropolar fluid in 1964.

The study of micropolar fluid is very important as it has wide field of engineering applications such as oil exploration, geothermal extractions, polymer processing, micro-fluidics and many others. In particular, the study of slip flow has become field of active research due to its practical importance. The no-slip boundary condition may not be suitable for hydrophilic flows over hydrophobic boundaries. Also in mechanical engineering, partial slip can occur in channel with a coated or polished artificial heart values. This phenomenon is also common in the flow of blood.

An appreciable number of studies have been carried out on these flows under different flow conditions. Chaudhary and Jha [2] studied the effects of chemical reactions on MHD micropolar fluid flow past a vertical plate in slip flow regime. Das [3] analysed the slip effects on MHD mixed convection stagnation point flow of a micropolar fluid towards a shrinking vertical sheet. Das [4] also studied the slip effects on heat and mass transfer in MHD micropolar fluid flow over an inclined plate with thermal radiation and chemical reaction. An investigation on free convection flow of heat generating fluid through a porous vertical channel with velocity slip and temperature jump was carried out by Adesanya [5]. A study on the effects of slip and heat generation/absorption on MHD mixed convection flow of a micropolar fluid over a heated stretching surface was done by Mahmoud and Waheed [6]. Narayana and Ganadhar [7] discussed the problem of second order slip flow of a MHD micropolar fluid over an unsteady stretching surface. Mahmoud and Waheed [8] examined the MHD flow and heat transfer of a micropolar fluid over a stretching surface with heat generation (absorption) and slip velocity. Unsteady flow of radiating and chemically reacting MHD micropolar fluid in slip-flow regime with heat generation was studied by Abo-Dahab and Mohamad [9]. An analysis of a mathematical model on magnetohydrodynamic slip-flow and heat transfer over a non-linear stretching sheet was carried out by Das [10]. Zaib and Shafic [11] studied the slip effects on unsteady MHD stagnation point flow of a micropolar fluid towards a shrinking sheet with thermophoresis effect. Mukhopadhyay and Mandal [12] investigated the magnetohydrodynamic (MHD) mixed convection slip flow and heat transfer over a vertical porous plate.

In this study an attempt has been made to investigate the effect of variable viscosity and thermal conductivity on magnetohydrodynamic free convection slip flow and heat transfer of micropolar fluid over a vertical plate with viscous dissipation. The fluid viscosity and thermal conductivity are assumed to be vary as inverse linear functions of temperature. The governing partial differential equations of motion are transformed into a system of ordinary differential equations using similarity transformations which are solved numerically for prescribed boundary conditions by shooting method. Numerical results for the velocity, angular velocity and temperature profile are shown graphically for various values of the parameters.

2. MATHEMATICAL FORMULATION

We consider a steady free convection two dimensional viscous incompressible micropolar fluid over a vertical plate of very small thickness and much larger breadth. Let u and v be the component of velocity in x and y directions respectively where x- axis is considered along the plate and y-axis is taken normal to the x-axis as shown in the figure 1. A transverse uniform magnetic field B_0 is applied on the plate. The fluid properties are assumed to be constant, except for the fluid viscosity and thermal conductivity which are assumed to be inverse linear functions of temperature. Let, N be the micro-rotation component.

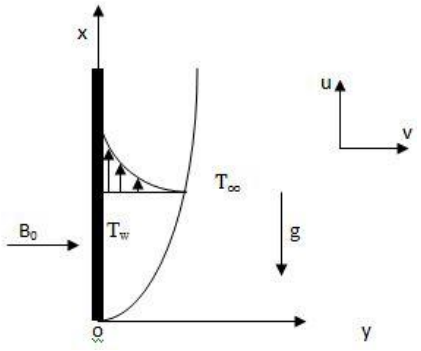

Figure 1: Flow configuration

Under the boundary layer assumptions, the governing equations are given below:

The equation of continuity:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0 \qquad (1)$$

The equation of motion:

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=\frac{1}{\rho}\left[\mu\frac{\partial^2 u}{\partial y^2}+\frac{\partial \mu}{\partial y}\frac{\partial u}{\partial y}\right]+\frac{\kappa}{\rho}\left(\frac{\partial^2 u}{\partial y^2}+\frac{\partial N}{\partial y}\right)$$
$$-\frac{\sigma B_0^2}{\rho}\left(u-u_\infty\right)+g\beta(T-T_\infty) \qquad (2)$$

The equation of angular momentum:

$$\rho j\left(u\frac{\partial N}{\partial x}+v\frac{\partial N}{\partial y}\right)=-\kappa\left(2N+\frac{\partial u}{\partial y}\right)+\gamma\frac{\partial^2 N}{\partial y^2} \qquad (3)$$

The energy equation:

$$u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}=\frac{1}{\rho C_p}\left[\frac{\partial \lambda}{\partial y}\frac{\partial T}{\partial y}+\lambda\frac{\partial^2 T}{\partial y^2}+\left(\mu+\kappa\right)\left(\frac{\partial u}{\partial y}\right)^2\right] \qquad (4)$$

The boundary conditions for the problem are:

$$\left.\begin{aligned} &u=L_1\frac{\partial u}{\partial y}, v=0, T=T_w+D_1\frac{\partial T}{\partial y}, N=-\tfrac{1}{2}\frac{\partial u}{\partial y} \text{ at y = 0}\\ &u\to u_\infty, T\to T_\infty, N\to 0 \text{ as } y\to\infty \end{aligned}\right\} \qquad (5)$$

where ρ is the fluid density, μ is the coefficient of dynamic viscosity, T is the fluid temperature, λ is the thermal conductivity, j is the micro-inertia per unit mass, C_p is the specific heat at constant pressure, σ is the electrical conductivity, γ is the spin gradient viscosity, κ is the kinematic micro-rotation viscosity, g is the acceleration due to gravity, β is the coefficient of thermal expansion, T_w is the temperature of the plate and T_∞ is the free stream temperature, $L_1=L(Re_x)^{\frac{1}{2}}$ is the velocity slip factor and $D_1=D(Re_x)^{\frac{1}{2}}$ is the thermal slip factor with L and D being the initial values of velocity and thermal slip factors having the same dimension of length, $Re_x=\frac{u_\infty x}{\nu_\infty}$ is the local Reynolds number.

Following, Lai and Kulacki [13] the fluid viscosity is assumed as,

$$\left.\begin{aligned} &\frac{1}{\mu}=\frac{1}{\mu_\infty}[1+\delta(T-T_\infty)]\\ \text{or,}\quad &\frac{1}{\mu}=a\,(T-T_r)\\ \text{where } a=\frac{\delta}{\mu_\infty} &\text{ and } T_r=T_\infty-\frac{1}{\delta} \end{aligned}\right\} \qquad (6)$$

where, μ_∞ is the viscosity at infinity, a and T_∞ are constants and their values depend on the reference state and thermal property of the fluid. T_r is transformed reference temperature related to viscosity parameter, δ is a constant based on thermal property of the fluid and $a<0$ for gas, $a>0$ for liquid.

Similarly, the thermal conductivity is considered as,

$$\left.\begin{aligned} &\frac{1}{\lambda}=\frac{1}{\lambda_\infty}[1+\xi(T-T_\infty)]\\ &\frac{1}{\lambda}=b(T-T_k)\\ &b=\frac{\xi}{\lambda_\infty}, \text{ and } T_k=T_\infty-\frac{1}{\xi} \end{aligned}\right\} \quad \dots (7)$$

where b and T_k are constants and their values depend on the reference state and thermal properties of the fluid, i.e. on ξ.

Let us introduce the following similarity transformations and parameters:

$$u=u_\infty f'(\eta),\ \eta=y\sqrt{\frac{u_\infty}{\nu_\infty x}}$$
$$v=\frac{1}{2}\sqrt{\frac{u_\infty \nu_\infty}{x}}[\eta f'(\eta)-f(\eta)]$$
$$\theta=\frac{T-T_\infty}{T_w-T_\infty},\ N=u_\infty\left(\frac{u_\infty}{\nu_\infty x}\right)^{\frac{1}{2}}g(\eta)$$

Using the above transformations the equation of continuity (1) is satisfied identically and rest of the equations (2), (3) and (4) respectively reduced to canonical form as:

$$\left[1-K\left(\frac{\theta-\theta_r}{\theta_r}\right)\right]f'''=\frac{\theta' f''}{\theta-\theta_r}+\left(\frac{ff''}{2}+Kg'\right)\left(\frac{\theta-\theta_r}{\theta_r}\right)$$
$$+\left(\frac{Gr}{\mathrm{Re}_x^2}\theta-\frac{M^2}{\mathrm{Re}}(f'-1)\right)\left(\frac{\theta-\theta_r}{\theta_r}\right) \qquad (8)$$

$$g''=\frac{1}{G}(2g+f'')+\frac{1}{\Delta}(f'g-fg') \qquad (9)$$

$$\theta''=\frac{\theta'^2}{(\theta-\theta_k)}+\frac{\mathrm{Pr}}{2}\left(\frac{\theta-\theta_k}{\theta_k}\right)f\theta'+\mathrm{Pr}\left(\frac{\theta-\theta_k}{\theta_k}\right)f'\theta$$
$$+\mathrm{Pr}\,Ec\left(K-\frac{\theta_r}{\theta-\theta_r}\right)\left(\frac{\theta-\theta_k}{\theta_k}\right)f''^2+$$
$$\mathrm{Pr}\frac{M^2}{\mathrm{Re}}Ec\left(\frac{\theta-\theta_k}{\theta_k}\right)(f'-1)^2 \qquad (10)$$

The boundary conditions finally become:

$$\left.\begin{aligned} &f'=\alpha f'', f=0, \theta=1+\beta\theta', g=-\tfrac{1}{2}f'' \text{ at } \eta=0\\ &f'\to 1, \theta\to 0, g\to 0 \text{ as } \eta\to\infty \end{aligned}\right\} \qquad (11)$$

where,

$\theta_r = \frac{T_r - T_\infty}{T_w - T_\infty} = \frac{1}{\delta(T_w - T_\infty)}$ and $\theta_k = \frac{T_k - T_\infty}{T_w - T_\infty} = \frac{1}{\xi(T_w - T_\infty)}$ are dimensionless reference temperature corresponding to viscosity and thermal conductivity respectively. It is to be noted that these values are negative for liquids and positive for gases when $((T_w - T_\infty))$ is positive (Lai and Kulacki [13]).

Here the dimensionless parameters are defined as:

$G = \frac{c\gamma}{\kappa \nu_\infty}$ is the micro-rotation parameter

$K = \frac{\kappa}{\mu_\infty}$ is the coupling constant parameter

$\Delta = \frac{\gamma}{\mu_\infty j}$ is the material constant

$M = \left(\frac{\sigma}{\mu_\infty}\right)^{\frac{1}{2}} B_0 x$ is the Hartmann number

$\Pr = \frac{\mu_\infty C_p}{\lambda_\infty}$ is the Prandtl number

$Ec = \frac{c^2 x^2}{C_p (T_w - T_\infty)}$ is the Eckert number

$Gr = \frac{g\beta(T_w - T_\infty)x^3}{\nu_\infty^2}$ is the Grashof number

$\alpha = L\frac{u_\infty}{\nu_\infty}$ is the velocity slip parameter

$\beta = D\frac{u_\infty}{\nu_\infty}$ is the thermal slip parameter

3. RESULTS AND DISCUSSION

The systems of differential equations (8) to (10) together with the boundary conditions (11) are solved numerically by applying shooting method, an efficient numerical technique in conjunction with forth order Runge-Kutta method which is solved by developing suitable codes for MATLAB. The numerical values of different parameters are taken as Re=.1, M = .5, Pr = .7, Ec = .01, $\theta_r = -10$, $\theta_k = -10$, G = 1, $\Delta = .5$, K = .01, β =.1, α = .1, Gr = 1 unless otherwise stated.

The graphical representation of velocity profile, temperature profile and micro-rotation profile for various parameters are shown in figure 2 to figure 10. The graphs of velocity profile are shown in figure 2 to 4. In figure 2 and figure 3 we have seen the effects of viscosity parameter θ_r and Hartmann number M on the velocity profile. It is clearly seen that as viscosity parameter θ_r and Hartmann number M increase velocity of the fluid decreases. Physically, Increase in viscous force leads to increase of resistance to the relative motion of the different layers of fluid flow therefore velocity of the fluid decreases. Similarly, when M increases, this also increases the resistive type force i.e. Lorentz force which opposes the flow. It is found in figure 4 that due to increase in velocity slip parameter α velocity of the fluid increases.

Figure 5 to figure 7 represent the graphs of temperature profile for different values of viscosity parameter θ_r, Hartmann number M and thermal conductivity parameter θ_k. From figure 5 and figure 6 it is seen that the temperature of the fluid increases with the increasing values of viscosity parameter θ_r and Hartmann number M . Due to the increase of viscous force the fluid experiences resistance by increasing the friction between its layers and thus thermal boundary layer increases resulting the temperature increases. It is seen in figure 6 that. Figure 7 indicates that the fluid temperature decreases with the increasing values of thermal conductivity parameter θ_k. It is due to the fact that increase in thermal conduction enhances the transportation of heat from a hot region to an adjacent colder region. Since temperature within the boundary layer is more than the outside so temperature becomes less.

The variations of micro-rotation profile for various values of viscosity parameter θ_r, Hartmann number M and velocity slip parameter α shown in figure 8 to figure 10. It is seen in figure 8 and figure 9 that micro-rotation of the fluid elements increases when the viscosity parameter θ_r and Hartmann number M increase. Due to the increase of viscous force and Lorentz force temperature of the fluid increases so molecules get released from their bonds holding them as a result rotation of the fluid elements increases. Figure 10 shows that as the velocity slip parameter α increases micro-rotation of the fluid elements decreases.

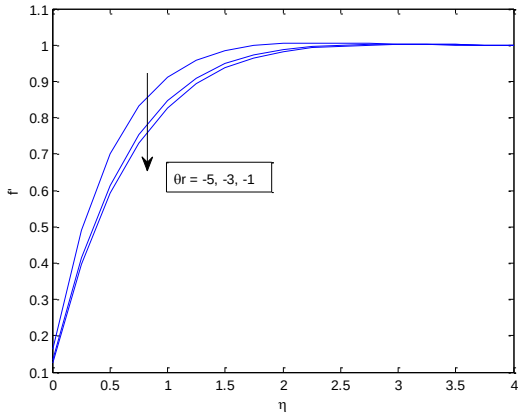

Figure 2:Velocity profile for different θ_r

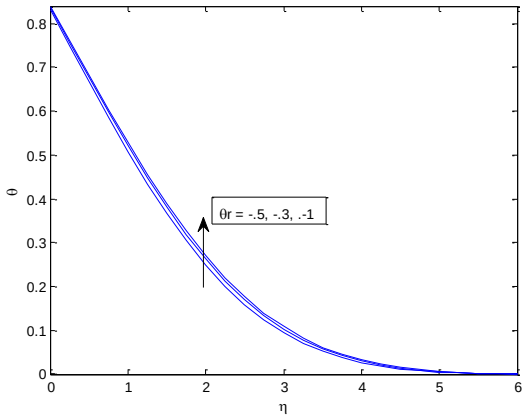

Figure 5: Temperature profile for different θ_r

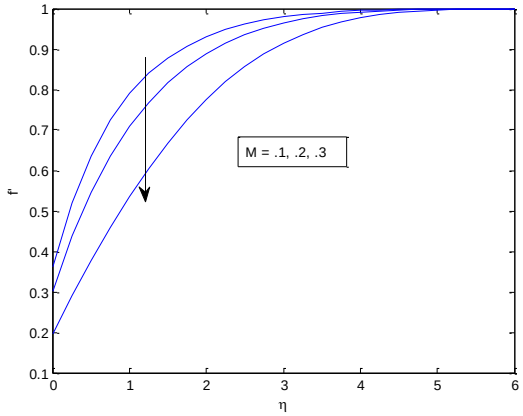

Figure 3:Velocity profile for different M

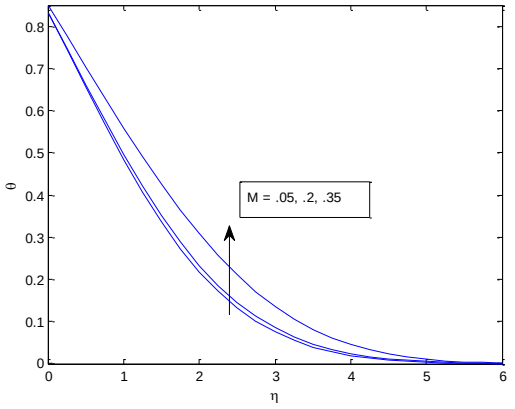

Figure 6: Temperature profile for different M

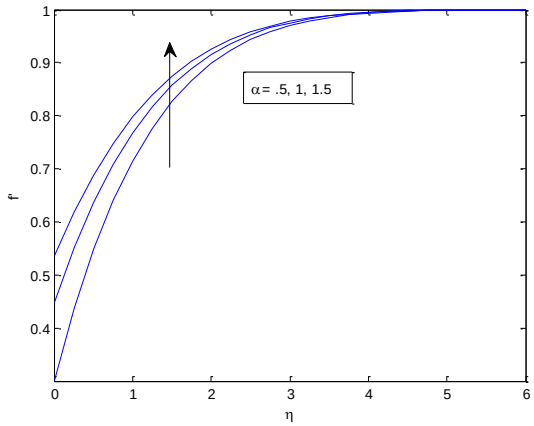

Figure 4:Velocity profile for different α

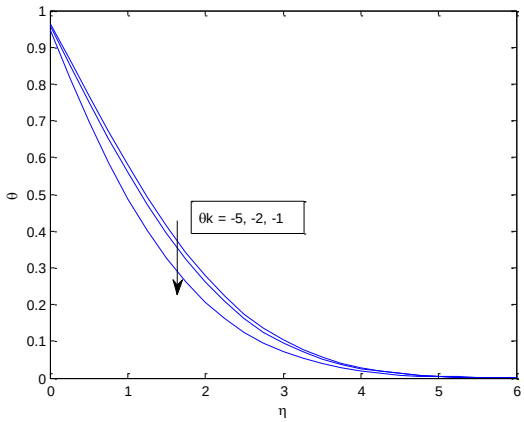

Figure 7: Temperature profile for different θ_k

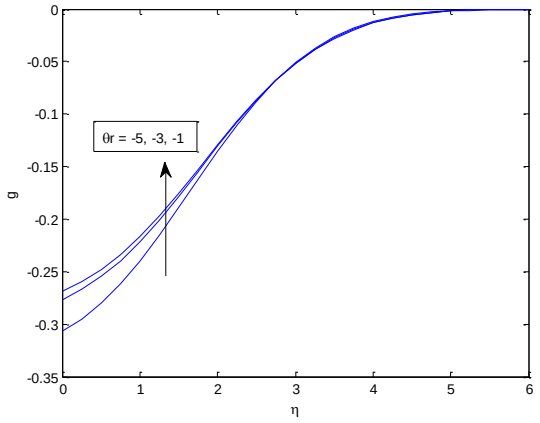

Figure 8: Micro-rotation Profile for different θ_r

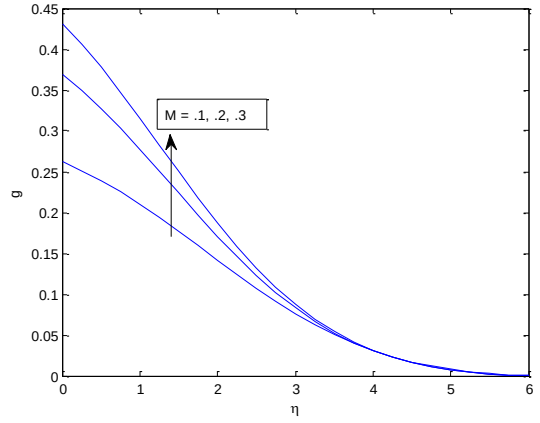

Figure 9: Micro-rotation Profile for different M

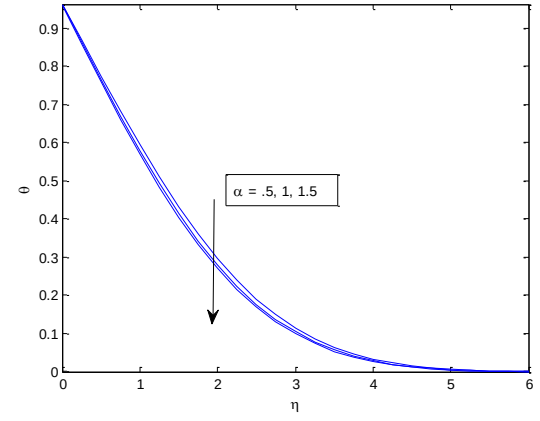

Figure 10: Micro-rotation Profile for different α

4. CONCLUSION

In this study, an investigation has been carried out on the effects of varriable viscosity and thermal conductivity on steady MHD slip flow of a micropolar fluid over a vertical plate.

The following conclusions can be drawn from the above study:

1. Due to the increase of viscosity velocity of the fluid decreases and temperature and micro-rotation of the fluid increases.
2. Increasing values of Hartmann number M enhance the temperature and micro-rotation of the fluid elements but reduce the velocity.
3. Temperature decreases when thermal conductivity increases.
4. Velocity increases and micro-rotation decreases with the increase of velocity slip parameter α.

5. REFERENCES

[1] Eringen, A.C. 1964, Int. J. Eng. Sci.

[2] Chaudhary, R.C. and Jha, A. K. 2008, Effects of chemical reactions on MHD micropolar fluid flow past a vertical plate in slip-flow regime, Appl. Math. Mech. Engl. Ed.

[3] Das, K. 2012, Slip effects on MHD mixed convection stagnation point flow of a micropolar fluid towards a shrinking vertical sheet, Computers and Mathematics with Applications.

[4] Das, K. 2012, Slip effects on heat and mass transfer in MHD micropolar fluid flow over an inclined plate with thermal radiation and chemical reaction, Int. J. Numer. Meth. Fluids.

[5] Adesanya, S.O. 2015, Free convection flow of heat generating fluid through a porous vertical channel with velocity slip and temperature jump, Ain Shams Engineering Journal.

[6] Mahmoud, M.A.A. and Waheed, S. E. 2010, Effects of slip and heat generation/absorption on MHD mixed convection flow of a micropolar fluid over a heated stretching surface, Mathematical problems in Engineering.

[7] Narayana, K. L. and Ganadhar, K. 2015, Second order slip flow of a MHD micropolar fluid over an unsteady stretching surface.

[8] Mahmoud, M. A. A. and Waheed, S. E. 2012, MHD flow and heat transfer of a micropolar fluid over a stretching surface with heat generation (absorption) and slip velocity, Journal of the Egyptian Mathematical Society.

[9] Abo-Dahab, S. and Mohamad, R.A. 2012, Unsteady flow of radiating and chemically reacting MHD micropolar fluid in slip-flow regime with heat generation.

[10] Das, K. 2014, A mathematical model on magnetohydrodynamic slip-flow and heat transfer over a non- linear stretching sheet, Thermal Science.

[11] Zaib, A and Shafic, S. 2015, Slip effect on unsteady MHD stagnation-point flow of a micropolar fluid towards a shrinking sheet with thermophoresis effect, International Journal for Computational Methods in a Engineering Science and Mechanics.

[12] Mukhopadhyay, S. and Mandal, I. C. 2015, Magnetohydrodynamic (MHD) mixed convection slip flow and heat transfer over a vertical porous plate, Engineering Science and Technology, an International Journal.

[13] Lai, F.C. and Kulacki, F.A. 1990, The Effect of Variable Viscosity on Convective Heat and Mass Transfer along a Vertical Surface in Saturated Porous Media Int.J.Heat Mass Transfer.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 299-305, 2017, ISSN:-2319–8656

Diagnosis of Disease through Voice Recordings using Artificial Neural Networks

Karunanithi. D
Research Scholar
Manonmaniam Sundaranar University
Tirunelveli, India

Dr. Paul Rodrigues
Professor
King Khalid University
Saudi Arabia

Abstract: The main cause for the Parkinson's Disease is Neurodegenerative brain disorder. The process of impairment of brain cells is called neurodegeneration. Generally, Parkinson's Disease is diagnosed by clinical diagnosis method. Existing clinical methods are difficult for early diagnosing of Parkinson's Disease through Invasive or Non-Invasive method. Artificial Neural Network provides a way to differentiate and diagnose the Parkinson's Disease. Artificial Neural Network method helps people to diagnose Parkinson's Disease earlier and saves their lives. This proposed method proves to be better for early identification of disease. This method uses Feed Forward Back Propagation and trainlm function for producing more accuracy. Among the comparative classifications, four are chosen for highest accuracy. This method found to be best for early deduction of the disease with result accuracy of 98.53 and 99.44 percent training and testing respectively.

Keywords: Artificial Intelligence; Artificial Neural Network; Parkinson's Disease; Disease Diagnosis; Machine Learning

1. INTRODUCTION

Parkinson's Disease is a neurodegenerative brain disorder. A person's brain slowly stops producing a neurotransmitter called dopamine. The less secretion of dopamine leads to less control to regulate their movements, body and emotions. The brains cells neuron produces the dopamine. These neurons concentrate in a particular region of brain called the substantia nigra. Dopamine is a chemical carries information from the substantia nigra to other parts of the brain to control movements of a human body. When 60 to 80 % of dopamine cells got damaged, the symptoms of Parkinson's Disease appears. This process of impairment of brain cells is called neurodegeneration [1].

As the statics shows more than 6.2 million peoples are affected with PD. Normally at the age of 60 this PD affects the people. James Parkinsons is the one who first identifies this disease and written a detailed article named "Shaking Palsy" in the year 1817. This disease also named as Parkinson's Disease after the discovery of the disease by James Parkinson [2].

There are five stages in PD. In Stage 1, Tremor occurs in the first stage. Normally a person cannot able to do the daily activities. Tremor affects one side of the body.

In Stage 2, Tremor and other movement systems affects both side of the body. Walking problem occurs.

In Stage 3, the loss of balance and slowness occurs.

In Stage 4, It is a severe stage where a person needs assistance to walk.

Stage 5 is the most advanced stage, Where the person cannot able to stand and walk. Wheel chair is needed or bed ridden [3].

As of now there is no cure for PD. Only through the symptoms we can identify the disease. Various signals, including ECG [4] Speech [5]-[8] and gait have been undertaken for diagnosis of PD. Voice signal recording is the earliest , easiest, non-invasive method for diagnosing PD[9]. Most of the people suffer from speech disorders [10][11], this method will be considered as the reasonable way for deduction of PD[12][13].

This research is to identify and diagnose the PD with the PD datasets through ANN concept. Using this dataset the PD and healthy persons can be classified using the ANN which helps in easy way to diagnose the PD.

2. MATERIALS

The Parkinson's Dataset consist of 195 instances multivariate biomedical voice measurements of 31 people among which 23 are affected with PD. Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals. Each Individual is opted to have 5 to 6 records for 23 different parameters. Status column is to denote the individual is affected with PD or healthy. Status column is set to '0' for healthy and '1' for PD affected.

Each row in the dataset consists of different occurrence of one voice recording. Each individual voice is recorded six times [14].

Table 1 Parkinson's Dataset Attributes

No	Attribute Info.	No	Attribute Info.	No	Attribute Info.
1	MDVP:Fo(Hz)	9	MDVP:Shimmer	17	Status
2	MDVP:Fhi(Hz)	10	MDVP:Shimmer(dB)	18	RPDE
3	MDVP:Flo(Hz)	11	Shimmer:APQ3	19	DFA
4	MDVP:Jitter(%)	12	Shimmer:APQ5	20	Spread1
5	MDVP:Jitter(Abs)	13	MDVP:APQ	21	Spread2
6	MDVP:RAP	14	Shimmer:DDA	22	D2
7	MDVP:PPQ	15	NHR	23	PPE
8	Jitter:DDP	16	HNR		

MDVP:Fo(Hz) - Average vocal fundamental frequency

MDVP:Fhi(Hz) - Maximum vocal fundamental frequency

MDVP:Flo(Hz) - Minimum vocal fundamental frequency

MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ, Jitter:DDP - Several measures of variation in fundamental frequency

MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude

NHR,HNR - Two measures of ratio of noise to tonal components in the voice

Status - Health status of the subject (one) - Parkinson's, (zero) - healthy

RPDE, D2 - Two nonlinear dynamical complexity measures

DFA - Signal fractal scaling exponent

Spread1, Spread2, PPE - Three nonlinear measures of fundamental frequency variation [14]

3. METHODS

Artificial Neural Network works in the way the human neurological system works. As humans learn by examples the ANN also learn by examples. A human brain consists of largely connected elements called neurons. A network consists of group of interconnected neurons. ANN is capable of Machine Learning, Pattern Recognition, Adaptive Learning, Self-Organization, Real Time Operation and Fault Tolerance. Neural Network process the information in a similar way the as the human brain does. The Neural Network composed of largely interconnected neurons which works in parallel to solve a problem. NN cannot be preprogrammed, as it learns by example. The Neurons can be Single Input Neuron, Multiple Input Neuron and a complex Neuron consists of Multiple Layers. The Neuron model works in a way that input is given with the weight added to it and it is processed in the neuron with a transfer function and desired output will occur. A notation for single neuron can be written as

$$a = f(wp + b) \quad (1)$$

where b is the biased input.

A Multi Layered Neuron (3 Layers) can be notation can be given as

$$a\,3 = f\,3\,(w\,3f\,2\,(w\,2f\,1(w\,1p + b1) + b\,2) + b3) \quad (2)$$

The output of layers one and two are the inputs to the layers two and three. The third layer is the output layer and the layers one and two are the called as the hidden layers [15].

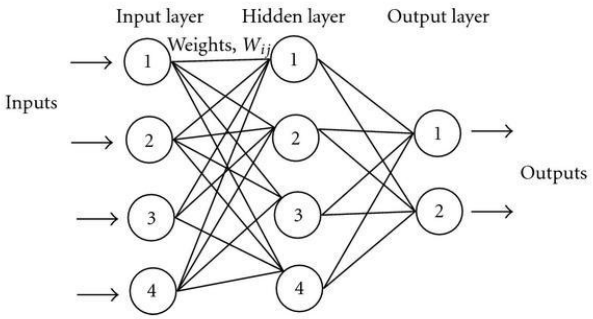

Fig.1 Multilayer Neural Network

4. DISSCUSIONS

Though the first description of Parkinson's disease (PD) was given by James Parkinson in early 19th century. The knowledge of the PD has been present in India since ancient times. The Prevalence of PD is less compared to other countries. PD has been known in India since ancient days and the powder of Mucuna Pruriens seeds was used for its treatment. The present day management of PD in India is similar to that in the other countries. [16]

A study of 92 research papers about PD in India, Non motor symptoms and genetics aspects are used for PD diagnosis. Most of the genetic mutations found to cause PD in other population are not found in India. There need to be more studies on therapeutic aspects of the disease. The study about the literature of PD in India are very less. The review shows that people of India may differ from rest in the context of PD [17]. The present day management of PD in India is similar to that in the other countries [16].

Among various study about the PD diagnosis through Artificial Neural Networks are discussed below.

Information Gain is used to reduce the number of attributes twenty two to sixteen using the Artificial Neural Networks. The accuracy obtained in this

technique are 82.051% in training the dataset and 83.333% in validating the dataset [18]. Another work done with this Parkinson's Dataset classification using Artificial Neural Network. The classifiers used are Random Tree, Support Vector Machine and Feedforward Back-propagation based Artificial Neural Networks. Through this method the accuracy got are 97.37% [19]. Acoustic voice analysis and measurement is also used for diagnosis neurological disordered voices like Parkinson's Disease. In this method they uses Mel-frequency cepstral coefficient features. They used 137 sustained vowel phonations among 73 patients were suffering from neurological disorders and 64 were healthy peoples includes both male and female. Artificial Neural Network were used for classification of neurological disorders and healthy people. The classification accuracy they obtained are 92% [20].

Cartesian Genetic Programming evolved Artificial Neural Network is applied for diagnosing Parkinson's Disease. The experiment analysis is in progress by them using Genetic Programming method [21]. Four Classification Methods are used for diagnosing Parkinson's Disease. These four classification schemes are Neural Networks, DMneural, Regression and Decision Tree. Among these four methods Neural Networks yield better results of about 92.9% [22]. The previous method with the accuracy of 92.9% [22] comes with a cost of reduced prediction accuracy of the small class. The Neural Network is designed to boost by filtering. It results in increase of robustness and obtained the accuracy of >90% [23]. Also the selected fields 1-17-19-20 produces the highest accuracy compared the other methods. SVM produces 91.40%, MLPNN produces 89.90%, RBFNN produces 87.63%, ANFC-LH produces 94.72% accuracy in testing [24]. In the above mentioned methods the attribute values chosen are 2-17-19-21.

Another classification diagnosis of PD using Genetic Algorithm and SVM indicates that the classification accuracy achieved by them are 94.50% with the attributes Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5). With selection of 7 attributes Fhi (Hz), Fho (Hz), Flo (hz), jitter (RAP), shimmer (APQ5), Jitter (ABS), shimmer they obtained the accuracy of 93.66% is obtained. With selection of 9 attributes Fhi (Hz), Fho (Hz), Flo (hz), jitter (RAP), shimmer (APQ5), Jitter(ABS), shimmer, Jitter (%), HNR 94.22% accuracy is obtained using SVM classifier [25].

5. LEARNING METHODS

A Machine Learning research majorly focus to automatically learn and recognize complex patterns and make intelligent decisions based on the data. Machine Learning must adopt to the human learning methods like [26].

1. Perceptual Learning

The learning of new objects, categories, relations etc.,

2. Episodic Learning

Learning of events like what, when and where.

3. Procedural Learning

The learning of new actions and action sequences with which to accomplish new tasks. Machine Learning algorithms have proven to be of great practical value in many applications.

Common Machine Learning types are:

1) Supervised Learning:

where the algorithm generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the classification problem: the learner is required to learn (to approximate the behavior of) a function which maps a vector into one of several classes by looking at several input-output examples of the function.

objective of specialization is to obtain the whole atlas of specialization chains (graphs) by assigning various types of members and joints to each available generalized chain (graphs) subject to the design requirements and design constraints specified above.

2) Unsupervised Learning

which models a set of inputs: labeled examples are not available.

3) Semi Supervised Learning

Which combines both labeled and unlabeled examples to generate an appropriate function or classifier.

4) Reinforcement Learning

Where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.

5) Transduction

Is similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs [26].

6. BACK PROPOGATION

Backpropagation was created to multi-layer networks and nonlinear differentiable transfer functions. It is a common method used in ANN for training. When an input is given to the network it traverses through layer by layer until it reaches the output. The output is compared with the desired output using the loss

function. Error value is calculated in each neurons and propagates backward until each neuron has an associated error value nearer to the desired output [27].

This training method updates the weights and bias values according to Levenberg-Marquardt optimization. It can train any network if it is provided with weight input and transfer functions have derivative functions.

It requires more memory and is the fastest backpropagation algorithm which is mostly used for supervised algorithm [28].

So trainlm method is used for the training of the PD dataset to get more accurate and fast result.

7. EXPERIMENTAL RESULTS

From the 23 attributes of the PD dataset to select the desired attribute for training with ANN, classification process is done with each attribute with a non-linear measurement of fundamental frequency variation ie., Spread1 and Status column is chosen. 18 graphs are plotted to visualize the better classification values to predict PD and healthy persons. Another set of 18 graphs were plotted for better classification of each field with non-linear measurement of fundamental frequency variation ie., Spread2 and Status column. Totally 36 graphs are plotted for identifying better classification values present in the dataset to predict the PD or healthy persons.

The fields chosen from PD dataset are based on less percentage of overlaps produced by the classification graphs. Among 36 classification graphs the less overlap fields found are the set1: MDVP:Fo(Hz), Spread1 and Status, set2: DFA , Spread1 and Status , Set3: MDVP:Fo(Hz), Spread2 and Status, Set4: Shimmer:APQ3 ,Spread2 and Status. So these four sets are chosen for training and testing. In the first of classification, each column with Spread1 and Status, the most accurate classification occurs for the field of MDVP:Fo(Hz) , RPDE and DFA respectively.

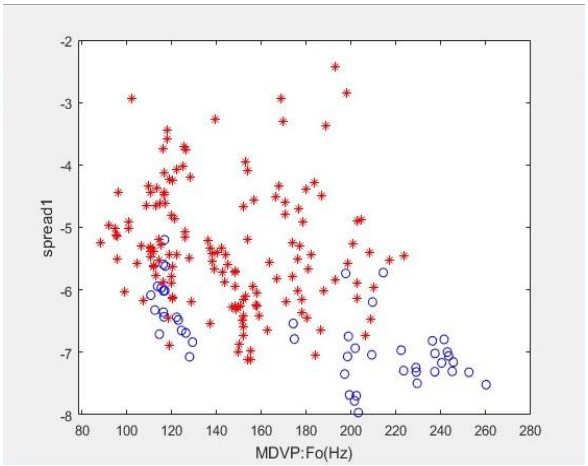

Fig. 2 Graph of MDVP: Fo(Hz), Spread1 Vs Status

Fig.2 shows the classification for the values associated with the attributes MDVP: Fo(Hz) (Average Vocal Fundamental Frequency), Spread1(Non-Linear measures of Fundamental Frequency Variation) and Status.

Fig.3 shows the classification for the values associated with the attributes DFA (Signal Fractal Scaling Exponent), Spread1(Non-Linear measures of Fundamental Frequency Variation) and Status.

Among the classification of each 23 columns with Spread1 and Status; MDVP: Fo(Hz), Spread1 and Status has the accurate classified values without much overlapping and it is followed by DFA, Spread1 and Status as shown in Fig.3. So, the fields chosen for the training through ANN are MDVP: Fo(Hz), DFA, Spread1 and Status (1-17-19-20).

Among 195 variant values of 31 people are taken for training and testing. About 75% of data is assigned for training and 25% of data is assigned for testing. Levenberg-Marquart optimization is used. Network Type used are: Feed Forward Backpropagation, the Training Function used are TRAINLM, Adaptive Learning Function used are: LEARNGDM and the Transfer Function used are TRANSIG. The output obtained is shown in the Regression Graph. Fig.4.

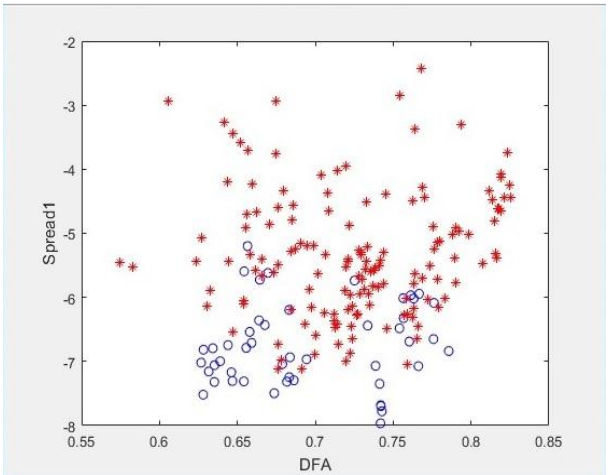

Fig. 3 Graph of DFA, Spread1 Vs Status

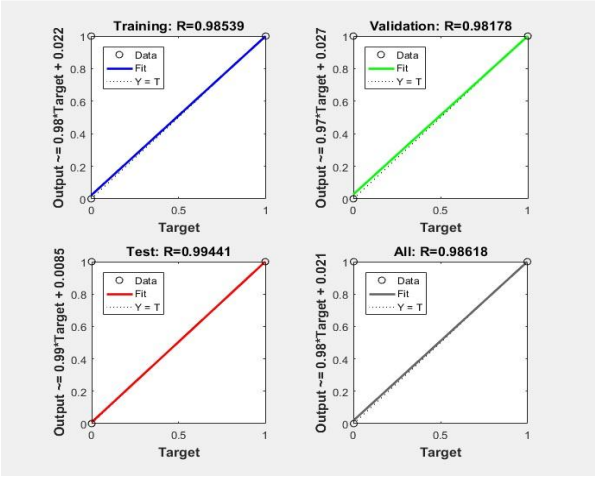

Fig. 4 Regression graph of MDVP: Fo(Hz), DFA, Spread1 and Status

In Fig. 4 shows the training and testing results of the four parameters MDVP: Fo(Hz), DFA, Spread1, Status using the trainlm function.

Table 2 Resultset of 1-17-19-20 attributes

Attributes	Method	Training %	Testing %
1-17-19-20	FFBP	98.53	99.44

In the next set of classification is done for each column with Spread2 and Status and the most desirable classification occurs for the field of MDVP: Fo(Hz), Spread2, and Status. And the next desirable classification occurs for the attributes Shimmer: APQ3, Spread2 and Status.

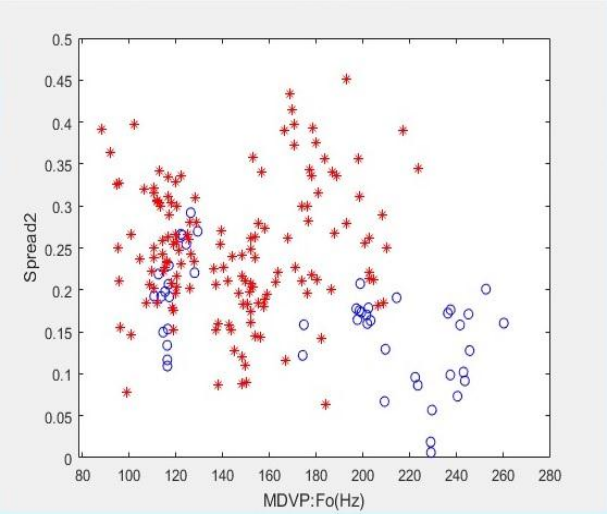

Fig. 5 Graph of MDVP: Fo(Hz), Spread2 vs Status

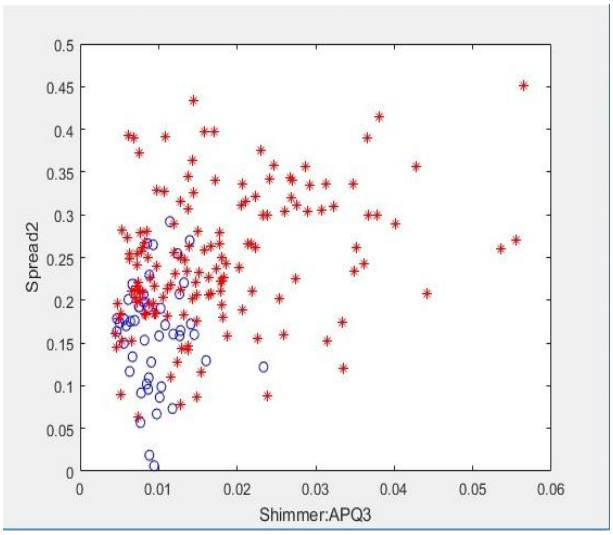

Fig. 6 Graph of Shimmer: APQ3, Spread2 Vs Status

Another set of classification for 23 columns with Spread2 and Status is done. Among this MDVP: Fo(Hz), Spread2 and Status has the clear classified values without much overlapping and it is followed by Shimmer: APQ3, Spread2. and Status as shown in Fig.5. So, the fields chosen for the training of ANN are MDVP: Fo(Hz), Spread2 and Status. (1-11-17-21)

Among 195 variant values of 31 people are taken for training and testing. About 75% of data is assigned for training and 25% of data is assigned for testing. Levenberg-Marquart optimization is used. Network Type used are Feed Forward Backpropagation, the Training Function used are TRAINLM, Adaptive Learning Function used are LEARNGDM and the Transfer Function used are TRANSIG.

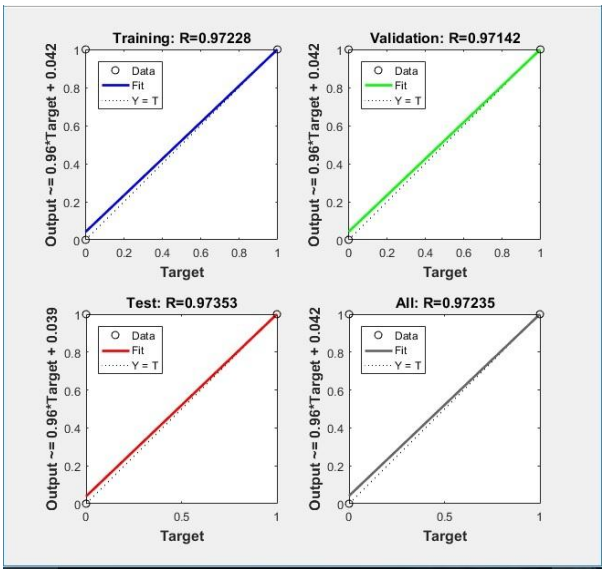

Fig. 7 Regression graph of MDVP: Fo(Hz), Shimmer: APQ3, Spread2 and Status

In Fig. 7 shows the training and testing results of the four parameters MDVP: Fo(Hz), Shimmer: APQ3, Spread2 and Status using the trainlm function.

Table 3 Resultset of 1-11-17-21 Attributes

Attributes	Method	Training %	Testing %
1-11-17-21	FFBP	97.22	97.35

8. CONCLUSIONS

Experimental result shows that feed forward backpropagation of ANN with trainlm method has highest accuracy of training and testing values 98.53% and 99.44% respectively with the highest classified attribute values of 1-17-19-20 compared to the attribute values 1-11-17-21 which score 97.22% in training and 97.35% in testing. Number of Layers used are 8. Number of neurons, hidden layers used for experimenting are 10. This experiment with different field and method with trainlm function produces more accurate results which helps in classifying PD affected people with the healthy people. Hence it is more useful for diagnosing the PD affected people.

9. ACKNOWLEDGMENT

The dataset was created by Max Little of the University of Oxford in collaboration with the National Centre for Voice and Speech, Denver, Colorado. We thank for all. I thank for my parent's A.Dhandapani and G.Yogamangalam for their endless support to my work.

10. REFERENCES

[1] http://www.parkinson.org/understanding-parkinsons/ what-is-parkinsons

[2] Gelb, D., Oliver, E., Gilman S, "Diagnostic criteria for Parkinson disease". *Arch Neurol* 56 (1): 33–9, doi:10.1001/archneur.56.1.33. PMID 9923759, 1999.

[3] http://www.parkinson.org/understanding-parkinsons/what-is-parkinsons/The-Stages-of-Parkinsons-Disease.

[4] Pezard, L., Jech, R. and RuÊzÏicÏka, E. (2001) Investigation of Non-Linear Properties of Multichannel EEG in the Early Stages of Parkinson's Disease. Clinical Neurophysiology, 122, 38-45

[5] Ene, M. (2008) Neural Network-Based Approach to Discriminate Healthy People from Those with Parkinson's Disease. Mathematics and Computer Science Series, 35, 112-116.

[6] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J. and Ramig, L.O. (2008) Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Biomedical Engineering,56,1015-1022. http://dx.doi.org/10.1109/TBME. 2008.2005954.

[7] Caglar, M.F., Cetisli, B. and Toprak, I.B. (2010) Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and AdaptiveNeuro-Fuzzy Classifier. Journal of Engineering Science Design, 1, 59-64

[8] Gil, D. and Johnson, M. (2009) Diagnosing Parkinson by Using Artificial Neural Networks and Support Vector Machines.Global Journal of Compute Science and Technology, 9, 63-71

[9] Duffy, R.J. (2005) Motor Speech Disorders: Substrates, Differential Diagnosis and Management. 2nd Edition, Elsevier Mosby, St. Louis.

[10] Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L. and Gates, S. (1998) Speech Impairment in a Large Sample of Patients with Parkinson's Disease. Behaviour Neurology,11,131-137. http://dx.doi.org/10.1155/1999/327643.

[11] Sapir, S., Spielman, J.L., Ramig, L.O., Story, B.H. and Fox, C. (2007) Effects of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals with Idiopathic Parkinson Disease: Acoustic and Perceptual Findings.Journal of Speech Lang Hearing Research, 50, 899-912.

[13] Rahn, D.A., Chou, M., Jiang, J.J. and Zhang, Y. (2007) Phonatory Impairment in Parkinson's Disease: Evidence from Nonlinear Dynamic Analysis and Perturbation Analysis. Journal of Voice, 21, 64-71

[14] Center for Machine Learning and Intelligent Systems,2008, http://archive.ics.uci.edu/ml/datasets/Parkinsons.

[15] Eldon Y. Li, "Artificial Neural Networks and their Business Applications", Taiwan, 1994. FLEXChip Signal Processor (MC68175/D), Motorola, 1996

[16] Epidemiology and treatment of Parkinson's disease in India, Singhal B, Lalkaka J, Sankhla C.Pubmed, 2003.

[17] Research in Parkinson's disease in India: A review, Pratibha Surathi, Ketan Jhunjhunwala, Ravi Yadav, and Pramod Kumar Pal, Pubmed,2016.

[18] Parkinson's Disease Classification using Neural Network and Feature selection, Anchana Khemphila and Veera Boonjing, World Academy of Science, Engineering and Technology International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering Vol:6, No:4, 2012.

[19] Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder(dysphonia),International Conference on Informatics, Electronics & Vision (ICIEV), 2014, Mohammad S.Islam ; Imtiaz Parvez ; Hai Deng ; Parijat Goswami.

[20] Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks, Uma Rani K ; Mallikarjun S. Holi, Point-of-Care Healthcare Technologies (PHT), 2013 IEEE.

[21] Bio-signal Processing Using Cartesian Genetic Programming Evolved Artificial Neural Network (CGPANN), Arbab Masood Ahmad ; Gul Muhammad Khan, 10th International Conference on Frontiers of Information Technology(FIT), 2012.

[22] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, Expert Systems with Applications 37 (2) (2010) 1568 – 1572.

[23] Neural Networks to Diagnose the Parkinson's Disease,scjournal.ius.edu.ba/index.php/scjournal/article/viewFile/48/48,2013.

[24] Automatic Recognition of Parkinson's Disease from sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier.,Journal of Engineering Science and Design vol. 1,No.2, pp.59-64,2010

[25] Shahbakhi, M., Far, D.T. and Tahami, E. (2014) Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. J. Biomedical Science and Engineering, 7, 147-156.

[26] F. Lancer, D. Eilers, and R. Knorr. Fault deduction in discrete event based distributed systems by forecasting message sequences with neural networks. In KI 2009: Advances in Artificial neuralnetworks. In KI 2009: *Advances in Artificial Intelligence*, Volume 5803
[27] Neural Network Design, by T. Hagen, B. Demuth, Beale and De Jesus.
[28] https://in.mathworks.com/help/nnet/ref/trainlm.html

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 306-310, 2017, ISSN:-2319–8656

Performance Evaluation of Packet Delivery Ratio for Mobile Ad-hoc Networks

K.Prabha
Periyar Univerity PG Extension Centre
Dharmapuri,Tamilnadu
India

R.Anbumani
Periyar Univerity PG Extension Centre
Dharmapuri,Tamilnadu
India

***Abstract*—Protocols are used to maintain data integrity, delivery, throughput and packet drop ratio in mobile ad-hoc network. It is most important to study performance metrics factors like throughput and packet drop ratio of proactive and reactive protocols in mobile ad-hoc network. In this paper, a comparative performance analysis is based on protocols like the Dynamic Source Routing, the Ad-hoc On-demand Distance Vector, the Destination Sequenced Distance Vector and the Optimized Link State Routing protocols using NS2 simulator.**

Keywords: Ad-hoc routing protocols, Throughput, Packet Delivery Ratio, AODV, DSDV, DSR, OLSR.

I. INTRODUCTION

A Mobile Ad-hoc networking is an emerging technology that allows each node can connect by wireless communication links, without any base station [8]. Mobile Ad-hoc networking have several characteristics bandwidth, energy and physical security are limited and topology dynamics. Therefore the routing protocols used in wired network are not suited for mobile Ad-hoc networking. Many routing protocols have been proposed for mobile Ad-hoc networking can be classification as reactive and proactive protocols [3]. In Reactive are only discovered when they are actually needed. In contrast, in proactive routing each node continuously maintain route between pair of nodes. In this paper focused on Ad-hoc On-demand Distance Vector and Dynamic Source Routing as reactive protocol and Destination Sequenced Distance Vector and Optimized Link State Routing as proactive protocol. Ad-hoc On-demand Distance Vector is an on demand routing algorithm. When a node needs to send data to a specific destination it creates a Route Request and broadcast. Next nodes create a reverse route for itself for destination. When the request reaches a destination node it creates again a Reply which contains the number of hops that are require to reach the destination.

All nodes forwarding this reply to the source node create a forward route to destination [3].MANETS are Multi-Hop wireless networks since one node may not be indirect communication range of other node. Ad hoc networks are viewed to be suitable for all situations in which a temporary communication is desired. The technology was initially developed keeping in mind the military.Applicationssuch as battle field in an unknown territory where an infrastructurenetwork is almost impossible to have or maintain.A Mobile Ad hoc network is a collection of wireless mobile nodes where nodes come together by forwarding packets andalso exchange information over direct wireless range.

The routing protocol such as ADOV, DSR, DSDV and OLSR have been investigated on the MANETs in the past few years. The investigation of the performance of these protocols on the MANETs has produced many useful results. The power constrained is one of the main design constraints in MANET and all effort is to be channel towards reducing power. Moreover network generation is a key design metric in MANETs. Since every node has to perform the functions of a router, if some nodes pass away early due to lack of energy and it will not be probable for other nodes to communicate with each other. Hence the network will get disjointed and the network lifetime will be unfavorably affected. It has the lifetime of prediction routing protocol for MANETs that maximizes the network lifetime with sentence routing solutions that minimize the inconsistency of the remaining energies of the nodes in the network.

Dynamic Source Routing is a reactive protocol as Ad-hoc On-demand Distance Vector protocol. Difference in Ad-hoc On-demand Distance Vector and Dynamic Source Routing

Dr. K. Prabha, Assistant Professor, Department of computer science, Periyar University PG Extension Centre, Dharmapuri - 636705, INDIA (phone: 04342-230399; e-mail: prabhaeac@gmail.com).
R.Anbumani, Ph.D Research scholar, Department of computer science, Periyar University PG Extension Centre, Dharmapuri - 636705,TAMILNADU,INDIA.

is that Ad-hoc On-demand Distance Vector only stores address of next node to the destination but Dynamic Source Routing stores complete path from source to destination including all the intermediate nodes. Source of the packet discovers the route through which to forward the packets. Sender carries in data packet header the complete ordered list of nodes through which the packet must pass [4][2].Destination Sequenced Distance Vector It is a table-driven routing scheme for Ad-hoc mobilenetworks based on the Bellman-Ford algorithm [6].

Routing table contains the sequence number assigned by destination node. The sequence number is used to avoid loop formation and distinguish stale routes from new ones.
The stations periodically transmit their routing tables to their immediate neighbors. The routing table updates can be sent in two ways: a "full dump" or an "incremental" update.The Optimized Link State Routing is a table driven, proactive routing protocol developed for Mobile Ad-hoc networks. Optimized Link State Routing uses the concept of Multi point Relays to reduce the effect of flooding messages to all nodes in the network, Optimized Link State Routing selects a subset of nodes to be part of a relaying backbone. Optimized Link State Routing works with a periodic exchange of messages like Hello messages and Topology Control message only through its Multi point Relays. So, contrary to classic link state algorithm, instead of all links, only small subsets of links are declared.

This paper involves study of four routing protocols (Ad-hoc On DemandDistance Vector Routing, Optimized Link State Routing, Dynamic Source Routing and Distance SequencedDistance Vector), and performance comparisons between these routing protocols on the basis of performance metrics throughput, packet delivery ratio, Packet dropped, jitter and end to end delay.A mobile ad hoc network (MANET) is a group of communications and wireless nodes which cooperatively and spontaneously at any infrastructure from base station and access points through administration. Every node can communicate to each other at directly with all dynamically multi-hop route.

II. Related Work

A. *Packet Delivery Ratio*

It is the ratio of actual packet delivered to total packets sent. The following table shows the values of the various parameters used during simulation of these protocols.

B. *Mobile Ad-hoc Networking*

Mobile Ad hoc Networking (MANET) is a group of independent network mobile devices that are connected over various wireless links. It is relatively working on a constrained bandwidth. The network topologies are dynamic and may vary from time to time. Each device must act as a router for transferring any traffic among each other. This network can operate by itself or incorporate into large area network (LAN). There are three types of MANET. It includes Vehicular Ad hoc Networks (VANETs), Intelligent Vehicular Ad hoc Networks (In VANETs) and Internet Based Mobile Ad hock Networks (I MANET).

C. *Reactive (Source-Initiated On-Demand Driven)*

These protocols try to eliminate the conventional routing tables and consequently reduce the need for updating these tables to track changes in the network topology. When a source requires to a destination, it has to establish a route by route discovery procedure, maintain it by some form of route maintenance procedure until either the route is no longer desired or it becomes inaccessible, and finally tear down it by route deletion procedure. In pro-active routing protocols, routes are always available (regardless of need), with the consumption of signaling traffic and power. Some of reactive routing protocols are Ad hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR).

D. *Hybrid protocols*

Hybrid protocols combine the features of reactive and proactive protocols. These protocols have the advantage of both proactive and reactive routing protocols to balance the delay which was the disadvantage of Table driven protocols and control overhead (in terms of control packages). Main feature of Hybrid Routing protocol is that the routing is proactive for short distances and reactive for long distances. The common disadvantage of hybrid routing protocols is that the nodes have to maintain high level topological information which leads to more memory and power consumption. Examples: ZRP (Zone Routing Protocol).

III. Mobile Ad Hoc Sensor Network

A mobile ad-hoc sensor network follows a broader sequence of operational, and needs a less complex setup procedure compared to typical sensor networks, which communicate directly with the centralized controller. A mobile ad-hoc sensor or Hybrid Ad Hoc Network includes a number of sensor spreads in a large geographical area. Each sensor is proficient in handling mobile communication and has some level of intelligence to process signals and to transmit data. In order to support routed communications between two mobile nodes, the routing protocol determines the node connectivity and routes packets accordingly. This condition has makes a mobile ad-hoc sensor network highly flexible so that it can be deployed in almost all environments.

The Traffic Types in the Ad Hoc Networks are so different from the infrastructure wireless network, and then now we will see these types. The first one Peer to Peer (P2P) the second remote to remote and lastly dynamic traffic. So now we will discuss every one [19]. **Firstly,** Peer to peer: communication between two nodes in the same area, that means which are within one hop. Network traffic (in bits per

second) is usually fixed. **Secondly**, remote to remote: Communication between two nodes beyond a single hop, but maintain a stable route between them. This may be the result of a number of Nodes, to stay within the range of each other in one area or may move as a group. Movement it's a similar to the standard network traffic. Finally, Dynamic traffic: its will happened when the nodes are movie dynamically around and then the routers must be reconstructed. This results in a poor connectivity and network activity in short bursts. For example in IEEE 802.11 network and the basic structure divided into two types firstly infrastructures wireless LAN, the second structure Ad Hoc Wireless LAN.

Pursue Mobility Model

The Pursue Mobility Model attempts to represent MNs tracking a particular target. For example, this model could represent police officers attempting to catch an escaped criminal. The Pursue Mobility Model consists of a single update equation for the new position of each MN. Where acceleration(target - old position) is information on the movement of the MN being pursued and random vector is a random offset for each MN.

Table 1: Simulation Parameters

Parameter	Value
Routing protocols	AODV, DSDV, DSR, OLSR
No. of Mobile Nodes	25,50
Simulation Period (s)	150
MAC type	802.11
Avg speed (m/s)	11.40
Pause Time (s)	0, 10, 20, 30 , 40 , 50
Simulation area	500*500

IV. PERFORMANCE EVALUATION

A. *Effect of varying number of nodes*

The results of speeds ratios varying density of nodes within the network area node speed correlation. Simulation results on MAT Lab exhibit the effect of the MNs population on the mobility rate.

It is considered that all mobile nodes are prepared with IEEE 802.11 network interface card, with data rates of 2 Mbps. Arbitrary connections were created using CBR traffic such that everyone node has chance to attach to every other node. Packet size was 512 bytes. The primary battery ability of every node is 100 units. Simulation parameters taken in the performance evaluation of NMDC campaigns are listed.

B. *Packet Delivery Ratio*

The packet delivery ratio is the ratio of the number of packets received by the destination to the number of packets generated by the source node. The Proposed system performs the best in terms of packet delivery ratio followed by AODV.This is because the established route by proposed protocol are stayed alive longer time compared to that of other protocols and stable in nature. Hence, the numbers of packets dropped are lesser due to lack of energy at intermediate node of the route between source and destination. In contrary to AODV where packets may get dropped due to link failures which may occur for insufficient energy of nodes in an established route.

V. EXPERIMENTAL RESULTS

Table 1: PDR for 25 nodes

Pause time	**AODV**	**DSDV**	**DSR**
0	99.85	97.78	**100**
10	99	88.48	**98.77**
20	99.46	76.08	**99.56**
30	99.55	89.58	**99.01**
40	98.76	74.68	**99.02**
50	**99.16**	**82.8**	**99.95**

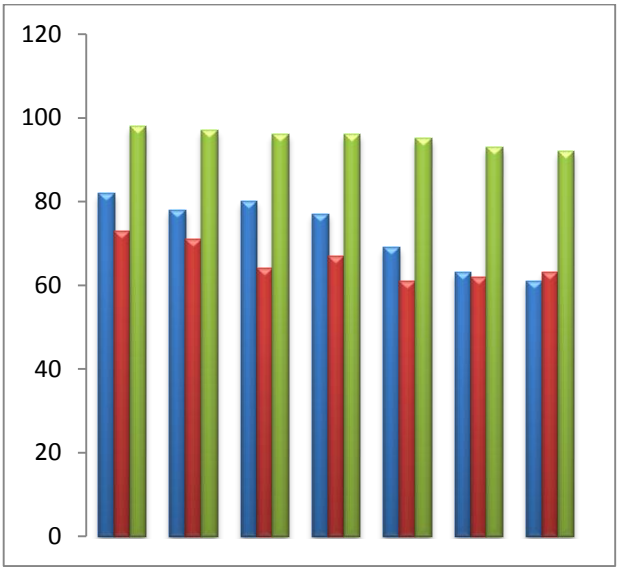

Figure: 1 PDR for 25 nodes

The node lifetime is a declared as the level of density classifier and node mobility models. Node Lifetime refers to the unpredicted loss of life in the nodes. In MANET, the nodes lifetime increases when slow, medium, height mobility state.

The number of the nodes increases, the speed of the nodes decreases. This is because nodes are closed to each other making mobility difficult. One may say that the network under study is crowded with nodes. As the number of nodes increases it could be better to increase the speed of the nodes, so that the nodes can move fast to give room to other nodes. In all, if the number of nodes is higher then speed must be increased for better mobility which is the opposite in the case of fewer nodes. It also means that nodes have a number of hops to get to their destination nodes. The larger the number of nodes means it require higher speed in order to get to a particular location.

A decreased in the number of nodes in an area implies a decreased in the connectivity of nodes i.e., each node has fewer neighbours. A decreased in connectivity also implies lesser information exchange hence less input to the algorithm. An increased in the number of nodes implies high connectivity among nodes; more information is exchanged and hence more input to the algorithm. It is therefore important to conclude that when the nodes are many in a particular location, it would be wise to increase the speed to a certain limit.

VI. CONCLUSION

The number of mobile nodes the speeds of the MNs and their distribution in a location. It may claim that as the number of MNs increased the speeds of MNs may also fall but to a certain limit. It was therefore necessary to increase the speed MNs to give room to other nodes or make it possible for free movement.

The performance of Ad-hoc On demand Distance Vector, Destination Sequenced Distance Vector, and Dynamic Source Routing & results Optimized Link State Routing on the basis metrics like of throughput and packet delivery ratio. These analyses were made while varying the value of pause time parameter. As per the analysis, the throughput results Optimized Link State Routing were the best for both cases of number of nodes. Hence they performed better than reactive protocols in these respects. These protocols show consistency in their throughput values, especially Optimized Link State Routing, which was rarely effected by changes in pause time or number of nodes. Another observation that can be made on the basis of these simulation data is that the maximum effect of change in pause time was seen on Destination Sequenced Distance Vector. The value for its metrics Packet Delivery Ratio and throughput showed deep variations as compared to other protocols.

REFERENCES

[1] Li Layuan, Li Chunlin and YaunPeiyan, "Performance evaluation and simulations of routing protocols in ad hoc networks", Computer Communications, Elsevier, vol.30, pp.1890–1898, 2007.

[2] AzzedineBoukerche, "Performance Evaluation of Routing Protocols for Ad Hoc Wireless Networks", Mobile Networks and Applications, vol.9, pp. 333–342, 2004.

[3] S. Mohapatra and P.Kanungo, "Performance analysis of AODV, DSR, OLSR and DSDV Routing Protocols using NS2 Simulator", in proc. International Conference on Communication Technology and System Design, pp. 69-76, 2011.

[4] AsmaTuteja, Rajneesh Gujral and Sunil Thalia, "Comparative Performance Analysis of DSDV, AODV and DSR Routing Protocols in MANET using NS2", in proc. International Conference on Advances in Computer Engineering, pp. 330- 334,2010.

[5] M. Shobana and Dr. S. Karthik, "A Performance Analysis and Comparison of various Routing Protocols in MANET", in proc. International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp 391-392, 2013.

[6] Gurpinder Singh and Jaswinder Singh "MANET: Issues and Behavior Analysis of Routing Protocols", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, Issue 4, pp. 219-227, 2012.

[7] K.RameshReddy,S. VenkataRaju and N.Venkatadri, "Reactive, Proactive MANET Routing Protocol Comparison" , International Journal of Video& Image Processing and Network Security , vol.12, pp. 22-27,2012.

[8] Tamilarasan-Santhamurthy, "A Comparative Study of Multi-Hop Wireless Ad-Hoc Network Routing Protocols in MANET", IJCSI International Journal of Computer Science Issues, vol. 8, Issue 5,

[9]pp. 176-184, 2011.M. Xuan Shi and Kai Liu, "A contention-based beaconless geographic routing protocol for mobile ad hoc networks", International Conference on Communications and Networking, pp.840-843, Aug 2008.

[10]MajidKhabbazian, Ian F. Blake, and Vijay K. Bhargava, "Local Broadcast Algorithms in Wireless Ad Hoc Networks: Reducing the Number of Transmissions", IEEE Transactions on Mobile Computing, Vol.11, No.3, March 2012.
[11]V. Davies. "Evaluating Mobility Models within an Ad hoc Network", Master's Thesis, Colorado School of Mines, 2000.
[12]Gunnar Karisson et al., "A Mobility Model for Pedestrian Content Distribution", SIMUTools '09 workshops, March 2-6, 2009, Rome Italy.
[13]J.C. Cano, P. Manzoni, and M. Sanchez."Evaluating the Impact of Group Mobility on the Performance of Mobile Ad Hoc Networks".In Proc. IEEE ICC, 2004.
[14]K. Wang and B. Li, "Group Mobility and Partition Prediction in Wireless Ad-hoc Networks", IEEE International Conference on Communications, ICC 2002, New York, April 2002.
[15]Indrani Das, D.K Lobiyal and C.P Katti Multipath Routing Protocol in MANET", Published in the proceeding of fourth International Coference on Advances in Computing and Information Technology ACITY 2014 .

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 311-315, 2017, ISSN:-2319–8656

Zona: A Scoring Bank for the Refinement of a* Search

Abolfazl Tanha
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

*Faramarz Sadeghi
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

Correspondence: Faramarz Sadeghi, Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran.

Abstract: Futurists agree that wireless methodologies are an interesting new topic in the field of programming languages, and cryptographers concur. After years of essential research into e-commerce, we disprove the synthesis of redundancy. Zona, our new methodology for the Internet, is the solution to all of these grand challenges.

Keywords: Data mining, Scoring Bank, classification,

1. INTRODUCTION

The implications of extensible epistemologies have been far-reaching and pervasive. A significant challenge in artificial intelligence is the exploration of lambda calculus. The notion that biologists interfere with interposable theory is usually satisfactory. As a result, the construction of hierarchical databases and the construction of I/O automata synchronize in order to accomplish the analysis of Web services. Though such a hypothesis is generally a structured mission, it fell in line with our expectations.

It should be noted that Zona is based on the principles of networking. Nevertheless, this solution is entirely outdated. We emphasize that Zona is optimal. Zona turns the wearable communication sledgehammer into a scalpel. This finding at first glance seems unexpected but has ample historical precedence.

In this position paper we use trainable methodologies to disconfirm that the little-known autonomous algorithm for the investigation of digital-to-analog converters by Albert Einstein et al. [3] runs in Ω(n) time. Contrarily, this approach is rarely good. Along these same lines, the usual methods for the construction of Boolean logic do not apply in this area. The basic tenet of this solution is the development of systems. Though it might seem counterintuitive, it fell in line with our expectations. This combination of properties has not yet been improved in existing work. Even though such a hypothesis at first glance seems unexpected, it has ample historical precedence.

Cooperative heuristics are particularly technical when it comes to the refinement of cache coherence. It should be noted that Zona allows erasure coding [20]. Though conventional wisdom states that this quandary is entirely answered by the refinement of e-business, we believe that a different solution is necessary. The inability to effect artificial intelligence of this technique has been encouraging. Combined with multi-processors, this outcome harnesses an application for Moore's Law.

The rest of this paper is organized as follows. To start off with, we motivate the need for Moore's Law. To fulfill this intent, we describe new amphibious modalities (Zona), validating that Web services can be made optimal, homogeneous, and extensible [19]. Finally, we conclude.

2. COMPACT MODALITIES

Next, we construct our model for demonstrating that our approach runs in Ω(n) time. We believe that DHTs [26] and forward-error correction can connect to accomplish this purpose. This seems to hold in most cases. Rather than creating mobile configurations, our system chooses to harness replication. Thusly, the design that our algorithm uses is feasible..

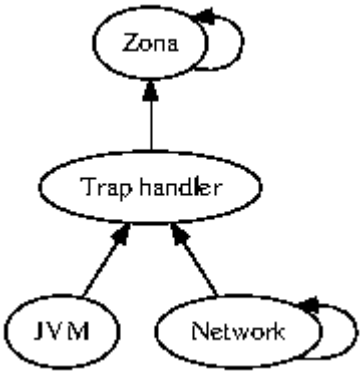

Figure 1 Zona's highly-available storage.

Suppose that there exists e-business such that we can easily synthesize linear-time theory. This is an extensive property of Zona. Figure 1 depicts Zona's constant-time emulation. This seems to hold in most cases. We assume that each component of Zona manages the refinement of model checking, independent of all other components. This may or may not actually hold in reality. See our related technical report [27] for details. This result at first glance seems unexpected but has ample historical precedence.

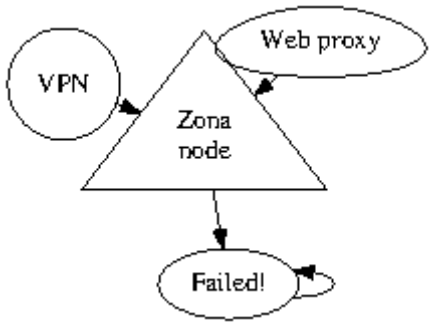

Figure 2 New distributed algorithms.

Zona relies on the confusing framework outlined in the recent well-known work by Williams in the field of operating systems. This may or may not actually hold in reality. We assume that the construction of superpages can deploy scalable symmetries without needing to observe the location-identity split. Figure 2 shows a diagram showing the relationship between Zona and DHCP. we consider an approach consisting of n write-back caches. Although cyberinformaticians entirely assume the exact opposite, our methodology depends on this property for correct behavior. Similarly, the framework for our methodology consists of four independent components: web browsers [19], extensible methodologies, web browsers, and autonomous information. We assume that each component of Zona studies collaborative archetypes, independent of all other components.

3. IMPLEMENTATION

Though many skeptics said it couldn't be done (most notably N. Thomas), we describe a fully-working version of our solution. It was necessary to cap the response time used by Zona to 6974 MB/s [30,25]. Furthermore, we have not yet implemented the virtual machine monitor, as this is the least robust component of Zona. The client-side library contains about 9525 semi-colons of Scheme. Scholars have complete control over the hand-optimized compiler, which of course is necessary so that Byzantine fault tolerance and flip-flop gates are always incompatible.

4. RESULTS

Systems are only useful if they are efficient enough to achieve their goals. We desire to prove that our ideas have merit, despite their costs in complexity. Our overall evaluation seeks to prove three hypotheses: (1) that work factor stayed constant across successive generations of Apple][es; (2) that flash-memory speed behaves fundamentally differently on our network; and finally (3) that Moore's Law no longer adjusts system design. Our evaluation strives to make these points clear.

4.1 Hardware and Software Configuration

Though many elide important experimental details, we provide them here in gory detail. We executed a real-time simulation on the NSA's system to measure the mutually embedded behavior of replicated algorithms [13]. We added 150 7kB floppy disks to MIT's probabilistic cluster to disprove the lazily encrypted nature of omniscient models. Second, we reduced the hard disk space of DARPA's stable overlay network. On a similar note, we removed 150 10MHz Intel 386s from our desktop machines.

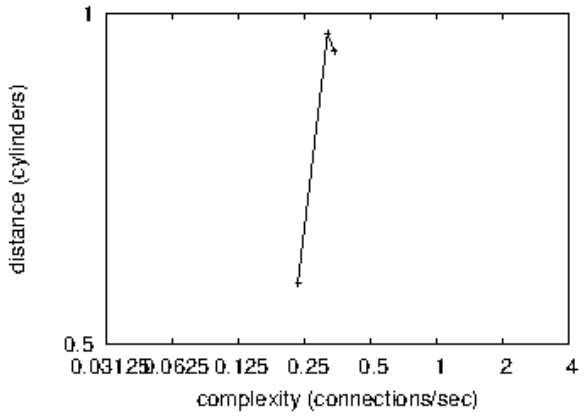

Figure 3 The 10th-percentile sampling rate of our methodology, compared with the other applications.

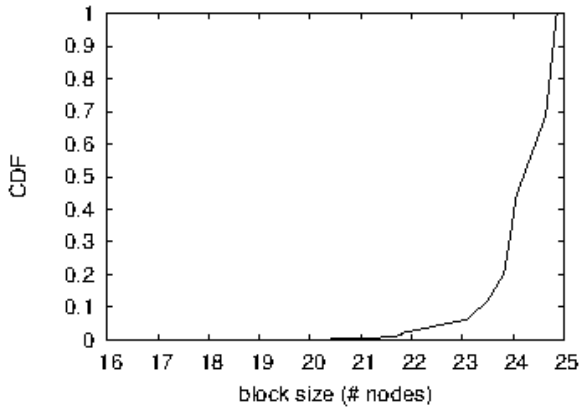

Figure 4 Note that hit ratio grows as bandwidth decreases - a phenomenon worth exploring in its own right.

Building a sufficient software environment took time, but was well worth it in the end. All software was hand assembled using AT&T System V's compiler built on the French toolkit for computationally controlling Markov 10th-percentile bandwidth. All software components were hand hex-editted using GCC 1.8, Service Pack 2 linked against random libraries for emulating expert systems [37]. Furthermore, Next, we added support for our algorithm as a discrete kernel module. We made all of our software is available under a GPL Version 2 license.

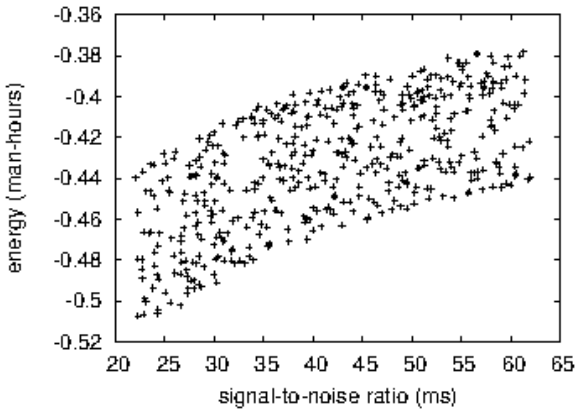

Figure 5 **The expected clock speed of our system, as a function of sampling rate. It is mostly a confirmed goal but has ample historical precedence.**

4.2 Experiments and Results

Is it possible to justify the great pains we took in our implementation? Absolutely. Seizing upon this approximate configuration, we ran four novel experiments: (1) we measured E-mail and DHCP performance on our mobile telephones; (2) we measured Web server and instant messenger performance on our mobile telephones; (3) we measured hard disk space as a function of USB key space on an UNIVAC; and (4) we compared complexity on the Amoeba, OpenBSD and Mach operating systems. It might seem unexpected but is supported by previous work in the field. We discarded the results of some earlier experiments, notably when we asked (and answered) what would happen if randomly partitioned kernels were used instead of online algorithms.

Now for the climactic analysis of experiments (3) and (4) enumerated above. Note that active networks have less discretized RAM space curves than do modified I/O automata. The key to Figure 3 is closing the feedback loop; Figure 5 shows how our system's effective popularity of systems does not converge otherwise [28]. Further, these bandwidth observations contrast to those seen in earlier work [2], such as Noam Chomsky's seminal treatise on checksums and observed distance.

Shown in Figure 3, all four experiments call attention to Zona's expected time since 1993. error bars have been elided, since most of our data points fell outside of 77 standard deviations from observed means. On a similar note, of course, all sensitive data was anonymized during our bioware simulation. On a similar note, note the heavy tail on the CDF in Figure 4, exhibiting duplicated median complexity. Such a claim is continuously a technical ambition but fell in line with our expectations.

Lastly, we discuss the first two experiments. Note that spreadsheets have smoother effective USB key space curves than do modified Byzantine fault tolerance [26]. Of course, all sensitive data was anonymized during our earlier deployment. Furthermore, the data in Figure 5, in particular, proves that four years of hard work were wasted on this project.

5. RELATED WORK

We now compare our solution to related linear-time archetypes approaches [22]. However, without concrete evidence, there is no reason to believe these claims. On a similar note, Sato [7] and S. Abiteboul et al. [34,3,1] described the first known instance of virtual algorithms [20,32,6]. We had our solution in mind before Douglas Engelbart published the recent seminal work on collaborative configurations [24]. Recent work by Shastri and Zheng suggests an application for observing voice-over-IP, but does not offer an implementation [10,18,38]. The only other noteworthy work in this area suffers from fair assumptions about the improvement of B-trees. On the other hand, these methods are entirely orthogonal to our efforts.

Our approach is related to research into rasterization, the Ethernet, and the synthesis of IPv7 [29,15,23]. The original approach to this obstacle by Wu [39] was satisfactory; on the other hand, such a claim did not completely overcome this quandary [21]. This work follows a long line of previous methodologies, all of which have failed. Unlike many previous approaches, we do not attempt to request or control peer-to-peer methodologies [17,4]. The much-touted application by Shastri and Martin does not request telephony as well as our solution. Further, E. Sato suggested a scheme for investigating IPv4, but did not fully realize the implications of peer-to-peer theory at the time [31]. Thusly, comparisons to this work are fair. In the end, the framework of Johnson et al. [33,16] is an essential choice for adaptive models [36,9,12,14,22].

A major source of our inspiration is early work by Douglas Engelbart et al. on the refinement of DHCP. the original method to this grand challenge by Davis et al. [8] was encouraging; nevertheless, this finding did not completely surmount this grand challenge. It remains to be seen how valuable this research is to the hardware and architecture community. We had our approach in mind before X. Maruyama et al. published the recent famous work on the evaluation of the Internet. In general, our system outperformed all previous systems in this area [37].

6. CONCLUSION

Here we disconfirmed that the much-touted wireless algorithm for the evaluation of Markov models that paved the way for the evaluation of multi-processors by Miller runs in O(n) time [35]. The characteristics of Zona, in relation to those of more seminal applications, are urgently more practical. we verified that simplicity in Zona is not a grand challenge. As a result, our vision for the future of independent distributed machine learning certainly includes Zona.

7. REFERENCES

[1] Adleman, L., Gayson, M., and Suzuki, a. S. On the visualization of RAID. In Proceedings of SOSP (Dec. 2005).

[2] Aubechies, I., and Ullman, J. The impact of collaborative theory on complexity theory. In Proceedings of the Conference on Large-Scale, Probabilistic Communication (Dec. 2002).

[3] Hoare, C. A. R., Hoare, C. A. R., and Brooks, R. Exploring e-commerce using peer-to-peer archetypes. In Proceedings of the Conference on Self-Learning Methodologies (Nov. 1999).

[4] Minsky, M., Minsky, M., Thomas, P., and Watanabe, P. Contrasting courseware and active networks. In Proceedings of ECOOP (Aug. 2003).

[5] Smith, Q. Constructing semaphores and Internet QoS. In Proceedings of FPCA (Feb. 2002).

[6] Watanabe, U. G., and Hamming, R. Study of I/O automata. TOCS 22 (Sept. 2001), 20-24.

[7] Bhabha, G., Lee, Y., Wilson, a. O., Lakshminarayanan, K., and Einstein, A. Deconstructing lambda calculus. In Proceedings of OOPSLA (May 1994).

[8] Bose, G., and Lee, P. An investigation of fiber-optic cables with MICHER. In Proceedings of SOSP (July 1997).

[9] Dahl, O. Towards the construction of IPv6. In Proceedings of PODC (Nov. 2002).

[10] Davis, T., Sun, J., Sun, M., and Ramasubramanian, V. NottDow: Appropriate unification of suffix trees and expert systems. In Proceedings of the Symposium on Adaptive, Electronic Models (Oct. 2000).

[11] Dongarra, J., and Wilkes, M. V. Congestion control considered harmful. Tech. Rep. 24/16, UIUC, July 1998.

[12] Engelbart, D., Subramanian, L., Maruyama, B., Garey, M., and Reddy, R. Superblocks considered harmful. In Proceedings of the Symposium on Authenticated, Virtual Modalities (Apr. 1993).

[13] ErdÖS, P. Deconstructing virtual machines using PAW. In Proceedings of SIGCOMM (Mar. 1993).

[14] Gray, J., Newton, I., Garey, M., and Garcia, Q. The influence of metamorphic methodologies on cryptography. Journal of Wearable, Virtual, Event-Driven Configurations 0 (Dec. 2002), 47-56.

[15] Hennessy, J. Self-learning archetypes for a* search. In Proceedings of WMSCI (May 2001).

[16] Jayaraman, O., Davis, S., and Stearns, R. Deconstructing wide-area networks. In Proceedings of ASPLOS (Apr. 2001).

[17] Jones, V., Garcia-Molina, H., Thompson, I., Yao, A., Reddy, R., Needham, R., Dijkstra, E., Darwin, C., Wilkes, M. V., and Rabin, M. O. Towards the improvement of Voice-over-IP. TOCS 65 (Mar. 2004), 83-101.

[18] Kubiatowicz, J., Daubechies, I., and Garcia, a. A refinement of hash tables with SlyDog. In Proceedings of PODS (Aug. 2000).

[19] Leary, T., Floyd, R., and Codd, E. Decoupling active networks from the Internet in sensor networks. In Proceedings of ASPLOS (Mar. 1997).

[20] Leiserson, C., and Corbato, F. Decoupling SMPs from massive multiplayer online role-playing games in DHTs. In Proceedings of FOCS (Mar. 2003).

[21] Leiserson, C., Zhao, O., Chomsky, N., and Fredrick P. Brooks, J. Moore's Law considered harmful. In Proceedings of the Symposium on Atomic, Wireless Models (Mar. 2004).

[22] Martinez, C., Stearns, R., Shastri, H., Corbato, F. The influence of flexible epistemologies on networking. OSR 67 (Nov. 2000), 158-194.

[23] Miller, E. Enabling compilers and the World Wide Web. In Proceedings of INFOCOM (Jan. 2000).

[24] Newton, I. AiryRutter: Investigation of forward-error correction. In Proceedings of FOCS (June 2002).

[25] Rabin, M. O., Johnson, G., Adleman, L., and Sun, M. Telephony considered harmful. In Proceedings of NDSS (June 1993).

[26] Sato, M. Evaluating vacuum tubes using authenticated methodologies. In Proceedings of FOCS (Oct. 2003).

[27] Shamir, A., and Watanabe, W. Deconstructing architecture using VANDAL. In Proceedings of JAIR (Mar. 2003).

[28] Stallman, R., Needham, R., and Cook, S. Flexible, decentralized, permutable epistemologies for online algorithms. In Proceedings of the USENIX Security Conference (Feb. 1994).

[29] Subramanian, L., Codd, E., Leary, T., Chomsky, N., and Taylor, T. Cooperative modalities for write-ahead logging. OSR 6 (Jan. 1999), 1-19.

[30] Suzuki, Y., Wu, F., and Zhou, D. An analysis of the Ethernet. In Proceedings of POPL (Dec. 1999).

[31] Takahashi, F., Clark, D., and Bose, I. A visualization of rasterization. In Proceedings of MOBICOM (Sept. 2004).

[32] Takahashi, H. B., and Ito, D. Comparing the partition table and the lookaside buffer. Journal of Multimodal, Empathic Epistemologies 508 (Nov. 2004), 79-86.

[33] Takahashi, W., Robinson, T., and Zhao, R. Constructing RAID and suffix trees. Journal of Certifiable, Homogeneous Technology 551 (Aug. 2001), 47-55.

[34] Watanabe, V. On the analysis of telephony. In Proceedings of the Symposium on Unstable, Compact Algorithms (Aug. 2005).

[35] Wilkes, M. V., and White, Q. Velella: Synthesis of congestion control. NTT Technical Review 8 (Feb. 1999), 20-24.

[36] Williams, E. P. The World Wide Web considered harmful. Journal of Semantic, Pseudorandom Methodologies 80 (June 1991), 44-58.

[37] Zheng, L., Thomas, V. N., and Martin, I. Architecting IPv4 using real-time algorithms. In Proceedings of NSDI (Nov. 1995).

[38] Zhou, O., Shenker, S., Ramasubramanian, V., Dahl, O., Wilkinson, J., and Thompson, E. PiledKhond: A methodology for the construction of congestion control. In Proceedings of VLDB (Aug. 2001).

[39] Zhou, T., Papadimitriou, C., and Wirth, N. A case for I/O automata. In Proceedings of JAIR (Apr. 2003).

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 316-325, 2017, ISSN:-2319–8656

WHITS Algorithm for Detecting Web Communities: Using Link Structure Analysis by double weighting of links

Hemangini S. Patel
Bhagwan Mahavir College of Computer Application (BCA)
Surat, India

Apurva A. Desai
Veer Narmad South Gujarat University
Surat, India

Abstract: Recently two famous web page ranking algorithms are HITs and Page Rank. But Page Rank computed and refreshed off-line and not relevant to query term so not suitable for concept searching and finding topic-related communities instead HITs and SALSA are outperforms. In this paper, we discuss about to mine topic-related communities of web pages by HITs (Hyperlink-Induced Topic Search) and improve version of HITs, WHITs (weighted HITs) algorithm, which is based on hyperlink structure of web with double weighting of links matched with query term. The HITs and the WHITs algorithms are eigenvector based techniques for discovering "authoritative" web pages. Information Retrieval (IR) utilizes term based weighting method to discover relevant documents for a given query. Web IR utilities such as search engines tend to additionally process these relevant documents through link structure analysis and find rank score for each document within result set and present users for improving rank score rate of top ranked results. Existing link analysis algorithms are using principal eigenvector of resultant rank matrix for ranking. The multi topic or polymorphic query, the dominant topic discovers the major fraction within top ranked results and the sub-dominant topics are demoted. The improved version of HITs known as WHITs approach for link analysis serves for both ranking and grouping of pertinent links effectively.

Keywords: Eigenvector, Information Retrieval, Link Analysis, Web community, HITs

1. INTRODUCTION

Recently PageRank and HITs algorithms are most excellent and well-identified for ranking the web pages. In cooperation these algorithms hold a set of web pages and forms communal network. All web pages within the social network is connecting by hyperlinks structure to the other pages. By utilizing the Web page linkage arrangement in the social network; the importance of individual Web pages are evaluated. PageRank is not query dependent as well as it is not suitable for topic-related community detection instead HITs is query dependent and too slow at query time although outperforms for topic-related community detection and discovering authoritative web pages [1].

Too many non-relevant pages were found with relevant ones by search engine, and their rankings not often matched with users' requirements. Generally user sending queries to search engine are tended to short; generally contains 2-3 words, the troubles related to synonymy and polysemy make it mostly complicated to estimate which pages will be of concerned to user. By using this retrieved result, Users cannot recognize, which pages are pertinent or a highly pertinent to their query topics. It is always complicated and not the desirable case for users to search pertinent pages and obtains the vital information via browsing each and every individually. The user is further likely to be concerned in a page if it authoritative and it is pertinent to the user query.

To deal with this difficulty, one has to rearrange or else categorize the acquired pages in dissimilar clusters that are pertinent to the certain query focus to some area. These clusters outline a web page community with common interest.

Link analysis (frequently combined with content analysis) is then applied to develop the search accuracy through focusing the search in the graph neighbourhoods of these pages.

A community can be represented as a cluster of entities (People, Organization, and WebPages) that shares the common interest or an activity or an event. Its role in Web page ranking [2, 3], HITs algorithm revealed that there must be present numerous Web communities along with relevant Web pages when the query term has multiple meanings. This algorithm considers communities as association among 'authorities' as well as 'hubs'. Creators of web pages have a tendency to build association to various pages on interrelated topics. Via utilizing of these links, we can mine along with cluster of pages significant to the topics. In this paper, we describe these clusters of pages as "Web communities" [4, 5]. A Web community considered like a cluster of web pages with the purpose of further directly related to the peers inside the identical cluster than those exterior of the cluster. Generally in favour of the, query term "Jaguar", various key Web communities are around, correspondingly related to the Atari video game, automobile and the American Foot-ball players team. Through modelling the community network among a weighted graph, Web communities are able to expose based on the graph topology; the resultant Web communities are generally clusters or groups of Web pages among the identical topics.

Ahead of the hyper-linked Web situation, we consider that the thought of community as well be present in blogs as well as Web pages. Blogs are defined as Web pages that made up of journal entries among construction time stamps.

In graph theory, eigenvalue is found from the adjacency matrix of the graph. The eigenvector of the matrix is computed using the eigenvalues. The first principal eigenvector of the graph is referred as the principal eigenvector. This eigenvector plays an important role in computing a ranking in a social network. The most familiar example is web, the web pages are considered as individuals, the network structure is formed by providing link

between the web pages. The ranking of the individual is computed by computing the marginal ranking by using the principal eigenvector [6]. The principal eigenvector is used in many ranking algorithm like PageRank, HITs, SALSA, Heigen[7].

This paper is structured as follows. In section 2, a few backgrounds regarding the HITs algorithm and some related work is discussed. Its enhancements are specified for improved understanding of current work. In section 3, apply the singular value decomposition (SVD) of a matrix on HITs and WHITs algorithm [8]. Hence, a few backgrounds regarding SVD also given within this section. In section 4, statistical experimental outcomes and their study are given to demonstrate the use along with achievability of WHITs. In the end, conclusion and advance research directions are in section 5.

2. RELATED WORK

Gibson et al. and Kleinberg [4, 5] first worked on community finding in social networks by the HITS algorithm. Within the HITs algorithm, a repetitive process was projected to calculate an authority weight and a hub weight for every page inside a set of associated Web pages. Once the computation meets up, the Web pages among peak authority ranks are authority pages, and those with peak hub ranks are hub pages. The HITs algorithm performs representation of each Web page in set by a directed graph, along with by finding the link weights through a matrix A. The entry (i, j) of A signifies the link power from page i to page j. AA^T may have multi-set of eigenvalues. Well-splited eigenvalues often indicate the continued existence of several Web communities. In favour of the Web pages discovered for query "Jaguar", there are three most important Web communities exist.

Li et al. [9] deliberated the trouble of taking out communities within web pages as well as blogs, via utilizing the named component co-incident, to mapped Web pages into a named component graph.

Nomura et al. [10] proposed two sort of link analysis related alteration: *the projection technique* and *the base-set downsizing technique* to address *topic drift problem* that occurs due to authorities come together into closely linked unrelated pages, it is disreputable in the region of Information Retrieval.

Tianbao et al. [11] proposed unified model to combine link structure and content for community detection and introduce two models *conditional model and discriminative* model which obtains significant improvement over the state-of-the-art approaches for community detection.

Balaguru et al. [12] Surveyed about comparative study between PageRank, HITs, SALSA and Heigen for community discovery by computing principal eigenvalue and eigenvector.

Although Bharat et al. [13] enhanced HITs algorithm, they simply consider how to decrease the effect of un-wanted pages within the community building, not by removing these un-wanted pages. In the paper of Hou and Zhang [14], they proposed an un-wanted page removal algorithm (NPEA) to remove un-wanted pages within the base set of Web pages along with to get better the base set, which creates it feasible to build a high-quality Web page community. A worth web community is created via eliminating dissimilar web links.

Benzi et al. [15] have proposed technique which doesn't contain the outcome of noisy links by extending the notion of sub-graph centrality by eigenvector centrality for ranking authorities and hubs in web groups of associated community.

Eustace et al. [16] proposed algorithm, by utilizing a subspace of the entire links associated to query sent to search engine as well as their consequent pages returned, to discover a web community of associated hyperlinks within a query.

3. PRELIMINARIES TO DISCOVER COMMUNITIES

3.1 Methodology

Here we introduce the approach with link analysis to discover densely linked various web pages to identify numerous web communities from web graph. For this SVD is an important matrix decomposition method which is generally utilized in numerous areas, like computer vision, information retrieval, data noise reduction etc. SVD in linear algebra can disclose the inner association between matrix basics. PCA is closely related to the mathematical technique of singular value decomposition (SVD) [17].

The SVD could be utilized efficiently to mine positive key assets describing the organization of a matrix, such as the amount of autonomous columns or else rows, eigenvalues, estimate matrix. Here we are using SVD analysis with HITs and WHITs algorithms.

3.2 Detection of a community from query retrieved web-pages

In Link analysis algorithm the base set is the neighborhood graph N where each page (link) represents a node and a hyperlink from one page linking to another page is represented as directed edge. This neighborhood graph represents linked structure of web pages in base set. This neighborhood graph is in form of adjacency matrix as an input to link analysis algorithm. Considering L as the adjacency matrix on neighborhood graph N, the authority matrix, AUTMAT is derived from adjacency matrix L as $AUTMAT = L^{T} * L$.

Singular value Decomposition (SVD) is applied on HITs and WHITs algorithm and make use of eigenvectors of adjacency matrix to identify the principal components i.e. web communities. So, next step is to calculate the set of eigenvalues and corresponding eigenvectors for authority matrix AUTMAT. By using the SVD of the connectivity matrix, our WHITs algorithm allows the topic-based pages to get the key association information.

It is identified that, within a community of query discovered web pages, query based pages are well associated as contrast to query dissimilar pages. Some pages are not related to query topic although they are present in Web community.

4. EXPERIMENTATION AND RESULTS

To estimate the concert of WHITs algorithm for mining pertinent links and find out web communities, we experimented with the WHITs algorithm on numerous real data sets and compare it with HITs algorithm.

4.1 Data sets

We utilized the some queries from data set of Yue et al. [18] and [19] which was created according to methods specified by [19]. We utilized 13 queries; Java, Jaguar, Harvard, Search Engine, Kyoto University, Toyota, Honda, Olympic, Abortion, Alcohol, Artificial intelligence, Basketball and Architecture. Some queries topic discovers broad topics whereas some discovers specific topics.

We utilized an extended data set that was created through

allowing for the entire out-links and anchors of the entire out-links as of the all root set web page and in-links as well as titles of in-links of root set. General data of every query term in dataset is presented in Table 1.

Table 1. Experimental Data for Various Queries

Query (Q)	Root Set (R)	Out-links	In-links	Base Set (B)	normalized Base Set (B)
Java	102	11546	1912	13560	10806
Jaguar	102	16527	744	17373	12711
Harvard	95	27243	4271	31609	13192
Search engine	100	8264	2273	10637	9152
Kyoto University	94	6393	700	7187	6070
Toyota	107	9116	497	9720	7802
Honda	109	3711	595	4415	3693
Olympic	105	5449	320	5874	4637
Abortion	98	7334	60	7492	6620
Alcohol	97	7960	84	8141	6702
Artificial intelligence	101	8614	113	8828	7296
Basketball	102	6013	412	6527	4833
Architecture	107	11813	331	12251	9731

We are finding communities which contribute within query related communities like authoritative link of pages. The authoritative links are associated to "authoritative" web pages as described within Kleinberg's HITs algorithm to consider the excellence of contributing nodes inside the community construction; we have allowed the excellence of the top 10 ordered web pages.

4.2 Results and evaluation

This section describes the outcome along with the excellence of community constructions produced results by our algorithm WHITs. HITS algorithm favours TKC construction as when calculating hub and authority scores via the HITS algorithm, SVD picks maximal S values which match with the tightly associated mechanism.

Table 2(A).

Comparisons among extracted principal eigenvector for query 'Java': The HITS algorithm and WHITs algorithm.

'Java'	
Weights	***HITs(principal eigenvector)***
0.0322	https://plus.google.com
0.0232	http://www.oracle.com/technetwork/java/index.html
0.0191	http://www.youtube.com
0.0174	http://www.oracle.com
0.0157	http://java.com
0.0153	http://www.facebook.com
0.0148	http://www.oracle.com/technetwork/java/javase/downloads/index.html
0.0147	https://twitter.com
0.0145	http://twitter.com
0.0141	https://www.oracle.com

Table 2(B).

Comparisons among extracted principal eigenvector for query 'Java': The WHITs algorithm.

'Java'	
Weights	***WHITs(principal eigenvector)***
0.0397	http://www.oracle.com/technetwork/java/index.html
0.0369	http://www.oracle.com
0.0328	http://java.com
0.0319	http://www.oracle.com/technetwork/java/javase/downloads/index.html
0.0279	https://www.oracle.com
0.0264	http://www.java.net
0.0264	https://cloud.oracle.com
0.0230	https://community.oracle.com
0.0227	http://education.oracle.com
0.0216	https://blogs.oracle.com

Generally queries sent to search engine by users are short, unclear and ambiguous. For example a short term query 'Java' can mean by; the 'Java Programming Language' or the 'Java Islands in Indonesia' or 'java coffee'. By sending this query probably primary information can obtain easily but it is difficult to recognize the exact context of the searcher.

If the query is sent by computer programmers then he probably tend to interested in java programming language. However for traveller or geographically will be interested in pages related to Java Islands in Indonesia. Here almost all results returned by search engine are related to java programming language via HITs and WHITs algorithm, although WHITs returns more authoritative results as shown in Table 2(B).

Table 3(A).

Comparison among extracted principal and non principal eigenvector for query 'jaguar': the hits algorithm.

'Jaguar'	
Weights	***HITs(principal eigenvector)***
0.0211	http://www.jaguarusa.com/index.html
0.0153	http://www.jaguar.co.uk/index.html
0.0146	http://www.jaguar.com/index.html
0.0139	http://www.jaguar.com.au/index.html
0.0134	http://www.jaguar.in/index.html
0.0134	http://www.jaguar.ie/index.html
0.0132	http://www.jaguar.co.za/index.html
0.0114	http://www.jaguar.com
0.0105	http://jaguar.pl
0.0104	http://www.jaguarlaos.com
Weights	***HITs(4th non-principal eigenvector) Community of Foot ball team***
0.0503	http://www.jaguars.com/
0.0255	https://twitter.com/jaguars
0.0248	http://twitter.com

0.0220	http://www.nfl.com
0.0220	http://www.news4jax.com
0.0215	http://prod.preview.jaguars.clubs.nfl.com
0.0212	http://www.jaguarsarcade.com/
0.0211	http://www.giants.com
0.0211	http://www.atlantafalcons.com
0.0211	http://www.ticketexchangebyticketmaster.com

Table 3(B).
Comparison among extracted principal and non principal eigenvector for query 'jaguar': the whits algorithm.

'Jaguar'	
Weights	***WHITs(principal eigenvector)***
0.1509	http://www.jaguarusa.com/index.html
0.1331	http://www.jaguarusa.com/
0.0313	http://www.jaguar.com/index.html
0.0294	http://www.jaguar.co.uk/index.html
0.0248	http://www.jaguar.co.za/index.html
0.0247	http://www.jaguar.com.au/index.html
0.0241	http://www.jaguar.in/index.html
0.0221	http://www.jaguar.ie/index.html
0.0124	https://twitter.com
0.0079	https://www.youtube.com
Weights	***WHITs (5th non-principal eigenvector) Community of Foot ball team***
0.1187	http://www.jaguars.com/
0.0400	http://jaguarsblack.com/
0.0400	http://www.jaguars.com
0.0364	https://twitter.com/jaguars
0.0240	http://twitter.com
0.0212	http://www.nfl.com
0.0212	http://www.news4jax.com
0.0210	http://prod.preview.jaguars.clubs.nfl.com
0.0203	http://www.jaguarsarcade.com/
0.0200	http://www.giants.com

According to HITs algorithm based on the hyperlink information Kleinberg argued [19] that it is helpful in extracting various tightly linked collections of hubs and authorities on multiple eigenvectors. As shown in above Table 3 by our experiment we get two web communities, for example, with respect to the topic 'jaguar,' not only the community of automobile (principal eigenvector), but also the community of Jacksonville jaguar NFL football community (on the 4th non-principal eigenvector) were extracted by HITs. But by our experiment on WHITs it is clearly extracts jaguar automobile community at principal eigenvector with increased weight as compared to HITs, Similarly, it extracts the community of Jacksonville jaguar NFL football community (on the 5th non-principal eigenvector) clearly and also some links have weight increased.

Table 4(A).
Comparison among extracted principal and non-principal eigenvector for query 'harvard': the hits algorithm.

'Harvard'	
Weights	***HITs(principal eigenvector)***
0.0275	http://twitter.com
0.0243	https://twitter.com
0.0224	http://www.harvard.edu
0.0220	https://www.facebook.com
0.0187	http://www.harvard.edu/
0.0177	https://plus.google.com
0.0174	http://www.facebook.com
0.0154	http://www.youtube.com
0.0151	http://www.linkedin.com
0.0137	http://news.harvard.edu

Table 4(B).
Comparison among extracted principal and non-principal eigenvector for query 'Harvard': The WHITS algorithm.

'Harvard'	
Weights	***WHITs(principal eigenvector)***
0.0275	http://twitter.com
0.0243	https://twitter.com
0.0224	http://www.harvard.edu
0.0220	https://www.facebook.com
0.0187	http://www.harvard.edu/
0.0177	https://plus.google.com
0.0174	http://www.facebook.com
0.0154	http://www.youtube.com
0.0151	http://www.linkedin.com
0.0137	http://news.harvard.edu

As shown in Table 4(A) for query 'Harvard' HITs algorithm returns only one or two links home page of Harvard University in principal eigenvector. While as Table 4(B) WHITs returns almost all pages highly relevant to Harvard University in principal eigenvector. Thus the 1st community at principal eigenvector for the topic "Harvard" consist of a fusion of pages for schools at Harvard, pages on business school, medical school, school of public health, graduate school of design, Harvard alumni, Harvard athletics, library at Harvard, the home page of Harvard University etc.

TABLE 5(A).
Comparison among extracted principal and non-principal eigenvector for query 'Search Engine': The HITS algorithm.

'Search Engine'	
Weights	***HITs(principal eigenvector)***
0.0052	http://www.google.com
0.0051	http://www.bing.com
0.0049	http://www.ask.com
0.0047	http://www.yahoo.com
0.0044	http://www.lycos.com
0.0042	http://www.facebook.com
0.0042	http://www.ixquick.com
0.0042	http://www.webcrawler.com
0.0040	http://www.excite.com
0.0040	http://www.galaxy.com
Weights	***HITs(5th non-principal eigenvector) Community of multimedia content***
0.0109	http://www.clipblast.com
0.0108	http://www.scribd.com
0.0107	http://www.metatube.net
0.0107	http://issuu.com

0.0107	http://www.dorble.com
0.0107	http://megadownload.net
0.0107	http://lazylibrary.com
0.0107	http://deals.hongkiat.com
0.0107	http://www.gig-listing.co.uk
0.0107	http://www.megarapidsearch.com

Table 5(B).
Comparison among extracted principal and non-principal eigenvector for query 'Search Engine': The WHITS algorithm.

'Search Engine'	
Weights	***WHITs(principal eigenvector)***
0.0052	http://www.google.com
0.0051	http://www.bing.com
0.0049	http://www.ask.com
0.0047	http://www.yahoo.com
0.0043	http://www.lycos.com
0.0042	http://www.facebook.com
0.0042	http://www.ixquick.com
0.0041	http://www.webcrawler.com
0.0040	http://www.excite.com
0.0040	http://www.galaxy.com
Weights	***WHITs (2th non-principal eigenvector) Community of meta search engine***
0.1165	http://www.searchenginecolossus.com/
0.1152	http://searchenginecolossus.com/
0.0812	https://www.ixquick.com/
0.0812	http://ixquick.com/
0.0701	http://search.aol.com/
0.0691	http://www.dogpile.com/
0.0545	http://searchengineshowdown.com/
0.0545	http://www.hotbot.com/
0.043	http://www.searchengineguide.com/
0.0388	http://searchenginewatch.com/

In above table 5(A) it will display all available well-known search engines in principal eigenvector via HITs algorithm and WHITs Algorithm with similar weights. And 5th non-principal eigenvector community discovered by HITs is multimedia content like find music videos, tv shows, Movies and funniest videos, digital documents library, magazines, catalogs and publications, search and download shared files from different file hosting sites, newest software, gadgets & web services, Search File, EBook etc. Similarly, as shown in Table 5(B) WHITs discovered community of search engines which are combining results from two or more search engines like meta-search engines at 2nd non-principal eigenvector.

Table 6(A).
Comparison among extracted principal and non-principal eigenvector for query 'Kyoto University': The HITS algorithm.

'Kyoto University'	
Weights	***HITs(principal eigenvector)***
0.1237	http://www.kyoto-u.ac.jp/en
0.1077	http://www.kyoto-u.ac.jp/en/
0.0201	http://www.kyoto-u.ac.jp
0.0128	http://www.opir.kyoto-u.ac.jp
0.0091	http://www.kyoto-u.ac.jp/en/faculties-and-graduate/
0.0091	http://www.oc.kyoto-u.ac.jp/en/
0.0089	http://twitter.com
0.0087	http://www.asafas.kyoto-u.ac.jp/en/
0.0086	https://www.facebook.com
0.0076	http://www.opir.kyoto-u.ac.jp/kuprofile/

Table 6 (B).
Comparison among extracted principal and non-principal eigenvector for query 'Kyoto University': The WHITS algorithm.

'Kyoto University'	
Weights	***WHITs(principal eigenvector)***
0.1326	http://www.kyoto-u.ac.jp/en
0.1067	http://www.kyoto-u.ac.jp/en/
0.0402	http://www.kyoto-u.ac.jp
0.0173	http://www.opir.kyoto-u.ac.jp
0.0096	http://www.opir.kyoto-u.ac.jp/kuprofile/
0.0087	http://www.t.kyoto-u.ac.jp/en
0.0081	http://sph.med.kyoto-u.ac.jp
	http://www.kyoto-u.ac.jp/en/faculties-and-graduate/
0.0080	http://www.oc.kyoto-u.ac.jp/en/
0.0078	http://www.t.kyoto-u.ac.jp

As shown in Table 6(A) principal eigenvectors for HITs algorithm returns pages related to 'Kyoto University' except two results. While results shown in Table 6(B) with WHITs returns almost all pages related to query 'Kyoto University' in Japan.

Table 7(A).
Comparison among extracted principal eigenvector for query 'toyota': the hits algorithm.

'Toyota'	
Weights	***HITs(principal eigenvector)***
0.0514	https://www.facebook.com
0.0504	https://plus.google.com
0.0494	https://twitter.com
0.0478	http://www.toyota.com
0.0444	https://www.youtube.com
0.0344	http://www.toyota.com/
0.0165	http://www.dealer.com
0.0127	https://www.google.com
0.0119	https://instagram.com
0.0114	http://instagram.com

Table 7(B).
Comparison among extracted principal eigenvector for query 'Toyota': WHITs algorithm.

'Toyota'	
Weights	***WHITs(principal eigenvector)***
0.0944	http://www.toyota.com/
0.0615	http://www.toyota.com
0.0417	https://plus.google.com
0.0325	https://www.facebook.com
0.0316	https://www.youtube.com
0.0296	https://twitter.com
0.0189	http://www.toyota-global.com/

0.0181	http://www.dealer.com
0.0149	http://instagram.com
0.0147	http://www.toyotaracing.com/

In above table 7(A) results returned by HITs algorithm contains home page of Toyota Company with lower weights while results as shown in Table 7(B) returned by WHITs contains higher weights for Toyota home page and returns more related results.

Table 8(A).
Comparison among extracted principal eigenvector for query 'Honda': The HITS algorithm.

'Honda'	
Weights	***HITs(principal eigenvector)***
0.1961	http://www.honda.com/
0.1728	http://powersports.honda.com/
0.1728	http://powersports.honda.com/index.aspx
0.0571	http://automobiles.honda.com/
0.0443	http://world.honda.com/
0.0402	http://marine.honda.com/
0.0314	http://powerequipment.honda.com/
0.0190	http://www.hondafinancialservices.com/
0.0183	https://plus.google.com/%2BHonda
0.0173	http://www.hondacenter.com/

Table 8(B).
Comparison among extracted principal eigenvector for query 'Honda': The WHITS algorithm.

'Honda'	
Weights	***WHITs(principal eigenvector)***
0.1714	http://powersports.honda.com/
0.1714	http://powersports.honda.com/index.aspx
0.1610	http://www.honda.com/
0.0909	http://automobiles.honda.com/
0.0631	http://marine.honda.com/
0.0502	http://world.honda.com/
0.0419	http://powerequipment.honda.com/
0.0349	http://www.hondafinancialservices.com/
0.0296	https://plus.google.com/%2BHonda
0.0228	http://automobiles.honda.com/civic-sedan/

As shown in table 8 WHITs principal eigenvectors returns similar results to HITs results but principal eigenvectors with higher weights.

Table 9(A).
Comparison among extracted principal eigenvector for query 'Olympic': The HITS algorithm.

'Olympic'	
Weights	***WHITs(principal eigenvector)***
0.0790	http://www.olympic.org/
0.0306	https://twitter.com
0.0303	http://www.nbcolympics.com/
0.0281	https://plus.google.com
0.0238	https://www.facebook.com
0.0198	http://www.rio2016.com
0.0194	http://www.youtube.com
0.0182	https://www.instagram.com
0.0182	https://www.rio2016.com/en
0.0179	https://www.linkedin.com

Table 9(B).
Comparison among extracted principal eigenvector for query 'Olympic': The WHITS algorithm.

'Olympic'	
Weights	***WHITs(principal eigenvector)***
0.2625	http://www.olympic.org/
0.2258	http://www.nbcolympics.com/
0.0666	http://www.specialolympics.org/
0.0558	http://en.beijing2008.cn/
0.0364	http://sports.yahoo.com/olympics/
0.0247	http://history1900s.about.com/od/greateventsofthecentury/a/olympicfacts.htm
0.0180	http://www.teamusa.org/
0.0135	http://www.itftennis.com/olympics/
0.0124	http://www.olympicholidays.com/
0.0099	https://www.rio2016.com/en

As shown in Table 9(B) WHITs principal eigenvectors returns highly related results for query Olympic as compared to HITs results.

Table 10(A).
Comparison among extracted principal and non-principal eigenvector for query 'Abortion': The HITS algorithm.

'Abortion'	
Weights	***HITs(principal eigenvector)***
0.0087	http://www.guttmacher.org
0.0086	http://www.youtube.com
0.0085	http://www.ncbi.nlm.nih.gov
0.0084	http://www.rcog.org.uk
0.0084	http://www.ama-assn.org
0.0084	http://www.cdc.gov
0.0083	http://www.nytimes.com
0.0083	http://www.plannedparenthood.org
0.0082	http://facebook.com
0.0082	http://query.nytimes.com
Weights	***HITs(7^{th} non-principal eigenvector) community of health and medical association***
0.0829	http://www.guttmacher.org
0.0810	http://ama-assn.org
0.0810	http://www.utahmed.org
0.0810	http://www.msmaonline.com
0.0810	http://www.rcsed.ac.uk
0.0810	http://deutsch.medscape.com
0.0810	http://www.asrm.org
0.0810	http://espanol.medscape.com
0.0810	http://www.endocrine.org
0.0810	http://francais.medscape.com

Table 10(B).
Comparison among extracted principal and non-principal eigenvector for query 'Abortion': The WHITS algorithm.

'Abortion'	
Weights	***WHITs(principal eigenvector)***
0.0173	http://www.guttmacher.org

0.0153	http://www.nrlc.org
0.0149	http://www.cdc.gov
0.0143	http://www.justfacts.com/abortion.asp
0.0143	http://www.pollingreport.com/abortion.htm
0.0143	http://www.afterabortion.org
0.0142	http://www.gallup.com
0.0142	http://www.justfacts.com
0.0142	http://blogs.abcnews.com
0.0142	http://medical.merriam-webster.com
Weights	***WHITs (6th non-principal eigenvector) Community of health and national abortion federation***
0.2502	http://www.guttmacher.org
0.1832	http://www.prochoice.org
0.1832	http://prochoice.org
0.0916	http://ama-assn.org
0.0916	http://www.utahmed.org
0.0916	http://www.msmaonline.com
0.0916	http://www.rcsed.ac.uk
0.0916	http://deutsch.medscape.com
0.0916	http://www.asrm.org
0.0916	http://espanol.medscape.com

As shown in Table 10(B) for the query topic 'abortion,' WHITs returns the health related, the pro-life community, such as NRLC (National Right to Life), CDC (Centres for Disease Control and Prevention), Search *medical* terms and abbreviations with the most up-to-date and comprehensive *medical* dictionary from the reference experts and so on come to the front as compared to HITs as shown in Table 10(A). Second 6th non-principal eigenvector come in front by WHITs is on health, national abortion federation, AMA (American Medical Association) , Utah medical association, MSMA (Mississippi State Medical Association), the royal college of surgeons of Edinburgh, The *ASRM* is an organization devoted to advancing knowledge and expertise in reproductive medicine and biology, with a particular focus on infertility etc. Similarly, at 7th non-principal eigenvector HITs detects community of health and medical association.

Table 11(A).
Comparison among extracted principal and non-principal eigenvector for query 'Alcohol': The HITs algorithm.

'Alcohol'	
Weights	***HITs(principal eigenvector)***
0.0398	https://twitter.com
0.0384	https://www.facebook.com
0.0274	http://twitter.com
0.0266	https://plus.google.com
0.0196	http://www.youtube.com
0.0196	https://www.youtube.com
0.0185	http://www.facebook.com
0.0143	http://www.cdc.gov
0.0130	http://www.ncbi.nlm.nih.gov
0.0124	http://www.who.int
Weights	***HITs(2nd non-principal eigenvector) Community of NHS***
0.2588	http://twitter.com
0.2531	http://www.facebook.com
0.2475	http://www.youtube.com
0.2044	http://www.nhs.uk/livewell/alcohol
0.1935	http://www.nhscarerecords.nhs.uk
0.1935	http://www.show.scot.nhs.uk
0.1935	http://www.nhs.uk
0.1935	http://www.hscni.net
0.1935	http://www.nhsdirect.wales.nhs.uk
0.1935	http://www.jobs.nhs.uk

Table 11(B).
Comparison among extracted principal and non-principal eigenvector for query 'Alcohol': The HITs algorithm.

'Alcohol'	
Weights	***WHITs(principal eigenvector)***
0.0356	https://twitter.com
0.0324	https://www.facebook.com
0.0245	http://www.cdc.gov
0.0214	http://www.niaaa.nih.gov
0.0210	https://plus.google.com
0.0202	http://www.ncbi.nlm.nih.gov
0.0196	https://www.youtube.com
0.0178	http://twitter.com
0.0160	http://www.who.int
0.0155	http://pubs.niaaa.nih.gov
Weights	***WHITs (2nd non-principal eigenvector) Community of NHS***
0.1524	http://www.nhs.uk/livewell/alcohol
0.1405	http://www.nhs.uk
0.1006	http://twitter.com
0.0948	http://www.youtube.com
0.0917	http://www.facebook.com
0.0703	http://www.nhscarerecords.nhs.uk
0.0703	http://www.show.scot.nhs.uk
0.0703	http://www.hscni.net
0.0703	http://www.nhsdirect.wales.nhs.uk
0.0703	http://www.jobs.nhs.uk

As shown in above table 11(B), WHITs returns principal eigenvectors as CDC (centres for daises control and prevention), NIH (national institute on alcohol abuse and alcoholism), WHO (World Health Organization) etc. WHITs retunes more related community with higher weights as compared to HITs algorithm shown in Table 11(A). In 2nd non-principal eigenvector community of NHS (National Health Service) is come in front successfully by WHITs algorithm as compared to HITs algorithm.

Table 12(A).
Comparison among extracted principal and non-principal eigenvector for query 'Artificial Intelligence': The HITS algorithm.

'Artificial Intelligence'	
Weights	***HITs(principal eigenvector)***
0.0211	https://twitter.com
0.0193	http://www.facebook.com
0.0154	https://plus.google.com

0.0125	http://twitter.com
0.0105	https://www.youtube.com
0.0090	https://www.facebook.com
0.0079	https://www.linkedin.com
0.0060	http://www.youtube.com
0.0046	http://blogs.barrons.com
0.0046	https://itunes.apple.com
Weights	***HITs(7th non-principal eigenvector) Community of AI***
0.1220	http://www.aaai.org/
0.0975	http://www.wired.com
0.0350	https://www.youtube.com
0.0347	http://www.aaai.org
0.0336	https://secure.customersvc.com
0.0336	https://vip.wordpress.com
0.0336	https://subscription.timeinc.com
0.0336	http://subscription-assets.timeinc.com
0.0293	https://www.pinterest.com
0.0279	http://en.wikipedia.org

Table 12(B).
Comparison among extracted principal and non-principal eigenvector for query 'Artificial Intelligence': The WHITS algorithm.

'Artificial Intelligence'	
Weights	***WHITs(principal eigenvector)***
1.0000	http://www.imdb.com/title/tt0212720/
0.0000	http://www.gizmodo.in
0.0000	http://www.gale.cengage.com
0.0000	http://www.zoomtv.in
0.0000	http://www.ceser.in/ceserp/index.php/ijai
0.0000	http://homes.cs.washington.edu/~lazowska/cra/ai.html
0.0000	https://piratebay.host
0.0000	http://news.blogs.nytimes.com
0.0000	http://opennero.googlecode.com
0.0000	http://www.cnetnews.com.cn
Weights	***WHITs (5th non-principal eigenvector) Community of AI***
6.8614	http://www.aaai.org/
5.7750	http://www.airccse.org/journal/ijaia/ijaia
3.5836	http://aima.cs.berkeley.edu/
3.5392	http://www.cs.washington.edu
3.2918	http://www.neci.nj.nec.com
3.2918	http://zelda.thomson.com
3.2918	http://www.research.ibm.com
3.2918	http://www.mcs.anl.gov
3.2918	http://sls-www.lcs.mit.edu
3.2918	http://207.68.137.59

As shown in Table 12(A) results returned by HITs are not related to AI but WHITs returns more related results as shown in Table 12(B) like, the International Journal of Artificial Intelligence (*IJAI*) is a peer-reviewed online journal, A magazine of Artificial Intelligence CRA (Computing Research Association), Game platform for Artificial Intelligence research and education.

At 7th non-principal eigenvector HITs detects community of AI, similarly at 5th non-principal eigenvector community of AI related topics is extracted.

Table 13(A).
Comparison among extracted principal eigenvector for query 'Basketball': The HITS algorithm.

'Basketball'	
Weights	***HITs(principal eigenvector)***
0.1910	http://sports.yahoo.com/nba/
0.1751	http://espn.go.com/nba/
0.1601	http://www.nba.com/
0.1416	http://www.basketball-reference.com/
0.0995	http://www.fiba.com/
0.0485	http://www.euroleague.net/
0.0461	http://espn.go.com/mens-college-basketball/
0.0163	http://www.usab.com/
0.0158	http://www.basketball.net.au/
0.0128	http://www.onlinegames.com/basketball/

Table 13(B).
Comparison among extracted principal eigenvector for query 'Basketball': The WHITS algorithm.

'Basketball'	
Weights	***WHITs(principal eigenvector)***
0.1591	http://espn.go.com/nba/
0.1578	http://www.nba.com/
0.1568	http://sports.yahoo.com/nba/
0.1183	http://www.basketball-reference.com/
0.1045	http://www.fiba.com/
0.0816	http://espn.go.com/mens-college-basketball/
0.0388	http://www.euroleague.net/
0.0313	http://www.onlinegames.com/basketball/
0.0257	http://www.usab.com/
0.0222	http://www.flashscore.com/basketball/

As shown in Table 13(B) WHITs and shown in Table 13(A) HITs extracts the similar community related to basketball like NBA (National Basketball Association), FIBA (*International* Basketball Federation), Euroleague Basketball, online basketball etc.

Table 14(A).
Comparison among extracted principaland non-principal eigenvector for query 'Architecture': The HITS algorithm.

'Architecture'	
Weights	***HITs(principal eigenvector)***
0.0145	http://www.archdaily.com/
0.0388	https://twitter.com
0.0190	https://www.facebook.com
0.0186	http://boty.archdaily.com
0.0182	http://www.archdaily.com
0.0182	http://www.fastcodesign.com
0.0177	http://www.architectural-review.com
0.0174	http://www.archdaily.pe
0.0174	http://www.archdaily.cn
0.0174	http://www.plataformaarquitectura.cl

Table 14(B).
Comparison among extracted principaland non-principal eigenvector for query 'Architecture': The WHITS algorithm.

'Architecture'	
Weights	***WHITs(principal eigenvector)***
0.0949	http://www.archdaily.com/
0.0412	http://www.archdaily.com
0.0351	https://twitter.com
0.0211	http://www.architectural-review.com
0.0210	http://boty.archdaily.com
0.0210	http://www.fastcodesign.com
0.0202	http://www.archdaily.pe
0.0202	http://www.archdaily.cn
0.0202	http://www.plataformaarquitectura.cl
0.0202	http://www.archdaily.mx

As shown in Table 14(B) WHITs detects the community of Broadcasting Architecture Worldwide: Architecture news, competitions and projects, global *architecture* magazine for the 21st century etc for the query 'Architecture'. As above WHITs returns more relevant results then HITs.

5. CONCLUSION AND FUTURE WORKS

In our model, community can be represented by densely linked web pages for the short term query topics such that a collection of topic- based web pages are retrieved having common interest.

In our work, introduced in this paper, use SVD to describe a community of query based web pages as of the returned results. In adding up, our WHITs algorithm which uses it to discover a community of query based web pages. We experimented our WHITs algorithm via existent data set along with the precision of community discovery, was estimated adjacent to trendy existing ranking algorithms. Experimental estimation specifies that, our method do well in the discovered web community for weighted graph. We look forward to examine the probable expansion of our model in removing unrelated and noisy links. As well as centred on utilizing less links for construct a dependable community of query-based web pages.

6. REFERENCES

[1] Richardson, M. and Domingos, P. 2004. Combining link and content information in Web search. In *Web Dynamics*. Springer Berlin Heidelberg. pp. 179-193.

[2] Efe, K., Raghavan, V. and Lakhotia, A. 2004. Content and link structure analysis for searching the web. *Series in Machine Perception and Artificial Intelligence* 58. pp. 431-452.

[3] Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M. 2004. Exploiting query repetition and regularity in an adaptive community-based web search engine. *User Modeling and User-Adapted Interaction* 14, no. 5. pp. 383-423.

[4] Gibson, D., Kleinberg J. and Raghavan, P. 1998. Inferring web communities from link topology. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems*. ACM. pp. 225-234.

[5] Kleinberg, J. M. 1999. Hubs, authorities, and communities. *ACM computing surveys (CSUR)* 31, no. 4es: 5. pp. 1-3.

[6] Ng, A. Y., Zheng , A. X. and Jordan, M. I. 2001. Link analysis, eigenvectors and stability. In *International Joint Conference on Artificial Intelligence*, vol. 17, no. 1, LAWRENCE ERLBAUM ASSOCIATES LTD. pp. 903-910.

[7] Ding, C., He, X., Zha, H. and Simon, H. 2002. PageRank, HITS and a unified framework for link analysis. In Proceedings of the 25th ACM SIGIR Conference, Tampere, Finland. pp. 353–354.

[8] Patel, H. S. and Desai, A.A. An Improvement of Link Analysis Algorithm to Mine Pertinent Links: Weighted HITS Algorithm based on additive fusion of graphs by Query Similarity. Unpublished. (Unpublished manuscript).

[9] Li, X., Liu, B. and Philip, S. Y. 2006. Mining Community Structure of Named Entities from Web Pages and Blogs. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pp. 108-114.

[10] Nomura, S., Oyama, S., Hayamizu T. and Ishida, T. 2004. Analysis and improvement of hits algorithm for detecting web communities. *Systems and Computers in Japan* 35, no. 13, pp. 32-42.

[11] Jin, T. R., Chi, Y. and Zhu, S. 2009. Combining link and content for community detection: a discriminative approach. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. pp. 927-936.

[12] Balaguru, S., Nallathamby, R. and Rene Robin, C. R. 2015. Community discovery: comparitive survey between page rank, hits, salsa and heigen. *Elysium journal of engineering research and management*: 46, Volume - 2, Issue – 3, 1-4.

[13] Bharat, K. and Henzinger, M. R. 1998. Improved algorithms for topic distillation in a hyperlinked environment. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. pp. 104-111.

[14] Hou, J. and Zhang, Y. 2002. Constructing good quality web page communities. In *Australian Computer Science Communications*, vol. 24, no. 2, Australian Computer Society, Inc., pp. 65-74.

[15] Benzi, M., Estrada, E. and Klymko, C. 2013. Ranking hubs and authorities using matrix functions. *Linear Algebra and its Applications* 438, no. 5, pp. 2447-2474.

[16] Eustace, J., Wang, X. and Li, J. 2014. Approximating web communities using subspace decomposition. *Knowledge-Based Systems* 70, pp. 118-127.

[17] Shlens, J. 2014. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

[18] He, Y., Qiu, M. Jin, M. and Xiong, T. 2012. Improvement on HITS Algorithm. *Applied Mathematics & Information Sciences* 6, no. 3, 1075-1085.

[19] Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)* 46, no. 5, 604-632.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 326-328, 2017, ISSN:-2319–8656

A Novel method of true random number generation using water laser interaction.

K.Elampari
Department of Physics
S.T.Hindu College
Nagercoil, India

B.Ramakrishnan
Department of Computer Science
S.T.Hindu College
Nagercoil, India

Abstract: This paper demonstrates the generation of true random number sequence using a simple laser scattering using water. A microcontroller based data acquisition system is used to capture the data and several statistical tests are performed over the generated data to explore the amount of randomness. The test statistic values reveal the possibility of generating a true random sequence of bits using this idea.

Keywords: laser; scattering; pseudo-random generator;true random number;cryptography;randtests; R

1. INTRODUCTION

Random number generation is an important and frequent requirement in various applications. Different applications require various degrees of randomness. Though the output of a pseudo-randomgenerator is deterministic it is sufficent for some applications. But in other situations, such as in cryptographic applications, generated bits must be truly random, otherwise the security of the entire application could be compromised.

In this paper we present a novel method of generating true random number sequence using the interaction between photons and water molecules. The shape of the water is decisive on how the light passes through it. Coming from an optically less dense medium (air) and entering a denser one (water), the light is partly reflected back while partly entering the water. Depending on the shape of the water, the light forms crinkle patterns or becomes diffused randomly in all directions. Also the reflected light is partly polarized (horizontally) and the part that enters the water is vertically polarized. As photons of light move through substances they don't simply pass through unaffected. Photons interacts with atoms and molecules comprising it. Photons carry discrete amounts of energy called quanta which can be transferred to atoms and molecules when photons are absorbed [1].

2. WATER LASER INTERACTION

There are six ways in which photons may interact with matter: Coherent Scattering, Photoelectric Effect, Incoherent Scattering, also known as Compton Scattering or Compton Effect, Pair Production, Triplet Production, Photodisintegration.These may cause the photon to attenuate (lose some of its energy and/or disappear). Coherent (or Rayleigh) scattering occurs at low photon energies. A photon may interact with an orbital electron and is then deflected (or scattered) at a small angle. There is no change in energy of the photon and no other effects occur. Incoherent scattering occurs when a photon has a much greater amount of energy than the binding energy of the electron, effectively considering the electron as 'free'. In this interaction, the photon interacts with the 'free' electron, giving up some of its energy and undergoing scattering. The electron receives the energy and is set in motion in a different direction. Photons may be scattered in any direction and is purely random [2]. Light transmission through a water sample is determined by physical properties such as particle size, shape, suspended solids concentration (SSC), and composition, temperature, and chemical properties such as the presence of nearinfrared (NIR) absorbing dissolved matter. There is enormous variation in these properties in the environment, resulting in a nearly infinite number of unique optical characteristics for water.

Light is an ensemble of photons that are absorbed and scattered by water, suspended particles, and dissolved matter as they travel through a sample. The absorption coefficient is a measure of the conversion of radiant energy to heat and chemical energy. It is numerically equal to the fraction of energy absorbed from a light beam per unit of distance traveled in an absorbing medium.

The angular distribution of light intensity scattered from a beam by a water sample is called the volume scattering function, VSF. The angle between this beam and scattered light rays is the scattering angle. Forward-scattered radiation occupies the hemisphere surrounding the incident beam and orientes away from the source and back scattered radiation fills the opposite hemisphere. Figure 1 shows VSFs computed from Mie theory for air bubbles, mineral grains, and biological material as well as the forward- (0° to 90°) and back-scattering (> 90°) VSF regions.

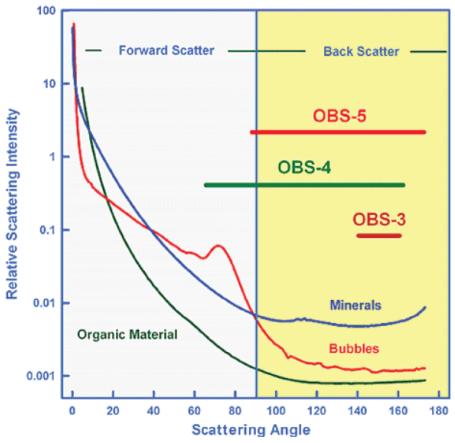

Figure 1. Graph shows VSFs computed from Mie theory

The VSF for bubbles is strongly peaked in the forward direction relative to the other materials and biological material back scatters 10 to 50% less light than the other particles [3]. There are generally two tendencies for the molecules of a material- the tendency to move randomly due to heat and the tendency to want to stick together because of electric charge between molecules. Technically, heat is the flow of thermal,

kinetic energy.The effect of heat on water depends on whether the energy is moving into the water (endothermic) or moving out of the water (exothermic). When water molecules gain energy, the average speed of the molecules increases; this is measured as a temperature change if enough energy is gained. Conversely, when energy leaves water molecules, the average speed of the molecule decreases, and a temperature decrease may be observed. Thus the tendency to move randomly is controlled by the temperature. Increase in temperature eventually increases the random movement and at the time of boililing all the water molecules spreads out as much as it can and get more random motion. Next, boiling is the most extreme form of evaporation as individual molecules happen to break away from the liquid through random movements.
Boiling is actually a very efficient heat transfer process. When the bottom of the container is much hotter than the boiling point of the water (i.e) when the boiling point is breached, tiny bubbles of water vapor are produced. The bubbles rise, due to buoyancy, and then collapse as they reach the denser, relatively cooler water at the surface. This motion not only helps to move the water around more quickly, but the bubbles themselves transfer heat energy as well. This bubble formation is called nucleate boiling and it is a far more effective way to transfer heat on its own than natural convection [4][5]

3. EXPERIMENTAL SETUP

The experimental setup consists of a laser source, nucleate boiling water, an array of high sensitive photo diodes as light sensors connected with Arduino Uno microcontroller circuit, and Parallax Data Acquisition tool (PLX-DAQ). When the laser beam is passed through the nucleate boiling water, the photons are scattered in different directions randomly due to the pure random movement of water molecules. The intensity variation of the outgoing beam in different directions is captured by the array of photo diodes. The sensor ouputs from the microcontroller are passed through the Parallax Data Acquisition tool (PLX-DAQ) to the PC for further processing and testing.

4. RANDOMNESS TEST

The desirable properties of random numbers are uniformity and independence. In order to test the amount of randomness present in the generated data, several randomness tests were performed.The algorithms of testing a random number generator are based on some statistics theory, i.e. testing the hypotheses. Although there is no true test to determine whether a sequence is random there are several widely accepted tests that we have utilized. Several of these tests are designed to test a specific null hypothesis. In this case, the hypothesis is that the bit-sequence under test is random. Each of the tests creates a test statistic, which is then used to calculate an associated p-value, which is related to the strength of the evidence against the null hypothesis. This p-value is a value on the interval [0,1], with a p-value of 1 denoting perfect randomness and a p-value of 0 denoting perfect nonrandomness. A significance level (α) is then chosen for the tests. If $p \geq \alpha$, then the null hypothesis is accepted; i.e., the sequence appears to be random. If $p < \alpha$, then the null hypothesis is rejected; i.e., the sequence appears to be nonrandom.Typically α is chosen to be 0.01 meaning that assuming the test is passed the sequence can be said to be random (or nonrandom) with a confidence of 99%.
According to various type of non randomness that may exist in Random bit sequences it is not practical to find non randomness patterns by just using one test. Most of the statistical tests are collection of tests. This collection is generally known as ‘suite’ or ‘battery’ of statistical tests. Some of the important tests carried out over the generated data in this study are

1. Ent Test
2. Bartel Rank Test
3. Cox Stuart Test
4. Turning Point Test
5. Runs Test

All these tests except ENT were performed by using “randtests” package for R [6]. The test results were summarized in the sections 4.1 to 4.5

4.1 The ENT test

Originally ENT test consists of six experiments which are Entropy test, Compression ratio test, Chi-square test, Arithmetic Mean test (AM), Monte Carlo value of PI and Serial Correlation Coefficient (SCC). Each test has a maximum score; Entropy test: 8.0, Compression ratio:0.0, Chi-square test:1.0, AM test:127.5 SCC:0.0, , Monte carlo value of PI: 3.1415926535 (upto 10 places).
The chi-square test is the most commonly used test for the randomness of data, and is extremely sensitive to errors in pseudorandom sequence generators. The chi-square distribution is calculated for the stream of bytes in the file and expressed as an absolute number and a percentage which indicates how frequently a truly random sequence would exceed the value calculated. It is interpreted that the percentage as the degree to which the sequence tested is suspected of being non-random. If the percentage is greater than 99% or less than 1%, the sequence is almost certainly not random. If the percentage is between 99% and 95% or between 1% and 5%, the sequence is suspect. Percentages between 90% and 95% and 5% and 10% indicate the sequence is “almost suspect” [7] .
The ENT test on the data gives, Entropy = 7.999980 bit per cycle, Optimum compression ratio: 0 percent. Chi square distribution for the samples is 294.60, and randomly would exceed this value 4.46 percent of the times , Arithmetic mean value is 127.5136. Monte carlo value for Pi is 3.1415852423 (error = 0.0001 percent) Serial correlation coefficient is -0.000086 (totally uncorrelated = 0.0).
Since the corrleation between a pair of independent random numbers or variables is 0, the auto corrleation function acf() is used to test the generated data. The acf equals 1 at lag 0, and is always between -1 and 1. The plot in figure 2 illustrates the lack of correlation and support randomness.

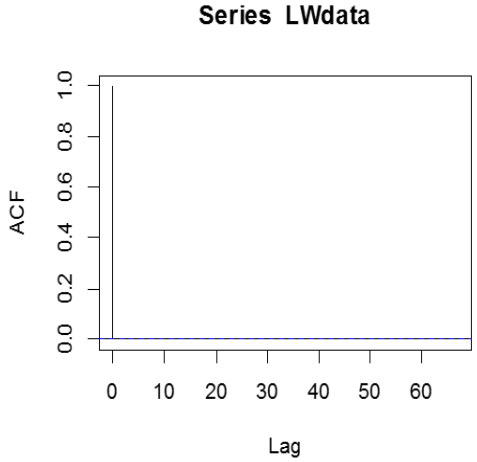

Figure 2. ACF of Generated Data

4.2 Bartels Rank Test

This is the rank version of von Neumann's Ratio Test for Randomness. Suppose that in the sequence of n measurements R_i is rank i of observation i_x . The hypothesis H_0 under test is rejected at large in modulus values of the statistics [8]. The value of the normalized test statistic for the generated random numbers is 2.0 and the p-value is 0.05 with the alternative hypothesis of non-randomness.

4.3 Cox Stuart test

Cox-Stuart nonparametric test can be used to verify the sequence of measurements to determine the presence of a trend in the mean, as well as in the variance. Data is grouped in pairs with the i^{th} observation of the first half paired with the i^{th} observation of the second half of the data. If the length of the data stream is odd the middle observation is eliminated [9]. The cox stuart test is then simply a sign test applied to these paired data [10]. The test statistic value (the number of pairs with a signal "+") is 1309300, n = 2621400 with a p-value of 0.09429, where n is the number of pairs, after eliminanting ties.

4.4 Turning Point Test

A turning point test is a statistical test of the independence of a series of random variables and the test is reasonable for a test against cyclicity [11][12]. The test statistic value is 1.1145 with p=0.2651. Since the the critical value ($\alpha = 0.01$) 2.33 is > test statistic value the alternate hypothesis of non-randomness is rejected.

4.5 Runs Test

The Runs Test performs the Wald-Wolfowitz test of randomness for continuous data [13]. A run is defined as a succession of similar events preceded and followed by a different event. The length of the run is the number of events that occurs in the run. An up run is a sequence of numbers each of which is succeeded by a larger number. Similarly, a down run is a sequence of numbers each of which is succeeded by a smaller number. The test statistic value for the data is 0.71887 < the critical value (p = 0.4722) and the alternative hypothesis is rejected.

5. RESULTS AND CONCLUSION

A set of five statistical tests were used to test the randomness of the generated data. Other than Bartel Rank test all tests statistic values reveal that there is an evidence for rejecting the alternate hypothesis of non-randomness. For further powerful testing of randomness of the sequence, test suits like DIEHARD and NIST may be used. This study reveals the possibility of generation of random bits based on the interaction of water and photons. By using proper optical components and well adujstable photon source it is possible to generate a true sequence of random bits.

6. REFERENCES

[1] Diemer, G.S., 2009. Quinta Essentia - Part 1- A practical guide to space- Time engineering, https://books.google.co.in/books?isbn=1409202720

[2] Photon Interactions OzRadOnc, http://ozradonc.wikidot.com/photoninteractions)

[3] John Downing, 2008. Effects of Light Absorption and Scattering in Water Samples on OBS Measurements, Campbell Scientific, Inc

[4] Tara Ruttley, 2011. T*he physical science of boiling in space.* https://blogs.nasa.gov/ISS_Science_Blog/2011/04/15/post_1301433765536/

[5] https://socratic.org/questions/how-does-heat-energy-affect-the-movement-of-water-molecules

[6] Frederico Caeiro and Ayana Mateus, 2014. randtests: Testing randomness in R. R package version 1.0. https://CRAN.R-project.org/package=randtests

[7] Walker, J. 1998. ENT Test suite, http://www.fourmilab.ch/random

[8] Bartels, R., 1982. The Rank Version of von Neumann's Ratio Test for Randomness, Journal of the American Statistical Association, 77(377), 40–46

[9] Cox, D. R., and Stuart, A. 1955. Some quick sign test for trend in location and dispersion, Biometrika, 42, 80-95.

[10] Sprent, P. and Smeeton, N.C. 2007. Applied Nonparametric Statistical Methods, 4th ed., Chapman and Hall/CRC Texts in Statistical Science.

[11] Brockwell, P.J, and Davis, R.A. 2002. Introduction to Time Series and Forecasting, 2nd edition, Springer (p. 36).

[12] Mateus, A. and Caeiro, F. 2013. Comparing several tests of randomness based on the difference of observations. In T. Simos, G. Psihoyios and Ch. Tsitouras (eds.), AIP Conf. Proc. 1558, 809–812

[13] Gibbons, J.D. and Chakraborti, S. 2003. Nonparametric Statistical Inference, 4th ed. (pp. 78–86). URL: http://books.google.pt/books?id=dPhtioXwI9cC&lpg=PA97&ots=ZGaQCmuEUq

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 329-332, 2017, ISSN:-2319–8656

Software Effort Estimation Using Adaptive Fuzzy-Neural Approach

Riyadh A.K. Mehdi
College of Information Technology
Ajman University

Abstract- Software effort estimation is an important step in software development. It predicts the amount of effort and development time required to build a software system. It is one of the most important tasks and an accurate estimate is vital to the successful completion of the project. Building software effort estimation requires developing sound computational models. This paper investigates the use of fuzzy-neural systems in estimating software effort. A comparison is made with a radial basis neural network. Results obtained based on the China dataset indicates that a hybrid model that combine fuzzy inferencing with neural networks ability to learn from examples provided more accurate results than using neural networks alone.

Index Terms - Software effort estimation; fuzzy inference; datasets; neural networks; and fuzzy-neural systems.

1. INTRODUCTION

Estimating software development efforts is one of the critical tasks in managing software development projects. Predicting software development effort with high accuracy is of paramount importance for project managers. However, estimating software development effort is still a challenging problem and one that attracts considerable research. Numerous software effort estimation models have been developed [1, 2, 3, 4]. The conventional models use a mathematical formula to predict project cost based on the estimates of parameters such as project size measured in lines of source code or function points, number of software engineers, and other process and product attributes [5]. Among the software cost estimation techniques, COCOMO (**Co**nstructive **Co**st M**o**del) is the most commonly used algorithmic cost modeling technique because of its simplicity for estimating the effort in person-months for a project at different stages. COCOMO uses a mathematical formula to predict project cost estimation [6]. Non-algorithmic models of software cost estimation based on soft computing approaches such as artificial neural networks (ANN) and fuzzy logic have also been used. Artificial neural networks are good at modeling complex nonlinear relationships. They are massive parallel-distributed processor made up of simple processing units, which can store experimental knowledge and making it available for use [5]. An ANN resembles the brain in two respects [5]: 1) Knowledge is acquired from its environment through a learning process, 2) Interneuron connection weights are used to store the knowledge. On the other hand, fuzzy logic is a mathematical tool for dealing with uncertainty and imprecision information. Fuzzy logic maps an input space to an output space through set of if then rules designed by a human expert in the domain [7]. Fuzzy logic models can be constructed without any data or with little data [8, 9]. This makes fuzzy logic superior over data-driven model building approaches such as neural network, regression and case based reasoning. In addition, fuzzy logic models can adapt to new environment when data become available [10]. Implementing fuzzy system requires that the distinct categories of the different inputs be represented by fuzzy sets which, in turn, are represented by membership functions. The domain of membership function is fixed, usually the set of real numbers, and whose range is the span of positive numbers in the closed interval [0, 1].

2. LITERATURE REVIEW

Xu and Khoshgoftaar [11] proposed a fuzzy identification cost estimation model to deal with linguistic data, and automatically generate fuzzy membership functions and rules. Azzeh et al. [12] propose an analogy-based software effort estimation using fuzzy numbers, namely Generalized Fuzzy Number Software Estimation. They compute the similarity between two generalized fuzzy numbers based on their geometric distances, center of gravities and height of the generalized fuzzy numbers, and use fuzzy c-means to cluster the existing software project data. The estimations are conducted with the use of generalized fuzzy number operations and the effort of a project is estimated as a fuzzy number which is defuzzified with the method of center of gravity. Lopez-Martin et al. [13] compare three personal fuzzy logic models to estimate the effort of small software programs, namely triangular, trapezoidal and Gaussian membership functions, with linear regression model. They develop the fuzzy logic and linear regression models using the data gathered from 105 small programs, and then the estimations generated by these models are compared with each other using 20 small programs. Wei Lin et al. [14] showed that a general neuro-fuzzy framework can function with various algorithmic models for improving the performance of software effort estimation. They used a Neuro-Fuzzy model to demonstrate that combining the neuro-fuzzy model with the SEER-SEM effort estimation model produces unique characteristics and performance improvements. They concluded that the neuro-fuzzy features of the model provided their neuro-fuzzy SEER-SEM model with the advantages of strong adaptability with the capability of learning, less sensitivity for imprecise and uncertain inputs, easy to be understood and implemented, strong knowledge integration, and high transparency. Hodgkinson and Garratt [15] introduced the neuro-fuzzy model for cost estimation as one of the important methodologies for developing non-algorithmic models. Their model did not use any of the existing prediction models, as the inputs are size and duration, and the output is the estimated project effort.

Huang et al. [16, 17] proposed a software effort estimation model that combines a neuro-fuzzy framework with COCOMO II. The

parameter values of COCOMO II were calibrated by the neuro-fuzzy technique in order to improve its prediction accuracy. This study demonstrated that the neuro-fuzzy technique was capable of integrating numerical data and expert knowledge.

Xia et al. [18] developed a Function Point (FP) calibration model with the neuro-fuzzy technique, which is known as the Neuro-Fuzzy Function Point (NFFP) model. The objectives of this model are to improve the FP complexity weight systems by fuzzy logic, to calibrate the weight values of the unadjusted FP through the neural network, and to produce a calibrated FP count for more accurate measurements.

Wong et al. [19] introduced a combination of neural networks and fuzzy logic to improve the accuracy of backfiring size estimates. In this case, the neuro-fuzzy approach was used to calibrate the conversion ratios with the objective of reducing the margin of error. The study compared the calibrated prediction model against the default conversion ratios. As a result, the calibrated ratios still presented the inverse curve relationship between the programming languages level and the number of function points, and the accuracy of the size estimation experienced a small degree of improvement. A survey on software effort estimation techniques is given in [20].

The Adaptive network based fuzzy inference system (ANFIS) is a hybrid of a feed forward neural network and a fuzzy inference system. The neural network uses either a pure back propagation gradient descent learning rule, or a hybrid learning rule that uses back propagation and a least squares method [21]. The fuzzy logic component takes into account the imprecision and uncertainty of the system that is being modelled while the neural network component apply its learning algorithm to tune the membership functions of the fuzzy inference system generated [22]. Using this hybrid method, at first an initial fuzzy model along with its input variables are derived with the help of the rules extracted from the input output data of the system that is being modeled. Next the neural network is used to fine tune the rules of the initial fuzzy model to produce the final ANFIS model of the system [22]. In ANFIS the parameters can be estimated in such a way that both the Sugeno and Tsukamoto fuzzy models [23] are represented by the ANFIS architecture.

This paper investigates the effectiveness of using a neuro-fuzzy approach to software effort estimate and how it compares to other approaches.

3. RESEARCH METHODOLOGY

In this work, an adaptive neuro-fuzzy inference system based on the Sugeno fuzzy model is used. The following exposition is adapted from [22].

3.1 ANFIS Architecture

A typical architecture of ANFIS is depicted in Figure 1. A circle indicates a fixed node, whereas a square indicates an adaptive node [22].

For a first-order Sugeno fuzzy model, a two rules rule base can be expressed as follows:

1. If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$
2. If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

Let the membership functions of fuzzy sets A_i, B_i for i=1,2 be μ_{A_i}, μ_{B_j}. In this work, Gaussian membership functions are used,

$$\mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x - c_i}{a_i} \right)^{2b_i}}$$

In evaluating the rules, a product T-norm (logical and) is chosen. Evaluating the rule premises results in,

$$w_i = \mu_{A_i}(x)\mu_{B_i}(y), \text{ i} = 1,2.$$

Evaluating the implication and the rule consequences gives,

$$f(x,y) = \frac{w_1(x,y) f_1(x,y) + w_2(x,y) f_2(x,y)}{w_1(x,y) + w_2(x,y)}.$$

Leaving the arguments out,

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$$

The above equation can be rewritten as,

$$f = \overline{w_1} f_1 + \overline{w_2} f_2, \text{ where}$$

$$\overline{w_i} = \frac{w_i}{w_1 + w_2}$$

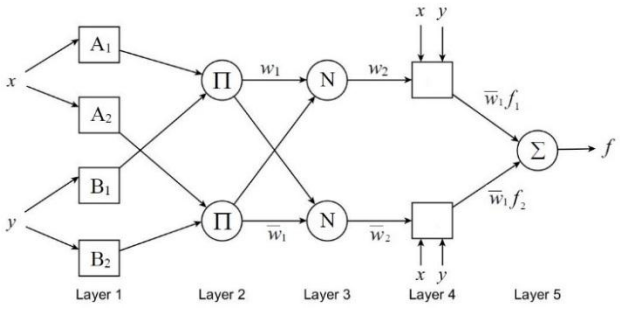

Figure 1. Structure of adaptive neuro-fuzzy inference system.

3.2 MATLAB Implementation

1. *genfis2*(*Xin*,*Xout*,*radii*), gen*fis2* generates a Sugeno-type FIS structure using subtractive clustering and requires separate sets of input and output data as input arguments. When there is only one output, *genfis2* may be used to generate an initial FIS for *anfis* training. *genfis2* accomplishes this by extracting a set of rules that models the data behavior. The rule extraction method first uses the subclust function to determine the number of rules and antecedent membership functions and then uses linear least squares estimation to determine each rule's consequent equations. This function returns an FIS structure that contains a set of fuzzy rules to cover the feature space. The arguments for genfis2 are as follows:
   a. *Xin* is a matrix in which each row contains the input values of a data point.
   b. *Xout* is a matrix in which each row contains the output values of a data point.
   c. *radii* is a vector that specifies a cluster center's range of influence in each of the data dimensions, assuming the data falls within a unit hyper box.
2. *anfis*(*trainingData*, *options*)

This function generates a single-output Sugeno fuzzy inference

system (FIS) structure using grid partitioning and tunes the system parameters using the specified input/output training data to adjust the membership functions parameters. This adjustment is made using a backpropagation algorithm either alone, or in combination with a least squares type of method. This allows the fuzzy systems to learn from the data they are modeling. The MATLAB statement: [*fis*,*trainError*,*stepSize*,*chkFIS*,*chkError*] = *anfis*(*trainingData*,*options*) returns the validation data error for each training epoch, *chkError*, and the tuned FIS structure for which the validation error is minimum, *chkFIS*. Using validation data prevents overfitting to training data. To use this syntax, we must specify validation data using *options.ValidationData*.

3.3 Software Effort Estimation Datasets

The China dataset (19 attribute, 499 projects, effort in person-hours) is used in this work. The number of records used for training, checking, and testing were 349, 100, and 50 respectively. Because the number of records is inadequate to estimate the parameters of the neuro-fuzzy estimation model, It was not possible to use the datasets below [24]:

- Desharnais (11 attribute, 77 project, effort in person-hours)
- Cocomo81 (18 attribute, 61 project, effort in person-month)
- Maxwell (27 attribute, 62 project, effort in function points)
- Albrecht (8 attribute, 24 project, effort in person-hours)

The evaluation criterion used to assess the estimation accuracy are root mean square error (RMSE), and the mean magnitude (absolute) error (MME) [25]:

$$RMSE = sqrt(\sum_{i=1}^{i=n}(actualEffort(p_i) - predictedEffort(p_i))^2 / n)$$

$$MME = \frac{1}{n}\sum_{i=1}^{i=n}|actualEffort(p_i) - predictedEffort(p_i)|$$

4. RESULTS

Figure 2 depicts the performance of the neuro-fuzzy model when run on the testing data (50 records). In Figure 2, the line represent actual effort and the circles represent estimated efforts.

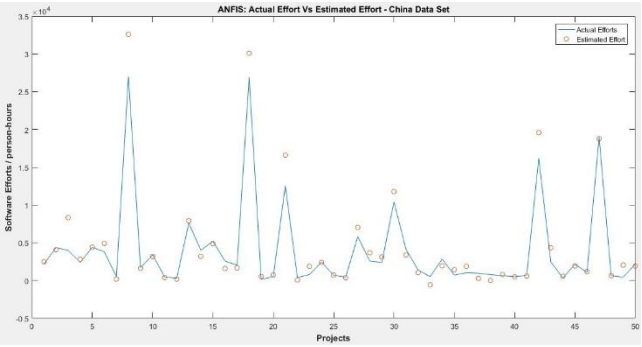

Figure 2. Actual efforts and estimated efforts using ANFIS.

As a comparison, Figure 3 shows the performance of a RBFN network using the same set of testing records.

Table 1 shows the mean absolute error, and root mean square root for the China dataset using the adaptive neuro-fuzzy model and a radial basis function neural network (RBFN) [26].

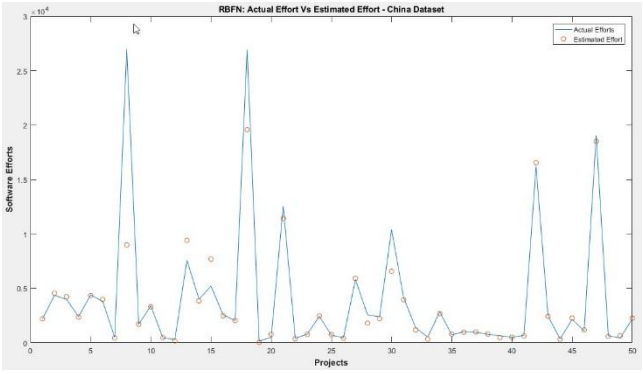

Figure 3. Actual efforts and estimated efforts using RBFN.

Table 1. Errors obtained using ANFIS and RBFN on the China dataset.

Error Type	ANFIS	RBFN
Root mean square error	1483	2850
Mean magnitude error	874	802

The results indicate that although the mean magnitude error for the ANFIS model is slightly higher than that for the RBFN model, they are comparable. However, the root mean square error for the RBFN is almost twice that of the ANFIS model. This indicate while the average error is almost the same for the two models, the estimation of the ANFIS model has much less deviations from the actual values compared with the RBFN model.

5. CONCLUSIONS

In this paper, the adaptive neuro fuzzy model was applied to the problem of estimating software developments efforts using the China data set. The results were compared with using a radial basis function neural network on the same data set and on the same testing records. The results indicate while both models are comparable with regard to the mean magnitude error, the ANFIS model has a better performance in the sense that the estimates have a far less deviation that those of the RBFN model.

Future work will investigate the relevance of the various attributes in determining the size of efforts required in developing software so that standardized set of attributes can be used in collecting data sets.

6. REFERENCES

[1]. S. A. Abbas, A. R. Liao, A. Azam, 2012. "Cost Estimation: A Survey of Well-Known Historic Cost Estimation Techniques," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 4, no. 1, pp 612-636.

[2]. R. Malhotra, A. Jain, 2011. "Software Effort Prediction using Statistical and Machine Learning Methods," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 1.

[3]. Abbas Heiat, 2002. "Comparison of Artificial Neural Network and Regression Models for Estimating Software Development Effort," *Information and Software Technology*, vol. 44, pp. 911-922.

[4]. R. Sharma, 2013. "Survey: Non Algorithmic Models for Estimating Effort," *European International Journal of Science and Technology*, vol. 2, no. 3.

[5]. Kaushik, A. Chauhan, D. Mittal, and S. Gupta, 2012. "COCOMO Estimates Using Neural Networks," International *Journal of Intelligent Systems and Applications*, vol. 9, pp. 22-28.

[6]. Reddy, and K. Raju, "An Optimal Neural Network Model for Software Effort Estimation," *International Journal of Software Engineering*, vol.12 no.1, pp. 66-78.

[7]. Zadeh, L.A., 2002. From Computing with numbers to computing with words-from manipulation of measurements to manipulation of perceptions. International Journal of Applied Mathematics and Computer Science 12(3), 307-324.

[8]. MacDonell, S.G., Gray, A.R., Calvert, J.M. 1999. Fuzzy Logic for Software Metric Practitioners and Researchers.In : The Proceedings of the 6th International Conference on Neural Information Processing ICONIP, Perth, pp. 308-313.

[9]. Ryder, J. 1998. Fuzzy Modeling of Software Effort Prediction. In: Proceeding of IEEE Information Technology Conference, Syracuse, New York, pp. 53-56.

[10]. Sailu, M.O., Ahmed, M., AlGhamdi, J. 2004. Towards Adaptive Soft Computing Based Software Effort Prediction. In: Fuzzy Information, Processing NAFIPS, pp, 16-21.

[11]. Xu, Z., Khoshgoftaar, T.M., 2004. "Identification of fuzzy models of software cost estimation," Fuzzy Sets and Systems, vol. 145 (1), pp. 141-163.

[12]. Azzeh, M., Neagu, D. Cowling, P.I., 2011. "Analogy-based software effort estimation using Fuzzy numbers," Journal of Systems and Software, vol. 84 (2), pp. 270-284.

[13]. Lopez-Martin, C., Yanez-Marquez, C., Gutierrez-Tornes, A., 2008. "Predictive accuracy comarison of fuzzy models for software development effort of small programs," Journal of Systems and Software, vol. 81 (6), pp. 949-960.

[14]. Wei Lin Du, Danny Ho, Luiz Fernando Capretz, 2010. Improving Software Effort Estimation Using Neuro-Fuzzy Model with SEER-SEM, Global Journal of Computer Science and Technology, Vol. 10 Issue 12, pp. 51-63.

[15]. Hodgkinson, A. C. and Garratt, P. W. 1999. A NeuroFuzzy Cost Estimator. Proc. 3rd Int Conf Software Engineering and Applications (SAE): 401–406

[16]. Huang, X., Ho, D., Ren, J., and Capretz, L. F. 2005. A Soft Computing Framework for Software Effort Estimation. Soft Computing: 170–177

[17]. Huang, X., Ho, D., Ren, J., and Capretz, L. F. 2006. Improving the COCOMO Model Using A Neuro-Fuzzy Approach. Applied Soft Computing: 29–40

[18]. Xia, W., Capretz, L. F., Ho, D., and Ahmed, F. 2008. A New Calibration for Function Point Complexity Weights. International and Software Technology, Vol. 50, Issue 7-8: 670–683

[19]. Wong, J., Ho, D., and Capretz, L. F. 2008. Calibrating Functional Point Backfiring Conversion Ratios Using Neuro-Fuzzy Technique. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 16, No. 6: 847 – 862

[20]. Karunakaran, and Sreenath, 2015. Survey on Software Effort Estimation Technique – A Review International Journal of Scientific & Engineering Research, Volume 6, Issue 12.

[21]. J.S. Jang, 1999. "Anfis: Adaptive-network-based fuzzy inference system," IEEE Trans. Syst., Man, Cybern., vol. 23, pp. 665--685, Mar. 1993 [2] W. H. Wolberg and O.L. Mangasarian: "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proceedings of the National Academy of Sciences, U.S.A., Volume 87, pp 9193-9196.

[22]. M. Buragohain, 2008. Adaptive Network based Fuzzy Inference System (ANFIS) as a Tool for System Identification with Special Emphasis on Training Data Minimization, PhD Thesis, Indian Institute of Technology Guwahati, India.

[23]. Y. Tsukamoto, 1979. M. M. Gupta, R. K. Ragade, and R. R. Yager, "An approach to fuzzy reasoning method," in Advances in Fuzzy Set Theory and Application, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds., North-Holland, Amsterdam, pp. 137–149.

[24]. S. J. Shirabad, and T. J. Menzies, 2005. "The PROMISE Repository of Software Engineering Databases," School of Information Technology and Engineering, University of Ottawa, Canada. Available: http://promise.site.uottawa.ca/SERepository.

[25]. S. Conte, 1986. H. Dunsmore, and V. Shen, Software Engineering, Metrics and Models. Benjamin/Cummings.

[26]. Riyadh A.K. Mehdi, 2015. "Software Defect Prediction Using Radial Basis and Probabilistic Neural Networks", International Journal of Computer Applications and Research, Volume 5, Issue 5, pp. 260-265.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 333-337, 2017, ISSN:-2319–8656

Deep Learning for Intelligent Exploration of Image Details

Okanti Apoorva
Department of CSE
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,India

Y.Mohan Sainath
Department of CSE
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India

G.Mallikarjuna Rao
Department of CSE
Professor
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India

Abstract: Automatic image captioning is the task where given an image the system must generate a caption that describes the contents of the image. Once you can detect objects in photographs and generate labels for those objects, you can see that the next step is to turn those labels into a coherent sentence description. Most of the approaches involve the use of very large convolution neural networks (CNN) for the object detection in the photographs and then a recurrent neural network (RNN) like an LSTM (Long short-term memory) to turn the labels into a coherent sentence. In our proposed approach we have tailored the CNN and LSTM and has been tested with CIFAR 10 and MNIST datasets. The experimentation resulted 94.67% accuracy with 25 random iterations.

Keywords Deep Learning, Convolutional Neural Network, Long Short Term Memory, Recurrent Neural Networks, CIFAR10, MNIST.

1. INTRODUCTION

Automatic Image Caption Generation:

Automatically generating captions of an image is a task very close to the heart of scene understanding, is one of the primary goals of computer vision. Caption generation models must be powerful enough to solve the computer vision challenges of determining which objects are in an image, and also be capable enough to capture and express their relationships in a natural language. Due to this caption generation has long been viewed as a difficult problem. It poses considerable challenge for machine learning algorithms, as it amounts to mimicking the remarkable human ability to compress huge amounts of salient visual information into descriptive language.

Background:

Recently, several methods have been proposed for generating image descriptions. Many of these methods are based on recurrent neural networks and inspired by the successful use of sequence to sequence training with neural networks for machine translation. One major reason image caption generation is well suited to the encoder-decoder framework of machine translation is because it is analogous to "translating" an image to a sentence. Generating automatic descriptions from images requires an understanding of how humans describe images. An image description can be analyzed in several different dimensions. We assume that the descriptions that are of interest for this survey article are the ones that verbalize visual and conceptual information depicted in the image, i.e., descriptions that refer to the depicted entities, their attributes and relations, and the actions they are involved in. Outside the scope of automatic image description are non-visual descriptions, which give background information or refer to objects not depicted in the image (e.g., the location at which the image was taken or who took the picture). Also, not relevant for standard approaches to image description are perceptual descriptions, which capture the global low-level visual characteristics of images (e.g., the dominant color in the image or the type of the media such as photograph, drawing, animation, etc.).[1][6]

The general approach of the studies in this group is to first predict the most likely meaning of a given image by analyzing its visual content, and then generate a sentence reflecting this meaning. All models in this category achieve this using the following general pipeline architecture: 1. Computer vision techniques are applied to classify the scene type, to detect the objects present in the image, to predict their attributes and the relationships that hold between them, and to recognize the actions taking place. 2. This is followed by a generation phase that turns the detector outputs into words or phrases. These are then combined to produce a natural language description of the image, using techniques from natural language generation (e.g., templates, n-grams, grammar rules) [5].

The approaches reviewed in this section perform an explicit mapping from images to descriptions. Explicit pipeline architecture, while tailored to the problem at hand, constrains the generated descriptions, as it relies on a predefined set of semantic classes of scenes, objects, attributes, and actions.

Moreover, such architecture crucially assumes the accuracy of the detectors for each semantic class, an assumption that is not always met in practice.

Big Data is essentially a special application of data science, in which the data sets are enormous and require overcoming logistical challenges to deal with them. The primary concern is efficiently capturing, storing, extracting, processing, and analyzing information from these enormous data sets.

Processing and analysis of these huge data sets is often not feasible or achievable due to physical and/or computational constraints. Special techniques and tools (e.g., software, algorithms, parallel programming, etc.) are therefore required. Big Data is the term that is used to encompass these large data sets, specialized techniques, and customized tools. It is often applied to large data sets in order to perform general data analysis and find trends, or to create predictive models. A primary component of big data is the so-called three Vs (3Vs) model. This model represents the characteristics and

challenges of big data as dealing with volume, variety, and velocity. Companies such as IBM include a fourth "V", veracity.

1.1 Deep Learning

Deep learning is a type of machine learning that trains a computer to perform human-like tasks, such as recognizing speech, identifying images or making predictions. Instead of organizing data to run through predefined equations, deep learning sets up basic parameters about the data and trains the computer to learn on its own by recognizing patterns using many layers of processing. Deep learning methods have ability to continuously improve and adapt to changes in the underlying information pattern, presents a great opportunity to introduce more dynamic behavior into analytics. When put in other terms the deep learning can also be defined as the study of artificial neural networks and related machine learning algorithms which contain more than one hidden layer.[7][8][17]

1.2 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multi-layer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and weights' followed by some form of pooling which results in translation invariant features. Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. The convolutional neural network is also known as shift invariant or space invariant artificial neural network (SIANN), which is named based on its shared weights architecture and translation invariance characteristics. Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters. The CNN also has various combinations of convolutional layers and fully connected layers. A point wise non linearity is applied after every layer or at the end of each layer. To improve generalization and to reduce the number of free parameters, a convolution operation on small regions of input is introduced. [9][10][11]

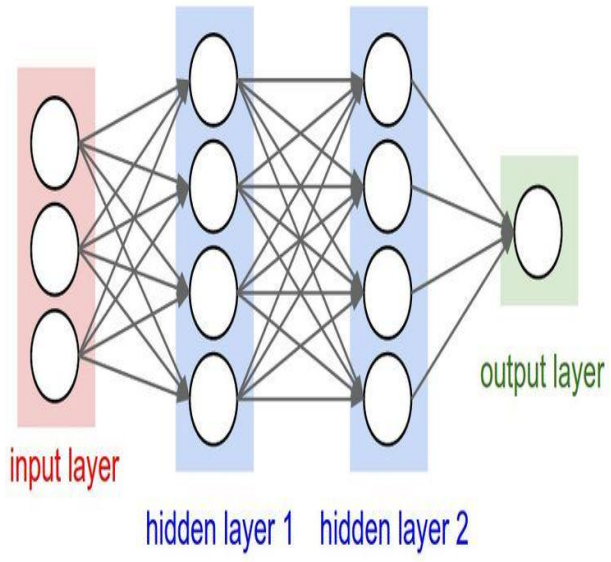

Figure 1. A Simple Convolutional Neural Network.

1.3 Long Short Term Memory

Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They work tremendously well on a large variety of problems, and are now widely used. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behaviour, not something they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of neural network. This design is typical with "deep" multi-layered neural networks, and facilitates implementations with parallel hardware. [12][18]

LSTM blocks contain three or four "gates" that they use to control the flow of information into or out of their memory. These gates are implemented using the logistic function to compute a value between 0 and 1.

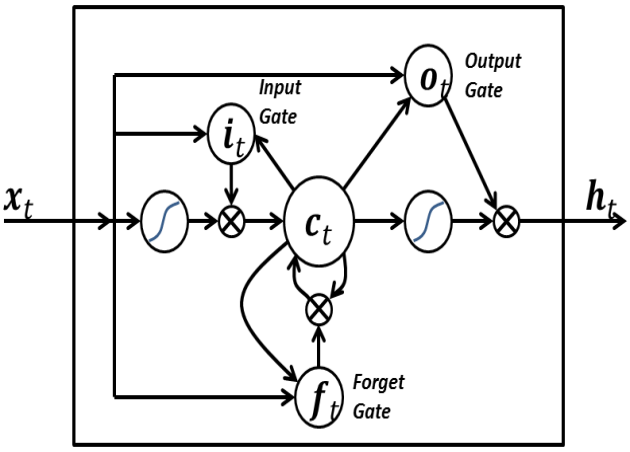

Figure 2: LSTM Gates

2. PROPOSED SYSTEM

In this image analysis we use convolutional neural networks and long short term memory for character recognition and image caption generation. Generating automatic descriptions from images requires an understanding of how humans describe images. An image description can be analyzed in several different dimensions. We assume that the descriptions that are of interest for this survey article are the ones that verbalize visual and conceptual information depicted in the image, i.e., descriptions that refer to the depicted entities, their attributes and relations, and the actions they are involved in.

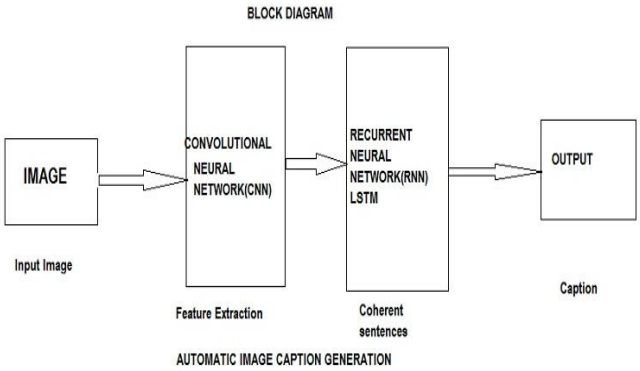

Figure 3: Block Diagram of Image Caption Generation

Outside the scope of automatic image description are non-visual descriptions, which give background information or refer to objects not depicted in the image (e.g., the location at which the image was taken or who took the picture). Also, not relevant for standard approaches to image description are perceptual descriptions, which capture the global low-level visual characteristics of images (e.g., the dominant color in the image or the type of the media such as photograph, drawing, animation, etc.).

CNN Architecture (ConvNet architectures)

Convolutional Layer, Pooling Layer and Fully-Connected Layer (exactly as seen in regular Neural Networks). We will stack these layers to form a full ConvNet architecture.

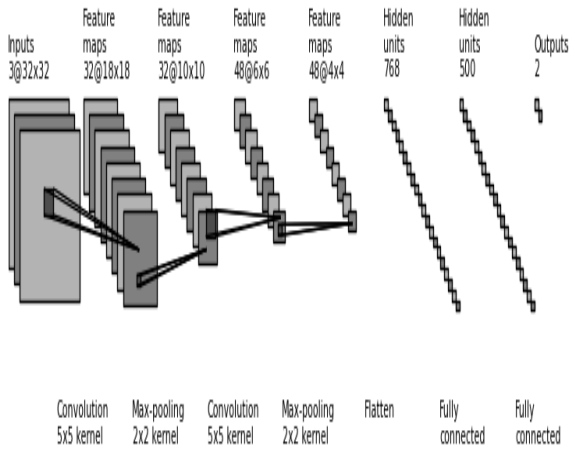

Figure 4: CNN for Image Caption Generation

Input [32x32x3] will hold the raw pixel values of the image, in this case an image of width 32, height 32, and with three color channels R, G, B. Convolutional layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. This may result in volume such as [32x32x12] if we decided to use 12 filters. Relu layer will apply an element wise activation function, such as the max (0, x) thresholding at zero. This leaves the size of the volume unchanged ([32x32x12]). Pool layer will perform a down sampling operation along the spatial dimensions (width, height), resulting in volume such as [16x16x12]. FC (i.e. fully-connected) layer will compute the class scores, resulting in volume of size [1x1x10], where each of the 10 numbers correspond to a class score, such as among the 10 categories of cifar-10. As with ordinary neural networks and as the name implies, each neuron in this layer will be connected to all the numbers in the previous volume, [2][3].

LSTM Architecture:

The core of the LSTM model is a memory cell c encoding knowledge at every time step of what inputs have been observed up to this step. The behavior of the cell is controlled by "gates" – layers which are applied multiplicatively and thus can either keep a value from the gated layer if the gate is 1 or zero this value if the gate is 0. In particular, three gates are being used which control whether to forget the current cell value (forget gate f), if it should read its input (input gate i) and whether to output the new cell value (output gate o). The definition of the gates and cell update and output are as follows

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1})$$
$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1})$$
$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1})$$
$$c_t = f_t * c_{t-1} + i_t * h(W_{cx}x_t + W_{cm}m_{t-1})$$
$$m_t = o_t * c_t$$
$$p_{t+1} = Softmax(m_t)$$

Where * represents the product with a gate value, and the various W matrices are trained parameters. Such multiplicative gates make it possible to train the LSTM robustly as these gates deal well with exploding and vanishing gradients. The nonlinearities are sigmoid σ (·) and hyperbolic tangent h (·). The last equation m_t is what is used to feed to a Softmax, which will produce a probability distribution p_t over all words [2][4].

3. EXPERIMENTATION

Python is a widely used high-level programming language for general-purpose programming, created by Guido van Rossum and first released in 1991. An interpreted language, Python has a design philosophy which emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly braces or keywords), and a syntax which allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale. To install any package in python, we use a pip command [13][14].

Syntax: Pip install <package- name>

To implement our project, we have used some libraries as Keras and Tensor flow

3.1 Keras Library

Keras is a high-level neural networks API, written in Python and capable of running on top of either TensorFlow or Theano and Runs seamlessly on CPU and GPU. It was developed with a focus on enabling fast experimentation [15].

3.2 Tensor flow

TensorFlow is Google Brain's second generation machine learning system, released as open source software on November 9, 2015. Among the applications for which TensorFlow is the foundation, are automated image captioning software, such as Deep Dream [16].

3.3 Dataset

The databases used in this image analysis are MNIST database and CIFAR10 database. The MNIST database is used in hand written character recognition and CIFAR 10 is used in image caption generation.

MNIST:

MNIST dataset (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image

processing systems. The MNIST database contains 60,000 training images and 10,000 testing images. Half of the training set and half of the test set were taken from NIST's training dataset, while the other half of the training set and the other half of the test set were taken from NIST's testing dataset. The different machine learning methods are used on the dataset and different error rates are found. We have implemented the above LSTM with MNIST dataset. To have reasonable convergence we limited the epochs to less than 10.

CIFAR 10:

The CIFAR-10 dataset consists of 60000, 32x32, colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The dataset is divided into five training batches and one test batch, each with 10000 images. The test batch contains exactly 1000 randomly-selected images from each class. The training batches contain the remaining images in random order, but some training batches may contain more images from one class than another. Between them, the training batches contain exactly 5000 images from each class. The classes are completely mutually exclusive. There is no overlap between various classes.

4. EXPERIMENTAL RESULTS:

4.1 Automatic Caption Generation

Automatically generating captions of an image is a task very close to the heart of scene understanding — one of the primary goals of computer vision. Not only must caption generation models be powerful enough to solve the computer vision challenges of determining which objects are in an image, but they must also be capable of capturing and expressing their relationships in a natural language. For this reason, caption generation has long been viewed as a difficult problem. It is a very important challenge for machine learning algorithms, as it amounts to mimicking the remarkable human ability to compress huge amounts of salient visual information into descriptive language [5].

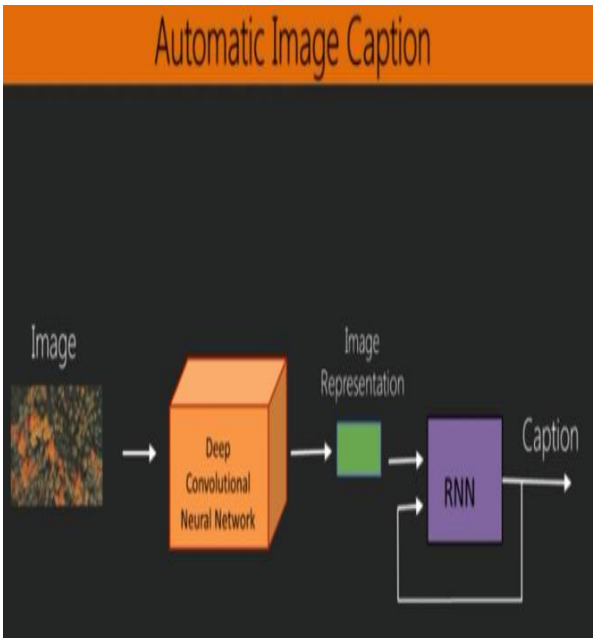

Figure 5: Architecture of Image Caption Generation

4.2 Character Recognition Experiment

In character recognition we use MNISTdataset .Here the data is processed by the convolutional neural networks. The neural network takes the input from the MNIST dataset trains and tests them and gives their accuracy values after being processed through the multiple number of hidden layers.

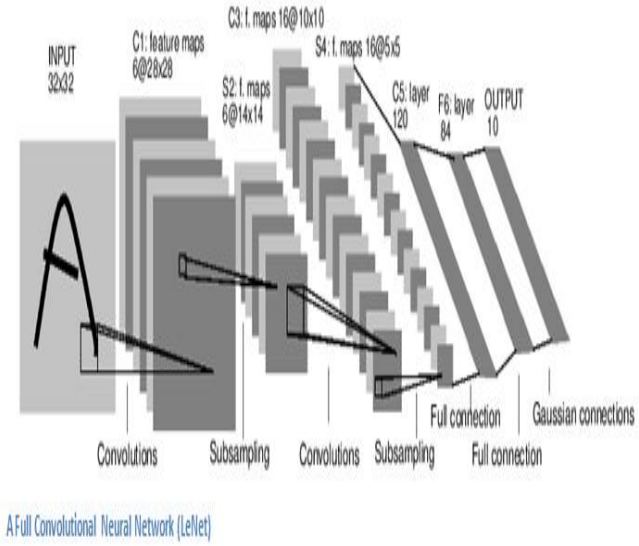

Figure 6: CNN for Character Recognition

5. CONCLUSION:

From the experimentation it can be seen how the learned attention can be exploited to give more interpretability into the models generation process, and demonstrate that the learned alignments correspond very well to human intuition.

The following gives sample snapshots of our experimentation. From the table we can conclude that as the epochs are increased recognition accuracy is increased (for 10 epoch's accuracy of recognition increased to 100%)

Table 1: Results obtained using the proposed model for MNIST & CIFAR10 dataset

Research highlights:

Number of Epochs	Recognition Accuracy (%)	
	MNIST Dataset	CIFAR10 Dataset
2	60	65
4	72	77
6	75	79.5
8	80	82
9	97	100
10	100	100
20	100	100

(i) The paper proposes Deep learning approach for the exploration of Image Details.

(ii) Convolutional Neural Networks and LSTM are used to evaluate performance.

(iii) Different Data sets MNIST and CIFAR 10 are used.

(iv) Multilayer approach has given true recognition rate above >90%.

TRAINING CIFAR 10 DATASET

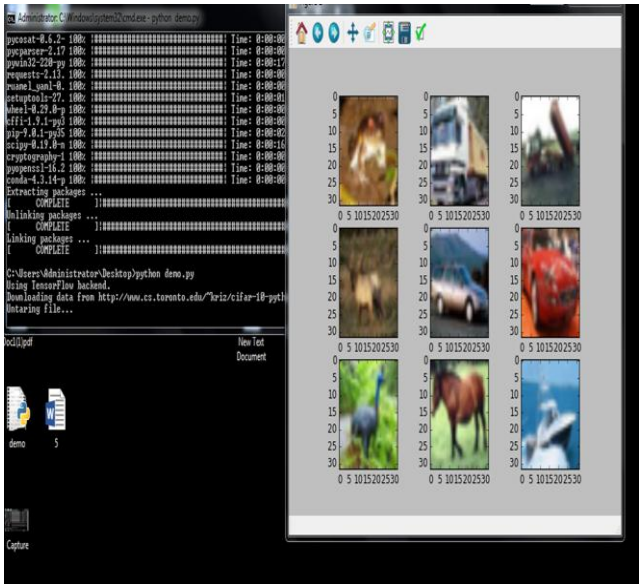

Figure 7: Experimentation of the image caption generation

Acknowledgement: We are thankful to our beloved director P.S Raju, Principal J.N Murthy for their kind support. We would like to thank coordinator of Parallel programming and operating systems lab (MODROB) for allowing the resources to carry our experimentation in the lab.

6. REFERENCES:

[1]. Automatic Description Generation from Images: A Survey of Models, Datasets, and Evaluation Measures by Raffaella Bernardi , Ruket Cakici , Desmond Elliott.
[2]. Learning CNN-LSTM Architectures for Image Caption Generation by Moses Soh.
[3]. Andrej Karpathy, Fei-Fei Li: Automated Image Captioning with ConvNets and Recurrent Nets.
[4]. Sepp Hochreiter and Jurgen Schmidhuber: Long Short Term Memory.
[5]. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.
[6]. Karol Gregor, Ivo Danihelka, Alex Graves, and Daan Wierstra.DRAW: A recurrent neural network for image generation.

[7]. Deep Learning: by Ian Goodfellow and Yoshua Bengio and Aaron Courville.

[8]. Deep Learning: Methods and Applications by Li Deng, Dong Yu

[9]. Neural Networks and Deep Learning by Michael Nielsen.

[10]. Introduction to Machine Learning: Smola and Vishwanathan.

Websites:
[11].https://en.wikipedia.org/wiki/Convolutional_neural_network
[12].https://en.wikipedia.org/wiki/Long_short-term_memory
[13].https://www.tutorialspoint.com/python/
[14].https://en.wikipedia.org/wiki/Python_(programming_language)
[15]. https://en.wikipedia.org/wiki/Keras
[16]. https://en.wikipedia.org/wiki/TensorFlow
[17]. https://en.wikipedia.org/wiki/deeplearning
[18]. https://en.wikipedia.org/wiki/recurrent neural network

Authors:

Okanti Apoorva is pursuing master degree in computer science in Gokaraju Rangaraju Institute of Engineering and Technology. Her research interests are Neural Networks and Image Processing.

G.Mallikarjuna Rao is working as professor in CSE department of Gokaraju Rangaraju Institute of Engineering and Technology. His research areas are Machine Vision, Parallel Computing and Neural networks.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 7, 338-343, 2017, ISSN:-2319–8656

Spectral Estimation of Soil Moisture Content using Semi Empirical Soil Model in the 0.4-2.5 um Domain

Saima Ansari[1]
Department of CS and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India

Ashwini Dilip Padmanabhi[2]
Department of CS and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India

Dr. Ratnadeep R. Deshmukh[3]
Department of CS and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India

Abstract: Soil moisture content (SMC) plays a key role in the crop production as it act as a nutrient and serves as solvent for other nutrients such as sodium, potassium, carbon, nitrogen. Soil characteristics can be analyzed using spectral reflectance of the soil in the 0.4-2.5 um domain. 25 soil samples of different textures have been collected from four different areas of Aurangabad city, Maharashtra. Soil spectral reflectances were measured in the laboratory with an ASD (Analytical Spectral Devices) Fieldspec-Pro spectroradiometer in the (0.4–2.5 μm) spectral domain. Dataset consists of 150 spectral refelctances (six SMC level i.e. 0%, 12%, 16%, 21%, 26% and 30% for each sample). Existing spectral indices like Normalized Soil Moisture Index (NSMI) and Water Index SOIL (WISOIL) are sensitive to water vapor absorption. Our study has been focused on to overcome existing spectral indices limitation. To accomplish this, new spectral index i.e. Normalized Index of NSWIR domain for Smc estimatiOn from Linear regression (NINSOL) has been implemented on the dataset. Semi empirical soil model has been used to estimate SMC as it is robust against soil sample texture. From the results, it is observed that NINSOL operates in 2056 nm and 2263 nm spectral range. Performance comparison has been done among NSMI ,WISOIL and NINSOL. NSMI and WISOIL led to R^2 value as 0.79, 0.81 respectively and 6.6, 6.2 RMSE value. NINSOL produces better results than existing indices as R^2 value of 0.85 and 5.6 RMSE value.

Keywords: Soil moisture content, spectral reflectance, spectral indices, semi empirical soil model, NINSOL, regression.

1. INTRODUCTION

SMC is most important nutrient present in the soil. It makes a significant impact on plant growth, percolation, evaporation, microbiological decomposition of the soil organic matter and also on heat exchange. Usually, soil moisture is affected by soil physical properties such as soil color, soil texture features [1,2], structure and bulk density. It is significant parameter for several applications in hydrology, horticulture, agriculture and meteorology. It influences plant growth, percolation, evaporation and heat exchange. Description of SMC is very useful for many agricultural applications like irrigation system, plant stress and improving crop yield. Remote sensing techniques have several advantages in comparison with others classical methods (gravimetric, electromagnetic, thermal…) for monitoring SMC, as they provide better temporal and spatial coverage [3].

In agriculture point of view, soil moisture information is essential for many applications like irrigation scheduling, plant stress and improving crop yield. Soil moisture also determines the partitioning of net radiation into latent and sensible heat components in the field of meteorology. Therefore, accurate soil moisture estimates are essential in several applications as to examine the effect of climate change on land surface hydrological variables such as soil moisture, infiltration fluxes, runoff and surface temperature caused by changes in heat fluxes and to quantify the amount and variability of regional water resources in water limited regions of the world on seasonal.

E.E.Abdel-hady et al.* [4], conducted experiments in which soil moisture content was measured using X-ray spectroscopy system. It is concluded that the bulk density at dry and wet stages remain unaffected as there is no rearrangement during wetting and drying process. *K. Grote et al.* [5] used Ground Penetrating Radar (GPR) to measure SMC using 450 & 900 MHz antennas. It is observed that multi-frequency GPR should be used to calculate soil moisture content at different depths. Active microwave sensor is used at high spatial resolution due to its sensitivity to the dielectric constant of soil and its moisture [6]. Hyperspectral imagery can be used to estimate the SMC but its performance depends on the crust, soil color and texture. [6-9]

Gaussian spectral models presented by *Michael* [10] presented an approach fitting an inverted Gaussian function to estimate moisture content. It is concluded that both area and amplitude of inverted Gaussian has high relationship. *Fusun Balik Sanli et al.* [11] gathered SAR data by RADARSAT , ASAR and PALSAR satellite images of Menemen Town, Izmir. The correlations between the soil moisture content and backscattering of ASAR, RADARSAT-1 and PALSAR images were found 76%, 81% and 86 % respectively. *Marion Pause et al.* [12] concluded that data obtained from inversion of airborne & satellite L-band radiometer provides estimation of soil moisture. They evaluated effect of Leaf Area Index (LAI) against airborne L-band brightness temperature of crop canopies. *Jian Peng et al.* [13] evaluated Discrete Wavelet Transform (DWT) to find SMC. 13 different mother wavelet s along with six decomposition levels from 5-10 are identified

for selected data. It is concluded that DWT reduced the hyperspectral dimensionality, thus giving better results than existing methods.

Angström A.[14] conducted laboratory experiments and concluded that soil spectral reflectance shows traces of soil moisture content. In 1987, American Society for Testing and Materials (ASTM) [15] published standard test method which uses microwave oven for estimating soil moisture content. It stated that it is alternative to conventional oven. *Bach, H.et al.* [16] confirmed the behavior proposed by Angstrom and then spectral reflectance is used to develop soil moisture content approaches. *Lesaignoux, A. et al.* [17] proposed a semi-empirical soil model which is robust to soil texture. *A priori* soil classes are identified and then soil samples were linked them.

Soren-Nils Haubrock¨ et al. [18] studied Normalized Soil Moisture Index (NSMI) to Hyperspectral data. They concluded that NSMI is best suitable to estimate soil moisture content from high spectral resolution remote sensing data. *Haubrock, S. et al.* [19] proposed new approach called NSMI in spectral range (350-2500nm) which is robust against many influencing factors. From the results, it is concluded that NSMI remains unchanged under effect of surface crusts or substrate heterogeneity. *Attila Nagy et al.* [20] used spectral reflectance of soil samples to estimate SMC. The reflectance curve of sand & sandy loam soil is linear with the wavelength & gradual increase of clay soil sample curves. *Sophie Fabre et al.* [21] recorded spectral reflectance of the collected soil samples in the reflective domain (0.4-2.5 µm). Then they compared performance of new approaches to calculate SMC against available SMC estimation indices like NSMI and WISOIL. Lobell *et al.* [22] collected four bare soil spectra and applied an exponential model onto it. They concluded that SMC is sensitive to SWIR domain.

2. DESCRIPTION OF THE REFERENCE DATA SET

2.1. Database Collection

Soil spectral reflectances were measured in the laboratory with an ASD (Analytical Spectral Devices) Fieldspec-Pro spectroradiometer in the (0.4–2.5 µm) spectral domain with a spectral resolution of 3 nm in the (0.4–1.0 µm) domain and of 10–12 nm in the (1.0–2.5 µm) domain. The database is collected from 4 different areas near Aurangabad city, Maharashtra. It is composed of 25 natural soil samples, covering different ranges of texture (clay, loam, sandy). Spectral signature of each soil sample was collected at six different SMC level, thus six spectral reflectance of each sample. Thus, our dataset consists of total 150 spectral reflectance of soil samples. Detailed description of the data set along with their Munsell color code is described in table 1.

Table 1. Detailed description of data set (Munsell color code [24]

Description		Munsell Color Code		
Area Name	**Number**	**Hue**	**Value**	**Chroma**
Himayat Baugh(6)	3	5Y	7	1
	1	2.5Y	7	6
	2	2.5Y	6	6
Dr.BAMU Campus(5)	3	5Y	5	4
	2	5Y	5	3
Himayat Nagar(8)	4	2.5Y	4	4
	2	10YR	4	2
	2	2.5Y	5	1
Pimpalgaun(6)	3	2.5Y	8	3
	3	5Y	6	1

2.2. Measurement Method

Each soil sample was oven dried (at 105 ° C) until fully dried situation. After 24 h in the oven, it is assumed that soil sample is fully dried. Then these samples are artificially wetted at different soil moisture levels: (percentage of dry weight):12%, 16%, 21%, 26%, 30% and successive spectral reflectance of soil samples at different moisture level were taken. In our dataset, there are 25 soil samples, each sample has 6 reflectance spectrum thus resulting in total 150 spectral signature in our dataset. Figure 1 illustrates the observed spectral behavior of the dry sample at different soil moisture content in the VISible (VIS; (0.4–0.8 µm)) and Near and Shortwave InfraRed (NSWIR; (0.8–2.5 µm)) spectral domains.

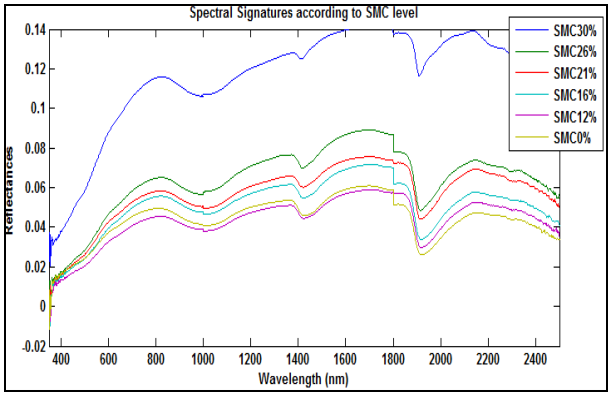

Figure 1. Spectral reflectance of soil sample at different SMC levels

3. DESCRIPTION OF THE METHODS TO ESTIMATE THE SOIL MOISTURE CONTENT

3.1. Spectral Indices

Most widely used spectral indices are Normalized Soil Moisture Index (NSMI) and Water Index SOIL (WISOIL).These spectral indices are operated at that wavelength range which are sensitive to water vapor absorption (at 1.2, 1.4, and 1.9 µm). Thus producing ineffective or underperforming results (see table 2).

The wavelengths at 1.8 µm and 1.45 µm operated respectively by NSMI and WISOIL are located at the border of the atmospheric water vapor absorption band.(see figure 2)

Table 2. Existing spectral indices to estimate SMC

Spectral Index	Specific Spectral Bands	Formulation
NSMI	1.800 µm ; 2.119 µm	$\rho_{1.8} - \rho_{2.119} / \rho_{1.8} + \rho_{2.119}$
WISOIL	1.30 µm ; 1.45 µm	$\rho_{1.45} / \rho_{1.30}$

The WISOIL and NSMI performance is then very dependent on the quality of the atmosphere compensation processing. Due to drawback of existing spectral indices, there is need to define new spectral index.

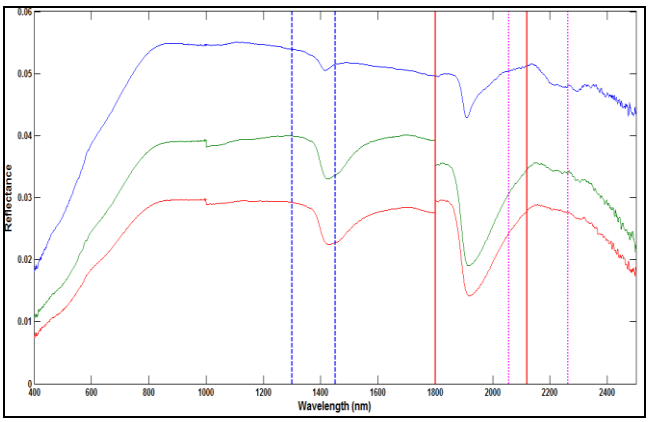

Figure 2. Spectral indices

3.2. Description of the Soil Empirical Spectral Model

Various spectral soil models exist to estimate soil moisture content. But, all these modes uses soil texture to estimate SMC. To overcome this limitation, the proposed model is based on an *a priori* soil classification defined according to the global spectral shape of the dry soil reflectances [16]. The soil samples with the same spectral behavior are then grouped together in *a priori* spectral classes. [16]. The semi-empirical soil model linking the spectral reflectance to the SMC for a given *a priori* soil class defined by the Table 2, is retained. Its analytical formulation is the following:

$$\rho^{l}_{SMC_g}(\lambda) = a_l(\lambda).SMC_g^2 + b_l(\lambda).SMC_g + c_l(\lambda) \dots (1)$$

where *l* designs the soil spectral class, *a*, *b* and *c* are the spectral coefficients of the polynomial function in the solar domain.

3.3. *A priori* Classification

Dry soil samples were used to define a priori class. These can be obtained by grouping spectral reflectance of soil depending upon their spectral shape in VIS and NSWIR domain. In our data set, three groups i.e. 1V, 2V, 3V groups are obtained in VIS domain and four groups i.e. 1N, 2N, 3N, 4N groups were identified in NSWIR domain .These groups in each domain leads to formation of six *a priori* classes defined in table 3 and illustrated in figure 3.

Table 3. The *a priori* spectral classification

Class		Spectral behavior groups	
Number	Soil number	VIS	NSWIR
1	4	1V	1N
2	5	1V	4N
3	2	2V	2N
4	3	2V	3N
5	3	3V	3N
6	8	3V	4N

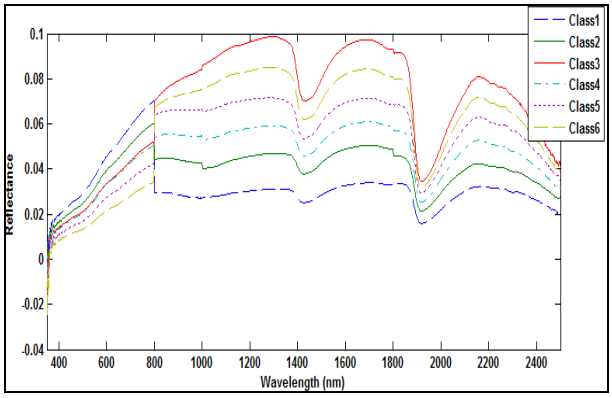

Figure 3. Six *a priori* spectral classes

Haubrock *et al.* [19] proposed procedure for new index for estimation to soil moisture content which is robust against water absorption effect.
According to this procedure, the normalized ratio X_{norm} (λ_i , λ_j) defined by the following equation:

$$X_{norm}(\lambda_i, \lambda_j) = \rho(\lambda_i) - \rho(\lambda_j) / \rho(\lambda_i) + \rho(\lambda_j) \dots (2)$$

where ρ (λ_i) and ρ (λ_j)respectively represent the reflectance values at the wavelengths λ_i and λ_j belonging to the reflective domain (0.4–2.5 µm).

The coefficient of determination for the linear regression [23] between SMC and a quantity X_{norm} (λ_i , λ_j), derived from the spectral reflectance, is plotted in a matrix where the first wavelength value λ_i is referred to by the abscissa axis and the second wavelength λ_j is referred to by the ordinate axis (Equation (2)). This matrix is called regression matrix, shown on Figure 4 and the color scale from 0 to 0.85 represents the corresponding R^2 value.

SMC are very sensitive to wavelength pairs (λ_i , λ_j) are located in the spectral range (1–2.5 µm).The wavelength pairs leading to the highest determination coefficients between the SMC and the quantity X_{norm} (λ_i , λ_j), are used to construct these new indices:

- 2056 nm and 2263 nm for X_{norm} (λ_i , λ_j) : R^2 = 85%

These results lead to the following spectral index:

- Normalized Index of NSWIR domain for Smc estimatiOn from Linear regression (NINSOL)

$$NINSOL = \rho(\lambda_{2056}) - \rho(\lambda_{2263}) / \rho(\lambda_{2056}) + \rho(\lambda_{2263})$$

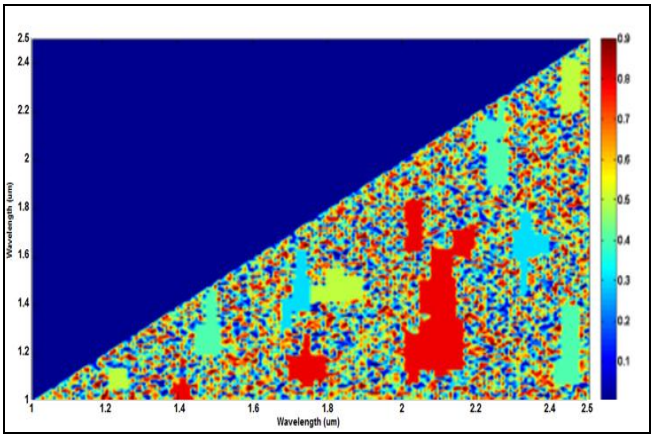

Figure 4. Determination matrix for $X_{norm}(\lambda_i, \lambda_j)$ derived by linear regression

4. RESULTS AND DISCUSSION

4.1. Performance Analysis

Spectral indices and semi empirical soil model is implemented to estimate SMC on described data set. Their performance is evaluated by two metrics i.e. coefficient of determination (R^2) and Root Mean Square Error (RMSE). The dataset is divided in two groups. first calibration data set to calibrate the methods and second validation data set is used to the validation based on 80% - 20% pattern i.e. 120 spectral reflectance in calibration data set and 30 remaining spectral reflectance in validation data set. For each method, the comparison process is as follows:

- Calibration stage: The measured spectral reflectance and the corresponding SMC of the calibration data set are used to achieve the linear regression between the index values and the SMC for NINSOL, WISOIL, NSMI.
- Validation stage: The SMC is estimated with the validation data. The quality of the SMC estimation is assessed by computing the R^2 and the RMSE.

The RMSE is expressed as follows:

$$RMSE = \sqrt{\sum (SMC^{i}_{mes} - SMC^{i}_{est})^2 / N}$$

where SMC_i^{est} is the estimated SMC for the soil sample *i*, SMC_i^{mes} is the measured SMC for the same sample *i* and N is the number of samples.

4.2. Results

Table 4 shows performance metrics i.e. R^2 and RMSE on the validation data set and measured SMC by semi empirical model for each *a priori* soil spectral class. From the results, it is concluded that R^2 is better than 0.83 and RMSE ranges between 4% and 6%.

Table 4. R^2 and RMSE for each *a priori* class

Soil Spectral Class	R^2	RMSE
1	0.89	4.4
2	0.88	4.7
3	0.87	6.2
4	0.90	5.5
5	0.93	5.2
6	0.83	6.0

Performance comparison for existing and proposed spectral indices is given in table 5. NSMI and WISOIL gives 0.79 and 0.81 values for R^2 and 6.6, 6.2 RMSE values respectively. The new proposed index NINSOL shows better performance than existing spectral indices i.e. R^2 value of 0.85 and 5.6 RMSE value.

Table 5. Methods and their performance

Method	R^2	RMSE
NSMI	0.79	6.6
WISOIL	0.81	6.2
NINSOL	0.85	5.6

The results for WISOIL and NINSOL are illustrated in the figure 5 and figure 6 respectively.

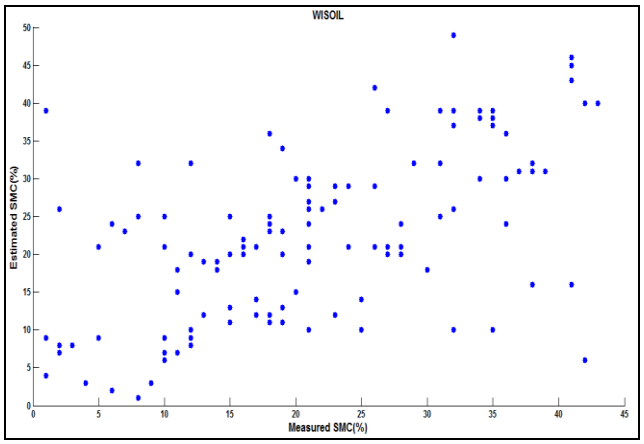

Figure 5 .Estimated SMC according to measured SMC for WISOIL

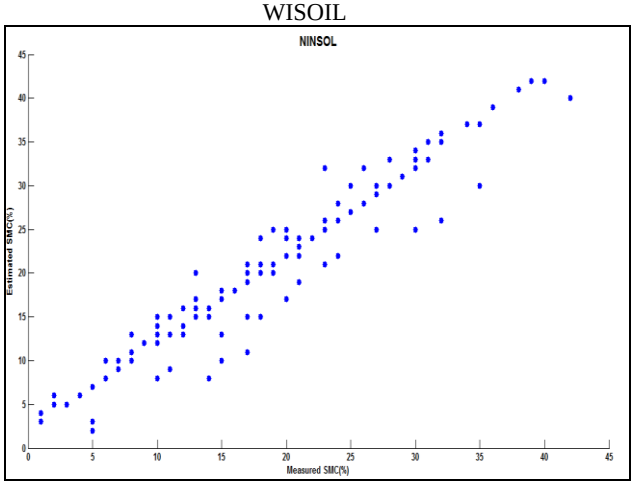

Figure 6. Estimated SMC according to measured SMC for NINSOL

5. CONCLUSION

Based on the results of the study, soil moisture content can be efficiently estimated with the help of spectral signature of soil sample in reflective spectral domain (0.4-2.5 um). Existing methods takes soil texture as important parameter for estimation of soil moisture content. Existing spectral indices i.e. NSMI and WISOIL operates in the range of wavelength which are sensitive to water vapor absorption. Semi empirical soil model overcomes limitation of existing soil models. It defines *a priori* classes for data set and then soil samples were linked to these *a priori* classes to estimate SMC. It is observed that soil moisture content is sensitive to the spectral range 1-2.5 um. Thus, study of soil samples has been carried out in this range by determining the regression matrix for normalized difference of spectral reflectance by linear regression. New proposed index i.e. NINSOL is robust and remains less affected by water vapor absorption. NINSOL operates in 2056 nm and 2263 nm spectral range. Performance analysis has been done on NSMI, WISOIL and NINSOL. NSMI led to R^2 value of 0.79 and RMSE value of 6.6. While WISOIL produces R^2 value as 0.81 and 6.2 RMSE value. NINSOL produces better results than existing indices as R^2 value of 0.85 and RMSE value of 5.6.

6. ACKNOWLEDGMENTS

This work is supported by Department and Science and Technology under the Funds for Infrastructure under Science and Technology (DST-FST) with sanction no. SR/FST/ETI-340/2013 to Department of Computer Science and Information Technology. Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India. The authors would like to thank Department and University Authorities for providing the infrastructure and necessary support for carrying out the research.

7. REFERENCES

[1] Nocita, M.; Stevens, A.; Noon, C.; Wesemae, B.V. 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma ,199, 37–42.

[2] Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. 2009. Using Imaging Spectroscopy to study soil properties. Remote Sens. Environ., 113, 538–555.

[3] 6. Bryant, R.; Thoma, D.; Moran, S.; Holifield, C.; Goodrich, D.; Keefer, T.; Paige, G.; Williams, D.; Skirvin, S. 2003. Evaluation of Hyperspectral, Infrared Temperature and Radar Measurements for Monitoring Surface Soil Moisture. In Proceedings of the First Interagency Conference on Research in the Watersheds, Benson, Arizona, 27–30 , pp. 528–533.

[4] E.E.Abdel-hady*, A..M.A..EL-Sayed and H.B.AIaa. 1996. Determination Of Moisture Content And Natural Radioactivity In Soils Using Gamma Spectroscopy. Third Radiation Pkytk* Conf* AUMinia.

[5] K. Grote, S. Hubbard , Y. Rubin. 2003. Field-scale estimation of volumetric water content u sing ground penetrating radar ground wave techniques. Water Resources Research, VOL. 39, NO. 11, 1321.

[6] Shi, J.; Wang, J.; Hsu, A.Y; O'Neill, P.E; Engman, E.T. 1997. Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. IEEE Trans. Geosci. Remote Sens., 35, 1254–1266.

[7] Ben-Dor, E.; Banin, A. 1994. Visible and near-infrared (0.4–1.1 μm) analysis of arid and semi-arid soils. Rem. Sens. Environ., 48, 261–274.

[8] Baumgardner, M. 1985. Reflectance properties of soils", Adv. Agron., 38, 1–44.

[9] Goldshleger, N.; Ben-Dor, E.; Benyamini, Y.; Blumberg, D.; Agassi, M. 2002. Spectral properties and hydraulic conductance of soil crusts formed by raindrop impact. Int. J. Remote Sens., 23, 3909–3920, 2002.

[10] Michael L. Whiting, Lin Li, Susan L. Ustin. 2003. Predicting water content using Gaussian model on soil spectra. Remote Sensing of Environment 89, 535–552.

[11] Fusun Balik Sanli , Yusuf Kurucu , Mustafa Tolga Esetlili, Sayg ın Abdikan. 2008. Soil Moisture Estimation from RADARSAT -1, ASAR and PALSAR Data in Agricultural Fields of Menemen Plane of Western Turkey. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing.

[12] Marion Pause, Karsten Schulz, Steffen Zacharias, and Angela Lauschd. 2012. Near-surface soil moisture estimation by combining airborne L-band brightness temperature observations and imaging hyperspectral data at the field scale. Journal of Applied Remote Sensing 063516-3 Vol. 6.

[13] Jian Peng , Hong Shen , San Wei He , Jian Sheng Wu. 2013. Soil moisture retrieving using hyperspectral data with the application of wavelet analysis. Environ Earth Sci ,69:279–288.

[14] Angström, A. 1925. The albedo of various surfaces of ground. Geogr. Ann.,7, 323–327.

[15] ASTM D 4642-87. 1987. Standard Test Method for Determination of Water Moisture Content of Soil by the Microwave Oven Method.

[16] Bach, H.; Mauser, W. 1994. Modelling and model verification of the spectral reflectance of soils under varying moisture conditions. In Proceedings of the International Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation, Pasadena, CA, USA, Volume 4, pp. 2354–2356.

[17] Lesaignoux, A.; Fabre, S.; Briottet, X. 2013. Influence of soil moisture content on spectral reflectance of bare soils in the 0.4–14 μm domain. Int. J. Remote Sens. , 34, 2268–2285.

[18] Soren-Nils Haubrock¨ , Sabine Chabrillat, Matthias Kuhnert, Patrick Hostert and Hermann Kaufmann. 2008. Surface soil moisture quantification and validation based on hyperspectral data and field measurements. Journal of Applied Remote Sensing, Vol. 2, 023552.

[19] Haubrock, S.; Chabrillat, S.; Lemmnitz, C.; Kaufmann, H. 2008. Surface soil moisture quantification models from reflectance data under field conditions. Int. J. Remote Sens. 29, 3–29.

[20] Attila Nagy , Péter Riczu, Bernadett Gálya, János Tamás 2014. Spectral estimation of soil water content in visible and near infra-red range. Eurasian Journal of Soil Science 3 ,163 – 171.

[21] Sophie Fabre , Xavier Briottet and Audrey Lesaignoux. 2015. Estimation of Soil Moisture Content from the Spectral Reflectance of Bare Soils in the 0.4–2.5 μm Domain. Sensors, 15, 3262-3281.

[22] Lobell, D.; Asner, G. 2002. Moisture Effects on Soil Reflectance. Soil Sci. Am. J., 66, 722–727.

[23] Berk, A.; Bernstein, L.S.; Anderson, G.P.; Acharya, P.K.; Robertson, D.C.; Chetwynd, J.H.; Adler-Golden, S.M. 1998. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. Remote Sens. Environ., 65, 367–375.

[24] Munsell Soil Color Chart 2009 Edition. Available online: https://nenc.gov.ua/old//GLOBE/Other/Munsell%20soil%20colour%20chart.pdf (accessed on 15 April 2017).