ISBN: 978-93-5300-455-2

SHRI SHANKARLAL SUNDARBAI SHASUN JAIN COLLEGE FOR WOMEN

A Unit of Sri S. S. Jain Educational Society | Affiliated to University of Madras
Accredited with 'A' Grade by NAAC | An ISO 9001:2015 Certified Institution

International Conference on

Recent Advances in Computing and Communication

ICRACC 2018

(23rd–24th February 2018)

Organized by

Department of Computer Science
Shri Shankarlal Sundarbai Shasun Jain College for Women
Chennai, India

Publication Partner

Genxcellence Publications & IJARCS
www.ijarcs.info, India

INTERNATIONAL CONFERENCE

ON

RECENT ADVANCES IN COMPUTING AND COMMUNICATION

ICRACC 2018

23rd and 24th February, 2018

PROCEEDINGS

ISBN: 978-93-5300-455-2

Volume 9, Special Issue No. 1, February 2018

International Journal of Advanced Research in Computer Science

(ISSN: 0976-5697)

Available Online at www.ijarcs.info

IN ASSOCIATION WITH

Organized By

Department of Computer Science

Shri Shankarlal Sundarbai Shasun Jain College for Women

3, Madley Road, T.Nagar,

Chennai - 600 017.

Telephone: 044 2432 8506/07

Website: www.shasuncollege.edu.in

Email id: info@shasuncollege.edu.in

PROCEEDINGS

ABOUT THE INSTITUTION

Shri Shankarlal Sundarbai Shasun Jain College for Women, an educational institution par excellence, is dedicated to bring about women empowerment through education. It is affiliated to the University of Madras. It has been accredited with A Grade by National Accreditation and Assessment Council (NAAC) and ISO 9001:2015 certified by Bureau Veritas. NIRF announced the ranking on 3rd April 2017. Our college secured 59th Rank at the National Level. Our college received "Best Accredited Student Branch Award" given by Computer Society of India in 2014 and 2016 during CSI Annual Convention. The management has made available the latest technology for both faculty and students to make teaching and learning experience a desirable one. Our college is a perfect knowledge centre for the future.

ABOUT ICRACC 2018 CONFERENCE

The two-day International Conference on Recent Advances in Computing and Communication (ICRACC) 2018 organized by the Department of Computer Science aims to bring together leading academic scientists, researchers and research scholars to exchange and share their experiences and research results on all aspects of recent trends of Computer Science. It also provides a premier interdisciplinary platform for researchers, practitioners and educators to present and discuss the most recent innovations, trends, and concerns as well as practical challenges encountered and solutions adopted in the fields of Computer Science.

The main theme of the conference is "Recent Advances in Computing and Communication". This International Conference makes sure to support and motivate 'Young Researchers' by establishing their academic and professional relationships.

MESSAGE FROM THE SECRETARY

Shri. S. Abhaya Kumar Srisrimal Jain

Secretary

It is a great pleasure to know that the Department of Computer Science is organizing an International Conference on “Recent Advances in Computing and Communication” (ICRACC 2018).

The purpose of this conference is to bring together researchers, experts from industry, academia, and other interested organizations to meet, exchange information and ideas on developments in the computing and communication field.

I hope this conference would certainly help the participants to have the latest updates and better understanding to contribute more towards the recent trends in technology.

I wish the best to the participants and the faculty members.

MESSAGE FROM THE ASSOCIATE SECRETARY

Shri. R. Ashok Mehta

Associate Secretary

I am really happy to know that the Department of Computer Science is organizing an International Conference on "Recent Advances in Computing and Communication" on 23rd & 24th February 2018.

I hope that this conference would surely encourage innovative ideas among the participants paving way for new inventions in the Computing field.

MESSAGE FROM THE PRINCIPAL

Dr. B.Poorna

Principal

Welcome to ICRACC 2018.

I am very glad that the Department of Computer Science is organizing an International Conference on "Recent Advances in Computing and Communication" on 23rd and 24th February, 2018.

Research activities across all the computing fields pave the way for the industrial world to strive forward with huge advancements. As an educational institution, encouragement and support to research can be provided by establishing a suitable platform for the research community, to interact with each other and to share the knowledge. Having this objective in mind, ICRACC 2018 has been organized to provide the same benefits and learning experience to all the participants. Sessions on different domains, keynote addresses from eminent professors and opportunity to network with the researchers will help the participants immensely in their research career.

This technical International Conference will provide a prestigious international platform by bringing together local and overseas technical researchers and students to exchange their experienced knowledge and expertise in issues relating to the dominating trends in technology.

With great pleasure and pride, I welcome all the participants and convey my best wishes for ICRACC 2018.

ORGANIZING COMMITTEE

CHIEF PATRONS	:	**Shri. S. Abhaya Kumar Jain** – Secretary
		Shri. R. Ashok Kumar Mehta – Associate Secretary
CO-PATRONS	:	**Dr. B. Poorna** – Principal
		Dr. S.T. Deepa – Associate Dean
CONVENER	:	**Ms. B. Gomathi,** HOD
		Department of Computer Science
CO-CONVENER	:	**Dr. M. Anita Indu,** Asst. Professor,
		Department of Computer Science
MEMBERS	:	**Ms. K. Suma,** Assistant Professor
		PG Department of Computer Science
		Ms. S.G. Packiavathy, HOD,
		Department of Computer Applications (Shift I)
		Ms. X. Jose Suganya, HOD,
		Department of Computer Applications (Shift II)
		Ms. T.Yegammai, Asst. Professor
		Ms. S. Sridevi, Asst. Professor
		Ms. N. Latha, Asst. Professor
		Ms. N. Rehna, Asst. Professor
		Ms. V. Venkateswari, Asst. Professor
		Ms. N. M. Kavitha, Asst. Professor
		Ms. S. Sasikala, Asst. Professor
REVIEWERS	:	**Dr. S.T. Deepa** – Associate Dean
		Ms. S.G. Packiavathy, HOD – BCA (Shift I)
		Ms. X. Jose Suganya, HOD – BCA (Shift II)
		Ms. S. Sridevi, Asst. Professor
		Dr. M. Anita Indu, Asst. Professor
		Dr. S. Prasanna , Asst. Professor

INDEX

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A Method for Classifying the Germination of Green Gram Image using Neural Networks

Harish S Gujjar
Asst. Professor and Head, Dept. of Computer Science, SSA GFGC Ballari, Karnataka, India
gujjarh@gmail.com

ABSTRACT:

This paper deals with a computer vision system based on machine learning techniques in the field of image processing and germination of the green gram is spontaneously assessed the rate of germination. The germination test is most important and trusted method to determine the speed and successfulness of germination. It gives us the important information regarding the successful of converting the germinated seed into plant. On a whole all green grams are not able to germinate. In this paper we use Artificial Neural Network (ANN) which uses Multilayer Perception structures are used. In between 30 ^{0}C to 40^{0}C almost all the seed germination takes place. On an average only 90% of the seeds can germinate. This work is able to classify accurately of 96% of germinated seeds.The Green gram samples are collected from APMC in Ballari districts of Karnataka for the growing year 2018.

Keywords: artificial neural networks, seeds, germination, green gram, image processing

INTRODUCTION

To increase the growth and supply of money is based on the quality of the seed which is the most fundamental part of any agriculture. Earlier, the detection of many physical and interpretations of seeds was focused, But today a large variety of techniques are available for seed testing green gram seed assessment and judgment which correlate well with certain vigor and germination parameters (McDonald 1998). In the present context seed testing is carried out in scientific experiment by expert humans. The experiments are sketched to assess the quality of the seed. Large number of tests is carried out. For example, the highest germination capability can be found out by the germination test or successful of germination of seed. The germination speed of certain seed bulk seeds is an important sign which represents the seed performances in the farm and it is indicated as percentage (for example 89% germination pace means out of 100 , 89 seeds are expected to germinate subjected to the good growth atmosphere). The above measure is important for estimating maximal seeding speed and to estimate whether a certain seed lot has the ability to process a quality crop.

From then to now the manual counting is tedious and takes lot of time and human resource, the efficiency of the process can be boosted by very large number of ways. On the basis of Machine learning and Image processing and with the help of Computer vision system, we can design a machine which tests the approximation of germination of seeds.

Analysis of images was introduced in the area of seed technology by Howarth and Stanwood (1994) those are created a database of color images to identify the variation of both environmental and genetic phenomena.

In the field of seed Identification and classification in the field of image processing it has a very good result (Uchigasaki et al. 2000, Granitto et al. 2002) and judging the germination (McDonald et al. 1998). Dell' Aquila et al. (2000) he has taken analysis of image to signalized the objective of seeds of White cabbage while Geneve and Kester (2001) assessed size of the seeding following germination by digital image analysis which are assesed by computer by using scanner (Ducournau et al. 2004). Urena et al. (2001) suggested a Machine vision system which included a process of gathering data by using a system with a technique of logic based fuzzy to assess the quality of germination. Ducournau et al. (2004) produced a Machine vision system modeled to count the number of radical tips which were appeared on the seeds below temperature, lighting and measurement of moisture in air conditions. A mechanical process system which works on the algorithm which has an ability to count the seeds which are germinated and give the average germination time on the point of difference between the two successive pictures.

An image processing and computer vision uses famous and renowned software called MATALAB. A digital camera of Nikon D810 FX 36.3MP Digital SLR (Single Lens Reflex) Camera sigma 18 -200 mm lens zooming. The camera is fixed to a iron stand having movement towards vertical which gives a firm support. The camera was fixed at a distance of 450 mm. The images were having a resolution of 3873 X 2593 pixels, a horizontal and vertical resolution of 300 dpi respectively having a bit dept of 24. A 210 mm diameter white light of 22 Watts fluorescent tube was fixed which glown at a voltage of 230 volts was placed over the cell culture dish having the seed samples placed on a white tissue paper. A 270 mm diameter semi spherical steel utensil is used to cover bulb to get the diffused light to eliminate the impact of external factors (Figure 1). The images thus obtained from the above are transferred from digital camera to PC (quad core processor having 4 GB RAM) through a USB cable.

IMAGE AQUISITION

The samples of green grams are bought from the APMC from the areas of ballari districts of Karnataka, 2018 as the growing year. At the start of the experiment, the green grams are preserved for 30 days in a controlled air and temperature at around 4 C°, the seed is moisturized to balance the relative humidity at 50%. The 700 seed samples are randomly chosen from the bag. Further, a white filter paper is put to a cell cultured dish which is moisturized with a 3ml of pure water. The optimal contrast between the seed was minimized by using the white filter paper. An approximately 25 seeds were spread over the top of the misted filter paper in every dish and was positioned at a relatively similar distances. The dish was inculpated with cover. The germination of seeds were carried out in an well directed condition and in very low light at 20 to 30 °C and relative humidity at 75% in an incubator invented by Jacobsen. In a day 8 hours were used to illuminate the seeds.

A cool white fluorescent light of 750 flux was used to provided. The seizing of the images was carried out by a digital camera with Nikon D810 FX 36.3MP Digital SLR (Single Lens Reflex) Camera sigma 18 -200 mm lens zooming. The camera is fixed to a iron stand having movement towards vertical which gives a firm support. The camera was fixed at a distance of 450 mm. The images were having a resolution of 3873 X 2593 pixels, a horizontal and vertical resolution of 300 dpi respectively having a bit dept of 24. A 210 mm diameter white light of 22 Watts fluorescent tube was fixed which glows at a voltage of 230 volts was placed over the cell culture dish having the seed samples placed on a white tissue paper. A 270 mm diameter semi spherical steel utensil is used to cover bulb to get the diffused light to eliminate the impact of external factors(Figure 1). The images thus obtained from the above are transferred from digital camera to PC (quad core processor having 4 GB RAM) through a USB cable.

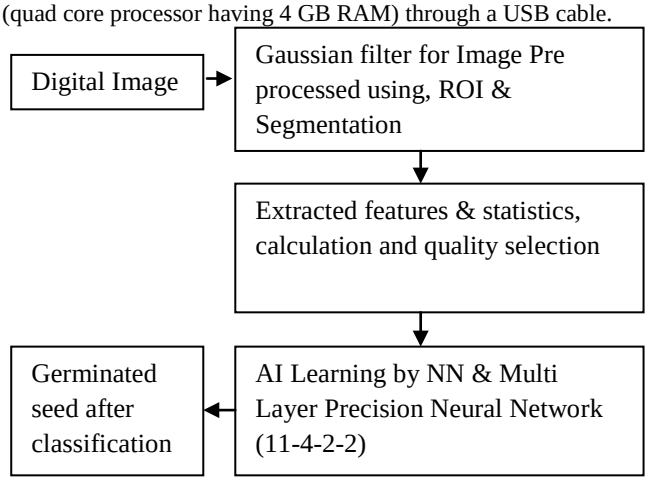


Figure1: Proposed methodology.

PRE PROCESSING OF AN IMAGE

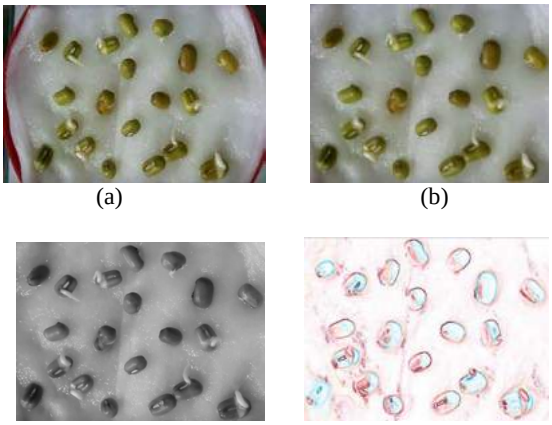
Firstly, original RGB images were used to extract the features with the help of MATLAB which helps in image processing (Figure 2a). The individual images thus received were cropped by a fixed known radius in the region of interest (ROI) (Figure 2b). To efficiently manipulate all the images, by cropping of images the size of the image was reduced which is helpful in manipulations. The original image has a size of 3873 X 2593 pixels was diminished to 1854 X 1836 pixels. Secondly, the images were smoothed with a σ limited at 2 which uses a Gaussian filter (Eq. 1) drawn by Rasband (2008):

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu^2)}{2\sigma^2}} \qquad (1)$$

Where, the pixel intensity is represented as *X*, mean is indicated μ , Standard deviation is notified as σ, variance is signified as σ^2 =3.18 and **e**=2.718. The conversion of the image from color space RGB to an 8-bit image of a gray scale (Figure 2c) and was ultimately converted by thresholding into a binary image (Figure 2d). The predefined testing of the thresholding were carried out by applying the lower limit as 85 and the highest limit as 255grayscales.

The maximum and minimum size of a particle was set to 1500 to 14,000 pixels respectively. By doing this we can just ignore the smaller and larger areas which cannot be considered as seeds. Lastly the seed was extracted from the background and was named by a device itself. The external tracing of the boundaries of the seeds was considered as yellow (Figure 2e).

The minimum particle size was set to 1500 pixels and the maximum to 14.000 pixels. With this we additionally avoided the smaller and the bigger areas that could not be accounted as seeds. Finally, presented seeds were separated from the background and automatically labeled with the integer. The external perimeter of the seed was traced in yellow .



(a) (b) (c) (d)

Figure2: (a) Original color image of RGB, (b) Image is cropped, (c) 8-bit grey scale image, (d) Binary image after thresholding.

After the Thresholding and Labeling of the seeds the upcoming step is to examine the particles which were labeled. The detailed description of the steps in image processing is shown in Table 1.

The culmination of the different features which was extracted together from different Petri dish was tabulated, where row represents the seed and the column represents the individual parameter of a single seed. For the examination by expert a 28 different images and 25 different seeds was bestowed for germinated and non germinated classification over a 700 seeds. These results were also included in the table.

PROPOSED METHODOLOGY

Further the analysis was carried out by software which uses machine learning called WEKA (Waikato Environment for Knowledge Analysis), A formatted file is created using *csv* (comma-separated values) and send once the feature characteristics are generated. WEKA is a data mining and machine learning tool which consists of large number of algorithms for operating by means of a program, categorization, Measurement of relation between the mean and variance, grouping the number of things of same kind, amalgamation rules and representation of objects.

The information identification from main data was taken from main streams of agricultural, to do the above the University of Waikato, New Zealand proposed WEKA and further it was used for other fields also. Initially it was available i Java version and was translated to programming in MATALAB.

The amalgamation of the two tools was primarily described by Mayo et al. (2007) in this method feature vectors were used for classification of months by using species for drawing out of the each month of the image. Further choosing and grading was carried out by WEKA. The assessment process used info grain attribute evaluator which was used to estimate by attributing the grade by grading and rating by a single assessment. The edge filtering and small mean from the outer length of the closed boundary is the largest result of classification which is the gross value of each and every pixel in the classified region of pixels (Table 1).

Parameter	Description
Perimeter	The end to end are of the outside area is called a Perimeter.
Kurtosis	The measurement of the degree of curved arck is known as Kurtosis.
Max	The highest value of grey level within the selected boundary.
Skewness	The lack of equality of distribution of measure of the degree is considered as Skewness.
Standard Deviation	The average grey value is obtained by the grey values of standard deviation.
Major	Major is the primary axis of the best fitting ellipse.
Area	A four sided are of pixels is called the selected area.
Mode	The largely repeated value in the area of grey mode level. It refers to the top most point in the histograph.
Median	The median value is the average of two middle values of pixels in the particular area.
Minor	Minor is the subsidiary axis of the prime suited ellipse.
Mean	It is the central value of grey inside the selected area. It is the summation of the gray values in the wanted area which is divided by the number of pixels.

Table 1: Measured Parameters.

The 10 sets of data were used for learning of modules of classification, each having 70 occurrences. Then, 9 sets of these were used for training and the testing was carried out on the remaining sets (training seeds of 630 and testing of 70 seeds for each execution. The above process was successive used for 10 times on the structure thus designed.

CLASSIFICATION MODEL

The best method for categorization is Artificial Neural networks (ANN) with multilayer perceptions model (MPL) was used and compared with the human made counting. The back propagation algorithm model was used for training. At 0.3 was the value set for rate of learning and 0.2 as momentum rate. The input layer has 11 neurons and where as the output layer has w neurons as the number of features and classes was 11 and 2 respectively. Inside the ANN there are large number of hidden layers and neurons which were trained at the speed of 100 to 2000 epochs.

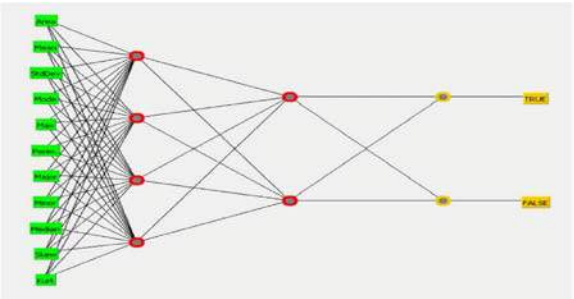


Figure3: Model of Artificial Neural Network

PERFORMANCE EVALUATION

To judge the performance of the ANN, we have cauterization accuracy, correctness, recollect and F-calculate which are extracted from confusion of matrix. Accuracy of classification is the process of the judgment of the accuracy of the model of neural network in comparison made with the human beings and their correctness over 25 data sets. The calculation is carried out by the number of accurately categorized occurrences which were slashed by the number of occurrences.

$$\text{Accuracy} = \frac{\sum(TN,TP)}{\sum(TP,FP,TN,FN)} * 100\ \% \qquad (2)$$

where true positive is abbreviated as TP, True Negative is abbreviated as TN, FP & FN is unfolded as false positive false negative respectively. The summation of TP+TN+FP+FN is the total number of instances occurred in the test set TN + TP is the accurate count cauterization of instances (Witten and Frank 2005).

In this paper TP is constitute as real seeds of germination which were also assumed as germinated seeds. TN is not constitute as real seeds of germination which were also assumed as non germinated seeds. The actual not germinated seeds are notified as FP and FN are assumed as not germinated and were actually germinated.

Precision is the amount of instances predicted positively which are positive which were really assumed as positively among the total which is estimated as follows :

$$p = \frac{TP}{\sum(TP,FP)} \qquad (3)$$

where, TP is abbreviated as true positive and FP is abbreviated as false positive. A FP happens when the class is incorrectly estimated as positive when it is actually obstructive (Witten and Frank 2005).

model	Accuracy		Precision	
	Mean	Std. dev	Mean	Std. dev.
ANN	95.99	3.17	0.995	0.0344

model	Recall		F-measure	
	Mean	Std. dev.	Mean	Std. dev.
ANN	0.9955	0.0200	0.959	0.0264

Table 2: Calculated Performance for cauterization

Recall is the summation of the TP and TF which is explained as true positive to summation. It is sometimes named as sensitivity in different areas. It is measured as follows:

$$\text{recall} = \frac{TP}{\sum(TP,FN)} \qquad (4)$$

FN is incurred when approximation is incurred as negative when it is really positive (Baeza and Riberio 1999, Witten and Frank 2005).

F-measure is stated as the recall and harmonic mean of precision. It is mesured as follows:

$$\text{F-measure} = \frac{2\,(Precision\ *recall\ )}{(\ Precision\ +Recall\ )} \qquad (5)$$

When both precision and recall have high values the net result is also has high values and it is decide as the fine method of agreement between them (Baeza and Riberio 1999).

REASULT AND DISCUSSION

The finest correctness of 95.99 % (Table 2) was extracted with two hidden layers that included four and two hidden neurons respectively which included a training set of 500. By algorithms we obtained the pace of TP at round 97% with 3 % FN, 93 % TP

and 7% FP respectively. We obtained the recall at 0.9955 and F measure at 0.959 respectively by using the ANN.

The importance of our model has gained over the research of (Dell' Aquila et al. 2000), person observed the germinated seeds the method of difference in seed position with respect to XY point of seed on superimposition of two successive images of before and after germination. A single image was used to extract 11 different features to estimate the germination of the seed. A single image is also advantageous because there may be a chance of fungal infection caused by the removal of the incubator by Jacobsen.

The very important dissimilarity between our research and the jossen et al (2010) is that the automation was entirely carried out without the human interventions for exchanging the data between various software areas.

CONCLUSION

In this paper, a automatic judgment between the seeds germination speed was implemented by using the image processing by computer vision and machine learning techniques. The outcome indicates that the accuracy of using ANN is very high (96%). The model thus developed has the capability of cauterization of germinated seed which has out raided other methods. It has also illustrated the great signs of time consumption in automated process and the human process in cauterizing the germination of seed quality.

REFERENCES

1. Hiroyuki Nonogaki, Seed biology updates - Highlights and new discoveries in seed dormancy and germination research, International Journal of Frontiers in Plant science, Volume 8, Apr 11, 2017.
2. Sangeeta Mishra, Effect of Temperature and Light on the seed germination of *Sida cordifolia* International Journal of Scientific and Research Publications, Volume 6, Issue 1, January 2016, ISSN 2250-3153.
3. Dushyant, K. S., Showkat A. G., Gurpreet S., Rajneesh K. A. and Rajendra S. 2015. Reproductive Phenology of Sida cordifolia L. Ann. Bio. Sci. 3 (1):10-15.
4. Jossen R, Kodde V, Willems L, Ligterink W, Van der Plas L, Hilhorst H. Germinator: a software package for high-throughput scoring and curve itting of Arabidopsis seed germination. Plant J. 2010; 62(1):148-159.
5. Rasband WS. Image. US National Institutes of Health, Bethesda, Maryland, USA, 2008.
6. Mayo M. Watson AT. Automatic species identification of live moths. Knowledge-Based Systems, 2007;20:195-202.
7. Witten, I., and Frank, E. (2005).Data Mining: Practical Machine Learning Tools Techniques (2nd Ed.).
8. Morgan Kaufmann. Ducournau S, Feutry A, Plainchault P, Revollon P, Vigouroux B, Wagner MH. An image acquisition system for automated monitoring of the germination rate of sunflower seeds. Comp. Electron. Agricult. 2004;44:189-202.
9. Granitto PM, Navone HD, VerdesPF, Ceccato HA. Weed seeds identification by machine vision. Comp. Electron. Agricult. 2002;33:91-103.
10. Geneve RL, Kester ST. Evaluation of seedling size following germination using computer-aided analysis of digital images from flat-bed scanner. Hortscience 2001;36(6):1117-1120.
11. Urena R, Ro driguez F in B erenguel M. A machine vision system for seeds germination quality evaluation using fuzzy logic. Comp. Electron. Agricult., 2001,32:1-20.
12. Dell'Aquila A, van Eck JW, van der Heidjen GWAM. The application of image analysis in monitoring the imbibitions process of white cabbage *(Brassica oleracea* L.) seeds. Seed Sci. Res. 2000;10:163-169.
13. Uchigasaki M, Serata K, Miyamoto S. An automated machine vision system for classification of seeds using color features. J. Agricult. Struct. 2000;30(4):325-332.
14. Baeza R, Ribeiro B. Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Boston, 1999: 544 p.
15. McDonald MB. Seed quality assessment. Seed Sci. Res.1998;8:265-275.
16. Howarth MS, Stanwood PC. Extracting 3-D information using 2-D images of seeds. Comp. Electron. Agricult. 1994;10:175-188.

Harish S. Gujjar is working as an Assistant professor and Head in department of computer science, SSA GFGC, Ballari, Karnataka, India. He did his M.C.A (Master of Computer Application) from Vishveswaraiah Technical University, Belgaum, Karnataka, India. M. Phil (Computer Science) from Allagappa university, Karaikudi, Tamilnadu, India. And pursuing Ph. D (Computer Science) from Bharathiar University, Coimbatore , Tamilnadu, India.

Volume 9, Special Issue No. 1, February 2018
ISBN: 978-93-5300-455-2
International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER
ICRACC 2018
(23rd–24th February 2018)
Available Online at www.ijarcs.info

IRIS IMAGE PREPROCESSING AND COMPRESSION FOR HIGHLY SECURED AUTHENTICATION

Rajapriya C.
Asst. Prof, department of computer science
Sri Muthukumaran Arts and Science College, Mangadu, Chennai, India

ABSTRACT

The authentication of humans the use of iris-based popularity is an extensively growing era. Iris popularity is feasible to be used in differentiating between same twins. Even though the iris coloration and the overall statistical first-class of the iris texture may be depending on genetic factors, the textural information are independent and uncorrelated for genetically same iris pairs. The function extraction and class are heavily primarily based on the rich textural details of the iris. With the need for protection structures going up, Iris authentication is rising as one of the essential techniques of biometrics-based totally identity structures. This undertaking essentially explains the Iris popularity device advanced through Daugman and tries to put in force this set of rules, with some modifications. Firstly, image preprocessing is completed followed by way of extracting the iris portion of the attention image. The extracted iris element is then normalized, and iris is constructed the use of 1D Gabor filters. Later iris and pupil are as compared to find the Hamming Distance that is a fractional measure of the dissimilarity.

Keywords: Preprocessing, Normalization, Gabor Filter, Hamming Distance, Pupil, Iris

INTRODUCTION

Biometrics refers to the identification and verification of human identity based on certain physiological developments of someone. The typically used biometric capabilities include speech, fingerprint, face, handwriting, gait, hand geometry etc. The face and speech strategies were used for over 25 years, even as iris method is a newly emergent approach. The iris is the colored part of the eye in the back of the eyelids, and in the front of the lens. It's miles the only inner organ of the body that is commonly externally visible. Those seen patterns are specific to all individuals and it has been observed that the opportunity of finding people with same iris patterns is almost 0. Although there lies a hassle in taking pictures the image, the splendid sample variability and the stableness over the years, makes this a dependable safety recognition machine.

An iris-based totally biometric identity scheme involves studying capabilities that are discovered within the tissues that surrounds the student. complicated iris patterns can comprise many one-of-a-kind functions together with ridges, crypts, jewelry, and freckles[1]. Iris scanning makes use of a reasonably conventional camera and requires no close contact between the issue and the reader. The iris is specific from man or woman to person because there are such a lot of exclusive styles that surround the pupil. The iris-scanning procedure is straightforward and painless. Iris recognition is a tested, correct method to discover people. It examines automatic iris reputation as a biometrically based generation for non-public identification and verification. Iris popularity machine consists of the pre-processing system, segmentation, function extraction and reputation.in comparison with different biometric capabilities, iris can gain high accuracy due to the wealthy texture of iris styles. The performance of iris reputation machine relatively depends on segmentation. most commercial iris recognition structures use patented algorithms developed via Daugman, and these algorithms are able to produce ideal popularity rates [2]. The Canny side Detector is one of the maximum usually used image processing tools, detecting edges in a completely strong way[3]. This paper presents an technique for segmenting the iris patterns.The used approach determines an automated worldwide threshold and the pupil middle. Experiments are performed the usage of iris pix obtained from CASIA database (Institute of Automation, chinese Academy of Sciences) and MATLAB application for its easy and green equipment in image manipulation.The system is to be composed of a number of sub-systems, which correspond to each level of iris reputation. these degrees are segmentation – locating the iris location in an eye fixed image, normalisation – developing a dimensionally consistent representation of the iris area, and characteristic encoding – developing a template containing handiest the most discriminating capabilities of the iris. The input to the gadget might be an eye image, and the output might be an iris template, in order to offer a mathematical illustration of the iris vicinity.performance analysis parameters used for proposed system are Computational time, false attractiveness charge, false Rejection rate and Accuracy.performance parameters are evaluated among two iris popularity techniques.We evaluate the accuracy among the proposed system and Wavelet rework method.All snap shots tested in this task had been taken from the chinese language Academy of Sciences Institute of Automation (CASIA) iris database. In practical utility of a plausible gadget, an image of the eye to be analyzed ought to be acquired first in virtual form appropriate for analysis.

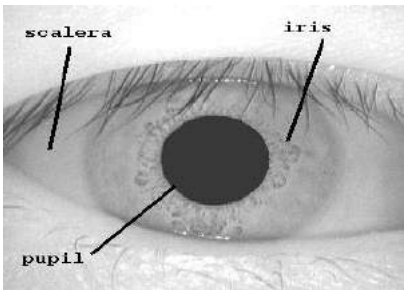


Figure 1 : Image of the eye

Iris Preprocessing

Image enhancement is the process of adjusting digital image so that the results are most suitable for display.It is used to improve the quality of the image. The **Adaptive Mean Adjustment**[2] is used to enhance the image.Adaptive Mean Adjustment[2] is a computer image processing technique used to improve contrast in images. It modifies the allocation of the pixels to become more consistently increase out more than the obtainable pixel variety. In histogram dealing out, a histogram displays the sharing of the pixel intensity values. Dark image will have low pixel values whereas a bright image will have high pixel values.

Figure 2: Canny edge detection

CLAHE[2] formula is given by,

CLAHE = X(I,j)-Xmin(I,j)/Xmax(I,j)-Xmin(I,j)

Where ,X is the image,Xmin-Minima of the image,Xmax-maxima of the image.

The main aim of the image enhancement is to improve the contrast and brightness of the image in order to improve the quality of the image.

The image is considered as a function z=f(x,y), it is an 2D matrix.

Where z is the gray level of the image.

f(x,y) = a1*x +a2*y + a3 +e(x,y), --(2)

Calculate the values a1, a2, a3, i.e. $\hat{a}1$, $\hat{a}2$, $\hat{a}3$,

We have to reduce the sum of square of residuals at each pixel,

$$S^2 = \sum_x \sum_y [\hat{a}1 * x + \hat{a}2 * y + \hat{a}3 - f(x,y)]^2, \quad (3)$$

It gives

$$F = \frac{[(\hat{a}1 - a1)^2 \sum_x \sum_y x^2 + (\hat{a}2 - a2)^2 \sum_x \sum_y y^2]/2}{S^2/(n-3)} \quad (5)$$

$\hat{a}1$, $\hat{a}2$, $\hat{a}3$ by

$$\hat{a}1 = \frac{\sum_x \sum_y x * f(x,y)}{\sum_x \sum_y x^2},$$

$$\hat{a}2 = \frac{\sum_x \sum_y y * f(x,y)}{\sum_x \sum_y y^2},$$

$$\hat{a}3 = \frac{\sum_x \sum_y f(x,y)}{\sum_x \sum_y 1}. \quad (4)$$

Thus,

F has an F distribution with 2,n-3 degree of freedom.

When considering 3*3 window, n-3= 6.

Now we derive the following. Change f(x,y) by equation (2)

,

Ie).,

a1*x +a2*y + a3 + e(x,y) ,

$$\hat{a}1 = a1 + \frac{\sum_x \sum_y x * e(x,y)}{\sum_x \sum_y x^2},$$

$$\hat{a}2 = a2 + \frac{\sum_x \sum_y y * e(x,y)}{\sum_x \sum_y y^2},$$

$$\hat{a}3 = a3 + \frac{\sum_x \sum_y e(x,y)}{\sum_x \sum_y 1}. \quad (6)$$

From above equation and noise equation, it drives variances of $\hat{a}1$, $\hat{a}2$, $\hat{a}3$

$$\sigma^2_{\hat{a}1} = \frac{\sigma^2}{\sum_x \sum_y x^2},$$

$$\sigma^2_{\hat{a}2} = \frac{\sigma^2}{\sum_x \sum_y x^2},$$

$$\sigma^2_{\hat{a}3} = \frac{\sigma^2}{\sum_x \sum_y x^2}, \quad (7)$$

the covariance is 0, nN

Now noises are un-correlated for pixels.

From equations (2) (4) (6):

$$S^2 = \sum_x \sum_y e^2(x,y) - (\hat{a}1 - a1)^2 \sum_x \sum_y x^2 - (\hat{a}2 - a2)^2 \sum_x \sum_y y^2 - (\hat{a}3 - a3)^2 \sum_x \sum_y 1.$$

(8)

Now e(x,y) ~ N(0),

$$\frac{\sum_x \sum_y e^2(x,y)}{\sigma^2} \sim \chi^2_n, \quad (9)$$

The χ^2_n → for the chi-squared distribution with n degrees of freedom, n can be calculated as,

$$n = \sum_x \sum_y 1. \quad (10)$$

Because e(x,y) is a normal distribution function, based on the equation (6), $\hat{a}1$, $\hat{a}2$, $\hat{a}3$ gives the normal distribution :

$\hat{a}1 \sim N(a1, \sigma^2_{\hat{a}1})$, $\hat{a}2 \sim N(a2, \sigma^2_{\hat{a}2})$, $\hat{a}3 \sim N(a3, \sigma^2_{\hat{a}3})$, (11)

with the variance given in equation (23), so

$$\frac{(\hat{a}1 - a1)^2}{\sigma^2_{\hat{a}1}} = \frac{(\hat{a}1 - a1)^2 \sum_x \sum_y x^2}{\sigma^2} \sim \chi^2_1,$$

$$\frac{(\hat{a}2 - a2)^2}{\sigma^2_{\hat{a}2}} = \frac{(\hat{a}2 - a2)^2 \sum_x \sum_y y^2}{\sigma^2} \sim \chi^2_1,$$

$$\frac{(\hat{a}3 - a3)^2}{\sigma^2_{\hat{a}1}} = \frac{(\hat{a}3 - a3)^2 \sum_x \sum_y 1}{\sigma^2} \sim \chi^2_1. \quad (12)$$

Following the equations (9), (10), (12) ,

$$\frac{S^2}{\sigma^2} \sim \chi^2_{(n-3)}, \quad (13)$$

U~ χ^2_j , V ~ χ^2_k , then

$$\frac{U/j}{V/k} \sim F_{j,k.}$$

$$\frac{[(\hat{a}1 - a1)^2 \sum_x \sum_y x^2 + (\hat{a}2 - a2)^2 \sum_x \sum_y y^2]/2}{S^2/(n-3)} \sim F_{2,n-3.}$$

(14)

DN (Digital number)

Estimate the Digital Number (DN) as

DN = WF * RV + (1-WF) * DN(old), (15)

RV - > Reference value for the pixels

WF - > Weight vector

WF = max (WF1, WF2) (16)

The contrast of the image is calculated as follows,

(Image contrast enhancement)

$$WCON = \frac{DN\max(window) - DN\min(window)}{DN\max(image) - DN\min(image)}$$

(17)

Iris Localization

Due to computational ease, the image became scaled down via 60%. The picture became filtered the usage of Gaussian clear out, which blurs the image and reduces outcomes because of noise. The diploma of smoothening is determined by using the same old deviation, σ and it is taken to be 2 in this example.

The part of the attention sporting data is handiest the iris part. It lies among the sclera and the student. Consequently the subsequent step is isolating the iris component from the eye image. The iris internal and outer barriers are positioned with the aid of locating the brink picture the use of the canny facet detector.

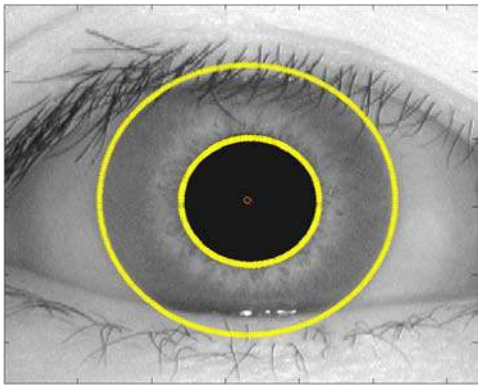


Figure 3: Image with boundaries

The Canny detector mainly entails 3 steps, viz. finding the gradient, non-most suppression and the hysterisis thresholding. As proposed by means of Wildes, the thresholding for the attention picture is finished in a vertical path simplest, in order that the have an impact on due to the eyelids can be reduced. This reduces the pixels at the circle boundary, however with using Hough rework, successful localization of the boundary can be received even with the absence of few pixels. it's also computationally quicker since the boundary pixels are lesser for calculation.

The use of the gradient picture, the peaks are localized the use of non-maximum suppression. it really works inside the following manner. For a pixel imgrad(x,y), within the gradient image, and given the orientation theta(x,y), the brink intersects two of its eight related buddies. The factor at (x,y) is a maximum if its cost is not smaller than the values at the 2 intersection points.

The subsequent step, hysteresis thresholding, gets rid of the vulnerable edges underneath a low threshold, but not if they may be connected to a edge above a excessive threshold thru a series of pixels all above the low threshold. In other words, the pixels above a threshold T1 are separated. Then, those factors are marked as side factors most effective if all its surrounding pixels are extra than some other threshold T2. the brink values were found by way of trail and blunders, and have been obtained as zero.2 and 0.19. Area detection is accompanied via finding the limits of the iris and the pupil. Daugman proposed the use of the Integra-differential operator to stumble on the bounds and the radii. This behaves as a

circular side detector by means of searching the gradient image alongside the boundary of circles of growing radii. From the probability of all circles, the maximum sum is calculated and is used to discover the circle facilities and radii.

The Hough remodel is some other way of detecting the parameters of geometric items, and in this case, has been used to find the circles inside the part image. For each part pixel, the points at the circles surrounding it at extraordinary radii are taken, and their weights are expanded if they're edge factors too, and these weights are introduced to the accumulator array. Thus, in any case radii and area pixels were searched; the most from the accumulator array is used to locate the center of the circle and its radius. The Hough remodel is carried out for the iris outer boundary using the whole image, and then is performed for the scholar simplest, instead of the entire eye, due to the fact the student is usually in the iris.

There are a few issues with the Hough transform. First of all, the brink values are to be discovered by trial. Secondly, it is computationally intensive. This is improved by means of just having 8-manner symmetric factors at the circle for every search factor and radius. The eyelashes have been separated through thresholding, and people pixels have been marked as noisy pixels, considering they do no longer include within the iris.

Iris Normalization

As soon as the iris place is segmented, the next degree is to normalize this component, to allow technology of the iris and their comparisons. By Considering the fact that versions in the attention, like optical size of the iris, position of student within the iris, and the iris orientation trade man or woman to individual, it's far required to normalize the iris photograph, so that the representation is not unusual to all, with comparable dimensions. Normalization manner involves un-wrapping the iris and changing it into its polar equal. it's far done the use of Daugman's Rubber sheet model. The center of the pupil is considered as the reference point and a remapping formula is used to transform the points on the Cartesian scale to the polar scale.

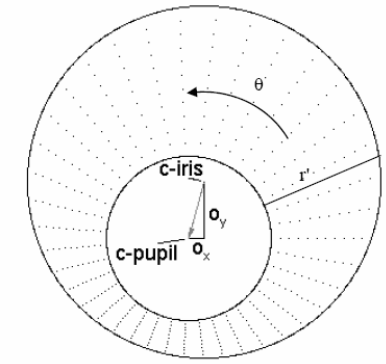


Figure 4: Normalization process

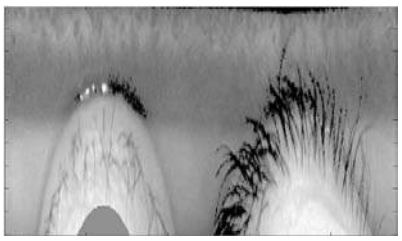


Figure 6: Normalized iris image

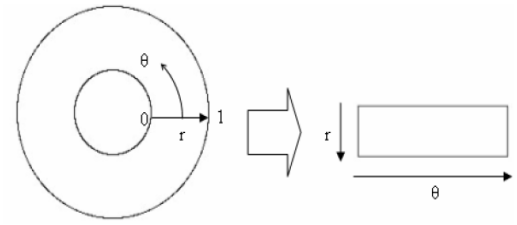


Figure 5: Unwrapping the iris

Encoding

The final technique is the generation of the iris. For this, the maximum discriminating feature in the iris pattern is extracted. The phase facts inside the pattern handiest is used because the segment angles are assigned irrespective of the photograph comparison. Amplitude records are not used since it depends on extraneous elements. Extraction of the segment information, in line with Daugman, is carried out the use of second Gabor wavelets. It determines which quadrant the resulting pharos lies the use of the wavelet:

$$h_{\{Re,Im\}} = \mathrm{sgn}_{\{Re,Im\}} \int_{\rho}\int_{\phi} I(\rho,\phi) e^{-i\omega(\theta_0-\phi)} \cdot e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} \rho d\rho d\phi$$

where, $h_{\{Re,Im\}}$ has the actual and imaginary part, every having the fee 1 or 0, depending on which quadrant it lies in.

A simpler manner of using the Gabor filter out is via breaking up the 2nd normalized pattern into some of 1D wavelets, and then those indicators are convolved with 1D Gabor wavelets.

Gabor filters are used to extract localized frequency records. However, due to some of its barriers, log-Gabor filters are extra broadly used for coding natural photographs. It become suggested by discipline, that the log filters (which use Gaussian transfer features regarded on a logarithmic scale) can code herbal pix higher than Gabor filters (viewed on a linear scale). Information of herbal iris implies the presence of high-frequency additives. Because the regular Gabor filters beneath-represent high frequency additives, the log filters grow to be a better desire. Log Gabor filters are constructed using

$$G(f) = \exp\left(\frac{-(\log(f/f_0))^2}{2(\log(\sigma/f_0))^2}\right)$$

For the reason that strive at enforcing this characteristic become unsuccessful, the gabor- convolve function written by using Peter

Kovesi was used. It outputs a cellular containing the complex valued convolution consequences, of the same length as the enter image.

Iris Image compression

The three colour RGB model is not suited for image processing purpose. To compress the image, the luminance-chrominance values are taken due to the higher value than the RGB colour format. Consequently, RGB 2D data are transformed to one of the luminance-chrominance models, despite the fact that acting the compression method and then transform the 2D signals returned to RGB version because the displays are most usually presents output 2D signals with direct RGB model [3]. The chrominance components represent the colour information in the images [6]. To provide such quality transmission, wireless is more convenient, but is not perfect. There is limitation as well as difficulties such as bandwidth, signal attenuation, co-channel interferences, and time varying channel too. Consequently, the quality of the transmitted images gets degraded [5]. The acquired sign inside the wireless verbal exchange channel is characterised through the joint impact of the two impartial techniques small scale fading because of the contemplated and scattered sign or else due to the shadowing from diverse barriers inside the propagation direction. numerous form of statistical fashions are evolved to examine the result of noising and shadowing separately, but noising and shadowing occurs simultaneously, it is important to convey complex illustration that would be used to model their results simultaneously.

Shannon's Nyquist sampling theorem derives that a data must be sampled at a higher data than double the maxima frequency of the data for reliability of signal reconstruction. For maximum bandwidth images, such as image, the needed sampling rate is very high. Some small coefficients of the transforms like DCT and DWT coefficients less coefficients can be deleted with small Impact the quality of the data significantly. The above basic theory is used in all image compression algorithms [9]. The algorithm of Compression Sensing (CS) is to obtain important information directly without first sampling the data in the old sense. It is shown that if the data is "sparse" of compressible, then they obtained data is enough to recreate the original data with high possibility. The Sparsity is denoted with respect to a suitable source, such as DCT or DWT for the given data [10]. The concept of the CS is also acquired dimension of the data through a development that is incoherent with the data. In CS, a sensing method should give a enough number of CS capacity in a non-adaptive way, so that enable by great reconstruction In the figure 1, the image shows that RGB colour space that the image can be compressed using CS technique to transmit the data through the OFDM transceiver [11].

According to CS technique 3 major steps for CS application:

a) Sparse symbol of the data.

b) Design MxN dimension matrix unconnected to transform root to calculate the data and expand M - dimensional quantity vector.

c) Reconstruction the data by -dimensional quantity vector.

The principle of CS can be described as below:

The following notation as we have $f = \{f_1, \dots\dots\dots, f_N\}$ be real- magnitude samples of a data, which can be correspond to by the transform coefficients, . That is,

$$f = \Psi x = \sum_{i=1}^{N} x_i \psi_i \quad (1)$$

where $\Psi = [\psi_1, \psi_2, \dots\dots\dots \psi_N]$ is an × transform source matrix, which resolve the field where the signal is sparse and too $x = [x_1, x_2, \dots\dots\dots x_N]$ is an - measurement vector of coefficients with $x_i =< x, \psi_i >$. We presume that is S sparse, sense that 3 are only important fundamentals in with S<<N .

STD-BCS-SPL ALGORITHM

The STD-BCS-SPL (Standard Bi-directional Compressive Sensing special) algorithm is based on the concept of the wavelet transform. It restricts the necessity of the arbitrary access of the whole image to small sub images. The STD-BCS-SPL algorithm will work in the principle of partial ordering by the magnitude with a set partitioning sorting method [10].

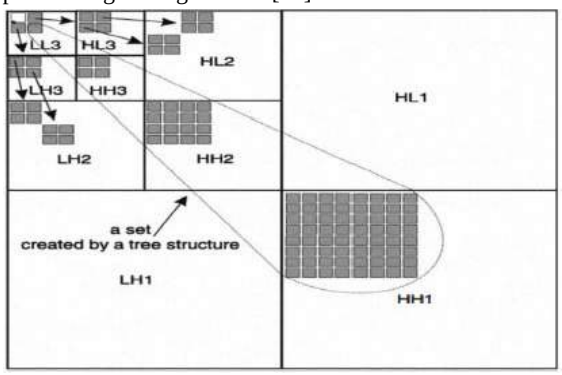


Fig.2: Wavelet sub-bands in STD-BCS-SPL.

The figure above shows a spatial orientation tree and parent-children contribution which is defined by the STD-BCS-SPL algorithm in all the sub-bands in wavelet image. The tree is defined in the way that of each and every node has either no offspring or four offspring at the same spatial location in the four sub-band level. The 2D signalsels which are in the lowest frequency sub-band tree roots grouped into the blocks 2 adjacent 2D signalsel and in each of the block one of the block is marked by star as shown in fig. The STD-BCS-SPL can also describe this type of collocation with one to four parent-children relationships.

SOURCE CODING

STD-BCS-SPL algorithm is used for source coding of a full image. It is used for image compression and wavelet decomposition.

The figure 5 shows the image decomposition analysis using wavelet transform. The decomposition implies the low and high frequency coefficients to estimate LIS, LIP and LSP.

Image decomposition: The proposed method first decomposes a data into coefficients called sub-bands after which the consequent coefficients are evaluated with a threshold. Coefficients beneath the edge are set to zero. The coefficients over the edge worth are encoded with a lossless compression. The first step in STD-BCS-SPL coding which will decompose the original data into wavelet decomposed still image format.

The figure 5 shows the image decomposition analysis using wavelet transform. The decomposition implies the low and high frequency coefficients to estimate LIS, LIP and LSP.

Image decomposition: The proposed method first decomposes a data into coefficients called sub-bands after which the consequent coefficients are evaluated with a threshold. Coefficients beneath the edge are set to zero. The coefficients over the edge worth are encoded with a lossless compression. The first step in STD-BCS-SPL coding which will decompose the original data into wavelet decomposed still image format.

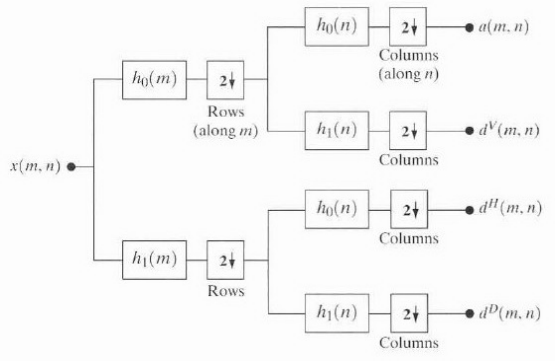


Figure 5: Image decomposition using Wavelet Transform.

CONCLUSION

The private identity method developed via John Daugman turned into carried out, with some adjustments regarding due to processing pace. It's been tested best for the CASIA database picture. Due to computational performance, the search vicinity in multiple components has been decreased, and the elimination of mistakes due to reflections in the eye image has not been applied. Because of unsuccessful strive in the filtering section of the code, a characteristic via Peter Kovesi changed into used. Since the iris for the eye snap shots have been no longer to be had, accuracy of the consequences couldn't be determined. even though, a sample of the iris from John Daugmans papers is presented under.

REFERENCES

[1] John Daugman, University of Cambridge, *How Iris Recognition Works*. Proceedings at International Conference on Image Processing.

[2] C.H.Daouk, L.A.El-Esber, F.D. Kammoun, M.A.AlAlaoui, *Iris Recognition,*

[3] Tisse, Martin, Torres, Robert, *Personal Identification using human iris recognition*, Peter Kovesi, Matlab functions for Computer Vision and Image Processing, *What are Log-Gabor filters ?*

[4] Nor'aini Abdul Jalil, Rohilah Sahak, Azilah Saparon *Faculty of Electrical Engineering,* University Teknologi MARA, 40450 Shah Alam, Selangor "A Comparison of Iris Localization Techniques for Pattern Recognition Analysis", Selangor, Malaysia 2012 Sixth Asia Modelling Symposium.

[5] P. Radu, K. Sirlantzis, G. Howells, S. Hoque, F. Deravi "Are Two Eyes Better than One? An Experimental Investigation on Dual Iris Recognition", 2010 International Conference on Emerging Security Technologies.

[6] Carlos A.C.M. Bastos, Tsang Ing Ren and George D.C. Cavalcanti, " Analysis of 2D log-Gabor filters to encode iris patterns," 2010 22nd International Conference on Tools with Artificial Intelligence.

[7] Jaehan Koh, Venu Govindaraju, and Vipin Chaudhary "A Robust Iris Localization Method Using an Active Contour Model and Hough Transform", 2010 International Conference on Pattern Recognition.

[8] Chai Tong Yuen,Saied Ali Hosseini Noudeh, Mohammad Shazri and Mohamed Rizon, "A Fusion Technique for Iris Localization and Detection", 2010 International Conference on Technologies and Applications of Artificial Intelligence.

[9] Sharat Chikkerur_, Venu Govindaraju, and Alexander N. Cartwright," Fingerprint Image Enhancement Using STFT Analysis", Springer-Verlag Berlin Heidelberg 2005

[10] Shimna. P. K,Neethu. B," Fingerprint Image Enhancement Using STFT Analysis", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 2, Ver. III (Mar - Apr.2015), PP 61-68.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

VIRTUAL LEARNING ENVIRONMENT

Ms. Kanimozhi
Assistant Professor, Department of Computer Science & PG department of IT,
Dr. MGR Janaki College of Arts and Science College for Women. Chennai, India
Kanimozhi135@gmail.com

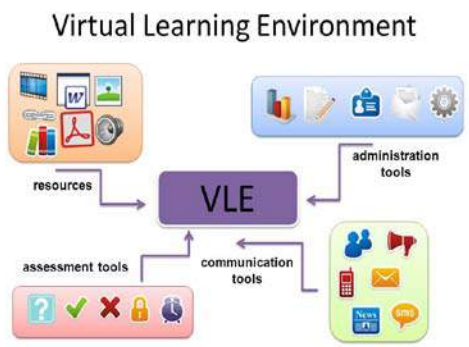
ABSTRACT

Virtual learning uses computer software, the Internet or both to deliver instruction to students. This minimizes or eliminates the need for teachers and students to share a classroom. Virtual learning does not include the increasing use of e-mail or online forums to help teachers better communicate with students and parents about coursework and student progress; as helpful as these learning management systems are, they do not change how students are taught.

Keywords : Computer Based, Internet Based, Remote Teacher Online, Blended Learning

INTRODUCTION

Virtual education refers to instruction in a learning environment where teacher and student are separated by time or space, or both, and the teacher provides course content through course management applications, multimedia resources, the Internet, videoconferencing, etc.Personal computers and the Internet have revolutionized entire sectors of American society. Facebook, Twitter, YouTube, Skype and other online communications media have allowed billions of people around the world to share ideas in a matter of seconds, mostly at a very low cost. These advances in computer technology are as remarkable as they are familiar.But most people are not aware of how computers and Internet technology are transforming the way students learn. This emerging education paradigm is often called "virtual learning," and it has the potential to improve student achievement, educational access and schools' cost-effectiveness. Specifically, virtual learning uses computer software, the Internet or both to deliver instruction to students. This minimizes or eliminates the need for teachers and students to share a classroom. Virtual learning does not include the increasing use of e-mail or online forums to help teachers better communicate with students and parents about coursework and student progress; as helpful as these learning management systems are, they do not change how students are taught.



Virtual learning comes in several forms:

- ***Computer-Based****:* Instruction is not provided by a teacher; instead, instruction is provided by software installed on a local computer or server. This software can frequently customize the material to suit the specific needs of each student.
- ***Internet-Based****:* This is similar to *computer-based* instruction, but in this case, the software that provides the instruction is delivered through the Web and stored on a remote server.
- ***Remote Teacher Online****:* Instruction is provided by a teacher, but that teacher is not physically present with the student. Instead, the teacher interacts with the student via the Internet, through such media as online video, online forums, e-mail and instant messaging.
- ***Blended Learning****:* This combines traditional face-to-face instruction, directed by a teacher, with *computer-based*, *Internet-based* or *remote teacher online* instruction. In effect, instruction comes from two sources: a traditional classroom teacher, and at least one of the forms of virtual learning described above.
- ***Facilitated Virtual Learning****:* This is *computer-based*, *Internet-based* or *remote teacher online* instruction that is supplemented by a human "facilitator." This facilitator does not direct the student's instruction, but rather assists the student's learning process by providing tutoring or additional supervision. The facilitator may be present with the learner or communicating remotely via the Web or other forms of electronic communication.
- ***Online Learning***: This is any form of instruction that takes place over the Internet. It includes *Internet-based* instruction; *remote teacher online* instruction; and *blended learning* and *facilitated virtual learning* that involves these two virtual learning methods. It excludes *computer-based* learning.
- ***Full-Time Online****:* This is online learning with no regular face-to-face instruction or facilitation. It is *Internet-based* and *remote teacher online* learning only, though it may include some occasional interaction with human teachers and facilitators.

PURPOSE Of VLE:

- Our society is changing. A new paradigm of education is developing, one that integrates the technology of computers and the Internet in education. We do not only learn from books. We have many technological tools available to us. The use of computers, and especially the Internet, opens a new world of potential. With the use of technology, education can surpass the physical boundaries of the classroom and provide students the opportunity to experience more. Since Gutenburg, the Internet represents the largest transfer of information to occur in history. According to Robert B. Cummings, Director Learning Resources Center, SHRP-SON at University of Alabama at Birmingham market research indicates that we can make the following assumptions:
- 50% of learning will continue to be "in person", involving things only available in person, although most of this activity will be facilitation • 50% of learning will take place on the Internet, which is a better vehicle for cognitive learning due to the extent of information, low cost, and convenience.
- Employers will expect to hire people who know how to learn on-line .
- Education will become more student oriented (convenient), rather than faculty oriented
- Internet will dominate teleconferencing, because it's cheaper (lower technological investment) than video codecs, offers universal access, and has a high level of interactivity.
- Personal computers will be ubiquitous.Following the emergence of the Internet in the early 1990s, many new tools and products have been developed to exploit its benefits fully. Since the mid-1990s the Virtual Learning Environments (VLEs) have appeared with the aim of supporting learning and teaching activities across the Internet. Traditionally the school has been the place where teachers and pupils meet each other. It has been the setting where the institutional teaching/learning process takes place. However, various forms of computer-mediated communication are adding interesting new dimensions to regular school learning. The Internet offers such advantages as flexible access and new ways of communicating and assessing for students and teachers. The Internet also has some disadvantages such as reliance of information service providers, viruses and low speed of connections. However, for the teacher, creating Internet resources that are stimulating, appealing, easy to use and educationally sound is time consuming. The VLEs allow teachers to create resources quickly and without the need to develop technical skills. VLEs provide an integrated set of Internet tools, allow easy upload of materials and offer a consistent look and feel that can be customized by the user.

REASONS FOR VLE

Communication – opens up an infinite number of channels in the format of forums, discussion threads, polls, surveys – instant feedback either as a group or individually

- **Producing work** – students do not physically have to find their teacher to hand in work due to secure virtual 'hand-in' folders that have time windows

- **Resource hub** – teachers have infinite online storage space for ppts, docs, worksheets etc. that can either be secure or shared with students

- **Dynamic home pages** – teachers have the opportunity to create an exciting virtual space to represent their room/subject

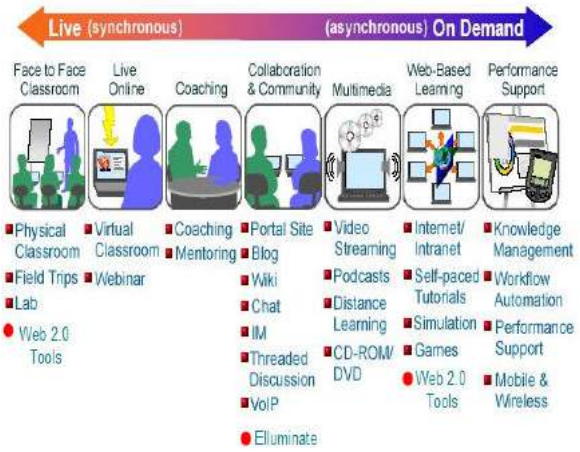
- **Links to outside sources** – pathways to all other online learning spaces are linked via the VLE

- **Embedded content** – YouTube, BBC, newspapers can all be embedded as the dynamic feed of the homepage

- **Podcasts & videos** – both teacher- and student-produced podcasts and videos have a shared platform; again, either secure or shared.

VLE TOOLS

Most virtual-learning technologies fall into three broad categories. These are not precise divisions — technologies and functionalities overlap as each category evolves.



Lecture capture:

These technologies have come a long way from their roots in rough audio and videotape recordings of class sessions. Lecture capture system (LCS) use involving hardware and software tools to record fully integrated presentation began at colleges and universities, but K–12 districts are catching up.

Lecture capture can stretch teaching resources and enrich the curriculum by building a growing store of reusable digital resources. The Cornwall–Lebanon School District in Lebanon County, Pa., adopted an LCS two years ago as a way for its more than 4,700 students to both receive information and create presentations.

An LCS records every aspect of the speaker's presentation, including all the additional materials, such as Microsoft PowerPoint slides, interactive whiteboard annotations or output from a document camera. The recordings are then edited and annotated to create rich, complex presentations for asynchronous viewing by students. Many lecture capture systems also stream live audio and video, providing remote real-time access to the presentation.
Lecture capture lets students catch up on or review class content whenever it's convenient for them. The edited recording can be integrated into a virtual-learning environment and thus becomes a component of a fully online or blended course.

In a software-based LCS, an agent is downloaded on the presenter's computer, which is networked with the other hardware (microphone, video camera and interactive whiteboard) used for the session. The software agent integrates the output from the various tools, including keystrokes on the speaker's computer.

When the edited recording is complete, the LCS automatically distributes a link to students registered in the course and others on a predetermined distribution list. Teachers can also release the lectures on a set schedule. Many systems include tools that promote student interaction, such as polls or requests for responses to the captured content. Results of the polling and student commentary are then integrated into the presentation. Current LCS systems offer high-definition recording and playback at a pixel resolution of 1920x1200 or better.

Webinars:

These interactive online presentations are usually delivered first in real time and then recorded and made available for review or first-time viewing by a new audience. In K–12 districts, webinars are most often used as training vehicles for instructors, though creating a webinar is a common assignment for students in virtual courses in the higher grades. With their highly structured format, webinars offer an excellent platform to focus or expand on important topics.
Most webinars consist of PowerPoint slides that are accompanied by audio explanation by the teacher. Audio is delivered over a standard phone line or streamed via Voice over Internet Protocol (VoIP). Using remote desktop sharing, teachers can talk students through complex topics while using a variety of tools and applications to display information on their computer screens.
The technology needed to support a webinar varies with the technical complexity of the presentation. Webinars work best if everyone in the audience has a high-speed Internet connection. There are many stand-alone software offerings that let schools or instructors create and deliver webinars; that functionality is also available in many course or learning management systems. Hosted webinar applications are also available as cloud services.

Interactive web conferencing: This technology takes many forms, but the main focus is on two-way communication over a distance, with the Internet providing the link between locations. Interactive web conferences can range anywhere from an online chat about homework to a lecture delivered via telepresence. But even in its most basic forms, these technologies deliver real-time interactivity over distance.
School districts often use interactive web conferencing to extend the geographic reach of classes. Web conferencing can let a teacher or expert speaker deliver a lecture simultaneously to multiple classrooms anywhere in the world and respond in real time to questions from students at all locations. The North Slope Borough School District, which is administered from Barrow, Alaska, but includes seven villages spread over 88 square miles, uses video conferencing to deliver courses from Barrow to the secondary schools in each small community.
The requirements for the most basic forms of interactive web conferencing are pretty simple — a software application and an Internet connection. Some districts use web conferencing for virtual-learning courses, virtual review sessions for traditional or blended classes, or collaboration among teachers or students at separate sites.

Positive aspects of virtual classroom/e-learning:

"Independence and time Management"

Students who take courses online often sharpen their ability to work on their own, and they also expand experience in managing their time efficiently. With nobody to stand over them and make them work, virtual learners tend to develop these skills more rapidly than if they were to learn strictly in a traditional classroom.

"Advanced and specialized classes"

In many cases, small school and rural schools simply cannot offer advanced or specialized classes. Virtual education gives students the opportunity to gain experience in areas that would otherwise remain out of reach.

"Emphasis on the written work"

Strong writing skills are essential to success in secondary and higher education as well as in the workplace. Virtual learning/distance education teaches students to communicate more effectively through writing, because the questions they ask and the work they complete is based almost solely on the written word. Virtual learning clearly gives students the chance to widen writing skills.

"Knotty aspects of virtual/distance education"

"Lack of face-to face interaction"

Some educators dispute that both teacher-student and student-student contact are essential to the learning process, and online classes eradicate these elements of education altogether. Also, students who excel at class participation need to deem that this piece of the learning puzzle will be missing as well.

"The need to self-start"

Virtual education actually lets students make their own schedules. Those who have a hard time with self-motivation will undoubtedly have problem in this type of educational setting"

CONCLUSION

Virtual learning environments can provide relevant and rewarding experiences. Although currently underused in some curriculum areas, particularly the arts, new technologies will provide more effective means of delivery. Many emerging technologies and networks can be used to enrich and provide greater interactivity within the virtual learning environment. Advances in technology ensure that almost all traditional classroom equipment can be emulated in the virtual learning environment.

The future of virtual learning environments has many innovative and exciting possibilities. New networks can allow students more opportunities way beyond those offered by the Web in its current state but careful planning and innovation will be required to ensure

that the potential for the scope of delivery is reached. The importance of mobility should also be considered so that learning can take place in the most appropriate context. If issues of cost and programming were resolved students would be given access to the range of additional hardware and software required.

One of the main disadvantages of the virtual learning environment is the lack of face-to-face personal interaction and the student social contact, which traditional educational contexts provide. It is because of these factors, and the lack of evidence of how they will impact on student personal and social development, that virtual learning environments may not entirely replace traditional classrooms and teacher pupil contact.

REFERENCES

[1] Britain, Sandy; Liber, Oleg (1999). "A Framework for Pedagogical Evaluation of Virtual Learning Environments" (PDF). JISC Technology Applications Programme (Report 41). Archived from the original (PDF) on 2014-06-14. Retrieved 1 February 2015.

[2] Davis, C. (April 2014). Virtual Learning Rubric. Retrieved from http://www.doe.mass.edu/odl/standards/VLPrubric.pdf

[3] Holyoke, M (2011), "What is virtual learning environment (VLE) or managed learning environment (MLE)", WhatIs.com

[4] Posey, Burgess, Eason, & Jones. "Advantages and Disadvantages of the Virtual Classroom and the Role of the Teacher" (PDF).

[5] https://www.researchgate.net/

[6] First Virtual Communications, Inc (2001) [WWW document] URL http://www.cuseeme.com/ (visited 2 January, 2002)

[7] Flickerman, R (2001) NoCatAuth:Authentication for Wireless Networks [WWWdocument] URL http://oreilynet.com/pub/a/wireless/2001/11/09/nocatauth.html (visited 7 December, 2001)

[8] Knowledge Media Institute, The Open University (2000), Open University of the UK's Knowledge Media Institute's Stadium. [WWW document]URL http://kmi.open.ac.uk/stadium/welcome.html (visited 20 December, 2001)

[9] Microsoft Corporation (2001) NetMeeting Home [WWW document] URL http://www.microsoft.com/windows/netmeeting/ (visited 2 January 2002)

[10] NCSA (1994) NCSA Collage [WWW document] URL http://archive.ncsa.uiuc.edu/SDG/Software/Brochure/UNIXSoftDesc.html (visited 20 December, 2001)

[11] PlaceWare Web Conferencing Provides Live (2002), Interactive Business Meetings and Presentations Over the Internet [WWW document] URL http://www.placeware.com/

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

REAL TIME BUS TRACKING SYSTEM

G. Gayathri,
Dept. of Information Science & Technology ,
1stYear M.C.A., Anna University,
College of Engineering Guindy
Chennai, India
Email:demand4great@gmail.com

V. Vidhya,
Asst. Professor,
Dept of Computer Applications,
Shri. SSS Jain College for Women,
Chennai, India.
Email:vidhya2982@gmail.com

V. Vinodh
Management Consultant,
London,United Kingdom
+447449790796
praveenvino@gmail.com

ABSTRACT

When it comes to taking the public transportation, time and patience are of essence. In other words, many people using public transport buses have experienced time loss because of waiting at the bus stops. So, we require one tracking system to track the Complete Transport System. Every GPS tracking system is a common approach to get vehicle location information in real- time. The system includes a GPS/GPRS module for location acquisition and message transmission, AT&T's cellular data service to transfer of location information. It will show the correct position of the vehicle to the user on the basis of the location information sent by the GPS Device.

Keywords: GPS, AT&T, AVL Database, AVL, EAT, LED, LCD, Vehicle Tracking.

1. INTRODUCTION

Real-time vehicle tracking and management system has been the focus of many researchers, and several studies have been done in this area. In this project, the main area of concentration is tracking buses for which people were waiting for a long time. People have to know where the bus is at present and the time of the bus to reach bus stop. It will help the passengers to track the vehicles, to get real time position of the vehicles, changed routes (If any); it can also act as an anti- theft application by detecting the exact position of the vehicles[1].

Real time tracking is becoming more and more popular as devices utilizing the Global Positioning System (GPS) become more readily available. In this system, using AT&T's cellular data service the data will be sent by the buses with their coordinates among other data. This data allows the dispatcher to know where all the buses are at any given time[2][3].

The proposed system will show user the real time location of the vehicle on the Google Map by using GPS (Global Positioning System) & AT&T.

The application will ask the user (Passenger) to enter the bus number in which he/she wishes to travel. Then the user will enter the source and the destination of their journey. After entering all the necessary information, the user will click on the locate button. Upon clicking the locate button, the user will get all the detailed information about the location of the bus. By using this project passengers will be able to easily access data related to the bus that they are interested.

2. PROPOSED APPLICATION

The system involves many different parts that worked together to accomplish the requirements. The source of all the data that is presented originates in the AVL Database. This database is where the buses send data every 30 seconds. The buses send information such as longitude, latitude, heading, route, and speed. From there the data is accessed on two different web servers. The estimated arrival time server grabs the data and does calculations on each bus and outputs a comma separated string. The AVL data is mapped to the buses on a real time Google map by the second server.

[4]The second server also communicates with the estimated arrival time server to request the ETA string and parse it to display in a table

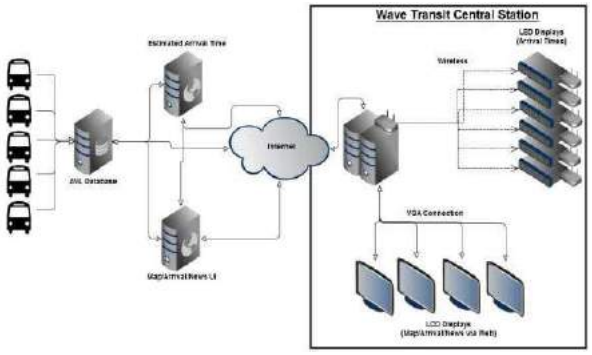


Fig 1:Architecture Diagram

that is presented to the user. There is a Java application running on the inside
Server that makes a request for the ETA string, parse it, and sends the times to each specific LED display, mounted above each bus stop. The arrival time will be sent by the Java application's sockets to each LED display individually. The communication uses

the wireless local area network at the station. Two PCs control two of the liquid crystal displays (LCD) via dual video graphics array (VGA) ports.

The mobile friendly version has a very simple user interface to cut down on load times. It has the option to view a table which is a text based display of a specific route or view a map of a specific route. A map with all buses was left out of the mobile version because it was unusable on a small screen, due to the large number of buses to display. Along with all these improvements a simple implementation of Google's direction service was used to give the user an ETA for the bus to arrive at its next stop. This information includes with the route name, heading, speed, and next stop on all versions of the map along with the table display[3].

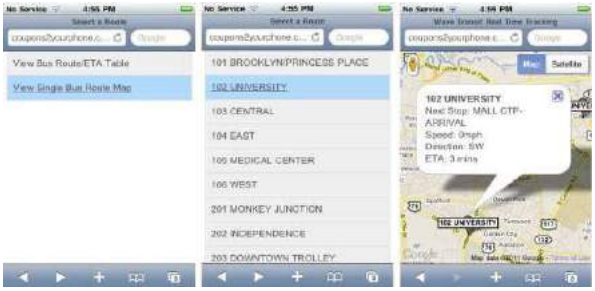


Fig 2. Final Mobile Implementation

3.LED PROGRAM

The Java program will be called every twice minute by the LED Java code that resulted in a shell program which is done in the final implementation. CRON must call the Java program as it has a minimum scheduling resolution of 1 minute for every 30 seconds[4].The ETA values will be called with the help of Mobile Education's server for every sixth route which sends a command to the display to show that routes name and ETA. Each bus has a specific LED sign that it stops under which has a static IP address. The Java program the IP addresses are hardcoded to display appropriate information on the correct LED sign[4][5].

Fig3. Countdown Sign

4. DESIGN

There were few issues that arose when going over the project requirements[6]. The first issue to overcome was the programming language; lack experience in Java Server Pages presented issues that slowed down the projected implementation of this project. This was a requirement so that the system can be maintained in the future by Dr. Vetter. JSP was released in 1999 by Sun as a direct competitor to ASP and PHP. JSP helps to bridge the gap between java and the web. Throughout the project the syntax and integration with java presented a steep learning curve. I was able to become proficient in JSP by learning the best practices, so now it can be used in future.

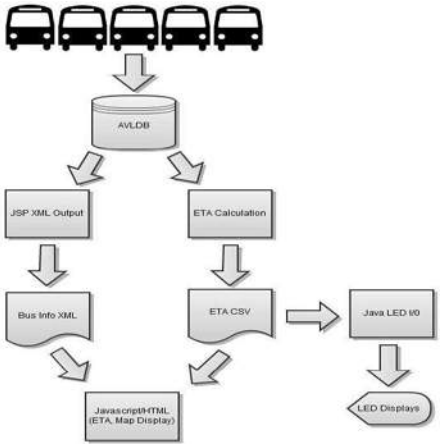


Fig 4: Data Flow Diagram

The major issue was to present the data to the user. At the bus station the data on the page changes without refreshing the entire page. Previously a meta refresh was used to update the information. That wasn't acceptable because it would cause a noticeable blink of the user interface.[5]The only possible way around this was to change the data using a client side language like JavaScript as JSP is server side. Also, the displays are all 1o8op so all of the interfaces needed to be optimized for that resolution. The public site should be accessible to all resolutions which means it cannot be optimized for the more common resolution. The news page to arrive at its next stop. This information is available along with route name, heading, speed, and next stop on all. was merged with the arrivals display in prototype 3 so, this concern was eliminated.

6.SOFTWARE AND SYSTEM QUALITY METRICS

The international Organization or Standardization defines quality as "the totality of characteristics of an entity to bear on its ability to satisfy state or implied needs". This definition itself is very vague, but in essence the quality will be measured by how well it conforms to the requirements and that it can be used as it was intended[3]. In other words this protect would be considered successful if it meets the scope in the allotted time, satisfies the customer and reaches the ultimate goal of providing a benefit to the passengers of Wave Transit. Meeting these big goals the system must be ensured to met the requirements set forth by Wave Transit.

The software was developed with good quality that includes readability, maintainability, low complexity and robust error handling because readability and maintainability go hand in hand, the software was written in a way that another individual can easily understand and update it if necessary[2]. The software was tested module by module each focused on a particular task to have the least complexity possible. The robust error handling system was enforced which was capable of running the entire life time that provides the user with a system and the tools which operates successfully.

5. FUTURE SCOPE

In this system, there is a possibility of the overall system malfunction due to a particular type of attack, it is termed as Denial of Service (DoS) attack by malicious agents who might try to disrupt the function of the system. A Distributed Security Scheme for AdHoc Networks can be used and to prevent this kind of attack. Such methodology will be analyzed to make this Real Time Passenger Information System more robust. A novel data hiding technique, based on Steganographic mechanism can also be used for security purposes. Here, the advantage lies in the fact that computationally costly encryption-decryption mechanism is avoided, thus making it suitable for a heterogeneous combination of processing elements, which are being used in present system. Here, many processing elements e.g. Mobile phone etc. lacks the processing
power and battery power, which is required for traditional encryption-decryption system.

7. CONCLUSION

In this paper, the partial implementation detail of Real Time Bus Tracking was stated. This system tracks the current location of all the buses and estimates their arrival time at different stops in their respective routes. Estimates are updated every time the bus sends an update and the information is passed to the passengers by display terminals at bus stops, web based GUI and smart phone application which is android based. This research serves the needs of passengers, vehicle drivers and administrators of the transport-system. With the help of GPS and the ubiquitous cellular network, real time vehicle tracking for better transport management has become possible.

REFERENCES

1. "MTA BusTime." Metropolitan Transportation Authority, March 12, 2011.
2. "Extreme Programming." Wikipedia, The Free Encyclopedia. March 9, 2011.
3. "Organizational Structure & Analysis." *Wave Transit - Wilmington,NC*. TJRAdvisors, 12/2009. Web. 14 Mar 2011.
4. "Spiral Model." Wikipedia, The Free Encyclopedia. March 8, 2011.
5. "Online CS Modules: The Spiral Model." N.p., n.d. Web. 14 Mar 2011.
6. Melanson, Donald. "Brooklyn bus riders get real-time bus tracking via cellphone." *Engadget*. N.p., 05 02 2011. Web. 15 Mar 2011.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A SURVEY ON WASTEWATER TREATMENT (WWT) ANALYSIS USING VARIOUS TECHNIQUES

Dr. C. Victoria Priscilla ,
Associate professor & Head,
P.G. Department of computer science,
S.D.N.B Vaishnav College for women,
Chennai, India

A. Anusuya,
Student, P.G. Department of computer science,
S.D.N.B Vaishnav College for women,
Chennai, India
email:anusuyaarumugam1@gmail.com

ABSTRACT

Wastewater is contaminated water that has been affected by human use and also causes environmental pollution. Waste Water Treatment (WWT) is a process of removing contaminants from wastewater and reuses the water in various applications such as Hydroelectric Power Generation, Agriculture, and Radioactivity etc. The developments of various techniques for WWT have been implemented by many researchers. The choice of the suitable treatment techniques is dependent on the wastewater pollutant concentrations such as Biochemical oxygen demand (BOD) and Chemical oxygen demand (COD). This paper explores various techniques like Data Mining, Machine learning, Principal Component Analysis (PCA), Support Vector Machine (SVM), Regression Trees (RT) and provides the estimation of the wastewater quality characteristics.

Keywords: Waste Water Treatment; Data Mining; Machine learning; SVM; PCA; RT;

INTRODUCTION

Wastewater is used water which includes substances such as human waste, solid waste, chemicals, oils, and storm sewer. When wastewater is not properly treated it causes health problems and environmental pollution. [2] In every year 1.8 million people die due to waterborne disease. So the permanent solution is to reduce the threat of water pollution in wastewater treatments. Wastewater treatment is an essential scheme which has to be taken more seriously and also make improvements in our society and future generation. Wastewater management with water and nutrient loop is shown in figure1.Wastewater treatment is a process used to convert wastewater into an effluent which can be returned through the water cycle with minimum contact with the environment or directly reused. It is also called as water reclamation. Place of wastewater treatment is called as wastewater treatment plant (WWTP), repeatedly referred as Water Resource Recovery Facility (WRRF) or Sewage Treatment Plant. Wastewater Treatment Methods are categorized into three subdivisions such as Physical, Biological, and Chemical. Various types of treatment plant Help to reduce pollutants in wastewater. Types of Wastewater Treatments are shown in figure 2.

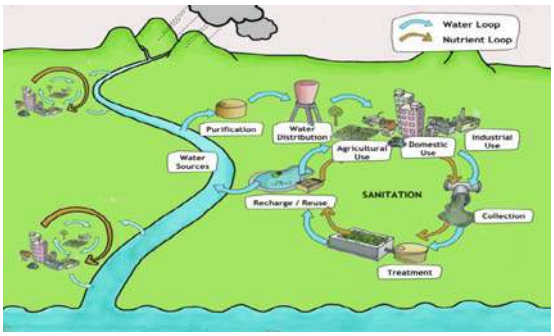


Figure 1: Waste Water Management [1]

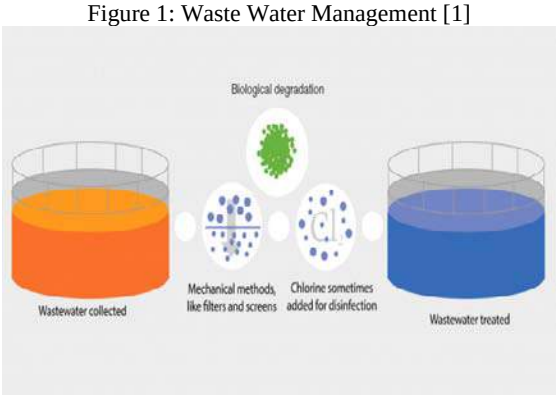


Figure 2: Types of Wastewater Treatments [3]

A. Types Of Wastewater Treatments Process:

a. Physical Water Treatment Process [4]:

- Screening used to remove large suspended solids.
- Sand Filtration is used to filter small suspended solids and dissolved solids.
- Membrane Filtration is used to remove total suspended solids and total dissolved solids.

b. Chemical Water Treatment Process [5]:

- Chemical Precipitation is used to remove dissolved metals in wastewater.
- Chemical Coagulation is used to destabilizing wastewater particles.
- Chemical Oxidation involves adding or generating oxidants in wastewater.
- Ion Exchange is used to soften the water.
- Chemical Stabilization is the process where sludge is mixed with lime to raise the pH to the high level (>12).

c. Biological Water Treatment Process [6]:

- Biological Anaerobic Wastewater Treatment process is used to maintain the organisms in the absence of oxygen.

- Biological Aerobic Wastewater treatment process is used to remove up to 98% of organic contaminants in the presence of oxygen. It is classified as more process as follows:
  - ❖ Activated Sludge Process (ASP) highly concentrates on the degradation of microorganisms and also removes nutrients from wastewater to produce a high-quality outflow.
  - ❖ Trickling Filters are used to remove compounds in the water such as ammonia after primary level treatment.
  - ❖ Aerated Lagoon Process (ALP) introduces oxygen into the oxidation pond in wastewater.
  - ❖ Oxidation Pond contains partially treated wastewater and also reduces the growth of microorganisms such as algae and bacteria.

B. Types Of Wastewater Treatments Plants [7]:

- Effluent Treatment Plants (ETP) are mostly used by various leading companies such as pharmaceutical, chemical industry, Leather industry, steel industry, and tanneries.ETP process used to purify water and removes toxic and non-toxic materials or chemicals from wastewater. It provides environmental protection for various leading companies.

- Sewage Treatment Plants (STP) is used to remove contaminants from wastewater and produce a waste stream and a solid waste. It uses physical, chemical, and biological processes to remove physical, chemical, and biological contaminants in wastewater.
- The Ministry of Environment & Forest, Govt. of India has launched a new scheme called Common Effluent Treatment Plant (CETP) and provides more benefit to industrial technologies.

C. Various Treatment Levels Of ETP [8]:

- Preliminary Level used to do physical separation operations such as screening, sedimentation, and clarification.
  - ❖ Screening: A Screen opens with uniform size used to remove large solids.
  - ❖ Sedimentation used to remove suspended solids from water.
  - ❖ Clarification used for separation of solids from fluids
- Primary Level used to remove suspended and settleable materials. It uses both physical and chemical method treatments.
- Secondary Level uses both Biological and chemical method treatments.
- Tertiary (advanced) Level used for the final cleaning process that improves water quality before recycled, reused or discharged to the environment.

DATA MINING TECHNIQUES:

Francesco et.al [9] has proposed two models, Support Vector Regression (SVR) and Regression Trees (RT) to estimate Wastewater Quality Indicators (WQI). Wastewater Quality indicators (WQI) were biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), and total dissolved solids (TDS) in WWT. Support Vector Regression (SVR) and Regression Trees (RT) provide robustness, reliability and high capability. SVR is same as Support Vector Machine (SVM) which used to analyze non-linear input data. Regression Trees (RT) is used to predict an outcome of real-valued functions. Both machine learning models are compared with statistical measures which are correlation coefficient (R^2) and root-mean-square-error (RMSE) that is shown in table 1.

Table 1: Comparison between SVR and RT [9]

Statistical Measures	**RMSE**		**R^2**	
WQI	**SVR**	**RT**	**SVR**	**RT**
TSS	**1049**	3486	**0.97**	0.906
TDS	**1549**	1826	**0.851**	0.828
COD	**1172**	2395	**0.893**	0.784
BOD	**104**	103	**0.87**	0.871

Finally, the result shows that SVR has better performance than RT in forecasting water quality indicators in WWT.

Andrew et.al [10] proposed Data-Mining models such as Support Vector Machine (SVM), Adaptive Neuro-Fuzzy Inference System (ANFIS), neural network (NN), k-nearest neighbor (K-nn) and Random Forest Tree for methane production in WWT. Biological anaerobic wastewater treatment process is used here. The Architecture of ANFIS is shown in figure 3. Both models are compared with statistical measures like percentage error (PE), fractional bias (FB), root mean square error (RMSE), normalized mean square error (NMSE), and index of agreement (IA) which is shown in table 2.

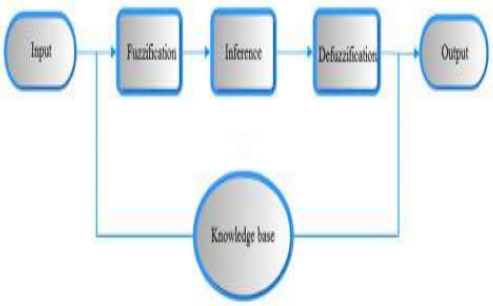


Figure 3: Architecture of ANFIS [10]

Finally, the result shows that Adaptive Neuro-Fuzzy Inference System (ANFIS) model provides better performance for methane production.

Table 2: Statistical metrics for all algorithms [10]

Models	PE	FB	RMSE	NMSE	IA
ANFIS	**0.08**	**0.00**	**74144**	**0.01**	**0.99**
NN	0.16	0.07	144787	0.04	0.99
SVM	0.10	0.03	84089	0.01	0.99
Random Forest Tree	0.10	0.06	92629	0.02	0.99
K-nn	0.13	0.02	123363	0.03	0.99

Hamed et.al [11] has proposed two models, ANN and ANN with PCA and two methods, multilayer perceptron (MLP) feed-forward

neural network and stop training method used to predict industrial wastewater. Principal Component Analysis (PCA) is used to modify and improve the performance of the neural network and also used for reduction of dimensionality. This model used to estimate pH, Chemical Oxygen Demand (COD), Total Dissolved Solids (TDS), total nitrogen (Total N) and total phosphorous (Total P). Both models are compared with statistical methods such as Mean Absolute Error (MAE), determination coefficients (R2) and (RMSE) root mean square error which shown in figure 4.

Davut et.al [12] proposed Intelligent model which is based on wavelet packet decomposition, entropy and the neural network is used to predict the total suspended solids (TSS) in wastewater treatment. Wavelet packet decomposition is used to reduce the input vectors dimensions of the intelligent model. Structure of Intelligent model is shown in figure 5. Wavelet expansion function, wavelet basis function, mother wavelet function, scaling function and J-level wavelet decomposition were used in wavelet transforms. Wavelet packet and NN structure used to compose two layers such as wavelet packet layer and multilayer perceptron layer. Wavelet packet layer has two stages that are wavelet packet decomposition and wavelet entropy. Multilayer perceptron layer used to classify the features of wavelet packet layer. Wavelet packet and NN structure combine as WPNN.

Model	Parameter	Target		PCA-ANN			ANN		
	COD		Data	R^2	RMSE	MAE	R^2	RMSE	MAE
	pH								
Model 1	TDS	pH	All	0.76			0.63		
	N		Train	0.8	0.1569	0.1214	0.74	0.1850	0.1348
	P		Test	0.71	0.2606	0.2107	0.59	0.2240	0.1773
	COD		Data	R^2	RMSE	MAE	R^2	RMSE	MAE
	pH								
Model 2	TDS	TDS	All	0.72			0.68		
	N		Train	0.82	0.99e3	0.76e3	0.80	1.07e3	0.83e3
	P		Test	0.40	1.86e3	1.63e3	0.14	1.85e3	1.49e3
	COD		Data	R^2	RMSE	MAE	R^2	RMSE	MAE
	pH								
Model 3	TDS	N	All	0.84			0.76		
	N		Train	0.90	2.7207	3.4785	0.80	4.0307	2.9149
	P		Test	0.62	4.0246	4.9533	0.78	5.4452	4.3869
	COD		Data	R^2	RMSE	MAE	R^2	RMSE	MAE
	pH								
Model 4	TDS	p	All	0.7			0.66		
	N		Train	0.80	0.2479	0.1925	0.74	0.2676	0.2074
	P		Test	0.59	0.2828	0.2189	0.38	0.2796	0.2192
	COD		Data	R^2	RMSE	MAE	R^2	RMSE	MAE
	pH								
Model 5	TDS	COD	All	0.59			0.53		
	N		Train	0.71	43.673	32.706	0.64	43.765	31.8084
	P		Test	0.44	48.164	36.788	0.50	50.026	37.7631

Figure 4: Statistical parameters of various models [11]

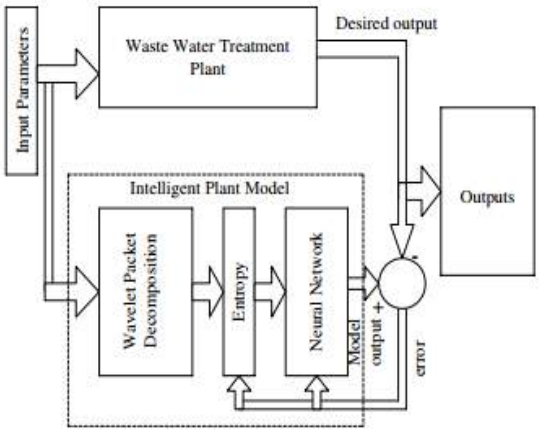


Figure 5: Structure of Intelligent model [12]

WPNN model used 50 different input data for training and 12 input data for testing. Finally, the result shows that WPNN model provides effectiveness and reliability in extracting features of input data.

Djeddou et.al [13] has proposed ANN model for prediction of sludge Volume Index (SVI) of activated sludge process (ASP) to avoid high non-linearity and non-uniformity task issues in Municipal Waste Water.ANN used a standard feed-forward back propagation neural network with three layers are input layer, one hidden layer, and an output layer with Log-sigmoid transfer function. Fifteen variables were used as input parameters. These variables were TSS, COD, BOD, pH, conductivity, NH^+_4, NO_3, P,TSS efficiency, COD efficiency, BOD efficiency, N-NH^+_4 efficiency,N-NO_3 efficiency, P-PO_4efficiency.Kolmogrov theorem used to determine no of hidden layer and no of hidden nodes (NHN).ANN model is evaluated through statistical methods like correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE) and mean absolute error (MAPE) that is shown in table 3.ANN model 15-13-1 has one input layer with fifteen input variables, one hidden layer with thirteen nodes and one output layer with one output variable. Finally, the result shows that ANN model 15-13-1 is a useful tool for activated sludge process (ASP) and also increases the performance of Waste Water Treatment Plant.

'Table 3: Performance of ANN prediction models [13]

ANN model	$R_{TRAINING}$	R_{ALL}	MAE	RMSE	MAPE (%)
ANN 15-5-1	0.9490	0.8091	0.364	0.548	18.76
ANN 15-6-1	0.9705	0.9126	1.141	1.346	14.79
ANN 15-7-1	0.8164	0.7565	0.394	0.484	28.43
ANN 15-9-1	0.9922	0.8246	0.292	0.479	21.35
ANN 15-10-1	0.977	0.8476	0.268	0.446	18.67
ANN 15-11-1	0.9191	0.8217	0.347	0.454	22.82
ANN 15-13-1	**0.9993**	**0.8432**	**0.186**	**0.443**	**10.98**
ANN 15-15-1	0.983	0.8784	0.254	0.393	17.12

Samaneh et.al [14] proposed hybrid ANN-COA model which offers a high degree of accuracy for predicting and controlling WWTP. The model was designed through Artificial Neural Network (ANN) and Cuckoo Optimization Algorithm (COA).

Selected input variables were T, pH, DO, BOD, COD, TSS, TDS, NO_3 and PO_4. Neural network with three layers is input layer, one hidden layer and an output layer with two functions such as hyperbolic tangent sigmoid and linear transfer function. An Aggregated measure (AM) has three parameters are model bias (MB), NS and r_{mod}. Here ANN and ANN-COA models are evaluated through statistical methods were Regression coefficient (R), mean square error (MSE), root mean square error (RMSE) and aggregated measure (AM) results shown in table 4. Finally, the result shows that ANN-COA model provides better performance than ANN.

Table 4: Performances of ANN and COA-ANN models [14]

	Models	R	MSE	RMSE	MB	NS	r_{mod}	AM
ANN	ANNBOD	0.72	36.41	6.03	0.07	0.7	0.62	0.75
	ANNCOD	0.8	29.7	5.44	0.06	0.82	0.79	0.8
	ANNTSS	0.7	48.29	6.94	0.08	0.6	0.54	0.68
COA-ANN	**COA-ANNBOD**	0.58	82.1	9.06	0.05	0.85	0.8	0.86
	COA-ANNCOD	0.68	66.75	8.17	0.032	0.9	0.85	0.9
	COA-ANNTSS	0.36	73.11	8.55	0.06	0.85	0.74	0.84

Hossien et.al [15] proposed Neural Network model used for prediction of Perchlorate wastewater treatment. Perchlorate (C104) is a strong electrochemical oxidant. The variables were speed of particle, pH particle, temperature of particle, density, wastewater, sewage of TSS, flue and waste rate of the tank. These variables used to predict Perchlorate density of wastewater. The entire dataset was normalized with two functions such as logsin and tansig. Different trail network models are evaluated through statistical methods were correlation coefficient (R), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).the statistical measurement is shown in table 5. Finally, the result shows that Trail 2 model provides better performance.

Table 5: Statistical metrics of different models [15]

Frequency Network	**No of neurons**	**MAE**	**R**	**MAPE (%)**	**RMSE**
Trail 1	16	7.453	0.8507	4.595	11.964
Trail 2	**17**	**7.373**	**0.8846**	**4.831**	**10.827**
Trail 3	16	7.457	0.8642	4.745	11.55
Trail 4	15	7.517	0.8723	4.625	11.15
Trail 5	16	7.561	0.863	4.515	11.709

Abdullah et.al [16] proposed Artificial Neural Network model to predict the performance of WWTP. Data was taken from Konya WWTP. Neural network with three layers is input layer, one hidden layer, an output layer and many neurons in each layer. The input values are pH, temperature, BOD, COD, TSS, and TDS. The entire dataset was normalized with three functions such as logsin, Purelin, and tansig. ANN model was evaluated through statistical methods are correlation coefficient (R) and mean square error (MSE).ANN 3-3-1 model shows maximum correlation coefficient and minimum mean square error.ANN 3-3-1 model is used as a valuable performance assessment tool for Wastewater plant operations. Summary of ANN 3-3-1 model is shown in table 6.

Table 6: Summary of ANN 3-3-1 model [16]

No of neurons	Transfer functions	Training R	Testing R	MSE
7	Logsig	0.98	0.93	0.0078904
	Purelin	0.89	0.58	0.0052331
	Tansig	0.93	0.86	0.0044014
8	Logsig	0.99	0.87	0.0075730
	Purelin	0.88	0.58	0.0058282
	Tansig	0.96	0.57	0.0044315
9	**Logsig**	**0.99**	**0.95**	**0.0023310**
	Purelin	0.61	0.38	0.0084474
	Tansig	0.99	0.79	0.0058679
10	Logsig	0.99	0.86	0.0071530
	Purelin	0.79	0.34	0.0049574
	Tansig	0.99	0.83	0.0067905
11	Logsig	0.99	0.79	0.0067391
	Purelin	0.61	0.50	0.0043681
	Tansig	0.81	0.58	0.0443300

Xiupeng et.al [17] proposed Neural Network model for short-term prediction of influent flow rate in WWTP. Multilayer Perceptron Neural Network Algorithm (MLP) used to build the prediction model by different time horizons with various data such as influent flow rate, rainfall rate, and radar reflectivity. Data was taken from WRF Iowa. Different time horizons were current time (t), t+15, t+30, t+60, t+90, t+120, t+150, t+180. Multilayer Perceptron with three layers is an input layer with 34 inputs, an output layer, and one or more hidden layers. NN model is built by various data mining algorithms such as MLP, random forest, boosted tree and SVM .all data mining algorithm is evaluated through statistical methods are correlation coefficient (R), mean absolute error (MAE) and mean square error (MSE).The entire dataset was normalized with five functions such as logistic, identity, hyperbolic, exponential and sine function. Comparison of all algorithm with statistical metrics proves that MLP model has MSE and MAE values are smallest one and highest correlation coefficient (R=0.988). Comparison of all algorithms with statistical metrics is shown in table 7. Based on comparison MLP model was constructed with statistical metrics and different time horizons as t to t+180.Prediction of influent flow rate with two models were graphically designed and shown in figures (7, 8). Finally, the result shows that prediction model predicted influent flow rate well until t+150.

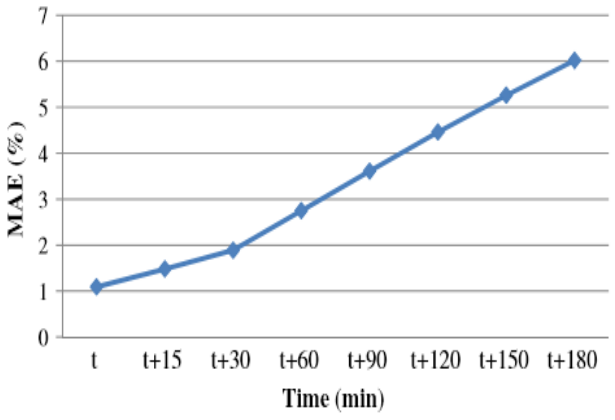


Figure 5: MAE of model for influent rate [17]

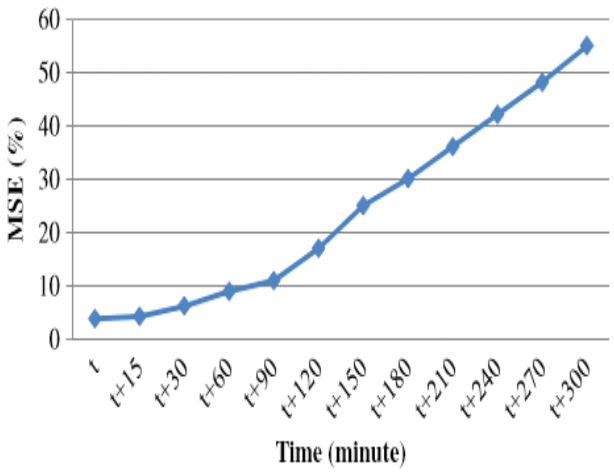


Figure 6: MSE of model for influent rate [17]

Table 7: Prediction Accuracy of various algorithms [17]

Algorithm	MAE	MSE	R
MLP	**1.09**	**4.21**	**0.988**
Random forest	3.04	20.69	0.945
Boosted tree	1.77	11.16	0.97
SVM	1.47	5.46	0.985

CONCLUSION AND FUTURE WORK:

Water plays big role in world economy. Water management and Wastewater treatment practice is most important for future generation to avoid environmental pollution and health problems. In this paper, an analysis is presented for Wastewater Treatment Plant (WWTP) using various techniques at different locations. Based on this survey several techniques are analyzed for the implementation of the waste water treatment. this study shows when the waste water is recycle the quantity of the waste water is reduced in urban area. The aim of the future plan is to analyses the data set from UCI Machine Learning Repository and implements the water quality indicator to evaluate various data mining algorithms which examine the performance of WWTP.

REFERENCES:

1. http://archive.epi.yale.edu/case-study/primary-vs-secondary -types-wastewater-treatment
2. https://en.m.wikipedia.org/wiki/wastewater
3. https://www.sswm.info/category/implementation-tools/implementation-tools-introduction
4. http://www.waterprofessionals.com/learning-center/articles/physical-water-treatment/
5. https://www.thomasnet.com/articles/chemicals/wastewater-chemical-treatment/
6. http://www.neoakruthi.com/blog/biological- treatment-of- wastewater.html
7. http://www.yourarticlelibrary.com/water/types-of-wastewater-treatment-process-etp-stp-and-cetp/27418
8. Rakesh Singh Aiswal, Dr.Santosh Kumar ,Shweta Singh and Megha Sahu, “Wastewater Treatment by Effluent Treatment Plants”, SSRG International Journal of Civil Engineering,Vol-3, ISSN 2348-8352,32-38,2016
9. Francesco Granata Stefano Papirio, Giovanni Esposito, Rudy Gargano and Giovanni de Marinis, “Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators”,www.mdpi.com/journal/water, Vol-9, DOI 10.3390/w9020105, 1-12, 2017.
10. Andrew Kusiak, Xiupeng Wei, “Prediction of methane production in wastewater treatment facility: a data-mining approach”, Annals of Operations Research, Vol-216, ISSN 0254-5330, 71-81, 2014.
11. Hamed Hasanlou, Naser Mehrdadi, Mohammad Taghi Jafarzadeh, Hamidreza Hasanlou, “Performance Simulation of H-TDS Unit of Fajr Industrial Wastewater Treatment Plant Using a Combination of NN and PCA ”, Journal of Water Resource and Protection,Vol-4,DOI 10.4236/jwarp.2012.45034 ,311-317 ,2012
12. Davut Hanbay a, Ibrahim Turkoglu a, Yakup Demir b, “Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks”, Expert Systems with Applications, DOI 10.1016/j.eswa.2006.10.030,1038–1043,2008.
13. Djeddou M, Achour B, “The Use of A Neural Network Technique for the Prediction of Sludge Volume Index in Municipal Wastewater Treatment Plant”, Larhyss Journal, ISSN 1112-3680, 351-370, 2015.
14. Samaneh Khademikia a, Ali Haghizadeh c, Hatam Godini d, Ghodratollah Shams Khorramabadi b, “Artificial Neural Network-Cuckoo Optimization Algorithm for Optimal Control of Khorramabad Wastewater Treatment Plant, Iran”, Civil Engineering -- Journal, Vol-2, 2016.
15. Hossien Fakhraee, Seyed Shayan Akbarpour, “Application of Artificial Neural Network for Prediction of Perchlorate Wastewater Treatment”, Computational Research Process in Applied Science and Engineering, Vol 01(03), ISSN 2423-4591, 103-111, 2015.
16. Abdullah Erdal TUMER, Serpil EDEBALI, “An Artificial Neural Network Model for Wastewater Treatment Plant of Konya”, IJISAE, ISSN: 2147-6799, 2015.
17. Xiupeng Wei, Andrew Kusiak and Hosseini Rahil Sadat, “Prediction of Influent Flow Rate: Data Mining Approach”, Journal of Energy Engineering, Vol-139, ISSN 0733-9402, 118-123, 2013

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

A SURVEY ON IMPACT OF SOCIAL MEDIA ON DIFFERENT DIMENSIONS

Dr. M. Mahadevi
Assistant Professor, P.G. Department of Computer Science,
S.D.N.B Vaishnav College for women,
Chennai, India
email:maxsaran@gmail.com

R. Pavithra
Student, P.G. Department of Computer Science,
S.D.N.B Vaishnav College for women,
Chennai, India
email:letshopepavi@gmail.com

Abstract: Social Media had acquired a large popularity and user-ship. The social media had impacted users directly or indirectly in many ways. It has both positive and negative effects on the users. The users are of varying age groups and from different social and cultural backgrounds. Due to its vast usage and popularity among the people, the scholars have a keen interest on analyzing the impacts caused by social media. This review tries to focus on different existing articles and works which employed various data **mining and statistical techniques to** find the impact caused by social media. It mainly focuses on the impact of social media on variables such as society, education, student's community, cybercrime, and security.

Keywords: Social Media, data mining, statistical techniques, academic performance, privacy and security threats.

INTRODUCTION:

Social Media is Internet-based software and interfaces which are used by people to establish communication between one another, sharing information about their lives such as their biographical details, their interests, hobbies, picture and multimedia and also their thoughts and views. Social media was initially viewed as personal tools used by individuals to communicate with friends and family but was later used by businesses for economic purposes, advertisements. It gained popularity among youngsters.

Fig 1: Usage of social media [1]

Social media had become a part and parcel of people's life. It had not only connected millions of people online but also paved a way to new problems. The social media had a major growth spurt of users on the advent of smartphones. It is easier to access these sites and also launched as attractive apps which are easy to download and available free of cost. The user count on social media platforms is increasing constantly. There are 2. 7 billion users in social media as of 2017, with Facebook having the highest user count of 2061 million as of 2017.

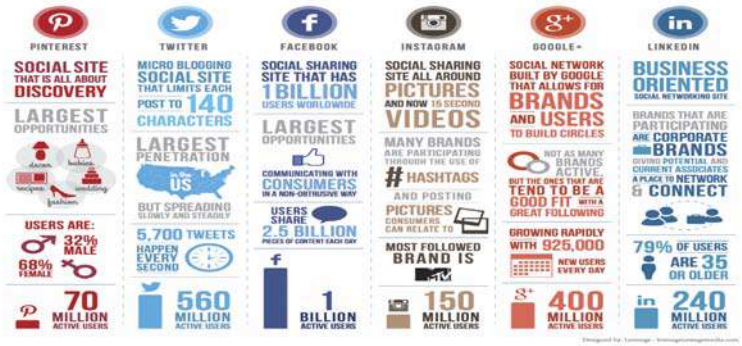


Fig-2: Popular social networking sites [2]

I. IMPACT OF SOCIAL MEDIA:

The social media's popularity and widespread usage have kindled many scholars to conduct studies and write articles based upon the impact of social media on various aspects. This literature review reveals the major purpose of these studies and its findings in terms of impact or effect of social media and also enables us to understand the different dimensions of the way social media have affected people and society. The impacts can be sub-categorised as follows.

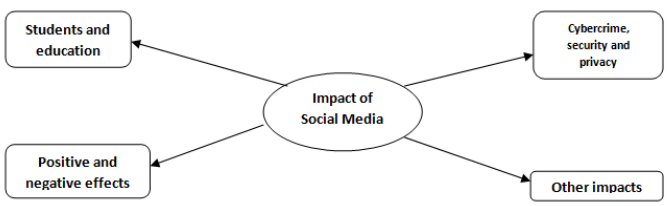


Fig-3: Impact of social media

A. Impact on Students and Education:

The student's community constitutes the majority of the users in social media and hence they are more impacted by **Social media** in comparison with the other users **.**Fig-4 shows the college students' usage of internet

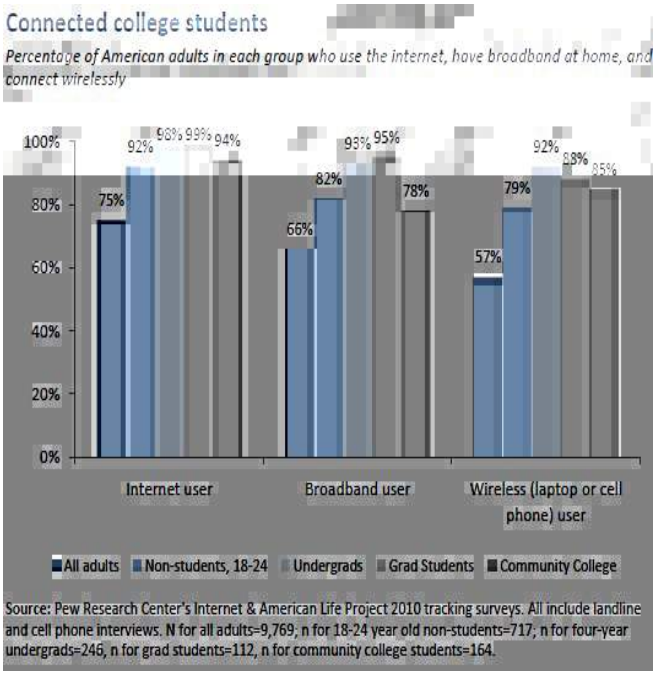


Fig-4: College student's internet usage [3]

Jayshree Jha, et al [4] had conducted a study to find the effect caused by social media in academics among students. They had adopted the questionnaire survey method .The samples were drawn at random and analyzed using statistical **tools** such as pie chart and graphs. Based on the findings it is evident that it affects the grades of the students. It also proves that the social media acts as a platform to release their pressure. This study also suggests us to find an approach on how to balance the relationship between the usage of social media and academics.

The study conducted by **Waqas Tariq, et al** [5] dealt with the impact caused by social media on students and education, as well as the life of the teenagers. They also tried to describe the danger of social media usage among youngsters and teenagers. Based upon various surveys conducted between 1991-2011 and **statistical analysis** on graphical and other visual representations of data it is evident that there is no third party or community to keep track or check the user related activities in social media. Hence social media usage among teenagers without any guidance could be dangerous.

The study conducted by **Aida Abdulahi, et al [6]** dealt with the negative effects of social networking sites such as Facebook and its impact upon Asia Pacific University Scholars in Malaysia. They examined the link between usage of Social Media and health threats .They also comprehensively analyzed the law and privacy of social media and risks involved by them**. Survey methodology** is adopted and they had employed **Pearson moment correlation and regression** to examine the relationships among the variables. According to the results, it is evident that as time spent on **Social Media** increases, the academic performance of the students declined. Secondly, people also unaware of information sharing policies in social media and doesn't know that their personal data can be shared among other parts of social media. As a result, they end up sharing their private information with unauthorised people. The author suggested users to have the awareness of these security policies while using **Social media**.

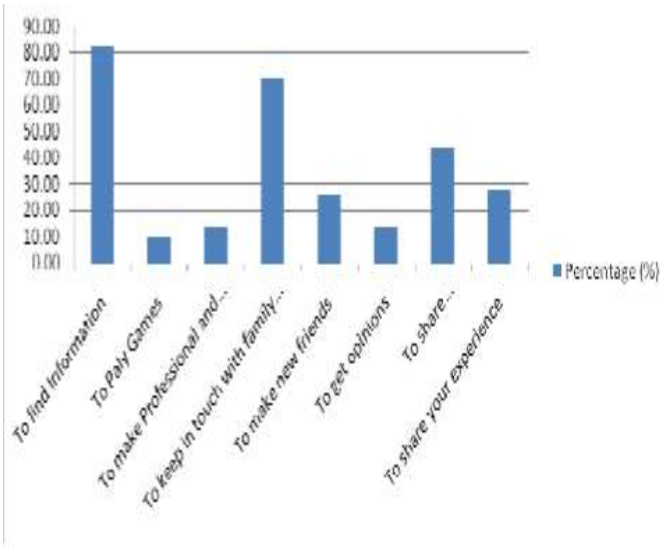


Fig-5: Purpose of using Social Media [7]

According to Tejas**. B shah,** [7] based on their study; there exists positive as well as negative effects caused by social media among students. The study employed the method of **questionnaire survey** which was used for data collection .The analysis on the collected data showed that facebook is the most used Social Media among the sample and about **38%** of the samples had used it for more than a year. About 74% of the students believe Social Media have a positive effect on academics. And 76% thinks that Social Media is helpful for their assignments related to their academics .The fig-5 show the purpose of Social Media usage among students. The research concludes that students should have the balance between usage of social media and their education.

B. Impact on social cybercrime, security and privacy:

Cybercrime is another issue prevails among Social Media and acts as a threat to the security of social media users .The privacy of the users is in danger if Social Media is used without caution. Fig-6 shows the cyber bullying in social media.

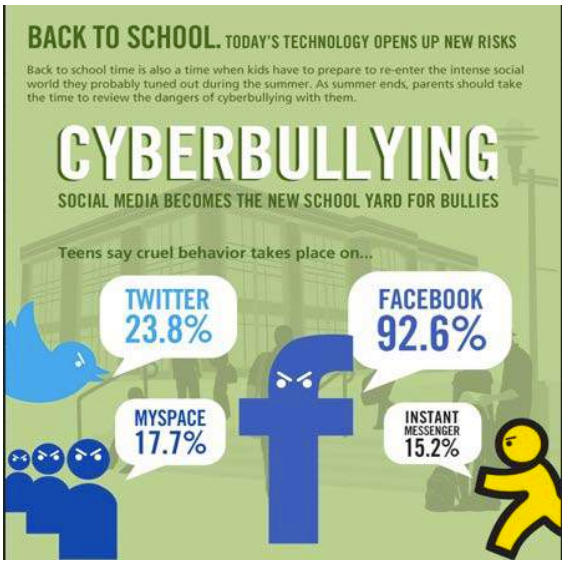


Fig-6: Cyberbullying [8]

Ganesan and P.Mayilvahanan [9] have conducted a study on cybercrime using a data mining technique .the cybercrime analysis was performed on social media data**. Clustering and k-means methods** are used for exploring the data set. The **k-means clustering algorithm** is used to obtain unstructured data. Since the dataset regarding **Social Media** is of unstructured format such as text, video, images etc. It proposes the clustering technique for analyzing the crime data from the database. And this technique is faster to obtain data from the vast social networking data available and types of crime can also be classified and using k-means to retrieve data.

Roshan jabee et al [10] had studied the view of Social Media users regarding security and privacy settings improvements, particularly on facebook. He adopted a **survey method based on questionnaire** and analyzed the data using **statistica**l and visual representation of data such as graphs and diagrams .The results shows that 92% of users are students belong to age group (15-25) years and 24% of users have experienced a security breach .Table-1 showed the responses given by the users about the need for improvement in default privacy settings in facebook. He strongly recommends that the user should be more concerned in security settings in facebook.

Table-1: Default privacy setting of Facebook need improvement.[10]

Response	No. of users	% of users
Yes	145	85.29
No	17	10
Depends	8	4.70
Total	170	100

C. Positive and negative impact:

Usage of social media had its own impact on the users. It can be revealed to have both positive as well as negative effects based on how they used social media and in which way it had impacted the users

Dr. Indrajit Roy chowdhury, [11] had focused on negative and positive impacts caused by social media among youths. A **survey method** is adopted and **statistical** and visual representations of these data are used for analysis. The results showed that 80% of female Facebook users regularly access their accounts .Over 10% of female FB users in Kolkata had experienced harassment by someone in social media .About 60% of the users have used facebook more than 4 years. And 85% are already aware of the negative impact caused by Social Media.

Anurag Sarkar et al [12] explored the merits and demerits of social media. They adopted a **survey and statistical** methods to visually represents data in the form of graphs .The result concludes that social media will become a significant part of human life and it can be brought under control with proper planning and management. It also suggests that more detailed study should be required in focus on health issues. Their study states that cyberbullying, identity theft, negative impact on mental and physical health, poor language and grammar as the demerits. The positive impacts were distinguished as worldwide connectivity, a community of interest, real-time sharing of information, free advertising and business advantages.

The study conducted by **A.T.M Shahjahan, et al [13]** depicts the effects of social media on the people. The **questionnaire survey method** is adopted and the analysis indicated that it has both positive and negative effects such as meeting new people,sharing idea beyond landscapes to reach more people for business ,education purposes etc. The negative impact is spending more time in Social Media than necessary

D. Other impacts caused by social media:

The study conducted by **Nicole B. Ellison, et al [14]** adopted the **empirical method of survey** and found that about 94% of the sample uses facebook . Usage of facebook helped them to overcome barriers faced by students who have low satisfaction and low self-esteem. But overall **regression model** indicates facebook being less useful for maintaining or creating bonding social capital. Strong linkage between facebook usage and highschool connection suggests Social Media helped them to maintain their relationship as people move from offline community to another.

Annapoorna Shetty, et al [15] in their study adopted a questionnaire method to find the effects caused by social media. The survey was being conducted on a sample of 100 youngsters. The average age groups were between 18-30 years. The data was analyzed and it shows that the Positive use of social media was developing the academic career, better lifestyle, to adopt new trends, fashion, and anthropology .Table-2 shows the positive impact of social media among youngsters based upon 100 responses in which 66 people told yes and 34 people told no.

Table-2: Social media have positive impact on youth[15]

Response	F%	Valid%	Cumulative%
Yes	66	66	66
No	34	34	34
total	100	100	100

The study conducted by Samaneh **Beheshti-Kashi and Baharak Makki** presents **[16]** the use of social media in the context of news. The **online survey** was conducted among both users and non-users of facebook and the statistical analysis indicated that social media was only an additional path to traditional media than being its alternative .it also raised the question regarding the credibility of news published by a friend on social media. The non-Social Media users intentionally avoids the content generated by Social Media users and being sceptical of truthfulness of news from social media

II. CONCLUSION:

The above literature review shows the impact caused by social media on various aspects. Fig-7 depicts the statistics among the usage of social media by both men and women. It is ascertained that the recent women use social media such as facebook more than men. Thus social media has more impact on women than men. Many researchers have tried the impact of this study based on statistical measures. So the aim of the proposed study is mainly focused on data mining techniques for the analysis of the dataset. The proposed work gives emphasis upon finding the impact of social media among women in terms of physical, mental and emotional aspects using data mining techniques. The dataset can be obtained by employing a questionnaire survey method and can be analyzed using suitable data mining model. The results may show how much women are affected by using social media in a positive and negative way in terms of physical, mental and emotional aspects.

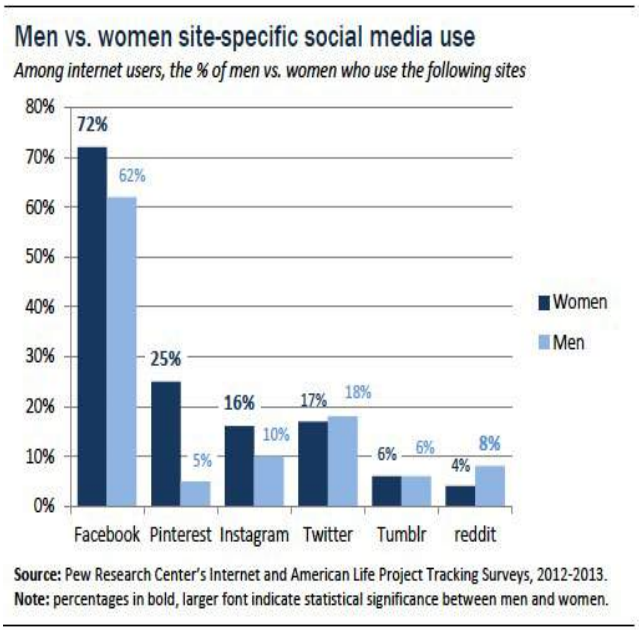


Fig-7: Usage of social media [17]

IV. REFERENCES

[1]https://www.Smartinsights.Com/social-media-marketing/social-media-strategy/new-global-social-media-research/.

[2] http://www.digitalvidya.com/blog/impact-of-social-media-marketing-on-your-business-dmblog-0702/.

[3] http://www.pewinternet.org/2011/07/19/college-students-and-technology/.

[4] Jayshree Jha, Neelam Jaipuria ,shivesh jha, Priya Sinha, The Effects of Social Media on Students- International Journal of Computer Applications (0975 – 8887), International Conference on Advances in Information Technology and Management ICAIM – 2016.

[5]Waqas Tariq1, Madiha Mehboob2, M. Asfandyar Khan1 and faseeullah3, The Impact of Social Media and Social Networks on Education and Students of Pakistan ,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, 2012.

[6] Aida Abdulahi, Behrang Samadi ,Behrooz Gharleghi ,A Study on the Negative Effects of Social Networking Sites Such as Facebook among Asia Pacific University Scholars in Malaysia, International Journal of Business and Social Science,Vol. 5, No. 10, 2014.

[7] Tejas. B shah, Marten. H. Patel, The Effects of Social Media on College Students, RESEARCH HUB – International Multidisciplinary Research Journal, Volume-2, Issue-4, 2015.

[8]https://rampages.us/peasedn200/wp-content/uploads/sites/13742/2015/12/cyberbullying-social-media-becomes-the-new-school-yard-for-bullies-1.jpg.

[9] M.Ganesan , P.Mayilvahanan ,Cybercrime Analysis in Social Media Using Data Mining technique s,International Journal of Pure and Applied Mathematics Volume 116 No. 22 2017, 413-424 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version).

[10] Roshan Jabee Department of Computer Science and Engineering, Jamia Hamdrad New Delhi, India M. Afshar alam, Issues and Challenges of Cyber Security for Social Networking Sites (Facebook), International Journal of Computer Applications (0975 – 8887) Volume 144 – No. 3, 2016.

[11] Dr. Indrajit Roy chowdhury, Mr. Biswajeet saha, Impact of Facebook as a Social Networking Site (SOCIAL MEDIA) On Youth Generations: A Case Study of Kolkata City, International Journal of Humanities and Social Science invention volume 4 Issue 6, 2015.

[12] Anurag Sarkar1, Prof. Shalabh Agarwal2 Abir Ghosh3 Dr. Asoke Nath4, Impacts of Social Networks: A Comprehensive Study on Positive and Negative Effects on Different Age Groups in a Society International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 5, 2015.

[13] A.T.M Shahjahan, Kutub Uddin Chisty, Social Media Research and Its Effect on Our Society, International Journal of Information and Communication Engineering Vol: 8, No:6, 2014

[14] Nicole B. Ellison Charles Steinfield Cliff Lampe, The Benefits of Facebook ‘‘Friends:’’ Social Capital and College Department of Telecommunication, Journal of Computer-Mediated Communication 12 (2007) 1143–1168 ª 2007.

[15] Annapoorna Shetty1, Reshma Rosario2, Sawad Hyder2 International Journal of Innovative Research in Computer and Communication Engineering.(An ISO 3297: 2007 Certified Organization) Vol. 3, Special Issue 7, 2015.

[16] Samaneh Beheshti-Kashi and Baharak Makki,Department of Engineering, Faculty of Technology and Science, University of Agder, Grimstad, Norway social media news: motivation, purpose and, usage International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 2, 2013.

[17] http://www.pewresearch.org/fact-tank/2013/09/12/its-a-womans-social-media-world/

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A Study Paper on Wireless Sensor Secure Routing

T. Yegammai,
Assistant Professor, Department of Computer Science,
yegammai@shasuncollege.edu.in

S.G Packiavathy
Head, Department of Computer Applications,
packiavathypaul@hotmail.com

ABSTRACT

An important purpose is that the wireless sensor routing security networks have many sensor routing protocols and nodes but have no security. Security goals for routing in sensor networks show us how crippling attacks have been made and attacks have been made and attacks against ad-hoc and peer-to peer networks. Two undocumented attacks such as sinkhole and hello floods which have been described and analyze the security of all secure routing in wireless sensor networks and protocols used for disseminating controls and information's network called sinks.

INTRODUCTION:

Routing security in wireless networks is an important purpose in the routing network which has limited nodes and application networks but have no security. Although there is no security available, we make the

security properties. In insecure wireless communication, limited nodes[1] and insider threats, where when designing the network secure routing adversary people has laptops with energy and long range communication, where the routing becomes non-trivial. Crippling attack is provided for all the major routing protocols because they have no security on the routing of sensor network and is insecure.

BACKGROUND:

Sensor network refers to the sensors and general computing elements. A sensor network consists of hundreds and thousands of low power costs and nodes but only at fixed locations which affects the environment[2]. The sensor networks which consist of one or more point of control is the base stations. The sensor node is the access point for the securing routing which is used to disseminate the control information on networks. Sensor network routing might have laptop, memory- storage, and high-bandwidth links for communication among the sensor network. Sensor network use low power and bandwidth that would communicate to the nearest base station for sensor network. The aggregation networks where the total numbers of messages sent, the energy is saved in the network where from the aggregation point they collect the readings from surrounding nodes. Sensor networks differ from other systems where it has a great challenge [2]. The value of sensor networks comes from many nodes where they will have to develop cheaper sensor nodes. Security is critical when networks are at risk where we have sensor nodes such as[3]:- High bandwidth Sensor node Base station Low latency and Only laptop and base stations use low latency and high bandwidth.

RELATED WORK:

Security issues are similar to sensor networks and are developed for ad-hoc networks. The secure routing protocols [4] for ad-hoc networks and sensor network reasons are that they sense security in ad-hoc network for authentication and secure routing protocols. The routing protocols are based on public key cryptography[5] which is used for sensor nodes.

SECURITY GOALS:

Secure routing protocols which have integrity, power and availability of messages in presence of arbitrary power. Security is not relevant to the application data and not responsible for routing protocol.

ATTACKS ON SENSOR ROUTING:

There is attack against the ad-hoc sensor networks but quite simple for the following categories:

SINK HOLE:

The goal is to take away all the traffic (nodes) from the particular area through nodes creating a sinkhole with adversary at the centre [5]. Sinkhole attacks which enables many other attacks where only one single node provides single high quality information which influences large number of nodes. The laptop class with transmitter which provides a high quality route for transmitting with power.

HELLO FLOOD:

The hello packets which are available to the bound channel are available to the attackers. An advisory situated close to the base station may be completely disrupted[8]. Protocols which depend on the local information exchange between neighboring nodes for maintenance nodes.

Attacks on specific sensor network protocols:

The main attacks of the sinkhole and hello flood is that the tiny OS protocols which can increase latency or disable thread [7]. The attacks on the link and sensor networks are against the network routing by line layers encryption. Sinkhole attacks on the network. Protocols which defend against protocols and the provided information, such as

HELLO FLOODS ATTACKS:

The hello floods attacks which verify the links of the nodes based on messages over the link. The hello floods attacks verify the link between two nodes, even if the advisory has high sensitive networks which will verify neighbors for each node to prevent hello flood attack.

CONCLUSION:

Securing routing which becomes vital to acceptance and the use of sensor networks against various protocols and attacks against the ad-hoc and peer to peer network where the attacks and routing protocols which defeat the security goals against the adversary. But we have

demonstrated the currently routing protocols against the sinkhole and hello flood attacks. A design of the sensor network routing protocols which satisfies the security goals and also where the authentication and sensor routing protocols which can be used as security nodes and where the key cryptography which cannot depend the laptop class adversary and the protocol can be designed well for sensor network to be secured.

REFERENCES

[1] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Wormhole detection in wireless ad hoc networks,"Department of Computer Science, Rice University, Tech. Rep. TR01-384, June 2002.

[2] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *IEEE INFOCOM '97*, 1997,pp. 1405–1413.

[3] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.

[4] F. Stajano and R. J. Anderson, "The resurrecting duckling: Security issues for ad-hoc wireless networks," in *Seventh International Security Protocols Workshop*, 1999, pp. 172–194.

[5] J. Hubaux, L. Buttyan, and S. Capkun, "The quest for security in mobile ad hoc networks," in *Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC 2001)*, 2001.

[6] L. Zhou and Z. Haas, "Securing ad hoc networks," *IEEE Network Magazine*, vol. 13, no. 6, November/December 1999.

[7] J. Kong, P. Zerfos, H. Luo, S. Lu, and L. Zhang, "Providing robust and ubiquitous security support for mobile ad-hoc networks," in *ICNP*, 2001,pp. 251– 260.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

ANALYSIS AND PREDICTIONS ON BLENDED LEARNING READINESS AMONG INDIAN STUDENTS AT UNIVERSITIE SUSING DECISION TREE CLASSIFIER IN SCIKIT-LEARN ENVIRONMENT

Rajalakshmi R
MSc Information Technology
M.O.P Vaishnav college for women
Chennai- 600034, India
e-mail: mail2raji.cs@gmail.com

ABSTRACT

Decision trees classifiers are modest and hasty data classifiers, usually used in data mining to study the data and generate the tree and its rules that will be used to formulate predictions. One of the key challenges for knowledge discovery and data mining systems stands in developing their data analysis capability to discover out of the conventional models in data. Since the Union budget 2017, the debate around the quality of Higher Education in India has been acquiring momentum, which laid emphasis on skills development, employability and digitisation of the education process. The key lies in blended learning, a model that is fast gaining pace in the Indian context,where online tools are combined with classroom and instruction to provide an overall improvement in educational outcomes activities.The explanatory variables are students' attitude towards learning flexibility, online learning, classroom learning, degree and stream. This paper represents an implementation of the decision tree classifier algorithm using scikit-learn library for python on data collected from the survey with the purpose of predicting whether a particular student is ready for blended learning throughdecision trees.

Keywords: Data mining; Classification; Decision trees; Blended Learning; scikit-learn

I. INTRODUCTION

Data mining is a process used to extract usable data from a larger set of any raw data.Two main objectives can be distinguished in the data mining process integrated in the management system: a description objective consisting in establishing the eloquent variables and its influences; and a prediction objective.

Online educational tools are used in a wide range of context for many different goals and motives, but there is anincreasing focus on blended learning where online tools are combined with classroom and instruction to provide an overall improvement in educational outcomes activities.Students in the developing world are frequently quoted as being among the most important heirs of online education initiatives such as MOOCs.Blended learning is a hybrid form of teaching and learning which involves both classroom and online learning. Blended learning incorporates information via online courses, developed by experts from different fields, and helping students access globally developed and industry relevant course material. The approach mixes concept building and enquiry-based learning which retains human interaction in education and allows students to combine traditional classroom methods with online digital medium.

The main objective of this paper is an attempt to use data mining methodologies to study students' readiness towards blended learning. Data mining provides many tasks that could be used to study this analysis. In this research, the classification task is used to evaluate student's readiness and as there are many approaches that are used for data classification, the decision tree method is used here. An online questionnaire is used to collect data for this research and the respondents were students from different colleges in India belonging to different streams of study. The decision trees are generated using Decision tree classifier algorithm built using scikit-learn library for python in Anaconda Navigator (Spyder) Environment.

II. REVIEW OF LITERATURE

The paper [1] has presented findings of a study of mobile learning readiness among Malaysian students from two public universities. This describes about the basic readiness, skill readiness, psychological readiness, budget readiness of students from different universities. The research concludes that it is essential for educators to integrating mobile learning into academic programs and that the respondents welcomed this idea.[2] Reports on blended learning environment approach to enhance the performance of the students learning science. Students' learning experience outside the school and learning outcomes can be improved by adequately preparing for both pre- and post-visit of blended learning courses. The concept about the particular subject is strongly covered by multi-faced roles played by students and teachers. [3] Provides an insight into how the students really work and learn using technologies. Qualitative and quantitative techniques were used to gain an appreciation of the students' experience with ICT as a supporting mechanism and the blended delivery medium for the module.E-learning is becoming popular in educational institutions because ICT (information and communication technologies) [4] .Linear Regression analysis is used as a means to find that the relationship between E-learning and student is high. To find out the student's preferences, a commonly used approach is to implement a decision model that matches some relevant characteristics of the learning resources with the student's learning style. Adaptive machine learning algorithms are used as a tool to learn about the student's preferences towards E learning over time. At first, all the information available about a particular student is used to build an initial decision model based on learning styles [5].

III. SYSTEM MODEL

To build a reliable classification model, the following methodology is adopted. The methodology consists mainly of five steps: Collecting the relevant features of the problem under study, preparing the data, building the classification model, evaluating the model using one of the evaluation methods, and finally using the model for future prediction of the student performance. These steps are presented in the next subsections.

A. Collecting the relevant features

In this step, the relevant features are collected through a survey via Google Forms and the respondents were 100 students from different colleges in India. Initially 15 attributes have been collected of which only 7 conditional attributes and one class attribute have been considered. The attributes along with their descriptions and possible values are presented in Table I. The class attribute is the student's readiness for blended learning and named BLREADY.

B. Feature selection and construction

New features were constructed from the existing features. They were Attitude towards Learning Flexibility, Attitude towards Online Learning, and Attitude towards Classroom Learning. The features were also discretized. All the respondents agreed that web is a useful platform for learning and that technology wasn't a hindrance to anyone in any way.

C. Building the classification model

The Classification model is built using the decision tree method. The decision tree is a very good and practical model since it is comparatively fast, and can be easily converted to simple classification rules. The decision tree method relies mainly on using the information gain metric which determines the feature that is most useful. The information gain depends on the entropy measure. The categorical values of the conditional variables are encoded using one-hot encoder. The data in csv format is split into training data (70%) and test data (30%). The tree module is used to build a Decision Tree Classifier. Accuracy metrics from the predicted class variable is computed using Accuracy_score module.

Table I. Symbolic Attribute Description

Sr.No	ATTRIBUTE	DESCRIPTION	POSSIBLE VALUES	ONE-HOT ENCODED VALUES
1	Degree	Degree pursued by the student	Undergraduate, Postgraduate, Others	2,1,0
2	Stream	Stream of Study	Science, Arts, Engineering, Commerce, Others	5,0,2,1,4
3	Year	Year of Study	I, II, III, IV, V, Graduated	1,2,3,4,5,0
4	OL	Have you done any online course?	Yes, No	1,0
5	ATLF	Attitude towards Learning Flexibility	Poor, Fair, Excellent	0,1,2
6	ATOL	Attitude towards Online Learning	Poor, Fair, Excellent	0,1,2
7	ATCL	Attitude towards Classroom Learning	Poor, Fair, Excellent	0,1,2
8	BLREADY	Readiness for Blended Learning	Yes, No	1,0

1) Data Slicing

Data slicing is a step that is used to split data into train and test set. Training data set can be used specifically for our model building. Test dataset should not be mixed up while building model. We should not standardise our test set even during standardisation.

2) Decision Tree Training

DecisionTreeClassifier(): This is the classifier function for DecisionTree defined by scikit-learn. It is the main function for implementing the algorithms. Some important parameters are:

- criterion: It defines the function to measure the quality of a split. Information gain is used as the criterion.
- splitter: It defines the strategy to choose the split at each node. Best split has been chosen.
- max_depth: The max_depth parameter denotes maximum depth of the tree.
- min_samples_leaf: The minimum number of samples required to be at a leaf

The following rules are inferred from the decision tree in Figure 1.

- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is fair or excellent and the student has done online course then the student is ready for blended learning.
- If the student's attitude towards classroom learning is fair or excellent and the student belongs to commerce or Humanities then the student is not ready for blended learning.
- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is fair or excellent, the student has not done online course and the degree is postgraduate or undergraduate then the student is ready for blended learning.
- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is poor and the attitude towards online learning is fair or excellent then the student is ready for blended learning.
- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is poor, the attitude towards online learning is poor and the degree is postgraduate or undergraduate then the student is not ready for blended learning.

IV. PERFORMANCE ANALYSIS AND RESULTS

In order to measure the performance of a classification model on the test set, the classification accuracy or error rate are usually used for this purpose. Accuracy is the ratio of the correctly predicted data points to all the predicted data points. Accuracy as a metric helps to understand the effectiveness of our algorithm. The classification accuracy is computed from the test set where it can also be used to compare the relative performance of different classifiers on the same domain. However, in order to do so, the class labels of the test records must be known. Moreover an evaluation methodology is needed to evaluate the classification model and compute the classification accuracy. The decision tree classifier resulted in an accuracy of 78% and Naïve Bayes

algorithm resulted in an accuracy of 72%. The sample training data is shown in Table II. The sample test data is shown in Table III.

V. BIVARIATE ANALYSIS

Bivariate Analysis finds out the relationship between two features. The association between stream and readiness for blended learning is shown in Figure 2. Figure 3 shows the scatter plot between stream and attitude towards online learning. Figure 4 shows the scatter matrix of all the features

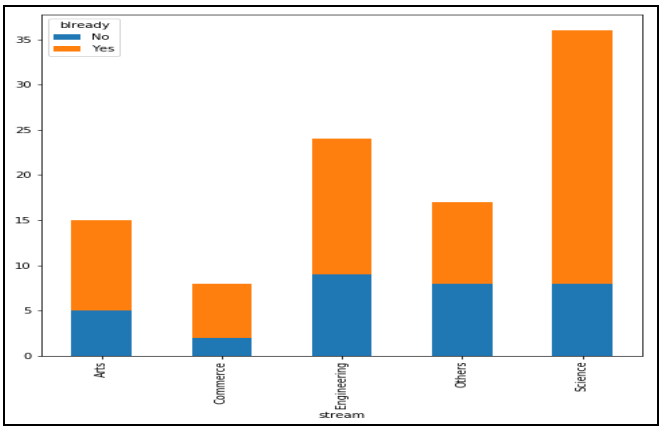


Figure 2. Association between stream and blended learning readiness

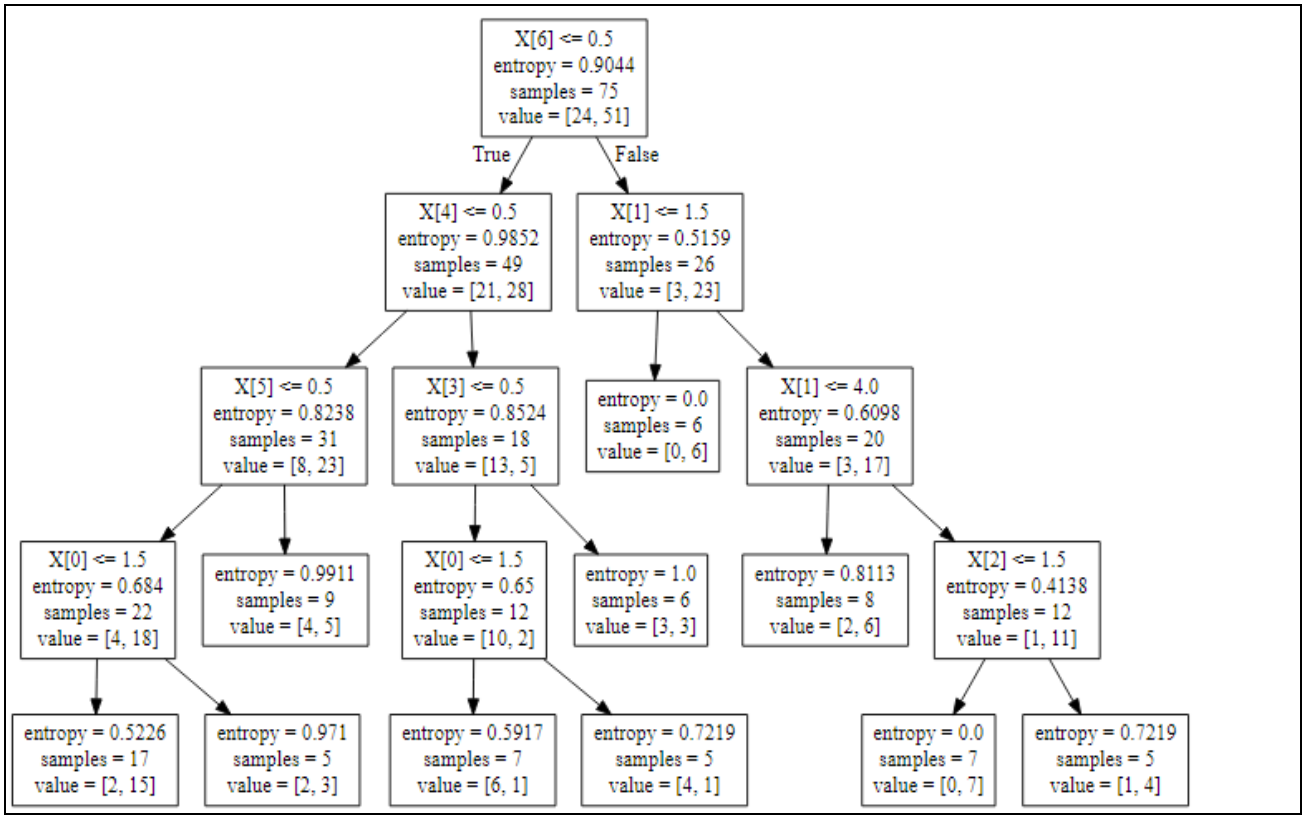


Figure 1. Decision tree

.Table II. Sample Training Data

Sr.No	DEGREE	STREAM	YEAR	OL	ATLF	ATOL	ATCL	BLR
1	Postgraduate	Science	I	No	Poor	Fair	Poor	Yes
2	Postgraduate	Science	I	No	Excellent	Excellent	Excellent	Yes
3	Undergraduate	Engineering	Graduated	No	Excellent	Poor	Excellent	No
4	Postgraduate	Science	I	Yes	Excellent	Excellent	Fair	Yes
5	Undergraduate	Engineering	IV	Yes	Fair	Fair	Poor	Yes
6	Postgraduate	Science	I	No	Excellent	Excellent	Excellent	No
7	Undergraduate	Others	III	No	Excellent	Fair	Excellent	No
8	Undergraduate	Engineering	IV	No	Fair	Excellent	Excellent	No
9	Postgraduate	Science	I	Yes	Poor	Poor	Poor	Yes
10	Undergraduate	Engineering	II	No	Excellent	Excellent	Excellent	No

Table III. Sample Test Data

Sr. No	DEGREE	STREAM	YEAR	OL	ATLF	ATOL	ATCL	BLR
1	Undergraduate	Engineering	IV	No	Excellent	Excellent	Fair	No
2	Postgraduate	Science	II	No	Excellent	Excellent	Excellent	Yes
3	Undergraduate	Engineering	IV	No	Fair	Excellent	Excellent	Yes
4	Undergraduate	Engineering	IV	No	Excellent	Excellent	Fair	Yes
5	Postgraduate	MBA	I	Yes	Fair	Excellent	Excellent	Yes
6	Postgraduate	Engineering	IV	Yes	Excellent	Excellent	Excellent	No
7	Undergraduate	Commerce	Graduated	No	Excellent	Excellent	Poor	Yes
8	Undergraduate	Commerce	Graduated	No	Excellent	Excellent	Excellent	No
9	Postgraduate	MBA	I	No	Excellent	Excellent	Excellent	Yes
10	Undergraduate	Engineering	IV	No	Excellent	Excellent	Excellent	No

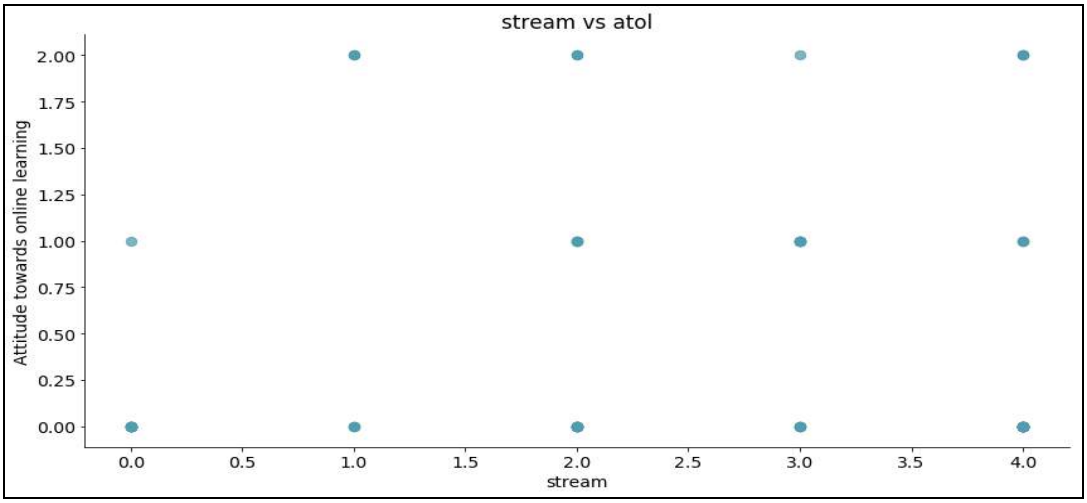


Figure 3. Scatter plot of stream Vs. attitude towards online learning

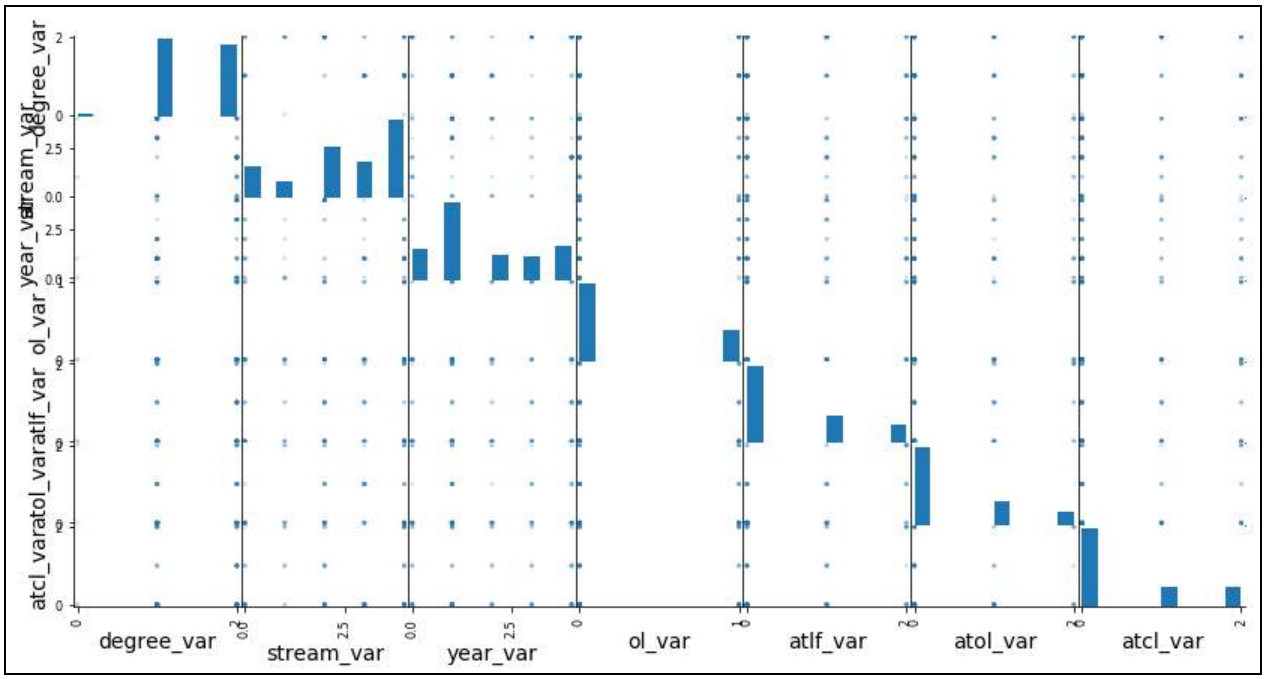


Figure 4. Scatter matrix

VI. CONCLUSION

This research is a preparatory attempt to use data mining functions to analyse and evaluate students' readiness towards blended learning. The higher managements can use such classification model to enhance the methodology of learning according to the extracted knowledge. Such knowledge can be used by the management system to improve their policies, enhance their strategies, and improve the quality of education system. One of the most striking future works is to collect a real and large data set from the students across India and apply the model using such data. Also, it could be narrowed down by taking a particular institution in hand and collecting responses from the students and thereby predicting the preference towards blended learning. Moreover, several other classification methods can also be applied to test the most suitable method that suit the structure of the data and give a better classification accuracy.

VII. REFERENCES

[1] Supyan Hussin1, Mohd Radzi Manap1, Zaini Amir1, Pramela Krish & Pramela Krish1. Mobile Learning Readiness among Malaysian Students at Higher Learning Institutes - Published by Canadian Center of Science and Education,2012

[2] Sandhya Devi Coll, Research Scholar, Curtin University, Perth, WA, Australia David Treagust**.** Blended Learning Environment: An Approach to Enhance Student's Learning Experiences Outside School (LEOS) –MIER Journal of Educational Studies, November 2017.

[3] Fiona Concannon, Antoinette Flynn and Mark Campbell, What campus-based students think about the quality and benefits of e-learning, British Educational Communications and Technology Agency, 2005

[4] The Impact Of E-Learning On Students Performance In Insitution , Oye, N. D.; 2A.Iahad, N., Madar, M. J. and Ab.Rahim, N., Department of Information System Universiti Teknologi Malaysia

[5] Cristina Carmona1, Gladys Castillo2, Eva Millán1, Discovering Student Preferences in E-Learning

[6] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, Mustafa I. Al-Najjar, Mining Student Data Using Decision Trees, International Arab Conference on Information Technology (ACIT'2006), Nov. 2006, Jordan.

[7] Vasile Paul Bre_felean, Babe_-Bolyai University, Faculty of Economics and Business Administration, Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment, https://www.researchgate.net/publication/4266105 (Article in a conference proceedings)

[8] Subitha Sivakumar, Rajalakshmi Selvaraj, Predictive Modeling of Students Performance Through the Enhanced Decision Tree Advances in Electronics, Communication and Computing pp 21-36, 2017

[9] Sachin, R.B., Vijay, M.S.: A survey and future vision of data mining in educational field. In: 2012 Second International Conference on Advanced Computer Engineering & Communication Technologies, pp. 96–100 (2012)

[10]Norwawi, N.M., Abdusalam, S.F., Hibadullah, C.F., Shuaibu, B.M.: Classification of students performance in computer programming course according to learning style. In: 2nd Conference on Data Mining and Optimization, pp. 37–41 (2009)

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

CHANGING TRENDS OF PREFERENCES IN MODE OF TRANSACTIONS-A PREDICTION USING ROUGH SET THEORY

Preethi M.
MSc Information Technology
M.O.P Vaishnav College for Women
Chennai-,600034 India
e-mail: m.preethiy@gmail.com

ABSTRACT

Rough Set Theory is a new technique that deals with fuzziness and improbability stressed in decision making. Data mining is a discipline that has an important contribution to data analysis, discovery of new significant knowledge, and independent decision making. The rough set theory offers a feasible approach for decision rule extraction from data. The introduction of demonetisation resulted in elimination of high valued currency notes. It aimed to achieve the goal of a 'less cash' society. Digital trades bring in better scalability and responsibility. Recently RBI has also disclosed its document- "Payments and Settlement Systems in India: Vision 2018" boosting the electronic payments and to help INDIA grow from cash to cashless society in the long run. Thus giving this model an overlook, this paper focuses on studying the views of people on evolution of cashless economy and their comfort level with it. The study was conducted in Chennai; data was collected with the help of organised questionnaire and analysed using rough set theory.

Keywords: Data mining, analysis, Rough Set Theory, cashless transaction, objects

I. INTRODUCTION

To generate information it needs enormous collection of data. It is essential, to develop powerful tool for interpretation of such data and for the abstraction of exciting knowledge that could help in decision-making. The only solution is 'Data Mining'. Data mining is a process of extraction of useful information and patterns from huge and unstructured data. The motive of data mining effort is either to generate a descriptive model or a predictive model. A descriptive model presents the main features of the data set and a predictive model is to predict an unknown value of a specific variable; the target variable. There are many techniques available in data mining in which regression is of the method. This strategy is adopted for prediction. Independent variables are features which are already known and are used for prediction of decision values for a new object. Rough Set Theory is one of the powerful tools which can be implemented to find hidden patterns and generate rules. This technique is used to deal with imprecise data.

Cash is any legal medium of exchange that is free of restrictions. The demonetisation effect has given incredible enhancement to the cashless transactions because of the hindrance of high denomination currencies. The circulation of hard cash became minimal. The spread of digital payment started to fulfil the goal of demonetisation. Apart from that, it has become very advantageous to common people like convenience, discounts, budget discipline etc. Some of the cashless modes are cheque, Demand draft, debit card, credit card, net banking, mobile and e-wallets etc.

The main objective of the paper is to find the preferences of people regarding the mode of transactions and their comfort level with the cashless transactions. People preferences were studied by collecting data from them in the form of questionnaire and analysed by means of data mining approaches. Regression method is used to identify the type of attributes and rough set theory has been used for prediction purposes.

II. REVIEW OF LITERATURE

[1] Discusses about the situation of the country after the arrival of demonetization. The nature of the paper is very descriptive. [1] has covered the consequences, advantages and the disadvantages which were brought by demonetisation. [1] also deliberates the impacts of going cashless and concludes by giving suggestions on what kind of steps should the government take for improving the cashless economy. [2] remarks the exact information about how and when did the tremendous change evolved. [3] gives the picture of many countries of the world which has cashless economy and its features. Then the strengths of India going cashless, weaknesses faced and the available opportunities are conversed and concluded with the threats in going cashless [5][7] explains the rough set theory basic concepts and terminologies with suitable examples and enlightens the process of deriving decision patterns for predictions.

III. METHODOLOGY

Knowledge Discovery process consists of various steps. The steps are collecting data, analysing data sets to discover patterns, building model, evaluation the model using one of the data mining techniques. In this paper, Rough Set Theory is used for evaluating the model and for prediction of the mode preferred by people.

IV. ROUGH SET THEORY

One of the mathematical approaches in data mining is rough set theory. Observations show that the application of this theory has grown world-wide. ZdzislawPawlak projected this theory. This concept is considered as a mathematical tool for imperfect data analysis. It is first non-statistical approach in data mining. With every object in the universe, we subordinate some other object which is comparable to it. This acts as the basic concept for the rough set theory. Objects are categorised based upon the information available with that of the objects whose data are already available. All such objects form a set which is known as elementary sets. With the available knowledge of data, a boundary has to be assigned for each set. The boundary region of the rough

set is given by the variance of the lower and upper approximation. Lower approximation is the depiction of objects which belongs to the subset whereas the upper approximation is the description of objects which most likely belong to that set. The main benefit of rough set theory is that it helps in generating rules from data. In this method, data is structured as a table which is known as information systems. The records denote the objects and the columns represent the attributes. The attributes are broadly classified as condition and decision attributes. The decisions are derived from the rows of the table. This method is very easy to understand. Indiscernibility Relation acts as the chief concept in Rough Set Theory which is an equivalence relation.

V. FROM BARTER TO DIGITAL

In earlier days, barter system used goods as a medium of exchange for purchasing goods or for services. Upon the introduction of currency, the barter system slowly transformed to an economy where currency played a vital role for transactions. The transactions were done using currency notes and coins. It greatly facilitated the transactions. Another way of transaction was the digital method which included internet banking, mobile banking etc. Transactions by hard cash paved the way for growth of black economy. This led to severe economic downfall in India. The government made many efforts to bring all the transactions into banking channel for the prevention of circulation of black money and, systematic and orderly growth of India's economic development. One such effort was the introduction of demonetisation which resulted in elimination of high valued currency notes. This gave rise to the control of cash economy. The cashless transaction is not only safer than the cash transaction but is less time consuming, convenient and lowers risk. People are witnessing the effects of demonetization and many people are left in confusion. Going cashless not only facilities our lives but also helps authenticate and formalize the transactions. This results in increasing the economic growth by controlling corruption and the flow of black money.

VI. PROPOSED SYSTEM

In this paper, we put forward a technique to predict the preference of people in the mode of transaction after the implementation of demonetisation. Euclidean distance measure is used along with the rough set theory in this proposed approach. Rendering the RST, the collected data is organised in the form of table in which the column represents the attributes. The characteristics considered are: Preference of amount of cash having in hand, Mobile banking transactions, Mode of Payment for monthly bills, Convenience, Amount spent as cash in a month, Safety opinion.

Each person's response is considered as an object. For each object 'i' a vector $V_i$is produced consisting of n-tuples, where n is the count of variables. $V_i = (a_1, a_2, a_3,\ldots, a_n)$ where $a_i$varies from 0 to n. Using the RST, the redundant and inconsistent values are eliminated and the decision table is formed.

Each value is assigned with the corresponding 5 scale point and the threshold value is determined. To predict the mode preferred for any given object X, the distance formula is used to find the closeness of the object with the member of the vector.

The Euclidean distance formula is

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

Where x1, x2, y1, y2 are the points in vectors V_i andV_j

The decision attribute of vector V_i is assigned to the object X whose d is minimum.

VII. PERFORMANCE ANALYSIS AND RESULTS

For the proposed system, data was collected from various students, IT professionals and business people. The data objects are associated with both data and knowledge. The preference of mode of transaction by people is the decision variable. The abbreviations and the description of the attributes used in the technique are given along with their values in Table I.

Table I: Attributes and their values

SR. NO	ATTRIBUTE	DESCRIPTION	POSSIBLE VALUES	CORRESPONDING VALUES
1	NBA	Number of Bank Accounts	1,2,3,4,>4	1,2,3,4,5
2	PACH	Preferred Amount of Cash in Hand	<500,500-1000,1000-2000,2000-3500,>3500	5,4,3,2,1
3	MBTR	Mobile Banking Transaction Regularity	Never, Rare, Sometimes, Mostly, Always	1,2,3,4,5
4	RWAA	Regularity in Withdrawal of Amount from ATM	Never, In case of emergency, monthly once, weekly, daily	5,4,3,2,1
5	OPMB	Option for Paying Monthly Bills	Cash, Online Payment	1,2
6	AASE	Average Amount Spent in a month through E-Payment	<1000,1000-2000,2000-2500,2500-4000,>4000	1,2,3,4,5
7	PASC	Percentage of Amount Spent as Cash in a month	20%,20-35%,35-50%,50-70%,>70%	5,4,3,2,1
8	CCDE	Convenience for Cashless payment for Daily Expenditure	Yes, No, Maybe	3,1,2
9	PIE	opinion about Positive Impact in Environment about minimising cash	Yes, No, Maybe	3,1,2
10	CRT	Opinion on Reduction in Theft and reduction in fake currency	Yes, No, Maybe	3,1,2
11	OCD	Opinion on Cashless transactions increment after Demonetisation	Yes, No, Maybe	3,1,2
12	OUCF	Opinion on Usage of Cash in Future	Strongly disagree, disagree ,neutral, agree, strongly agree	1,2,3,4,5

.

Table II: Training data

SR.NO	NBA	PACH	MBTR	RWAA	OPMB	AASE	PASC	CCDE	PIE	CRT	OCD	OUCF	Average	Mode Preferred
1	2	1	1	4	1	1	1	2	3	3	3	5	2.25	Cash
2	1	2	2	4	1	2	2	3	2	3	3	5	2.5	Cash
3	2	1	3	4	2	5	4	3	3	3	3	3	3	Cashless
4	1	4	2	2	1	1	1	1	1	2	2	3	1.75	Cash
5	2	5	4	2	2	5	4	3	3	2	3	4	3.25	Cashless
6	2	4	4	4	2	3	4	2	3	3	3	5	3.25	Cashless
7	2	3	2	4	1	1	1	2	3	3	1	3	2.166	Cash
8	1	5	2	4	1	2	1	2	3	2	2	4	2.41	Cash
9	2	4	4	4	2	2	5	2	2	1	1	3	2.66	Cashless
10	2	4	4	4	2	4	4	2	2	3	3	4	3.16	Cashless

Table II is obtained by applying the assigned values to the collected data objects and the average is calculated. Threshold value should be determined for the training data. It is the minimum value to decide the preference. The threshold value for the sample data is 2.5. If the average is greater than 2.5, then the preference is predicted as cashless mode.

From table II, the following vectors are constructed. For this table, the vectors are V_1, V_2, .., V_{10}.

V_1=(2,1,1,4,1,1,1,2,3,3,3,5)=2.25

V_2=(1,2,2,4,1,2,2,3,2,3,3,5)=2.5

V_3=(2,1,3,4,2,5,4,3,3,3,3,3)=3

V_4=(1,4,2,2,1,1,1,1,1,2,2,3)=1.75

V_5= (2,5,4,2,2,5,4,3,3,2,3,4)=3.25

V_6= (2,4,4,4,2,3,4,2,3,3,3,5)=3.25

V_7=(2,3,2,4,1,1,1,2,3,3,1,3)=2.16

V_8=(1,5,2,4,1,2,1,2,3,2,2,4)=2.41

V_9=(2,4,4,4,2,2,5,2,2,1,1,3)=2.66

V_{10}=(2,4,4,4,2,4,4,2,2,3,3,4)=3.16

The decision attribute for the objects are determined based on the threshold value. Given a new object X, the distance between X and V_i is calculated using Euclidean distance formula. The vector with the minimum distance from X is chosen and the decision variable value of vector V_i is the decision for X.

Let X contain the values (1,4,4,4,2,5,4,3,2,3,3,4). By calculating the distance using the Euclidean distance formula, it is found that the distance between X and V_{10} is minimum.

The vectors and object X are represented in the form of graph to identify the patterns.

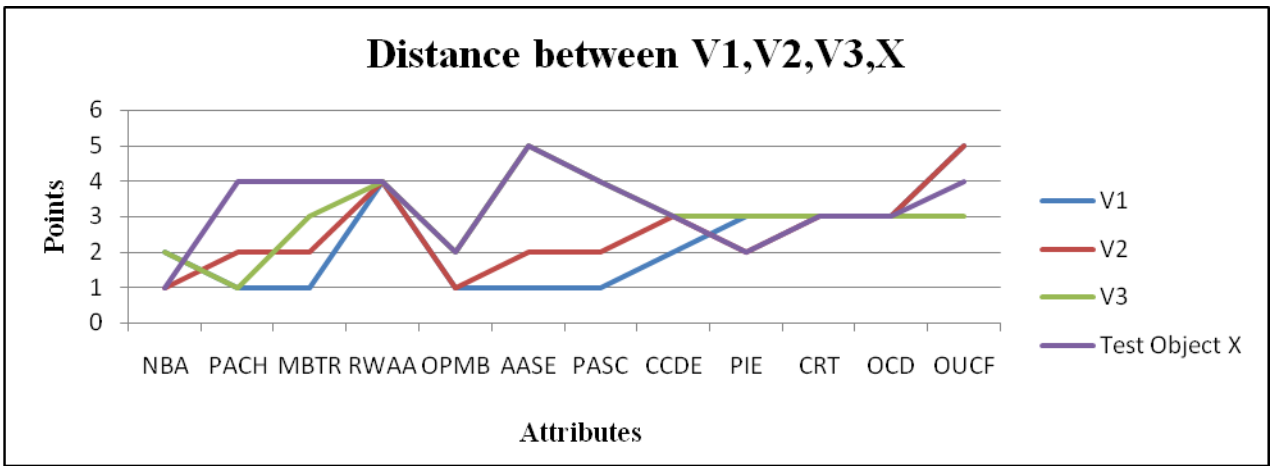


Figure1. Representation of vectors in the form of graph

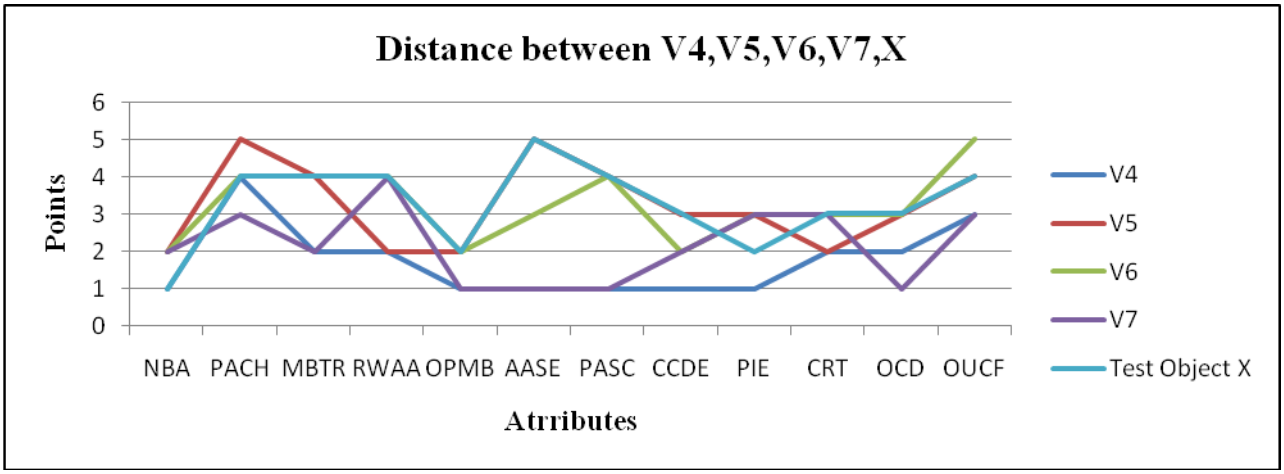


Figure2. Representation of vectors in the form of graph

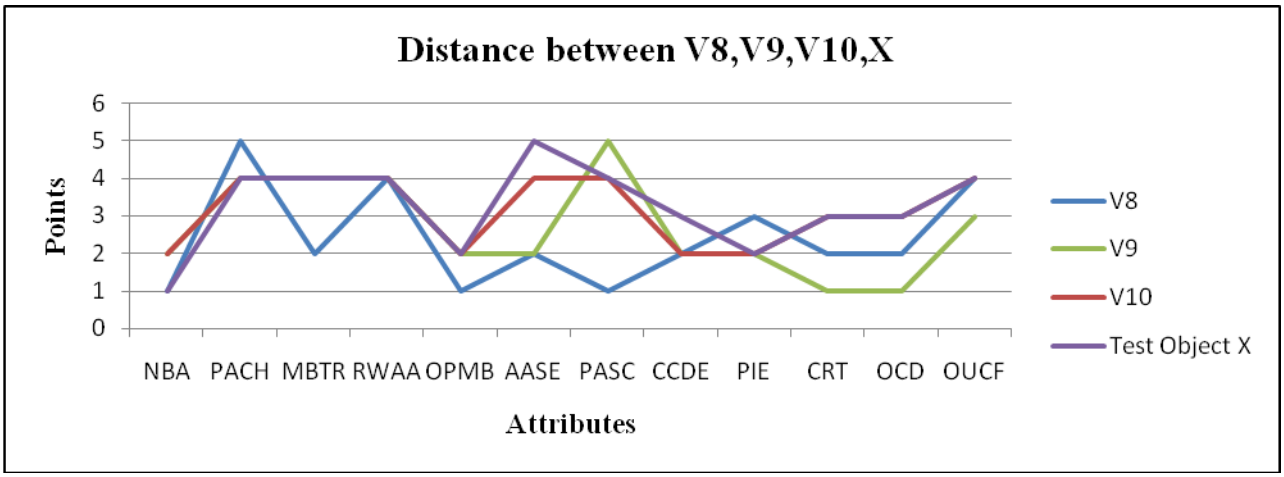


Figure3. Representation of vectors in the form of graph

Figure1, Figure2 and Figure3 shows that vector V_{10} and object X has similar pattern. Both Euclidean distance and the pattern prediction shows that vector V_{10} and object X are similar.

So the decision value of V_{10} is the decision of X. The decision variable for object X is Cashless.

In the similar manner, more test data are evaluated and the decision variable is predicted.

Table III: Sample Test Data

SR.No	NBA	PACH	MBTR	RWAA	OPMB	AASE	PASC	CCDE	PIE	CRT	OCD	OUCF	Mode Preferred
1	2	4	3	4	1	1	1	1	1	1	1	3	Cash
2	2	4	4	4	2	5	5	3	2	3	3	4	Cashless
3	2	4	2	3	1	1	3	1	1	2	3	3	Cash
4	1	5	4	4	2	3	4	2	3	2	3	1	Cashless
5	1	2	4	3	2	5	5	3	2	3	3	2	Cashless
6	3	4	3	4	2	2	4	2	2	3	1	5	Cashless
7	4	1	4	4	2	5	5	3	3	3	2	4	Cashless
8	3	4	5	2	2	5	4	3	3	3	3	4	Cashless
9	1	5	2	4	1	1	1	2	2	2	1	2	Cash

The above model is evaluated for 100 people.

VIII. CONCLUSION

From the above study, 60% of the people have preferred cashless transactions and are convenient with it. It gives the impression that many people agree with the government on the attempt to create a cashless economy which helps to fight against corruption, money laundering. But one of the biggest problems is security. It is important to strengthen Internet Security to protect against online fraud. For smooth implementation of cash less system, some measures are recommended to the Government to bring in transparency and effectiveness in e- payment system. Strategies should be used by government and RBI to encourage cashless transactions by promoting mobile wallets and withdrawing service charge on cards and digital payments. A financial literacy campaign should be conducted by government to make people aware of benefits of electronic payments.

IX. REFERENCES

[1] Pradeep H. Tawade(2017),"FUTURE AND SCOPE OF CASHLESS ECONOMY IN INDIA", IJARIIE , Vol-2 Issue-3 , ISSN (O)-2395-4396.

[2] Ramya, N & et.al(2017), "Cashless transaction: Modes, advantages and disadvantages".

[3] Mahima, "CASHLESS ECONOMY: SWOT ANALYSIS FROM INDIAN PERSPECTIVE", ISBN: 978-93-86171-37-5

[4] Preetigarg & manvipanchal(Apr. 2017),"Study on Introduction of Cashless Economy in India 2016: Benefits & Challenge's", e-ISSN: 2278-487X, p-ISSN: 2319-7668, Volume 19, Issue 4. Ver. II, PP 116-120

[5] Mohammad(2008)," Application of Rough Set Theory in Data Mining for Decision Support Systems (dsss) ", Journal of Industrial Engineering 1 (2008), PP. 25 - 34

[6] Silvia Rissino1 ,"Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications "

[7] Smita, Priti Sharma(May-Jun. 2014), "Use of Data Mining in Various Field: A Survey Paper", e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 16, Issue 3, Ver. PP 18-21

[8] Bharati M & et. al. ,"DATA MINING TECHNIQUES AND APPLICATIONS", Vol. 1 No. 4 301-305.

[9] Deshpande (2010),"DATA MINING SYSTEM AND APPLICATIONS: A REVIEW" ,International Journal of Distributed and Parallel systems (IJDPS) Vol.1, No.1

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

DYNAMIC ENABLEMENT OF LSO (LARGESEND OFFLOAD) IN NETWORK VIRTUALIZED ENVIRONMENT FOR BETTER NETWORK THROUGHPUT

S.Annie Christilla
Associate Professor
Department of Computer Science, St. Francis De Sales College, Bangalore, Karnatka, India

ABSTRACT

LSO is a very important feature of a GigE/10G NICs providing a good amount of performance benefits. Using this feature Network layers can send bigger packets instead of smaller packets. On virtualized environmentNetwork Interface Card(NIC) will be attached to VIOS(Virutal IO servers) and logical partitions will be sharing this NIC through virtualization technique. In these kind of environment, LSO feature can be turned on/off on the NIC, the bridge layerSEA (Shared Ethernet Adapter) present in VIOS and in the virtual adapter present in logical partition. LSO need to be enabled in all these components to exploit this feature. If LSO is turned off on the bridge(SEA) present in the VIOS, will cause poor network performance. In this paper, method to achieve better network throughput when LSO feature is turned off/on dynamically is being proposed in this draft.

***keywords*:** LSO(Largesend Offload), VIOS(Virtual IO Server), SEA(Shared Ethernet Adapter), NIC (Network Interface Card), LPAR (Logical PARtition)

1. INTRODUCTION

Maximum Transmission Unit (MTU) of a network is the maximum protocol data unit that can be transferred on the physical medium. MTU is an inherent property of the physical media. For instance MTU in Ethernet is 1500 bytes. In a Network Protocol Stack, Network Layer or Internet Protocol (IP) layer implements datagram fragmentation so that packets with size larger than the network interface's MTU are fragmented to MTU size before being delivered to the data link layer. Transport protocols such as TCP negotiate MSS (Maximum Segment Size) during connection establishment which is the largest amount of data that TCP is willing to send in a single segment. To avoid IP fragmentation, MSS is always set lower than MTU. Large Send or TCP Segmentation Offload (TSO) is a feature supported by Network Adapters in which the job of fragmenting a larger packets into MTU size is done by the Network Interface Card (NIC) in hardware. network protocol stack can send larger size packets (without having to do the job of fragmentation in software) to NICs and the NIC hardware will do the fragmentation in hardware which would help in improving performance.LSO is a very important feature of a GigE/10G NICs providing a good amount of performance benefits. Using this feature Network layers can send bigger packets instead of smaller packets. However if the network applications have been written in a way to send smaller packets, this hardware feature cannot be exploited well. This is because if applications send smaller packets through the network stack, the stack will end up sending smaller packets to the NIC. With this there would be a lot of packets being sent down from the application to the NIC through the network protocol stack in kernel. This can be avoided by making applications send large size packets so that the lesser number of packets are being sent down by the protocol stack to the NIC

2. PROBLEM STATEMENT

On virtualized environment, LSO is turned on NIC, SEA and Virtual Adapter on the logical partition. When TCP connection is established from LPAR to host that resides outside the system, LPAR will be sending bigger packets upto 64k to the VIOS. The NIC on the VIOS will segment this packet based on MTU and send the data out. When the connection is established if LSO is turned off on SEA, LPAR will keep sending large packet with IF_DF flag on. As a result SEA will not be able to fragment this packet and send ICMP back to LPAR. However LPAR will still continue to send bigger packets. When the retransmission timer expires LPAR will send one packet with size of 1500 bytes. This will result poor network throughput as after every retransmission time out one packet will be sent to the remote host.

3. EXISTING METHOD

When TCP connection gets establishedLPAR will be sending TCP options0x0E0303 along with SYN packet. If SEA has LSO option turned on while it gets response (SYNACK) back from the destinationit will piggy pack the same TCP option. Upon receiving SYNACK and ifTCP option present, the Transport layer onLPAR will identify LSO is turned on SEA and turns on LSO flag for the connection.Post that transport layer will be sending larger packetsto the VIOS. NIC on VIOS will in turn fragmentthe packet and send it to the destination.TCP connection on the LPAR will keep sending largePacketsupto 64k. The checksum field of the packet will have the MSS value that adapter will use the packet to fragment. Later if LSO is turned off on SEA, and largePacket is received from LPAR(>1500 bytes), packets will be fragmented in the SEA layer and sent down to adapter as small packets. If packet also has IF_DF flag, then those packets will get dropped in SEA layer. In this case SEA willbe sending ICMP “fragmention needed” packetback to the LPAR. On receiving this packet LPAR will turn off LSO feature for the connection. Later if LSO is enabled on SEA, there is no wayTCP connections on the LPAR will be aware of and still will be sending smaller packets. This Results in poor network performance.

4. PROPOSED METHOD

In this paper, an algorithm to achieve better network throughput is proposed. On VIOS at SEA layer information (Source IP, Source Port, Destination IP, Destination Port, Largesend capable connection or not) about each connection will be maintained. When LSO is turned off, ICMP message with type 42 will be generated to Source IP, Source Port with information whether LSO is turned off or not. ICMP message will be generated for all the connections. On LPAR once it receives the ICMP message, it turns

of the LSO feature in Transport layer for the specified connection. ICMP message will also have Destination IP and Destination Port and protocol no in the Data section. This will help the LPAR to find the correct connection and turn off the LSO feature. Later lets assume LSO is turned on on SEA then again the ICMP packets will get generated for the LPARS with the information LSO is turned on. Upon receiving this message LPARs will turn on the LSO for the particular TCP connection. At SEA layer, when it receives SYN Packet the entry about connection will be added. This entry will be removed upon receiving FIN for the same connection.

ICMP message used for communication between SEA and LPAR will be type 42 and code will have value either 0 or 1. 0 represents LSO is turned off on SEA and 1 represents LSO is turned on at the SEA layer. LSO at SEA can only be turn on if the LSO is enabled on the underlying network adapter. Data will have SourceIP, Source Port, Destination IP, Destination port and protocol no. Refer Figure 2 for the ICMP format.

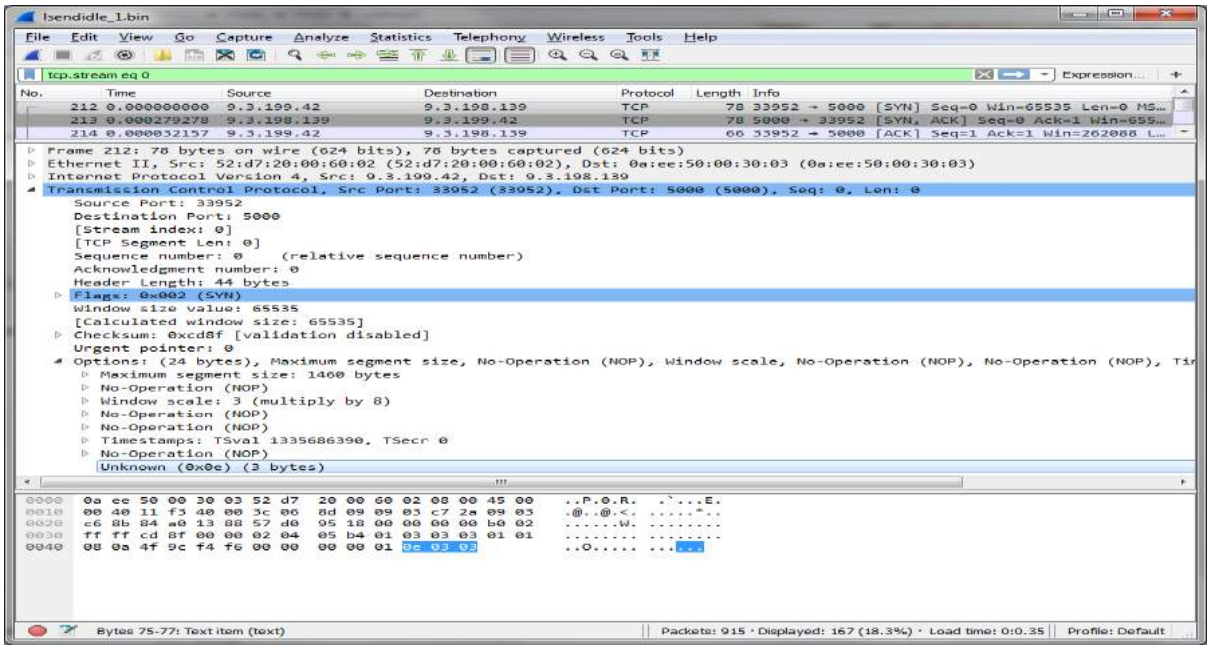


Figure 1. TCP Option to negotiate LSO between LPAR and SEA

Algorithm:

1) During connection establishment phase both LPAR and SEA will exchange the TCP option 0x0E0303 if they are LSO capable.
2) If they are LSO capable, TCP layer on LPAR will send large amount of data (> 1500 bytes) and upto 64K
3) On VIOS entry will be added in the connection cache table with information SourceIP, source Port, Destination IP, Destination port, LSO capable and Protocol number
4) If LSO is turned on the virtual adapter on LPAR, call back will be called and it will search the TCP control block and will turn off LSO for the tcp connections.
5) If LSO is turned off on the SEA, SEA will go through control connection cache table and send ICMP for each entry. ICMP packet type will be 0x42 and code 0 as LSO is turned off. This packet will also have information Source IP, Source Port, Destination IP, Destination port and protocol number.
6) Upon receiving this ICMP packet, LPAR will disable the LARGESEND flag on TCP control block.
7) While sending data, TCP layer on LPAR will not send bigger packet if LARGESEND flag is not set. (As a result SEA need not fragment the packet).
8) Now lets consider a case when LSO is turned on at SEA, then again SEA layer will go through control connection cache table and send ICMP for each entry. ICMP packet type will be 0x42 and code will be 1 as LSO is enabled. This packet will also have information Source IP, Source Port, Destination IP, Destination port and protocol number.
9) Upon receiving this ICMP packet, LPAR will enable the LARGESEND flag on TCP control block.
10) While sending data, TCP layer on LPAR will send bigger packet if LARGESEND flag is set.

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Type								Code								ICMP header checksum															
Data :::																															

Figure 2. ICMP Protocol format

5. CONCLUSION

The objective of the proposed method is to improve the network throughput by transferring the correct sized data to NIC on the VIOS. The proposed method updates the LSO capability of the SEA/underlying NIC to LPAR dynamically as and when it happens, thereby making sure LPAR sends the correct sized data to the NIC on VIOS. This eliminates the fragmentation effort on

VIOS and also when LSO is enabled uses the feature to full extent. Hence the network throughput is improved using this method.

REFERENCES

1. Retreived from https://tools.ietf.org/html/rfc792.
2. Hai Lin, Lucio Correia, & et. al., IBM PowerVM Virtualization
3. Gary R. Wright &[' W. Richard Stevens," TCP/IP Illustrated-Implementation", Vol. 2
4. Kumar Reddy, "Network Virtualization".

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

CINEMA CLOUD: AN ENABLING TECHNOLOGY FOR THE MOVIE INDUSTRY

B. Rasika
Bsc. Computer Science
M.O.P. Vaishnav College for Women
Chennai, India
rasikabalakumar@gmail.com

S. Sonaali
Bsc. Computer Science
M.O.P. Vaishnav College for Women
Chennai, India
sonaali.sonu3@gmail.com

ABSTRACT

With the promise of high definition and advanced special effects, movie theaters worldwide have evolved from traditional film projection to digital cinema projection. There are many things people think about when it comes to the film industry: glamour, money, success, movies stars; but IT is probably low on the list. However, it is cloud computing that is causing a revolution in the film industry. The film industry has decided to fully embrace cloud computing. This makes sense: the film industry is vast and sprawling, it is not just based in Hollywood, but all over the world. The film industry is more like a global village.

Keywords: Motion pictures, entertainment industry, cloud computing, films, web and internet services.

I. Introduction

Technology, such as the 3D one, has been rapidly advancing and changing the film industry. But it's not the only one. Cloud computing and film aren't necessarily two things together; one deals with entertaining people whereas the other aids people with saving and transferring of data from one place to another. From a business perspective, one of the biggest advantages of the cloud is the cost savings, but there actually ends up being a creative advantage, too.

II. Moving to cloud

Nonetheless, the cloud offers some tempting advantages, with budget pressures limiting the resources production companies can spend on media management, the cost savings of replacing on-premise hardware with cloud infrastructure can be substantial. Additionally, the importance of cloud

delivery is growing, and organizations that can enable seamless file distribution will be best positioned to move quickly in a changing environment. Cloud will enable media companies to keep pace with turbulent times by providing:

- Faster time to market
- Increased sales by increasing exposure to content
- A richer flow of information to adapt quickly to changing consumer interests and demand
- Decreased labor, inventory, and working capital costs
- Faster, fresher content packaged, identified, and available to the right consumer anywhere, anytime

III. How does it help?

At first glance it can seem quite hard to see how cloud technology and film fit hand in hand, but if you look at the techniques that are used within the film industry to actually create the feature film; the reasoning behind why many studios are turning to cloud-based systems becomes apparent. The rendering process is one of the most power-hungry aspects of film making; compositing all the scenes of a film with visual effects and audio can take a very long time and requires a huge amount of computing power. The savings can be greater if compared to purchasing the computer equipment needed instead of leasing it. Cloud services can substantially reduce both the capital expenditure and the fixed costs of a visual effects company. They also can get things done a lot faster. Visual effects company Atomic Fiction advocates of using this cloud technology. For example one technology Lionsgate (Entertainment Company) is utilizing to manage the release of its films can be found in the cloud. Distributing productions through services such as Amazon Prime, Netflix, and others is one way this company is adapting to a new era in media consumption. In addition, a great deal of Lionsgate's infrastructure is managed in the cloud, saving the company a great deal of headache of managing its own data centers and bandwidth allocation.

IV. Cloud Computing 3D rendering and animation:

Today, if you look at a nicely 3D rendered image, it's quite difficult to distinguish it from a real photography. Yes the computer which renders it needs a huge system resource. The main barrier for creating such realistic, even more real than a camera can capture is the lack of computing power and partially lack of more advanced software needed for 3D rendering. Just one realistic image consumes more than 4 GB physical RAM and 512 MB Graphical RAM. One minute running video needs at least 30 x 60 = 1800 such images. This needs a huge resource. Four seconds of animation frames, requires about 10K computing cycles to ensure a precise and realistic animated frame. So basically the current barrier to create **Cloud computing 3D Rendering** is the cost and lack of more advanced software.

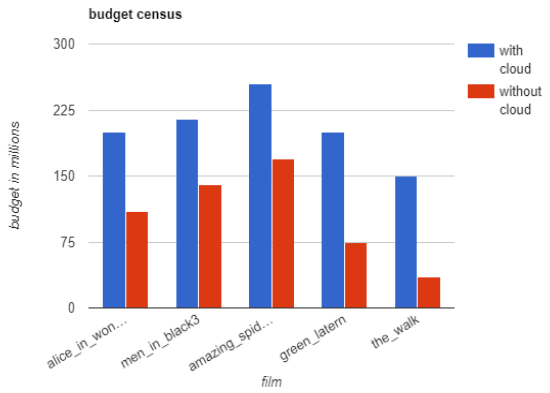
V. Rendering with & without cloud

Cloud-based performance has the potential to cut costs for everyone, if service providers can cater to small studios by thinking big. The 3D animation & visual effects (VFX) industries require substantial data storage and processor-intensive resources. An animated film has taken a larger market share of distributed media. Large production studios have access to substantial resources such as unprecedented server power to support the incredibly expensive and time consuming process of rendering animation data. And since then, for every advancement in technology that creates a

more efficient render system to speed up render times, an increase in data, resolution, and output capability slow them back down again. Cloud-based performance allows multiple workflow streams to benefit from a single cloud-infrastructure. Cloud-based performance certainly benefits large film studios and VFX studios that employ hundreds of animators working towards a similar goal. By queuing performers to the cloud, the animators do not idle their machines rendering frames locally, or slow down their local network with server overload. Studios also do not incur the direct costs of building servers, maintaining them, and hiring staff to manage the workflows. With pay-as-you-use terms, the animation companies can also budget more appropriately the processing needs of a project to their clients.

An animation company believes that company would not have been able to compete with larger studios without access to outsourced cloud computing power. The rendering process of assembling all the component elements of a film such as video, audio, graphics, filters and so on into one final version, can take an agonizingly long time and requires vast computing power. This can be solved with the computing resources of cloud at beneficial cost.

Film	Budget in million(without cloud)	Budget in million(with cloud)
Alice in wonderland	200	110
Men in black 3	215	140
Amazing spider man 2	255	170
Green lantern	200	75
The walk	150	35



Source: 1: http://getwrightonit.com/how-much-does-3d-animation-cost/

Source: 2: https://library.creativecow.net/article.php?author_folder=cow_news&article_folder=VFX_The-Walk-Cloud-Post-Production&page=1

VI. VFX In The Cloud

Overall there are two chief uses VFX artists and firms have for cloud computing: storage and rendering. Storage is important – but for many VFX artists it's incidental. Rendering is the vital — time-consuming — part of the equation. And rendering is increasingly a task moved onto the cloud. Cloud rendering solves one of, if not the largest barrier to entry in the VFX industry. We have some featured VFX tools such as Maya (3D animation, VFX simulation), MotionBuilder (3D character animation and virtual production software), MubBox (3D digital sculpting and texture painting software), 3ds Max (3D software for modeling, animation and rendering).

A. Flexible Computing

Atomic Fiction movies use the cloud for 90-95% of its "heavy data lifting" and have completed notable feature film projects. Zync cloud interface. Zync (maker of visual effects software), one of several similar tools, is a plug-in (is a **software** component that adds a specific feature to an existing computer program. When a program supports plug-ins, it enables customization.) compatible with commercial VFX software that lets an artist access cloud-based rendering functions much in the same way as a local render farm. Zync uses the idle computer time of Amazon's vast data centers they've built all over.

B. Real Time Rendering

One of the most interesting technologies that will propel cloud render farms to even better performance is graphics processor unit or GPU (Graphics processing unit) rendering. GPUs have long been used by gamers to produce stunning graphics in real time that harnesses a GPU's computing power in most commercial 3D packages.

C. Scalability

Decreased cloud render times aren't just time and money saver; there are creative benefits as well. For smaller VFX shops and freelancers access to cloud services allow them to implement more aggressive, complex 3D shots that normally would require hundreds of thousands of hours of computing time with a small render farm.

VII. Cloud gains momentum in media

From the surveys cited, it can be observed that despite the challenges, cloud computing, mainly due to its economic attractiveness, continues to grow. There are good reasons to switch to cloud: low costs, low barriers to entry, increased mobility and scalability. There has been an emergence of content clouds recently. As technical and cost barriers fall and security issues are addressed, the cloud has become a viable platform not only for back-end operations, but also for key business practices, including content management and distribution. During the 2008 presidential election in the United States, the New York Times online was able to handle record traffic using cloud technology. The on-demand nature of massively scalable clouds has enabled media companies to provide more video on demand (VOD) without having to make investments in content delivery networks. By harvesting, hosting and combining their content with other content in the cloud, publishers and media companies can answer a higher level of questions for the customer. Consumer demand rules—getting content to the consumer fast is the key to cloud success.

VIII. Security concerns

While cloud computing is a huge help for the industry when it comes to compiling the final cut, there is still some doubts with regards to the security of this technology, especially in an industry which relies on complete secrecy until the film is ready to be released.

Studios are combating this worry by creating their own, private cloud systems, meaning that they don't necessarily have to worry about the public – or other studios – gaining access to their systems like they would if they used the same cloud systems that Netflix, Amazon and Google do.

This is essentially a huge investment for studios to undertake, but for the sole reason that they can essentially guarantee their work's safety – which is more than enough of an excuse to spend so much when cheaper cloud solutions are available to them.

Security is something that we take very seriously here at Video2DVD, and while we don't use cloud technology to process your VHS (Video Home System) tapes to DVD format.

IX. Conclusion

Cloud computing is here to stay in the film industry, for every frame of film 24 GB of data is processed. The truth is: what is really pushing the film industry forward and lowering costs is possibly the least glamorous thing of all: the cloud. Storing information in the cloud gives you almost unlimited storage capacity. Cloud computing is probably the most cost-efficient method to use, maintain and upgrade. From the graph, we can conclude that the cost spent on creating the movie with cloud is lower compared to the movie without cloud. There are many one-time-payments, pay-as-you-go and other scalable options available, which make it very reasonable for the company in question.

X. References

[1] http://www.bbc.com/news/business-37636099
[2] https://rctom.hbs.org/submission/cloud-based-rendering-for-the-animation-industry/
[3] https://cloudtweaks.com/2012/07/is-cloud-computing-changing-the-film-industry/
[4] https://www.video2dvdtransfers.co.uk/blog/2016/11/23/film-technology-how-cloud-computing-is-revolutionising-the-film-industry/
[5] http://www.ervik.as/how-cloud-computing-is-changing-the-movie-industry/
[6] https://thecustomizewindows.com/2011/10/cloud-computing-3d-rendering-and-scope-in-film-industry/
[7] https://www.autodesk.com/redshift/cloud-computing-adoption-entertainment-studios/
[8] http://www.indiewire.com/2013/05/so-how-exactly-is-cloud-computing-changing-the-vfx-industry-38000/
[9] http://www-935.ibm.com/services/multimedia/fr_FR_Cloud_Computing_for_Media.pdf
[10] https://www.computer.org/csdl/mags/it/2014/05/mit2014050050-abs.html
[11] http://sknr.net/2017/04/05/media-entertainment-industries-turning-cloud-computing/
[12] https://www.zadarastorage.com/blog/tech-corner/film-industry-needs-different-cloud-storage-solution/
[13] https://www.fasthosts.co.uk/blog/cloud/cloud-pixar-and-hollywood-computing
[14] https://www.flandersinvestmentandtrade.com/invest/en/sectors/digital-society/cloud-computing
[15] http://radicalhub.com/cloud-computing-and-the-entertainment-industry/
[16] https://www.youtube.com/watch?v=RaVmo1G6O7k "How Cloud Computing is Changing the Movie Industry"

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

DIABETES DATA ANALYSIS USING MAPREDUCE AND CLASSIFICATION TECHNIQUES

M. Ashok Kumar
Ph.D Research Scholar
Dept. of Computer Science
Periyar University
Salem-11
williamashok@gmail.com

Dr. I. Laurence Aroquiaraj
Assistant Professor
Dept. of Computer Science
Periyar University
Salem-11
laurence.raj@gmail.com

ABSTRACT

Data mining techniques can be applied to extract valuable knowledge from data repositories, e.g. through clustering, classification or association rule mining. Mapreduce is a programming technique which is suitable for analyzing large data sets that otherwise cannot fit in your computer's memory. The research hopes to propose a quicker and more efficient technique of diagnosing the disease, leading to timely treatment of the patients. After mapreduce the classification techniques such as KNN and SVM are applied. The performance of classification techniques are analyzed and interpreted.

Keyword: Classification**,** MapReduce, KNN, SVM.

1. Introduction

Data mining is an extraction of hidden predictive information from large database. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods.

Consequently, data mining consists of more than collecting and managing data; it also includes analysis and prediction[1]. With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important to develop powerful means for analysis and perhaps interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. Limitation of Data Mining are primarily data or personnel-related rather than technology-related. Data mining is one step in the KDD process. It is the most researched part of the process. In that effects of diabetes have been reported to have a more fatal and worsening impact on women than on men because of their lower survival rate and poorer quality of life. WHO reports state that almost one – third of the women who suffer from diabetes have no knowledge about it. The effect of diabetes is unique in the case of mothers because the disease is transmitted to their unborn children. Strokes, miscarriages, blindness, kidney failure and amputations are just some of the complications that arise from this disease [5]. For the purposes of this paper, the analyses of diabetes cases have been restricted to pregnant women.

Generally a person is considered to be suffering from diabetes, when blood sugar levels are above normal (4.4 to 6.1 mmol/L) [1]. Pancreas present in the human body produces insulin, a hormone that is responsible to help glucose reach each cell of the body. A diabetic patient essentially has low production of insulin or their body is not able to use the insulin well. There are three main types of diabetes, viz. Type 1, Type 2 and Gestational [2].

- Type 1 – The disease manifest as an autoimmune disease occurring at a very young age of below 20 years. In this type of diabetes, the pancreatic cells that produce insulin have been destroyed.
- Type 2 - Diabetes is in the state when the various organs of the body become insulin resistant, and this increases the demand for insulin. At this point, pancreas doesn't make the required amount of insulin. Gestational diabetes tends to occur in pregnant women, as the pancreas don't make sufficient amount of insulin. All these types of diabetes need treatment and if they are detected at an early state, one can avoid the complications associated with them.

Now a day, a large amount of information is collected in the form of patient records by the hospitals. Knowledge discovery for predictive purposes is done through data mining, which is an analysis technique that helps in proposing inferences [6]. This method helps in decision-making through algorithms from large amounts of data generated by these medical centers. Considering the importance of early medical diagnosis of this disease, data mining techniques can be applied to help the women in detection of diabetes at an early stage and treatment, which may help in avoiding complications.

2. Literature Review

A literature review reveals many results on diabetes carried out by different methods and materials of diabetes problem in India. Many people have developed various prediction models using data mining to predict diabetes. Combination of classification-regression-genetic-neural network, handles the missing and outlier values in the diabetic data set, and also they replaced the missing values with domain of the corresponding attribute [13].

The classical neural network model is used for prediction, on the pre-processed dataset. In predictive analysis of diabetic treatment using regression based data mining techniques to diabetes data, they discover patterns using SVM algorithm that identify the best mode of treatment for diabetes across different age [14]. They concluded that drug treatment for patients in the young age group can be delayed whereas; patients in the old age group should be prescribed drug treatment immediately. Prediction and classification of various type of diabetes using C4.5 classification algorithm was carried out in Pima Indians Diabetes Database [15].

A hybrid combination of Classification and Regression Trees (CART) and Genetic Algorithms to impute missing continuous values and Self Organizing Feature Maps (SOFM) to impute categorical values was improved in [18]. Deploying a health information exchange (HIE) repository promote and integrate the data within a single point of robust data sharing. This

sharing of information and electronic communication systems enable access to health services and also promotes additional care over dual eligible patients. It recognizes which patient is requiring more care and attention than others. It gives needed data to determine which strategies should be put in place to maximize positive behavior modification [19].

The predictive analytics works in three areas such as Operations management, Medical management and biomedicine, and System design and planning. Healthcare predictive analytics system can help one of the issues that is to address the cost of patients being repeatedly admitted and readmitted to a hospital for chronic diseases which is similar or multiple. The survey of New England Journal of Medicine tells that one in five patients suffer from preventable readmissions. Therefore,1% of the population accounts for 20% of all US healthcare expenditures almost and 25% for over 80% of all expenditures [20].

Various big data technology stack and research over health care combined with efficiency. Cost savings, etc., are explained in better healthcare [21]. The hadoop usage in health care became more important to process the data and to adopt the large scale data management activities. The analytics on the combined compute and storage can promote the cost effectiveness to be gained using hadoop [22].

In [24] Fuzzy Ant Colony Optimization (ACO) was used on the Pima Indian Diabetes dataset to find set of rules for the diabetes diagnosis. The paper [8] approached the aim of diagnoses by using ANNs and demonstrated the need for preprocessing and replacing missing values in the dataset being considered.

Hence, there is a requirement of a model that can be developed easily providing reliable, faster and cost effective methods to provide information of the probability of a patient to have diabetes. In the present work, an attempt is made to analyze the diabetes parameters and to establish a probabilistic relation between them using Naïve Bayes and Decision Tree approach. For the purpose of analysis the models are tested depending on the percentage of correctly classified instances in the dataset.

3. Methods and Materials

3.1. MapReduce

To use an implementation of MapReduce to manage many large-scale computations in a way that is tolerant of hardware faults.

In brief, a MapReduce computation executes as follows:

1. Some number of Map tasks each is given one or more chunks from a distributed file system. These Map tasks turn the chunk into a sequence of key-value pairs. The way key-value pairs are produced from the input data is determined by the code written by the user for the Map function.
2. The key-value pairs from each Map task are collected by a master controller and sorted by key. The keys are divided among all the Reduce tasks, so all key-value pairs with the same key wind up at the same Reduce task.
3. The Reduce tasks work on one key at a time, and combine all the values associated with that key in some way. The manner of combination of values is determined by the code written by the user for the Reduce function.

3.1.1 Extensions to MapReduce

MapReduce has proved so influential that it has spawned a number of extensions and modifications. These systems typically share a number of characteristics with MapReduce systems:

1. They are built on a distributed file system.
2. They manage very large numbers of tasks that are instantiations of a small number of user-written functions.
3. They incorporate a method for dealing with most of the failures that occur during the execution of a large job, without having to restart that job from the beginning

3.2. Classification Technique

The classification of big data is becoming an essential task in a wide variety offields such as biomedicine, social media, marketing, etc. The recent advance in data gathering in many of these fields has resulted in an inexorable increment of the data that we have to manage. The volume, diversity and complexity that bring big data may hinder the analysis and knowledge extraction processes [9]. Under this scenario, standard data mining models need to be re-designed or adapted to deal with this data. The k-Nearest Neighbor algorithm (k-NN) [2] is considered one of the ten most influential data mining algorithms [10].

A medical diagnosis is a classification process. A physician has to analyze lot of factors before diagnosing the diabetes which makes physician's job difficult. In recent times, machine learning and data mining techniques have been considered to design automatic diagnosis system for diabetes [12]. Recently, there are many methods and algorithms used to mine biomedical datasets for hidden information, including Neural networks (NNs), Decision Trees (DT), Fuzzy Logic Systems, Naive Bayes, SVM, cauterization, logistic regression and so on [11]. These algorithms decrease the time spent for processing symptoms and producing diagnoses, making them more precise at the same time.

4. Proposed Work

MapReduce is a programming framework [9] to process largescale data in a massively parallel way. MapReduce has two major advantages: the programmer is oblivious of the details related to the data storage, distribution, replication, load balancing, etc.; and furthermore, it adopts the familiar concept of functional programming. The programmer must specify only two functions, a map and a reduce. The typical framework is as follows [15]: (a) the map stage passes over the input file and outputs (key, value) pairs; (b) the shuffling stage transfers the mappers' output to the reducers based on the key; (c) the reduce stage processes the received pairs and outputs the final result. Due to its scalability, simplicity and the low cost to build large clouds of computers, MapReduce is a very promising tool for large scale data analysis, something already reflected in academia (see [12] [11] for examples).

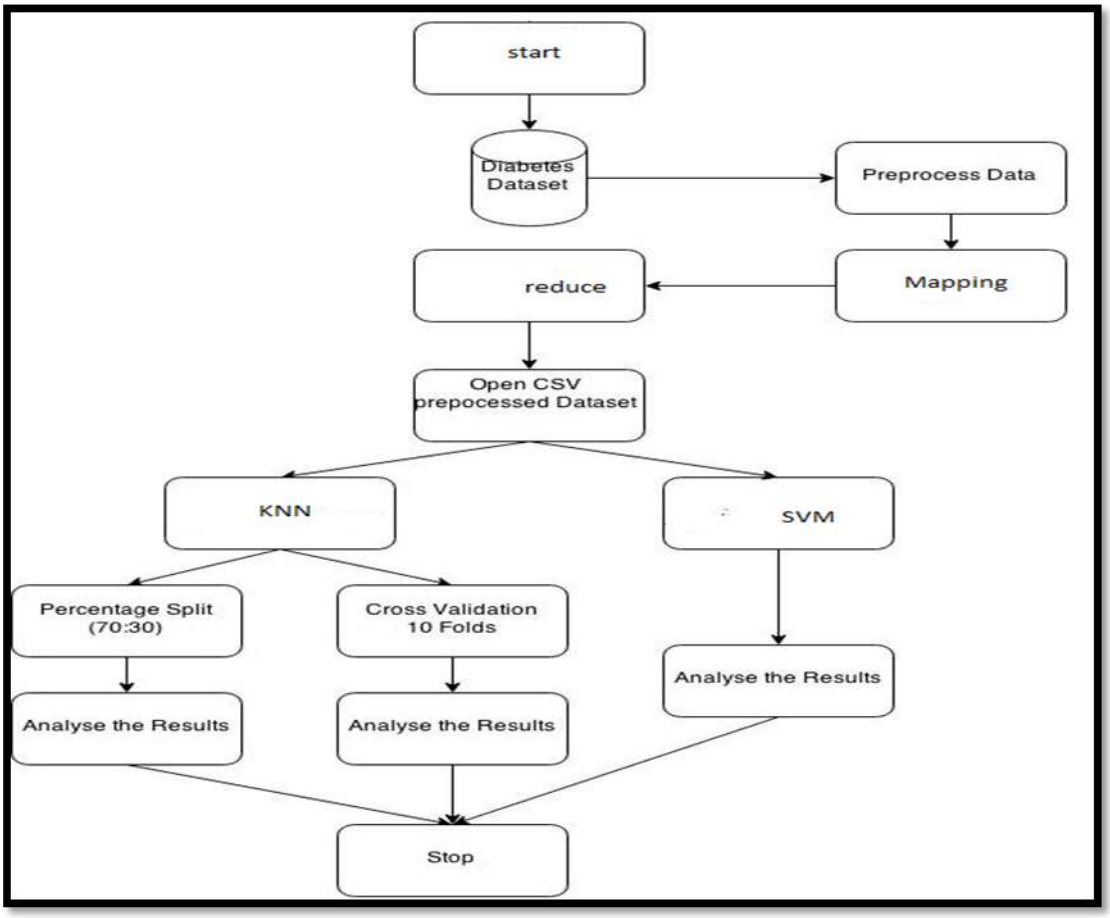


Figure 4.1: Proposed Methodology

4.1. MapReduce Framework

MapReduce [6] is a popular programming framework to support data-intensive applications using shared-nothing clusters. In MapReduce, input data are represented as key-value pairs. Several functional programming primitives including Map and Reduce are introduced to process the data. Map function takes an input key-value pair and produces a set of intermediate key-value pairs. MapReduce runtime system then groups and sorts all the intermediate values associated with the same intermediate key, and sends them to the Reduce function. Reduce function accepts an intermediate key and its corresponding values, applies the processing logic, and produces the final result which is typically a list of values.

MapReduce (MR) is a programming framework developed by Google to address the previous problems. An MR program requires (at least) two components:

1 A mapper is used to filter the input data.

2 A reducer performs a summary of the information provided by the mapper.

The MR framework takes charge of running in parallel multiple mappers/reducers, handles data redundancy, faults, etc.

Formal definition

Input data to the problem must be composed of key/value pairs (k, v), belonging to two generic domains k ∈ Min and v ∈ Vin. The data is initially filtered according to a function:

$$MAP(k, v) = list(k2, v2)$$

where the output data can belong to different domains k2 ∈ Mmap and v2 ∈ Vmap. The results from the map operations can be shuffled and collected, and finally reduced using a different function:

$$REDUCE(k2, list(v2)) = (k2, list(v3)),$$

with v3 ∈ Vout.

4.2. MapReduce workflow in MATLAB

- The input data is saved in a particular object called datastore, which handles data distribution and partitioning in chunks.
- Each data chunk is processed by a different map function, and the result is stored in an intermediated object of class KeyValueStore.
- The intermediate outputs are grouped by key (i.e. by k2 in our formal definition).
- Each group of KeyValueStore elements is processed by a reduce function.
- Final results are saved in an output datastore object.

4.2.1. Classification Techniques using MATLAB

Thedesigned model allows the k-Nearest neighbor classifier to scale todatasets ofarbitrarysize,justbysimplyaddingmorecomputingnodes if necessary. Moreover, this parallel implementation provides the exact classification rate as the original K-NN model.The conducted experiments, using a dataset with up to 1million instances, show the promising scalability capabilities ofthe proposedapproach.Diabetes mellitus is one of the most serious health challenges in both developing and developed countries. According to the International Diabetes Federation, there are 285 million diabetic people worldwide. This total is expected to rise to 380 million within 20 years.

The proposed method uses SupportVector Machine (SVM), a machine learningmethod as the classifier for diagnosis of diabetes.The machine learning method focuses onclassifying

diabetes disease from highdimensional medical dataset. The experimentalresults obtained show that support vectormachine can be successfully used for diagnosingdiabetes disease. SVM with Radial basis function kernel is used for classification. The performance parameters such as the classification accuracy, sensitivity, and specificity of the SVM and RBF have found to be high thus making it a good option for the classification process.

4.3. K - Nearest Neighbor Algorithm

KNN is a method which is used for classifying objects based on closest training examples in the feature space. KNN is the most basic type of instance-based learning or lazy learning. It assumes all instances are points in n-dimensional space.K-nearest neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-nearest neighbor category. The purpose of this algorithm is to classify a new object based on attributes and training samples. This algorithm used neighborhood classification as the prediction value of the new query instance. A distance measure is needed to determine the "closeness" of instances. KNN classifies an instance by finding its nearest neighbors and picking the most popular class among the neighbors.

4.3.1. Features of KNN

a) All instances of the data correspond to the points in an n-dimensional Euclidean space

b) Classification is delayed till a new instance arrives

c) In KNN, the Classification is done by comparing feature vectors of the different points in a space region.

d) The target function may be discrete or realvalued.

An arbitrary instance is represented by($a_1(x)$, $a_2(x)$, $a_3(x)$,.., $a_n(x)$), where $a_i(x)$ denotes features. Euclidean distance between two instances $d(x_i, x_j)$=sqrt (sum for r=1 ton $(a_r(x_i) - ar(x_j))^2$). The k-nearest neighbor algorithm is simplest of all machine learning algorithms and it is analytically tractable. In KNN, the training samples are mainly described by n-dimensional numeric attributes. The training samples are stored in an dimensional space. When a test sample (unknown class label) is given, k-nearest neighbor classifier starts searching the 'k' training samples which are closest to the unknown sample or test sample. Closeness is mainly defined in terms of Euclidean distance. The Euclidean distance between two points P and Q i.e. P (p_1,p_2, …. P_n) and Q (q_1, q_2,..q_n) is defined by the following equation:-

$$d(P,Q) = \sum_{i=1}^{n} (P_i - Q_i)^2$$

The Simple KNN algorithm is

- Take a sample dataset of n columns and m rows named as R .In which n-1th columns near the input vector and nth column is the output vector
- Take a test dataset of n-1 attributes and y rows named as P.
- Find the Euclidean distance between every S and T
- Then, Decide a random value of K. K is the no. of nearest neighbors.
- Then with the help of these minimum distance and Euclidean distance find out the nth column of each.
- Find out the same output values.

If the values are same, then the patient is diabetic, otherwise not. After this, the accuracy rate and the error rate of the data set are being calculated. The accuracy rate shows that how many outputs of the test dataset are same as the output of the data of different features of the training dataset. The error rate is showing that how many outputs of the data of the test dataset are not same as the output of the data of different features of the training dataset. KNN is a highly effective inductive inference method for noisy training data and complex target functions.

4.3.2. Algorithm: K Nearest Neighbor Approach

Let G = {g_1,g_2,…,g_n} be a set of n labeled objects. X={x_1,x_2,…,x_m} be a training vector with known class labels. Y={y_1,y_2,…,y_k} be a set of testing gene vector without class labels. The algorithm is defined as follows:

Algorithm: KNN

Input: Training Diabetics data With Class Labels.
Testing Diabetics data Without Class Labels.
Value For K = Number Of K Nearest Neighbors.

Output: Predicted Classes For Test Data.

begin
Input y, of unknown classification.
Set K, $1 \leq K \leq n$.
Initialize i=1.
do until (K-nearest neighbors found)
Compute distance from y to $x_i$using Equation[5.6],[5.7]
if ($i \leq K$) THEN
Include x_i in the set of K-nearest neighbors
else if (x_i is closer to y than any previous nearest neighbor)THEN
Delete farthest in the set of K-nearest neighbors
Include xiin the set of K-nearest neighbors.
end if
Increment i.
end do until
Determine the majority class represented in the set of K-nearest neighbors.
if (no tie occurs) then
Classify y in the class of minimum sum
else
Classify y in the class of last minimum found.

```
end if
else
 Classify y in the majority class
end if
end
```

4.4. Support Vector Machine

SVMs are set of related supervised learning methods used for classification and regression [2]. They belong to a family of generalized linear classification. A special property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separate the data. The separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalization error of the classifier will be [2]. We consider data points of the form

$$\{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)\ldots\ldots\ldots,(x_n, y_n)\}.$$

Where y_n=1 / -1 , a constant denoting the class to which that point x_n belongs. n = number of sample. Each xn is p-dimensional real vector. The scaling is important to guard against variable (attributes) with larger variance. We can view this Training data , by means of the dividing (or separating) hyper plane , which takes

$$w \cdot x + b = o$$

Where b is scalar and w is p-dimensional Vector. The vector w points perpendicular to the separating hyperplane. Adding the offset parameter b allows us to increase the margin. Absent of b, the hyperplane is forced to pass through the origin, restricting the solution. As we are interesting in the maximum margin, we are interested SVM and the parallel hyperplanes. Parallel hyperplanes can be described by equation

$$w.x + b = 1$$
$$w.x + b = -1$$

If the training data are linearly separable, we can select these hyperplanes so that there are no points between them and then try to maximize their distance. By geometry, We find the distance between the hyperplane is $2/|w|$. So we want to minimize $|w|$. To excite data points, we need to ensure that for all I either

$$w.\, xi - b \geq 1 \text{ or } w.\, xi - b \leq -1$$

This can be written as

$$yi\,(w.\, xi - b) \geq 1\,,\ 1 \leq i \leq n$$

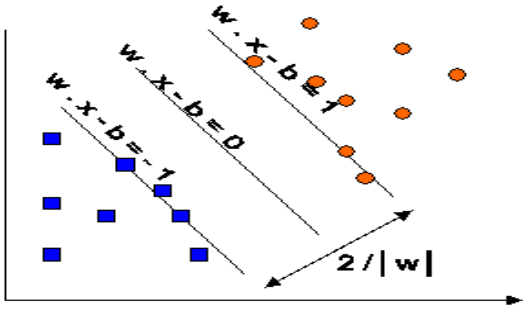


Figure 4.2: Maximum margin hyperplanes for a SVM trained with samples from two classes

5. Experiment Analysis and Result

The work explores the aspect of ANN and SVM Classifier as Data Mining techniques in determining diabetes in women. The main objective is to forecast if the patient has been affected by diabetes using the data mining tools by using the medical data available. The classification type of data mining has been applied to the Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases. Table 5.1 shows a brief description of the dataset that is being considered.

Table 5.1: Dataset Description.

Dataset	No. of Attributes	No. of Instances
Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases	8	768

Table 5.2 : Attribute Description.

Attribute	Relabeled values
Number of times pregnant	Preg
Plasma glucose concentration	Plas
Diastolic blood pressure (mm Hg)	Pres
Triceps skin fold thickness (mm)	Skin
2-Hour serum insulin	Insu
Body mass index (kg/m2)	Mass
Diabetes pedigree function	Pedi
Age (years)	Age
Class Variable (0 or 1)	Class

5.1.Validation measures

In this work, the accuracy measures Precision, Recall and specificity were used for measuring accuracy rate of three classification algorithms namely Fuzzy Soft set based classification, K-nearest neighbor approach and Fuzzy K-NN algorithm [4].

5.2. Precision

Precision is a measure of the accuracy provided that a specific class has been predicted. It is defined by:

$$\text{Precision} = tp/(tp+fp)$$

Where tp and fp are the numbers of true positive and false positive predictions for the considered class.

5.3. Recall

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. It is commonly also called sensitivity, and corresponds to the true positive rate.

$$\text{Recall/Sensitivity} = tp/(tp+fn)$$

5.4. Specificity

Specificity, which is a measure that is commonly, used in two class problems where one, is more interested in a particular class. Specificity corresponds to the True – negative Rate.

$$\text{Specificity} = tn/(tn+fp)$$

5.5. Overall classification Accuracy

Accuracy is the overall correctness of the model and is calculated as the sum of correct classifications divided by the total number of classification.

Accuracy=(True Classification)/(Total no of cases).

5.6.Performance Evaluation

Performance of Classification algorithms is evaluated by using accuracy measures for the diabetic's dataset on before and after dimensionality reduction using mapreduce.

The performance of Classification algorithms is analyzed based on Precision, Sensitivity and specificity validity measures on before and after MapReduceand the results are shown in Table 5.3.

Table 5.3: Performance evaluation

Accuracy Measures	**Before Mapreduce**		**After Mapreduce**	
	KNN	**SVM**	**KNN**	**SVM**
Precision	0.78	0.83	0.82	0.85
Sensitivity	0.75	0.78	0.80	0.82
Specificity	0.86	0.90	0.91	0.94

The performance of the KNN Classification algorithm is analyzed based on Precision , Sensitivity and specificity validity measures on before and after MapReduceand the results are shown in Fig 5.1.

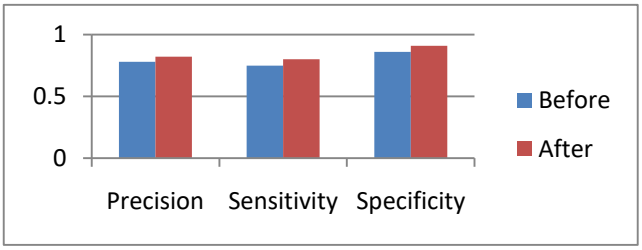


Figure 5.1: KNN Performance Evaluation

The performance of SVM Classification algorithms is analyzed based on Precision, Sensitivity and specificity validity measures on before and after MapReduceand the results are shown in Fig 5.2.

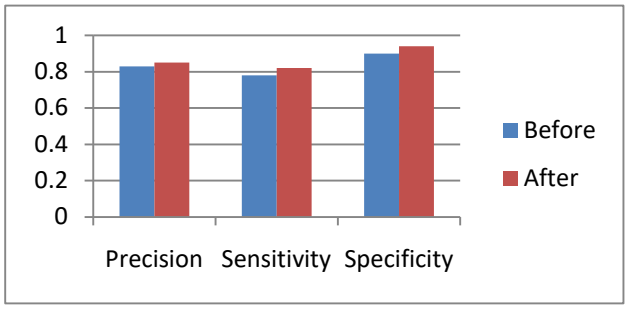


Figure 5.2: SVM Performance Evaluation

The comparative analysis of Classification algorithms is analyzed based on Precision, Sensitivity and specificity validity measures on after MapReduceand the results are shown in Table 5.4.

Table 5.4: Comparative Analysis of ANN and SVM

Accuracy Measures	Classification algorithms	
	KNN	SVM
Precision	0.82	0.85
Sensitivity	0.80	0.82
Specificity	0.91	0.94

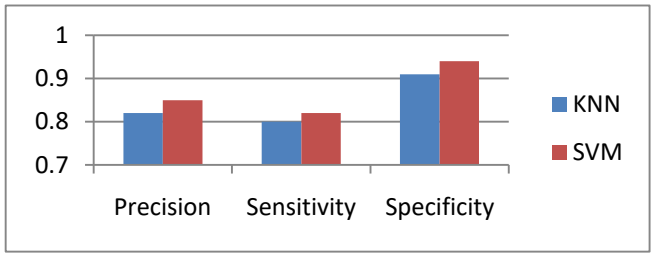


Figure 5.3: Comparative Analysis

Table 5.5: Overall Accuracy for Classification Algorithms

Classification Algorithm	**Accuracy**
KNN	88 %
SVM	93 %

6. Conclusion

The automatic diagnosis of diabetes is an important real-world medical problem. Detection of diabetes in its early stages is the key for treatment. This work shows how KNN and SVM are used to model actual diagnosis of diabetes for local and systematic treatment, along with presenting related work in the field on before and after MapReduce. Experimental results show the effectiveness of the proposed model. This research work also shows the importance of the MapReduceapproach for the performance of classification techniques after MapReduceis better than the performance before MapReduce. In future it is planned to gather the information from different locales over the world and make a more precise and general prescient model for diabetes conclusion. Future study will likewise focus on gathering information from a later time period and discover new potential prognostic elements to be incorporated. The work can be extended and improved for the automation of diabetes analysis.

References

1. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kauffman Publishers, 2001
2. S. Kumari and A. Singh, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Proceedings of Seventh International Conference on Intelligent Systems and Control, pp. 373-375, 2013.
3. C. M. Velu and K. R. Kashwan, "Visual Data Mining Techniques for Classification of Diabetic Patients", 3rd IEEE International Advance Computing Conference (IACC), 2013
4. S.Sankaranarayanan and Dr Pramananda Perumal.T, "Predictive Approach for Diabetes Mellitus Disease through Data Mining Technologies", World Congress on Computing and Communication Technologies, pp. 231-233, 2014.

5. Mostafa Fathi Ganji and Mohammad Saniee Abadeh, "Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease", Proceedings of ICEE 2010, May 11-13, 2010
6. T.Jayalakshmi and Dr.A.Santhakumaran, "A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks", International Conference on Data Storage and Data Engineering, 2010.
7. Sonu Kumari and Archana Singh, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Proceedings of 71th lnternational Conference on Intelligent Systems and Control ISCO – 2013.
8. White, A.P., Liu, W.Z.: Technical note: Bias in information-based measures in decision tree induction. Machine Learning 15(3), 321–329, 1994.
9. A.S. Manjunath, M.A. Jayaram, "Application of Genetic Algorithm Optimized Neural Network Connection Weights for Medical Diagnosis of Pima Indians Diabetes". International Journal on Soft Computing (IJSC), Vol.2, No.2, May 2011.
10. Changjing Shang and Qiang Shen, "Aiding Classification of Gene Expression Data with Feature Selection: A Comparative Study", International Journal of Computational Intelligence Research. ISSN 0973-1873 Vol.1, No.1, pp. 68–76, 2005.
11. Ping Chang and Jeng-Shong Shih, "The Application of Back Propagation Neural Network of Multi-channel Piezoelectric Quartz Crystal Sensor for Mixed Organic Vapours".Tamkang Journal of Science and Engineering, Vol. 5, No. 4, pp. 209-217, 2002.
12. Pradipta Maji and Sankar K. Pal, "Fuzzy–rough sets for information measures and Selection of relevant genes from microarray data", IEEE transactions on systems, man, and cybernetics—part b: cybernetics, vol. 40, no. 3, June 2010.
13. Qiang Shen, Alexios Chouchoulas, " A Rough fuzzy approach for generating classification rules", ww.elsevier.com/locate/patcog, Pattern R ecognition 35 (2002) 2425 – 2438.
14. Ronaldo C. Prati, Gustavo.E. A, Batista.P.A, and Maria C. Monard, "Class imbalances versus class overlapping: an analysis of a learning system behavior". Springer-Verlag Berlin Heidelberg 2004.
15. Sellappan Palaniappan, Tan Kim Hong, "Discretization of Continuous Valued Dimensions in OLAP Data Cubes".IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.11, November 2008.
16. K.Thangavel, P.Jaganathan,A.Pethalakshmi, Karnan.,"Effective Classification with Improved Quick Reduct For Medical Database Using Rough System", BIME Journal, Volume (05), Issue (1), Dec., 2005.
17. Karegowda, M.A. Jayaram, A.S. Manjunath, "Cascading K-means Clustering and KNearestNeighbor Classifier for Categorization of Diabetic Patients" IJEAT Vol.1 No.3 pp 147-151,2012.
18. Hardik Maniya, Mosin I. Hasan, Komal P.Patel "Comparative study of Naïve Bayes Classifier and KNN for Tuberculosis", IJCA pp 22-26,2011.
19. W. Yu, and W. Zhengguo (2007), "A Fast KNN algorithm for text categorization", Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, pp.3436-3441
20. Asha Gowda Karegowda ,MA.Jayaram'Integrating Decision Tree and ANN for Categorization of Diabetics Data' International Conference on Computer Aided Engineering, December 13-15, IIT Madras, Chennai, India in 2007.
21. Siti Farhanah Bt Jaafar and DannawatyMohdAli, "Diabetes mellitus forecast using artificial neural networks", Asian conference of paramedical research proceedings, 5-7, September, , Kuala Lumpur, MALAYSIA in 2005.
22. Rajeeb Dey and Vaibhav Bajpai and Gagan Gandhi and Barnali Dey, "Application of artificial neural network technique for diagnosing diabetes mellitus", IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10 in 2008.
23. Y. Angeline Christobel, P.Sivaprakasam, "A New Classwise k nearest Neighbor (CKNN) Method for the Classification of Diabetes Dataset".
24. "Forecast of Diabetes using Modified Radial basis Functional Neural Networks" International Conference on Research Trends in Computer Technologies (ICRTCT) Proceedings published in International Journal of Computer Applications® (IJCA) (0975 – 8887) ,G. Magudeeswaran and D. Suganyadevi, Sreesaraswathi Thyagaraja College Pollachi-642 107, Tamil Nadu in 2005.
25. "Diagnosis of Diabetes Mellitus based on Risk Factors" International Journal of Computer Applications (0975 – 8887) Volume 10– No.4, November 2010
26. Nahla H. Barakat, Andrew P. Bradley, and Mohamed Nabil H. Barakat,"Intelligible support vector machines for diagnosis of diabetes mellitus", IEEE transaction on information technology in Biomedicine, Vol. 14, No. 4, July 2010.
27. A. H. Project, "Apache hadoop," 2015. [Online]. Available: http: //hadoop.apache.org/
28. A. M. Project, "Apache mahout," 2015. [Online]. Available: http: //mahout.apache.org/
29. M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. USENIX Association, 2012, pp. 1–14.
30. A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml in 2010.
31. National Diabetes Information Clearinghouse (NDIC), http://diabetes.niddk.nih.gov/dm/pubs/type1and2/#signs
32. Global Diabetes Community, http://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

IMPENDING USE OF NAME DATA NETWORKING IN VEHICLE-TO-VEHICLE COMMUNICATION

P. Malathi
Asst. Professor, Dept. of Computer Science, Guru Nanak College, Chennai. India
E mail ID: malathii1989@gmail.com

R.Vinoth
Asst. Professor, Dept. of Computer Application, St. Joseph's College, Chennai. India
E mail ID: vinomogi@gmail.com

ABSTRACT

Specialists bring suggested a few substance spread strategies to tending to those monstrous Growth On substance trade. Named information systems administration (NDN) is a standout amongst the novel plans in which networks use named information for content spread As opposed to group personalities. In NDN, those substance itself may be introduced in the system layer dependent upon client hobbies. The utilization for NDN done Vehicle-to-Vehicle (V2V) correspondence need various possibility because of those preferences for named built information recovery against group built information seeking. This paper displays an in-depth survey of the possibility employments of NDN in V2V nature's domain for extraordinary accentuation once their preferences What's more Hindrances. The paper additionally gives future examination course that Might make embraced on the subject.

INTRODUCTION

Named information systems administration (NDN) (Jacobson, 2009) may be another web structural engineering which concentrates primarily on name-centric networking, On lieu of the universal host-centric approach. Recently, various Vehicle-to-Vehicle (V2V) provisions dependent upon NDN bring been recommended (Baid, 2013) (Grassi, 2013). Intuitively, the multi-source way and in-network caching offers from claiming NDN need aid steady of the majority of the data spread On broadly secured locales What's more irregular contact tests which are was troublesome with customary IP-based networks. To instance, An information recovery disappointment because of irregular contact will have the ability should recoup a greater amount fast through the preferences from claiming conveyed caches.

Vehicular systems administration (VN) is a standout amongst the A large portion paramount innovations in the broad network, Additionally Likewise about now; innovations need aid readied for usage Also dissemination. The utilization about vehicle sensors for natural monitoring, space, logistics industry, thus a lot of people different requisitions will be subsequently favorable element. The the vast majority inclined inquiry that analysts need done their personalities is, the thing that is the association between vehicle stage and majority of the data dissemination?.

This paper keeps tabs for moving forward the timing Furthermore promptness for majority of the data conveyance "around vehicles, to achieve shrewdly transportation framework (ITS) dependent upon the idea from claiming NDN of the vehicular organize. ITS consolidate those developments of majority of the data system, sensors, communication, Also calculations with move forward those remarkable execution about transportation. Another exploration region to remote telecommunication need opined from claiming ITS requisition by making correspondence between vehicles for example, such that V2V, which backing information gathering What's more trade for information majority of the data.

The paper may be further sorted out as takes after: V2V correspondence necessities What's more a few purpose are secured in the next segment. Further points for V2V inspiration Furthermore tests are examined. NDN architectures, layered protocol model, What's more investigate tests are likewise exhibited in the paper. Offers Furthermore preferences of NDN to V2V were laid open in the remaining parts of the paper. Theoretical outline from claiming named information vehicular systems administration (NDVN) Likewise commitment Also other proposals finishes up the paper.

V2V CORRESPONDENCE PREREQUISITES

On vehicular network, because of the confinements on the accessible range Furthermore remote networks, the necessities pointed during finer usability from claiming bandwidth, low inactivity. These will improve Dependability of the organize (Puvvala, 2012). Since those whole organize posesunpredictability of the vehicles on the network, hazards, dangers What's more safety majority of the data require sufficient conveyance in time. Those framework necessities very nearly a immaculate organization in the earth thereabouts Likewise on work Likewise a V2V should complete message conveyance Similarly as when necessary (Bhuvaneshwari, 2013).

Previously, V2V, because of its adaptable structure through topology, messages Furthermore data are predestined starting with a vehicle should in turn. Once a vehicle need information, it advances those data on a close-by vehicle for the would like of simultaneously re-forwarding until it gets of the last end. For the reason for the specified correspondence style in V2V, constantly on vehicles need with bring those V2V enabled frill should captivate in the act.

The underlining engineering organization behind V2V correspondence is committed short extend interchanges (DSRC). Units introduced around vehicles permit helter skelter pace correspondence between vehicles What's more infrastructures make up those DSRC. In place with bring a working correspondence for V2V, one needs spectrum, a trait empowers individuals to create provision with low inactivity. Those elementary objective for picking this requisition might furnish those prioritization for majority of the data for example, wellbeing provisions.

V2V correspondences mostly need aid performed for the remote entry vehicle earth (WAVE) innovations. WAVE innovation organization is joined with framework architecture, separate interfaces, Furthermore benefits it gives for example, such that those wild card essential administration set (BSS). WAVE permits those transmission Furthermore gathering from claiming information frames with the wildcard BSS. This characteristic empowers communication-group setup without substantially of the overhead necessary for itinerant IEEE 802. 11a/g. As stated by Puvvala done (Puvvala, 2012), WAVE standard will a adequate score transmit ahead 5. 9 GHz band with the transmission go about regarding 100-500 meters with respect to recurrence. Those worldwide standard over US/Europe that utilization this component may be introduced Concerning illustration a outline around Figure-1.

Concerning illustration An result, it plainly postures the specific prerequisites Previously, each part of the framework. DSRC What's more WAVE are those significant necessities necessary with distribute under V2V correspondence. An bound together V2V skeleton works around exactly critical requirements, to be specific Naming, Scalability, Mobility, capacity and Cache, movement characteristics, security What's more security which would delineated as takes after. Naming.

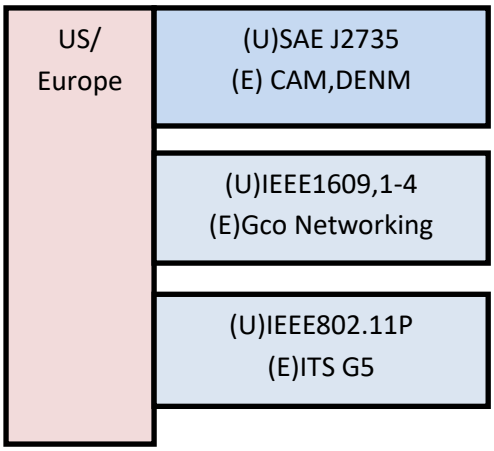


Figure-1.US/Europe standard for short range correspondence (Puvvala, 2012).

Naming sees Concerning illustration an open investigate hole about tests to issues for example, such that flat, progressive Also disseminated manifestations over planning NDN-based V2V provision. Those to start with activity should figure it out to a bound together V2V correspondence skeleton is the capacity on relegate exceptional names inside the extent from claiming every vehicle. Information data trade created Toward vehicles alternately an aggregation of vehicles, necessities fitting naming methodology clinched alongside configuration on empower movement message imparting when required (Wang &Wakikawa, 2010). Firstly, names must be interesting Concerning illustration a application-centric, Along these lines Concerning illustration with furnish those standard for requisition and benefits. Secondly, the names must be precisely intended with meet wellbeing approaches on secure In light of requisition prerequisites. Naming Previously, NDN will be picked as at-types Also hierarchies; however, Distributed-Hash Tables (DHT) would likewise utilized to exactly architectures for NDN e. G. Seen for (Grassi, 2013), (Jacobson, 2009). Versatility.

An significant challenge clinched alongside V2V is posed Toward those gigantic development for vehicular advancements Previously, future At an extensive amount from claiming vehicles would provided with sensors Likewise respects with adaptability Furthermore Growth of the organize. Those test posed is By what means would this innovation organization provides for effective correspondence that reveals to great guarantee for a large-scalable sending from claiming V2V helpful wellbeing frameworks (Puvvala, 2012) Furthermore (White, 2009). A coordinated circuit V2V correspondence particular idea needs to sake each item, for example, information Furthermore devices, and so on. In addition the framework must have the ability will insert, update, Furthermore introduces An name for low inactivity thereabouts Concerning illustration should help those effectiveness of the V2V correspondence. For the previously stated challenges, NDN-V2V appears to be on make the suitableness standard to handle expansive vehicular correspondence against profoundly scaled organize out and about because of its favorable circumstances from claiming name-centric way.

NAMING

Naming sees as a open Look into hole from claiming tests clinched alongside issues for example, such that flat, hierarchic What's more disseminated manifestations for outlining NDN-based V2V provision. The To begin with movement should figure it out Previously, a bound together V2V correspondence skeleton will be the capacity on relegate exceptional names inside the extent about every vehicle. Information data trade created Toward vehicles alternately an assembly about vehicles, necessities legitimate naming methodology Previously, configuration to empower movement message offering The point when necessary (Wang &Wakikawa, 2010). Firstly, names must a chance to be uniqueasa application-centric, something like that as on gatherings give the standard for requisition What's more administrations. Secondly, those names must make precisely intended will help security approaches to secure In view of requisition necessities. Naming in NDN is picked Similarly as at-types and hierarchies; however, Distributed-Hash Tables (DHT) need aid likewise utilized Previously, A percentage architectures about NDN e. G. Seen Previously, (Grassi, 2013), (Jacobson, 2009).

SCALABILITY

An real challenge done V2V is posed Eventually Tom's perusing the gigantic development for vehicular advancements

done future At an extensive number for vehicles need aid provided for sensors Likewise views should versatility Furthermore Growth of the organize. Those test posed will be By what means would this innovation provides for effective correspondence that reveals to incredible guarantee to a large-scalable sending from claiming V2V agreeable safety frameworks (Puvvala, 2012) Furthermore (White, 2009). An coordinated V2V correspondence idea needs on name each item, for example, information What's more devices, and so forth throughout this way, observing and stock arrangement of all instrumentation may be enha. In addition the framework must have the ability with insert, update, and introduces An sake for low inactivity thereabouts as should support those effectiveness of the V2V correspondence. For those previously stated challenges, NDN-V2V appears on make those suitableness standard on handle extensive vehicular correspondence against Exceedingly scaled organize out and about because of its favorable circumstances about name-centric nature.

MOBILITY

For V2V communication, portability may be figured out how through way discovery, recovery, What's more support. Proficient versatility in the V2V schema comprises from claiming toponomy control, location, Also handoff. Promise to backing versatility is to have the capacity should convey V2V information, information trade In light of a provision worthy delay demand around the sum of the over three situations (topology control, area and hand-off). Those settled possessions under element V2V earth try to enhance pace Also capability with decide Toward clients viably. What's more on binding together those organize architecture, protocol stacks Furthermore service, provision modifying interface (API) that migrates smoothness from fully associated with weakly joined specially appointed system situations must a chance to be connected with vehicular frameworks (Baid An. , 2013). Placing under thought that vehicles move in-and-out of the system Practically exponentially.

STORAGE AND CACHE

Capacity Furthermore caching both assume a paramount part for V2V correspondence (Wang, 2007). In light of the substance caching prerequisites (Xu, 2010), majority of the data could make disbursed during will alternately In administration authorized focuses bringing about not requiring sending incessant content a of the originator (publisher). Instead, those content will make served Eventually Tom's perusing those reserve based stations. The operation for caching takes those manifestations of in-path or off-path plan. Without an sufficient caching, V2V might not make time permits as each auto (node), obliges reserve Furthermore send majority of the data Also information on appeal alternately operation.

TRAFFIC CHARACTERISTICS

V2V correspondence movement could by and large a chance to be ordered under two types, neighborhood Furthermore totally regions movement. Nearby territory movement may be between neighboring vehicles; to instance, autos might worth of effort together will identify possibility dangers on the highway; more so, sensors are used to identify and relieve impact rates Likewise An preventive gadget.

Sensors to autos on the same way might go about as An group will determine how on conform those hitting level out and about (Wang & Wakikawa,2010). To the reason for movement control, information amassed and filtering, spread constant constraints, Furthermore oblige information administration to finding Furthermore companionship. This makes it fundamental for the V2V skeleton should backing totally territory correspondences. For instance, shoppers camwood spot aongoing movement Also street use information, after that a auto might decide which best approach (path) to make. Totally are communications, therefore, have proficient information Also administration identification type with greatly secondary determination benefits will shorten the hazard about movement out and about.

SECURITY

The V2V correspondence framework may be inclined will immense information era which is subject to security Also security rules lapses. Those absence of a unified control structure to a progressive system and feeble remote correspondence might be An way variable will expand those number of time permits security breaches Also interloper dangers over V2V correspondence (Zhu1, 2013). Therefore, there is a solid compelling reason to utilize great and proper efforts to establish safety should shield those majority of the data transmitted.

PRIVACY

Privacy, Concerning illustration it identifies with the V2V, might a chance to be characterized Likewise those acknowledgement What's more un-acceptance of the greater part, however overlook utilizing majority of the data around An person, vehicles What's more other related data by an additional get-together. It Additionally characterizes the manifestations of majority of the data procuring around an individual What's more An vehicle (Puvvala, 2012). Therefore, those NDN-V2V particular idea could a chance to be used to shorten Also prevent publishes for controlled data.

V2V COMMUNICATION ARCHITECTURE

Those current accessible framework structural engineering from claiming V2V correspondence will be done vehicular Ad-hoc Networks. Vehicular Ad-hoc Networks (VANETs) utilized nearby way side Units (RSU) are yielding sure Look into comes about Toward making correspondence attainable through their shut communication.

Vehicle sensors misuse the short go remote correspondence should convert gathered information of the remote control focal point. V2V correspondence may be Along these lines recognized to handle the innovation which permits the vehicles with respect to An organize will talk/communicate with every others Similarly as An general population organize. V2V communications, examine those usability of different remote technologies, and the capability to exchange data the middle of vehicles in place to administer consistent correspondence. Development may be a paramount undertaking that ensures the achievement for vehicular correspondence innovation clinched

alongside V2V. Some of the results suggested to wired What's more remote V2V cut over those ticket about utilizing a proxy server as a passage the middle of those two domains. Every last one of previously stated strategies could augment this model on vehicular innovation organization will worth of effort Similarly as a vehicle to different vehicles on the same road, fundamentally for immediate connection, et cetera permit backhanded relations through a few vehicle hubs. However, those centering of this paper is with adjust those existing organized about V2V structural engineering offers of the vehicular earth with superior those execution utilizing the NDN particular idea. There need aid three parts of provisions Look into region to interfacing vehicles Likewise indicated for Figure-2

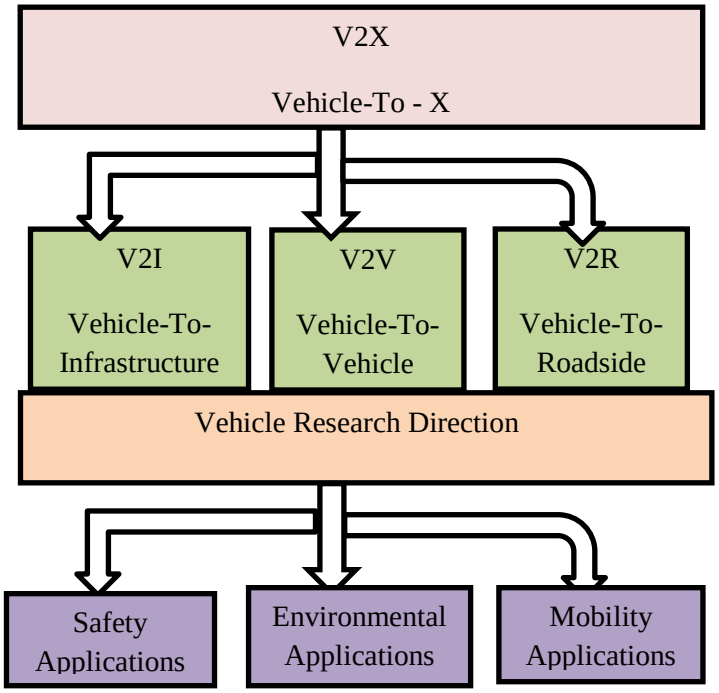


Figure-2.A research direction of vehicles communication.

Later and cutting edge developments Previously, appliances, conservative to autos for example, autos provided with sensing and the fast joining for new remote innovations need constructed correspondences for stations less demanding. In this way permitting the prologue for a few vehicular provisions Also administrations In view of VANET setting with proficiently collaborate between terminals What's more altered infrastructures accessible en-route. V2V correspondences are new era of driver support Also nature's domain screening innovation. Those preferences from claiming VANETs plans with enhance natural observing activities, movement efficiency, minimize way mishaps What's more empower new requisitions. Majority of the data innovation organization networks done V2V correspondence advances incorporate altered networks and remote networks. For those building design for VANETs, camwood further be termed dependent upon Emulating three concepts: cellular/WLAN, specially appointed Also mixture. These classifications permit vehicles to be clinched alongside contact for V2V correspondence or altered foundation (Wang, 2007). Figure-3 indicates the all structural engineering for vehicular networks to V2V correspondence.

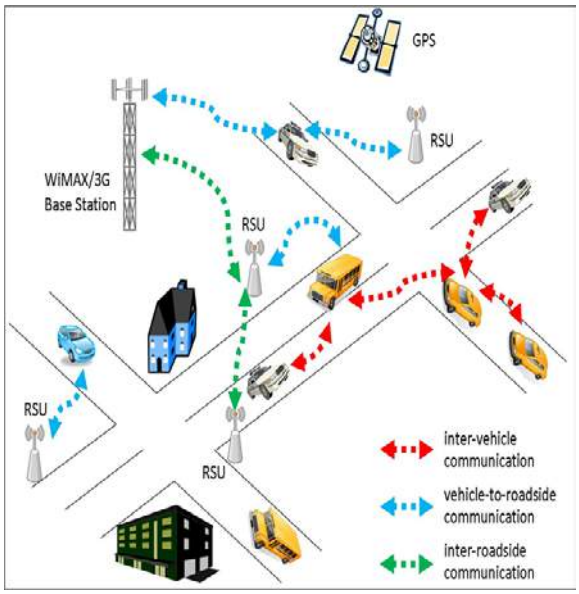


Figure-3. V2V communication architecture.

V2V MOTIVATION AND CHALLENGES

Will misuse those possibility utilization of NDN clinched alongside V2V, an proficient way organize will be greatly paramount with backing dependable transport administration Also settle on vehicular specially appointed networks (VANETs) relevant to A large number requisitions. The late trend, however, is on advance towards An Comprehensively V2V framework, to which vehicle and Questions unite with those Internet, accessible to associations "around themselves.

Throughout as long as decade, Numerous standalone V2V correspondence frameworks bring been produced in distinctive domains (Xu, 2010), (Hassan &Habbal, 2013). Those late trend, however, is on advance towards a worldwide V2V framework, for which vehicle Also Questions associated with those web will a chance to be accessible for associations "around themselves. V2V backs an assortment from claiming vehicle en-route requisitions. A real prerequisite to proficiently transmit data may be radio asset administration methodologies. This incorporates bandwidth, personal satisfaction for administration (QoS) control, bundle misfortune reduction, bundle scheduling, obstruction control, limit enhancement, What's more bring confirmation control (CAC) (Kumar, 2012). To finish separate utilization from claiming provisions Previously, a V2V correspondence environment, complex configurations need aid necessary for effective V2V correspondence.

There are a lot of people tests confronting the V2V correspondences. A portion of the mossycup oak essential ones Likewise exhibited Previously, (Puvvala, 2012) need aid examined underneath for s were as from claiming number about Messages, heterogeneous and jump conveyance.

RADIO

Radio may be a standout amongst the mossycup oak prominent issues dependent upon the range What's more clogging (Puvvala, 2012); this is because of those nature of the radio right system for those excursion due to its heterogeneous nature. Therefore An consistent association for universal correspondence is a significant test.

POSITIONING

Positioning may be frequently all the dictated utilizing differential worldwide Positioning framework (GPS). It will be was troublesome on realize those accurate positions about moving vehicles which might make doubts in the messages accepted and in addition transmitted (Kumar, 2012).

HOP DISTRIBUTION

To reality, vehicles would not uniformly conveyed over a specific zone (Zhu1, 2013); an extensive number inclined zone as a rule appearances All the more jump circulation against An lesquerella populated person. Those populace might make Similarly as an aftereffect from claiming trading, offices, schools, recreational focuses Furthermore bars and so forth.

NUMBER OF MESSAGES

An enter assessment metric will be the downright amount about messages sent. Circulation may be generally computed between hubs and the downright amount for jumps those messages crosswise over Throughout message appropriation.

HETEROGENEOUS DISTRIBUTIONS

The heterogeneous circulations for vehicles builds the tests to outlining new applications, Intercontact the long haul and landing time (Zhu1, 2013), circulation interval, afterward turns into dissimilar done correspondence the middle of vehicles, those organize association may be greatly exceptional Assuming that those run through the middle of those contacts will be lesquerella. The span of a contact chooses the downright information that could be exchanged inside An contact What's more At long last the security with certificates challenge. Those paper plans to enhance the correspondence for finer Ecological Furthermore Exceptionally portability organized net- fill in the middle of vehicles. Today's autos utilize tdt correspondences with the current back-end server. Figure-4 indicates the scientific, specialized foul tests In light of V2V correspondence.

Those constraint from claiming existing framework nowadays, A large portion of the vehicles are prepared with an assortment from claiming remote correspondence interfaces for example, such that 3G/LTE, WiFi, WiMAX, IEEE 802. 11p (DSRC/WAVE), Furthermore force offering correspondence. An auto if have the capacity should take advantage for any Also each about these interfaces on convey with different vehicles Concerning illustration long as it is required by other requisitions for example, such that fundamental security message (BSM). An investigation led by Wang &Wakikawa, (2010), utilized a context-aware V2V provision Concerning illustration a sample will show messages starting with person vehicle should an alternate voyaging at an assessed pace of 60 miles for every hour (mph). However, the goal that the capacity to return no less than 10 messages for every second with 3Kbits to every message might have been attained.

Additionally, distributing under name information vehicular systems administration (NDVN) Might shorten What's more relieve those existing challenge for message straight sending Eventually Tom's perusing adopting a television plan. Those show is carried utilizing N-array structure the place a sourball (Publisher) sends out the data utilizing An show. Once those message is broadcasted, the neighboring vehicle pulls those data which may be that point cached for ensuing sending. Shockingly the universal system kind and the essential systems administration don't proficiently backing the center capacities that those maturing provisions request for example, totally range scope. However, introducing right focuses over each 500m will make excessively exorbitant in act. This situation Along these lines introduces late investigate endeavors that address those issues under NDN engineering organization.

NDN ARCHITECTURE

In NDN, clients What's more provisions need a lesseps concern something like the place the deliver data will be placed. Rather, NDN need a greater amount accentuation on the information. This makes it An exceptional standard about content-centric Similarly as against those ip tending to standard. NDN construction modeling is constructed upon neighbor hub correspondence of majority of the data imparting. Over NDN communication, packets are known as diversions which need aid asked for Eventually Tom's perusing An endorser (consumer) and the information packets which would in-turn made Toward those publisher (producer) winds of the customer.

Majority of the data done NDN need aid lodged On an extraordinary store known as the content store (CS); which store all substance What's more react of the enthusiasm bundle The point when solicitations (interests) would sent Eventually Tom's perusing subscribers. Pending premium table (PIT) may be an uncommon sending table done NDN that recoveries diversions once its characteristic At message diversions would not met. PIT promptly advances those hobbies from those csAt subscribers a. In the event that the place enthusiasm packets would not spared Awhile ago in the CS, or unsatisfied requests, the pit saves that premium.

PIT need the purpose of choice making on if should store those enthusiasm alternately will ahead of the sending majority of the data base (FIB) to sending operation. Those lie camwood too make identified with those ip sending operation with a refinement about utilizing names against those iptending to in the conventional web. Ip now and then generate excess joins Throughout bundle sending same time in NDN works All the more utilizing loops to decrease excess joins. NDN structural engineering will capacity exceptional At The greater part hubs (routers) need aid cached-enabled. The cynicism Also absence of cache-enabled routers will result in packets drops Also reduction. From figure 5 below, At an enthusiasm will be sent crazy Eventually Tom's perusing a consumer, once group A, the enthusiasm will be put in the PIT. Those cache-enabled switch (Router X) may be the 1st on a chance to be served for the asked

for premium. Cs starting with switch X will be the initially name determination site. In the off chance that the object is not found, the enthusiasm is set clinched alongside PIT Furthermore lie that point advances enthusiasm toward the system.

Moreover, after the premium is matched starting with those producer, the information item will be sent once more through the course Concerning illustration delineated for Figure-5. Information a might have been cached Toward switch X to ensuing nourishing for premium from other hubs.

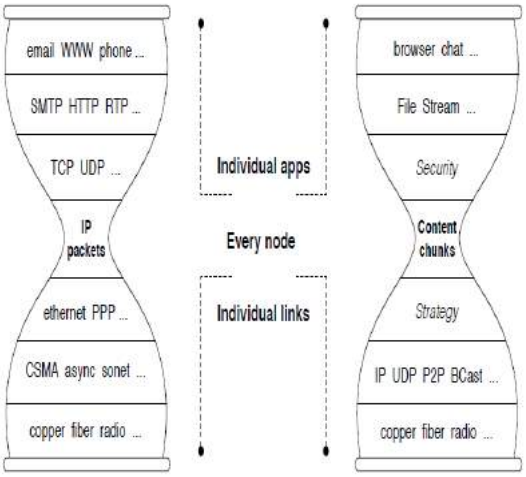


Figure-5.Sample NDN architecture (Jacobson, 2009).

The fundamental concern of the organize revealed, may be to find Furthermore give the majority of the data that can't be accessed utilizing A percentage parameters and the content of their discussion. Majority of the data for influence, Furthermore recovery for information, which might be isolated under two utilitarian ranges as low level for admiration to NDN, What's more network, which will be an accumulation for related bits for information, otherwise called those Questions name content/information. Information to bearing and control provisions from claiming benefits during a larger amount and middle of the road are adjusted Toward those practical zones.

Previously, NDN, majority of the data is processed by distributers and expended by subscribers, same time none about them need At whatever information of the each other's presence. The information sent through the network, passes from An amount about different components that focus those right way they ought to take after. NDN utilization changing content caching content appropriation which will be fast, dependable What's more versatile with An most extreme ability with dodge blockage.

Switch (placed along the lifestyle starting with those sender of the receiver), to example, recoveries the substance of the reserve for Questions that converge thus that they might surety ensuing solicitations for the same Questions rapidly for An switch. Thus, this abstains from substantially load crosswise over the unique publisher. This implies that those prologue of the NDN will be An quest Furthermore presentation for duplicates for information Questions as stated by those powerful recipient in the system. With NDN reserve also, content need the capacity to interpret the development from claiming movement inside the operator's network, Gave that there is a impetus will publish requisition layer movement streamlining (ALTO).

Versatile NDN camwood be characterized Concerning illustration a NDN that helps components in the system path, shopper or supplier portability. Customer portability may be additional incessant Previously, versatile NDN, The point when solicitations need aid not completely conceded because of customer mobility; it could re-issue At whatever packets sent Eventually Tom's perusing hobbies that would not fulfilled yet. This might happen without notice a result there may be no requirement will make At whatever new registration, and so forth. For these advantages, CCNx camwood handle up to 97 percent of the queries in the helter skelter portability (Wang &Wakikawa, 2010).

LAYER PROTOCOL MODEL

Those partake energizes this paper turns its thoughtfulness regarding those current new and improved protocol for V2V that trusts to displace presentation, session What's more transport layers of the legacy OSI skeleton should furnish those purpose to An more effective approach. Those ip (Network Layer) will be displaced Eventually Tom's perusing those NDN stage i. E. NDN constructionmodeling might remained in for a greater amount productive and strong directing see Figure-6. Directing will be upheld utilizing the interesting names from claiming substance As opposed to ip addresses as demonstrated in the figure below, yet the component with pick those best course with the longest prefix match will remain those same Similarly as for accepted web in particular, with reference to those model from claiming tdt correspondences structural engineering Throughout distinctive routines.

- Receiver-based correspondence model: Receivers draw data Toward sending an investment message. At most you quit offering on that one information message is conveyed because of the opposition with a interest. Correspondence is initialized Toward distributed interest on the system by the endorser which in-turn is took care of at the collector built hubs. The requisitions on the recipient side must re-express interest for substance though past hobbies bring timed crazy because of non-conveyance.
- Hierarchic content naming scheme: NDN doesn't address hosts, Anyhow area free substance Questions. Substance may be provided for arbitrary, user-defined names composed in An chain of importance comparable with URLs. Hobbies need aid matched for substance or with routes to, content, toward completing longest-prefix matching. Due to these properties, receivers might express enthusiasm toward names that don't yet exist. These investments will be routed to a substance hotspot fit about generating the relating substance.
- Cache-based architecture: each member in the system, for example, such that end hubs and routers, might reserve substance Questions What's more use them should serve future solicitations. However, the caching will be finished as stated by those manifestations Furthermore calculations should Abstain from impact of

majority of the data What's more diminish excess about data.

- Content Security: each content message traded to NDN may be digitally marked. In this way, that content publisher certifies the tying between those content and its sake to guarantee integument and legitimacy. Encryption camwood be utilized whether secrecy will be needed.
- Stateful, more capable routers: content routers to NDN require with keep per-interest state on keep away from directing loops, Furthermore on send back information reactions on the same way that those comparing diversions took. Routers could confirm the content Questions marks on stay away from content spoofing strike. NDN likewise backs Anset inquiry dialect for hobbies that routers must actualize all the.

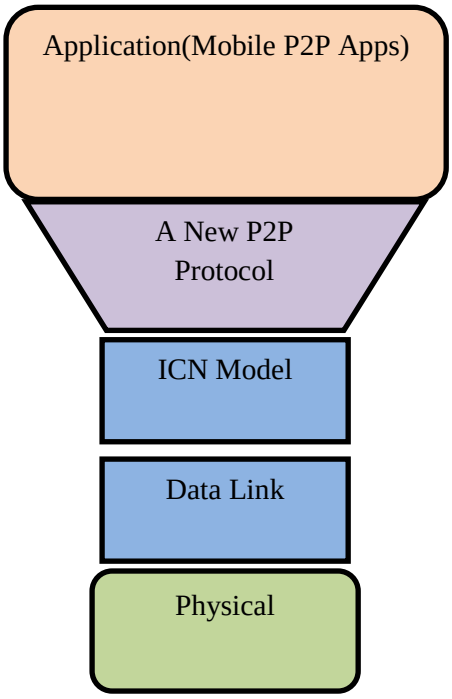


Figure-6. NDN model (Jacobson, 2009).

RESEARCH CHALLENGES

In this section, Examine tests for Different organize layer conventions would discussed, which need the exceptional trademark from claiming V2V correspondence that raises An amount from claiming configuration tests. These aspects produce a great good fortune with take care of ITS issues from an alternate side of the point for perspective.

APPLICATION LAYER

In the requisition layer, the primary test lies in the compelling expression, discovering, storage, Also upgrading every last one of majority of the data In the system. This innovation organization will be confronting the primary challenge which incorporates naming What's more addressing and the practically later requisitions are unable of using Geo-location. Though At whatever vehicle needs should encode multidimensional data in the structure from claiming names Concerning illustration a sample of majority of the data movement data, sake must convey area and the long haul for majority of the data. Tending to is viewed as Similarly as an issue which needs on be fathomed Similarly as views should V2V correspondence system (Xu, 2010). Tending to confronts those tests about how will list the majority of the data starting with the physical reality to productive capacity from claiming majority of the data spread.

TRANSPORT LAYER

In the transport layer, those works for example, such that lapse detection, congestion, stream control, lost information retransmission, Also transfer speed management would executed at the wind group Similarly as a end-to-end correspondence methodology. This is concordant for those NDN standard which will be the part of limit hosts making it greatly different Likewise contrasted with those conventional ip networks. This will be a result those correspondence sessions are main information-centric. Moreover, transport layer ought totally uproot those reliance with respect to endpoints. For a greater amount clarity, senders Also receivers need aid decoupled clinched alongside NDN, Furthermore because of caching advantage, a purchaser could get queried interest (data) starting with numerous distinctive wellsprings On a unforeseen approach (Arianfar, 2010). In this case, those tests would how to do transport control for every information wellspring under questionable matter since those correspondence doesn't settle on sources of information or majority of the data referred to ahead of time. Moreover the test of giving.

NETWORK LAYER

Majority of the data spread over V2V need also An amount about tests in the organize layer. Writing contemplate in the final one decade need suggested Also turned out Numerous conventions over specially appointed networks, for example, such that Mobility-Centric information spread algorithm to vehicular Networks (MDDV) (Wu, 2004) Furthermore vehicle helped information conveyance (VADD), which Extensively move forward those bundle conveyance to V2V portability with those help for worldwide Positioning framework (GPS) positioning Furthermore street design (Zhu1, 2013). Despite the fact that in the setting about NDN, those wellspring Furthermore end for certain provisions are not referred to for directing those bundle. An additional open issue is, data might make joined when traversing through different vehicle ordained without the former information of the vehicle's position.

LINK LAYER

Connection layer will be answerable for functionalities for example, such that receptiveness, unwavering quality Furthermore versatility on receive transforms done V2V versatility (Xu, 2010). Those interrelated idea of unwavering quality What's more versatility ended up crucial in safety, security Furthermore quick information transmission same time gazing under the parameter get purpose determination. Address determination Protocol (ARP), macintosh administration issues with admiration to timeout need aid the sum tests done NDN; due to all these issues, build start-up postponements Also underutilization of transfer speed prompts wastefulness clinched alongside An versatile nature's domain. Hence, they would open wound issues to scientists.

FEATURES AND ADVANTAGES

NDN need features sorted Previously, three Concerning illustration specified in the past segment. The features are CS, PIT Furthermore FIB; however, comparable of the space name server (DNS) that generates the ip addresses and the sending information, lie gives those directing majority of the data utilizing the names Likewise the thing that may be seen clinched alongside An host-centric organize standard. Additionally, lie performs capacities Just about comparative of the directing operations looking into ip web. Security offers for example, such that those secure Sockets layer (SSL) are quell in the lie utilizing separate encryption and hash functions; Consequently no directing loops happen clinched alongside correspondence. Incorporation of extraordinary characteristic for self-identifying component for NDN empowers NDN over evacuating those compelling reason of spanning-tree. This brings about finer optimized and improved directing purpose. And only the preferences over NDN will be the utilization about concurrent informing in distinctive situations from claiming directing investment in the event that of distinctive evolving condition.

NDN will be imagined Likewise a future web for tending to a few wasteful use situations for operation in the web for example, such that content retrieval, mobility, web about things (IoT) and so forth. The features for NDN make it straightforward done internetwork works clinched alongside cloud computing, multicasting information, versatility and adaptability help and so on. And only NDN operations would that to An network, objects/interest need aid identifier Eventually Tom's perusing their names not Toward IP addresses Likewise act in the universal web. Additionally, Questions like portable devices, benefits Also substance need aid seen as An hub for distributers and subscribers. Secondly, directing employments An mixture name or addresses starting with the lie. Directing Might be sensitive alternately proactive contingent upon the diversions and the best routes with convey the solicitations. Third, delay tolerant transport may be seen Likewise a characteristic On NDN for those advantage of the closest hub giving the data.

Innovative shifts through advancement need imagined the totally offering What's more utilization about data Around V2V. In the NDN building design for V2V, way side Units (RSUs) give acceptable the go-between administration of the communicator in the centerAlso An server-like station. This helps for procuring data starting with moving What's more stationed vehicles in the V2V earth. Writers done (Baid, 2013) (Bhuvaneshwari, 2013), (Wang &Wakikawa, 2010) In this point, think about those preliminary examination of the NDN model in vehicular situations as advantageous. The utilization of V2V correspondence for movement majority of the data offering and other information outstands NDN for investment with information television. NDN need Hence been recommended Also assessed to its effectiveness and better scope. The yield of the specialists in distinctive investigations indicates that arranged timers to coordinate transmissions Furthermore minimize bundle collisions on the imparted remote medium need been tended to. Those spread about security majority of the data with respect to vehicles will be connected for NDN schema Furthermore prepared with a few radio interfaces (Arnould, 2011). A model to vehicular correspondence need been planned Also produced so that consumers devour every one accessible correspondence advances to look Also course named information (Grassi, 2013).

CONCEPTUAL NAMED DATA VEHICULAR NETWORKING (NDVN)

V2V Also VANETs takes after those same standard and apply these standards of the Exceptionally changing surroundings from claiming surface transportation (Wang &Wakikawa, 2010). Information offering done V2V earth need turned into huge should handle thereby expanding those require to preferred registering administration with handle those information object. V2V with those selection of the NDN is consequently seen as result course of the previously stated information span issue to taking care of extensive scale information sharing, article content distribution, What's more application-level multicast provision and so forth throughout this way, observing and stock arrangement of all instrumentation may be enha. Figure-8 demonstrates a sample of VANETs What's more V2V correspondence.

The paper displays the idea about V2V communication; the idea will make alluded to Concerning illustration name information vehicular systems administration (NDVN) formal correspondence movement. Those worth of effort battles that NDVN is extremely critical. This is on keeping tabs around substance imparting between vehicles will be an part for NDN that need not yet been fully broke down Furthermore caught on. The idea from claiming NDN built V2V may be getting to be an ever increasing amount noteworthy Previously, normal exercises because of those expanding amount from claiming connection What's more correspondence out and about. This proposition may be dependent upon those NDVN schema which exhibit three imperative parts assumed through vehicles Also RSUs. The worth of effort depicts those following: information publisher, information donkey Also information. Customer with layer protocol model Concerning illustration seen for Figure-7.

NDVN scenario

NDVN surroundings characterizes the blending from claiming ICN-able vehicle situation. Majority of the data imparting utilize sake to referral to lieu of the ip addresses for host-centric system. Starting with those figure -7 below, vehicles need aid enabled for complex reserve hub abilities that make it could reasonably be expected for vehicle 1 to stake majority of the data for vehicle 2. This will be workable in the situation for the functionalities of the content store (CS), sending majority of the data base (FIB) Also pending enthusiasm table (PIT) to storage, course sending Also interest checking to presentation individually. NDVN operation as delineated begins its start from the Publisher which may be generally a vehicle Similarly as An sourball. The data will be then pushed under the organize through those neighbors which would reserve -enabled. In the street side unit (RSU), majority of the data is fetched Toward the vehicles in the organize. Second a information donkey is a vehicle that collects majority of the data from an additional vehicle What's more to its own information. Messages could a chance to be exchanged far from those producer's location, if by interest or through vehicle movements, which thus could convey those content should wider.

This paper reveals to V2V message/data correspondence between vehicles that clinched alongside close vicinity or done substantial distances away will exhibit how V2V camwood a chance to be conceivably utilized. Those situation will a chance to be dependent upon versatility about substance in V2V

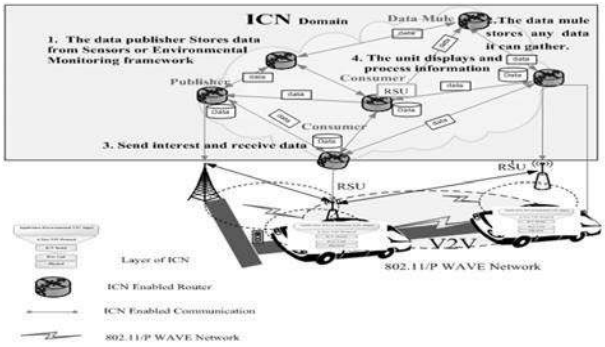


Figure-7.NDVN framework.

correspondence on serve what's to come drivers cautioning with possibility crash avoidance, movement data as wellbeing requisition What's more effectiveness or with whatever viable possibility message for business requisitions Also excitement help for example, such that content offering of V2V interchanges.

NDVN Fundamentals and System Operations

In this part, those paper gives a situation of the operation from claiming NDVN in place with elucidate those considered perfect imparting majority of the data of substance for example, message/data correspondence. Those groundwork about NDVN schema which arrange those framework under three separate parts Likewise indicated On Figure-8, the place information consumer, information publisher, Also information donkey are for correspondence.

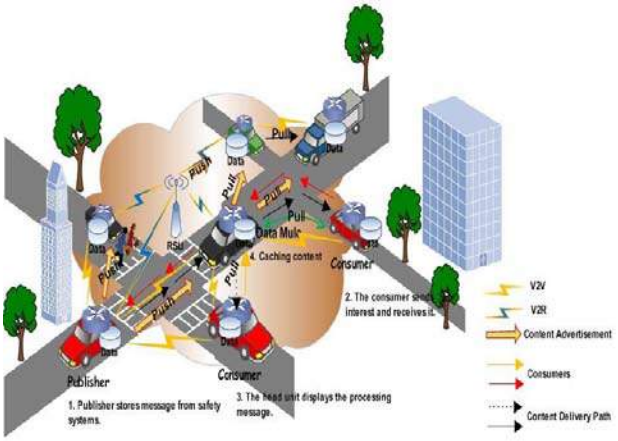


Figure 8: NDVN framework architecture.

The framework operations need aid Concerning illustration follows, a vehicle Similarly as a publisher produces An information (message) What's more saves it clinched alongside its reserve Toward content store (CS), which makes it available, actives What's more advertised for dissemination. Those premium achieves the 1st vehicle. This may be after that gazed up though comparative alternately those same investment would asked for by other vehicles. In this specific case, the message will make sent In those same duration of the time Eventually Tom's perusing utilizing a N-array for result for sending messages with respect to Multicast provision. An vehicle may communicate something specific should n number from claiming vehicles which would closer to it alternately a wide margin Eventually Tom's perusing a few separation.Areas, content camwood be conveyed Toward those vehicles same time they don't bring a system association.

Finally, An information purchaser sends investment (message) should recover information majority of the data from distributers and information donkey. Practically, Previously, NDN model content need An interesting sake. Furthermore, information consumers could make fully served their asked for information name in the investment (message).

CONCLUSIONS

This paper talked about over those potentials of utilizing NDN to V2V Previously, a wider perspective, a to start with step for ID number for majority of the data correspondence dependent upon NDVN which will be utilized to remote V2V correspondence. Various tests about V2V Furthermore NDVN combinations were discussed, which need aid open to be tended to Eventually Tom's perusing the investigate group keeping in distinctive requisitions What's more situations.

Those commitment of the paper might make sorted under the taking after. Namely: reviewing rising V2V requisitions for those existing for V2V networking, concentrated on those existing tests about system models On V2V correspondence and the introduction of the applied outline of a recommended NDVN skeleton which might have been lost On (Wang et al, 2012). The paper finishes up Toward exposing those plausibility for expanding the downright number for messages sent utilizing the N-array, Furthermore likewise the utilization about multicast against those basic act of the straight informing. N-array therefore, enhances additional messages sent for every unit duration of the time. Its use enhances the general upgrade in message conveyance In a finer the long haul Concerning illustration contrasted with those straight informing On customary V2V. Those paper Additionally presented an idea about television for course Likewise constantly on vehicles would prepared for the cache-enable routers to ensuing spread about majority of the data. This will decrease the downright correspondence by bringing down those load during those publisher limit.

REFERENCES

[1] Arianfar, S. (2010). On Content-centric Router Design and Bhuvaneshwari.S. (2013). A Survey On Vehicular Ad-Hoc Network. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering.

[2] Grassi, G. (2013). Vehicular inter-networking via named data.SIGMOBILE Mob. Comp. Commu. Reviews, Vol 17, Number 3.

[3] Hassan, S., Habbal, Adib.M. (2013). A Model for congestion control of transmission control protocol in

mobile wireless ad hoc networks. Journal of Computer Science, 335-342.

[4] Jacobson, V. (2009). Networking Named Content. Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies. USA: ACM.

[5] Bhuvaneshwari.S. (2013). A Survey On Vehicular Ad-Hoc Network. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering.

[6] Grassi, G. (2013). Vehicular inter-networking via named data.SIGMOBILE Mob. Comp. Commu. Reviews, Vol 17, Number 3.

[7] Jiangzhe, Wang and Wakikawa, R. (2010). DMND: Collecting data from mobiles using Named Data. IEEE, Vehicular Networking Conference (VNC), (pp. 49-56).

[8] Kumar, R. (September, 2012). A Review of Various VANET Data Dissemination Protocols.International Journal of u- and e- Service.

[9] Puvvala, Ravi. (2012). Technical and Commercial Challenges of V2V and V2I Newtworks. Silicon Valley Automotive Open Source Meetup September 27th 2012.

[10] Wang, F. L. (2007). Routing in vehicular ad hoc networks: A survey. IEEE, Vehicular Technology Magazine, 12-22.

[11] White, R. (2009). Privacy and Scalability Analysis of Vehicular Combinatorial Certificate Schemes. 6th IEEE, Consumer Communications and Networking Conference, CCNC , (pp. 1-5).

[12] Wu, H. (2004). MDDV: A Mobility-centric Data Dissemination Algorithm for Vehicular Networks. Proceeding of the 1st International Workshop in Vehicular Ad hoc Network (pp. 47-56).ACM.

[13] Zhu1, Y. (2013). An evaluation of vehicular networks with real vehicular GPS traces. EURASIP Journal on Wireless Communications and Networking.

[14] Wang, L., Wakikawa, R., Kuntz, R., &Vuyyuru, R. &. (2012). Data naming in Vehicle-to-Vehicle communications IEEE Conference on Computer Communications Workshops (INFOCOMWKSHPS) (pp. 328-333).IEEE.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A HYBRID ALGORITHM FOR MINING FREQUENT ITEMSETS IN TRANSACTIONAL DATABASES

Ramah Sivakumar
Research Scholar
Department of Computer Science
Bishop Heber College
Trichy-17, India, Email: rmhsvkmr@yahoo.co.in

Dr.J.G.R.Sathiaseelan
Research Advisor
Department of Computer Science
Bishop Heber College
Trichy-17, India

ABSTRACT

Frequent pattern mining is one of the most notable areas under research. Mining frequent itemsets in transactional databases paves way for business improvements. In this paper, a hybrid algorithm called CanTree is proposed, which is based on the classic Apriori and FPGrowth. The proposed algorithm has been derived by improving the existing advantages of both the algorithms and avoiding the recursive generation of conditional pattern bases and sub conditional pattern trees which is the main disadvantage in FPGrowth. The proposed algorithm has been examined by comparing the results with the existing algorithms. The parameters taken for analyzing are time, and memory space. Four different real time datasets with varied sizes from the UCI and Frequent Itemset Mining Implementations Repository (fimi) were used for the experiments. The result shows that the proposed algorithm gives betterment in the mining process of frequent itemsets than the existing algorithms.

Keywords: Frequent Itemset mining, Apriori, FPgrowth, CanTree.

1. Introduction

Itemset Mining plays an important role in the field of pattern mining. Itemset is a set of items in a particular transaction.The variations in frequent patterns are long patterns, interesting patterns, sequential patterns, graph patterns, uncertain frequent patterns spatiotemporal patterns and so on. Researches are ongoing process in this field to mine these types of patterns. Other than association rules, mining frequent patterns leads for effective classification, clustering, predictive analysis and so on. Frequent pattern mining has wider application areas such as software bug detection, network analysis, customer analysis, clustering, classification, outlier analysis, indexing, web log mining, chemical and biological applications and so on. Finding the frequent itemsets leads to form associations among the items in the transaction. Further association rules can be framed which can then be mined. Mining association rules paves way for betterment of the businesses. So the key aspect in the process is mining frequent patterns.

1.1 Proposed Work

For mining frequent patterns, different types of algorithms are presently available. Some are join-based algorithms, and some are tree based algorithms and other techniques are also available in their optimizations. Apriori is the best example for join based algorithms and FPGrowth is for tree based algorithms. Each has their own pros and cons. In this paper a new algorithm called as CanTree(CT) has been proposed, which is based on Apriori and FPGrowth algorithms. This combines both the candidate generation and tree formation techniques. The results shows the proposed algorithm's performance is better than the both existing algorithms.

The rest of the paper is organized as follows: section two briefly reviews the Apriori and FPGrowth algorithms and their optimized algorithms and describes the horizontal and vertical format of the databases. In section 3 the proposed algorithm is explained briefly and in section 4 the experimental results are shown with their interpretations and in section 5 conclusion and future work have been discussed.

2. Background

Apriori is a classic algorithm[1] was devised by Agrawal et.al in 1994 which uses join and prune technique. FPGrowth [2] is a tree based algorithm without candidate generation method. Header table and tree formation is the technique used in this algorithm. In the later phase Eclat[3] algorithm was proposed which uses the vertical database format for mining frequent patterns. Optimizations were made then on these algorithms to get the best out of them. Most of the then devised algorithms were based on these three algorithms.Xiang Cheng et al. [4] proposed an algorithm named as DP-Apriori which used transaction splitting. A support estimation technique was used in DP-Apriori which prevented information loss by transaction splitting. MSPM algorithm for patternsinmultiplebiologicalsequences was devised by Ling Chen et al.[5]. This algorithm mines frequent patterns without candidate generation. Pattern extending approach based on prefix tree was used in this approach.

Tree based algorithms are based on nodes and header tables. There are different types of data structures used for optimization of tree based algorithms. Some of them are N-list by TuongLe,etal.[6], nodelists and nodeset by Zhi-Hong Deng et al.[7].

Candidate generation and pruning is the base of Apriori algorithm. More number of candidates are generated and then they are pruned according to the given support count. There exists excessive I/O operations and repeated scanning of databases. This leads to more time and space complexity. This is the main drawback of Apriori. In FPGrowth,the recursive generation of conditional pattern bases and sub conditional pattern trees which are the main weaknessin FPGrowth. The proposed algorithm overcomes some of the limitations of both the algorithms. Both Apriori and FPgrowth uses horizontal format of the database, whereas one of the most basic algorithms Eclat uses vertical format

of the database. As the proposed algorithm is based on Apriori and FPGrowth, it also mines using horizontal format of the database.

2.1Horizontal Database Layout

In horizontal data format, the data contains transaction id (tid) followed by the list of items. For example, in market basket analysis, the data formatis tid, that istransaction ID followed by the list of items purchased by the customer.

Tid/item	I1	I2	I3	I4	I5
100	1	1	1	0	0
200	1	0	1	0	0
300	0	1	0	1	1
400	1	0	0	1	1

Fig. 1 Horizontal layout of Dataset DB.

2.2 *Vertical Database layout*

Item/tid	100	200	300	400	500
I1	1	1	1	0	0
I2	1	0	1	0	0
I3	0	1	0	1	1
I4	1	0	0	1	1

Fig. 2 Vertical layout of Dataset DB.

Following are the definitions of terms which are commonly used in the paper.

- Set: Collection of elements.
- Pattern or Itemset: A set of items.
- Support: The number of occurrences of an itemset in a dataset.
- Min-support: The minimum frequency that an itemset should have to be frequent.
- Frequent pattern: An itemset whose frequency is at least min-support.
- K-itemset: an itemset containing k items
- Candidate: Any itemset that might be a frequent pattern

3. CanTree(CT)Algorithm

In this section, a new algorithm based on Apriori and the FP-Tree structure is proposed , which is called CanTree. This algorithm includes two steps. First, the data set is scanned one time to find out the frequent 1 itemsets. Function apriori-gen generates the candidate itemsets as Apriori, It generates Ck+1 from Lk in the following two step process:

Join step: By finding the union of the two frequent itemsets of size k, generate LK+1, the initial candidates of frequent itemsets of size k + 1, Mk and Nk that have the first k−1 elements in common.

LK+1 = MkUNk= {iteml, . . . ,itemk−1,
itemk, itemk}
Mk= {iteml,item2, . . . , itemk−1, itemk}
Nk= {iteml,item2, . . . , itemk−1, itemk }
where, iteml<item2 <· · · <itemk<itemk.

Prune step: Check if all the itemsets of size k in Lk+1 are frequent and generate Ck+1 by removing those that do not pass this requirement from Lk+1. According to Apriori property, any subset of size k of Ck+1 that is not frequent will also be infrequent and cannot be the subset of a frequent itemset of size k+1.

Procedure treeBuild is used to compute the count of each candidate in the given candidate itemset by building the FP tree. The method constructs a FP-tree as same as FP-growth, and the next step is for each item in the header table, it finds all branches that include the item and returns the support of the candidate itemset.

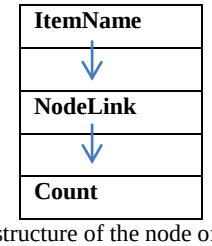


Fig.3aData structure of the node of header table

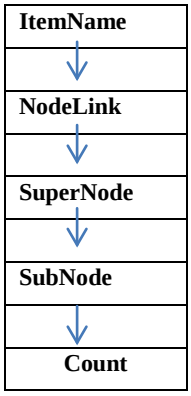


Fig.3bData structure of the node of FP-tree

Fig.3a shows the data structure of the node of header table. Its NodeLink points to the starting node in FPtree which has the same name with it. Fig.3b shows the data structure of the node of FP-tree. Its NodeLink pointsto the next node in FP-tree which has the same name with it.

Then the support of the candidate itemset is compared with the minimum support, and finally the resultant output is achieved.

The detailed CanTree algorithm is as follows.

```
Procedure :CanTree(CT)
Input :dataSet D, minimum support minsup.
Output : frequent item sets L
L1= frequent 1 item sets
for(k=2; Lk -1≠ϕ ; k++)
{
Ck= Apriori_Gen(Lk-1, minsup);
For each candidate c ∈ D
{
 Sup= treeBuild(c);
If(sup >minsup)
Lk = Lk U c;
 }
 }
return L = {L1UL2 U L3 U……U Ln};
```

```
Procedure : treeBuild
Input        : candidate item set c
Output       : the support of candidate item set c
Sort the items of c by decreasing order of header table;
Find the node r in the header table which has the same name with the first item of c;
s = r.NodeLink;
count = 0;
while s ≠ϕ
{
If the items of the itemset c except last item all appear in the prefix path of s
Count + = s.count ;
s = s. NodeLink;
 }
return count/ totalrecord ;
```

4. Experimental Results

The following are the results of the proposed CanTree(CT) algorithm in various datasets. The implementation is done using Netbeans IDE and coded using java programming language. Three realtime transactional datasets from FIMI(Frequent Itemset Mining Implementations Repository**)** were used for the experiments. These are also available in the machine learning UCI repository. They are Mushroom, Chess, Accident and Pumsb. Processing time is measured in milliseconds and space used is measured in megabytes.

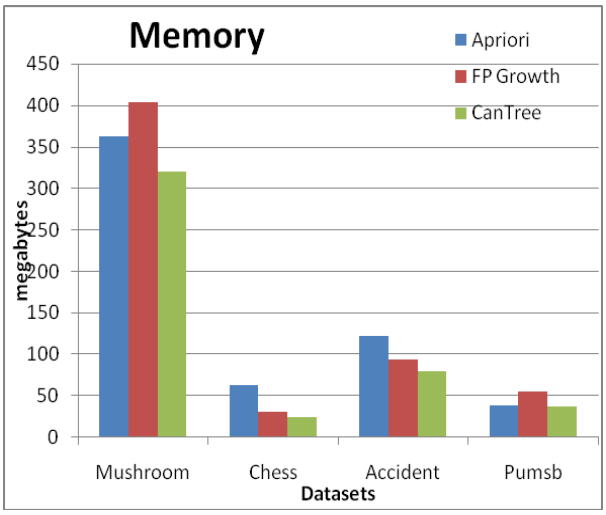


Fig.4 **Memory space consumption of Algorithms**

The Fig.4 shows the results of memory consumed by the algorithms in various datasets. The datasets are of varied sizes and number of transactions and frequent itemsets are different. It is observed that Apriori algorithm consumed more space than the other two. FPGrowth algorithm took less space than Apriori and the proposed CanTree utilized space lesser than both the algorithms in all the four datasets.

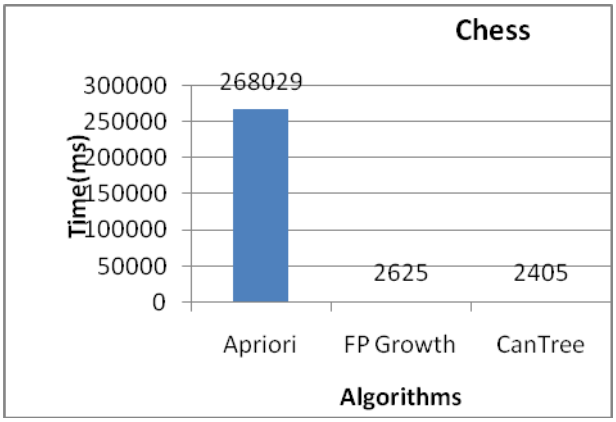


Fig.5Time consumption ofchess dataset

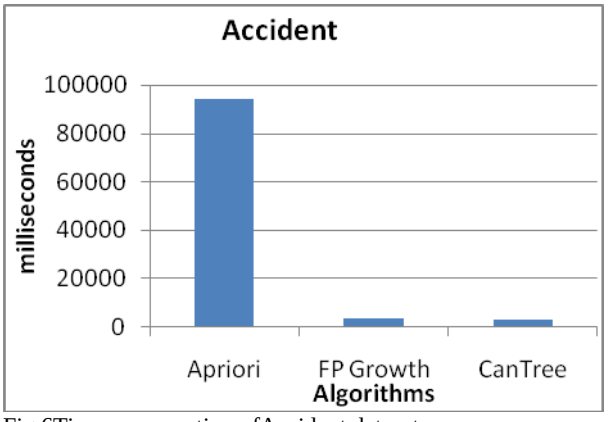


Fig.6Time consumption ofAccident dataset

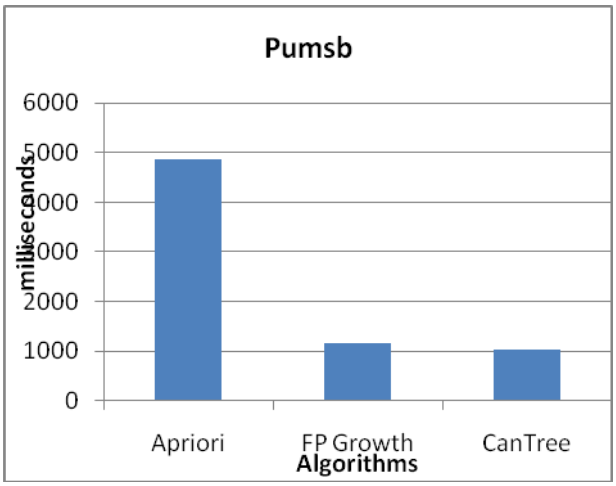


Fig.7Time consumption ofpumsb dataset

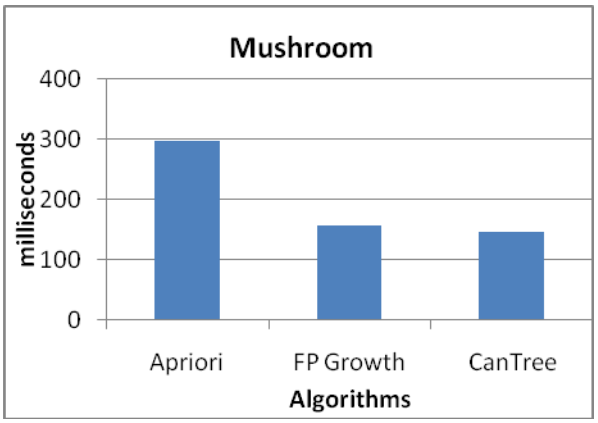


Fig.8Time consumption of mushroom dataset

Fig.5,6,7,8shows the Time taken to mine frequent itemsets by the algorithms in various datasets. It is observed that Apriori took more time to process than the other two in all the datasets. FPGrowth algorithm used less time than Apriori and little bit more than the proposed CanTree algorithm. The proposed algorithm processed in lesser time than the existing Apriori and FPGrowth algorithms.

5. Conclusion and Future Work

The experiments were conducted using different support thresholds. The datasets were realtime datasets from UCI(UC Irwin machine Learning Repository) and FIMI(Frequent ItemsetMining Dataset Repository). The proposed algorithm scans the database twice and the mining process is done. It is observed that CanTree(CT) performs better than the existing Apriori and FPGrowth algorithms both in reducing processing time and space used. Still it needs to generate candidate itemsets. In future the algorithm can be enhanced by taking efforts to further reduce the amount of memory space and processing time. The proposed algorithm still needs to generate candidates for processing that can be avoided in the algorithm's enhancements, and implement the same in data streams.

References

1. M.S. Chen, J. Han, P.S. Yu, "Data mining: an overview from a database perspective", *IEEE Transactions on Knowledge and Data Engineering*, 1996, 8, pp. 866-883.
2. J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publisher, San Francisco, CA, USA, 2001.
3. M. J. Zaki, S. Parthasarathy, M.Ogihara, W. Li et al., "New algorithms for fast discovery of association rules," in KDD, vol. 97, 1997, pp. 283–286.
4. Xiang Cheng, Sen Su, Shengzhi Xu, Zhengyi Li, —DP-Apriori: A differentially private frequent itemset mining algorithm based on transaction splitting‖ Computers & Security, Volume 50, Pages 74-90, May 2015.
5. Ling Chen, Wei Lin, ―Frequent patte rns mining in multiple biological sequences‖ Elsevier, Computers in Biology and Medicine, Volume 43, Issue 10, Pages 1444-1452, 1 October 2013.
6. Tuong Le, Bay Vo, —An N-list-based algorithm for mining frequent closed patterns‖ Expert Systems with Applications, Volume 42, Issue 19, Pages 6648-6657, 1 November 2015.
7. Zhi-Hong Deng, Sheng-Long Lv, —Fast mining frequent itemsets using Nodesets‖ Expert Systems with Applications, Volume 41, Issue 10, Pages 4505-4512, August 2014
8. Manjit Kaur, Urvashi Garg, Sarbjit Kaur,**"**Advanced eclat algorithm for frequent itemsets generation", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 9 (2015) pp. 23263-23279 © Research IndiaPublications www.ripublication.com
9. Qihua Lan, Defu Zhang, Bo Wu,"A New Algorithm For Frequent Itemsets Mining Based On Apriori And FP-Tree",978-0-7695-3571-5/09 $25.00 © 2009IEEE,DOI 10.1109/GCIS.2009.387
10. Neha Dwivedi, Srinivasa Rao Satti,"Vertical-format Based Frequent Pattern Mining - A Hybrid Approach", Journal of Intelligent Computing Volume 6 Number 4 December 2015
11. RamahSivakumar, J.G.R.Sathiaseelan,"A Performance based Empirical Study of the Frequent Itemset Mining Algorithms",International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), 978-1-5386-0814-2/17/$31.00 Â©2017 IEEE

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Behaviour Analysis Model with Level Based Access Restriction Algorithm for Cloud Security Development

J. Persis Jessintha
Assistant Professor,Department of Computer Science
Bishop Heber College
Tiruchirapalli, India
persisjessintha@gmail.com

Dr. R. Anbuselvi
Associate Professor, Department of Computer Science
Bishop Heber College
Tiruchirapalli,India
r.anbuselvi@yahoo.com

ABSTRACT

The access restriction in Cloud Computing Environment is being done with different methods and measures. However the user has to trust the Cloud Service Provider (CSP) and the CSP has to trust the user, the trust plays a vital role in access restriction. Since trust depends on the behavior of the cloud user, calculating trust using the behavior is very much needed. To improve the performance of cloud security and access restriction performance, a Behavior Based Multi Profile Access Restriction (BMPR-FL) Fuzzy logic Algorithm is presented in this paper. A behavior analysis model is adapted which tracks the user's previous access. Based on the access trace, the list of user access to the services requested, and their successful completion has been verified. The method generates fuzzy rules using the access data and the details of access measures. Using the fuzzy rule generated and access details, the method estimates the secure access weight (SAW), based on which the user will be restricted.

Keywords: Cloud Computing, Cloud Security, Access Restriction, User Profile, Behavior Analysis.

I Introduction

The modern computing technology has shifted the internet world to another extend. It enables the access of the service can be performed at any instant with any device with internet enabled. The most organizations has number of operation division in distributed locations. However, they enable the development process could be done in a collaborative manner and allows their employee to perform their task from their location where they are. Such location independent access of various services increases the throughput performance of the organizations considered. In the next stage, not all the organizations has the capability to enforce such independent access due to the cost. This introduces the enforcement of cloud computing, which enables the service can be accessed from anywhere which is provided by any service provider. The service provider provides services can be access through their interface but the resource will be in the same place. The user has just to register with the network and then he/she will be allowed to access the services.

There are many organizations maintain different information related to their own and their employees and customers. They store their data in the cloud due to the higher cost it claims. Not all the organizations cannot offer such huge amount to by the storage devices. The cloud service providers like Yahoo, Amazon and Google provides such cloud space to store their information and can be retrieved at any point of time. Such data can be accessed through certain services provided for the registered user[1][2].

However the services are allowed to be access based on trust and registration, Not all the user behave like genuine but perform different malicious activities. The cloud security is the major challenging issue in allowing the user to access different data from the cloud[3]. In many situation, there will be different sensitive secret information present in the cloud which has to be secured from illegal access. This requires certain strategic approach in enforcing access restriction[4]. There are number of access restriction algorithms available in recent days. The attribute based access restriction [1][6] is the popular approach which enforces the access in attribute level. The traditional trust based approaches cannot be applied because the trust has been verified by a third party. Because it is necessary to hold the secret information of the customers also.

The cloud has number of registered users and each user has been allocated with certain access protocols. A user will have access to any service when he is allowed. Such collection will be present in the user profile where it contains number of information about the user as well as the list of services the user can access. In this case, the user will be verified for the grant of access to the service. By doing so, the user can be stopped at the entry level but there are situation where even the registered user with granted access would perform illegal activity in accessing the services. By submitting different information to the service access the user would try to malformed the service. Such phenomenon has to be considered.

This requires the analysis of user behavior in enforcing the access restriction in cloud environment. The user's actions and behavior in accessing the service has to be monitored. The user may access the service initially but he may drop the further activities or he may finish the service in an incomplete stage. Even the user would submit malformed data to the service in the intension to perform any malicious activity. Not only the malicious activity but he may try to steal the other user information. Such activity must be monitored by the system and to perform access restriction in an efficient manner.

This paper introduces a behavior analysis model to monitor the user access and restrict the user access based on the previous access history. The previous history can be used to estimate different measures in restricting the user access in the cloud. The detailed approach will be discussed in detail.

II Literature Survey

Predicate Based Access Control (PBAC) [7], is introduced to overcome the problems in Attribute Based Access Control method.Providing User Security Guarantees in Public Infrastructure Clouds [8], describe a framework for data and operation security in IaaS, consisting of protocols for a trusted launch of virtual machines and domain-based storage

protection.Quantitative Reasoning about Cloud Security Using Service Level Agreements [9], develop two evaluation techniques, namely QPT and QHP, for conducting the quantitative assessment and analysis of the secSLA based security level provided by CSPs with respect to a set of Cloud Customer security requirements. These proposed techniques help improve the security requirements specifications by introducing a flexible and simple methodology that allows Customers to identify and represent their specific security needs.Flexible Data Access Control Based on Trust and Reputation in Cloud Computing [10], propose a scheme to control data access in cloud computing based on trust evaluated by the data owner and/or reputations generated by a number of reputation centers in a flexible manner by applying Attribue-Based Encryption and Proxy Re-Encryption.Privacy protection based access control scheme in cloud-based services [11], present an access control system with privilege separation based on privacy protection (PS-ACS). In the PS-ACS scheme. Towards temporal access control in cloud computing [12], resent an efficient temporal access control encryption scheme for cloud services with the help of cryptographic integer comparisons and a proxy-based re-encryption mechanism on the current time. Keyword Search With Access Control Over Encrypted Cloud Data [13], propose a scalable framework where user can use his attribute values and a search query to locally derive a search capability, and a file can be retrieved only when its keywords match the query and the user's attribute values can pass the policy check. Using this framework, we propose a novel scheme called KSAC, which enables keyword search with access control over encrypted data. KSAC utilizes a recent cryptographic primitive called hierarchical predicate encryption to enforce fine-grained access control and perform multi-field query search. Meanwhile, it also supports the search capability deviation, and achieves efficient access policy update as well as keyword update without compromising data privacy.

Secure and Efficient Attribute-Based Access Control for Multiauthority Cloud Storage [14], resent secure and cost-effective attribute-based data access control for cloud storage systems. Specifically, we construct a multiauthority CP-ABE scheme that features the system does not need a fully trusted central authority, and all attribute authorities independently issue secret keys for users. Then each attribute authority can dynamically remove any user from its domain such that those revoked users cannot access subsequently outsourced data; Also cloud servers can update the encrypted data from the current time period to the next one such that the revoked users cannot access those previously available data; and the update of secret keys and ciphertext is performed in a public way.

Fine-Grained Data Access Control Systems with User Accountability in Cloud Computing [15], present a way to implement, scalable and fine-grained access control systems based on attribute-based encryption (ABE). For the purpose of secure access control in cloud computing, the prevention of illegal key sharing among colluding users is missing from the existing access control systems based on ABE. This paper addresses this challenging open issue by defining and enforcing access policies based on data attributes and implementing user accountability by using traitor tracing. Furthermore, both the user grant and revocation are efficiently supported by using the broadcast encryption technique.

III Behavior Analysis Model with Level Based Access Restriction

The behavior analysis model maintains number of user profile where each profile contains number of information about the users and the list of services the users has access. Also, the method monitors the access of different services the user performs and stores them to the access traces. Using the access trace the method compute the secure access weight for different services. Based on the service access weight computed, the method grants or restricts the user access.

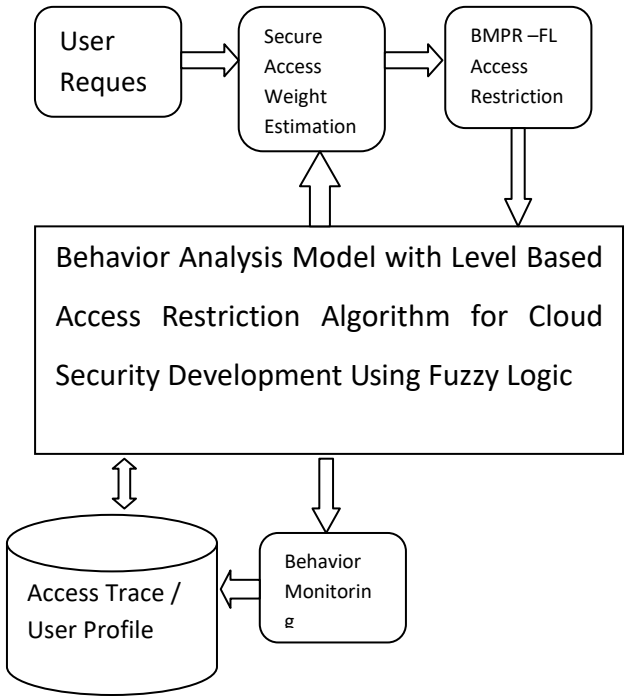


Figure 1. General Architecture of BMPR Access Restriction

The Figure 1, shows the general architecture of BMPR access restriction approach and shows different functional stages involved.

Consider N number of services present in the service pool, where K number of users has been registered to the cloud. Among N services, each service N_i would access the data attributes $Ø\{D_i,..D_n\}$ where only K-p number of user has access to all the attributes of the set Ø. Similarly only K-p number of user has access to all the services of N. Restricting the user k from accessing the service s to which he has no access is the key issue here. There are number of approaches has been discussed to solve this problem and each uses different measures and parameters. This paper consider the level of access for the user which is being computed based on the behavior of the user in accessing the service.

A. Behavior Monitoring:

The behavior of user in accessing the service would vary between different users. For example, consider there exist a service S_i, being accessed by the user u_k, which contains v number of states involved. The user u_i, would access the service and terminate in the v-n state blindly. Such intrinsic behavior can be used in to identify the users trustworthy while providing access to the user. The behavior of the user activity in accessing the services has been monitored here. This is an independent activity which tract the user request, how the request moving and it monitors each stage of the request. At each state of the request, the method tracks the status. Once the service has been terminated then it logs the status and other information to the access trace. Generated access trace being used to perform different activity in access restriction.

Initially, the service requested by the user can be identified from the service request Ur, as follows:

Service requested Sr = Service-ID $\in Ur$ -- (1)

It is necessary to identify and extract the data being submitted to the service Sr. It has been extracted as follows:

Sd =Service-Data$\in Ur$ -- (2)

Now, the number of states the request has been followed has to be identified using the following equation.

State Channel Sc = $\sum States \forall (Sr)$ --(3)

Now for each state present in the state channel, the service data submitted can be extracted using the equation (2).

The status of the request can be identified as follows:

Identify status of request Rs = Ur.Status.

Once all the states and their data and their status has been extracted, the access log can be generated and added to the access log set. The generated access log will be used to perform access restriction by computing the secure access weight.

Generate Access log Acl = $\sum_{i=1}^{size\,(State)} \left((S, Rs, Sed) \in Acl\right) \cup S, Rs, Sed$ --(4)

Here Sed is the service data, Acl-Access log, Rs- Status of request.

Generated log can be added to the trace as follows:

Access trace AT= $\sum (Traces \in AT) \cup Acl$ -- (5)

The behavior monitoring algorithm monitors the state of the request and status of the request to produce the access log to the trace. The traces generated have been used to compute the secure access weight for different user in the next stage to support access restriction to be performed in efficient manner.

B.Secure Access Weight Estimation:

The secure access weight is the measure which shows the trustworthy of user in accessing the list of attributes belongs to the service claimed. The method estimate the secure access weight at each stage of the service and by identifies the list of attributes the service accessing. Using all these, the method compute the secure access weight by computing the number of time the user has accessed the service or the attributes and the number of times it has been completed in success. This is estimated for each time window of the log. Also the factor of attribute level access is computed based on the number of attributes the user has access and the number of attributes the user does not have access. Using all these information, the method generates the fuzzy rule. Using the fuzzy rule generated and the user access details, the secure access weight has been estimated.

To compute the secure access weight, the user profile Up has been taken as the key with access trace AT.

For the service Sr being requested by the user Uk, the presence of access has been verified in the user profile Up. The verification of the access has been performed as follows:

$\forall (Profile\ p \in Up) if\ U_p$@Sr ,1,0 -- (6)

The equation verifies the presence of the service with the user profile and based on that it returns a value 1 or 0.

If the user has access to the service then, the list of attributes being accessed by the service has been identified as follows:

Identify list of attributes to be accessed Ats = $\sum Attributes@SER$ -- (7)

Further, the user would have access to limited attributes from the attribute set Ats. It is necessary to identify the list of attributes the user has access. It can be performed as follows:

Identify list of attributes the user has access UAAs = $\sum Attributes(Up) \in Ats$ --(8)

Using the values of equation (7) and (8), the attribute level factor can be measured as follows:

Compute Attribute level factor ALF = $size(UAAs) \Big/ Size(Ats)$

--(9)

Now the user access behavior value has to be computed. It can be measured by computing the number of times the user has accessed the service and number of times he has accessed in a proper manner. It can be measured as follows:

Compute Total Service Access TSA.

TSA= $\sum_{i=1}^{size\,(AT)} AT(i).User = UR.User \;\&\&\; AT(i).Service == SER$

--(10)

Compute Number of Successful Access NSA.

NSA = $\sum_{i=1}^{size\,(AT)} AT(i).User = UR.User \;\&\&\; AT(i).Service == SER \;\&\&\; AT(i).Status == Success$

--(11)

The NSA and TSA are estimated for each time window log. The using the values of different time window, the method generate the fuzzy rule.

To generate the rule, the method estimates the Minimum and maximum values of both the measures.

Generate Fuzzy Rule R.

Rule R = <TSA.Min, TSA.Max><NSA.Min, NSA.Max>

Using the values of (10) and (11), the secure access weight can be measured as follows:

Compute Secure Access Weight SAW.

SAW = $\frac{(NSA<NSA.Min,NSA.Max>(1,0))\times NSA}{(TSA<TSA.Min,TSA.Max>(1,0))\times TSA} \times ALF$

-- (12)

The secure access weight estimation algorithm computes the attribute level factor and secures access weight. The computed weight has been used to perform access restriction later.

C.BMPR-FL Access Restriction:

The behavior model based profile orient access restriction algorithm has been performed for each request being received from the user. The method identifies the user request and computes the secure access weight at each level. The service would have N number of levels or it may access different other services internally. For each service identified in the service life cycle, the method computes the secure access weight. Based on computed weight, the method performs access restriction in the cloud environment.

BMPR-FL Access Restriction Algorithm:

Input: User Request Ur, User Profile Up

Output: Boolean

Start

Read User request Ur.

Read User profile Up.

Identify the service ID SID = Ur.ServiceID.

Compute secure access weight SAW.

If SAW> WTh // weight threshold

Return Boolean

Else

Return false

End

Stop.

The BMPR-FL access restriction algorithm identifies the service being requested and computes the secure access weight. Based on the weight being computed, the method returns the Boolean value to allow or deny the request.

IV.Results and Discussion

The proposed behavior analysis model based hierarchical access restriction scheme has been implemented and evaluated for its performance. The method has produced efficient results in different parameters considered.

Parameter	Value
Protocol	BMPR-FL
Tool Used	Advance Java
Number of Services	100
Number of Attributes	500

Table 1: Details of Simulation

The Table 1, shows the details of simulation being used to evaluate the performance of the proposed BMPR algorithm.

Method	Access Restriction Performance %		
	50 Services	75 Services	100 Services
PBAC	81	85	89
BMPR-Fl	92	95	97.2

Table 2: Comparative Result on Access Restriction Performance

The Table 2, presents the comparative result on access restriction performance produced by different methods on varying number of services. The results show that the proposed BMPR algorithm has improved the access restriction performance in all the number of services considered.

Techniques	Time Complexity in seconds		
	50 Services	50 Services	50 Services
PBAC	56	56	56
BMPR-Fl	31	31	31

Table 3: Comparative Result on Time Complexity

The Table 3, presents the comparative result on time complexity performance produced by different methods on varying number of services. The results show that the proposed BMPR algorithm has reduced time complexity in all the number of services considered.

Techniques	Throughput Performance %		
	50 Services	75 Services	100 Services
PBAC	82	86	91
BMPR-Fl	85	91	98.3

Table 4: Comparative Result on Access Restriction Performance

The Table 4, presents the comparative result on throughput performance produced by different methods on varying number of services. The results show that the proposed BMPR-FL algorithm has improved the throughput performance in all the number of services considered.

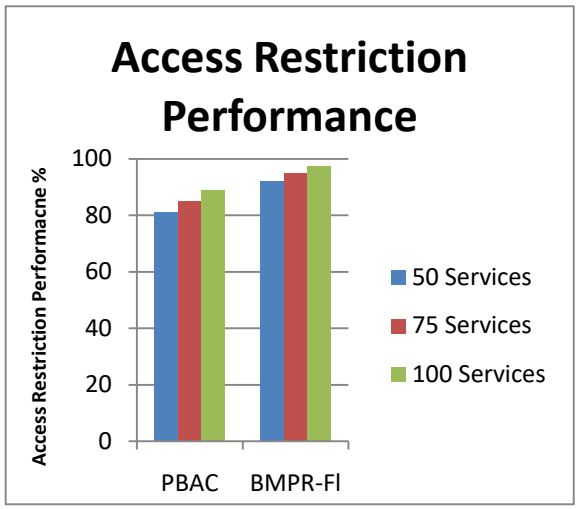


Figure 2: Comparison on Access Restriction Performance

The Figure 2, shows the comparative result on access restriction produced by different methods. The result shows that the proposed BMPR-FL algorithm has produced higher access restriction performance than other methods considered.

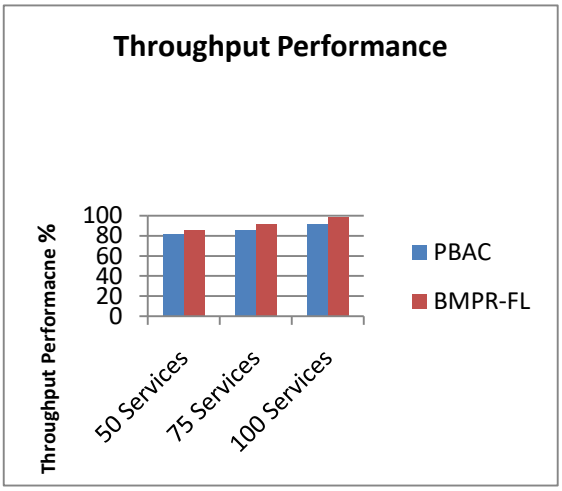


Figure 3: Comparison on throughput performance

The Figure 3, shows the comparison on throughput performance produced by different methods and shows that the proposed BMPR-FL algorithm has produced higher throughput than other methods.

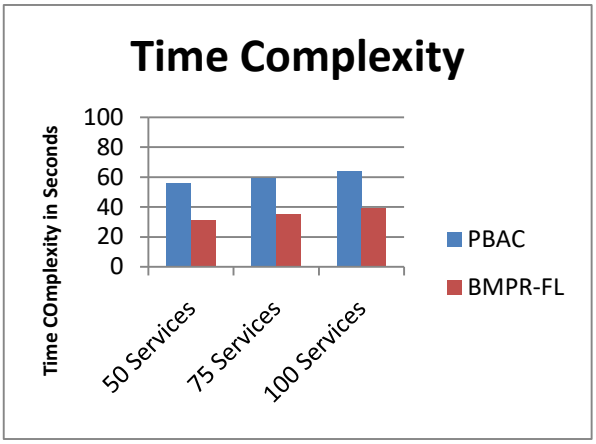


Figure 4: Comparison on time complexity

The Figure 4, shows the comparison on time complexity produced by different methods and shows clearly that the proposed BMPR-FL algorithm has produced less time complexity than others.

Conclusion:

In this paper, a behavior analysis model for access restriction in cloud environment has been presented. The method monitors the service access of the users and logs different state and their status to the access trace. Using the access trace available the method generates the fuzzy rule. Using the rule generated and access trace available, the method compute the secure access weight at each service state and based on that the method grant or deny the service access. The method produces efficient results in access restriction upto 97.2 % and throughput performance has been increased up to 98.3%. Also the time complexity of access restriction has been hugely reduced.

References

[1] Wei Teng ; Geng Yang Attribute-based Access Control with Constant-size Ciphertext in Cloud Computing, IEEE Transactions on Cloud Computing (Volume: PP, Issue: 99), Page(s): 1 – 1, 2015.

[2] Jun Luo, A Novel Role-based Access Control Model in Cloud Environments, JournalInternational Journal of Computational Intelligence Systems Volume 9, 2016 - Issue 1, 2016.

[3] Lixia Xie and Chong Wang, Cloud Multidomain Access Control Model Based on Role and Trust-Degree, Hindawi, Journal of Electrical and Computer Engineering Volume 2016 (2016).

[4] Kan Yang, Time-Domain Attribute-Based Access Control for Cloud-Based Video Content Sharing: A Cryptographic Approach, IEEE Transactions on Multimedia, Vol 18, Issue: 5, May 2016.

[5] Jongkil Kim, Surya Nepal, A Cryptographically Enforced Access Control with a Flexible User Revocation on Untrusted Cloud Storage, Data Science and Engineering , Volume 1, Issue 3, pp 149–160, 2016.

[6] Mehdi Sookhaka,, F. Richard Yua, Attribute-based data access control in mobile cloud computing: Taxonomy and open issues, Elsevier, Future Generation Computer Systems 72 (2017) 273–287.

[7] B.Srinivasa Rao, A Framework for Predicate Based Access Control Policies in Infrastructure as a Service Cloud, Int. Journal of Engineering Research and Applications, Vol. 6, Issue 2, (Part -6) February 2016, pp.36-44.

[8] Nicolai Paladi, Providing User Security Guarantees in Public Infrastructure Clouds, IEEE Transaction on Cloud Computing, Vol. 5, Issue 3, 2017.

[9] Jesus Luna, Quantitative Reasoning about Cloud Security Using Service Level Agreements, IEEE Transaction on Cloud Computing, Vol. 3, Issue 5, 2017.

[10] Zheng Yan, Flexible Data Access Control Based on Trust and Reputation in Cloud Computing, IEEE Transaction on cloud computing, Vol.5 Issue 3, 2017.

[11] Kei Fan, Privacy protection based access control scheme in cloud-based services, IEEE Transaction on China Communications, vol. 14, Issue 3, 2017.

[12] Yan Zhu, Towards temporal access control in cloud computing, IEEE, INFOCOM, 2012.

[13] Zhirong zen, Keyword Search With Access Control Over Encrypted Cloud Data, IEEE Transaction on sensor journal , vol 17, issue 3, 2017.

[14] Jianghong Wei , Secure and Efficient Attribute-Based Access Control for Multi-authority Cloud Storage, IEEE System Journal vol. issue 99, 2017.

[15] Jin Li, Fine-Grained Data Access Control Systems with User Accountability in Cloud Computing, Cloud Computing Technology and Science (Cloud-Com), 2010.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

AN APPROACH FOR ROAD TRAFFIC MANAGEMENT TO REDUCE TRAFFIC CONGESTION IN VANET

S.Angelin sophy
Research scholar,
Department of Computer Science,
Periyar University,
Salem, India

Dr.I.Laurence Aroquiaraj
Assistant Professor,
Department of Computer Science,
Periyar University,
Salem, India

ABSTRACT

Vehicular Ad-hoc network (VANET) is one of the best solutions which permit the vehicle to vehicle communication between nearby vehicles and nearby fixed equipment. In recent times transport efficiency plays an important part in the economic growth in advanced cities. Road traffic management involves monitoring of the actual traffic situation in real-time. With the goal of controlling the information from the vehicles, it is used in the traffic flows to reduce traffic congestion. This information is used to address with accidents and provide accurate and reliable traffic information in order to make predictions for drivers and transport authorities. Vehicle to Vehicle communication (V2V) and Vehicle to roadside Infrastructure communication (V2I) network is used to send and obtain the messages. The end result is simulated by using Intelligent Based Clustering Algorithm (IBCAV) and indicates this is one of the powerful ways to control congestion. The proposed technique guarantees reliable and well-timed delivery of messages to recognize congestion and avoids it.

Keywords: VANET, V2I, V2V, IBCAV algorithm.

1. INTRODUCTION

The vehicular ad-hoc network is a new form of network where nodes (i.e. vehicles) communicate with each other [1]. The ad-hoc network is a group of wireless mobile nodes without any fixed base station infrastructure and centralized control. In the Ad-hoc network, VANET is used for the continuous creation of a wireless network for data exchange in the domain moving vehicles. This fast growth within the number of vehicles on the roads has created a plethora of challenges for road traffic management authorities such as traffic congestion, increasing number of accidents, air pollution and so forth. VANET is a self sufficient and self-organizing wireless communication network. In this network, cars are called as nodes.

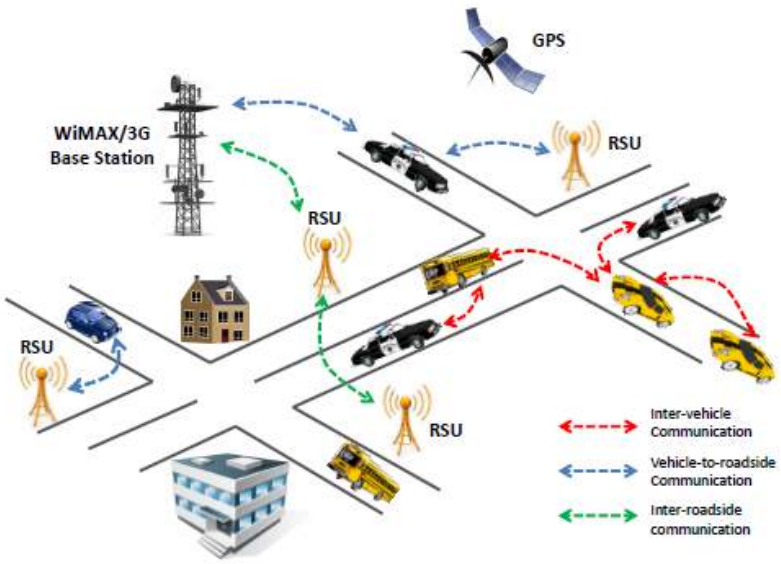


Figure 1. VANET Scenario

It involves themselves as servers or client for exchanging and sharing information. Each node acts as both host and router which moves arbitrarily and communicates with each other. The purpose of VANET is to switch data with high speed and to reduce the delay in communication range. The nature of the mobile node is continuously changed in VANET. The development of an effective system in the vehicular network has numerous advantages for road operators, as well as drivers point of view. Efficient movement

alarms and data about traffic incidents will diminish car influxes, enlarge street well- being and enhance the sheltered driving on the Highways. The mobility patterns of VANET are confined with road maps. The two main forms of common routing protocols are Proactive and Reactive. Proactive algorithms contain nodes as data. To maintain the validity of data, it should be communicated between nodes which use algorithms for high bandwidth. Proactive algorithms are also called as table-based algorithms. DSDV, OLSR and STAR are referred as the algorithms used for routing purposes [2, 3].

Reactive algorithms operate to detect a route only when there is a need for a route. This method of routing is utilized by AODV, DSR and TORA algorithms. The disadvantage of the proactive method relates to its poor scaling system and the problem with the reactive method is that before transmission of the primary packet it needs to detect the route. This ends at longer transmission time and when the route is detected from its authentic point of departure to its destination, it could disappear even before the transmission of the first packet due to the high motion of the nodes. The active duration of a path declines with the wide variety of hops and the route also fade away at the same speed it appeared [4, 5]. VANET has the highest number of nodes and the clustering approach is an efficient solution to the scalability issue. The remainder of this paper is organized as follows: Section II presents the related works, Section III discuss proposed work and Section IV describes the simulation and the results and Section V concludes the paper.

2. LITERATURE REVIEW

The traffic congestion is one of the biggest issues around the world. There are many factors causing traffic congestion, such as rush hour, road construction, accident and bad weather. Those factors can cause traffic congestion and drivers who are unaware of congestion. The more extreme congestion is, the more time it will take to clear. This [6] might lead to a more efficient use of road infrastructure. There is a need for a system that provides useful information to drivers about traffic situations, Information such as congestion type, location and boundaries. The traffic Congestion system must relay this information to drivers within the congestion and those heading close to it. Congestion information may be useful for many VANET applications, such as route planning or traffic. Congestion information is collected as the number of vehicles passing a point per unit time by some roadside equipment and transmitted to different places for broadcasting to vehicles. The aggregate number of vehicles in the global has encountered an astounding development, increasing activity thickness and bringing on more mishaps.

The primary issue with the networks is spreading messages to vehicles at high speed. Integrating V2I, V2R and V2V method can provide the solution to this and it is achieved with a weighted cluster algorithm (WCA) and computing the overall performance on different parameters of the network [7]. The congestion information [8] is commonly available only at a single macroscopic level for all vehicles and is not customized for the requirements of each vehicle. For vehicles within the congestion to form their own picture, they need to collaborate using Vehicle-to-Vehicle (V2V) or vehicle-to-infrastructure (V2I) [9] communication. The information needs to be relayed to vehicles far away from the congestion and the vehicles heading closer to it may take actions [10].

VANET allows a tight connection between physical driving and the communication system, which requires considering the impact of each area and it is termed as a cyber-physical system [11]. An application-oriented technique that regards the specific goals of traffic for VANET protocol design has been recently stated for huge-area transportation networks with respect to simulators [12] and for understanding the influence of communication to driving safety at the microscopic level [13]. This allows the fastest way based on the tragic state to avoid congestion. However the overall performance of location-based traffic services can fail sometimes due to the disadvantages of inability to use in urban areas with tall buildings, where the satellite signals may be blocked. Another strategy to reduce congestion is to optimize traffic signals [14] deployed at intersections by analyzing the data collected in real time traffic. Collision avoidance systems [15, 16] are designed to detect a traffic incident in real-time and rapidly relay this information to nearby vehicles to prevent a collision. These systems are very different from traffic congestion systems. Vehicle-based GPS systems are used to discover and disseminate traffic congestion information [17] and the system is known as COC for VANET. This system maintains and disseminates three types of information: Raw Information (level 1), density information (level 2) and congestion area information (level 3).

A versatile activity sign control system focused on V2V communication is introduced. It permits the holding up time of the vehicles at the crossing point with queue duration. The idea of clustering is utilized for the vehicles approaching the convergence [18]. The timing cycle of traffic signals is controlled through vehicles in groups and it uses DBCV algorithm. This algorithm is a mix of cluster and dissemination procedure and is utilized to accumulate the obliged dense data. Clusters are formed around the heading of the vehicles in a given geographic district approaching the convergence. Street-smart uses clustering algorithms that work over a distributed network where each node analyzes the collected statistics and eliminating the need for a central entity [19]. IBCAV is a combination of cluster size, velocity and density, it is compared to a manner in which each of the elements is considered separately for header selection and reduce cluster head selection operation followed by a reduced usage of network resources and decline end-to-end delay, throughput is increased [20].

3. PROPOSED WORK

The proposed work includes the idea for detection of congestion and provides information to the driver and also communicates those to other vehicles. Clustering strategies are normally applied in VANETs to reduce beneficial aid consumption and enhance the overall performance of the network. Vehicles have the unique role of clustering and only a few of them chosen as cluster heads and transmit packets of information. At the same time as the vehicles are positioned as a cluster, it is very important to select a cluster head. In IBCAV, RSU is selected as a cluster head when it is within the limits of the cluster. RSUs have stronger processing capabilities than other nodes and when they are motionless, there is no need to change the cluster head. Traffic load is shaped by many factors such as street development, container necks climate and so on. The drivers are unaware of traffic load and the traffic burden is the additional time it will take to clear. The capacity of a driver is to recognize the road conditions and it will allow them to look for the alternate ways. In order to provide useful information to drivers about traffic, a system must recognize the traffic load, position and relay the data to drivers. These requirements are satisfied by traffic detection system.

To find the traffic load, an observer needs to identify the vehicles which are away from each other. Vehicles in the congestion have collaborated the usage of V2V (vehicle-to-vehicle) or V2I (vehicle-to-infrastructure) communication. Vehicle to Vehicle Communication is a technology which enables the vehicles to communicate with each other. Vehicle to Infrastructure communications is the wireless exchange of data between vehicles and highway infrastructure. In vehicle to infrastructure communication, the vehicle and the decision server communicate

with each other for providing the traffic. In Vehicle to Road Side Unit communication, Road Side Units (RSUs) use the data of moving nodes like speed, distance from RSU, and route information of vehicles. This is a communication infrastructure used to support the route information during traffic. The different system parameters are,

Data Collection: Collect data from the environment such as the current location and speed.

Information Sharing: The information is shared between V2V and V2I communication.

Decision Server: Appropriate decision has to be taken so that the congestion can be detected.

Message Broadcast: The warning message is broadcast continuously and provides information to the driver and also communicates these to other vehicles.

4. SIMULATION RESULT

The proposed approach has distinctive scenarios. First, the appropriate communication method for the vehicle 2 vehicle communication using Intelligent Based Clustering Algorithm (IBCAV) is used. The main aim of using this algorithm is to enhance the performance of the network. In IBCAV a few essential elements are considered such as cluster size, velocity and density. In clustering techniques, moving nodes are divided into different groups and put together in a single cluster based on certain rules. The size of each cluster represents the range of vehicles in a highway, within a communication range and form a cluster. In each cluster, a vehicle should be designated as a cluster head. IBCAV locate the quality of communicating nodes (v) by which the stability of network parameters is stable. The key parameters are connectivity, mobility and speed. Connectivity parameter indicates that the vehicles are in identical range or not. Mobility indicates the distance between the vehicles and the speed indicates the traveling time of the vehicle.

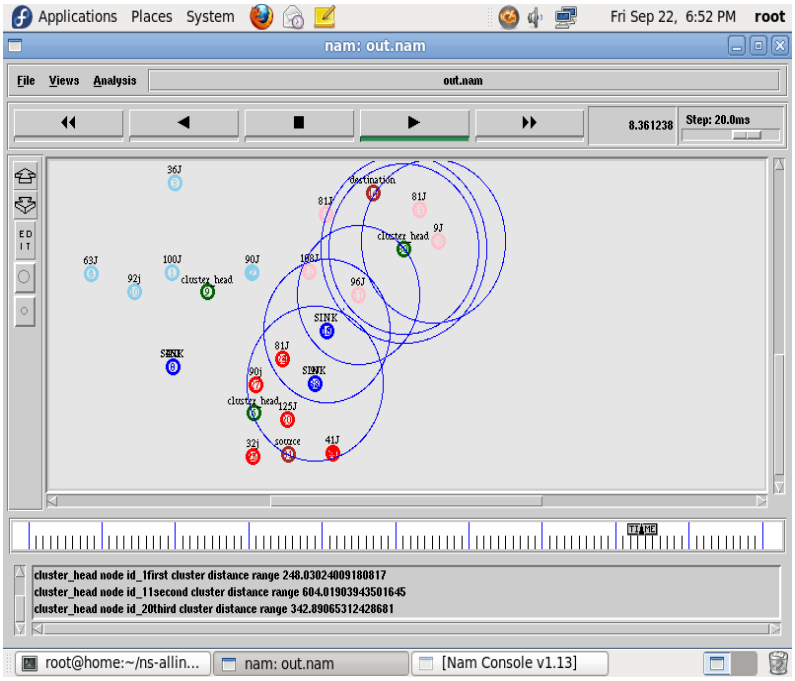


Figure 1. Simulation Scenario of nodes

The performance of the vehicular network is calculated by means of parameters such as Packet Delivery Ratio, Routing Overhead and End to End Delay. Throughput is defined as an average number of bits, bytes or packets per unit time.

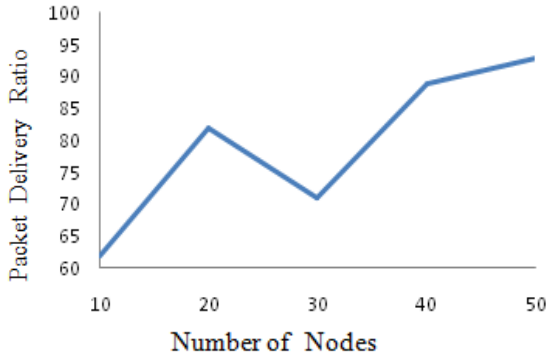


Figure 2. Packet Delivery Ratio

The Packet Delivery Ratio is the ratio of the received packet and sum of dropped and received packets in the network.

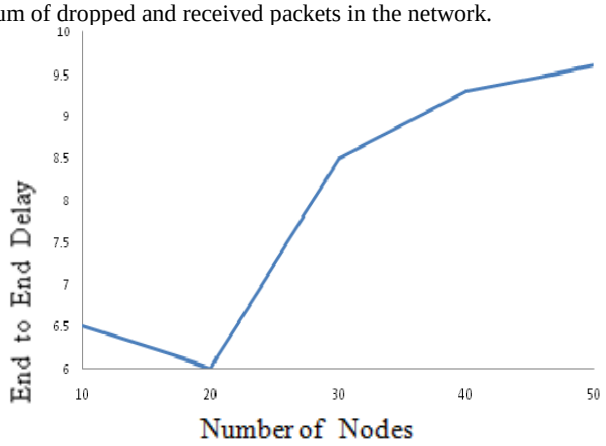


Figure 3. End to End Delay

The End to End delay is a time required by a packet to reach its destination

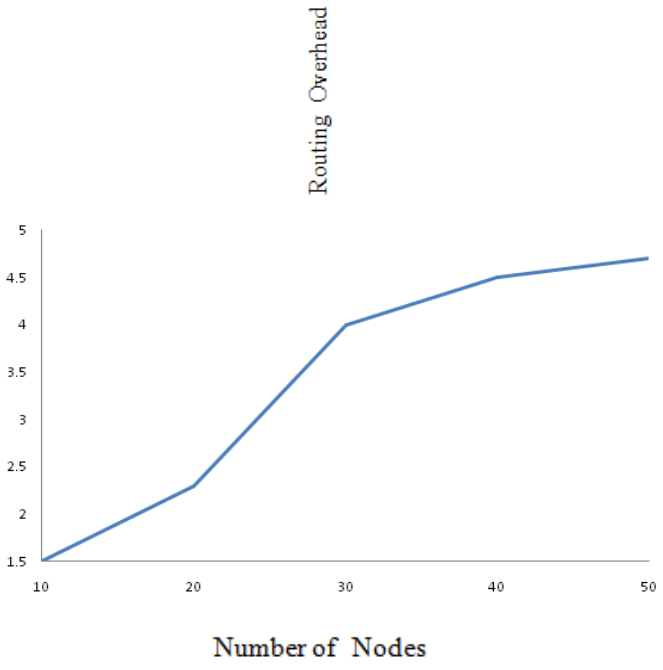


Figure 4. Routing Overhead

The Routing Overhead is the total number of routing packets travel in the network at simulation time.

5. CONCLUSION

VANET is a network of vehicles where nodes are capable to communicate with each other. This work evaluated the performance of VANET with parameters, i.e. routing overhead, packet delivery ratio, end-to-end delay and throughput. The different events happen on the road and its condition information is broadcasted to alert the drivers about the congestion. This timely information is useful for taking a decision and to the broadcasted message. The main issue in the network is to spread messages in vehicles at high speed. In IBCAV, control packets of network decreases and packet delivery ratio increases. Simulation results indicate that each of the elements is considered separately for header selection and reduce the cluster head selection operation by a reduced usage of network resources with the decline of end-to-end delay and throughput is increased.

REFERENCES

[1] Azzedine Boukerche, Horacio A.B.F. Oliveira, Eduardo F. Nakamura, Antonio A.F. Loureiro, "Vehicular Ad Hoc Networks: A New Challenge for Localization-Based Systems", A.Boukerche et al./Computer Communication (2008).

[2] Sh. XU, S. Lee, "A Direct Routing Algorithm with Less Route Built Throughput", International Conference on Traffic and Transportation Engineering (ICICN 2012).

[3] U. Nagaraj, Dr. M. U. Kharat, P. Dhamal, "Study of Various Routing Protocols in VANET", International Journal of Computer Science & Technology, IJCST, Vol. 2, Issue 4, 2011.

[4] K. Pandy, A. Swaroop, "A Comprehensive Performance Analysis of Proactive, Reactive and Hybrid MANETs Routing Protocols", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, 2011.

[5] R. Kumar, M. Dave, "A Comparative Study of Various Routing Protocols in VANET", International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, 2011.

[6] Pritesh Patwa Rudra Dutta," Joint Modeling of Mobility and Communication in a V2V Network for Congestion Amelioration", Computer Science, North Carolina State University, (2006).

[7] Prashant Panse,Dr.Tarun Shrimali,Dr.Meenu Dave - "An Approach for preventing Accidents and Traffic Load Detection on Highways using V2V Communication in VANET", International Journal of Information,Communication and Computing Technology (IJICCT), 2016.

[8] Reshma R, Nayak, Sahana S K," Smart Traffic Congestion Control Using Wireless Communication", Journal of Advanced Research in Computer and Communication Engineering, (September 2013).

[9] Mohamed Watfa," Advances in Vehicular Ad-Hoc Networks: Developments and Challenges", University of Wollongong, UAE, (2010).

[10] Tarun Prakash, Ritu Tiwari," Counter-based Traffic Management Scheme for Vehicular Networks", Journal of Emerging Trends in Computing and Information Sciences, JUNE (2011).

[11] E. A. Lee,"Cyber physical systems, Design challenges, Elect. Eng, Comput.Sci.Dept, Univ.California, Berkeley, CA,Tech. Rep.UCB/EECS-2008.

[12] T. Gaugel, F, Schmidt-Eisenlohr, J, Mittag, and H. Hartenstein, "A change in perspective: Information-centric modeling of inter-vehicle communication," in Proc.8th ACM Int.Workshop VANET,2011, pp. 61–66.

[13] D.Baselt, M.Mauve and B.Scheuermann, "A top-down approach to inter-vehicle communication (poster)," in Proc. 3rd Annu. IEEE VNC, 2011, pp. 123–130.

[14] Barba, C.T, Mateos, M.A, Soto P, Mezher A.M, Igartua, M.A, "Smart city for VANETs using warning messages, traffic statistics and intelligent traffic lights," Intelligent Vehicles Symposium (IV), 2012 IEEE, vol., no., pp.902,907, 3-7 June 2012.

[15] Rawashdeh Z.Y and Mahmud S.M," Intersection Collision Avoidance System Architecture,5th IEEE Consumer Communications and Networking Conference, 2008, pp. 493 - 494.

[16] Mak, T.K, K.P Laberteaux and R Sengupta," A Multi-Channel VANET Providing Concurrent Safety and Commercial Service," Proceedings of the 2nd ACM international workshop on Vehicular Adhoc networks, January 2005.

[17] Fukumoto, Junya, et al,"Analytic Method for Real-Time Traffic Problems by Using Content Oriented Communications in VANET", 7th International Conference on ITS, 2007, pp. 1-6

[18] Maslekar N, Boussedjra M, Mouzna J, Labiod H," VANET based Adaptive Traffic signal Control",IEEE 73rd Vehicular Technology Conference (VTC Spring), pp.1-5,2011.

[19] Dornbush S and Joshi A," StreetSmart Traffic: Discovering and Disseminating Automobile Congestion using VANET", Vehicular Technology Conference. 2007, pp. 11-15.

[20] Mehrnaz Mottahedi, Sam Jabbehdari and Sepideh Adabi," IBCAV: Intelligent Based Clustering Algorithm in VANET", International journal of computer science issues, vol 10,issue 1, January 2013.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A STUDY ON EFFICIENT ENERGY MANAGEMENT SYSTEM IN ADHOC WIRELESS NETWORKS

A . Emmanuel Peo Mariadas
Research Scholar,
Department of Computer Science and Engineering,
Annamalai University, Tamil Nadu , India.
emmanuelpeo@gmail.com

Dr. R .Madhanmohan
Assistant Professor,
Department of Computer Science and Engineering,
Annamalai University, Tamil Nadu , India.
madhanmohan_mithu@yahoo.com.

ABSTRACT

A Mobile Ad-hoc Network (MANET) is a self-configuring network composed of mobile nodes without any fixed infrastructure. Wireless adhoc networks can be used to manage services anywhere and anytime. Ad hoc networks enable users to spontaneously construct a dynamic communication system. They do users to retrieve the services offered by the fixed network over multihop communications, without requiring infrastructure in the user immediacy. Wireless networks are come at from all sides by demand of communication bandwidth and correctly a key issue is to become user requests with minimum service delay. Network nodes have limited energy resources; the energy expended for transferring information across the network has anticipated minimized. Adhoc wireless networks are constrained by restrictive battery power, which makes energy management an important issue. Adhoc Networks are more active to these issues where each mobile device is act like a router and consequently, routing delay adds considerably to everywhere end-to end delay. In this paper , focus on energy management schemes, energy efficient routing Protocol which tries to approach the challenge of using battery power efficiently.

Keywords: Adhoc Networks, Energy Management, Routing Protocols.

I.INTRODUCTION

An Efficient battery management and transmission power management hugely deals with Energy management. In few years, there has been an explosive growth of interest in mobile computing [11], as well as in delivering World Wide Web content and streaming traffic to radio devices. There is a huge potential market for providing personal communication systems with access to airline schedules, weather forecasts, or location-dependent information. However, to offer high-quality and low-cost services to ad hoc network nodes, several technical challenges still need to be addressed. Mobile ad-hoc networks can turn the vision of getting connected "anywhere and at any time" into reality. Recent advancements such as Bluetooth introduced a new type of wireless systems known as mobile ad-hocnetworks. Mobile ad-hoc networks or "short live" networks operate in the absence of fixed infrastructure. Nodes in mobile ad hoc networks are constrained by limited battery power for their operation. Hence, energy efficiency is an important issue in adhoc networks.

The characteristics of mobile networks are summarized as follows:

- Communication via wireless means.
- Performance of nodes can be the roles of both hosts and routers.
- No centralized controller and infrastructure.
- Dynamic network topology.
- Frequent routing updates.
- Autonomous, no infrastructure needed.
- Can be set up anywhere.
- Energy constraintsand Limited security.

A. Why Energy Management isneededin Ad hoc Networks

In adhoc wirelessnetworks,mobile computation devices are usually battery powered. A limited energy budget constrains the computation and communication capacity of each device. Energy resources and computation workloads have different distributions within the network. Devices that expend all their energy can only be recharged when they leave the network. In wireless networks, the ratio of computation energy consumption to communication energy consumption varies in a wide range, depending on application type.

In some applications like micro sensor networks, communication dominates energy consumption. In other application domains and applications like simulation, artificial intelligence, target detection, handwriting recognition, and voice recognition computation energy consumption generally dominates communication energy consumption.

B. Reasons for energy management in Ad hoc networks

1) Limited energy reserve:

The ad hoc networks have limited energy reserve. The improvement in battery technologies is very slow as compared to the advances in the field of mobile computing and communication.

2) Difficulties in replacing the batteries:

In situations like battlefields, natural disasters such as earthquakes, and so forth, it is very difficult to replace and recharge the batteries.Thus, in such situations, the conservation of energy is very important.

3) Lack of central coordination:

Because an ad hoc network is a distributed network and there is no central coordinator, some of the nodes in the multihop routing should act as a relay node. If there is highest relay traffic, this leads to preferably power consumption at the respective relay node.

4) Constraints on the battery source:

The weight of the nodes may increase with the weight of the battery at that node. If the weight of the battery is decreased, that in turn will lead to less power of the battery and thus decrease the life span of the battery. Thus, energy management techniques must deal with this issue; in addition to reducing the size of the battery, they must utilize the energy resources in the best possible way.

5) Selection of optimal transmission power:

The increase in the transmission power increases the consumption of the battery charge. Because the transmission power

decides the reachability of the nodes, an optimal transmission power decreases the interference between nodes, and that in turn increases the number of simultaneous transmissions [17].

II.CLASSIFICATION OF ENERGY MANAGEMENTSCHEMES

A better understanding of the capabilities and limitations of the energy resources of the nodes is maintained to improve the life of an ad hoc wireless network. A longer lifetime of the node can be achieved by increasing the battery capacity. Increasing the capacity of the battery at the nodes can be achieved by either battery management, which concerns the internal characteristics of the battery, or power management, which deals with utilizing the battery capacity to the maximum possible extent. The Figure (a) shows an overview of Energy management system.

The main three categories of Energy management system can be divided into:

- **Battery management system**
- **Transmission power management**
- **System power management**

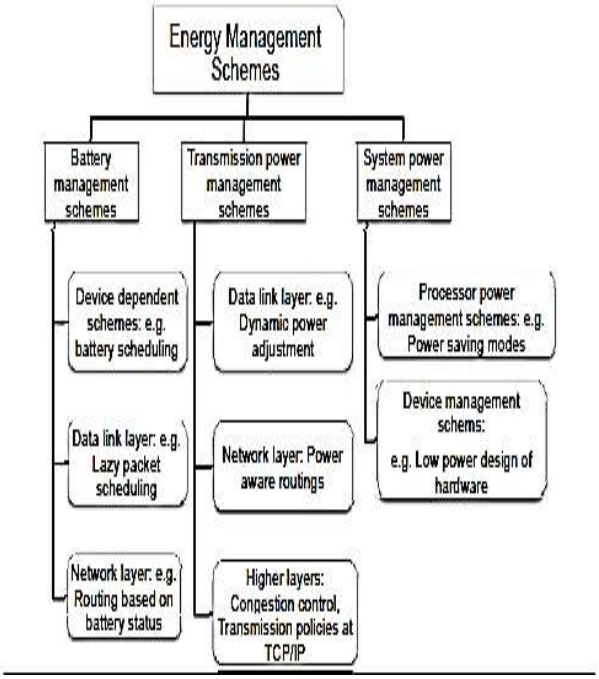


Figure (a) : Classification of Energy Management System

III.POWER MANAGEMENT OF ADHOC NETWORKS

The equipment in ad hoc networks permanently uses exhaustible energy as their power supply such as batteries .[16][17].

The genuine fact that mobile computing is evolving rapidly with advances in wireless communications and devices getting smaller and more efficient, advances in battery technology have not yet reached the stage where a mobile computer can operate for days without recharging[18].

Therefore, advancedpower conservation techniques are necessary. A variety oftechniques can be used to cope with power scarcity.

The Table1 shows the power management at various protocol layers.

Protocol Layer	Power Conservation Techniques
Data-Link Layer	Avoid unnecessary retransmission. Avoid collision in channel access whenever possible. Put receive in standby mode wheneverpossible. Use or allocate contiguous slots for transmission and reception whenever possible.Turn radio off (sleep) when not transmitting orreceiving.
Network Layer	Consider route relaying load.Consider battery life in route selection. Reduce frequency of sending control message. Optimize size of control headers. Efficient routereconfiguration techniques.
Transport Layer.	Avoid repeated retransmissions. Handle packetloss in a localized manner. Use power-efficienterror control schemes.
Application Layer	Adopt an adaptive mobile quality of service (QoS) framework. Move power-intensive computation from a mobile host to the basestation. Use proxies for mobile clients. Proxiescan be designed to make applications adapt topower or bandwidth constraints. Proxiescanintelligently cachefrequently used information, suppress video transmission and allow audio, and employ a variety of method to conservepower.

Table 1: Power Management at Various Layers

IV.ENERGY MANAGEMENTIN AD HOC NETWORKS

A very important aspect of the overall management of ad hoc networks is the energy management [2]. The mobile wireless sensor nodes in the field need to conserve energy and use it optimally in order to play the assigned role in an ad hoc network for a longer period of time.A various levels of Energy management are: Component level ,System level and Network level.

1. Component Level Energy Management

Component level energy management (CEM) [3] gives an opportunity to control the energy utilization by various components of a system. There are several components that are part of a system that get used during initialization, and other components that get used at irregular intervals. By a suitable design, if the energy consumed by these components during idle time can be reduced close to zero, the main operation of the system can be better sustained.

2. System Level Energy Management

Adhoc wireless networks carrying different kinds of traffic. A node in such a network is the system under consideration and considers a scenario in which multiple tasks handle traffic of different kinds. If we use the available energy in appropriateway it may be better to trade-off low priority traffic so as to be able to handle high priority traffic [10].

3. Network Level Energy Management

The issues related to energy management are deled at Network level. The objective is to conserve energy at network level by cooperating loading of neighboring nodes.

V. ENERGY EFFICIENT PROTOCOL

The major challenge in wireless networks is obviously Energy efficiency. To facilitatecommunication, most wireless

network devices are portable and battery-powered, and thus operate on an extremely constrained energy budget. However, progress in battery technology shows that only small improvements in battery capacity can be expected in the near future.Likewise recharging or replacing batteries is expensive or, under some situation, impossible, it is appropriate to keep the energy dissipation level of devices low.

An Adhoc network is a collection of two or more nodes equipped with wireless communications and networking capabilities without central network control, namely, an infrastructure less mobile network.

The important and challenging networks is Energy-efficient design in ad hoc networks is preferably than with other wireless networks.

First, due to the absence of an infrastructure, nodes in an ad hoc network must act as routers and join in the process of forwarding packets. Therefore, traffic loads in ad hoc network are heavier than in other wireless networks with fixed access points or base stations, and thus ad hoc network have more energy consumption.

Second, the trade-offs between different network performance criteria is needs directs to energy-efficient design.

Third, no centralized control implies that energy-efficient managementmust be done in a distributed and cooperative manner, whichis difficult to achieve [5] [7]. In the wireless interface, energyconsumption in idle mode is only slightly less than in transmitmode and almost equal to that of receive mode. Therefore, anetwork protocol is needed to maximize the time a device is insleep mode and also maximizes the number of wirelessdevices that can be in sleep mode. Many protocols have beenproposed to deal with this challenge. Here we are discussingabout the Energy Efficient Medium Access Control Protocol.

A. Energy-Efficient Medium Access Control (EE-MAC)Protocol

The Energy-Efficient Medium Access Control (EE-MAC) Protocol is based on the circumstance that most applications of ad hoc networks are data driven, which approach that the sole purpose of forming an ad hoc network is to make and resolve data.Hence, keeping all network nodes awake is costly andunnecessary when some nodes do not have traffic to carry [7].The protocol conserves energy by turning on and off theradios of specific nodes in the network. The goal is to reduceenergy consumption without significantly reducing networkperformance. The key idea of EE-MAC is to select masternodes from all nodes in the network. Master nodes stay awakeall the time and act as a virtual backbone to route packets inthe ad hoc network. Other nodes, called “slave nodes,” remainin an energy efficient mode and wake up periodically onlyduring signal intervals to check whether they have packets toreceive [5].

B. Features of EE-MAC Protocol:

In EE-MAC the mastersdon’t operate in power saving mode and forward packets allthe time, the packet delivery ratio and packet delay can beimproved.

The features of EE-MAC are

- Entering Sleep Mode Earlier
- Priority processing of packets to slaves
- Prolonging the sleep period for slaves

Performance:

Our main concern is energy efficiency; energylevel is given higher weight than connectivity. considersome metrics to evaluate the network performance which willdiffer from those used by others.

- Data Packet Delivery Ratio.
- End-to-end delay.
- Energy efficiency.

VI. CONCLUSION

In this paper, the study of energy consumption behaviour of various routing protocols is being analysed and discussed how to perform energy efficiently in wireless ad hoc networks. Because the nodes are mobile and can be used for emergency purposes like military or natural disasters, each node should employ its battery efficiently. Some of the problems which are faced meanwhile managing energy and are limited energy reserve, difficulties in replacing batteries, demand of central coordination, and constraints on the battery source. Furthermore, energy management in Ad hoc Networks at various levels like Component, System and Network, as well as the energy management schemes, power management in Ad hoc Networks. The challenge is not to grant each node with higher battery power but to handle the available battery power in a very efficient manner. We dealt with Energy-Efficient Medium Access Control (EE-MAC) Protocol which tries to approach the challenge of using battery power efficiently.

VII. REFERENCES

[1] Wang Yu, “Study on Energy Conservation in MANET”, Journal of Networks, Vol. 5, June 2010.

[2] Sridhar G and Sridhar V “Energy Management in Ad Hoc Mobile Wireless Networks”.

[3] Yuen, W.H.and C.W. Sung,“On Network Connectivity and Energy Efficiency of Mobile AdHoc Networks,” Proceedings of IEEE ICDCS 2003.

[4] A.Ephremides,“Energy concerns inwireless networks. IEEE WirelessCommunications” 2002.

[5] Subir Kumar Sarkar, T G BasavaRaju , “AdhocMobile Wireless Networks”

[6] A. J. Goldsmith, S. B. Wicker, “Design challenges for energy constrainedadhoc wireless networks, IEEE Wireless Communications” 9(4): 8–27,2002.

[7] C. E. Jones, K. M. Sivalingam, P. Agrawal, J. C. Chen, “A survey of energy efficient network protocols for Wireless Networks” 7(4):343–358,2001.

[8] T.A. ElBatt, S.V. Krishnamurthy, D. Connors, and S. Dao, “Power Management for Throughput Enhancement in Wireless Ad-Hoc networks,” Proceedings of IEEE ICC’00, New Orleans, 2000.

[9] V. Rudolph, T. H. Meng, “Minimum energy mobile wireless networks, IEEE Journal of Selected Areas in Communications” 17(8):1333–1344, 1999.

[10] Lettri. P and M.B. Srivastava , “Advances in wireless Terminals” IEEE 1999.

[11] Ephremides, Energy concern in wireless networks. IEEE WirelessCommunications,9(4):48–59, 2002.

[12] C. Perkins, Ad Hoc Networking, Addison-Wesley: Reading, MA, 2001,1–28.

[13] G. Forman, J. Zahorjan, The challenges of mobile computing, IEEE Computer, 27(4):38–47,1994.

[14] A. J. Goldsmith, S. B. Wicker, Design challenges for energy constrained adhoc wireless networks, IEEE Wireless Communications, 9(4): 8–27, 2002.

[15] Y. Xu, J. Heidemann, D. Estrin, Geography-informed energy conservation for ad hoc routing, Proceedings of International Conference on Mobile Computing and Networking(MobiCom’2001),2001, pp. 70–84.

[16] H. Woesner, J-P. Ebert, M. Schlager, A. Wolisz, Power-saving mechanisms in emerging standards for wireless LANs: the MAC level perspective, IEEE Personal Communications, 5(3):40–48, 1998.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

ANSWERING PATTERN QUERIES USING VIEWS

Dr.S.T.Deepa
Associate professor, Department of Computer Science
Shri S.S.Shasun Jain College for Women
Chennai, India
email: deepatheodore@gmail.com

Ms.G.S. Shailaja
Computer Instructor
Chennai Higher Secondary School
Chennai, India
e-mail: shakthi.s2012@gmail.com

ABSTRACT

Information is playing an important role in our lives. One of the major sources of information is databases. Databases and database technology are having major impact on the growing use of computers. In order to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output. The Structured Query Language (SQL) norms are been pursued in almost all languages for relational database systems. However, not everybody is able to write SQL queries as they may not be aware of the structure of the database. So there is a need for non-expert users to query relational databases in their natural language instead of working with the values of the attributes. The idea of using natural language instead of SQL, has promoted the development of Natural Language Interface to Database systems (NLIDB). The need of NLIDB is increasing day by day as more and more people access information through web browsers, PDA's and cell phones. In this paper we introduce an intelligent interface for database. We prove that our NLIDB is guaranteed to map a natural language query to the corresponding SQL query. We have tested our system on Northwind database and show that our NLIDB compares favourably with MS English Query product

Keywords: Natural language, pattern, containment, item set, queries.

I. INTRODUCTION

Data mining is the process of discovering interesting patterns from massive amounts of data. As a knowledge discovery process, it typically involves data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation. The major dimensions of data mining are data, knowledge, technologies, and applications. The book focuses on fundamental data mining concepts and techniques for discovering interesting patterns from data in various applications. Prominent techniques for developing effective, efficient, and scalable data mining tools are focused on. This chapter discusses why data mining is in high demand and how it is part of the natural evolution of information technology. It defines data mining with respect to the knowledge discovery process. Next, data mining from many aspects, such as the kinds of data that can be mined, the kinds of knowledge to be mined, the kinds of technologies to be used and targeted applications are discussed which helps gain a multidimensional view of data mining. Data mining can be conducted on any kind of data as long as the data are meaningful for a target application, such as database data, data warehouse data, transactional data, and advanced data types. Finally major data mining research and development issues are outlined.

II. LITERATURE REVIEW

The problem of answering queries using views is to find efficient methods of answering a query using a set of previously defined materialized views over the database, rather than accessing the database relations[1]. The problem has recently received significant attention because of its relevance to a wide variety of data management problems. In query optimization, finding a rewriting of a query using a set of materialized views can yield a more efficient query execution plan. To support the separation of the logical and physical views of data, a storage schema can be described using views over the logical schema. As a result, finding a query execution plan that accesses the storage amounts to solving the problem of answering queries using views. Finally, the problem arises in data integration systems, where data sources can be described as precomputed views over a mediated schema. This article surveys the state of the art on the problem of answering queries using views, and synthesizes the disparate works into a coherent framework. We describe the different applications of the problem, the algorithms proposed to solve it and the relevant theoretical results.

Data integration is the problem of combining data residing at different sources, and providing the user with a unified view of these data[2]. The problem of designing data integration systems is important in current real world applications, and is characterized by a number of issues that are interesting from a theoretical point of view. This document presents on overview of the material to be presented in a tutorial on data integration. The tutorial is focused on some of the theoretical issues that are relevant for data integration. Special attention will be devoted to the following aspects: modeling a data integration application, processing queries in data integration, dealing with inconsistent data sources, and reasoning on queries.

The problem of answering tree pattern queries using views was revisited[3].They first show that, for queries and views that do not have nodes labeled with the wildcard *, there is an alternative to the approach of query rewriting which does not require us to find any rewritings explicitly yet which produces the same answers as the maximal contained rewriting. Then, using the new approach, they give a simple criterion and a corresponding algorithm for identifying redundant view answers, which are view answers that can be ignored when evaluating the maximal contained rewriting. Finally, for queries and views that do have nodes labeled *,they provide a method to find the maximal contained rewriting and show how to answer the query using views without explicitly finding the rewritings.

View-based query processing requires answering a query posed to a database only on the basis of the information on a set of views, which are again queries over the same database[4].This problem is relevant in many aspects of database management, and has been addressed by means of two basic approaches: query rewriting and query answering. In the former approach, one tries to compute a rewriting of the query in terms of the views, whereas in the latter, one aims at directly answering the query based on the view extensions. They study view based query processing for the case of regular-path queries, which are the basic querying mechanisms for the emergent field of semi structured data. Based on recent results, They first show that a rewriting is in general a co-NP function wrt to the size of view extensions. Hence, the problem arises of characterizing which instances of the problem admit a rewriting that is PTIME. A second contribution of the work is to establish a tight connection between view based query answering and constraint satisfaction problems, which allows us to show that the above characterization is going to be difficult. As a third contribution, we present two methods for computing PTIME rewritings of specific forms. The first method, which is based on the established connection with constraint

satisfaction problems, gives us rewritings expressed in Data log with a fixed number of variables. The second method, based on automata-theoretic techniques, gives us rewritings that are formulated as unions of conjunctive regular-path queries with a fixed number of variables.

The problem of query rewriting for TSL, a language for querying semi structured data is addressed [5]. They develop and present an algorithm that, given a semi structured query q and a set of semi structured views V, finds rewriting queries, i.e., queries that access the views and produce the same result as q. The algorithm is based on appropriately generalizing containment mappings, the chase, and unication techniques that were developed for structured, relational data. We also develop an algorithm for equivalence checking of TSL queries. We show that the algorithm is sound and complete for TSL, i.e., it always finds every TSL rewriting query of q, and we discuss its complexity. They extend the rewriting algorithm to use available structural constraints (such as DTDs) to find more opportunities for query rewriting. We currently incorporate the algorithm in the TSIMMIS system.

The problem of maintaining materialized views of graph structured data was studied **[6]**. The base data consists of records containing identifiers of other records. The data could represent traditional objects (with methods, attributes, and a class hierarchy), but it could also represent a lower level data structure. They define simple views and materialized views for such graph structured data, analyzing options for representing record identity and references in the view. We develop incremental maintenance algorithms for these views.

[7]Developers of rapidly growing applications must be able to anticipate potential scalability problems before they cause performance issues in production environments. A new type of data independence, called scale independence, seeks to address this challenge by guaranteeing a bounded amount of work is required to execute all queries in an application, independent of the size of the underlying data. While optimization strategies have been developed to provide these guarantees for the class of queries that are scale-independent when executed using simple indexes, there are important queries for which such techniques are insufficient. Executing these more complex queries scale-independently requires precomputation using incrementally-maintained materialized views. However, since this precomputation effectively shifts some of the query processing burden from execution time to insertion time, a scale-independent system must be careful to ensure that storage and maintenance costs do not threaten scalability. In this paper, we describe a scale independent view selection and maintenance system, which uses novel static analysis techniques that ensure that created views do not themselves become scaling bottlenecks. Finally, we present an empirical analysis that includes all the queries from the TPC-W benchmark and validates our implementation's ability to maintain nearly constant high quantile query and update latency even as an application scales to hundreds of machines.

[8]To make query answering feasible in big datasets, practitioners have been looking into the notion of scale independence of queries. Intuitively, such queries require only a relatively small subset of the data, whose size is determined by the query and access methods rather than the size of the dataset itself. This paper aims to formalize this notion and study its properties. We start by defining what it means to be scale-independent, and provide matching upper and lower bounds for checking scale independence, for queries in various languages, and for combined and data complexity. Since the complexity turns out to be rather high, and since scale-independent queries cannot be captured syntactically, we develop sufficient conditions for scale independence. We formulate them based on access schemas, which combine indexing and constraints together with bounds on the sizes of retrieved data sets. We then study two variations of scale independent query answering, inspired by existing practical systems. One concerns incremental query answering: we check when query answers can be maintained in response to updates scale-independently. The other explores scale independent query rewriting using views.

The design and the evaluation of the ADMS optimizer was described**[9]**. Capitalizing on a structure called Logical Access Path Schema to model the derivation relationship among cached query results, the optimizer is able to perform query matching coincidently with the optimization and generate more efficient query plans using cached results. The optimizer also features data caching and pointer caching, different cache replacement strategies, and different cache update strategies. An extensive set of experiments were conducted and the results showed that pointer caching and dynamic cache update strategies substantially speedup query computations and, thus, increase query throughput under situations with fair query correlation and update load. The requirement of the cache space is relatively small and the extra computation overhead introduced by the caching and matching mechanism is more than offset by the time saved in query processing.

The prevalent use of XML highlights the need for a generic, flexible access-control mechanism for XML documents that supports efficient and secure query access, without revealing sensitive information to unauthorized users 10]. A novel paradigm for specifying XML security constraints and investigates the enforcement of such constraints during XML query evaluation is introduced. The approach is based on the novel concept of security views, which provide for each user group (a) an XML view consisting of all and only the information that the users are authorized to access, and (b) a view DTD that the XML view conforms to. Security views effectively protect sensitive data from access and potential inferences by unauthorized users, and provide authorized users with necessary schema information to facilitate effective query formulation and optimization. We propose an efficient algorithm for deriving security view definitions from security policies (defined on the original document DTD) for different user groups. We also develop novel algorithms for XPath query rewriting and optimization such that queries over security views can be efficiently answered without materializing the views. Our algorithms transform a query over a security view to an equivalent query over the original document, and effectively prune query nodes by exploiting the structural properties of the document DTD in conjunction with approximate XPath containment tests. Our work is the first to study a flexible, DTD-based access-control model for XML and its implications on the XML query-execution engine. Furthermore, it is among the first efforts for query rewriting and optimization in the presence of general DTDs for a rich class of XPath queries. An empirical study based on real-life DTDs verifies the effectiveness of our approach.

[11] Graph pattern matching is typically defined in terms of subgraph isomorphism, which makes it an np-complete problem. Moreover, it requires bijective functions, which are often too restrictive to characterize patterns in emerging applications. They proposed a class of graph patterns, in which an edge denotes the connectivity in a data graph within a predefined number of hops. In addition, we define matching based on a notion of bounded simulation, an extension of graph simulation. We show that with this revision, graph pattern matching can be performed in cubic-time, by providing such an algorithm. We also develop algorithms for incrementally finding matches when data graphs are updated, with performance guarantees for dag patterns. We experimentally verify that these algorithms scale well, and that the revised notion of graph pattern matching allows us to identify communities commonly found in real-world networks.

[12]They presented algorithms for computing similarity relations of labeled graphs. Similarity relations have applications for the refinement and verification of reactive systems. For finite graphs, we present an O(mn) algorithm for computing the similarity relation of a graph with n vertices and m edges (assuming m n). For effectively presented infinite graphs, we present a symbolic similarity-checking procedure that terminates if a finite similarity relation exists. We show that 2D rectangular automata, which model discrete reactive systems with continuous environments, define effectively presented infinite graphs with finite similarity relations. It follows that the refinement problem and the 8CTL model-checking problem are decidable for 2D rectangular automata.

Describing social positions and roles is an important topic within social network analysis. One approach is to compute a suitable equivalence relation on the nodes of the target network [13]. One relation that is often used for this purpose is regular equivalence, or bisimulation, as it is known within the field of computer science. In this paper we consider a relation from computer science called simulation relation. Simulation creates a partial order on the set of actors in a network and we can use this order to identify actors that have characteristic properties. The simulation relation can also be used to compute simulation equivalence which is a less restrictive equivalence relation than regular equivalence but is still computable in polynomial time. This paper primarily considers weighted directed networks and we present definitions of both weighted simulation equivalence and weighted regular equivalence. Weighted networks can be used to model a number of network domains, including information flow, trust propagation, and communication channels. Many of these domains have applications within homeland security and in the military, where one wants to survey and elicit key roles within an organization. Identifying social positions can be difficult when the target organization lacks a formal structure or is partially hidden.

III. METHODLOGY

Determining Pattern Containment

To do this, we first propose a sufficient and necessary condition to characterize pattern containment. We then develop a cubic time algorithm based on the characterization. Sufficient and necessary condition. To characterize pattern containment, we introduce a notion of view matches. Consider a pattern query Qs and a set V of view definitions. For each V 2 V, let VðQsÞ ¼ fðeV; SeVÞ j eV 2 Vg, by treating Qs as a data graph. Obviously, if V E Qs, then SeV is the nonempty match set of eV for each edge eV. We define the view match from V to Qs, denoted by MQsV , to be the union of SeV for all eV in V.

Minimal Containment Problem:

Given a pattern query Qs and a set V of view definitions,it returns either a nonempty subset V0 of V that minimally contains Qs, or ; to indicate that Qs 6v V.

Algorithm minimal initializes (1) an empty set V0 for selected views, (2) an empty set S for view matches of V0,and (3) an empty set E for edges in view matches. It also maintains an index M that maps each edge e in Qs to a set of views. Similar to algorithm contain, minimal first computes MQsVi for all Vi 2 V. In contrast to contain that simply merges the view matches, it extends S with a new view match MQsVi only if MQsVi contains a new edge not in E, and updates M accordingly. The for loop stops as soon as E ¼ Ep, as Qs is already contained in V0. If E 6¼ Ep after the loop, it returns, since Qs is not contained is V . The algorithm then eliminates redundant views, by checking whether the removal of Vj causes MðeÞ¼ ; for some e2 MQsVj. If no such e exists, it removes Vj from V0. After all view matches are checked, minimalreturns V0.

The algorithm is denoted as minimum. Given a pattern Qs and a set V of views,

minimum identifies a subset V0 of V such that (1) Qs v V0 ifQs v V and (2) cardðV0Þ _ log ðjEpjÞ _ cardðVOPTÞ, whereVOPT is a minimum subset of V that contains Qs. That is, the approximation ratio of minimum is OðlogjEpjÞ, where jEpj is typically small. The algorithm iteratively finds the “top” view whose

view match can cover most edges in Qs that are not yet covered.To do this, we define a metric aðVÞ for a view V, whereaðVÞ ¼ jMQsV n EcjjEpj:Here Ec is the set of edges in Ep that have been covered byselected view matches, and aðV Þ indicates the amount ofuncovered edges that MQsV covers. We select V with the largesta in each iteration, and maintain a accordingly. Similar to minimal, algorithm minimum computes the view match MQsVi for each Vi 2 V, and collects them in a set S. It then does the following. (1) It selects view Viwith the largest a, and removes MQsVifrom S. (2) Itmerges Ec with MQsViif MQsVicontains some edges that arenot in Ec, and extends V0 with Vi. During theloop, if Ec equals Ep, the set V0 is returned. Otherwise,minimum returns indicating that Qs 6v V.

Maximally Contained Rewriting

A pattern query Qs0 is a subquery of Qs, denoted as Qs0 _ Qs, if it is an edge induced subgraph of Qs, i.e., Qs0 is a subgraph of Qs consisting of a subset of edges of Qs, together with their endpoints as the set of nodes. Query Qs0 is called a contained rewriting of Qs using a set V of view definitions if_ Qs0 _ Qs, i.e., Qs0 is a subquery of Qs, and_ Qs0 v V, i.e., Qs0 can be answered using V. Such a rewriting Qs0 is a maximally contained rewriting ofQs using V if there exists no contained rewriting Qs00 such that Qs0 Qs00, i.e., there exists no larger contained rewritingQs00 with more edges than Qs0.Query-driven approximation scheme. When Qs is not contained in V, we can still efficiently answer Qs in a (possibly

big) graph G following two approaches. (1) One may firs tidentify a maximally contained rewriting Qs0 of Qs using V,and then compute Qs0ðGÞ as approximate answers to Qs, bysimply invoking the algorithm MatchJoin. (2) Alternatively, one may compute exact answers QsðGÞ by using Qs0ðGÞ and by accessing a small fraction GQs of G, such thatQsðGÞ ¼ Qs0ðGÞ [fðGQs Þ. Here fðGQs Þ first locates thematches of Qs0ðGÞ in the original graph G and then verifies the matches for Qs by visiting neighborhood of those matches, a small number of nodes and edges in G that constitute GQs ; this is the approach suggested, referred

to as scale-independent query answering using views there. Due to the space constraint, we focus on approximate answersQs0ðGÞ in this paper. That is, when limited views are available, we can still approximately answer pattern queries in

big graphs by relaxing Qs to maximally contained rewritingQs0, using those views. Accuracy. Given a graph G, we measure the quality of the approximate answers Qs0ðGÞ versus the true matches in the exact answers QsðGÞby following the F-measure:

$$Acc = 2 \cdot \delta recall \cdot precision \div \delta recall + precision;$$

where recall = #true matches found #true matches , and precision =#true matches found # matches . Here #matches is the number of all (edge)matches found by Qs0ðGÞ using views, #true matches is the number of all matches in QsðGÞ; and #true matches found is the number of all the true matches in both Qs0ðGÞ and QsðGÞ. Intuitively, a high precision means that many matches inQs0ðGÞ are true matches, and a high recall means Qs0ðGÞcontains most of the true matches in QsðGÞ. The larger Acc that can be induced by Qs0, the better. If Qs0 is equivalent toQs, i.e., Qs0ðGÞ = QsðGÞ for all G, Acc takes the maximum value 1:0. Observe that for any edge e in Qs, if e is covered by Qs0, then for any G, the match set Se of e in QsðGÞ is asubset of the match set S0eof e in Qs0ðGÞ; that is, Qs0ðGÞ finds all candidate matches of e in G.

IV. RESULTS AND DISCUSSIONS

The criteria for comparing the methods of XML queries and graph pattern queries

- **Accuracy** − Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

- **Speed** − This refers to the computational cost in generating and using the classifier or predictor.

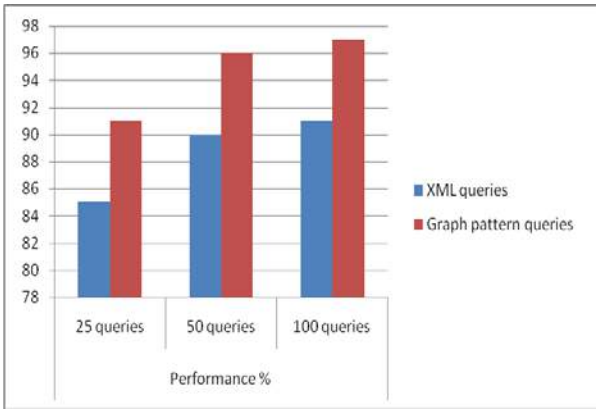
- **Robustness** − It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

- **Scalability** − Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

Interpretability − It refers to what extent the classifier or predictor understands

The performance of the both the queries are analysed.

Method	Performance %		
	25 queries	50 queries	100 queries
XML queries	85	90	91
Graph pattern queries	91	96	97



V. CONCLUSION

We have proposed a notion of pattern containment to characterize what pattern queries can be answered using views, and provided such an efficient matching algorithm. We have also identified three fundamental problems for pattern containment, established their complexity, and developed effective (approximation)algorithms. When a pattern query is not contained in available views, we have developed efficient algorithms for computing maximally contained rewriting using views to get approximate answers. Our experimental results have verified the efficiency and effectiveness of our techniques. These results extend the study of query answering using views from relational and XML queries to graph pattern queries. Finally, to find a practical method to query "big" social data, one needs to combine techniques such as view-based, distributed, incremental, and compression methods. The efficiency and effectiveness of the technique are verified through experimental results. The study of query answering is extended from These results extend the study of query relational to graph pattern queries.

VI. REFERENCES

[1] A. Y. Halevy, "Answering queries using views: A survey," VLDBJ., vol. 10, no. 4, pp. 270–294, 2001.

[2] M. Lenzerini, "Data integration: A theoretical perspective," inProc. 21st ACM SIGMOD-SIGACT-SIGART Symp. Principles DatabaseSyst., 2002, pp. 233–246.

[3] L. V. S. Lakshmanan, W. H. Wang, and Z. J. Zhao, "Answeringtree pattern queries using views," in Proc. 32nd Int. Conf. VeryLarge Data Bases, 2006, pp. 571–582.

[4] J. Wang, J. Li, and J. X. Yu, "Answering tree pattern queries usingviews: A revisit," in Proc. 14th Int. Conf. Extending Database Technol.,2011, pp. 153–164.

[5] X. Wu, D. Theodoratos, and W. H. Wang, "Answering XMLqueries using materialized views revisited," in Proc. 18th ACMConf. Inf. Knowl. Manag., 2009, pp. 475–484.

[6] D. Calvanese, G. D. Giacomo, M. Lenzerini, and M. Y. Vardi,"View-based query processing and constraint satisfaction,"in Proc. 15th Annu. IEEE Symp. Logic Comput. Sci., 2000, p. 361.

[7] Y. Papakonstantinou and V. Vassalos, "Query rewriting for semistructureddata," in Proc. ACM SIGMOD Int. Conf. Manag. Data,1999, pp. 455–466.

[8] Y. Zhuge and H. Garcia-Molina, "Graph structured views andtheir incremental maintenance," in Proc. 14th Int. Conf. Data Eng.,1998, pp. 116–125.

[9] M. Armbrust, E. Liang, T. Kraska, A. Fox, M. J. Franklin, and D. A.Patterson, "Generalized scale independence through incrementalprecomputation," in Proc. ACM SIGMOD Int. Conf. Manag. Data,2013, pp. 625–636.

[10] W. Fan, F. Geerts, and L. Libkin, "On scale independence for queryingbig data," in Proc. 33rd ACM SIGMOD-SIGACT-SIGARTSymp. Principles Database Syst., 2014, pp. 51–62.

[11] C.-M. Chen and N. Roussopoulos, "The implementation and performanceevaluation of the ADMS query optimizer: Integratingquery result caching and matching," in Proc. 4th Int. Conf. ExtendingDatabase Technol.: Adv. Database Technol., 1994, pp. 323–336.

[12] W. Fan, C. Y. Chan, and M. N. Garofalakis, "Secure XML queryingwith security views," in Proc. ACM SIGMOD Int. Conf. Manag.Data, 2004, pp. 587–598.

[13] W. Fan, J. Li, S. Ma, N. Tang, Y. Wu, and Y. Wu, "Graph patternmatching: From intractable to polynomial time," Proc. VLDBEndowment, vol. 3, no. 1, pp. 264–265, 2010.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

CREDIT CARD FRAUD RECOGNITION USING DATA MINING TECHNIQUES

R.Akila, MCA, M.Phil.
Assistant Professor in Computer Science
Guru Nanak College
Velachery- Chennai, India
akila.akilarchp@gmail.com

M.Bhuvaneswari, ME.
Assistant Professor in Computer Science
Guru Nanak College
Velachery- Chennai, India
me.bhuvaneswari@gmail.com

ABSTRACT

In recent world most of the people, small and large scale originations are moving their daily business activity to online and providing customers services via internet. Credit Card (CC) payment is playing major role in the business activity but same time CC fraud is one of the major concern and issues in online transaction. In recent year CC fraud frauds are increased in day to day activity. The Main reason is most of the customers are using CC for all kind of payment. So the aim of this paper is to identify the different types of CC frauds and review the alternative techniques to detect the CC frauds. So satisfying the customers all originations are moving to secured transaction for customer to make payment for purchasing goods. This study will help to understand the CC fraud and type of methodology can be used to detect the CC fraud. This study will help to understand the CC fraud and type of methodology used to detect the CC fraud.

1. INTRODUCTION

The internet becomes most popular mode of payment for online transaction. Banking system provides e-cash, e commerce and e-services improving for online transaction. This payment facilitates the acceptance of electronic payment for online transactions also known as a sample of Electronic Data Interchange (EDI), e-commerce payment systems have become increasingly popular due to the widespread use of the internet-based shopping. Now a day tremendous volume and value increase in credit card transactions and same time credit card frauds also increasing day by day.

1.1CREDIT CARD FRAUD IN BANKING

The CC is one of the most conventional ways of online transaction. It allows cardholders to purchase goods and services from the shopping websites or from the market. In case of risk of fraud transaction using CC has also been increasing. CC fraud detection is one of the ethical issues in the credit card companies, banks and financial Institutes. CC fraud system to find the fraud and remove duplicate from CC fraud application. Fraud can be identified two ways in banking sector. First validate exact match between duplicate data with fraud data base and compare duplicate data with fraud data base approximately with slightly altered spellings. This paper discuss with each successful fraud pattern to find the fraud with short time period.

2. RECOMMENDED SYSTEM

Fraud detection is one of the main and important goalsof this paper. Generally security based layer is proposed system for fraud detection in data mining. In security based layers CD and SD techniques are mainly used to find the fraud detection in real time.

1) Communal Detection (CD)
2) Spike Detection (SD)

CD technique is fixed set of attribute to find the fraud. It will compare with default list and match attribute with exact value. SD technique is same as CD but it will not match with exact value and compare with variable attributes.

2.1.Calculate CD Score

This is the method most of the places are using to detect the fraud in basic level. If any new application submitted from users or customer first it is taken as ainput to CD layer. CD layer compared with white list data provide the suspicious score, based on score it is decided as fraud or not. Basically CD layer used to compare with common relationship between new application and default list. If five or more attributes are matched the CD assign less number of suspicious score. If CD gives fewer score then it is considered as legal transaction and new application details added to white list. Suppose it gives more suspicious score then this transaction may be fraud. Even though if give less score new application data can be passed as a input to SD layer.

2.2 Calculate SD Score

This is another approach to find the fraud detection. SD layer taken input from CD layer output. SD algorithm have different step to find the suspicious score. Single step to find the scaled count value when compare with new application data.If single value similarity and time difference exceed then it is considered as fraud. Another step calculates current value score based on calculated weighted score. For example match with any one of the unique id, if it is matched then it gives more suspicious score and declare as fraud and reject the new application else added in the white list.

3. ARCHITECTURE DESIGN AND OVERVIEW

The architecture diagram represents complete flow of fraud detection system. Here first collected all input data from user or customers and all input data passed to fraud detection section to find the fraud based on suspicious score. Then different type algorithms and techniques such as case base reasoning, retrieval data methods, and diagnose the data. Finally find the data as fraud or not based on above techniques. If input data detected as fraud then details are stored to blocked list else if data is legal relationship then it is stored in original database and it is considered as genuine transaction.

3.1 Fraud Detection

Fraud detection system has used two different algorithms such as CD and SD to find the fraud. CD algorithms is purely relationship oriented and applied attributes to find the suspicious score. CD techniques called as adaptive based approach. Before

proceed the SD it is required to reinforce that CD and find real relationship to reduce the score. SD is not white list oriented approach it is attributed oriented process.

3.2 Fraud Verification

All new user details are taken to this section to identify the fraud, in the process applied two algorithms (CD & SD)and verified input data to make efficient credit card transaction. If the data is original it is proceed to further transaction else it is rejected the transaction and its added to black list. Next time easily find the fraud in first level without applied the CD & SD algorithm.

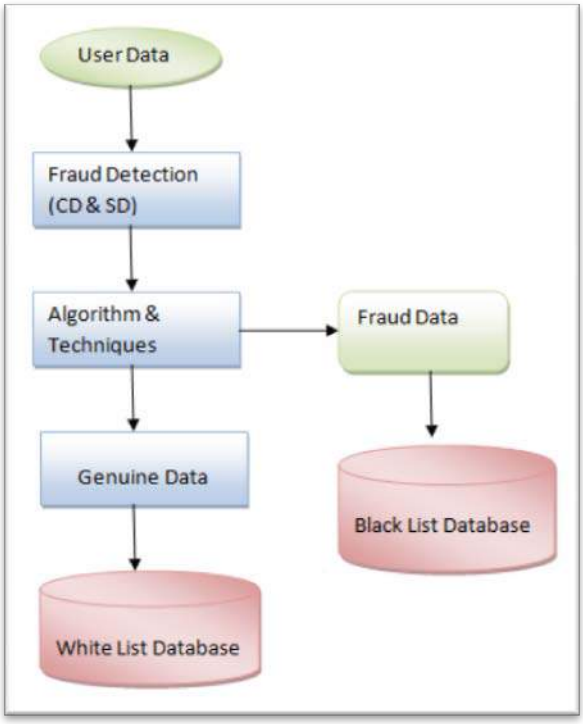


Fig1:System Architecture Diagram

4. BIOMETRICS PROCESS

Now a day's biometrics techniques is playing major role in security system. Biometrics based security is really challenged to fraudster.In the biometrics finger print authentication is one the way to stop the fraud. In this process electronic finger print scanners are scanned the finger print and stored in digital format. Digital format pictures are then processed in to digital template with unique value. Using this system only authorizes users or customers can make transaction. All the scanned digital templates are stored in biometric database.

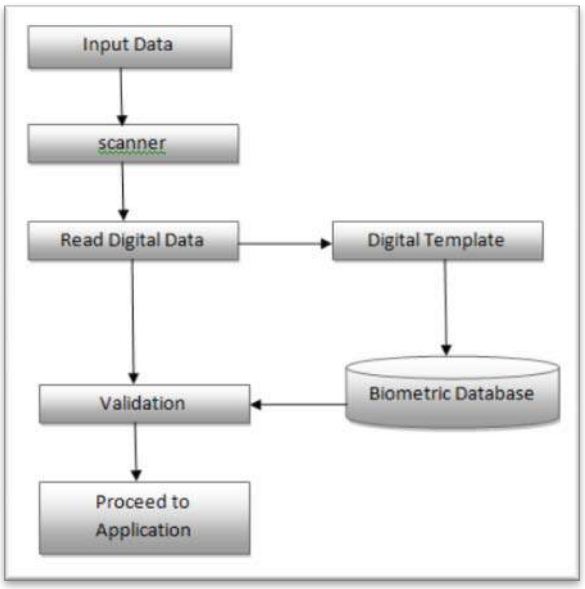


Fig. 2: Example for Biometrics Pattern Comparison and Retrieval

5. CONCLUSION AND FUTURE WORK

Data mining is a recognized platform fordefining rules, analyzing and predicting the data from large amount of data base.The aim of this paper is Identity the frauds in online transaction. This system detects the fraud detection in online system and it is used to avoid theduplicates application entry while applying new credit card.The CD and SD algorithm used to detect the fraud from multipleapplicants. In this proposed system combine with theexisting algorithm SD and CD and this system is well organized with more efficient and secure. Fraud verification layer is used to throw the fraudulent activity immediately because all black list data stored in the data base. Limited time period is required to identify the fraud using this proposed system. This technique can apply different industry to find the fraud activity.

6. REFERENCES

[1] ALKA HERENJ,SUSHMITA MISHRA, Secure Mechanism for credit card transaction fraud detection system issue 2,February 2013
[2] Richard J.Boltan and David J.Hand, "Statistical Fraud Detection", pp.1-54, 2002.
[3] ID Analytics, "ID Score-Risk: Gain Greater Visibility
[4] V.Dhecpa and Dr. RDhanapal, Analysis of credit-card frauddetection methods', International Journal of Recent Trendsin Engineering (2009), vol, 2. No. 3, pp.126-128.
[5] ID Analytics,ID Score-Risk: Gain Greater Visibility into Individual Identity Risk,"Unpublished, 2008.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

STUTTERING: Therapy for Kids Android Application

Keerthana B
Dept of Information Technology
4th Year, Sri Sairam Engineering College
Chennai, India
keerthanabalajiiyer@gmail.com

V.Vidhya
Asst Professor, Dept of Computer Applications
SSS Jain College for Women
Chennai, India
vidhya2982@gmail.com

ABSTRACT

Repetition of same word affects the flow of speech and it is termed as Stuttering. Stuttering makes the kid who suffers, nervous to face people and they will not be bold to talk aloud. The disorder of speech for a kid is a strong issue that has to be handled earlier. This paper, "Therapy for kids using an Application" is the proposed system of supporting the method of curing stuttering. Android App is a software application that runs on the android mobile. Android mobile is available with almost everyone, thus it makes it easier to assist through an app and help the doctors in monitoring the treatment of the kid.

Keywords: Stuttering, Android app, therapy, genetics, monitoring, voice recognition, record.

INTRODUCTION

[1] Stuttering is a speech disorder in which sounds, syllables or words are repeated disrupting the flow of speech. Development of stutter may be due to the genetics or the delay in speech or language problems. A long period treatment helps in curing the stuttering issues. Prolongation of a word also comes under stuttering. Affected count can be defined as 3:4 ratio of as many males as females. For kids, even the nervousness of something may cause stuttering. For example, reading a paragraph aloud in a class may lead a nervous child to stutter. Speech therapy is a method that brings out a huge difference. This application provides an easy way to help kids suffering from stuttering issues. It records the progress of the kid and the main factor of this is the voice recognition that supports in knowing the impact from the therapy.

STUTTERING

A. *Stuttering Issues*

Stuttering had a serious effect on the thoughts of a kid to face the world. [4]Several tests till date proves that most of the risk for stuttering onset is over by age five. [1] Many case-studies have shown that childhood speech impairments such as stuttering are mainly associated with lower test scores. Types of stuttering include blocks, prolongation, insertion, repetitions. M-mm is a kind of prolongation. Although it is less common around 1-5% it should also be treated. Treating with some regular exercises and therapies that does not bring discomfort is required.

GRAPH 1. Assessment of the Total Impact of Stuttering

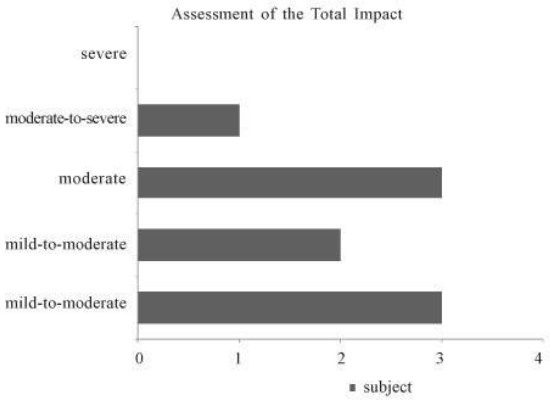


Fig1: Indicates a test that shows the impact of stuttering

B. *Stages*

At the initial stage of stuttering by kids when they try to join words to form a sentence is acceptable. Later if the same case continues in a frequent manner and leads to a worst case he/she should be evaluated by a speech-therapist in order to avoid further issues. Stuttering drops down when kids enter their school and start sharpening their skills.
Thus it may also result in a state of preventing stuttering.

EXISTING APPLICATIONS

Many applications were developed earlier that supports the speech therapy, but only few were successful. While creating an account, and log in to use the exercises were considered to be a huge development in the android app facilities. [5] Case studies defines that around 5% of children go through a period of stuttering that may last six-twelve months. Some cases will include few among many of those who begin to stutter might recover by late childhood. But few may have it as a long term problem. Apps were developed to aid parents to help their kid to overcome stuttering. The apps included an account creation for the kid to monitor and record his/her progress. The apps were seemed to be user friendly and also free of cost for utilization. Research activities have found that fluency disorders are not as popular as others. [3] Test cases were conducted which had speech, language tests with the obvious exception of stuttering for Children who stutter and Children's temperamental characteristics were determined using the Behavioural Style Questionnaire given.

PROPOSED APPLICATION

To Log in and to use the app is not a wonder. Latest technology has much more to support all issues with a solution of its kind. Regular practice improves the fluency. Basically, BUT-CAN-LIKE are few words that make a kid to stutter. It might be due to the pronunciation of the word or the formation of sentence in accordance with the tense. The correct pronunciation of a word may come out after a hard struggle of say half the word or a letter at most. [2] Exploring the stuttering treatment to an extent is the other way to assess, to which they address the cause of the disorder.

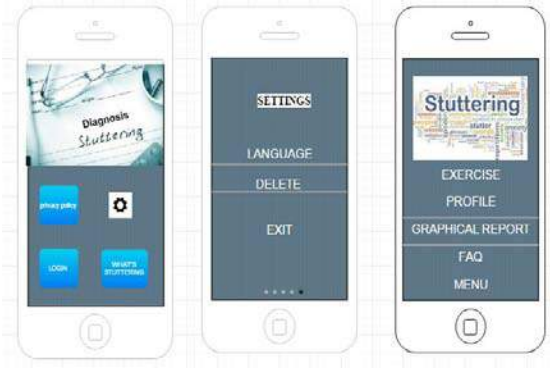


Fig2: A model representation

The application should be able to attain a child's good progress if and only if it is creative and attractive. The above figure 2 shows the model of an app that has an initial page followed by settings and also user's page after log in.

A. Application Features

In general, the app will consist of a login for user. Other than that, if an user is new to android and does not know how to access, there will be an option called guide that will help the user. It will be available for the user even if he didn't log in. But the only difference will be that it will not have the full virtual tour of the app if the user didn't login. Fig3 shows the flow chart that represents the way of giving exercise and evaluating a kid's progress. The system proposed is further more advanced with verbal as well as voice recognisable ways to give a successful progress that will help their respective doctor as well as parents to understand their problem. It will also help the kid with their pronunciation of words and aid them. When repetition of words, phrases becomes excessive, when the kid has fear of stutters and changes the phrase or when there is an increase in prolongation of words then it may result in consulting a therapist. For further practices advised by the therapist to work out at home, this app will be an essential tool. The proposed system will have word, phrases included and it will also be helpful in taking the exercises and practice sessions again and again. In general, the frustrations developed because of stuttering can be overcome with the help of this application. This system will ensure that kids will be interested to learn with this app as it will have attractive visuals that will not get them bored.

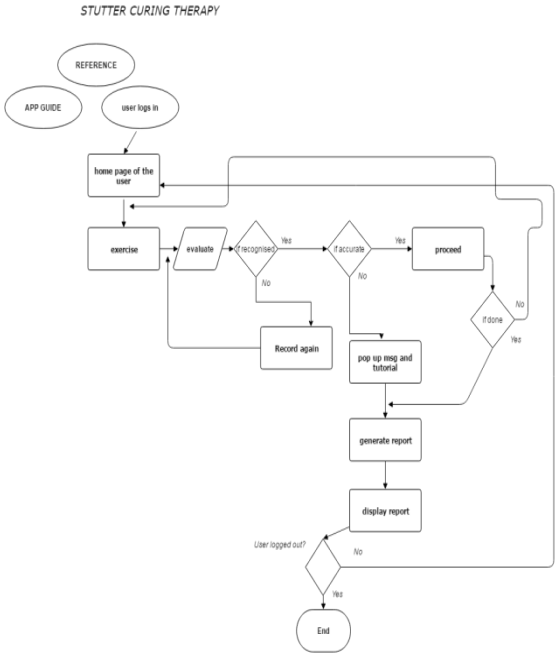


Fig3: Exercise and report generation flowchart

ACKNOWLEDGMENT

With the team work, the idea of proposing a flexible aid for kids who stutter is on progress. Inspired from various attempts of support to cure stuttering and would like to provide a better application to work on this speech aid or therapy.

FUTURE WORK

The scope for the future work is that this application will be user friendly with various new possible ideas such as live monitoring and it will also satisfy the requirements of the therapy system that can be implemented. The app that is to be developed will initially support English language which may have few additions in the next upgradable versions. It will also have a possibility of recording the voice and storing it for further references. It will also have an audio feedback system. With reference from professionals and therapists this app will be an aid that a therapist can recommend his/her patient.

REFERENCES

[1]" The Kid's Speech: The Effect of Stuttering on Human Capital Acquisition", by Joseph J. Sabia Daniel I. Rees, , June 2011 IZA in the paper Discussion . 5781.

[2] Ann Packman," Theory and therapy in stuttering: A complex relationship", Volume 37, Issue 4, December 2012, Journal of Fluency Disorders.

[3] "emperamental Characteristics of Young Children Who Stutter " by Julie D. Anderson, Mark W. Pellowski, Edward G. Conture, and Ellen M. Kelly", *Journal of Speech, Language, and Hearing Research*, October 2003, Vol. 46, 1221-1233. doi:10.1044/1092-4388(2003/095)

[4] Ehud Yairi, Nicoline Ambrose," Epidemiology of stuttering: 21st century advances", Journal of Fluency Disorders Volume 38, Issue 2, June 2013

[5] The Stuttering Foundation Since 1947

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A Study on Efficient Classification Model for Breast Cancer Prediction Based on Feature Selection Techniques

B. Tamilvanan
Research and Development Centre
Bharathiar University,
Coimbatore-641046, TN, India,

Dr. V. MuraliBhaskaran
Principal,
Dhirajlal Gandhi College of Technology
Salem-636290, TN, India.

ABSTRACT

Classification algorithms are efficiently utilized in the area of general medical diagnosis applications in order to identify the disorders in advance. One such disease, breast cancer is the most prevalent and earnest quandary with women in most of the developing countries. Many attempts are made in order to identify this problem with the objective of high precision and better accuracy. In this paper, an attempt is made with the most popular and efficient classification algorithms namely Naive Bayes, Best First Search(BFS), Random Search (RNS) and Genetic Search (GNS) to amend the efficiency of the detection, accuracy for the breast cancer dataset. As an objective of improving accuracy, an efficient dimensionality reduction technique is incorporated in this work. The performances of these approaches are evaluated using the metrics such as the precision, recall, f-measure and accuracy. From these measures it is clearly observed that Naive Bayes with Best First Search algorithm is able to achieve high accuracy rate along with minimum error rate when compared to other algorithms. The review can be stretched out to draw the execution of other characterization systems on an extended information set with more particular ascribes to get more exact outcomes.

Keywords: Classification, Feature selection, Naive Bayes, Best first search, Random search and Genetics search.

INTRODUCTION

Data mining systems and programming are used in a substantial differ of fields, together with securities exchange, ERP, media transmission, endeavor enterprises, , climate, social insurance and sizably voluminous information [1] [2]. These days wellness mind industry creates a tremendous measure of data about patients, disease conclusion, and so on. Some exceptional sorts of procedures to developing right groupings have been proposed (e.g., NB,BFS-NB, GNS-NB, RNS-NB). In characterization, we give a Breast Cancer informational index of case report or the info information, called the check informational index, with each archive comprising of different attribute.

An attribute can be both a numerical attribute and categorical attribute. If values of an attributes belong to an authoritatively mandated domain, the attribute is referred to as numerical attribute(e.g. Tumor-size, Deg-Malig, Menopause, Age, Inv-nodes). A categorical attribute (e.g. Irradiant, Breast, Node-cape, Breast-Quad, Class).Classification is the process of splitting a dataset into mutually exclusive groups, called a class, based on suitable attributes.

In this world, individual sorts of Breast Cancer maladies are a typical type of disease influencing all ladies of various ages. Bosom disease influences the bosom tissue and lobules. The classification of breast cancer is resulted from its beginning, if breast cancer is originate from milk ducts then it is known as ductal carcinoma while cancer cells found in lobules makes cancer termed as “lobular carcinoma.” The viewing of bosom malignancy is an essential stride which sifts through the manifestations that can be utilized to analyze the patient's real obsessive condition. Breast cancer is the most continuous reason for death in more established ladies however in the meantime, it is critical to note that more youthful ladies who don't go under tumor screening process stay in risk hover of breast cancer.

In this paper is designed accordingly: the relates works and show of the focused parts of the utilized data mining methods in section 1. The information of the dataset for Breast Cancer in section 2.The experimentation outcome and conversation in section 3. And finally, conclude the paper and future enhancements.

LITERATURE REVIEW

A multinomial logistic-regression model with a hill-like estimator generalizes logistic regression by using more than two distinct outcomes between the categorical and multinomial distributions [3].This model is mainly designed to predict the probabilities of different outcomes when using categorically dependent and independent variables.

Best First Search Algorithm:

The Best First Search is an important AI search strategy that allows back tracking along the search path. Like the best first search moves through the search space by making local changes to the current feature subset. However, unlike hill climbing method, suppose path being explored begins to look less promising, the best first search method can back-track to a more promising previous subset and continue the search from there. A best first search will explore the entire search space for specified time, so it is common to use a stopping criterion. Normally this involves limiting the number of the fullyexpanded subset and that results in no improvement [4][5].

Genetic Search Algorithm:

Search techniques traverse the attribute space to locate a decent subset and the quality is estimated by the property subset evaluator through CFS subset evaluator and hereditary pursuit is being utilized as a search techniques. The parameters of the genetic algorithm area number of generations, population size and the probabilities of mutation and crossover. A member of the initial population generates by specifying a list of attribute indices as a search point. For generating progress reports, every so many generation can be used [6][7].

PROPOSED METHOD

proposed BFSCFS-NB algorithm:

Step 1: To start with an OPEN list containing the start state, the CLOSED listempty and BEST← start state.

Step 2: Let assign s = arg max e(x) (get the state from OPEN with the highestevaluation).

Step 3: Eliminate s from OPEN and add to CLOSED.

Step 4: If e(s) ≥ e(BEST), then BEST← s.

Step 5: For every child t of s that is not in the OPEN or CLOSED list, evaluateand add to OPEN.

Step 6: If BEST changed in the last set of expansions, go to 2

Step 7: Return BEST.

Step 8: Obtain the new data set.

Step 9: Construct both training and test data discrete.

Step 10: Estimate the prior probabilities P(Cj), j=1,... k from the training data,where k is the number of classes.

Step 11: Estimate the conditional probabilities P(Ai= aℓ | Cj), i= 1,....,D,j=1,....,k, ℓ= 1,....,d from the training data, where D is the number offeatures, d is the number of discretization level.

Step12: Estimate the posterior probabilities P(Cj | A) for each test example xrepresented by a feature vector A.

Step 13: Assign x to the class C* such that C*=arg max j=1,2 P(Cj | A).

The first half of the algorithm from step one to eight is used to select thesubset using Best First Search and then the second half of the algorithm from nineto thirteen are for classification using Naive Bayes.

BREAST CANCER DATASET

The performance of these classification algorithms namely Naive Bayes, Best First Search, Genetics Search and Random Search was tested in a medical database for Breast Cancer Disease dataset from UCI machine learning repository (available at http://archive.ics.uci.edu/ml/datasets/Breast+Cancer [8]. The data set has ten features of the attributes. Table- 1 describes the data for Breast Cancer. The medical dataset contains data from reviews conducted among patients, each of which has ten features. All features can be considered as on indicators of Breast Cancer disease for a patient. The dataset holds records of the following attributes.

Table 1: UCI Dataset of Breast Cancer

Attributes Name	Attribute Type	Description
Age	Numeric	Age (years)
Inv-Nodes	Numeric	0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39
Node-Caps	Discrete	yes, no.
Menopause	Numeric	lt40, ge40, premeno
Deg-Malig	Numeric	1, 2, 3.
Tumor-Size	Numeric	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59
Breast	Discrete	left, right
Breast-Quad	Discrete	left-up, left-low, right-up, right-low, central.
Irradiat	Discrete	yes, no.
Class	Discrete	no-recurrence-events, recurrence-events

CONFUSION MATRIX

The confusion matrix shows how many occurrences have been given to each class and the elements of the matrix illustrate the number of test examples whose concrete class is the row and whose predicted class is the column. Tables 7, 8 and 8 illustrate the confusion matrix that is calculated for the Best First Search based NB, Genetic Search based NB and Random Search based NB algorithms.

Table 5 Different outcome of two class prediction

Actual Class	Predicated Class		
		a	b
	a	Ture Positive	False Negative
	b	False Positive	True Negative

Precision

It is utilized to speak to the portion of recovered information from associating datasets, which pertain to the search. Precision will be used to represent how many instance have been correctly classified in the confusion matrix table

(correct classified data is true positive and incorrect classified data is error positive).

$$\text{Precision} = \frac{tpA}{tpA + eBA}$$

Where tpA is represented as true positive for the class A and eBA are represented as false positive.

Recall
It is utilized to speak to the portion of recovered information from associating datasets; that are important to the inquiry that is successful. It is used to find out the ratio between the true positive and both true positive and false positive values.

$$\text{Recall} = \frac{tpA}{tpA + eAB}$$

Where tpA is represented as true positive for the class A and eAB are represented as error positive.

F-measure This is evaluated by the harmonic mean between precision and recall.

$$\text{F-Measure} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Accuracy This is calculated as the proportion of true positive, true negatives and true results from all the given data.

$$\text{Accuracy} = \frac{tpA + tpB}{tpA + eAB + eBA + tpB}$$

EXPERIMENT RESULTS AND DISCUSSION

In this section, we explain the test database and investigational analysis and the current evaluation results for four algorithms namely NB, BFS-NB, GNS-NB, RNS-NB classifier.

In this experimental analysis, NB, BFS-NBwith 5 potential attributes namely (Tumor-Size, Inv-Nodes, Node-Caps, Breast-Quad, Irradiat) GNS-NB5 potentialattributes(Tumor-Size, Menopause, Deg-Malig, Node-Caps, Irradiat), RNS-NBwith 6 potential attributes namely (Tumor-Size,Menopause, Inv-Nodes, Node-Caps,Deg-Malig, Irradiat)Algorithms performance were compared based on their application in medical datasets. Weka tool is is utilized for research area, share markets, bankingsector, education institute and climate datasets. It helps in composed exercises in machine learning, data mining, and text mining. It supports all the mining process to get a valid and clear visualization of accurate results. ten-fold cross-validation with feature selection attributes were to the input datasets in the experiments

Experimental Step Up

A brief description of the classification process by all algorithms, NB, BFS-NB, GNS-NB, RNS-NBare given below:

Table 10: Performance analysis related to accuracy

Method	Accuracy (%)
NB	72
GNS-NB	73
RNS- NB	77
BFS - NB	82

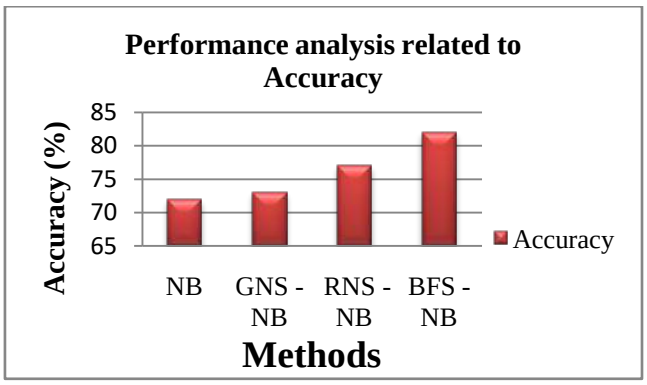


Fig 1 : Performance analysis related to accuracy

CONCLUSION

In this work popular classification algorithms along with feature selection method are used to calculate the breast cancer detection process more efficiently. The efficient classification algorithms namely NB, GNS-NB, RNS-NB, BFS-NB are used to develop the model and all are evaluated 10 fold cross-validation. The dimensionality reduction technique is able to select more efficient and relevant features from the ten original features and also observed that results obtained using relevant features are better than or equal to the results obtained using ten features with less effort. These classification algorithms are compared, and accuracy is evaluated for true positive and false positive rate. From the experiments, it is observed that Naive Bayes using Best First Search classification algorithm performs compare than other classification algorithms with 82% accuracy for both after feature selection and before feature selection using ten-fold cross validations.

REFERENCES

[1] B.Tamilvanan and Dr. V. MuraliBhaskaran, “A New Feature Selection Techniques Using Genetics Search and Random Search Approaches For Breast Cancer”, Biosciences and Biotechnology Research Asia, , vol. 14, no.1, pp. 409-414, March 2017.
[2] Sitar-Taut, V.A., et al, Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
[3] You-Shyang Chen, Modeling hybrid rough set-based classification procedures to identify hemodialysis adequacy for end-stage renal disease patients, Computers in Biology and Medicine, 2013, vol. 43, pp. 1590–1605.
[4] Hall, Mark A. and Lloyd A.Smith., "Feature subset selection: a correlation based filter approach", 1997.
[5] Hall. M, “Correlation based feature selection for machine learning”,Doctoral dissertation,University of Waikato, Dept. of Computer Science, 1999.
[6] Pallabi Borah et al.,“A statistical feature selection technique”, Netw Model Anal Health Inform Bioinforma, pp-3-55, 2014.
[7] ShashikantGhumbre, Chetan Patil and Ashok Ghatol, “Heart Disease Diagnosis using Support Vector Machine”, International Conference onComputer Science and Information Technology (ICCSIT'2011) , pp. 84-88,December 2011.
[8] UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets/Breast+Cancer

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Role of Information technology - A study on the Customer's Perspective towards Alternative Banking Operations

Dr. B.Dhanalakshmi
Asst. Professor, Dept. of Commerce(Honors),
SSS Shasun jain College for women,
T.Nagar,Chennai. Tamilnadu,India
dhanavkarthik@gmail.com

Dr. C.K.Deepa
Asst. Professor,
PG & Research Dept. of Commerce
SSS Shasun jain College for women,
T.Nagar,Chennai. Tamilnadu,India
ckdeepadhivya@gmail.com

Dr.Sambamurthy Padmavathi
Dean & Associate Professor
PG & Research Dept. of Commerce
SSS Shasun jain College for women,
T.Nagar,Chennai. Tamilnadu,India
gvsampad@gmail.com

ABSTRACT

Information Technology and the invent of computers have brought significant impact on the working of the banking sector. The lifestyles of the customers is been relooked by the society. Banking through alternate channels, especially internet banking is felt much comfortable by the customers. Banking through online is felt as a pride and the status quo of an individual is rated high. The credibility of the business men is also rated high as money transfer is made faster and easy. Thus the role of information technology in the Banking sector is expected to bring a remarkable change in the contributions of manufacturing and service industry by making a smoother and faster transfer of money between the trade and traders. The financial credentials of the Indian traders rise in the international market. A small attempt is made to understand the role of information technology on alternate banking and the customer's perspective towards the handling the banking transactions through online. A semi structured questionnaire was circulated among the users of online banking of various banks in Chennai city. The data collected was analyzed through SPSS and inferences were made.

I. INTRODUCTION

The inventors of the Globe have set the "Second Wave" of Industrial Revolution through the invention of Steam engine and the "Third Wave" is considered as Information Revolution through the development of the Information and Communication Technology, popularly known as ICT.
Information Technology and the invent of computers have brought significant impact on the working of the banking sector. The lifestyles of the customers is been relooked by the society. Banking through alternate channels, especially internet banking is felt much comfortable by the customers. Banking through online is felt as a pride and the status quo of an individual is rated high. The credibility of the business men is also rated high as money transfer is made faster and easy. Thus the role of information technology in the Banking sector is expected to bring a remarkable change in the contributions of manufacturing and service industry by making a smoother and faster transfer of money between the trade and traders. The financial credentials of the Indian traders rise in the international market. A small attempt is made to understand the role of information technology alternate banking and the customer's perspective towards the handling the banking transactions through online.

II. Review of Literature

Kumbhar, Vijay M.(2009), writes that information technology paves way for alternatives in banking industry. It has brought significant changes in the way of computerization of services and new way of internet banking. Modern banking rests on the strength of information technology It not only facilitates the customer delight but also on the operational efficiency of the banks. It also helps in reduction of the operational costs of the banker and the customer.
Hsueh-Ying Wu, et. al, (2010),indicates that the role of information technology is to create a new fabric where web is transforming the entire business activity. Electronic banking has become more diversified with specialization. The customers perceive usefulness and advantage over online banking. It is also expected that the online banking would provide abundance of information and guidance of handling the finance.
Ponnurangam Kumaraguru,(2012), states that information technology paves way healthy digital economy . Right from the purchase of mobile phones, the information are stored with the help of ICT. The e services are expected to increase the value for business and enable faster development.

Tejinderpal Singh et.al,(2012), has studied the various channels through which the banking products are aligned and in particular the online portals. Content Analysis was applied to study the websites of a few banks. The study revealed that the bankers vary significantly in presenting the features in the online portal.
Atul Bamrara,et.al, (2013), states that cyber crime is becoming a challenge to the National security. There is threat to safeguard the volume of data stored. It was found that Phising, vishing, spoofing, hacking threats are the major challenge faced by the bankers.

Geeta Sharma,(2014) states that information technology, enabled bankers to operate their transaction through alternate banking. It was found that the customers prefer to make payments through internet banking operations and feel happy that it enables them with current informations in the banking sector.
Sujoy Kumar Dhar (2015), thows light on the role of information technology in financial inclusion. The author analysed the issues and challenges of electronic banking and also the strategies which facilitates the customized banking.It was revealed that the electronic banking will be successful when the whole population of India is educated on the electronic banking and when they put it in to use.

III. Scope of the Study

Virtual banking is the concept spoken worldwide. LPG era has made the citizen to realize that change is the only thing which is constant. Bankers have also realized the need for adopting new and innovative practices not only interms of service delivery mechanism, and also devise new products and services. Internet banking is expected to bring reduction in operation costs, entry costs, remove the barriers to the entry of several traders simultaneously. In this context it is essential to make a study on the perspectives of customers towards the choice of internet banking as an alternate banking. This may bring clarity to understand whether there is a positive note for the role of information technology on the alternate banking transactions.

IV. Statement of Problem

The advent of electronic banking has paved way for utility for alternate banking. Demonetization in India has brought the need for approaching banking transaction through internet. In reality, many of the online banking transaction is popular among the customers. For example ATM operations have become the choice of withdrawal of money by all customers whether educated or not and whether they are from rural or urban India. Information technology has created a healthy atmosphere for variety of choices to do online banking. The utility and need for the online banking is also created by the Government and the Bankers. But the customers are the end users and hence there is a need for a study on how the customers look upon these services of bankers? What are their expectations?

V Research Methodology

The study is conducted with the primary data. Semi structure questionnaire was framed and circulated among the customers of various banks in the Chennai city. They were asked to give responses to their utility of online banking transactions. Random Sampling method was adopted to collect the data. The data was analyzed through SPSS. Inferences were made based on the results obtained from Chi square.

VI Objectives of the Study

Primary Objective: The study aims at understanding the effectiveness of the information technology in the form of alternate banking services based on the perspective with which the customers look upon it and get benefited.

Secondary Objectives:
To know the profile of customers who prefer alternate banking. To understand the nature of banking operations dealt by the customers To analyze the knowledge of the customers on alternate banking.

H_{01} There is no significant association between the Monthly Income of customers on the level of awareness on alternate banking.
H_{02} There is no significant association between the kinds of account and the level of awareness on alternate banking.

Analysis and Interpretations

Table 1
Socio Economic profile

Educational Qualification/Occupation and Income of the Respondents	
Educational Status	**Percentage**
Diploma	6
UG	52
PG	27
Professional courses	11
Other	4
Occupational Status of the Respondents	
Self-employed	13
Public sector	7
Private sector	37
Others	43
Annual Income of the Respondents	
Below Rs.2, 50,000	51
Rs. 2, 50,001-5, 00,000	25
More than Rs.5,00,000	24

From the Table 1 it is understood that the customers who prefers alternate banking are educated and they are mostly engaged in employment. It is noticed that 52% of the respondents have completed their under graduation. As regards their employment status 37% and 7% of them are employed in the Private sector and Public sector respectively. 43% have chosen either professional or entrepreneurial operations as their major source of income. This clearly indicates that the customers have constant and regular source of income. Since 51% of the respondents earn Rs.2.5Lakhs as their Annual Income, it may be understood that they belong to the segment of Indian Middle Class. From the profile of the customers we may observe that they have high potential for savings.

Table 2
Type of Banks in which customer's bank

Nature of Bank	Percentage
Public sector	45
Private sector	55

From the Table 2, it is observed that 55% of the respondents are the customers of the Private Sector banks where as the customers of the Public sector bank is 45%. It shows that the Sector of banks is not a barrier for the customers to prefer alternate banking.

Table 3
Nature of Accounts held by the customers

Nature of Account held	Percentage
Savings Account	79
Current Account	20
Fixed Deposit	1

From the Table 3 we observe that 79% of the customers have Savings A/C, while 20% of them have Current A/C. 1% of them have invested their money in the Fixed Deposits. It is quite obvious that the middle class prefer to have savings account.

Table 4
Level of Awareness on Alternate Banking Services

Alternate Banking Services	N	Min	Max	Mean	S.D
Online Banking Facilities	100	1	3	2.78	.440
Mobile Application of Banks(Mobile App)	100	1	3	2.58	.622
Special facilities provided by each Bank on e banking	100	1	3	2.48	.594
Security Measures offered by Banks on e banking	100	1	3	2.46	.610
Demat Accounts	100	1	3	2.24	.818
Forex services through e banking	100	1	3	2.01	.859

Source:Primary Data

From the Table 4, it is observed that the mean value (2.78) is highest for the awareness on the online banking facilities followed by Mobile applications (2.58), Special facilities (2.48) and the awareness on security measures offered by banks on e banking services at (2.46). The level of awareness is less with regard to the Demat accounts and Forex services whose mean value is 2.24 and 2.01 respectively.

Test of Hypothesis

The P value (0.009) obtained from the chi square test is significant at 5% level. The null hypothesis is rejected and there is a significant association between the income groups and the level of awareness on alternate banking services. Further it is also observed that 11% of the respondents of the income group who earn above 5 lakhs have high level of awareness on the alternate banking services. While 8% and 7% of the income group within 5 lakhs also feel the same. In general it is observed that, the more the earning capacity is the higher the level of awareness on the alternate banking services.

H_{02} There is no significant association between the Kinds of account and the level of awareness on alternate banking.

Table 5
Income and Level of Awareness on alternate banking services

Monthly Income (in Rs)	**Level of awareness on Alternate Banking Services**			**Total**	**Chi-square Value**	**P value**
	Low	Moderate	High			
Below 2.5 lakhs	23 (45.09%) [63.9%]	21 (41.17%) [52.6%]	7 (13.72%) [23.1%]	51	17.184	0.009*
2.5 lakhs –5 lakhs	10 (40.0%) [27.8%]	7 (28.0%) [18.4%]	8 (32.0%) [30.8%]	25		
Above 5 Lakhs	3 (12.5%) [8.3%]	10 (47.67%) [26.3%]	11 (45.83%) [42.30%]	24		
Total	36	38	26	100		

Source: Primary Data
Note: 1. The value within ( ) refers to Row Percentage
] refers to Column
2. The value within [Percentage
3. * Denotes significant at 5% level

Table 6
Types of Account and Level of awareness on Alternate Banking services

Types of Account	Level of awareness on Alternate banking services			Chi-Square Value	P Value
	Low	**Medium**	**High**		
Savings A/c	20 (25.3%) [55.6%]	33 (41.8%) [86.8%]	26 (32.9%) [100.0%]	20.569	<0.001**
Current A/c	15 (75.0%) [41.7%]	5 (25.0%) [13.2%]	0 (0.0%) [0.0%]		
Fixed Deposit	1 (100.0%) [2.8%]	0 (0.0%) [0.0%]	0 (0.0%) [0.0%]		
Total	36 (36.0%) [100.0%]	38 (38.0%) [100.0%]	26 (26.0%) [100.0%]		

Source Primary Data
Note: 1. The value within ( ) refers to Row Percentage
] refers to Column Percentage
2. The value within [
3. ** Denotes significant at 1% level

Table 6 indicates that the P value (<0.001) is significant at 1% and hence the null hypothesis is rejected and it is understood that there is a significant association between the types of account on the level of awareness on the alternate banking operations. The level of awareness is very high for the Savings account holders. While the current account holders and the fixed deposit holders do not have much awareness on the alternate banking.

VII Conclusion

The study reveals that there is a greater impact of Information Technology on the alternate banking operations. The profile of the respondents shows that most of them are undergraduates, engaged in employment and who earn constant and regular income. This enables them save a moderate amount. It is also seen that these customers are categorized under middle class. The common phenomenon of the middle class is basically domestic savings. The majority of 79% of them have savings account. The mean value on the level of awareness on the alternate banking facilities indicates that it is very high for these customers. This shows that the objective of implementing virtual banking in India may come true. It is also observed that there is a significant association between the income groups and types of account of the customers and the level of awareness on the alternate banking operations.

Information Technology has facilitated the concept of 'any time banking'. This paper clearly states that the primary duty of the Bankers and Government is to make awareness on the customers about the alternate banking so that they start availing the facilities of the online banking services. The

phase in which information technology is developing is evidenced by the customers appreciation on the utility of the services offered by Banking sector.

VIII References

[1] Aithal.P.S, (2015), Factors affecting Banker's perspective on Mobile Banking, International Journal of Management IT and Engineering, Vol 5, Issue 7, pp-28-38.

[2] Altschiller, Donald, The Information Revolution. New York: H.W. Wilson, 1995.

[3] Alvin Toffler, Powershift (New York: Bantam Books, 1990), xx. See also Donald Altschiller, ed. The Information Revolution, (New York: H.W. Wilson, 1995);

[4] Atul Bamrara,et.al, (2013), Cyber Attacks and Defense strategies in India; An empirical Assessment of Banking Sector. International Journal of Cyber Technology, Vol 7, Issue 1,pp-49-61.

[5] Geeta Sharma,(2014), An Empirical Investigation of Demography and Customer's Perception of Internet

[6] Banking in Indore District, Madhya Pradesh, Intercontinental Journal of Finance Research Review, Vol2, Issue II, pp-17-24.

[7] Hsueh-Ying Wu, et. al, (2010), A Study of Banks' customers perceived usefulness of adopting Online Banking, Global Journal of Business and Finance.Vol.4 No.3,pp 101-108.

[8] Kumbhar, Vijay M., Alternative Banking: A Modern Practice in India (September 15, 2009). Professional Banker, Vol. IX, No. 9, December 2009. Available at SSRN: https://ssrn.com/abstract=1473898.

[9] Ponnurangam Kumaraguru,(2012), Privacy in India, Attitudes and Awareness V 2.0,Indraprastha Institute of

[10] Technology, New Delhi, Funded by IDRC, http://ssrn.com/abstract=2188749.

[11] Sujoy Kumar Dhar (2015), Role of Electronic Banking in

[12] Financial Inclusion, Available at SSRN: https://ssrn.com/abstract=2574608 or http://dx.doi.org/10.2139/ssrn.2574608.

[13] Tejinderpal Singh et.al,(2012), Internet Banking, Content Analysis of selected Indian Public and Private sector Banks' online portals, Journal of Internet Banking and Commerce, Vol.17, No.1,pp 1-10.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

CLASSIFICATION OF FOOD GRAINS USING CLUSTERING ALGORITHMS

N. Minni
Assistant Professor
Dept of Computer Science
Avvaiyar Government College for women
Pondicherry University, Pondicherry, India
minnimca@yahoo.co.in

N. Rehna
Assistant Professor
Dept. of Computer Science
SSS Shasun College for women
Madras University, Chennai, India
rehnamca@yahoo.com

ABSTRACT

Image processing nowadays plays a vital role in automation in several domains like medical science, remote sensing, agriculture, environmental science, special science etc. In this paper, we present a survey of grading the agricultural products using image processing. A model for the automatic grading of food products is suggested by analyzing their quality. Quality is checked and analyzed using the classification and clustering algorithms. Neural network and image processing algorithms are becoming prominent in the field of agriculture. So we have proposed a model to detect the type of deficiencies in the food products with the help of image processing algorithms. The essential features such as shape, size, color, texture and mass are used to grade the quality of the products.

Keywords: Image Processing, Classification, Grading, Neural Network

INTRODUCTION

A. Need for grading and standardization

Grading and standardization is well practiced at all India level for engineering products and consumer goods. It is yet to become popular for rural producer. Efforts are made by standard organization to popularize the standards. Agmark is one of the important step in popularizing quality moment by gradation. There are many advantages of grading. The important one is to obtain fair price to producer and justice to the consumer. [6]

B. Food Grading

Grading of food is categorized into different lots, each containing similar characteristics. The characteristics could be one or more of the following types:

- Size – Big, medium, small, long, short, roundish, oblong etc.
- Flavour– which in turn speaks of taste or class
  Ripeness – raw, semi-ripe, ripe in case of fruits, oilseeds, pulses and cereals.
- Length of staple – in case of cotton and jute.
- Location oriented – like Goa Alfanso, Bydagichillies,Baiganpalli mango, and Nagpur orange
- Nasik grapes – having specific tastes, shape, colour etc.[8]

Food grading involves the examination, assessment and sorting of various foods regarding freshness, quality, legal conformity and market value [1][2]. Food grading often occurs by hand, in which foods are assessed and sorted. Machinery is also used to grade foods, and may involve sorting products by size, shape and quality. For example, machinery can be used to remove spoiled foodfrom fresh product. During post harvesting sorting or grading is the most time-consuming process [4][5].

C. Image processing in grading and standardization

Image processing remains an important area in the field of computer science and engineering. Image processing takes images as input , process and analyse the images and then produce the output.

First step is to capture or gather images of consideration. Their features are stored as two dimensional array. thecolor of the images are retrieved from each pixel. The pixels are stored in the form of array of binary digits. The second phase is to segment the images. The images are compressed or enhanced as needed and segmented. Then the features are extracted from their segments for analysis. Databases such as testing and training databases are used for analyzation. The last stage is output which is obtained after analyzation.[5][9][10]

The aim of the proposed model is to classify the agricultural products according to their quality. The quality of the products are analysed using the proposed image processing algorithm. The characteristics of the products are selected based on their nature. Image processing techniques and classifiers are used to extract the features of the images, identify their defeciencies, classify and grade them. The proposed model will ease the currently available procedure of food grading.

Section 2 gives the related work and literature review of grading of rice, grains, fruits and vegetables. Section 3 explains the proposed model in detail. It also analyses the characteristics of the products taken for grading and the training database is prepared. Section 4 discusses the conclusions and future works.

II. RELATED WORK

A. *Grading of rice*

Several researches had been done in this area. In Archana et al, Classification of four paddy grains is done on shape and their color. In this paper they have used pattern classification algorithms. The algorithm uses a two layer back propagation supervised neural networks with one hidden layer. They produced 98.7% accuracy of granule classification. To grade varieties of rice kernels an algorithm was developed by S.J.Mousavi et al. [13]

Another approach used is focused on providing a better approach for identification of rice quality by using neural network and image processing concepts. Today a great deal of effort is focused on the development of neural networks for applications such as pattern recognition and classification.[15]

This research has been done to pick out the relevant quality category for a given rice sample. It was based on texture and color feature extraction and are used to measure the quality of a rice sample [15].

B. *Grading of grains*

Grading is the ultimate aim of quality checking. To classify the grains , various standards are adopted for production of grains, breeding and quality checking and finally marketing. Image processing techniques such as neural networks, clustering algorithms and fuzzy logic are involved in the grading process. The trained network was used which identifies the unknown types of grains. It obtains 98% of accuracy[14].

Nandin et al. [15] developed a model to identify grains with 100% accuracy. It uses the image processing along with probabilistic neural network [16][17].

C. *Grading of Fruits*

Generally, the fruits are graded on the basis of size, weight,gravity, colour, variety, etc. Size grading is predominantly followed in almost all types of fruits on the basis of size. The fruits are graded as a small, medium, large and extra large. Fruits are classified on the basis of their maturity. They are graded into three categories i. immature ii. Properly mature iii. Over mature. Both quality and shelf life can be determined when grading is done on maturity. The mango fruits are graded by this method. They are categorized into three grades i. sp. gravity less than 1.0 ii. Sp. Gravity between 1.0-1.02 and iii. Sp.gravity more than 1.02. For Alphonso and Pairi fruits the second category is about 50%.[6][8]

D. *Grading of Vegetables*

As far as the vegetables are considered, they are mostly graded on the basis of size and color. The vegetables such as okra, brinjal, bitter guard, green chilli, bell pepper are graded on the basis of size. They are put into three grades i. small ii. Medium iii.Large . Grading on color are done in vegetables like tomato. Potato can be classified based on its plantlet segments. This is discussed in the paper color machine vision system by Alchanatis et.al.[17][18]

III. PROPOSED GRADING MODEL USING IMAGE PROCESSING

The proposed model is developed to grade the essential agricultural food products. The products taken for grading are fruits, rice and grains and vegetables. Each product have their own features for testing their quality. The features such as shape, size, color , texture and density are taken for testing in this model. The common features for each fruit or vegetable is collected from the agricultural database and stored in the training database. The image of the product is captured by an optical scanner or a digital camera. They are digitized and compressed and enhanced for preprocessing. The RGB image as captured is then converted into HSI (Hue, saturation, intensity) image. From this HIS image each pixel is extracted and stored in the pixel matrix. The pixel matrix stores the color at each pixel of the image. The image is segmented and features are extracted and classified from the pixel matrix. These features are compared with testing database and the products are graded accordingly[10].

The training database for sample fruits is given in Table 1. The training database for rice and grains is shown in Table 2.

Table 3 shows the training database for sample vegetables.

The framework of the proposed model is shown in Fig1.

Table 1: Training database for sample fruits [5][8][11]

Food Products (Fruits)	Features	Parameters to be tested	Classification method
Apple	Color, Texture , Wavelet	Broken skin, surface discoloration, spots, scars	Statistical and Syntactical classifiers
Orange	Color, Texture, intensity	Stem end, scars, density	K-means clustering
Dates	Physical , color	Flabbiness, size, shape, intensity	Feed forward MLP
Water Melon	Shape, size, mass	Mass, volume, dimensions, density	Shape based classification
Lemon	Color, size	Volume, shape, density	ANN classification

Table 2: Training database for rice and grains [4][9][11]

Food Products (Rice and Grains)	Features	Parameters to be tested	Classification method
Rice and Grains	Shape, Size, Texture	Length of grains, density, skin decay	Naïve Bayes classifier & ANN

Table 3: Training database for sample vegetables [6][8][11]

Food Products (Vegetables)	Features	Parameters to be tested	Classification method
Tomato	Shape, color, thickness	Firmness, defects, dirt, decay	Decision tree classifier
Mango	Size,color, maturity	Skin breaks, surface discoloration, overripe	ANN classifiers
Potato	Shape, firmness	Freezing injury, bacterial ring rot, loose sprouts	FMM neural networks

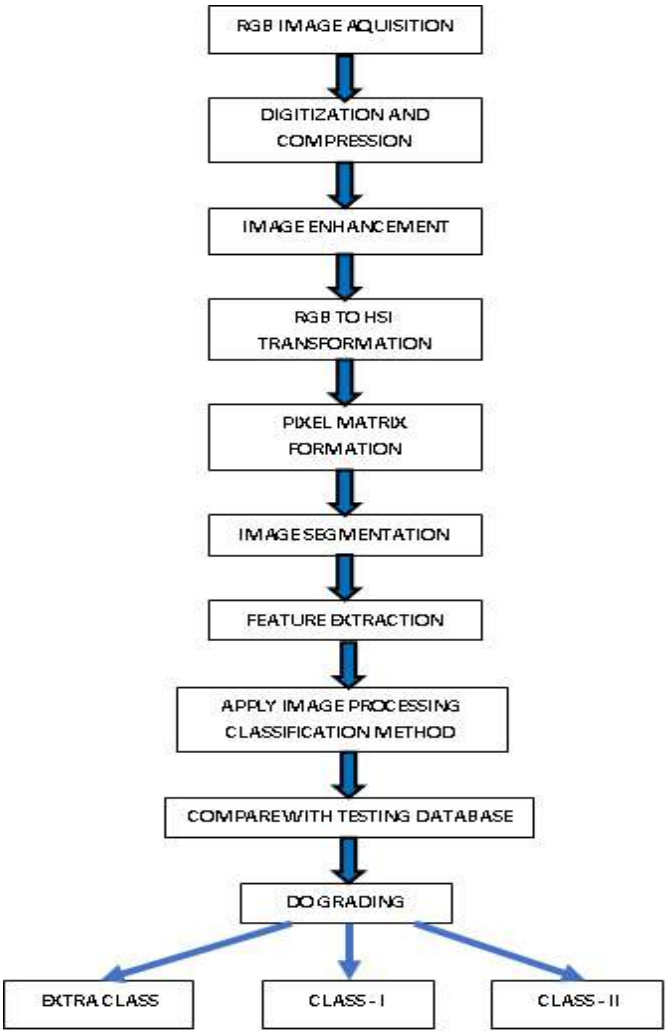


Figure 1.Grading Methodology using image processing classifiers

IV. Conclusion and Future Scope

This paper proposes an effective model for grading the quality of food products. It uses image processing applications which have been proved effective for various agricultural domains. The analysis of the parameters have proved to be accurate and less time consuming when compared to traditional methods. There are still some more features to be considered in each food product so that the grading can be done more effectively. This model can be improved by considering more combinations of features for better classification rather than identifying from one category of features. The proposed system consists of preprocessing, feature extraction, segmentation , training and classification and finally grading. This paper proposes a valuable approach which supports the accurate detection of deficiencies and lack of quality in food products and hence this model achieves efficient grading of food.

V. References

[1] Saravacos, George D, Maroulis, Zacharias B. "Food ProcessEngineering operations" pp.198-199, 2011

[2] Sivasankar, B. "Food Processing and Preservation.PHI Learning" pp. 175-177. 2002.

[3] A. Raji and A. Alamutu, "Prospects of computer vision automated sorting systems inagricultural process operations in Nigeria.," CIGR. vol.VII, pp. 1-2, 2005.

[4] S Abirami, P Neelamegam, and H Kala, "Analysis of Rice Granulesusing Image Processing and Neural Network Pattern Recognition Tool," IJCS, vol. 96, no.7, June 2014

[5] Mayur P. RajDr. Priya R. Swaminarayan "Applications of Image Processing for Grading Agriculture products" IJRITCC Vol3Issue: 3 March 2015.

[6] R.SwarnaLakshmi, B.Kanchanadevi A Review on Fruit GradingSystems for Quality Inspection" IJCSMCVol 3, Issue 7, July 2014

[7] http://agmarknet.nic.in/fveggmrules04.htm#rules

[8] Ms.SeemaBanot, Dr.P.M.Mahajan "A fruit detecting and grading system based on image processing – Review", IJIREEICE, Vol 4, Issue 1, Jan 2016

[9] Sneha S. Kausal , S.V.More "Review on Identification and Classification of Grains Using Image Processing " IJSR Volume 4 Issue 4, April 2015

[10] N.Minni, N.Rehna , "Detection of Nutrient Deficiencies in Plant Leaves using Image Processing " IJCOA , vol 5, Issue 2, December 2016

[11] Anderson Rocha, SiomeGoldenstein "Classifiers and Machine Learning Techniquesfor Image Processing and Computer Vision" Institute of Computing University of Campinas (Unicamp)13084– 851, Campinas, SP – Brazil

[12] ArchanaChaugule and Suresh N. Mali, "Evaluation of Shape andColor Features for Classification of Four Paddy Varieties," I.J. Image, Graphics and Signal processing, vol. 12, pp. 32-38, 2014.

[13] S. J. Mousavi Rad, F. Akhlaghian Tab, and K. Mollazade, "Design of an Expert System for Rice Kernel Identification using Optimal Morphological Features and back Propagation Neural Network,"IJAIS,vol. 3, no. 2, 2012.

[14] N.S. Visen, J. Paliwal, D.S. Jayas, and N.D.G. White, "Image analysis of bulk grain samples using neural networks," Canadian Biosystems Engineering, vol. 46, No. 7, 2004.

[15] Nandin Sidnal, Uttam V. Patil, and PankajaPatil, Grading and quality testing of food grains using neural network," IJRET ,vol. 02, no. 11, pp. 545-549, 2013.

[16] Yong Wu, Yi Pan, "Cereal grain size measurement based on image processing technology" ICICIP 2010

[17] Megha R. Siddagangappa1, A. H. Kulkarni " Classification and Quality Analysis of Food Grains." ISOR 2014

[18] L.A.I. Pabamalie, H.L.Premaratne "A Grain Quality Classification System" IEEE 2010. Sep 2011

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Semantic Similarity Measures: an Overview and Comparison

Dr.B.Poorna,
Principal, Shri.Shankarlal Sundarbai Shasun Jain college
For Women, Chennai, India
poornasundar@yahoo.com

A. Sudha Ramkumar,
Research Scholar, Bharathiar University
Coimbatore, India
sudharam99@gmail.com

ABSTRACT

Similarity measure is calculated based on the syntactical representation of terms. Similarity measure used in data mining task likes clustering, and classification returns irrelevant information. The semantic similarity calculated based on the relatedness between wordpairs of terms returns better result. Many researchers proposed approaches for getting word similarity by using different sources like ontologies, thesauri etc. This Paper provides an overview of six existing semantic similarity measures and compared those semantic similarity measures using the wordpairs of sports domain ontology. This paper describes about how WordNet is used to retrieve the synonyms and using synsets how semantic similarity measures are calculated. Finally, comparison of selected semantic similarity measures for the given wordpairs with respect to the considered knowledge base domain ontology and WordNet is presented.

***Keywords*:** Similarity measure; WordNet; Synonyms;

I. Introduction

In the digital era, the amount of electronic document has been increasing tremendously. The documents retrieved as a result of search query depends only on the terms. Hence, the retrieved result consists of a combination of both relevant and irrelevant content. If the machine understands the user requirement, it will fetch relevant content. In order to make machine understandable, ontologies become an integral part in today's information retrieval. Semantic Similarity measure is calculated based on the likeliness of the term's meaning. Many researchers proposed knowledge sources like WordNet Ontology in their work to prove how it can be used to calculate the semantic similarity between terms or concepts of ontology. WordNet is the lexical database developed at Princeton University and can be interpreted and used as ontology in the computer science. It is an online database which includes nouns, verbs, adjectives and adverbs grouped into sets of synonyms called synsets. Many researchers proposed that WordNet is widely used to compute the semantic similarity measure between the concepts and it reduces the dimensionality of the term matrix.

In this paper, an overview of six existing semantic similarity measures like wu&palmer, Leacock & Chodorow, HirstOnstonge, Resnik, Jiang & Conrath and Lin measures is provided. An experimental result shows the comparison of these six measures for the word pairs of sports domain ontology along with the WordNet. As a first step of this study an Overview of six existing semantic similarity measure is presented, followed by comparison of six similarity measure for the wordpairs of sports domain. In this study, new synonym retrieval algorithm is implemented to retrieve synonyms of all the selected terms using WordNet ontology and finally the similarity measure calculated to select only the most relevant synonym of the terms. This paper proves experimentally the performance of the six existing similarity measure. This paper is organized as follows Section 2 provides the Literature review of semantic similarity measure. Section 3 presents the overview of six semantic similarity measures. Section 4 compares the six semantic similarity measures followed by Conclusion in Section 5.

II. Literature Review

Zhang et al, presented a comparative study on different semantic similarity measures of term including path based measure, information content based measure and feature based semantic similarity measure affect document clustering. In their article, the domain ontology is integrated with the clustering process by reweighting the terms and proved that it has positive effects on document clustering. Mesh Ontology is used as knowledge source in this paper [15]. Zhang et al, presented nine semantic similarity measures with a term reweighting method on PubMed document. The experimental result shows that term reweighting has some positive effects on clustering and proved path based semantic similarity measures improves the performance significantly. Domain ontology is acting as a knowledge source in this paper [16].

Montserrat Batet et al. analyzed the existing semantic similarity measures by determining their advantage

and limitation based on the knowledge base. This paper proposed a new measure based on the exploitation of the taxonomical structure. SNOMED CT Ontology is used as knowledge source and accuracy of their proposed measure is compared with existing measure [2]. Lingling Meng et al. presented an effective algorithm for semantic similarity metric of word pairs. This new algorithm considers both path length and information content. This proposed algorithm outperformed traditional similarity algorithm. Here, the WordNet ontology is used as a knowledge source [10].

Gan et al, classified existing semantic similarity calculation method into 5 categories as Based on semantic distance, based on Information content, based on properties of terms, based on ontology hierarchy and hybrid method. In this paper, the knowledge resource is domain ontology. Finally, they provided a summary of characteristics, advantage and disadvantage of each category. Finally, concluded that these methods depend on 2 factors – the quality of annotation data and the correct interpretation of the hierarchical structure of ontology [4]. Thabet slimani et al, discussed about the existing semantic similarity measures based on path, information content and feature based. Based on two standard benchmarks, a calculation of all approaches is presented [12].

Mabotuwana et al, presented a semantic vector based approach to determine similarity between documents using domain ontology. This semantic algorithm improves classification accuracy when compared to non-semantic approach. Here, the domain ontology is used as a source [9]. Cui et al, proposed WordNet based semantic similarity clustering algorithm on the cluster analysis of complex network community. The proposed algorithm is compared with VSM and K-Means and proved with effective result. Here, the WordNet ontology is the knowledge resource [3]. Ali Hadj et al, proposed a new measure which combines the most significant parameters depth and hyponym of a concept. Experimental result shows that proposed measure outperforms existing path based, information content based and feature based approaches. The knowledge source used here is WordNet [13]. Ahmad Fayez et al, focuses only on the semantic similarity measure based on ontology as a knowledge source. In this paper, Al-Mubaid &Nyguan's method outperforms the other measure [1].

III. An Overview

In this section, an overview of six existing semantic similarity measures is provided. The semantic similarity measures considered in this section belongs to path based measure and information content based measure. The path based measures discussed here are wu&palmer measure, Leacock&Chodorow, Hirst &On-Sage measure and path based semantic similarity measure. The information content based measures considered for the comparison in this paper are Resnik semantic measure and Lin Semantic measure.

3.1 Path Based Measure

Wu et al, 1994 proposed a path based measure [14] that considers the depth of the concepts in the hierarchy. This measure calculates the similarity value by considering the depth of the two synsets in the WordNet, along with the depth of the least common subsumer. The wu&palmer measure ranges from 0 to 1.

$$SS_{W\&P} = \frac{2*\text{Depth (LCS)}}{\text{Depth(S1)+Depth(S2)}} \quad (1)$$

Leacock and chodorow [7] proposed a path based measure that depends on the length (C1, C2) of the shortest path between two synsets or wordpairs for their similarity measure. This measure considers IS-A links and scale the path length by the overall depth D of the taxonomy. The leacock and chodorow values lies from 0 to 4.

$$SS_{L\&C} = -\log \left(\frac{SP\ (C1, C2)}{2(max\_depth)} \right) \quad (2)$$

Hirst-St-Onge Measure [5] is a measure of semantic relatedness in that two lexicalized concepts are semantically close if their WordNet synsets are connected by a path are connected by a path that is not too long and that "does not change direction too often".

$$SS_{HS} = C - \text{Path Length} - K \times D \quad (3)$$

D is the number of changes of direction in the path. C and K are the constants. If SS_{HS} is zero then there is no path exists in between the concepts. The Hirst-St-Onge values ranges from 0 to 16.

3.2 Information Content Based Measure

Resnik similarity measure [11] considers the integration of ontology and corpus. Resnik defined the similarity between two concepts lexicalized in WordNet to be the information content of their lowest super coordinate that is the most specific common subsumer.

$$SS_{Res} = -\log P\ (LCS\ (C1, C2)) = IC\ ((LCS\ (C1, C2)) \quad (4)$$

Where $IC(C) = \frac{\log (\text{depth}(C)}{\text{Log}(\log(\text{deep}_{max})}$

Jiang and Conrath measure [6] is based on information content. Here, the distance between two concepts c1 and c2 is calculated as the difference between the sum of information content of the two concepts c1 and c2 and the information content of their most informative

subsume. Jiang & conrath measure is calculated using the following formula,

$$SS_{JC} = 2 * \ln P_{mis}(C1, C2) - (\ln(P(C1) + \ln(P(C2)) \quad (5)$$

This measure is the shortest path length between two concepts c1 and c2 and the density of concepts along the same path.

Lin Similarity measure [8] follows from his theory of similarity between arbitrary objects. It uses the same element as Jiang and Conrath. It is based on Resnik's similarity and it considers both the information content of lowest common subsume and two compared concepts. The Lin Similarity measure is calculated as follows,

$$SS_{Lin} = \frac{2 * sim_{Res}(C1, C2)}{IC(C1) + IC(C2)} \quad (6)$$

IV. COMPARISON

We conduct experiments on bbcsport dataset. We developed sports domain ontology using protégé tool. The concepts of sports domain ontology are extracted using Jena, a java framework for OWL ontology. The extracted concepts are mapped with the extracted terms of bbcsports dataset. After the term-concept match, the terms are selected for further processing. The selected terms are searched in WordNet for the synonyms. The synonyms returned consist of an array of words and is called as synsets. The pair of words of synsets is applied to the six semantic similarity measures. The similarity measures are calculated with the help of WordNet Similarity for Java (WS4J).

The experiment is conducted in eclipse luna, a integrated development environment. In order to compare six semantic similarity measures, the values are normalized to value ranges from 0 to 1. For the wordpair "centuries" and "hundred", the following table shows the semantic similarity values,

Table 1: Comparison of Six Semantic Similarity Value

Category	*Measure*	*Semantic similarity Value*	*Normalized value*
Path Based	*Wu&Palmer*	*1.0*	*1.0*
	Leacock&Chodorow	*3.688*	*0.92*
	Hirst and St-Onge	*16*	*1.0*
Information content based	*Resnik*	*8.4699*	*0.94*
	Jiang and Conrath	*1.2876*	*0.85*
	Lin	*1.0*	*1.0*

The value relies on the knowledge source like WordNet, thesauri and Domain ontology. If the knowledge source is domain ontology, then the value of the semantic similarity depends upon the correct interpretation of concept hierarchy. If the taxonomy is not correct, then there will be misinterpreted value will be returned as a result. So care must be taken during the development of domain ontology. If the WordNet is the knowledge source, then the value will be accurate. The following fig.1 shows the comparison of six semantic similarity values.

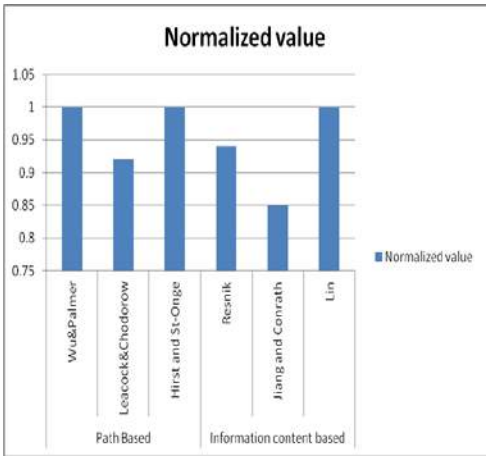


Figure.1 Comparison of six similarity measures

V. CONCLUSION

This paper compared six selected semantic similarity measure. The experimental result shows that Wu&Palmer, Hirst-St-Onge measure and Lin measure outperforms other semantic similarity measure. The Wu & Palmer and Hirst-St-Onge measure belongs to the path based category whereas the Lin measure belongs to the information content based measure. These measures play an important role in document clustering and classification process, because they reduce the complexity of clustering process by reducing the dimensionality of the term matrix. There are many other semantic similarity measures exists, depends upon the application and the knowledge source, the similarity measure improves the performance of the clustering process. This paper attempts to help the researcher to understand about the importance of semantic similarity measure in clustering process.

VI. REFERENCES

[1] Althobaiti, A. F. S. (2017). Comparison of Ontology-Based Semantic-Similarity Measures in the Biomedical Text. Journal of Computer and Communications, 5(02), 17.

[2] Batet, M., Sánchez, D., & Valls, A. (2011). An ontology-based measure to compute semantic similarity in biomedicine. Journal of biomedical informatics, 44(1), 118-125.

[3] Cui, L. Z., Lu, N., & Jin, Y. Y. (2014). Community Clustering Algorithm on Semantic Similarity in Complex Network. Lecture Notes on Software Engineering, 2(4), 348.

[4] Gan, M., Dou, X., & Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal, 2013.

[5] Hirst, G., & St-Onge, D. (1998). Lexical chains as representations of context for the detection and

correction of malapropisms. WordNet: An electronic lexical database, 305, 305-332.

[6] Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint cmp-lg/9709008.

[7] Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. WordNet: An electronic lexical database, 49(2), 265-283.

[8] Lin, D. (1998, July). An information-theoretic definition of similarity. In Icml (Vol. 98, No. 1998, pp. 296-304).

[9] Mabotuwana, T., Lee, M. C., & Cohen-Solal, E. V. (2013). An ontology-based similarity measure for biomedical data–Application to radiology reports. Journal of biomedical informatics, 46(5), 857-868.

[10] Meng, L., Huang, R., & Gu, J. (2013). An effective algorithm for semantic similarity metric of word pairs. International Journal of Multimedia and Ubiquitous Engineering, 8(2), 1-12.

[11] Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. J. Artif. Intell. Res.(JAIR), 11, 95-130.

[12] Slimani, T. (2013). Description and evaluation of semantic similarity measures approaches. arXiv preprint arXiv:1310.8059.

[13] Taieb, M. A. H., Aouicha, M. B., & Hamadou, A. B. (2014). Ontology-based approach for measuring semantic similarity. Engineering Applications of Artificial Intelligence, 36, 238-261.

[14] Wu, Z., & Palmer, M. (1994, June). Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics (pp. 133-138). Association for Computational Linguistics.

[15] Zhang, X., Jing, L., Hu, X., Ng, M., & Zhou, X. (2007, April). A comparative study of ontology based term similarity measures on PubMed document clustering. In International Conference on Database Systems for Advanced Applications (pp. 115-126). Springer, Berlin, Heidelberg.

[16] Zhang, X., Jing, L., Hu, X., Ng, M., Xia, J., & Zhou, X. (2008). Medical document clustering using ontology-based term similarity measures.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Fusion of Hybrid Optimization Algorithm and Fuzzy set for enhancing Information Retrieval using clustering

B. Gomathi*
Research & Development Centre,
Bharathiar University ,Coimbatore,
Tamil Nadu, India.
gomathiphd2k15@gmail.com

P. Sakthivel
Department of Electronics and Communication Engineering
Anna University, Chennai
Tamil Nadu, India

ABSTRACT

Fusion is the concept which deals with combination and the aim is to improve the overall performance. This paper proposes a novel approach named EPSOCS where Particle Swarm Optimization and cuckoo search are combined and fuzzy set is used for enhancing information retrieval. Clustering is accomplished and fuzzy c means algorithm is used for retrieval of information. This approach Enhanced Particle Swarm Optimization and Cuckcoo Search combines advantages of PSO and cuckoo with fusion of fuzzy set and clustering. The evaluation of the proposed algorithm shows that the new approach exhibits good optimization ability with fast convergence speed leading to efficient information retrieval.

Keywords: EPSOCS, Fuzzy C Means , PSO, Cuckoo Search, Swarm Optimization, Fusion

1. Introduction

Data is available enormously and the volume of data is increasing at a very high speed. Fusion is a methodology used in many areas of technology since they increase the performance rate. Fusion technique can be used along with information retrieval to improve the efficiency rate of data retrieved. Clustering is key feature that plays vital role in knowledge discovery. Fuzzy logic fused with clustering yields better results . Fuzzy C Means algorithm for clustering is an efficient way and this is commonly used if the data is fuzzy. There are various similarity measures and similarity computation is done using cosine similarity.

The degree of membership is directly proportional to the distance between data object to the cluster centers. The only pitfall when using Fuzzy c-means is randomly selected centre points reaching the local optimal solution. To overcome this issue swarm intelligence that are bio inspired can be applied. This paper focuses on hybridization of Particle swarm optimization and cuckoo search and the novel method proved to be efficient than applied individually.

This scope of the paper is to deliver a novel method called Enhanced Particle Swarm Optimization with Cuckoo Search EPSOCS which combines advantages of PSO and CS along with fusion of fuzzy set and clustering. Section 2 deals with Related work and section 3 deals with comparative study of genetic algorithm and about the Proposed method , section 4 describes about results based on metrics followed by conclusion.

2. Related Work

Data mining is always a flourishing field because data analysis can be easily accomplished using clustering approaches. Fuzzy clustering is of key importance in data analysis as most of the real time system prefer fuzzy than hard clustering. The membership value ranges between 0 and 1 in fuzzy and data falls between these values[1-2]. Fusion technique was applied by BogdanDit et al., [3] for combining information retrieval with link analysis algorithm to upgrade feature location in software. This is a fusion model for feature location and the new feature location techniques are based on integration of textual, dynamic, and web mining or link analysis algorithms applied to software. Swarm Intelligence is based on inspiration received from biological systems[4-5].

Every optimization algorithm is based on various inspiring agents. Ant Colony Optimization is based on behaviour of ants. Inspiring agent for Artifical Bee colony is honey bee. Particle Swarm Optimization is based on social behaviour of birds[6]. Cuckoo search is based on reproduction strategy followed by cuckoo and it is also a metaheuristic algorithm[7]. CS is a new approach and the popularisation of this algorithm is its implementation is easy. Hence CS is applied to solve many real world problem like scheduling problem etc. A detailed review of different Information retrieval technique is done and compared with proposed model. The proposed methodology combines PSO with CS and the enhanced hybrid optimization algorithm yield better result than used individually. In information retrieval similarity computation is major task to find relevancy and after careful study of various distance measure cosine similarity is used .The proposed model is a fusion of hybrid optimization algorithm with fuzzy set for efficient information retrieval.

3. Proposed System Architecture

This paper provides step by process by systematically splitting the work into three phases. During the first phase association among data is done and the statistical measure used to calculate distance is cosine similarity. In the second phase clusters are formed and evaluated using Fuzzy C Means algorithm. In the final phase Fusion process is done by combining hybrid optimization techniques with fuzzy data. The proposed system architecture is depicted in figure 1.

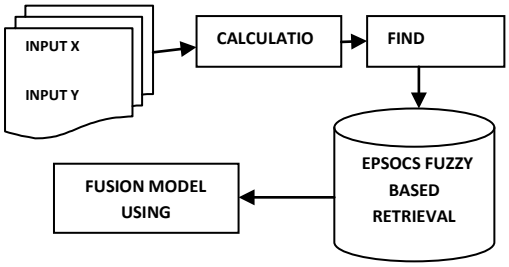


Figure1 Model of Proposed System Architecture

The Input Module deals about the data that needs to retrieved and this module is related to corpus creation. The proposed methodology uses health care data from UCI repository. Similarity computation is calculated and cosine similarity is used. The need for similarity computation is to find whether data belongs to same category. In this paper we use Fuzzy C means as many existing approaches proved that FCM provides better result compared to K-Means. The key factor is it permits data residng in multiple clusters with different membership values to participate in cluster analysis.

The objective function is given by equation (1)

$$K_m = \sum_{a=1}^{D}\sum_{b=1}^{N} {\mu_{ab}}^m \|x_a - c_b\|^2 \quad (1)$$

Step 1: Cluster membership values are initialised at random example .μ_{ab}

Step 2: Cluster distance are calculated.

$$C_b = \frac{\sum_{a=1}^{D} {\mu_{ab}}^m x_a}{\sum_{a=1}^{N} {\mu_{ab}}^m} \quad (2)$$

Step 3: calculate the objective function.

$$\mu_{ab} = \frac{1}{\sum_{j=1}^{N}\left(\frac{\|x_a - c_b\|}{\|x_a - c_j\|}\right)^{\frac{2}{m-1}}} \quad (3)$$

Step 4: Repeat steps 2-4 till objective function is minimised or for specified number of maximum iteration.

The evaluation is based on precision and recall value of the retrieved document.

$$\text{Precision} = \frac{|\{\text{relevant document}\} \cap |\{\text{Retrieved Document}\}|}{|\{\text{Retrieved Document}\}|} \quad (4)$$

The above equation 4 represents precision value.

The recall value is given by the equation 5.

$$\text{Recall} = \frac{|\{\text{relevant document}\} \cap |\{\text{Retrieved Document}\}|}{|\{\text{Relevant Document}\}|}$$

The below figure 2 represents Precision and Recall value using Fuzzy C Means algorithm.

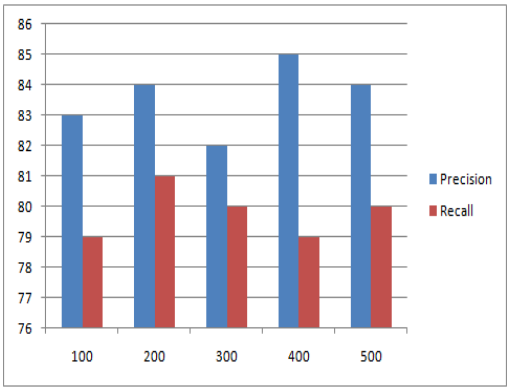


Figure 2 Precision and Recall value using Fuzzy C Means

3.1 Fusion of Fuzzy weight based system with Hybrid optimization

In this phase fuzzy based retrieval is performed using hybridization concept. The proposed method combines particle swarm optimization with cuckoo search and this enhancement on fuzzy based system proved to be efficient than used individually**.** In this paper, hybridization of Particle Swarm Optimization and Cuckoo search is done and classification based weight assignment with fuzzy approach is used in retrieving the information. Weights are calculated based on the importance such as very high, medium and related information and this is given in the figure 3.

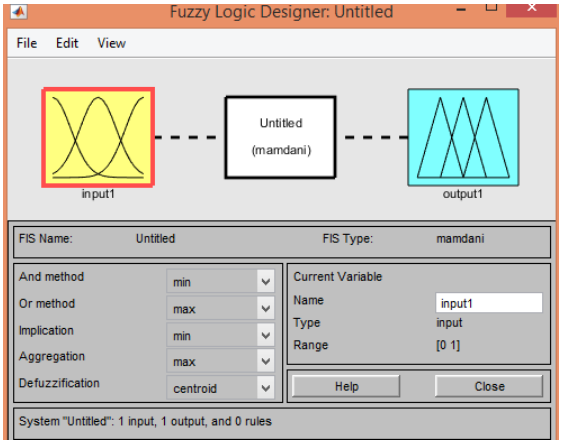


Figure 3 Fuzzy representation

Particle Swarm Optimization algorithm was discovered by James Kennedy and Russell Eberhart in the year 1995 It is based on the intelligence and movement of swarms. In this algorithm the work of agents also called as particles is to find the best solution in search space. The agents also referred as swarm must adjust flying and also they should consider about flying made by other agents. So personal best and global best position to be maintained. The advantages of PSO is the absence of selection operation normally used in genetic algorithms, and evolutionary programming. PSO also avoid crossover operation and it does not considers survival of fittest. Cuckoo search is based on the behavior of cuckoo. It is simple and easy to implement. The logic in CS is every cuckoo lays one egg and the egg is thrown at random in nest.

Only the best eggs will be forwarded to the next level. The total number of nest is fixed and the host bird discover the egg based on probability. For finding new solution levy flight and random walk is used. The usage of levy flight in local and global search plays a significant rule in cuckoo search[8].

3.2 Algorithm for Information Retrieval

Step 1: Read the input data

Step 2: From that information, the relation between the information's are identified

Step 3 :Measure the distance between the information using the cosine similarity

Step 4: Based on the distance similarity between the information is identified

Step 5:Form the cluster using Fuzzy C means clustering approach with hybrid optimization

4. Results

Different IR techniques are compared with the new EPSOCS method on the benchmark dataset . Comparison of different Information Retrieval technique with the novel method is given in the table 4.1.

IR Technique	Precision	Recall
Vector Space Model	.75	74
Latent Semantic Indexing	.83	.82
Latent Dirichlet Allocation	.861	.84
PSOCS weighted based Fuzzy Retrieval	.935	.87

Figure 4.1 Comparison of IR Techniques

The above table clearly depicts that our new approach EPSOCS Weight based Fuzzy Retrieval enhances the efficiency of information retrieval.

5. Conclusion

In this paper fusion technique is deployed by combining fuzzy set with hybrid algorithm. The integration of Particle swarm optimization with cuckoo search fetches better result than used individually. EPSOCS has the advantages of fast convergence speed, strong searching ability, and the capability to solve the problem of multidimensional continuous space optimization by using test functions. In future bat algorithm may be fused with EPSOCS in order to get better result.

REFERENCES

[1] Pal, N. R., Bezdek, J. C., and Tsao, E. C.-K. 1993, "Generalized clustering networks and Kohonen's self-organizing scheme", IEEE Trans. Neural Netw. 4, 549–557.

[2]Bezdek, J. C. 1981, " Pattern Recognition With Fuzzy Objective Function Algorithms", Plenum Press, New York, NY.

[3] Bogdan Dit, Meghan Revelle, Denys Poshyvanyk, "Using Data Fusion and Web Mining to Support Feature Location in Software", 18th IEEE International Conference on Program Comprehension, 2010, pp.14-23.

[4] Mohd Nadhir Ab Wahab, Samia Nefti-Meziani, and Adham Atyabi" A comprehensive Review of Swarm Optimization Algorithms", Published online 2015 May doi: 10.1371/journal.pone.0122827

[5] Ying-Chih Wu, Wei-Ping Lee , Ching-Wei Chien,"Modified the Performance of Differential Evolution Algorithm with Dual Evolution Strategy", 2009 International Conference on Machine Learning and Computing IPCSIT vol.3 (2011)

[6] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization: an overview," Swarm Intelligence, vol. 1, no. 1, pp. 33–57, 2007.

[7] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in Proceedings of the World Congress on Nature Biologically Inspired Computing, pp. 210–214, IEEE, Coimbatore, India, December 2009

[8] G. Zumofen, J. Klafter, and M.-F. Shlesinger, "Lévy flights and Lévy walks revisited," in Anomalous Diffusion From Basics to Applications: Proceedings of the XIth Max Born Symposium Held at Lądek Zdrój, Poland, 20–27 May 1998, vol. 519 of Lecture Notes in Physics, pp. 15–34, Springer, Berlin, Germany, 1999.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Big Data Analytics and Data Science – A Review on Tools and Techniques

Abirami.K, RajaMeenakshi.S and Supriya.R
Shri Shankarlal Sundarbai Shasun Jain College for Women, T.Nagar, Chennai, India

ABSTRACT

Big Data is a humongous amount of data in any form. Big data Analytics is the process to perform statistical analysis of data. Data science is go-ahead approaches that analyze past and current data in exploratory way to predict the occurrence of particular event in future. Tools used in big data and data science are to extract knowledge from the data in addition tools yield intelligence data solutions to business. This paper objective is to provide the essential details about Big Data, Data Science and tools in which through innovative analytics are achieved. Tools in Big data platform can be classified to accomplish few key tasks akin to Storage, Querying and Analysing. This paper also deals with Apache Hadoop, Hive, EXCEL, R-Programming, and Tableau these tools obviate the programming aspect and provide a GUI to build predictive models.

Keywords **-** Image Processing, Classification, Grading, Neural Network

I. INTRODUCTION

Big data deals with large amount of structured, semi structured and unstructured data. Data that is very large or unstructured must be converted to structured data which cannot be achieved by relational database engines. This type of data requires the concept of big data which provides structured data. Its process includes capturing of data, storing, search, updating and analyzing. Five dimensions of big data are volume (quantity), variety (difference), velocity (Speed), Veracity (accuracy) and Value (scope). Big data information arrives from different domains like healthcare, Banking Sectors and equity market. These data can be processed through many tools and techniques.

Data science is an interdisciplinary field of mathematics, statistics, computer science to extract knowledge and patterns from complex data. There is overwhelming flow of data from various sources like internet, e-commerce sites, cell phones and social media. This data can be structured or unstructured like videos, audio etc. Data science as a whole is related to compiling, purifying and analyzing the data. Data science is a composite of algorithms and tools from various disciplines to gather data, obtain insights, extract meaningful information and explicate it for decision making.

Data science is highly applied in various fields such as Marketing, Social media, healthcare, education, security, biological science etc.

Data Analytics also known as Analysis of data. Data analysis refines the data using diverse techniques. Analysis process evolves collection of data, applying statistical methods to derive data for decision making. Data analysis methods can be a quality or quantity based. How data analysis techniques imply to big data is a challenge in this era

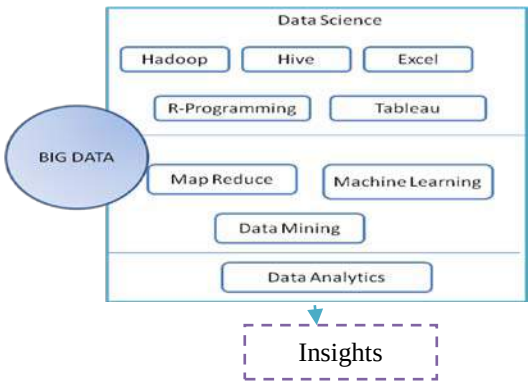


Fig: 1 stack of Bigdata, Data science, Tools and techniques

Above diagram depicts the flow of data from the various sources, also tools and techniques involve in data science and big data. When the enormous amount of data generated by the social media like Facebook, twitter, also data from healthcare domains turned into Big Data. Traditional tools and techniques fall short to handle big data. Thus, we couldn't derive knowledge from the data scattered around the system.

Both big data analytics and data science deals with unstructured data but big data analytics commonly used in financial services, retail marketing and healthcare whereas data science used to predict models in the new-fangled fields.

This paper explores the tools like Hadoop (storage), Hive (Querying), Excel (Analysing), R-Programming (Statistical Computing in Data Science), Tableau (Visualization) and the data mining, machine learning, Map Reduce techniques.

1. Hadoop

Hadoop is an open source framework from Apache.

Hadoop can handle data in huge volume. It is used to store the colossal amount of data in Big Data.

Hadoop Distributed File System and Map Reduce Engine are the main components in Hadoop. With the Hadoop Distributed File system the data is stored once on the server and consequently read and re-used many times thereafter.

It uses client/server structural design, with each cluster consisting of a single Name Node that manages file system

operations and along with Data Nodes that manage data storage on individual compute nodes.MapReduce is to schedule and process across the cluster.hadoop is a data warehouse whereas MapReduce process the data by dividing into smaller units.

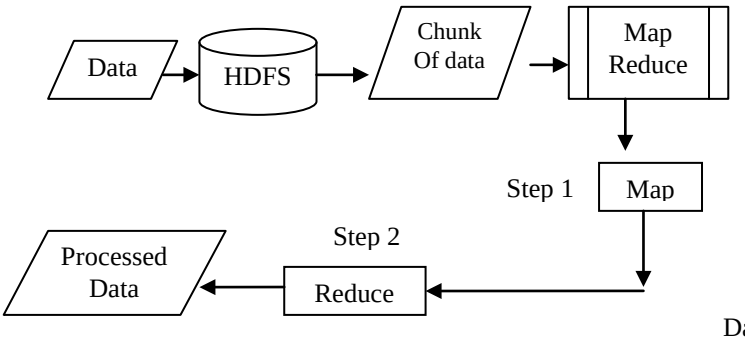


Fig 1.1

Data Process in Hadoop

Unlike traditional relational database systems (RDBMS) that can't scale to process large amounts of data, Hadoop enables businesses to run applications on thousands of nodes involving thousands of terabytes of data.

2. Hive

Apache provided another platform named as Hive to handle (querying) big data. HiveQL is a query language from Hive. Result set from this HiveQL is transferred to SQL Query set.

It does generate the MapReduce Scripts.

Like RDBMS hive facilitates storing of object in a binary stream and Meta data information through Hive Metastore.

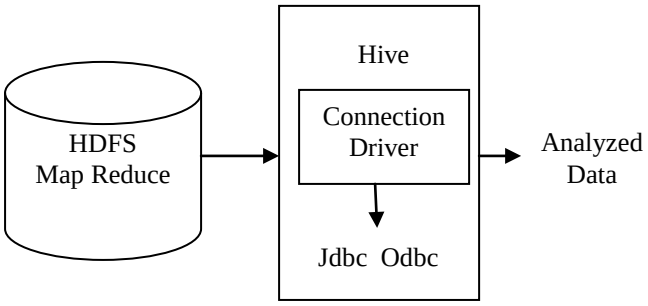


Fig: 2.1 Hive architecture

Hive is uniquely placed to come up with querying of data, powerful analysis, and data précising while working with huge volumes of data. The vital part of Hive is the HiveQL query which is an SQL-like interface that is used widely to query that is stored in databases.

Also Hive enables access data from Apache HBase storage system.

Connect Excel to Hadoop:through HiveODBC.it is possible to have a Hive add in to the Excel.

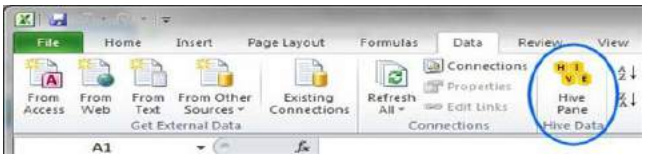


Fig 2.2 Hive Pane in Excel

Using the Hive Pane odbc connection is established and Hive data is imported to excel.

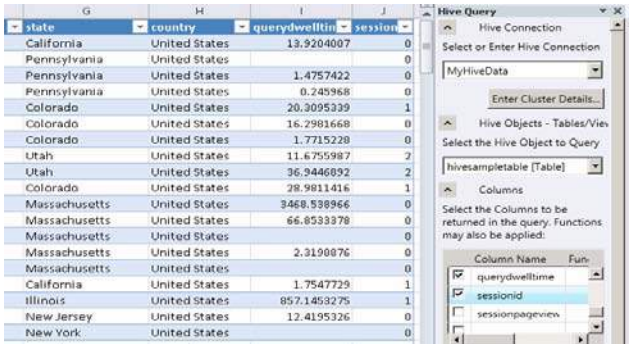


Fig 2.3 Hive data into Excel

3. Excel

Data stored in hadoop can be accessed by excel using ad hoc techniques. we can import data from hive to excel using HDInsight Services. Microsoft power query in excel used to extract the needed information from the data source. To accomplish this HDInsight cluster data is used.

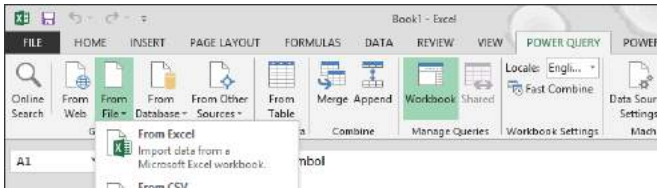


Fig 3.1 Power Query in Excel

Data cleaning and filtering are the crucial task of excel in big data.

4. R-Programming

It becomes the de facto programming language for data science. It performs statistical analysis of data.

RStudio framework is used to compile and execute R programming.

Data are manipulated using R programming. With the use of R Engine and IDE R-programming process the data and yields the data as statistical report.

Predominantly this language used to provide analyze patterns before the happening of the events.

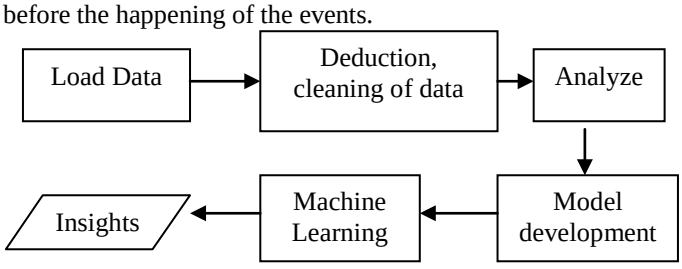


Fig 4.1 Data Process in R-Programming

5. Tableau

Tableau is primarily used in business intelligence platform which offer data visualization and exploration capabilities. When combined, Tableau and R offer one of the essential and complete data analytics solutions in the industry today,

providing businesses with incomparable abilities to see and understand their data.
A scalable object to various sources from tableau is the additional benefit.

Tableau has an interactive dashboard which renders the image as a result from analysis very quickly. The dashboard will also give you rich visualizations. The data visualization dashboard offer an in depth knowledge into the data.

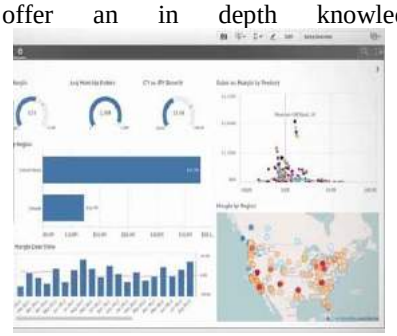


Fig 5.1 Tableau data output

It works with huge datasets and it includes more visualization tools.

6. Techniques

A. Machine learning (ML), a associate-field of artificial intelligence (AI), focuses on the task of enabling computational systems to learn from data about how to perform a desired task automatically. Machine learning applications including decision making, forecasting or predicting and it is a key enabling technology in the deployment of text mining and big data techniques in the diverse fields of engineering, healthcare, science, , business and finance.
Two types of Machine learning are:

- **Supervised ML:** The program is "practiced" on a pre-built set of "training examples", which then facilitate its ability to reach an exact conclusion when given new data.
- **Unsupervised ML:** finite amount of data should be given as the program input and must find patterns, relationships therein.

Working sample of Machine Learning algorithm:

ML described as learning a target function (f) that best maps input variables (M) to an output variable (N).

$$N = f(M)$$

In the learning process we develop predictions for the future (N) and provided (M) new inputs.

7. Data Mining

Mining is the analyse process from various view and give a summary to useful information. This task can be automatic or semi automatic on heavy data sets.
Data mining is a knowledge Discovery process which applies to the enormous set of data. Like Big Data.

8. Map Reduce

It's coupled with hadoop hdfs to handle big data. It incorporates stages.map stage, shuffle stage, reduce stage. Semantically the first two stages distribute the data and the reduce phase does the computation.:

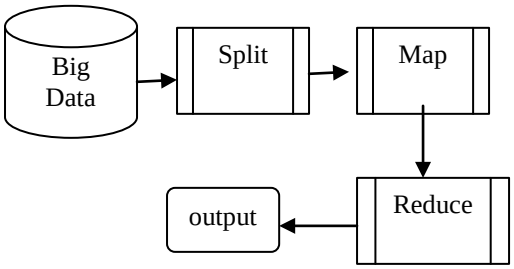


Fig 8.1 Steps in Map Reduce

Limitation of this technique is user has to follow a logic of his own.

9. Conclusion

This paper presents the overview on big data, data science and related tools, techniques. Tools discussed here provide the efficient way of handling big data. Understanding of this tools and its usage enables the researchers, business people, and academician to work with big data in an effective manner.

Future work:
Processing of ontology based data set using big data tools and to derive the insight for healthcare domain.

10. References

[1] M. R. Wigan and R. Clarke, ―Big Data's Big Unintended Consequences‖, IEEE Computer Society, , vol. 46, no. 6, (2013), pp. 46-53.
[2] Abdul Raheem Syed, Kumar Gillela, Dr. C. Venugopal, "The Future Revolution On Big Data", In International Journal of Advanced Research in Computer and Communication Engineering, 2013.Volume: 2 (Issue:6) , Page No. 2446- 2451.
[4] Daniel, B. K., Big Data and analytics in higher education: opportunities and challenges. British Journal of Educational Technology, 2015,46, 904–920.
[5] AmirGandomi, MurtazaHaider - Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, April 2015, Volume 35, Issue 2, Pages 137-144.
[6] Salisu Musa Borodo, Siti Mariyam Shamsuddin, Shafaatunnur Hasan - Big Data Platforms and Techniques, Indonesian Journal of Electrical Engineering and Computer Science Indonesian Vol. 1, No. 1, January 2016, pp. 191 -200.
[7] Proyag Pal,Triparna Mukherjee , Dr. Asoke Nath-Challenges in Data Science: A Comprehensive Study on

Application and Future Trends, August 2015,Volume 3, Issue 8.

[8] https://importoioweb.staging.wpengine.com/post/best-big-data-tools-use/.

[9] https://towardsdatascience.com/the-10-statistical-techniques-data-scientists-need-to-master-1ef6dbd531f7

[10] http://www.datascientists.net/what-is-data-science

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

ANDROID AND ITS BACKGROUND DEVELOPMENTS

R.Nandhini,
Department of Computer Applications,
S.S.S. Shasun Jain College for Women
T.Nagar Chennai-17

R. Aparna,
Assistant professor,
Department of Computer Applications,
S.S.S.Shasun Jain College for Women
T.Nagar Chennai-17

ABSTRACT

Android is basically for touch screen purpose in mobile phones. Google play store has lots and lots of apps that are being developed and updated day today. Apps that run in mobile phones are mostly developed using Android Studio or eclipse. Android Studio is Android's official Integrated Development Environment built on Jetbrains' intellj IDEA software. Android studio helps in building apps with highest quality. Android studio offers tools such as rich code editing, profiling tool, debugging and testing. The app developed can run either in virtual device such as emulator or any connected mobile phone while testing.
This paper gives you a familiarity on what is android, android studio, how to develop an application and run in mobile phone.

Keywords: android studio, debug, android runtime, activity, layout, virtual device.

INTRODUCTION

Android is open source mobile operating system originally developed by android Inc,butGoogle owned it on 2005.The initial release date of android is 23 September 2008.Google incorporate android's IDE android studio to develop android mobile application(app).The mobile application[1] can be developed using java language. Google-enabled java libraries help the code in java language to control mobile devices[2]. A platform is needed to develop mobile application. Therefore it is important to download the android SDK (Software development kit), that compile code. Secondly download JDK (Java development kit).JDK contain compiler, debugger, JDE (Java runtime environment) that includes java libraries, java virtual machine, components to run applet programming etc. To install the application in the mobile or any hardware device we should create android virtual device (AVD). APK (Android package kit) file contains all contents of android app[1].

STUDY

A) Application Framework
Android application framework is a technique of developing an android application in java coding in the android platform with the help of tools and API libraries provided by android SDK[3].

Application framework supports various media and features such as video,audio, device emulator, debugging tool etc.

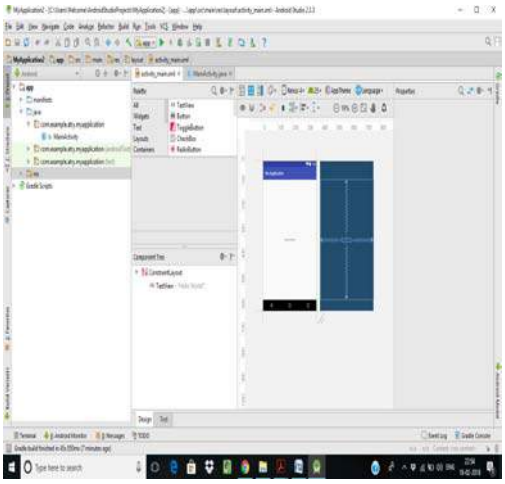
B) Android Runtime
The functionalities of Java coding is available in the set of core libraries of android. Dalvik virtual machine allows android application to run on its own process. To optimize minimum memory Dalvik virtual machine[4] execute file in (.dex) Dalvik executable format. Some developments for mobile devices can be made with DVM which is a register virtual machine.

c) Android Operating System
Based on the Linux kernel platform the mobile operating system android[5] was developed. For mobile devices android is a free downloadable open source software stack.The most used and convenient operating system is android. The applications developed using android OS has different delivering methods and computing platform. With the market share of 48% android is the most used operating system.

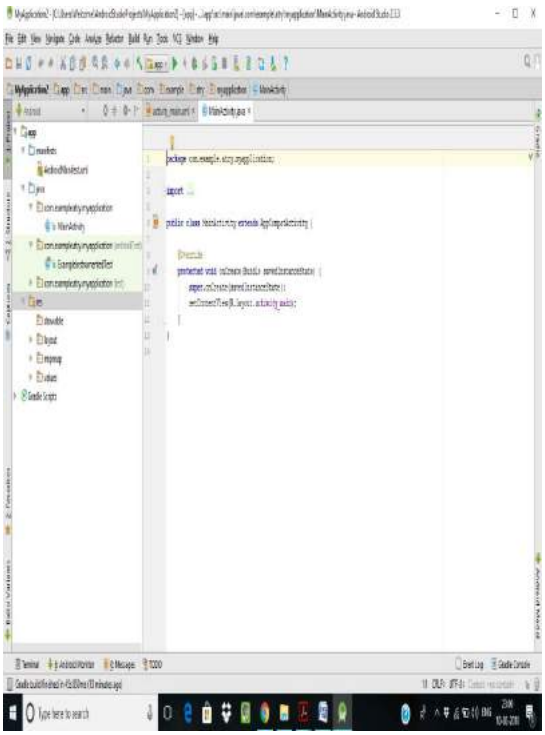
D) Android platform overview
The major components of android platform are system apps(such as Dialer, Calendar, Camera, Email,….), Java API framework->content provider-View system- Managers(such as activity, location, package, notification, resource, telephony, window,….),Native C/C++ libraries(Web kit, OpenMax AL, Media framework,….)Android runtime(Android runtime(ART) , Core libraries),Hardware abstraction layer(Audio, Bluetooth, Camera, Sensor,..),Linux kernel(drivers-Audio, Binder, Display, Keypad, Bluetooth, Camera, Shared memory, USB, wifi) and power management.

DEVELOPMENT OF AN APP

A) Working with layout and Activity
After opening the android studio,
- Click File -> New -> New Project
- Specify the application name, minimum SDK and type of activity.
- Then customize the activity and click finish.
- Once the gradle build is finished, go to res ->layout ->activity_main.xml ,in this we can design the layout.



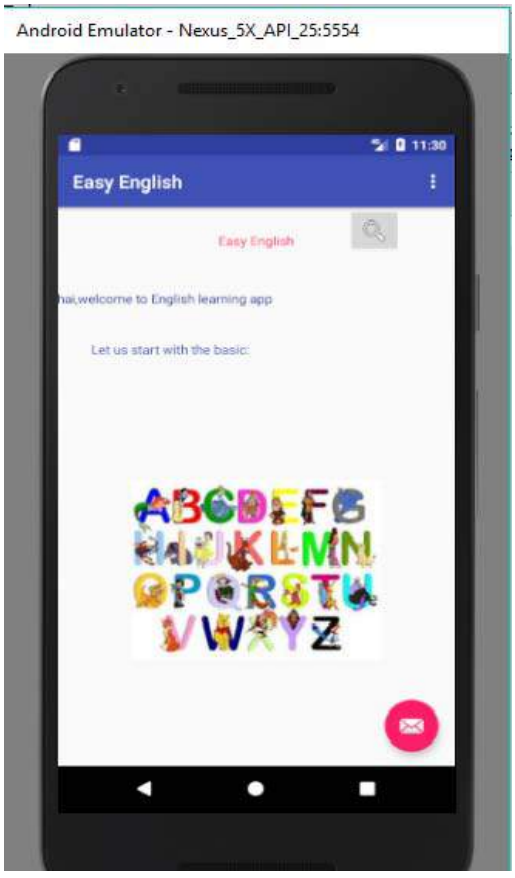
- For creating code in java file for actions, in project file go to java folder and select the application package and click on the MainActivity.java and specify the java coding and link with the layout.xml



```
package com.example.welcome.easyenglish;

import ...

public class MainActivity extends AppCompatActivity {

    @Override
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
    }
}
```

E) Running of app

To run[6] the app in virtual machine or in mobile phone, first click the Run 'app' button or press Shift +F10.
Now from the select deployment target dialog box select the connected device or available virtual device or create a new virtual device and click ok.
When target device comes online, the activity is launched in the emulator as follow.



If error is generated in the coding, then the activity will not be launched in the emulator. We must debug the bug and then proceed with running the activity. The type of bug will be given at the bottom in order to resolve the error easily.

F) Setting up android manifest

The icon of the application should be specified in the AndroidManifest.xml. The image that should be kept as icon is to be placed in mipmap folder. The activity that is launcher(default) is also specified in tAndroidManifest.xml

```
<?xml version="1.0" encoding="utf-8"?>
<manifest xmlns:android="http://schemas.android.com/apk/res/android"
    package="com.example.welcome.easyenglish">

    <application
        android:allowBackup="true"
        android:icon="@mipmap/ic_launcher"
        android:label="Easy English"
        android:roundIcon="@mipmap/ic_launcher_round"
        android:supportsRtl="true"
        android:theme="@style/AppTheme">
        <activity
            android:name=".Basics"
            android:label="Easy English"
            android:theme="@style/AppTheme.NoActionBar">
            <intent-filter>
                <action android:name="android.intent.action.MAIN" />

                <category android:name="android.intent.category.LAUNCHER" />
            </intent-filter>
        </activity>
        <activity android:name=".MainActivity"></activity>
    </application>

</manifest>
```

LIFECYCLE OF ACTIVITY

A) Activity states

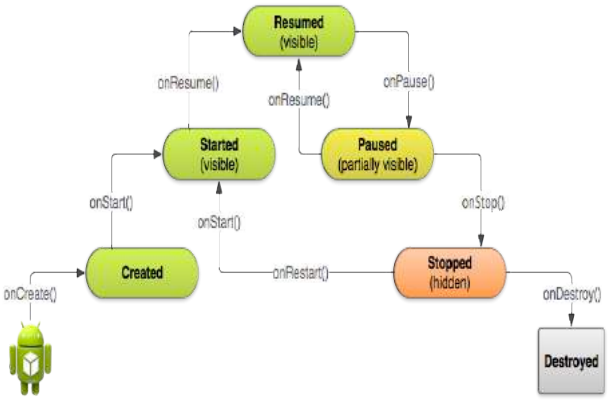
There are four states of activity, they are:

- **Running:** The activity is visible to the user and interactive.
- **Paused:** The activity is partially hidden but still running visibly and can be killed by the system at any time.
- **Stopped:** Activity is invisible but running and might be killed at anytime by the system.
- **Killed:** By calling finish() method, the system terminates the activity.

B) Lifecycle methods

When the following methods are called some operations are performed,

- **onCreate():** The activity is called
- **onResume():**The activity becomes visible and starts to interact with the activity again
- **onPause():** The activity becomes invisible, leaving space to some other activity.
- **onStop():** Called when no longer the activity is visible.



HISTORY OF ANDROID VERSIONS:

Code name	API level
✓ Alpha1	
✓ Beta	2
Internally known as (“Petit Four”)	
✓ Cupcake	3
✓ Donut	4
✓ Eclair	5-7
✓ Froyo	8
✓ Gingerbread	9-10
✓ Honeycomb	11-13
✓ Ice cream Sandwich	14-15
✓ Jellybean	16-18
✓ Kitkat	19-20
✓ Lollipop	21-22
✓ Marshmallow	23
✓ Nougat	24-25
✓ Oreo	26-27

SCOPE OF APP DEVELOPMET

Around the world mobile phones are used rapidly by huge crowd. Now-a-days it is impossible to imagine a world without mobile phones. With the mobile phones it is like all comforts are in our hand. There are lots of demands for advanced applications and updates.The mobile application development has just emerged so it has long way to vanish. It is a low cost app development platform so it has high scope. According to the global figure there are more than 85% smartphone running in android OS. Therefore there are advantage and scope in android development[4]. There are different OS[7](such as Apple, Windows 7) from which mobile application can be developed development. Till the utilization of mobile phones increase, the scope of android development increases.

CHALLENGES IN MOBILE APP DEVELOPMENT

All android applications packages are signed with certificate. The developer holds the private key of the certificate[5] since because it is authentic. The big challenge is that the app is not able to achieve high coverage. This is partial because it was designed for user input and cannot generate automatically such as login. Therefore the automated testing tool is unable to proceed. Another challenge is energy issue , since the state-of-the-art tool is not easily accessible to developers. The biggest challenge is that the app should be created on the demand in the in the market enterprise. The developer should look after whether the app have high performance, battery life, less memory utilizing, lack of transparency[5] by app store.

CONCLUSION

Development of android application has become very simple and easy. These applications are developed mainly to serve people with all the facilities in their hands and also to give entertainment by the means of gaming that includes lots of animations that interest them.Each and every day lots of apps are being created and updated. In future the demand in apps will be rapidly increasing and different advanced features will develop.Apps were created and will be created in an interactive manner since because artificial intelligence is also developing day by day rapidly. In future almost all mobile applications will have sensors, face recognization, etc. Though android application development has some challenges it has high scope and great market value. Developing android application is a easy task.

REFERENCES

[1] Josh Dehlinger and Jeremy Dixon, Department of computer and information sciences Towson university.

[2] Krithika.B, Prabhu.S, Vishalakshi.S, IT department,Sri Krishna arts and science college, Coimbatore.International Journal of trends in research development volume 2,ISSN 2394-9333.www.ijtrd.com

[3]MeiyappanNagappan, Department of software Engineering Rochester institute of techlogy,NY,USA International journal of computer trends and technology-volume3Issue3-2012 ISSN:2231-2803. http://www.internationaljournalssrg.org

[4] SuhasHollan and Mahima M Katti,Department of information science&engg,R.V college of Engineering Bangalore,India, International journal of computer applications(0975-8887) recent trends in future prospective in Engineering & Management Technology 2016.

[5] AbhinavKathuria et al, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.5, May- 2015, pg. 294-299

[6] Garimapandey (Noida) and DikshaDani(Ghaziabad) Department of computer science and engineering.

[7] Kirthika.B, Prabhu.S and Visalakshi.S, Android Operating System: A Review International Journal of Trend in Research and Development, Volume 2(5), ISSN 2394-9333 Sep - Oct 2015.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

Cryptography In Network Security: A Much Needed Technique

P.Srilakshmi
Department of Computer Applications,
S.S.S. Shasun Jain College for Women,
T. nagar, Chennai.

AparnaR.
Assistant Professor, Department of Computer Applications,
S.S.S. Shasun Jain College for Women,
T. nagar, Chennai.

ABSTRACT

In today's world we people have stated using the internet for many purposes be it for browsing, internet banking, E-mailing, social media and for many more things. But is our information safe and secure? Network security is very important and there are many algorithms which will help us to keep our data safe from people who hack into our system. In this paper we will study the different cryptographic algorithms which in keeping the data safe.

Keywords: cryptography,algorithms, cipher, encryption and decryption

I. INTRODUCTION

Network security is very important for the safety of our data because many of us use mobile phones and laptops (PDAs) in which we store most of our personal details like contacts, e-mails and even banking details in order to protect these files and information we use cryptography. The following are the most common algorithms which are used for network security[1] blowfish algorithm, RSA, RC4, data encryption standard, Diffie Hellman algorithm. These algorithms help us to keep the private data a secret. Cryptography encrypts[2] the details and only the person with the key will be able to access the information, the algorithms will help us to encrypt and decrypt the informations. In this paper we will study and compare these algorithms.

II. IMPORTANT TERMS IN CRYPTOGRAPHY

A. Cryptography

Cryptography is the method of keeping private information safe from other people. It is also known as cryptology. The practice and study of techniques for secure communication

B. Encryption

The process of converting plain text to a text which looks meaningless (cipher text) is called encryption.

C. Decryption

The process of converting the cipher text back to plain text (original text) is called decryption

D. Cipher(cypher)

Cipher is a group of algorithms which is used to encrypt and decrypt the messages and information. This is also called as cryptosystem.

E. Key

Key is a group of text which is used to encrypt and decrypt the message.

III. CRYPTOSYSTEMS AND ITS TYPES

Cryptosystems[3] are set ofalgorithms which are used to implement a security service. These cryptosystems are highly in demand to ensure security during data storage and transmission.

It consists of three process:

- Key generation

  Key generation is the process of generating keys for cryptography. The key is used to encrypt and decrypt data whatever the data is being encrypted or decrypted

- Encryption

  Encryption is the process of converting original information (called plaintext) into unintelligible text (called ciphertext).

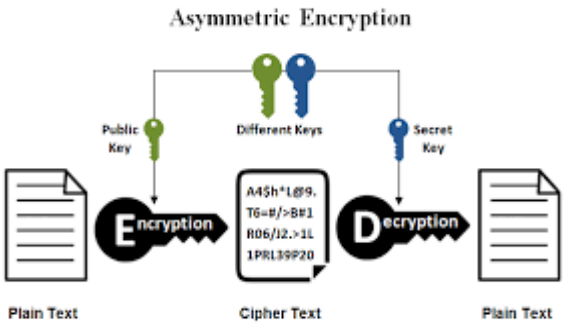
- Decryption

  Decryption is the process of transforming data that has been rendered unintelligible through encryption back to its original form.

This is of two types based the type of key used in Asymmetric cryptosystem[1] and symmetric cryptosystem [1].

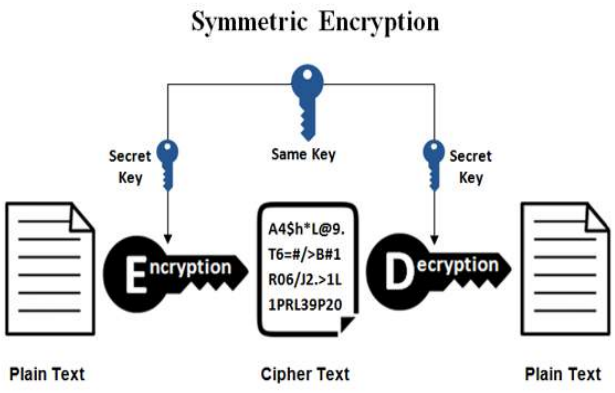
A. Asymmetric cryptosystem

This type of system uses one key for encryption and a different key for decryption. This is also known as public key cryptography. The keys used in this system are numbers. The keys are paired up. One key which can be shared is called public key, the other key which cannot be shared is called private key. Here only private key can decrypt the message. Some common algorithms are RSA, Diffie-Hellman key exchange.



B. Symmetric cryptosystems

This system uses the same key for both encryption and decryption. This is speeder and less difficult than asymmetric cryptosystem[2]. Since we use only one key the key is kept private and the information is kept safe. Most commonly used algorithms are AESand DES[4].



IV. ALGORITHMS

A. *RSA(Rivest Shamir Adleman)*

Ron Rivest, Adi Shamir, and Len Adleman created this RSA public key algorithm[2] in 1978. This is used to for signing and encryptionwithout exchanging the secret key. The key size should be more than 1024 bits for a good security. This algorithm is based on the "factoring problem"[3]. That is, it based on the difficulty of factorization[4] of product of twolarge prime number. The user creates two keys and releases the public key which is based on prime numbers.

B. *Diffie Hellman key exchange*

This algorithm is used for the secure transfer of the cryptographic keys. It is one of the practical examples of public key exchange[5]. It allows two people to have no previous information about each other to establish a secret key on an insecure network. It is used on many internet services such as speech synthesis [7].

C. *DES (Data Encryption Standard)*

This algorithm was developed by IBM. It is a symmetric cryptosystem. DES is used for the electronic encryption of data. It uses a 56-bit key[6]. It is considered insecure due to the less number of keys.

D. *AES (Advanced Encryption Standard)*

Its security is based on the intractability of certain discrete logarithm problems.*AES (Advanced Encryption Standard)*

This is also used for electronic encryption of data. It uses only one key. It has variable key length of 128/192/256 bits [7]. Each key could encrypt or decrypt a 128 bit data. AES is proven to be a relaiable algorithm.

Its security is based on the intractability of certain discrete logarithm problems.

V.APPLICATION AREAS OF CRYPTOGRAPHY

There are many applications which are currently being used. They are:

a. Secure Communication

This is used for communicating with other people securely. We can communicate such that the people trying to eavesdrop will not be able to do so. This is due to the use of public key that we are able to communicate peacefully.

b. Identification and Authentication

Identification is very important aspect. It is the process of verifying a person by cross checking with some proof like the ATM machines used a PIN number to identify the user. The same is applicable to cryptography the user and verified after which access is provided to that person.

c. Secret Sharing

This application allows us to share a secret with a group of people [8]. The actual secret is never disclosed.

d. E-commerce

This is a form of business conducted over the internet. Online shopping, booking tickets, transferring funds are a part of this. But giving credit cards away is not safe. Therefore, the card numbers are encrypted whenever it is entered and the information is thus secured.

VI.COMPARISON BETWEEN THE DIFFERENT ALGORITHMS

Algorithm	key size	Speed	Security
DES	56 bits	SLOW	INSECURE
AES	128,192, 256 bits	FAST	SECURE
RSA	1024 and above	FAST	SECURE
DIFFIE-HELLMAN	3072 BITS	FAST	SECURE

VII.CRYPTOGRAPHY IN DEFENCE

Cryptography is a huge asset to the military forces.With the help of enigma cipher the cryptanalysts attacked the Lorentz cipher.

In today's world, it is a necessity to keep the military[9] orders and plans a secret. The government is also providing funds for making the information secure [10]with the help algorithms. In 2010,stutex an elaborate computer worn was discovered.

Until recently cryptography has been of interest primarily to the defence and diplomatic personnel of governments, guarded over and directed by their national crypto logic services. The use of cryptography itself is not controlled i.e. sending an encrypted email or message, or making an encrypted phone call, is not subject to export control simply because it is encrypted.Hence,it is essential to focus more on strong algorithmic techniques to safeguard the information which carries a very secret message.

VIII.CONCLUSION

Network security has become a very important thing in today's world due to a growth in technology. This growth helps us in so many useful things like online shopping, e-commerce,e-banking,e-business but it also has its demerits like keeping our information safe from other people. Cryptography helps us in keeping this information safe. It has developed very rapidly in a short span of time. Cryptography is now-a-days used in many different fields like defense, air force and many in such places. Cryptography does not give 100% guarantee for keeping data safe at the same time the risk factor is reduced considerably.

IX. REFERENCES

[1] Prof. Mukund R. Joshi, RenukaAvinashKarkade, network security with cryptography

[2] RIVEST, R.L., SHAMIR, A., and ADLEMAN, L: 'A method for obtaining digital signatures and public-key cryptosystems', CACM, 1978, 21, pp. 120-126

[3] Dr. Sandeep Tayal1, Dr. Nipin Gupta2, Dr. Pankaj Gupta3, Deepak Goyal4, Monika Goyal5, A review paper on network security and cryptography.
[4] Coron, J. S. , " What is cryptography?", IEEE Security & Privacy Journal, 12(8), 2006, p. 70-73.
[5] R.Aparna ,Dr.P.l.Chithra,"A Review on Cryptographic Algorithms for Speech Signal Security" International Journal of Emerging Trends & Technology in Computer Science(IJETTCS), Volume 5, Issue 5, September - October 2016,pp 84-88.
[6] KritikaAcharya, ManishaSajwan, Sanjay Bhargava, Analysis of Cryptographic Algorithms for Network Security
[7] Aparna R.,Dr.P.L.Chithra,Role of Windowing Techniques in Speech Signal Processing For Enhanced Signal Cryptography,Chapter 28,Advanced Engineering Research and Applications.
[8] M. Shashanka and P. Smaragdis, "Secure sound classification: Gaussian mixture models," in Proc. ICASSP, vol. 3, Toulouse, France, 2006, p. 3.
[9] Laura Savu, Cryptography Role in Information Security, Recent Researches in Communication and IT,pg 36-41
[10] M. Quisquater, L. Genelle, and E. Prouff, "Thwarting higher-order side channelanalysis with additive and multiplicative maskings," in Cryptographic Hardwareand Embedded Systems 2011, Nara, Japan, 2011, pp. 240–255.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
RESEARCH PAPER

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

ENSEMBLE-OF-CLASSIFIERS APPROACH FOR DIAGNOSIS OF PERVASIVE DEVELOPMENTAL DISORDERS USING PSYCHO-METRIC PROFILES OF CHILDREN

Dr. Poorna B.
Principal
Shri S S Shasun Jain College for Women,
Chennai, India
e-mail: poornasundar@yahoo.com

Sumathi M.R.
Research Scholar, Department of Computer Science
Bharathiar University,
Coimbatore, India.
e-mail: sumathikuben@yahoo.co.in

ABSTRACT

Maintaining good health, mentally and physically, is important at every stage of life from childhood to adulthood to live a long and healthy life. But, due to various factors mental health problems have become common today than cancer, diabetes or heart disease. The onset of mental illness starts typically in the early stages of life i.e. infancy or childhood. The diagnosis of the onset of mental illness by the professionals is a complicated task as many factors are involved. An attempt has been made in this research to diagnose the early onset of pervasive developmental disorders from the psycho-metric profiles of children maintained by the psychologist. An ensemble-of-classifiers approach was used for the diagnosis. Individual classifier's balanced accuracies are used for weighing the classifiers and a final decision was made by averaging their predictions. This ensemble approach provided accuracy from 91% to 97%. Hence, this classifier may be used as an additional tool by the psychologists to diagnose the pervasive developmental disorder.

Keywords: Ensemble, Diagnosis, Pervasive Developmental Disorder, Weighted Average, Fuzzy Clustering

I. INTRODUCTION

Pervasive Developmental Disorders (PDD) refers to a group of conditions that involves delays in the development of many basic skills, ability to socialize with others, to communicate and to imagination. Children with these disorders are often confused in their thinking and generally have problems understanding the world around them. Typically, the disorder incepts before 3 years of age. So, it is difficult to diagnose the disorder. According to National Institute of Mental Health (NIMH), the symptoms of PDD include problems with communicating and interacting with others; unusual play with toys and other objects; difficulty with changes in routine or familiar surroundings; and repetitive body movements or behaviour patterns; little or inconsistent eye contact; failing to respond to someone calling their name. Other difficulties include sleep problems, digestion problems and irritability. But, the children may have above-average intelligence, ability to learn through visuals and audios, excellent mathematical and scientific knowledge, the ability to learn in detail and remember for long periods of time. The research suggested that genes and environment play important roles in the development of PDD. The PDD may first be identified by doctors in infants and toddlers by observing the child's behaviour and development. Diagnosing PDD in adults is not easy as the symptoms can overlap with symptoms of other mental health disorders such as schizophrenia or Attention Deficit Hyperactivity Disorder. Family history of PDD, premature or early birth with low birth weight may also have an impact in the development of PDD.

As the scientists do not know the exact causes of PDD and as the factors causing the mental health problems are overlapping, the diagnosis of PDD has become very difficult. Early diagnosis and treatment for PDD with proper care can reduce individual's difficulties and help them to learn new skills and make the most of their strengths. Although PDD is not curable, its symptoms can be addressed with appropriate interventions and many children with the disorder can be educated and integrated into community life.

The need for early identification has become more urgent by the accumulating evidence that intensive early intervention in optimal educational settings results in improved outcomes in speech and intellectual performance. For many years it was believed that individuals with PDD were not interested in human contact. They remain so aloof that it requires a great deal of effort to get a response. At the other end of the spectrum there are individuals who greatly enjoy and initiate social interaction, including hugging their parents and other shows of affection. Many children with the disorder may be restless because of an impairment of their imaginative and social skills. They do not know how to play with their toys and with other children meaningfully [1].

This research has made an attempt to diagnose the early onset of PDD using the psycho-metric profiles of children maintained by the psychologists. Machine learning techniques play a key role in predicting the mental health problems. Ensemble-of-classifiers approach is used to increase the predictive performance of individual classifiers.

I. RELATED WORK

A number of research works are going on in implementing machine learning techniques for predicting mental health problems. Nowadays, ensemble of classifiers is used to overcome the weaknesses of individual classifiers. The table 1 gives a sample list of ensemble of classifiers used for diagnosing the mental disorders.

Table I Literature Review on Ensemble of Classifiers in Mental health diagnosis

Sl.No.	*Year*	*Authors*	*Ensemble Technique used*	*Mental disorders diagnosed*
1	2007	Shen, Jess J. et al. [2]	Ensemble of three clustering methods	Three subtypes of PDD, namely Autism, PDD-NOS and Asperger's Syndrome

2	2012	Lin Manhua et al. [3]	Ensemble of Weak classifiers on subset of patches of brain images	Alzheimer's Disease with Mild Cognitive Impairment
3	2014	Farhan S. et al. [4]	Ensemble of SVM, MLP and J48 classifiers	Alzheimer's Disease
4	2014	Lebedev et al. [5]	Random Forest Ensemble technique	Alzheimer's Disease
5	2015	Gok et al. [6]	Ensemble of k-Nearest Neighbour Algorithms	Parkinson's Disease
6	2015	B. Ojeme et al. [7]	Ensemble of Bayesian Networks, Back Propagation Multi Layer Perceptron, Support Vector Machines, k-Nearest Neighbour algorithm and Fuzzy Logic	Depressive Disorders
7	2015	T. Latkowski et al. [8]	Ensemble using Rndom Forest	Autism Disorder
8	2016	Iftikhar M.A. and Idris A. [9]	Ensemble classification with SVM	Alzheimer's Disease with Mild Cognitive Impairment
9	2016	Husain et al. [10]	Random Forest ensembles	General Anxiety Disorder
10	2016	Ortiz et al. [11]	Ensemble deep learning architectures	Alzheimer's Disease
11	2016	Zhang et al. [12]	Multi-edit nearest neighbour and Ensemble Learning algorithm	Parkinson's Disease
12	2016	Vyskovsky et al. [13]	Random subspace ensemble method with Multi Layer Perceptron and SVM	Schizophrenia
13	2016	Z.A. Benselama et al. [14]	Ensemble of Sequential Minimization Optimization (SMO) algorithm, Random Forest and Feature-subspace aggregating approach (Feating)	Autism disordered speech
14	2017	Li et al. [15]	Combined Random Forest, SVM and Extreme Learning Machine algorithm	Parkinson's Disease
15	2017	Armananzas R. et al. [16]	Ensemble of different machine learning algorithms	Alzheimer's Disease
16	2017	Abou-Warda H. et al. [17]	Random Forest ensemble	Mental Disorders and Drug Abuse
17	2017	Lee E.S. [18]	Stacking based ensemble classifier of Logistic Regression, Decision Tree, Neural Networks, SVM and Naive Bayes networks	Depression

The sample list shows that a number of research works are going on in diagnosing the mental disorders like Alzheimer's, Parkinson's, Schizophrenia, Depressive disorders, etc. But, only a few attempts have been made to diagnose the mental health problems of children. This article has made an attempt to diagnose the Pervasive Developmental Disorder problem of children effectively using machine learning techniques. If the problem is diagnosed at an early stage, the intervention can be made and proper treatments can be provided to enhance the life of children.

Here, the classification models have been combined to reduce the model errors. Ensemble of classifiers is just like getting the advice of various experts and making a final decision based on the experts' opinions. Some of the advantages of ensemble classification are:

- Less noisy than single classification model.
- No room for over-fitting.
- Improvement in predictive accuracy.

II. MATERIALS AND METHODS

A. Dataset and its features

The dataset consisting of one hundred and thirteen psychometric profiles of children was collected from a clinical psychologist. The profiles are maintained by the psychologist in semi-structured text document format and these data are converted into attribute relation file format (.arff). The attributes considered important by the professionals were selected from the profile. This reduces sparseness of data. The name, address and other personal details that identify the child were excluded due to ethical reasons. The data were pre-processed in many ways. For example, numeric data like 'age' were converted into categorical data with four categories namely Infant, Early Childhood, Middle Childhood and Adolescent Childhood. The attributes whose values were missing, were filled with default values as prescribed by the psychologist. 46 attributes were represented from the psycho-metric profiles of children.

The age of the children ranged from 2 years to 16 years and there were 20 girl children and 93 boy children. 18 children had Autism Spectrum Disorder (ASD) and the rest 95 children did not have ASD. A glimpse of psycho-metric data as given by the psychologist is shown in Figure 1.

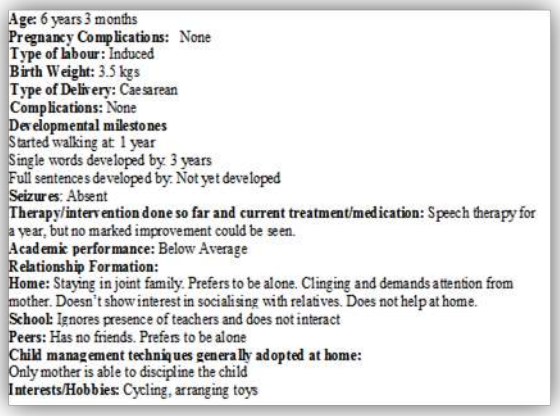

Age: 6 years 3 months
Pregnancy Complications: None
Type of labour: Induced
Birth Weight: 3.5 kgs
Type of Delivery: Caesarean
Complications: None
Developmental milestones
Started walking at: 1 year
Single words developed by: 3 years
Full sentences developed by: Not yet developed
Seizures: Absent
Therapy/intervention done so far and current treatment/medication: Speech therapy for a year, but no marked improvement could be seen.
Academic performance: Below Average
Relationship Formation:
Home: Staying in joint family. Prefers to be alone. Clinging and demands attention from mother. Doesn't show interest in socialising with relatives. Does not help at home.
School: Ignores presence of teachers and does not interact
Peers: Has no friends. Prefers to be alone
Child management techniques generally adopted at home:
Only mother is able to discipline the child
Interests/Hobbies: Cycling, arranging toys

Figure 1: Glimpse of the data as given by the psychologist

The semi-structured data was converted into tabular format i.e. .arff format, as in Table II.

Table II: Glimpse of dataset in .arff format

Academic Performance	*Affectionate to Others*	*Age*	*Being alone*	*Food Habits*	*Bowel Movement*	*Demands Attention of*	*Autism Spectrum Disorder (ASD)*
BA	Y	A	Y	I	R	Y	N
BA	Y	A	N	R	I	Y	N
A	Y	A	N	I	R	Y	N
A	Y	M	Y	I	R	N	N
BA	Y	M	N	R	R	N	N
A	Y	A	Y	R	R	N	N
A	Y	A	N	R	R	N	N
A	Y	M	N	R	R	N	N
A	Y	A	Y	I	R	N	N
A	Y	E	Y	R	I	Y	Y
A	Y	M	Y	I	R	N	N
A	Y	M	Y	I	R	Y	N
A	Y	M	N	R	R	Y	N
A	Y	I	N	I	I	N	N
A	Y	A	Y	R	R	N	N
A	Y	I	Y	R	R	Y	N
A	Y	I	Y	R	I	Y	N
BA	Y	A	Y	R	R	N	N

Note: A-Average; BA-Below Average; Y-Yes; N-No; I-Irregular; R-Regular

B. Feature Selection

Recursive Feature Elimination (RFE) method was applied to remove the attributes which are of less importance. Out of 46 features, 25 features were extracted and the experiment was done on the full feature set as well as on the reduced feature set. The details of the features are:

The Recursive Feature Elimination (RFE) algorithm was used to eliminate the less important features for the particular study. The twenty-five features that were extracted for diagnosing the ASD are mentioned below:

$$F_{RFE} = \{ F_1, F_3, F_5, F_7, F_8, F_{10}, F_{11}, F_{13}, F_{14}, F_{15}, F_{21}, F_{24}, F_{27}, F_{28}, F_{30}, F_{37}, F_{38}, F_{39}, F_{40}, F_{41}, F_{42}, F_{43}, F_{44}, F_{45}, F_{46}\}$$

C. Methodology

The methodology, using ensemble of classifiers, proposed in [19] was used for the early diagnosis of Pervasive Developmental Disorder. The methodology for diagnosing PDD using ensemble of classifiers has been shown in Figure 2. After final diagnosis is made, the ensemble model has been evaluated on various measures like Sensitivity, Specificity, Kappa-statistic value and Balanced Accuracy.

Table 3:Features collected from the psycho-metric profiles of children

S.No.	**Feature Name**	**S.No.**	Feature Name
F_1	Academic Performance	F_{24}	Moody
F_2	Affectionate	F_{25}	Has nightmares
F_3	Age	F_{26}	Mother had pregnancy complication
F_4	Aloof	F_{27}	Reading skill
F_5	Anxious	F_{28}	Completes school-work
F_6	Appetite	F_{29}	Has Seizures
F_7	Arithmetic Skill	F_{30}	Gender
F_8	Attention level	F_{31}	Sleeping Habit
F_9	Bowel Movement	F_{32}	Attracted to spinning objects
F_{10}	Concentration level	F_{33}	Stubborn
F_{11}	Demands attention of parents	F_{34}	Temper tantrums
F_{12}	Developmental delay	F_{35}	Under any medication
F_{13}	Distracted	F_{36}	Underactive
F_{14}	Maintains Eye-contact	F_{37}	Unusually loud
F_{15}	Psychiatric problem in family history	F_{38}	Whines/Screams
F_{16}	Fearful	F_{39}	Writing skill
F_{17}	Fidgets	F_{40}	Intelligence level
F_{18}	Fights with siblings/friends	F_{41}	Behavioural Emotional Problem
F_{19}	Friendly with elder children	F_{42}	Anxiety/Depression symptoms
F_{20}	Number of friends	F_{43}	Social/Language/Communication Deficit
F_{21}	Impulsive	F_{44}	Autism
F_{22}	Independent	F_{45}	Attention Deficit Hyperactivity Disorder
F_{23}	Listening skill	F_{46}	Pervasive Developmental Disorder

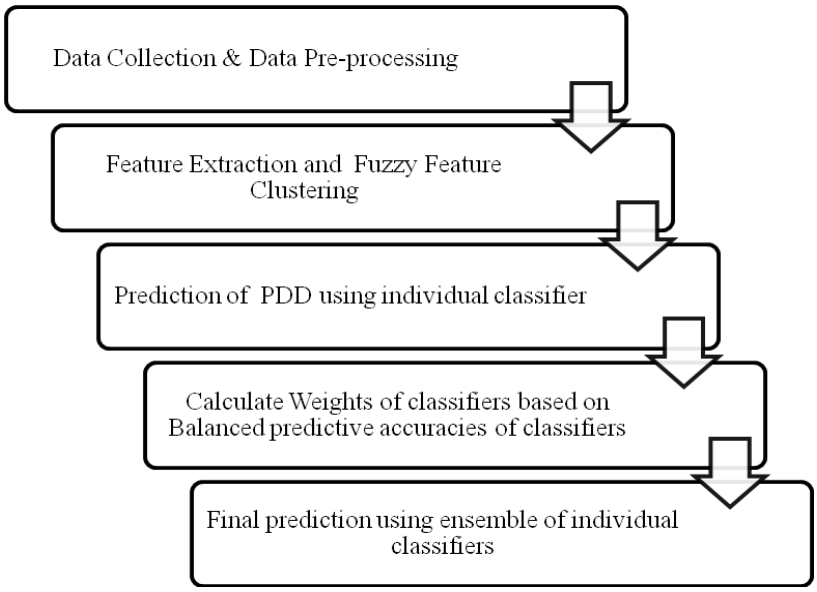


Figure 2: Ensemble methodology for diagnosing PDD

III. MODEL EVALUATION

The ensemble of classifier model had been evaluated on various measures like Sensitivity, Specificity, Kappa-value and Balanced Accuracy. Repeated *k*-fold cross validation is used to test the strength of the ensemble model. The comparisons were made on four models. The first two models used K-Medoids clustering method to cluster the features. The first model used full feature set and the second model used reduced feature set. The third and fourth models used K-Medoids fuzzy clustering method with full and reduced feature sets.

A. Performance Evaluation

A confusion matrix, also known as error matrix, was constructed to evaluate the performance of a machine learning model. In the present study, the sensitivity, specificity, kappa-value and balanced accuracy measures were calculated, from the confusion matrix, to evaluate the performance of the ensemble model. Table 4 shows the confusion matrix for diagnosing the PDD.

Table IV : Confusion Matrix

	Diagnosis made by the psychologists		
	Total Population	***With ASD***	***Without ASD***
Prediction made by the model	With ASD	True Positive (TP)	False Positive (FP)
	Without ASD	False Negative (FN)	True Negative (TN)

a. Sensitivity

Sensitivity, also called as true positive rate, recall or probability of detection, is a statistical measure that measures the proportion of positives which are correctly identified as positives by the classifier. A high sensitivity value indicates that the model will recognize all children with the disorder by testing positive. The formula for computing sensitivity is:

$$Sensitivity = \frac{\text{Number of true positives (TP)}}{\text{Number of true positives (TP)} + \text{Number of False Negatives (FN)}}$$

$$= \frac{\text{Number of children predicted with the ASD}}{\text{Total number of children truly affected by the ASD}}$$

b. Specificity

Specificity is related to the model's ability to correctly reject the healthy cases without a condition, i.e., it correctly rejects the children without the ASD disorder. It measures the proportion of healthy children known not to have the disorder, will also be tested negative by the model. High specificity value specifies that the model accurately exclude the children with ASD from the children without ASD. The formula for computing specificity is:

$$Specificity = \frac{\text{Number of true negatives (TN)}}{\text{Number of true negatives (TN)} + \text{Number of False Positives (FP)}}$$

$$= \frac{\text{Number of children predicted without ASD}}{\text{Total number of children truly not affected by the ASD}}$$

c. Kappa statistic value

Kappa statistic is a measure of agreement between the predictions and the actual labels. I can also be interpreted as a comparison of the overall accuracy to the expected random chance accuracy. The higher value of kappa statistic is preferred, The formula for calculating kappa statistic measure is :

$$\text{Kappa statistic value} = \frac{\text{Observed Accuracy} - \text{Expected accuracy}}{1 - \text{Expected accuracy}}$$

d. *Balanced Accuracy*

The conventional accuracy is high when the classifier takes advantage of an imbalanced data set and it is purely because of chance. To avoid this for an imbalanced data set, the balanced accuracy is used instead of conventional accuracy. Balanced accuracy is the average accuracy obtained on either class. From the confusion matrix, the balanced accuracy is given by:

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right)$$

If the classification model performs equally well on either class, the balanced accuracy reduces to the conventional accuracy.

B. Results

made between ensemble of classifiers using majority voting and ensemble of classifiers using Weighted Average. The values

The model has been evaluated on four measures Sensitivity, Specificity, Kappa Statistics Values and Balanced Accuracies and a comparison has been have been given in Table 5 and a graphical representation has been made in Figure 3.

Table V: Predictive performance of Ensemble of Classifiers

	Sensitivity		***Specificity***		***Kappa-Value***		***Balanced Accuracy***	
Attribute Set	***Majority Voting***	***Weighted Average***	***Majority Voting***	***Weighted Average***	***Majority Voting***	***Weighted Average***	***Majority Voting***	***Weighted Average***
K-Medoids Clustered - Full	0.96	0.96	0.45	0.88	0.48	0.79	0.71	0.92
K-Medoids Clustered - Reduced	0.96	0.94	0.27	1.00	0.29	0.81	0.61	0.97
Fuzzy K-Medoids Clustered - Full	0.93	0.94	0.73	0.88	0.66	0.74	0.83	0.91
Fuzzy K-Medoids Clustered - Reduced	0.94	0.94	0.88	1.00	0.74	0.81	0.91	0.97

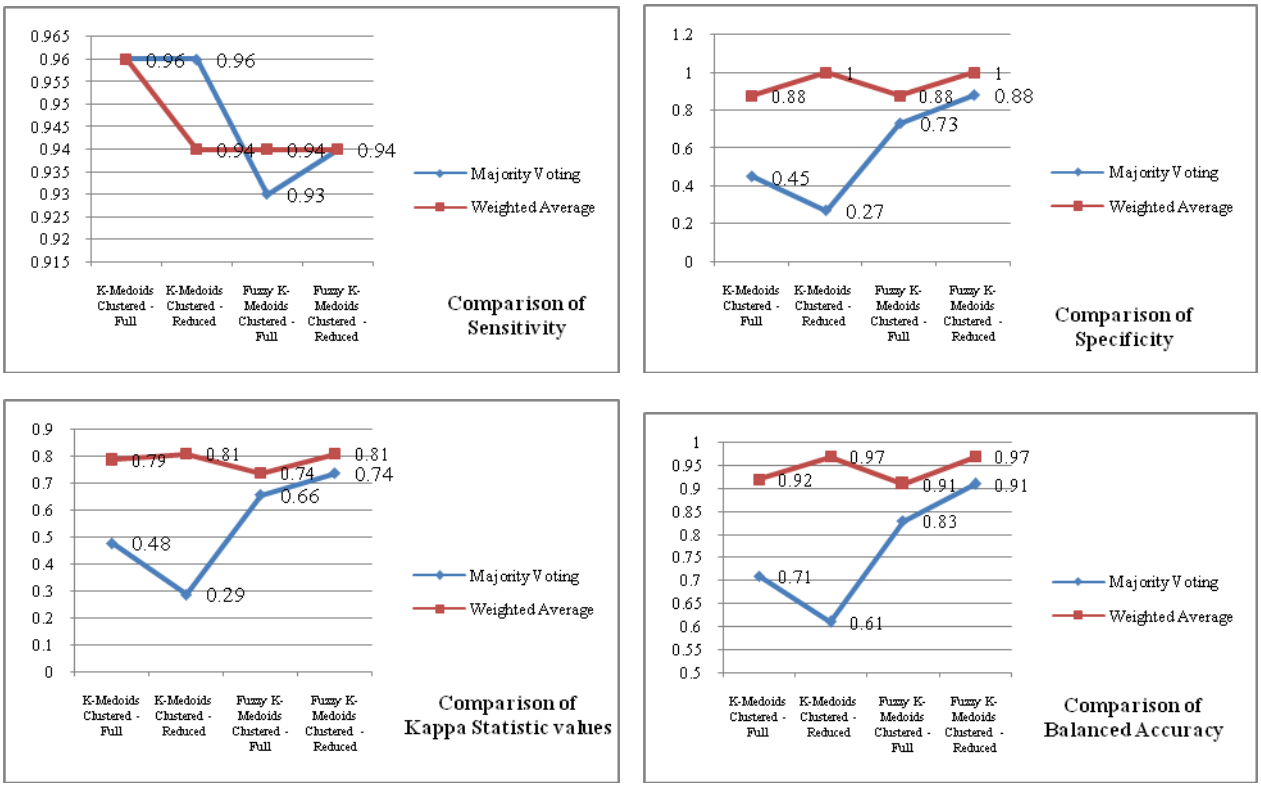


Figure 3: Evaluation of Ensemble of Classifiers

IV. DISCUSSION

The predictive performance of the ensemble of classifiers using Majoirty voting and Weighted average have been shown in Table V. The comparison has been made on full feature set and on the reduced feature set. Two clustering techniques namely *k*-Medoids and Fuzzy *k*-Medoids have been employed to cluster features. The comparison of the ensemble of classifiers shows that Weighted Average based Ensemble of classifiers is effective on various measures like Specificity, Kappa-statistics and Balanced Accuracy. According to sensitivity, there is only a slight difference between the both the methods.

V. CONCLUSION

Mental health is important at every stage of life from childhood to adulthood. Mental health problems have become common today among children. Mental health diagnosis is a challenging task as a number of factors are involved. The onset of mental illness starts typically in the early stages of life i.e. infancy or childhood. This research has attempted to diagnose the early onset of pervasive developmental disorders from the psycho-metric profiles of children maintained by the psychologist. An ensemble-of-classifiers approach was used for the diagnosis. Individual classifier's balanced accuracies are used for weighing the classifiers and a final decision was made by averaging their predictions. This ensemble approach provided accuracy from 91% to 97%. Hence, this classifier may be used as an additional tool by the psychologists to diagnose the pervasive developmental disorder.

VI. ACKNOWLEDGEMENT

We thank Dr. Sangeetha Madhu, Clinical Psychologist, Chennai Institute of Learning and Development (CHILD) Centre, Chennai, for providing data, insight and expertise throughout the study of the research article.

VII. REFERENCES

[1] Autism Spectrum Disorders – A Guide for Paediatricians in India Merry Barua and Tamara C. Daley – AAHAN – Action for Autism – 2008

[2] Shen, Jess J., et al. "Using cluster ensemble and validation to identify subtypes of pervasive developmental disorders." *AMIA Annual Symposium Proceedings*. Vol. 2007. American Medical Informatics Association, 2007.

[3] Liu, Manhua, et al. "Ensemble sparse classification of Alzheimer's disease." *NeuroImage* 60.2 (2012): 1106-1116.

[4] Farhan, Saima, Muhammad Abuzar Fahiem, and Huma Tauseef. "An ensemble-of-classifiers based approach for early diagnosis of Alzheimer's disease: classification using structural features of brain images." *Computational and mathematical methods in medicine*2014 (2014).

[5] Lebedev, A. V., et al. "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness." *NeuroImage: Clinical*6 (2014): 115-125.

[6] Gök, Murat. "An ensemble of k-nearest neighbours algorithm for detection of Parkinson's disease." *International Journal of Systems Science* 46.6 (2015): 1108-1112.

[7] B. Ojeme, M. Akazue & E. Nwelih (2015): Automatic Diagnosis of Depressive Disorders using Ensemble Techniques. Afri J Comp & ICTs Vol 8, No.3 Issue 2 Pp 31-38. 1.

[8] Latkowski T, Osowski S. Computerized system for recognition of autism on the basis of gene expression microarray data. Comput Biol Med 2015. Jan;56:82-88. 10.1016/j.compbiomed.2014.11.004

[9] Iftikhar, Muhammad Aksam, and Adnan Idris. "An ensemble classification approach for automated diagnosis of Alzheimer's disease and mild cognitive impairment." *Open Source Systems & Technologies (ICOSST), 2016 International Conference on*. IEEE, 2016.

[10] Husain, Wahidah, Lee Ker Xin, and Neesha Jothi. "Predicting Generalized Anxiety Disorder among women using random forest approach." Computer and Information Sciences (ICCOINS), 2016 3rd International Conference on. IEEE, 2016.

[11] Ortiz, Andres, et al. "Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease." International journal of neural systems 26.07 (2016): 1650025.

[12] Zhang, He-Hua, et al. "Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples." Biomedical engineering online 15.1 (2016): 122.

[13] Vyškovský, Roman, et al. "Random subspace ensemble artificial neural networks for first-episode schizophrenia classification." Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on. IEEE, 2016.

[14] Benselama, Zoubir Abdeslem, et al. "Ensemble classification methods for autism disordered speech." Med. J. Model. Simul 6.001 (2016): 011.

[15] Li, Yongming, et al. "Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms." Journal of Medical Imaging and Health Informatics 7.2 (2017): 444-452.

[16] Armañanzas, Rubén, Pedro Larrañaga, and Concha Bielza. "Ensemble transcript interaction networks: A case study on Alzheimer's disease." Computer methods and programs in biomedicine 108.1 (2012): 442-450.

[17] Abou-Warda, Horeya, et al. "A Random Forest Model for Mental Disorders Diagnostic Systems." International Conference on Advanced Intelligent Systems and Informatics. Springer, Cham, 2016.

[18] Lee, Eun Sung. "Exploring the Performance of Stacking Classifier to Predict Depression Among the Elderly." Healthcare Informatics (ICHI), 2017 IEEE International Conference on. IEEE, 2017.

[19] Sumathi, M. R., and B. Poorna. "Design and Development of Ensemble of Naïve Bayes Classifiers to predict Social and Communication Deficiency among Children." International Journal of Applied Engineering Research 12.24 (2017): 14190-14198.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

SHA GYAAN - A MOBILE APP FOR ENHANCING THE LEARNING CAPABILITY OF CHILDREN WITH HEARING AND SPEECH IMPAIRMENT

Dr. M. Anita Indu
Assistant Professor
Department of Computer Science
SSS Shasun Jain College for Women, Chennai.
e-mail: anitasaravanan78@gmail.com

Ms. N.M. Kavitha
Assistant Professor
Department of Computer Science
SSS Shasun Jain College for Women, Chennai.
e-mail: kavikarup@gmail.com

ABSTRACT

It has been observed that people with disability display lower education achievements when compared to normal peer groups. In the world of technology, the learning skill among differently abled students can be enhanced through information technology. People with disabilities have to recognize, understand, navigate and interact with IT so that they can participate equally in the economic and social aspects of society. The technology has been constantly used from mobile phones, to iPads, to laptops and TVs. Mobile devices provide accessibility features, flexibility, portability and customizability for people with disabilities (PwD). As a result of using mobile devices, Mobile Apps have been designed for people with disabilities, which help them connected with the world.

In this paper, survey of various mobile apps were discussed which helps in enhancing the learning skill of PwDs. A special mobile app named **SHA GYAAN** is proposed especially for hearing and speech impaired children. Keeping the needs of students, this mobile application enhances the learning capability of the children between the age group of 3 to 8. This app is intended to facilitate the joyful learning of the children and to evolve and build their learning skills. The major areas covered in this mobile app are Foundation language (Tamil & English), Environmental Science and Mathematics.

Keywords: Persons with Disabilities (PwD), Assistive device, Mobile Application, Information Technology, AAC (Augmentative and Alternative Communication), Hearing and Speech Impairment.

1. INTRODUCTION

Disability is an impairment that may be cognitive, developmental, intellectual, mental, physical, sensory, or some combination of these. It substantially affects a person's life activities and may be present from birth or occur during a person's lifetime. The first ever World report on disability which was produced jointly by WHO and the World Bank, explains that more than a billion people in the world today experience disability [1]. People with disabilities have generally poorer health, lower education achievements, fewer economic opportunities and higher rates of poverty than people without disabilities. To address the issues of lower education achievements and economic opportunities, the use of IT plays a vital role. Accessibility of Internet and other communication technologies became easier after the usage of smart phones by people of all income groups. Mobile devices not only provide a platform for communication but can also assist disability people with their daily tasks. For example, a blind person needs a talking GPS device (£750), a talking notetaker (£1500), a talking MP3 player (£250), a talking barcode scanner (£100) and many, many more specialist devices [2]. All of that had to be carried around in a backpack, each with its own charger. Smart phones provide all the above functionality in one device and are almost infinitely expandable with each new app or service that comes along.

2. REVIEW OF LITERATURE

There are many mobile apps designed to improve the learning skill of differently abled students in various fields.

2.1 VAAKYA

The picture based Augmentative and Alternative Communication (AAC) app designed for people with speech impairment is Vaakya. This app also helps students affected by autism, cerebral palsy and various other mental and physical conditions. The app is an AAC tool and can be used to practice during rehabilitation. It works as an effective tool for individuals who are unable to read or to communicate, as it depends on images and audio instead of text [3].

2.2 AVAZ

AVAZ app helps the children with learning disabilities. Large number of images is organized in such a way that users find it useful to communicate. The app supports multiple languages, both Indian and European, and users can also add their own 'words' - for example, a picture of your grandmother, along with the word grandma [4].

2.3 FREESPEECH

FreeSpeech, is the extended version of Avaz. FreeSpeech is a learning tool to teach the rules of grammar using semantic map. The app presents words like building blocks which can move them around, and it predicts the words and prompts you to expand the sentences. The words can be assembled into a grammatically correct sentence and helps teach abstract concepts like tenses [5].

2.4 LEARN WITH RUFUS

LEARN WITH RUFUS is an app for young learners and special needs students. This app is mainly

designed to identify facial expressions and recognizing emotions of others. It also includes the entertaining and many engaging activities to make students active [6].

2.5 MONTESSORI NUMBERS

Montessori Numbers is the mathematical app used for the students to understand the relationship between quantities and the numbers. It builds basic math competencies and introduces numeric order, the decimal system, counting up to 1000, comparing quantities, addition and subtraction. Additionally, it can pronounce numbers for better understanding and memorizing [7].

2.6 PROLOQUO2GO

Proloquo2Go is an AAC solution for students suffering with autism, cerebral palsy and brain injury. The main aim of this app is to give children and adults with speech impediments a voice. The visual vocabulary in the app helps to create sentences for communication. The app is flexible and customizable and allows choosing from a range of realistic accents for children and adults to match their "inner voice" [8].

2.7 VIDEO SCHEDULER

Video Scheduler is a visual schedule app with video model features. It allows creating checklists of steps necessary for achieving the goal or completing the task. This app helps ASD students and for students struggling with time and task management [9].

2.8 SOUNDING OUT MACHINE

The Sounding out Machine app is beneficial for learners who struggle with decoding. It sounds out difficult words and models how to pronounce them syllable by syllable. The students can take a snap of page and this app helps with challenging words in that page. There is also a typing mode, where a student can type in a particularly puzzling word [10].

2.9 SUPERWHY

SuperWhy offers interactive literacy games and engaging activities with words, letters, rhyming and spelling that improve reading and writing skills. Exercises with filling the gaps and choosing an ending to a story help to solve the particular problem.

The review of study shows that there are many apps specially designed for the differently abled students. These apps mainly includes mathematics and English to help them to solve simple problems using arithmetic operation, work with number pattern, simple games to solve day to day calculation, pronouncing the words, grammar, building sentences and so on. These apps were designed to improve the learning skill of children in a particular subject. Hence tutors or parents have to refer various apps to teach different subjects. This limitation has to be reduced in such a way that the four fundamental subjects (Tamil, English, Mathematics and Environmental Science) should be covered in a single mobile app for the benefit of students, parents and tutors [11].

3. PROPOSED STUDY

Parents of differently abled children want to admit their children in reputed school so that the students mingle easily with normal sighted peers. To improve their preschool activities, they should be encouraged to learn new skills in an enjoyable way. Understanding the concern of parents, the proposed study is aimed to design and develop the mobile application "SHA GYAAN" for enhancing the learning capability of the children with disability between the age group of 5 to 8. This app is intended to facilitate the joyful learning for the children and to evolve and build their learning skills. The initiatives taken in this app are:

- To express themselves
- To enable faster communication
- To workout activities with readiness
- To keep the content as close as feasible to local conditions and culture
- To keep the lessons child-friendly and allow them to enjoy learning
- To enable the use of languages in real life situations so that the language introduced is meaningful

The architecture of SHA-GYAAN is designed as in Fig 1. This is an android based app stored in Google Play Store and can be downloaded and accessed in any android smart phone. The scores of different activities taken up by the students will be stored in the local database and can be used to measure the improvement in their learning capability.

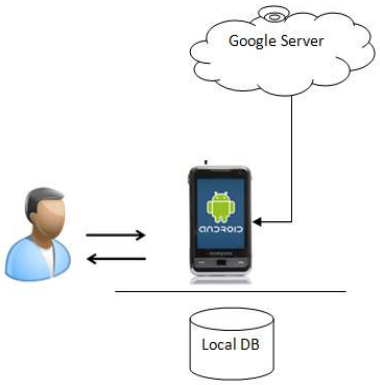


Figure 1 : SHA – GYAAN Architecture

The major areas covered in this mobile app are Foundation language (Tamil & English), Environmental Science and Mathematics.

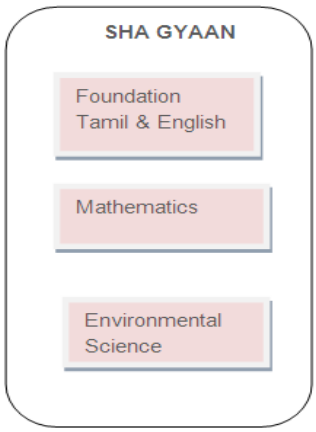


Figure 2: Subjects covered in SHA GYAAN

3.1 FUNDAMENTAL LANGUAGE

The fundamental language is used to teach alphabets and words and the activities in each lesson help language development by developing

- The four skills of listening, speaking, reading and writing through frequent repetition, picture support and gaming activities
- Vocabulary
- Awareness of language structure

This topic teaches alphabets (with vowels and consonants), Word formation, basic grammar, learning the object names with pictures along with the related gaming activities. To make the student to understand the writing pattern, the app provides "Learning by Overwriting" technique. This improves the writing capability of the children in an easy and efficient way. The grammar exercises like singular, plural, nouns, verbs, adjectives and prepositions are designed in a play way method. Innovative activities are included to improve the creative and communication skills of the child.

3.2 MATHEMATICS

Hearing impaired student's performance on problem solving tasks and word problems falls below that of their hearing counterparts [12]. Students who struggle with mathematics learning regardless of their mathematical knowledge prior to starting school exhibit the following characteristics:

- Demonstrate slow or inaccurate recall of basic arithmetic facts;
- Difficulty in representing mathematical concepts mentally;
- Poorly developed number sense; and
- Difficulty in storing information in their working memory.

The American National Council of Teachers of Mathematics has recommended Strategies for students with hearing impairment and their teachers. The learning strategy included technology usage, an approach and attack response to problem solving and use of diagrams, pictures, charts, mental images and dealing with vocabulary issues [13]. Keeping the above strategy in mind, the SHA GYAAN app is enriched by picture sequences, number games, life-oriented mathematics and gaming activities.

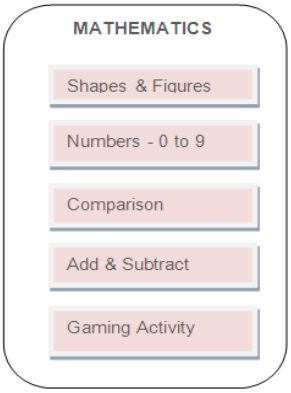


Figure 3: Activities for Mathematics

All these facilitate the learning of shapes, figures, numbers, measurements, patterns and data. The children are enabled to explore the mathematical concepts by touching, seeing, listening, practicing, talking and coloring. The Mathematical exercises given in the app related to the understanding of relationship between numbers and their basic operations provides opportunities for the children to imbibe the concepts and express them.

3.3 ENVIRONMENTAL SCIENCE

Children's appreciation for nature develops at a young age. So, environmental education plays a vital role in the life of kids. They watch and imitate the attitudes and actions of the adults around them towards living things. Science is a part of everyday life and should be learnt through pictorial representation with sounds. The children always enjoy nature around them – Animals, birds, Trees, sky, rain, etc. This category provides activities which are based on their experience in the outside world. The children should be aware that apart from the family members, many people help them in daily life. This app helps them to identify the persons with different jobs and build respect for their work. It includes playful games or forms of visual exercises that will excite the young minds and capture their interest.

4. CONCLUSION

Technology has created a revolution in possibilities for disabled learners. For years, differently-abled students have struggled with their assignments or been shut out of different classes or subjects because schools had accessibility or instructional problems. Assistive Technology is any item, piece of equipment, or product system that is used to increase, maintain, or improve functional capabilities of individuals with disabilities and the future of the educational system is practically determined by the development of technology. The teaching strategies designed in this app facilitate the hearing and speech impaired children to improve their capacity, productivity, and performance. Technology integration in education inspires positive changes in teaching methods. Parents or tutors can organize their time in a way that works for them, and they can easily help the children to gain the knowledge they are interested in. The four skills of listening, reading, writing and speaking can be improved in a better way by taking up the tasks provided in this app. This app gives the joy of learning through fun based activities and enhances the learning capability of children by the way of indianize education. This app is meant to create and offer content that is representative of the Indian ethos, mindset, practices and heritage.

5. REFERENCES

[1] World Report on Disability 2017. http://www.who.int/disabilities/world_report/2011/report.pdf

[2] Abdi Hassan, et. al(2017), A Smartphone and People with Disabilities: The Power and The Promise. Retrieved from https://digitalinclusion.blog.gov.uk/2017/01/24/smartphones-and-people-with-disabilities-the-power-and-the-promise/

[3] Jyoti Arora(2017), Vaakya : App To Help People With Speech Impairment. Retrieved from https://technotreats.com / 2017/02/18/vaakya/

[4] Retrieved from https://en.wikipedia.org/wiki/Avaz_app

[5] Retrieved from https://itunes.apple.com/in/app/avaz-freespeech/id1047763490?mt=8

[6] Learn with Rufus: Feelings and Emotions https://www.autismspeaks.org/autism-apps/learn-rufus-feelings-and-emotions

[7] http://lescapadou.com/LEscapadou_Fun_and_Educational_applications_for_iPad_and_IPhone/Montessori_Numbers_Math_Activities_for_Kids_on_iPad_and_iPhone.html

[8] http://learningworksforkids.com/apps/proloquo2go/

[9] Carol,L.H & et.al (2013),Video Scheduler – A Top Ten App for Special Education

[10] http://www.smartappsforkids.com/2016/02/review-avaz-freespeech-is-a-fantastic-new-learning-tool-its-our-newest-5-star-app.html

[11] https://en.wikipedia.org/wiki/Super_Why!

[12] Rochester Institute of Technology RIT Scholar Works Theses Thesis/Dissertation Collections 9-21-2005 Deaf students and problem solving in mathematics Heather Maltzan

[13] https://www.tes.com/teaching-resource/strategies-for-teaching-math-to-deaf-hard-of-hearing-students-6002506

[14] http://www.nctm.org/ Research-and- Advocacy / Research-Brief-and-Clips/ Learning-Difficulties-in-Mathematics/

[15] Jitka Vitova, et.al (2013), "Successes of students with hearing impairment in math and reading with comprehension", International Conference on Education & Educational Psychology 2013 (ICEEPSY 2013), Volume 112, Pages 725-729

[16] Jitka Vitova, Jana Balcarova (2012),"Language Competence versus the Mathematical Concepts of Pre-School children with Hearing Impairment", International Conference on Education & Educational Psychology 2013 (ICEEPSY 2012)

[17] Larry Medwetsky(2015),"Mobile Device Apps for People with Hearing Loss".

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

Combining Internet of Things and e-Learning Standards to Provide Pervasive Learning Experience

Dr. Mrs. S. Prasanna
Assistant Professor, Department of Computer Applications,
Shri Shankarlal Sundarbai Shasun Jain College for Women,
T. Nagar, Chennai 600 017
Email: s.prasana@shasuncollege.edu.in

ABSTRACT

(IoT) is the new technology developing in the recent years is quickly in the computing world. In Internet of Things, every day devices become smarter every day processing becomes intellectual day by day, and communication becomes very informative. This has transformed the way people interrelate and Internet of Things shaped a radical change in the field of education and has created new forms of communication between teachers and students. This lay concrete on the enhancement in the teaching and learning process and develop the perspective in which students learn. The incorporation of objects to the Internet leads to modernization that could assist the teaching-learning process. IoT provides more attractive learning atmosphere for students and more information about the learning process to aid teachers to augment their knowledge about the learning pace of their students and their learning complications. So, IoT and eLearning will have to be used to instruct more people in ICT as well as other domains.

Keywords: Internet of things, eLearning, ICT, Communication

I. INTRODUCTION

Internet of Things (IoT) is the arrangement of all kinds of things entrenched with sensors, electronics, software, and so on, connected to the Internet, based on the International Telecommunication Union's Global Standards Initiative [1, 2]. Internet of Things (IoT) is a element of the potential Internet which comprises of billions of sensor-based and actuator based smart devices, with data-processing capability [3]. According to the Gartner statement, over 26 billion of devices will have been associated to the Internet by the end of 2020 [4]. This paper will discuss utilities of IoT, architecture of IoT, six skills for IoT applications, IoT in eLearning and instructional design, Internet of Learning Things, IoT potentials to renovate education, and IoT to progress student performance.

II UTILITIES OF IoT

IoT may be categorized as the owner of key utility factors as below [5].

- Dynamic and self adapting: IoT devices and systems should have the capability to dynamically adapt with the varying contexts and take actions based on their functioning conditions, user's context, or sensed environment.
- Self-configuring: IoT devices may have self-configuring ability, allowing a large number of devices to work mutually to provide positive functionality. These devices have the facility to configure themselves in alliance with IoT infrastructure, setup the networking, and obtain latest software upgrades with smallest physical or user interference.
- Interoperable communication protocols: IoT devices may maintain a number of interoperable communication protocols and can communicate with other devices and also with the infrastructure.
- Unique identity: Each of IoT device has a unique identity and unique identifier such as IP address or URI. IoT systems may have intellectual interfaces which become accustomed based on the context, allow communicating with users and environmental contexts. IoT device interfaces let users to query the devices, observe their status, and manage them remotely, in association with the control, configuration and management infrastructure.
- Integrated into information network: IoT devices are typically integrated into the information network that permits them to communicate and exchange information with other devices and systems. IoT devices can be dynamically discovered in the network, by other devices and/or network, and have the potential to illustrate themselves to other devices or user applications.
- Context-awareness: Based on the sensed information about the physical and environmental parameters, the sensor nodes gain knowledge about the surrounding context. The decisions that the sensor nodes take thereafter are context-aware [6].
- Intelligent decision making capability: IoT multi-hop in nature. In a large area, this feature enhances the energy efficiency of the overall network, and hence, the network lifetime increases. Using this feature, multiple sensor nodes collaborate among themselves, and collectively take the final decision.

.

III THE ARCHITECTURES OF IoT

The two IoT architectures are (i) 3–layer architecture and (ii) 5-layer architecture.

3.1 THE 3-LAYER ARCHITECTURE

This comprises of three layers which are called perception, network, and application. The principle of perception layer is to recognize each object in the IoT system. This is done by collecting information about every object. This layer have RFID tags, sensors, cameras, etc. The second layer is the network layer. The network layer is the hub of the IoT. It sends the data collected by the perception layer. It comprises the software and hardware instrumentations of internet network along with the management and information centers. The third layer is the application layer. The application layer's objective is to congregate between the IoT social needs and industrial technology [7].

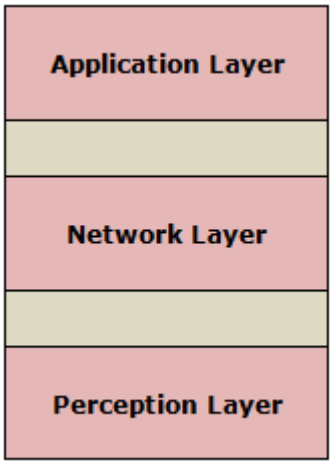


Fig 1: The IoT 3 – Layer Architecture

3.2 THE 5-LAYER ARCHITECTURE

The 3-layer architecture is not adequate due to the expected IoT development. Consequently, 5-layer architecture is proposed.

The primary layer is called business. The idea of this layer is to classify the IOT applications charge and management. It is also accountable for the user's privacy and all research associated to IOT applications. The second layer is called application. The objective of this layer is to resolve the types of applications, which will be used in the IoT. It also extend the IOT applications to be more intelligent, authenticated, and safe. The third layer is called processing. Its responsibility is to handle the information gathered by perception layer. The handling process comprises storing and analyzing. This layer utilizes method such as database software, cloud computing, ubiquitous computing, and intelligent processing in information processing and storing.

The fourth layer is called transport. It transmits and receives the information from the perception layer to the processing layer and vise versa. It contains many expertise such as infrared, Wi-Fi, and Bluetooth. IPV6 is used for addressing. The fifth layer is called perception. The objective of this layer is to define the physical meaning of each object in the IoT system such as locations and temperatures. It also gathers the information about each entity in the system and convert this data to signals. Technologies like RFID and the GPRS are adopted in the layer [8].

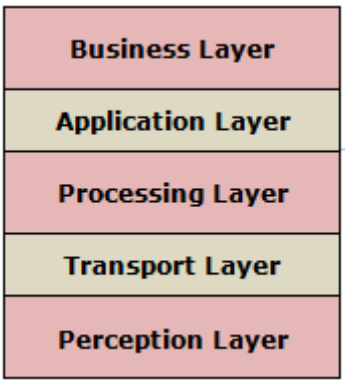


Fig2: The IoT 5 **– Layer Architecture**

IV. IoT IN ELEARNING AND INSTRUCTIONAL DESIGN

The article "IoT and e-Learning" [9] presented the potential ways to leverage IoT in eLearning. It also presented a good impact of the Internet of Things on eLearning. The activity tracker would send data about an employee's erroneous action back to the company's learning management system (LMS). The LMS then automatically assign a refresher course on safety procedures for that employee. It also stated that IoT can enhance eLearning in improving completion, reducing costs, and improving learning outcomes.

One of the most important part of eLearning is "Instructional Design" and IoT can definitely used in Instructional Design. The Internet of Things affect the instructional design in 1) Spatial information (place), 2) Temporal information (time), and 3) Persistence information (history). The three features of information should be used in instructional design to show the place and time of the student taking the course and compare with the historical data to select the most appropriate part of the learning to be done next, either in providing repeated learning of the previous topic or in proceeding to the next topic.

V. USEFUL E-LEARNING STANDARDS

5.1 IMS Learning Design (IMS-LD)

The IMS LD specification is developed by IMS GLC (IMS Global Learning Consortium) in 2003. It is the only available interoperability specification in the area of technology enhanced learning that allows the definition and orchestration of complex activity flows and resource environments in a multi-role setting . The IMS LD is based on the principle that considers a Learning process as a play metaphor [10]. Each person has a role and performs a set of activities. A method is the main element of an LD scenario. It helps coordinating the activities of each role. It consists of one or more play(s), acts and role-parts. A rolepart contains a reference to a role and a reference to a particular structured activity. An activity is either simple or composed. It has a learning or a supporting purpose.

5.2 IEEE Learning Object Metadata (IEEE LOM)

Learning Object Metadata [11] is an e-Learning standard published in 2002. It is considered as the most adopted content tagging specification. It is used for learning object annotation using metadata. LOM proposes classification of metadata elements into nine groups: General, lifecycle, metametadata, technical, educational, rights, etc. The advantages of adding metadata to learning objects aim to ensure interoperability and re-use of LOs, adaptability as well as sustainability.

VI. INTERNET OF LEARNING THINGS

The use of IoT in "learning" is called "Internet of Learning Things". From the article "Internet of Learning Things" [12], students and teachers would be taught to measure and share data – using new Internet of Things technology – in ways that help make learning fun, link directly to the curriculum, and ultimately inform the design of the next generation of schools". As an example of "Internet of Learning Things'', the Parrot AR.Drone2.0 enables students to survey an area using a mobile phone. HD video is shot and stored on a USB memory stick, or relayed directly back to the phone. In one package, Science (e.g. physics of flight); Technology (e.g. OS, networking, control); and Geography (e.g. surveys, observations) can be delivered, in a way that is completely engaging for children of all ages.

VII. IoT POTENTIALS TO RENOVATE EDUCATION

The article "Internet of Things in Education: The possibilities are numerous" [13], the author suggested four points to consider in using IoT to transform education. The first point is that "IoT will enable students to connect with teachers and access to full-time educational tools. It will also facilitate collaboration with teachers and other students. Parents can also have access to learning analytics through IoT". The second point is that "Schools are vulnerable places, with IoT possibly, we can reach a stage that with just the hit of a button a lockdown system can be initiated which can

be used in case of an emergency. Moreover, the system can send alerts to the police, fire stations and hospitals to fasten the response in case of an emergency. Surveillance will become extremely easy with IoT. The third point is that "IoT can help schools streamline mundane operations such as attendance, fee alerts, and student reports which can be automated easily. It can also bring down energy costs. ". The fourth point is that "Children with special needs can also benefit from IoT. Specialized software can help students with specific problems. For example, it can recognize visually impaired or hearing impaired students and make changes accordingly such as increasing font size or more visual cues. It will also save valuable time of the teachers which can be used to enhance the teaching experience."

VIII. IoT CAN PROGRESS STUDENT PERFORMANCE

In the article "Interaction System Based on Internet of Things as Support for Education" [14], it was stated that IoT could provide motivation and could allow students to be playful. IoT also allows teachers to teach students according to their aptitude. Teachers can choose the basic materials to suit students. Students also learn at their own pace according to their capabilities, so they are not limited by a one-size-fits-all program. The authors of the said article conducted an experimental validation which yielded evidence that IoT could improve the student's learning outcomes.

IX THE COURSE STRUCTURE

The goal of the IoT pilot course is to introduce and educate students with a background in business informatics in using the hardware, operating systems, software, and tools for automation of smart environments. The course may consists of four units as shown in below figure 3.

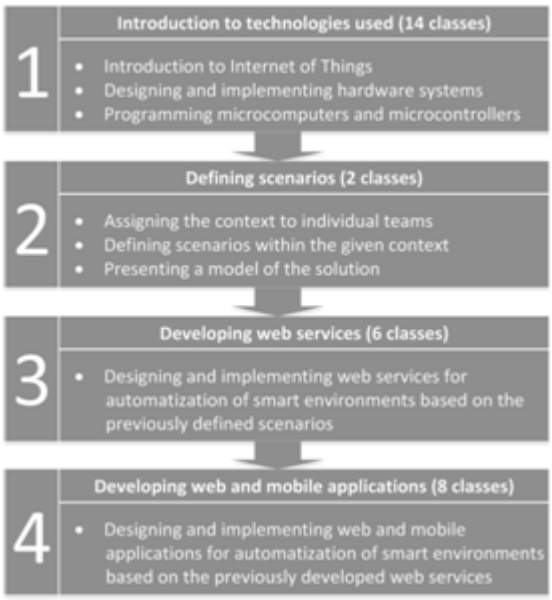


Fig3: The IoT Course Structure

X. CONCLUSION

Internet of Things (IoT) is the network of all kinds of things embedded with sensors, electronics, software, and so on, connected to the Internet, based on the International Telecommunication Union's Global Standards Initiative. The number of things in IoT will be 20.8 billion by 2020 and IoT spending will be 3,010 billion US$. IoT has also gained popularity in e-learning. This paper presented IoT in e-learning and instructional design, six skills for IoT applications, Internet of Learning Things, IoT potentials to transform education, and IoT to improve student performance.

XI REFERENCES

[1] Wikipedia.org. "Internet of Things". . Accessed 1 August 2015.
[2] International Telecommunication Union. "Internet of Things Global Standards Initiative".
[3] Perera, P. et al, 2014. Sensor Search Techniques for Sensing as a Service Architecture for the Internet of Things. IEEE Sens. J., Vol. 14, No. 2, pp. 406–420.
[4] Gartner, 2013. Forecast: The Internet of Things. Gartner.
[5] Sebastian and Ray, 2015 Sebastian, S., Ray, P.P., 2015. Development of IoT invasive architecture for complying with health of home. In: Proceedings of I3CS, Shillong, pp. 79–83.
[6]G. Yang, X. Li, M. Mäntysalo, X. Zhou, Z. Pang, L.D. Xu, S.K. Walter, Q. Chen, L.Zheng, "A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor and intelligent medicine box" IEEE Trans. Ind. Inf., 10 (4) (2014), pp. 2180-2191
[7] Miao W., Ting L., Fei L., ling S., Hui D., 2010.Research on the architecture of Internet of things. IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE), Sichuan province, China, Pages: 484-487.
[8] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future internet: the internet of things architecture, possible applications and key challenges," in Proceedings of the 10th International Conference on Frontiers of Information Technology (FIT '12), pp. 257–260, December 2012.
[9] Buzzconf.io. "IoT and e-Learning". <https:// buzzconf.io/sessions/iot-and-elearning/>
[10] IMS Learning Design specification. Retrieved March 04th, 2015, from http://www.imsglobal.org/learningdesign/
[11] "LTSC WG12: Learning Object Metadata",IEEE Learning Technology Standards Committee, 2002. Retrieved March 04th, 2015, from https:// ieee-sa.centraldesktop. com/ltsc/
[12] Mikelloydtech. "Internet of Learning Things". http://clwb.org/2013/08/21/internet-of-learning-things/
[13] Shashank Venkat. "Internet of Things in Education: The possibilities are numerous". http://blog.e-zest.com/internet-of-things-in-education-the-possibilities-are-numerous.
[14]. Huete, F., Oscar Hoyosa, L.P., and Grigori, D. "Interaction System Based on Internet of Things as Support for Education".

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

EFFICIENT ROUTING WITH INVERSE REINFORCEMENT LEARNING

Srikrishnan Subramanian
Student, Dept of Computer Science & Engg
SRM University, Chennai, India
Email: srikrishnan_subramanian@srmuniv.edu.in

Adithya Raam Sankar
Graduate Student Institute for Artificial Intelligence
University of Georgia Athens, GA, USA
Email: adithya.raam@uga.edu

ABSTRACT:

Transportation has undergone a lot of evolution since the invention of the wheel. With more and more sophisti-cated manufacturing methods being implemented, the production time of new vehicles has reduced drastically. This has led to a substantial increase in vehicular traffic in the last 3 decades. The beginning of personalized transportation has ushered in a new dimension in the understanding of traffic. The initial approach in managing spatial areas was to merge the roads and visualize it as a graphed network, where every stretch of road is visualized as an edge and the digressions/splitting of traffic being nodes leading to other edges. Increased vehicular traffic with almost constant mapped spatial area causes an unstable equilibrium, leading to congestion and gridlocks. This equilibrium requires an effective balancing/routing strategy to maintain stability among the network. The correlation between road networks and com-puter networks has been exploited to solve this problem, expecting minimal deviation from ideal behavior. Networking protocols are unable to handle deviations that occur due to natural human behavior. Machine Learning Techniques can be implemented to understand these deviations and obtain patterns in real time. The proposed system approaches the routing problem with the aim of learning optimal reward functions by observing regular human behavior for a set of actions. These functions are pivotal in maximizing utility for every agent involved in the procedure by adopting a cooperative and interactive approach.

Keywords—Inverse Reinforcement Learning, Traffic Regula-tion, Congestion Reduction, Re-Routing.

I. INTRODUCTION

Since the time man started moving around, he was con-stantly searching for a better method of transportation. The invention of the wheel was a major turning point in this jour-ney. It was followed by the development of various vehicles for personal as well as commercial commutation. Today, we have a variety of vehicles ranging from the eco-friendly bicycles for short distances to spaceships to reach other planets. Though both of these do not create much of a problem, the increase in the use of privately owned vehicles, like motorbikes and cars, has led to the increase in the congestion of roadways. In turn, this leads to the faster depletion of fossil fuels which provide the primary source of energy for the vehicles.

More than travelling a shorter distance and saving time, the need for saving fuel and eventually saving money has become a rising concern. Deciding the trade-off between time and fuel is a very crucial task. Any traveller will not want to compromise on time or distance as that is what they know as a factor for fuel consumption. Whereas in reality, traffic delays and poor routing[1] also influences the fuel efficiency. One of the possible solutions for this situation would be to create an ideal network of roads that reduce the number of traffic signals and in turn lessen the amount of vehicular crowding. But, given that the road network has already been laid out, any alterations would involve colossal destruction and a great deal of money. This raises the need for routing algorithms that act in real time to increase mileage in the existing system of roads. The proposed system intends to achieve the same by finding alternate paths effectively to eliminate on road congestion.

II. RELATED WORK

Researchers started approaching this problem from an essentially mathematical background. Two main approaches that included computer networking and economic models.The key aspect in which the current system is being proposed is the demarcation that exists between road traffic and network traffic. Network Traffic involves ideal participants while road traffic involves agents exhibiting erratic rational behavior.

The initial approach to route included baseline computer networking algorithms to chart possible paths. Dijkstra's Al-gorithm to obtain the shortest path is the most widely imple-mented algorithm. Location extraction and management can be achieved through Global Positioning Systems, which assure a good accuracy rate.Even if the network has a few sites that are down, this algorithm can be used to easily manage among the existing paths.

The second approach to solving traffic was to utilize eco-nomic principles. This was a feasible approach as economics

considers competition and social equilibrium. A Wardrop equi-librium [2] denotes a strategy profile in which all used paths by drivers between a given origin-destination pair have equal and minimal latency. The first rule of Wardrop Equilibrium states that no agent can decrease its experienced latency by unilaterally deviating to another path.[3]) Selfish routing would just bring about a string of latency issues for all agents involved. Also, conversion of economic principles into stable real-time self-balancing models requires additional computa-tional principles.

Application of decentralization and considering every ve-hicle as a potential data node was analyzed. Wireless Ad-hoc networks and Vehicular Ad-hoc networks(VANETs) have been implemented with such an overview [4]. Such networking approaches help in facilitating inter-vehicular transmission and aggregation of data that characterize the network under consideration.

Significant development began with creating mathematical definitions of traffic flow, both from an economic and a statistical standpoint. The measures of statistical aggregation were the initial metrics of defining the system. This was the essential characteristic of a Markov chain and understanding the transitions of the system. Hence, Markov chains were applied to further enhance the understanding of the traffic network.[5]

In recent studies, the properties of stochastic processes have been widely applied to understand complex human interac-tions. The same was extended to the understanding of traffic flows. The stochastic extension of Markov chains was Markov Decision Processes. were articulate in representing traffic flow. Given the essential characteristics of a stochastic process, such models were analyzed with greater detail for accuracy.

Markov Decision Processes explains state wise transitions and can help in an efficient maximization of the utility involved during the path, with solutions to the Markov decision process described in Bellman Equations [6]. These equations provide an effective solution for long term maximized utility. The primary requisite for solving a Markov Decision Process is to have a well-defined reward function in place, that can be maximized in the solution given by the Bellman equations. As-suming arbitrary reward functions, the results are not accurate and do not take into consideration, competition among agents that would pose changes to the rewards/payoff obtained.The branch of Inverse Reinforcement Learning began, as a result, to enable the learning of reward functions, over a long term period of observation of expert system, whose behaviour we intend to simulate.

III. PROPOSED SYSTEM

As we have seen the two pronged approaches of networking algorithms and economic models, it is increasingly clear about the shortcomings. These shortcomings can be overcome by merging both the systems. The proposed system is a hybrid approach to the routing of vehicular traffic. It employs a congestion aware mechanism that utilizes the knowledge of real time traffic. This availability of knowledge maps the traffic flowing in the streets and effectively merges the data into the decision making process of the system, thus making effective routing choices. The initial functioning of the system involves extraction of the source and destination under consideration. Following this, all possible paths are retrieved. These paths undergo an extensive analysis to identify possible levels of congestion at each “edge”. If found, systems will identify the alternative path to ensure the driver reaches the destination with no loss in rewards. The entire sequence of the system is governed by the second principle of Wardrop’s Equilibrium that drivers cooperate in a multi-agent environment for maxi-mized reward. The approach of inverse reinforcement learning to solve cooperative games enables in learning a more optimal reward function. Inverse Reinforcement learning algorithms as specified in [7],specifies that, reward functions are essentially quantifiers of tasks performed by the agents in the system.

Such systems would then alleviate road congestion, pre-vent future congestion and ensure that the agents that are participating in the routing mechanism have maximized utility. The methods of machine learning are utilized to enhance the routing efficiency and provide a real time best path for the agent under consideration. This system , as described in Figure 1, builds on the idea of applying reinforcement learning to learn the metrics of the flow. The system describes the real time updating and maintenance of the network of roads as main-tained.The set of Reinforcement learning algorithms balance exploration and exploitation. Exploration is trying different things to see if they are in face better than what has been tried before. Exploitation involves making greedy decisions based on local parameters. The advantages of Reinforcement Learning over standard supervised learning algorithms is that the latter doesn’t perform this balance. They generally are purely exploitative. (Bayesian algorithms implicitly balance exploration and exploitation by integrating over the posterior.) The approach of inverse reinforcement learning to solve co-operative games enables in learning a more optimal reward function. The crucial terminology involved in any reinforce-ment learning system include environment, agents, rewards and utility function. The environment is the network of roads under consideration upon which drivers function. Drivers are the agents that operate on the environment. Rewards would include any benefits that adds to the user’s usability levels. The system involves the context of real time information through regular inclusion into decision making , thus making the whole process of route prediction concurrent and progressive.

A. Initial Data

The starting module in the entire system involves initial-ization of the system structure and baseline training. Train-ing process involves the basic understanding of the system

processes and executing forthcoming processes with the same accuracy. The data for the initial training process is obtained by leveraging existing data collection methods. This data is used to create the initialization for the network and construction of a probability table with respect to the data observed. The most widely used methods of aggregating location based information with respect to vehicles are the availability of Geographical Positioning Systems(GPS). GPS data can be remotely tracked from mobile devices associated with the vehicle. A real-time aggregation of such data can happen with a regular poll of the GPS coordinates. Thus, obtaining periodic location information helps in analyzing the exact route taken by the vehicle and also time is taken to travel the distance. The environment handler accepts routing requests and performs the iterative process of routing . Environment Handler , as mentioned in Figure 1 , performs the process of apprenticeship learning where the requests are associated with an agent and a separate set of initializations performed.

A futuristic technique would be to take advantage of the developments in Internet of Things. This approach would involve the consideration that all vehicles are interconnected and there is consistent transfer of data among the network. Data collection using Internet of Things can be achieved using a ”checkpoint” strategy where vehicles log their traversal on a given road at data collecting devices positioned strategically. Increased availability of vehicles with built-in GPS system al-lows the decentralization of the path finding mechanism. Once the user enters the destination, the vehicles can find path to be taken by themselves and send it to the server. This minimizes the load on the algorithm and makes implementation much easier. This technique is advantageous in maintaining a stable persistent system when it comes to real time updates of traffic flows. The data gets mapped into a database to maintain the information with respect to the clusters.

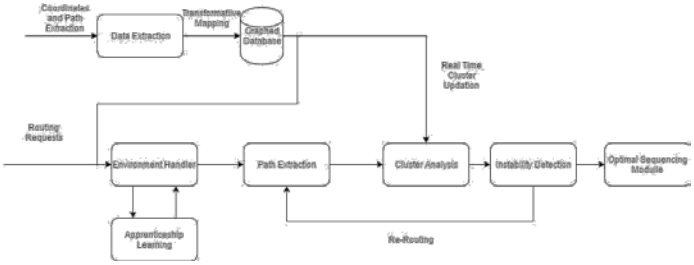


Fig. 1. Architecture of the System

B. Clustering Mechanism

This module is responsible for mining patterns and analyz-ing road clusters that exhibit similar traffic patterns. The train-ing data can be used to understand the historical patterns and deviations in traffic flow. Interesting patterns with relevance to this system would involve traffic flow among clusters. Once the system has been trained, it attains appropriate confidence levels that helps it get a solid understanding with respect to the reward functions used and the allocation of said rewards. The initial requirement in routing is to obtain all the possible paths from the source to the destination, achievable through single source shortest path algorithms. These paths would have common linkages with several clusters. The essentiality of the module is to ensure that the path optimally chosen provides a balanced traffic flow. These patterns may be based on a number of parameters that include the number of vehicles passing through, delay time, bottleneck issues etc. When the data with respect to the roads are obtained, the algorithm will cluster such roads to monitor for similar or deviant behaviour of traffic. Essential Analysis of such competing clusters is required to analyze possible sources of congestion .

Algorithm 1: Training and Reward Extraction

input: Coordinates of agents,C
1 M= Markov Decision Process;
2 C_{a1} = Monitored Coordinates of vehicle a_1 ;
3 T = Transition Probability Table of M;
4 foreach c_a 2 C do
5Extract the paths from coordinates;
6Update T
7 end
8 Constructfor the observed M such that
$E[^{P}_{t=0}\ ^{1}_{t}\ R(s_t)j\ )] > E[^{P}_{t=0}\ ^{1}_{t}\ R(s_t)j\ )]$8

Algorithm 2: Clustering Algorithm

input: graph G
1 P = All possible paths in the network;
2 initialize clusters;
3 foreach p 2 P do
4n_p = number of vehicles on p;
5t_p = average time spent on traversal
6 end
7 f lowrate$_p$=n_p/t_p; 8
while p do
9path$_p$ = paths with the same flowrate as current path ;
10 l_p=linking path between current path and paths in path$_p$;
11 if l_p then
12 append path p to cluster
13 end
14 create new cluster and assign path p to it.
15 end

C. Stability and Driver Interaction

The system functions essentially on the premise of the second rule of Wardrop Equilibrium [2].

Once the system receives the data with respect to the locations in consideration, the system checks all possible paths available for the journey. Once the path ideal for the user has been chosen, the algorithm now will analyze all possible allied paths that would influence the traffic throughout the network of paths that the vehicle would encounter. As by our initial definition, each vehicle is idealized as an agent in a multi-agent

environment. The common resource that all agents want would vary. It is least time and maximized fuel benefit that most drivers target. When the travel is considered, the resource would also include the right of traversing on a road. When agents compete for the same resource, we enter into a stand-off with two or more systems requesting access to the same resource. Given user preferences that change, certain agents may prefer optimizing on fuel costs over time at times. If the parameters considered in allocation are similar for the both the systems, it essentially reduces to a function requiring a feasible solution(maximized utility with respect to those parameters). Feasible solutions are defined as optimal allocation procedures. Hence, optimization procedures such as the particle swarm Optimization can be utilized to perform the analysis to obtain an optimized path. If the parameters that the agents prioritize are different in nature, we enter into a heterogeneous competition. Thus, resolving involves the concept of Nash's Equilibrium. We consider a bimatrix game with agents competing. Nash's Equilibrium will define the optimal utility that the agents can avail so as to not lose the possible utility.

When the introduction of the vehicle into the prescribed path causes an imbalance, the system would now have to route in an alternate manner. Thus, we would have to chart alternate paths from the driver's location towards the destination. This would require the passing of the geo-coordinates back to the path extraction. The entire systems now go through a rerouting phase, which is a second pass of the same algorithm. Once an alternate path that maximizes driver utility is obtained, the system enters the optimal sequencing module and the driver is guided along the same path and the stability of the network is restored.

Algorithm 3: Path Extractor Algorithm

```
input: source s, destination d
P = All possible paths from s to d;
foreach p 2 P do
n_p = number of vehicles on p; 4 end
Pick the path with least n_p;
while currentLocation is not d do
if inclusion into path causes no instability then
        n_p = n_p + 1;
    else
        Calculate utility ;
        if utility < threshold then
            route := True;
            pathExtractor(currentLocation, destination)
        end
    end
end
```

IV. CONCLUSION

The proposed system attempts to provide an congestion aware efficient routing mechanism using Inverse Reinforcement Learning. This routing system optimally provides us with the best possible path that can be learnt from the vehicular data. This system uses a hybrid model to determine congestion in the system by analyzing cluster based flows. The algorithm is devised in such a way that the complex characteristics of human driving are captured and reward functions are suitably formulated. This makes sure that efficiency of the system can be made better every time there is new areas or vehicles. In the process of traffic engineering, providing congestion-aware routing is always been a daunting task, this system takes a step forward in providing content which satisfies their interest and also helping the transport community to handle and construct traffic flows better.

ACKNOWLEDGMENT

The authors would like to thank the institution for allowing us to carry out this research work. The staff and coordinators were very much supportive in bringing out this system. Their queries, comments and suggestions were helpful in shaping the system to yield more effective results.

REFERENCES

[1] J. R. Pierce, "The Fuel Consumption of Automobiles," Scientific Amer-ican(ISSN: 0036-8733), vol. 232, 1975.

[2] J. G. Wardrop, "Some Theoretical Aspects Of Road Traffic Research," Proceedings of the Institution of Civil Engineers, vol. 1, 2002.

[3] F. S. G. Carlier, C. Jimenez, "Optimal Transportation with Traffic Congestion and Wardrop Equilibria," SIAM Journal on Control and Optimization, vol. 47, 2008.

[4] O. W. Y. B.K.Mohandas, R.Liscano, "Vehicle traffic congestion manage-ment in vehicular ad-hoc networks," Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on, vol. 47, 2009.

[5] E.Indrei, "Markov Chains and Traffic Analysis," The Rose-Hulman Undergraduate Mathematics Journal, vol. 8, 2006.

[6] P. E.Rachelson, F.Garcia, "Extending the Bellman equation for MDPs to continuous actions and continuous time in the discounted case," The International Symposium on Artificial Intelligence and Mathematics, 2008.

[7] S. R. Andrew Ng, "Algorithms for Inverse Reinforcement Learning," http://ai.stanford.edu/ ang/papers, vol. 1, 2000.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

Available Online at www.ijarcs.info

ICRACC 2018
(23rd–24th February 2018)

A SURVEY ON LOSSLESS AND LOSSY IMAGE COMPRESSION TECHNIQUES

Caran Hepsiba.B, Jeevitha.V,
Master of computer science,
University of Madras,
Shri Shankarlal Sundarbai Shasun Jain College for Women,
Tnagar, Chennai , Tamilnadu, India
caranbalraj888@gmail.com, jeevithav95@gmail.com

ABSTRACT

With the growth of multimedia technology over the past decades the demand for digital information increases dramatically. This Paper gives review of two different compression techniques. Digital images are compressed of an enormous amount of data. reduction in the size of the image data for both storing and transmission of digital images are becoming increasingly important as they find more applications. Image compression is an mapping from an higher dimension space to lower dimension space. The basic goal of image compression is to represent an image with minimum number of bits of an acceptable image quality.
In lossless compression, the image after compression and decompression is identical to the original image and every bit of information is preserved during decomposition process. In lossy compression the reconstructed image contains degradation with respect to the original image.

Keywords: Image compression, lossy compression and lossless compression, image compression standards.

I. INTRODUCTION

Image compression is a process of reducing the size in bytes of the image to an undesirable level. It plays an important role in many multimedia applications, such as image storage and transmission. It is different than compression of digital data. To lower the irrelevance and the redundancy of image data is the major target of the image compression is to enable them to get saved or transmit the data in the better form. Lossy compression methods can achieve and they reduce the accuracy of the reconstructed images by producing some distortions. It is generally used for video and sound, where a certain amount of information loss will not be detected by most users.
Lossless compression or error free compression is the data reduction method since there is no loss of data. It is used for medical imaging, technical drawings, clipart, or comics.

II. NEED FOR COMPRESSION

With the advancement in internet ,teleconferencing, multimedia and high –definition television technologies, the amount of information is handled by computers has been grown over the past decades hence storage of the digital image component of multimedia system is a major problem .the possible solution is to compress the information so that storage space can be reduced. Image compression is a way to represent an image in a more compact way, so that images can be stored in a compact manner and can be transmitted faster.

III. TYPES OF REDUNDANCY:

To reduce some data which is not relevant or provide no information is called as Redundancy. There are 3 types of data redundancy.

A. Coding Redundancy:

A code is a system of symbol used for representing and information. Each piece of information or event assigned a sequence of code symbols called code word[7]. The code length is defined as number of symbols in each code word.

A resulting image is said to have a coding redundancy if its gray levels are coded using more code symbols than actually needed to represent each gray level. The Huffman codes and the arithmetic coding technique are examples of image coding schemes that explore coding redundancy.

B. Inter pixel Redundancy:

In image neighboring pixels are not statistically independent. It is due to the correlation between the neighboring pixels of an image. This type of redundancy is called Inter-pixel redundancy. Inter pixel correlation are the structural and geometric relationships between objects in the image.

Other names:

- Spatial Redundancy
- Geometric Redundancy
- Interframe Redundancy

C. Psycho visual Redundancy:

The human eye does not respond to all information with equal sensitivity. Because, some information, may be given less importance when comparing to another information in normal visual processing. Such information is said to be psycho visually redundant.

Properties:

It has the following properties:
It is basically different from other redundancies.
It is related with real or quantifiable visual information.
Removal of this redundancy will not affect the perceived image quality, since the data is not essential for normal visual processing.

IV. CLASSIFICATION OF IMAGE COMPRESSION

Image compression can be classified into 2 types.
They are :

- lossless compression
- lossy compression

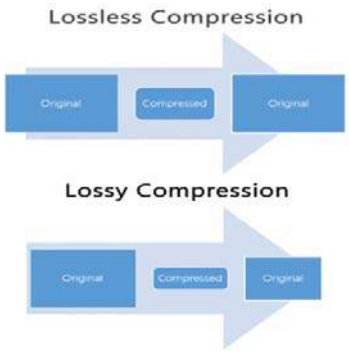


Fig:1

a. Lossless compression:

Lossless compression is the acceptable data reduction method since there is no loss of data. For both binary and grey-scale images lossless method can be applied.lossless compression retains raster values during compression and file size is also reduced. It is also known as entropy coding as it uses decomposition techniques to minimize loopholes. The original image can be perfectly recovered from the compressed image, in lossless compression techniques. These do not add noise to the signal. It is also known as entropy coding as it uses decomposition techniques to minimize redundancy.

Following techniques used in lossless compression are:

1. Run length encoding
2. LZW coding
3. Huffman coding
4. Arithmetic coding

i. Run Length Encoding:

The simplest data compression technique is run length encoding. It is effective when long sequences of the same symbol occur. Run-length coding exploits the spatial redundancy by coding the number of symbols in a run. The term run is used to indicate the repetition of a symbol, while the term run-length is used to represent the number of repeated symbols. This compression technique is useful in case of repetitive data. When we have sequence of same intensity pixel or symbols then this sequence is replaced by shorter symbols and it is represented by a sequence (Vi,Ri),where Vi is represented as the intensity of pixel and Ri is the no of consecutive pixel with same intensity.[2]

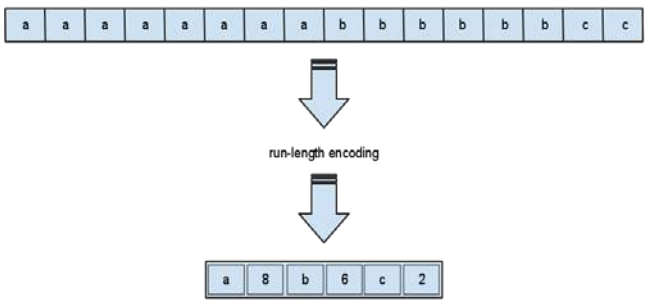


Fig2

ii. LZW Coding:

LZW (Lempel-ziv-welch) coding is an error free compression technique. It is dictionary based coding, which is used in computer industries[2].It replaces string of characters with single codes. LZW compression creates a table of strings commonly occurring in the data being compressed, and replaces the actual data with references into the table. The table is formed during compression at the same time at which the data is encoded and during decompression at the same time as the data is decoded[3].

iii. Huffman coding:

The Huffman coding algorithm is named after its inventor, David Huffman. Huffman coding today is often used as a "back-end" to some other compression methods[7]. The pixels in the given image are assigned some specific numbers. The pixel having lesser occurrences will be given higher number of bits and the pixel with higher frequency occurrences will get relatively lesser number of bits. It is a prefix code. No two symbols in an image can have exactly same binary set of numbers[8]. The pixels in the image are treated as symbols.

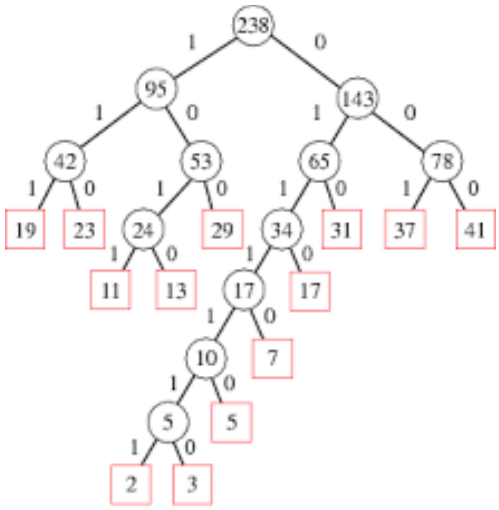


Fig3

iv. Arithmetic coding

Arithmetic Coding is the compression technique for lossless encoding that represents a message as some finite intervals between 0 and 1 on the real number line[7].AC does not generate individual codes for each character but performs arithmetic operations on a block of data, based on the probabilities of the next character. Using arithmetic coding it is possible to encode characters with the fractional number of bits. Arithmetic coding performs very well for sequences with low entropy.

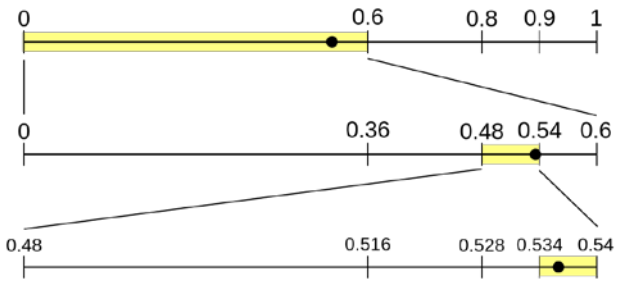


Fig 4

b. Lossy image compression:

In Lossy compression techniques reconstructed image contains loss of information which in turn produces distortion in the image. .It is a irreversible process. Image compression ratio can be very high.

Following techniques used in lossy compression are:

1. Transform coding.
2. Block truncation coding.
3. CodeVector quantization.
4. Wavelet Coding

i. Transform coding:

It is a common method for lossy image compression. It employs a reversible and linear. The linear transform is to decorrelate the actual image into a set of coefficients in transform domain. In transform domain the coefficients are then quantized and coded successively [5].

ii. Block Truncation Coding

Block truncation coding is well known technique for image compression, It (BTC) divides the original image into small sub blocks of size n x n pixels and after the division of image, it reduces the number of gray levels within each block. reduction of gray level is performed by a quantizer[2]. The threshold is normally the mean value of the pixel values in the vector. Then a bitmap of that vector is generated by replacing all pixels having values are greater than or equal to the threshold by a 1. Then for each segment in the bitmap, a value is determined which is the average of the values of the corresponding pixels in the original code vector[8].

iii. Code Vector Quantization

The basic idea in Vector Quantization is to create a dictionary of vectors of constant size, called code vectors. Code vector is the Values of pixels composed the blocks.A image is then parted into non-recurring vectors called image vectors. Dictionary is made out this information and it is indexed. Further, it is used for encoding the original image. Thus, every image is then entropy coded with the help of these indices[8].

iv. Wavelet Coding:

Wavelet coding technique is based on the discrete wavelet transform.

Discrete wavelet transform ,which transforms a discrete time signal to a discrete wavelet representation.

Digitize the source image to a signal s,which is a string of numbers. Decompose the signal into a sequence of wavelet coefficients. Use the Thresholding to modify the wavelet compression to w, to another sequence w. Use quantization to convert w to a sequence q. Apply entropy coding to compress q into sequence of e.

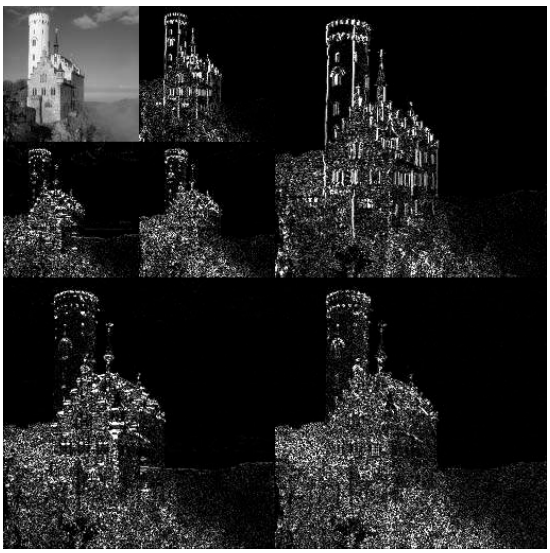


Fig5

V. CONCLUSION

In this paper we have concluded that two types of techniques can be used for compression. Lossy and Lossless techniques, the usability and efficiency of respective techniques are different. This review paper also gives the idea about various image types and performance parameter of image compression.

VI. REFERENCES

[1] S Jayaraman S Esakkirajan T .Veerakumar "Digital image processing" Mc Graw Hill Education (India) Private limited.

[2] Akhand Pratap Singh Dept. of electrical and electronics engineering, NITTTR Bhopal, M.P, India, Dr. Anjali Potnis Asst. professor, Dept. of electrical and electronics engineering, NITTTR Bhopal, M.P, India, Abhineet Kumar Dept. of electrical and electronics engineering, NITTTR Bhopal, M.P, India " a review on latest technique on image compression"journal of IRJET Volume: 03 Issue: 07 | July-2016

[3] Khobragade P. B., Thakare S. S. Department of Electronics and Telecommunication, Amravati University Govt. College of Engg. Amravati, Maharashtra, India" Image Compression Techniques- A Review" International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 272-275

[4] Manjari Singh Dept. Information Technology LNCT Sushil Kumar Dept. Information Technology LNCT Siddharth Singh Chouhan Dept. Information Technology LNCT Manish Shrivastava, PhD Dept. Information Technology LNCT" Various Image Compression Techniques: Lossy and Lossless",International Journal of Computer Applications (0975 – 8887) Volume 142 – No.6, May 2016

[5] Rajandeep Kaur Mtech Scholar (CSE) CTIEMT, Shahpur Jalandhar Pooja Choudhary Assistant Professor (CSE) CTIEMT, Shahpur Jalandhar" A Review of Image Compression Techniques" International Journal of Computer Applications (0975 – 8887) Volume 142 – No.1, May 2016

[6] Rafael C Gonzalez |Richard E Woods 3rd edition"Digital image processing"published by pearson education Inc.

[7] Bhavna Pancholi, Reena Shah, Mitul Modi Department of Electrical Engineering Faculty of Technology and Engineering The M. S. University, Baroda" Tutorial review on existing image compression techniques" International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 8 August, 2014 Page No. 7882-7889 Bhavna Pancholi, IJECS Volume 3 Issue 8 August 2014 Page No.7882-7889 Page 7882

[8] Dinesh V. Rojatkar, Nitesh D. Borkar, Buddhabhushan R. Naik, Ravindra N. Peddiwar BE, Electronics and Telecommunication Engineering, RTMN University, India brn2393@gmail.com Contact no:9096020390" Image Compression Techniques: Lossy and Lossless". International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015 ISSN 2091-2730 912 www.ijergs.org.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A PLATFORM OF INTERNET OF THINGS IN VARIOUS DOMAINS - A SURVEY

Ms. J. Jeba Priya,
Assistant Professor, PG Department of Computer Science
Shri Shankarlal Sundarbai Shasun
Jain College for Women
Chennai, Tamilnadu, India
j.jebapriya@shasuncollege.edu.in

Ms. V. Angala Parameshwari
Master of Computer Science
Shri Shankarlal Sundarbai Shasun
Jain College for Women
Chennai, Tamilnadu, India
angalams1994@gmail.com

ABSTRACT

An article is to address the features of Internet of Things in different sources in various region of a domain. Nowadays, learning is through electronic media, which is nothing but an Internet. This journal tells you about the application of IoT and how IoT change into an E-Learning. The applications of an IoT such as smart cities, smart homes, smart grid, smart health, smart transportation and mobility. The network of physical objects defines Internet of Things or "things" embedded with electronics, software, sensors, and connectivity allowing them to exchange data with the manufacturer, operator, or other connected devices.

Keywords: IoT, applications of IoT, IoT to e-learning, Internet of Things, smart cities.

I. INTRODUCTION

The Internet of Things is changing everything, and eLearning is no exception. As the world continues to change and become
More interconnected, watch out and plan for the following 6 changes, so you can stay at the top of the E-Learning industry. The International Telecommunication Union's Global Standards Initiative has been done through, network of all kinds of things embedded with sensors, electronics, software, and so on, connected to the Internet[3]. The term IoT encompasses an unbounded, growing set of devices and technologies, and as the IoT technologies gain traction globally, the need for experts that combine knowledge from various technical fields' increases. IoT projects are likely to need designers, system integrators, developers and technicians in order to take an idea from inception to execution. Such diverse requirements can create an understanding gap between business-oriented individuals and their ideas, and the actual implementers that deal with realistic constraints. [4]

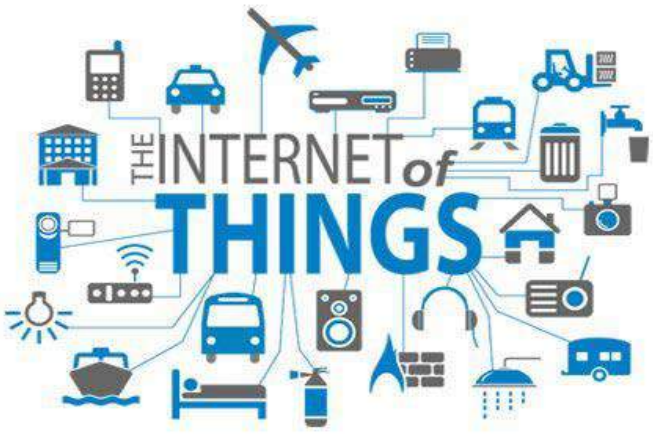


Fig1

It is an intelligent interconnection of all things via the Internet, to communicate and exchange information through information sensing devices in conformance with agreed protocols. Achieving the goal of intelligent identification, location-tracking, monitoring, and managing things. It intertwines many day-to-day objects surrounding us into networks in one or the other form. Varied smart technologies like RF identification [RFID] and sensor technology shall be embedded into a wide spectrum of application.

II. APPLICATION OF INTERNET OF THINGS

There are many applications in IoT such applications are:

- Smart cities
- Smart home
- Smart grid
- Smart Health and
- Smart Transportation and Mobility

A. Smart cities

Smart cities may still be viewed as cities of the future and smart life, and by the innovation. By the IoT, cities can be improved in many levels, by improving infrastructure, enhancing public transportation World Scientific News (2017). By connection all systems in the cities like transportation system, healthcare system, weather monitoring systems and etc., in addition to support people by the internet in every place to accessing the database of airports, railways, transportation tracking operating under specified protocols, cities will become smarter by means of the internet of things[8].

Fig 2

B. Smart Home

Wi-Fi have started becoming part of the home IP network and due the increasing rate of adoption of mobile computing devices like smart phones, tablets, etc
Security: Video door monitoring, motion sensors, gas leakage detection, intrusion sensors, curtain sensors, fire detection and control.

Features

Lighting control: Remote on/off, presence detection.

Electrical: Smart air conditioning, smart refrigerators.

Entertainment: AV controls, gaming consoles.

Communication: GPS navigation connected to entry and exit, proximity door unlocking.

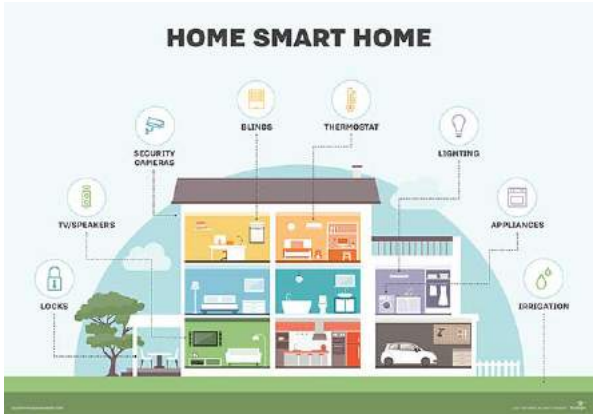


Fig 3

C. Smart grid

A smart grid is related to the information and control and developed to have a smart energy management. A smart grid that integrate the information and communications technologies (ICTs).Many applications can be handling due to the internet of things for smart grids, such as industrial, solar power, nuclear power, vehicles, hospitals and cities power control. Smart grid applications today's grid is very reliable and can deal with normal electricity fluctuations and it will take a step further towards using a low carbon energy system, by allowing integration between the renewable energy and green technologies, and offering many benefits to customer in cost savings through efficient energy use at home[8].

Fig 4

D. Smart Health

Smart health replaces the process of having a health professional come by at regular intervals to check the patient's vital signs, instead providing a continuous automated flow of information.

Fig 5

E. Smart Transportation and Mobility

The development in transportation is one of the factors to indicate the well being of the country. A road condition monitoring and alert application is one of the most important of IoT transformation application. The main idea of the concept of smart transportation and mobility is to apply the principles of crowd sourcing and participatory sensing. The process began with user identified the route wishes and marked some points as pothole in the smart phone's application.

Fig 6

III. IoT TO E-LEARNING

The Internet of Things is changing everything, and eLearning is no exception. ***"To study effectively through e-learning gives self-motivation and go beyond success".*** As the world continues to change and become more interconnected, watch out and plan for the following changes, so you can stay at the top of the eLearning industry.

A. People Will Consume Content Differently

It's estimated that the number of IoT-connected devices in 2018 will surpass the number of mobile devices for the first time in history. It's

projected that there will be over 6 billion smart phones and 50 billion IoT-enabled devices in the world by 2020. For reference, the world's population is 7.6 billion and growing. That means everyone-from teachers to students to marketers-will need to change the way they develop and write content.

Even in the past couple of years, the world has become much more visual. For example, video has trumped copy as the preferred way to consume content-and video itself has changed. One study found that adding an interactive aspect to a video gets 40% more engagement than those videos without this feature.

Video is easily integrated into small IoT-enabled devices and can say more at a glance than the written word, so eLearning students and teachers will have to change the way they deliver content. In short, that content needs to do a lot more than optimize for mobile in the coming months and years.

B. The Flexible Will Survive And Thrive

No business succeeds when it stands still-the same goes for teachers, students, and entrepreneurs in eLearning. It's critical to stay up to date with current IoT facts and trends; look at data that shows how, when, and where your audience is spending time online; and capitalize on that.

Achieving these goals might require more nimble movements over time, but that flexibility can help you succeed in this ever-evolving digital landscape.

C. Testing Will Change

With internet access on your phone, pens, and, who knows, maybe even sneakers someday, how will teachers test students in a fair, cheat-proof environment? Perhaps testing will change from memorized question-and-answer exams to research-based projects.

For instance, instead of answering multiple choice questions, students will have to use the internet to locate answers and expand on these findings. With the internet at our fingertips for practically everything, it makes sense to utilize that information in a productive way. As an eLearning teacher, you'll just have to think about the best way to test your students.

D. Expectations Will Shift

The IoT doesn't just change the way people connect to the internet-it also changes the lifestyles, expectations, and habits these individuals form. With Artificial Intelligence (AI) and IoT come home security, faster coffee, efficient energy use, and even faster streaming. Smart home automation is a great example of this shift.

As one of the biggest trends relating to the IoT right now, home automation gives people convenience and peace of mind like never before. Once people get used to this ease of living, it'll become the new norm for how they expect to access other information too.

When you're developing any new eLearning materials, think about how to make your eLearning courses cheaper, more interactive, and convenient to help attract more students and stand out from other classes.

E. New Majors Will Surface

With rising IoT-related developments and data points, new jobs will surface. Teachers will need to develop entirely new curricula and majors to account for this change—and students will need to look toward the future to plot out career prospects relating to these new fields, especially in technology-centered careers.

F. Job Competition Will Be Global

E-Learning is a globalization of education; people from anywhere can learn in virtual classrooms and get degrees. With the IoT also bringing the world closer together by connecting everyone to the web, it's safe to assume that jobs themselves will become more competitive and more students will be turning to eLearning to get the affordable and accessible education they need. Specific skills, advanced accreditation, and guidance will be in high demand. eLearning teachers may want to consider diversifying their classes and coursework while offering the best learning materials compared to their own competition. Doing so may help their students get an edge up on soon-to-be globally competitive career fields. It's incredible to imagine where the world will be after another 20 years of technological advancement. Stay ahead of the curve in eLearning with the merge of the IoT by remembering these points [7].

Conclusion

Thus Internet of Things gives new shapes to the living being by communicating among smart things. Internet of Things (IoT) is somehow a leading path to the smart world with ubiquitous computing and networking to ease different tasks around users and provide other tasks, such as easy monitoring of different phenomena surrounding us.

REFERENCES

[1] CSI Communications "INTERNET OF EVERYTHINGS" volume no:41/Issue No.4/July 2017
[2] M.U. Farooq, "A Review on Internet of Things (IoT)" International Journal of Computer Applications (0975 8887) Volume 113 - No. 1, March 2015.
[3] Prof. Dr. Srisakdi, International Journal of the Computer, the Internet and Management Vol.23 No.3 (September-December, 2015) pp. 1-4 1 "Applications of Internet of Things in E-Learning"
[4] Zorica Bogdanović, Konstantin Simić, Miloš Milutinović, Božidar Radenković and Marijana Despotović-Zrakić, "A PLATFORM FOR LEARNING INTERNET OF THINGS" International Conference e-Learning 2014
[5] Somayya Madakam," Internet of Things: Smart Things"
[6] Dr.R.P.Ram kumar, Hima Sampathrao, Vijay Kumar Burugari and Sanjeeva Polepaka, "Applications Domains of Internet of Things: A Survey" International journal of Engineering Technology science and Research IJETSR www.ijetsr.com ISSN 2394-3386 Volume 3,Issue 10 October 2016
[7] https://elearningindustry.com/internet-of-things-is-changing-elearning-6-ways
[8] "Internet of Things Applications, Challenges and Related Future Technologies", Zeinab Kamal Aldein Mohammeda , Elmustafa Sayed Ali Ahmedb, WSN 67(2) (2017) 126-148 EISSN 2392-2192.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

A Study Paper on Wireless Sensor Secure Routing

T. Yegammai,
Assistant Professor, Department of Computer Science,
yegammai@shasuncollege.edu.in

S.G Packiavathy
Head, Department of Computer Applications,
packiavathypaul@hotmail.com

ABSTRACT

An important purpose is that the wireless sensor routing security networks have many sensor routing protocols and nodes but have no security. Security goals for routing in sensor networks show us how crippling attacks have been made and attacks have been made and attacks against ad-hoc and peer-to peer networks. Two undocumented attacks such as sinkhole and hello floods which have been described and analyze the security of all secure routing in wireless sensor networks and protocols used for disseminating controls and information's network called sinks.

INTRODUCTION:

Routing security in wireless networks is an important purpose in the routing network which has limited nodes and application networks but have no security. Although there is no security available, we make the

security properties. In insecure wireless communication, limited nodes[1] and insider threats, where when designing the network secure routing adversary people has laptops with energy and long range communication, where the routing becomes non-trivial. Crippling attack is provided for all the major routing protocols because they have no security on the routing of sensor network and is insecure.

BACKGROUND:

Sensor network refers to the sensors and general computing elements. A sensor network consists of hundreds and thousands of low power costs and nodes but only at fixed locations which affects the environment[2]. The sensor networks which consist of one or more point of control is the base stations. The sensor node is the access point for the securing routing which is used to disseminate the control information on networks. Sensor network routing might have laptop, memory- storage, and high-bandwidth links for communication among the sensor network. Sensor network use low power and bandwidth that would communicate to the nearest base station for sensor network. The aggregation networks where the total numbers of messages sent, the energy is saved in the network where from the aggregation point they collect the readings from surrounding nodes. Sensor networks differ from other systems where it has a great challenge [2]. The value of sensor networks comes from many nodes where they will have to develop cheaper sensor nodes. Security is critical when networks are at risk where we have sensor nodes such as[3]:- High bandwidth Sensor node Base station Low latency and Only laptop and base stations use low latency and high bandwidth.

RELATED WORK:

Security issues are similar to sensor networks and are developed for ad-hoc networks. The secure routing protocols [4] for ad-hoc networks and sensor network reasons are that they sense security in ad-hoc network for authentication and secure routing protocols. The routing protocols are based on public key cryptography[5] which is used for sensor nodes.

SECURITY GOALS:

Secure routing protocols which have integrity, power and availability of messages in presence of arbitrary power. Security is not relevant to the application data and not responsible for routing protocol.

ATTACKS ON SENSOR ROUTING:

There is attack against the ad-hoc sensor networks but quite simple for the following categories:

SINK HOLE:

The goal is to take away all the traffic (nodes) from the particular area through nodes creating a sinkhole with adversary at the centre [5]. Sinkhole attacks which enables many other attacks where only one single node provides single high quality information which influences large number of nodes. The laptop class with transmitter which provides a high quality route for transmitting with power.

HELLO FLOOD:

The hello packets which are available to the bound channel are available to the attackers. An advisory situated close to the base station may be completely disrupted[8]. Protocols which depend on the local information exchange between neighboring nodes for maintenance nodes.

Attacks on specific sensor network protocols:

The main attacks of the sinkhole and hello flood is that the tiny OS protocols which can increase latency or disable thread [7]. The attacks on the link and sensor networks are against the network routing by line layers encryption. Sinkhole attacks on the network. Protocols which defend against protocols and the provided information, such as

HELLO FLOODS ATTACKS:

The hello floods attacks which verify the links of the nodes based on messages over the link. The hello floods attacks verify the link between two nodes, even if the advisory has high sensitive networks which will verify neighbors for each node to prevent hello flood attack.

CONCLUSION:

Securing routing which becomes vital to acceptance and the use of sensor networks against various protocols and attacks against the ad-hoc and peer to peer network where the attacks and routing protocols which defeat the security goals against the adversary. But we have

demonstrated the currently routing protocols against the sinkhole and hello flood attacks. A design of the sensor network routing protocols which satisfies the security goals and also where the authentication and sensor routing protocols which can be used as security nodes and where the key cryptography which cannot depend the laptop class adversary and the protocol can be designed well for sensor network to be secured.

REFERENCES

[1] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Wormhole detection in wireless ad hoc networks,"Department of Computer Science, Rice University, Tech. Rep. TR01-384, June 2002.

[2] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *IEEE INFOCOM '97*, 1997,pp. 1405–1413.

[3] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.

[4] F. Stajano and R. J. Anderson, "The resurrecting duckling: Security issues for ad-hoc wireless networks," in *Seventh International Security Protocols Workshop*, 1999, pp. 172–194.

[5] J. Hubaux, L. Buttyan, and S. Capkun, "The quest for security in mobile ad hoc networks," in *Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC 2001)*, 2001.

[6] L. Zhou and Z. Haas, "Securing ad hoc networks," *IEEE Network Magazine*, vol. 13, no. 6, November/December 1999.

[7] J. Kong, P. Zerfos, H. Luo, S. Lu, and L. Zhang, "Providing robust and ubiquitous security support for mobile ad-hoc networks," in *ICNP*, 2001,pp. 251– 260.

Volume 9, Special Issue No. 1, February 2018

ISBN: 978-93-5300-455-2

International Journal of Advanced Research in Computer Science
(ISSN: 0976-5697)
REVIEW ARTICLE

ICRACC 2018
(23rd–24th February 2018)

Available Online at www.ijarcs.info

Role of Soft Computing Techniques in Improving the Strength of Steganographic Systems

G.Umamaheswari
Research Scholar, Manonmaniam Sundaranar University
Asst. Professor, PG Dept. of Computer Science
SSS Jain College For Women, T-Nagar
Chennai, TamilNadu, India.
gumaganesh2011@gmail.com

Dr. C.P.Sumathi
Professor & Head, Dept. of Computer Science
SDNB Vaishnav College For Women, Chromepet,
Chennai, TamilNadu, India

ABSTRACT

In recent times steganographic systems that deal with concealing of secret data inside images have gained much attention. Robustness, capacity of embedding and imperceptibility are the major issues concerning embedding of secret data inside a cover. This paper analyses the recent advancements in this field by combining the soft computing techniques with steganographic algorithms for increasing the robustness, capacity and imperceptibility of the images used for embedding.

Keywords: Steganography, Least Significatn Bits (LSB), Soft Computing, Fuzzy Logic (FL), Neural Networks(NN), Genetic algorithm(GA), Support Vector Machines(SVM), Spatial domain, Transform Domain.

I. INTRODUCTION TO STEGANOGRAPHY

Steganography deals with secret communication of sensitive information. It deals with concealing secret data into other media like images, audio, video etc. When compared to cryptography, steganography provides an additional layer of security. Cryptography uses data scrambling with the intention of protecting it, so that it is not understood other than the intended recipient.

Steganography hides the very presence of data; in fact cryptography can be combined with steganography to increase the efficiency of steganographic systems. Although many algorithms exist for different ways of embedding secret data, the challenge in developing these kinds of systems is in protecting the secret embedded data that can withstand attacks. The combination of Soft computing techniques with steganographic algorithms helps to build a strong system that can withstand attacks. The Basic steganographic model is given in Figure 1 where the function f(x,m,k) represents the steganographic process.

The rapid growth in transmitting data through public channels has raised the bar for securing information. Steganography based information security is one such field that proposes techniques for hiding information that is being transmitted. An audio/ video/ image file can be used as a medium that is used to hide information. With regard to embedding data in images two popular techniques exist.

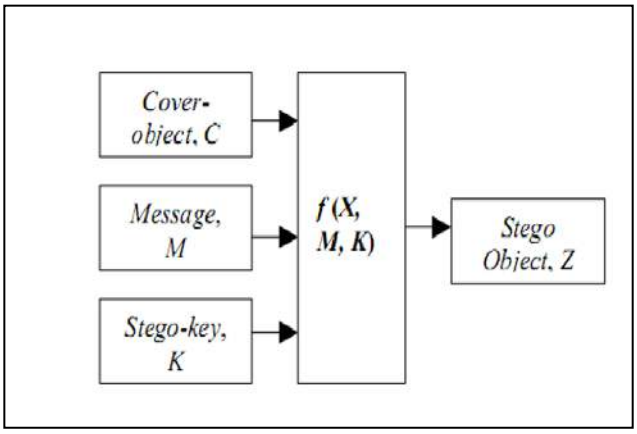


Figure 1: Basic Steganographic model [1].

1. Spatial domain secret embedding
2. Transform domain secret embedding

Spatial domain deals with embedding data straightforwardly in the spatial coordinates (i.e) the pixel values. The different techniques under the spatial domain include

- Least Significant Bit (LSB) substitution
- Pixel Value Differencing (PVD)
- Region of Interest methods [2].

The main features of spatial domain techniques are their

- Simplicity
- Less time complexity
- Ease of implementation

The main drawback of spatial domain techniques is that they cannot withstand attacks, a feature referred to as robustness.

The cover file is transformed to a different domain using any one of the transform domain techniques available and then data is hidden into the coefficients of the transformed domain.

The main highlight of these techniques is the robustness against attacks [3].

The Commonly used transform domain techniques are

- Discrete Cosine Transform (DCT)
- Discrete Fourier Transform (DFT)
- Discrete Wavelet Transform (DWT)

II. INTRODUCTION TO SOFT COMPUTING

Soft computing is a branch of computer science (sometimes referred to as computational intelligence) that is used to find approximate solutions to computationally hard tasks such as the solution of NP-complete problems, where it is found that there is no exact solution available in polynomial time. Soft computing is different from conventional (hard) computing in a sense that, it is tolerant to uncertainty, approximation, partial truth and imprecision. The main role model for Soft Computing is our Human mind.

The constituents of Soft Computing are Probabilistic Reasoning, Machine Learning, Fuzzy Logic, Evolutionary Computing etc.

Soft Computing techniques include Fuzzy logic, Genetic algorithms, Rough Sets, Neural networks and Support vector machines. These techniques intend at achieving robust, optimal, low cost and adaptive solutions for problems related to data hiding.

In recent times the combination of soft computing techniques along with the traditional ways of embedding secret data has gained much attention. In this paper we analyze several methods that propose to hide data using fuzzy logic, neural networks, genetic algorithm and support vector machines.

A. *Fuzzy Logic (FL)*

Fuzzy logic theory was introduced in 1965 by Zadeh [4]. Problems that require human like reasoning and inference are usually tacked by fuzzy logic.. Figure 2 represents the basic fuzzy logic model.

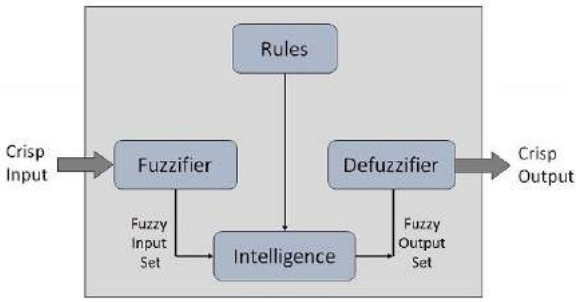


Figure 2: Basic Fuzzy Logic Model.

The process of fuzzification involves mapping of input values (mathematical) into fuzzy membership functions. To arrive at "crisp" output value they are mapped with fuzzy output membership functions by the process of de-fuzzification. The result can be used for decision and/or control purposes.

Process

1. All input values are fuzzified into fuzzy membership functions.
2. The fuzzy output functions are computed by application of all rules in the rule base.
3. The get "crisp" output values the process of De-fuzzification is done on fuzzy output functions [11].

Figure 3 shows the effect of information hiding fuzzy set.

The scheme of focus of attention is closely associated to information hiding. The collections of the elements that are hidden are viewed from the stand point of membership functions. Information hiding is achieved through normalization and is done by increasing or decreasing the level of α cut and is referred to as α information hiding. The hiding of elements about x is achieved [18].

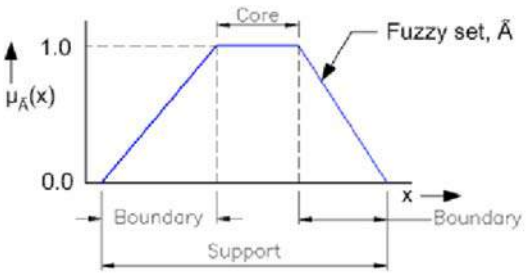


Figure 3: Effect of Information Hiding Fuzzy set.

A.saleema et.al [9] in their article suggests a hybrid fuzzy neural network model at the post processing phase that improves the quality of the stego image. The input features are subjected to fuzzy pre processing and the improved stego image is resilient to visual and statistical attacks and also the stego image is highly imperceptible

Sara sajasai et.al [10] has developed an intelligent novel adaptive steganography scheme that integrates Fuzzy Inference System (FIS) with the characteristics of Human Visual System (HVS). The scheme uses 36 rules that discover strong relationships between features and output classes.

Fatma Hassan Al-Rubbaiy in his article has proposed a system in which the cover image is divided into blocks of 10 x 10 and secret data in embedded only in the even sub images. Fuzzy Logic is used for dividing the blocks into odd and even sub images and also for encryption of secret message [12].

B. *Neural Network (NN)*

Neural Network is an optimization technique mainly used as a classifier. Neural networks are organised in layers and consists of neurons that are interconnected group of nodes. Neural networks can still continue in the event of the failure of an element, because of its parallel nature. Figure 4 shows the multi-layered perceptron model in which each node represented by a circle is a neuron and the arrows correspond to a link from the output of one neuron to the input of another neuron [6]. The working of the three layers in neural networks is as follows:

- Input Layer-> includes training set and trained target is passed as input to neural.
- Hidden Layer -> is concerned with the number of iterations at which the best result is achieved.
- Output Layer -> generates the final result.

Neural network accepts a signal as input to the neuron and is converted to a certain value using a function usually sigmoid function [7] given by equation 1 and outputs this value as output signal.

$$f(x) = \frac{1}{1+e^{-x}} \qquad (1)$$

Each synaptic link has a network weight. The network weight from unit i to unit j is represented by W_{ij}. O_i represents the output value for unit i. The Speed of the learning process normally has a value between 0 and 1 that is a constant and is referred to as Learning Rate. The main aim is to train the neural network to output a value of 1 or 0 as output signal.

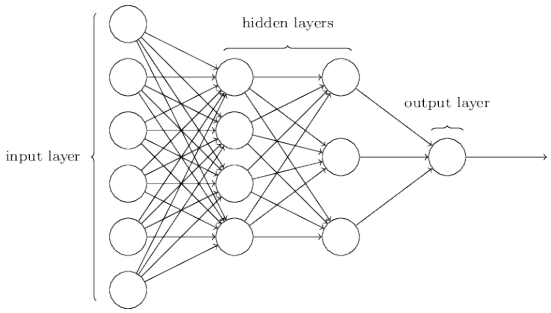


Figure 4: Multi-Layered perceptron model

In the article proposed by Anamika Sharma et.al. a new methodology in which wavelet transform is applied to both the cover image and the secret image is used. The process of fusion using neural network is used to get the stego image which is a merged image got from cover and secret image. The method is found to achieve robustness against image attacks like cropping, rotation and noising.

In Kensuke Naoe et. al's article [7] one network represents one binary digit for corresponding secret codes. The hidden layer consists of a number of neurons and the count of neurons is taken randomly as 10 neurons in this case. The proposed network uses the values that get converged from network weights and the extraction keys from coordinates of selected feature subblocks. The sender and the receiver shares the keys in order to extract proper hidden signals from the contents.

C. Genetic Algorithm (GA)

Genetic algorithms are modelled on evolutionary biology for search optimization. The natural theory of evolution which is based on the survival of the fittest approach is the basis of Genetic Algorithm. The next generation reproduces from only the fittest surviving individual.

In the field of steganography genetic algorithm is mainly used to find the best embedding position so that the image obtained after inserting secret message can withstand any type of attack.

A Component based architecture is proposed by Usha.B.A et. al. The main steps of genetic algorithm are

- [start] A random population of n chromosomes generated (suitable solution for the problem).
- [Fitness] For each chromosome in the population a fitness function f(x) is evaluated.
- [New population] the steps are repeated until a new population is found.
- [Selection] Two guardian chromosomes are selected as per their wellness using a fitness function.

  [Crossover] To frame another posterity (children) a hybrid likelihood is traversed. Posterity is a careful duplication of the parents when no cross over is performed.

  [Mutation] At every locus (position in chromosome) there is a possibility of a new posterity with mutation.

  [Accepting] Refers to the placing of new posterity in another general population.

  - [Replace] The newly created populace is utilized in the algorithm for the next run.
  - [Text] In the event of the end condition being satisfied stop and revisit the best solution in current population.
  - Goto the Fitness step. [5]

A secure steganographic framework is proposed by Shamimunnisabi et.al.[8]. Embedding of the secret text message is done in the coefficients of the Integer Wavelet Transform. The system is said to become robust by the usage of Genetic Algorithm. The method is found be tolerant to RS analysis. The message bits are embedded in 4 LSB's of the Integer Wavelet Transform coefficients using a mapping function. The process of Optimal pixel Adjustment is based on the fitness evaluation.

A secure and novel stegano-algorithm is suggested by dinesh kumar et.al. and is found to be extremely RS resistant. The combination of Genetic Algorithm with Integer Wavelet Transformis used in the proposed method. The coefficients of the wavelet transform are the locations for embedding the secret message. The process of optimal pixel adjustment using genetic algorithm and optimal pixel adjustment is done. A mapping function for all the blocks of the image is found using Genetic algorithm [13].

D. Support Vector Machine (SVM)

Support Vector machines are based on supervised learning models. Usually associated with learning algorithms SVM's

analyses data that is mainly used for regression and classification analysis.

In the article proposed by Akram AbdelQader et.al. [15] the cover and the secret images are divided into blocks of 8 x 8. Each block is subjected to Discrete Cosine Transform. Based on the linear SVM learning process for the DCT feature is done and bits of the cover image is replaced with the bits of secret data.

SVM is mainly used for characterization / relapse issues using a regulated machine learning calculation. A procedure referred to as bit trap is used to change information. The ideal limit of these changes is calculated. SVM has the high generalization ability to separate data into two classes, thus it is naturally suitable to classify the cover-image[16].

In the method proposed by Hanizan Shaker Hussain 1 et.al. [17] the cover image features are extracted. A SVM training dataset is constructed. The best parameter, the penalty parameter and the gamma value is found using cross validation technique and is used to train and generate the SVM function f(x).

TABLE I . ANALYSIS OF DIFFERENT SOFT COMPUTING APPROACHES.

S.No	Soft Computing Paradigm used	Cover Image & Size	PSNR (dB)	MSE / SSIM	Hiding Capacity / Speed/ payload ratio
1.	Genetic Algorithm [8]	Lena 512x512	46.83 (K=3)*	NA	1048576 (bits)
2.	Genetic Algorithm [13]	Lena 512x512	39.94	NA	50%
3.	Hybrid Fuzzy Neural Network [9]	Lena	39.47	0.8 (SSIM)	1.43 sec
4.	Fuzzy Logic [10]	Lena 512x 512	52.03	NA	0.5 bpp
5.	Genetic Algorithm [14]	Lena 512x 512	34.68	22.11	1048576 (bits) 50%
6.	Support Vector Machine [15]	Elephant 256 x 256	44.305	NA	64 x 64 image
7.	Support Vector Machine [17]	Lena NA	49.86	NA	NA

*K represents the no of LSB's used.

ACKNOWLEDGMENT

My heartfelt thanks to Dr.C.P.Sumathi who gave expert guidance and support in completing this article.

REFERENCES

[1] https://image.slidesharecdn.com/manika-150520175836-lva1-app6892/95/steganography-13-638.jpg?cb=1432144833

[2] Ratnakirti Roy, Anirban Sarkar, Suramoy Changder, " Chaos Based Edge Adaptive Image Steganography", International Conference on Computational Intelligence Modelling Techniques & Applications (CIMTA)2013, Proceedia Technology(2013), 138-146. (www.Sciencedirect.com)

[3] Taozhang, Shuai Ren, " Application of CL multi-wavelet transform and DCT in Information Hiding Algorithm", International Journal of Computer Networks and Information Security, 2011, 1, 11-17.

[4] Zadeh, L.A(2005), "The Concept of a Generalized Constraint – A Bridge from Natural Languages to Mathematics NAFIPS 2005 – Annual Meeting of the North American Fuzzy Information Processing Society.

[5] Usha.B.A, et.al " High Capacity Data Embedding Method in Image Steganography using Genetic Algorithm", International Journal of Computer Applications(0975 – 8887), Volume 121-No.14, July 2015, 30-33.

[6] Anamika sharma, Ajay Kushwaha, " Image Steganography Scheme Using Neural Network in Wavelet Transform Domain", International Journal of Scinetific Engineering and Research , ISSN (online) 2347-3878, Volume 3 Issue 10, October 2015, 153-158.

[7] Kensuke Naoe, Yoshiyasu Takefuji, "Damageless Information Hiding using Neural Network on YCbCr Domain", International Journal of Computer Science and Network Security, Vo8, No.9, September 2008.

[8] Shamimunnisabi, Cauvery N.K, "Empirical Computation of RS-Analysis for Building Robust Steganography Using Integer Wavelet Transform and Genetic Algorithm", International Journal of Engineering Trends and Technology, Volume 3 Issue 3 , 2012, ISSN:2231-5381, 448-456.

[9] Saleema.A, Dr.T.Amarunnishad, " A New Steganograhy Algorithm Using Hybrid Fuzzy Neural Network", International Conference on Emerging Trends in Engineering, Science and Technology- 2015, 2212-0173, Procedia Technology 24(2016), 1566 -1574.

[10] Sara Sajasi, Amir Masoud Eftekhari Moghadam, " A high Quality Image Steganography Scheme Based on Fuzzy Inference System", 13th Iranian Conference on Fuzzy Systems , 978-1-4799-1228-5/13 © 2013 IEEE.

[11] https://en.wikipedia.org/wiki/Fuzzy_logic

[12] Fatma Hassan Al-Rubbaiy, " Concealment of Information and encryption by using Fuzzy Technique", Journal of the college of Basic Education, Volume 16, Issue 69, 25-34.

[13] Dinesh Kumar, Narendra Yadav, "High Embedding Capacity and Secured Steganographic Model by Using RS based Genetic Algorithm and IWT", Research & Reviews : Journal of Engineering and Technology ISSN :2319-9873.

[14] Medisetty Nagendra Kumar, S.Srividya, "Genetic Algorithm Based Color Image Steganography using Integer Wavelet Transform and Optimal Pixel Adjustment Process", International Journal of Innovative Technology Exploring Engineering (IJITEE) ISSN :2278-3075, Volume 3, Issue 5, October 2013, 60-65.

[15] Akram AbdelQader, Fadel AlTamimi, " A Novel Image Steganography Approach Using Multi Layers DCT features based on Support Vector Machine Classifier", The International Journal of Multimedia & its Applications (IJMA) Vol.9, No.1, February 2017, 1-10.

[16] Rohit Tanwar, Sona Malhotra, "Scope of Support Vector Machine in Steganography", IOP Conf.Series: Materials and Engineering 225 (2017) 012077 doi:10.1088/1757-899X/225/1/012077.

[17] Hanizan Shaker Hussain, Roshidi Din, Aida Musthapa, Fawwaz Zamir Mansor, " LSB Algorithm Based on Support Vector Machine in Digital Image Steganography", Journal of Telecommunication, Electronic and Computer Engineering, E-ISSN :2289-8131 Vol 9, N0.2-12, 13-18.

[18] Witold Petrycz, Andrezej skowron, Vladik Kreinovich, "Handbook of Granular Computing", page 112.

Shasun Srrishti 2k17

CSI 2017

Workshop

Workshop on “Python Programming” for School Teachers

SHRI SHANKARLAL SUNDARBAI SHASUN JAIN COLLEGE FOR WOMEN

A Unit of Sri S. S. Jain Educational Society | Affiliated to University of Madras
Accredited with 'A' Grade by NAAC | An ISO 9001:2015 Certified Institution