N.R. Shetty
L.M. Patnaik
N.H. Prasad
N. Nalini *Editors*

Emerging Research in Computing, Information, Communication and Applications

ERCICA 2016

Springer

Emerging Research in Computing, Information, Communication and Applications

N.R. Shetty · L.M. Patnaik
N.H. Prasad · N. Nalini
Editors

Emerging Research in Computing, Information, Communication and Applications

ERCICA 2016

Springer

Editors
N.R. Shetty
Nitte Meenakshi Institute of Technology
Nitte Education Trust
Bengaluru, Karnataka
India

and

Central University of Karnataka
Gulbarga, Karnataka
India

L.M. Patnaik
IISc and INSA
National Institute of Advanced Studies
Bengaluru, Karnataka
India

N.H. Prasad
Master of Computer Applications
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka
India

N. Nalini
Nitte Meenakshi Institute of Technology
Bengaluru, Karnataka
India

ISBN 978-981-10-4740-4 ISBN 978-981-10-4741-1 (eBook)
https://doi.org/10.1007/978-981-10-4741-1

Library of Congress Control Number: 2017938319

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The Fourth International Conference on “Emerging Research in Computing, Information, Communication and Applications”, ERCICA 2016 is an annual event organized at the Nitte Meenakshi Institute of Technology (NMIT), Yelahanka, Bangalore, India.

ERCICA aims to provide an interdisciplinary forum for engineers and scientists to discuss and promote research and technology in the emerging areas of the core topics of Computing, Information, Communication, and their Applications. The conference brings researchers, educators, professional engineers, and technologists under a single forum to discuss developments in research.

ERCICA-16 received more than 450 papers from all over the world, viz. from China, UK, Africa, Saudi Arabia, and India. The ERCICA Technical Review Committee has followed all necessary steps to screen more than 450 papers by going through six rounds of quality checks on each paper before selection for Presentation/Publication in Springer proceedings. The acceptance ratio is only 1:4.

Bengaluru, India — N.H. Prasad
July 2016 — N. Nalini

Organizing Committee

ERCICA-2016
The Fourth International Conference on "Emerging Research in Computing, Information, Communication and Applications", (ERCICA-2016) was held during 29–30 July 2016 at Nitte Meenakshi Institute of Technology(NMIT), Bangalore and organized by the Departments of CSE and MCA, NMIT.

Chief Patrons
Dr. N.V. Hegde, President, Nitte Education Trust
Dr. N.R. Shetty, Advisor, Nitte Education Trust

Conference Chair
Dr. H.C. Nagaraj, Principal, NMIT, Bangalore

Program Chairs
Dr. Prasad N. Hamsavath, HOD, MCA, NMIT, Bangalore
Dr. Nalini N., Professor, CSE, NMIT, Bangalore

Publisher
Springer Nature

Advisory Chairs
Dr. Jharna Majumdar, Dean (R&D), NMIT, Bangalore
Prof. Ranganatha Setty, Dean (Academic), NMIT, Bangalore

Advisory Committee
Prof. L.M. Patnaik, INSA Senior Scientist, NIAS, Bangalore
Dr. Kalidas Shetty Founding Director of Global Institute of Food Security and International Agriculture (GIFSIA), Professor of Plant Science and Associate Vice-President of International Partnerships and Collaborations, North Dakota State University, Fargo, USA
Dr. K.D. Nayak, OS and CC, RandD (MED and MIST), DRDO, India
Dr. Sathish Udpa, Dean and Professor, Michigan State University, Michigan

Dr. B.S. Sonde, Vice Chancellor, Goa University, Goa, India
Dr. D.K. Subramanian, Dean and Professor (Rtd.), IISc, Bangalore, India
Dr. K.R. Venugopal, Principal, UVCE, Bangalore, India
Dr. K. Chidananda Gowda, Former Vice Chancellor, Kuvempu University, Shimoga, India

Program Committee

Dr. Savitri Bevinakoppa, Professional Development and Scholarship Coordinator, School of IT and Engineering, Melbourne Institute of Technology (MIT), Australia
Dr. Ramprasad P., Professor, Department of CSE and IT, Manipal University, Dubai
Prof. Uma Shama, Professor of Mathematics and Computer Science, Co-Director, GeoGraphics Laboratory, Science and Mathematics Center, Bridgewater State University, USA
Dr. P.V. Ananda Mohan, Advisor CDAC
Prof. Zuriati Binti Ahmad Zukarnain, Professor, Department of Communication Technology and Networks, University Putra Malaysia, Malaysia
Dr. Narushan Pillay, School of Engineering, University of KWAZULU-Natul, Durban, South Africa
Prof. Sonajharia Minz, Professor, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi
Dr. Vikram Goyal, Associate Professor, CSE, IIIT-New Delhi
Dr. K.N. Hari Bhat, Department of ECE, NMIT, Bangalore

Organizing Co-Chairs

Dr. Venkatesh K., Professor and Head, CSE, NMIT
Prof. Sitaram Yaji, Professor, ECE, NMIT
Dr. Sanjay H.A., Professor and Head, ISE, NMIT
Dr. M.N. Thippeswamy, Professor, CSE, NMIT
Prof. Sankar Dasiga, Professor, ECE, NMIT
Dr. Kavitha Sooda, Associate Professor, CSE, NMIT

Acknowledgements

First of all, we would like to thank Prof. N.R. Shetty who has always been the guiding force behind this event's success. It was his dream that we have striven to make a reality. Our thanks to Prof. L.M. Patnaik, who has monitored the whole activity of the conference from the beginning till its successful end.

Our special thanks to Springer Nature and especially the editorial staff who were patient, meticulous, and friendly with their constructive criticism on the quality of papers and outright rejection at times without compromising the quality of the papers as they are always known for publishing the best international papers.

We would like to express our gratitude to all the review committee members of all the themes of Computing, Information, Communication, and Applications and the best paper award review committee members.

Finally, we would like to express our heartfelt gratitude and warmest thanks to the ERCICA 2016 organizing committee members for their hard work and outstanding efforts. We know how much time and energy this assignment demanded, and we deeply appreciate all the efforts to make it a grand success.

Our special thanks to all the authors who have contributed to publish their research work in this conference and participated to make this conference a grand success. Thanks to everyone who have directly or indirectly contributed to the success of this conference ERCICA 2016.

Regards
Program Chairs
ERCICA 2016

About the Conference

ERCICA 2016

The Fourth International Conference on “Emerging Research in Computing, Information, Communication and Applications”, ERCICA-2016 is an annual event jointly organized by the Departments of CSE and MCA during 29–30 july, 2016 at Nitte Meenakshi Institute of Technology (NMIT), Yelahanka, Bangalore, India. The conference is supported by the Technical Education Quality Improvement Program (TEQIP-II).

ERCICA-2016 is organized under the patronage of Prof. N.R. Shetty, Advisor, Nitte Education Trust, Dr. L.M. Patnaik, Technical Advisor, NMIT, Dr. H.C. Nagaraj, Principal served as the Conference Chair and the Program Chairs of the conference were Dr. N.H. Prasad, Professor and Head, MCA and Dr. N. Nalini, Professor, CSE, NMIT, Bangalore, Karnataka.

ERCICA aims to provide an interdisciplinary forum for researchers, engineers, and scientists to discuss and promote research and technology in the thrust areas of computing, information, communication, and applications. The conference will bring researchers, educators, professional engineers, and technologists into a single forum in order to discuss and debate the emerging research in the above areas.

For ERCICA 2017, authors are invited to submit the manuscripts of their original and unpublished research contributions to ercica.chair@gmail.com (ERCICA website: http://nmit.ac.in/ercica.html). All the submitted papers will go through a peer review process, and the corresponding authors will be notified about the outcome of the review process. There will be six rounds of quality checks on each paper before selection for Presentation/Publication. Authors of the selected papers may present their papers during the conference.

Contents

Part II Information

About the Editors

Prof. N.R. Shetty is a Chancellor of Central University of Karnataka, Kalaburagi, and Chairman of the Review Commission for the State Private University, Karnataka. He is currently serving as an advisor to the Nitte Meenakshi Institute of Technology (NMIT), Bangalore. He is also founder Vice-President of the International Federation of Engineering Education Societies (IFEES), Washington DC, USA. He served as Vice Chancellor of Bangalore University for two terms and President of the ISTE, New Delhi, three terms. He was also a Member of the Executive Committee of the AICTE and Chairman of its South West Region Committee.

Prof. L.M. Patnaik obtained his Ph.D. in Real-time Systems in 1978, and his D.Sc. in Computer Systems and Architectures in 1989, both from the Indian Institute of Science, Bangalore. From 2008 to 2011, he was Vice Chancellor of the Defence Institute of Advanced Technology, Deemed University, Pune. Currently, he is an Honorary Professor with the Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore, and INSA Senior Scientist and Adjunct Professor with the National Institute of Advanced Studies, Bangalore.

Dr. N.H. Prasad is currently working as Professor and Head of the Department of Master of Computer Applications at Nitte Meenakshi Institute of Technology, Bangalore. He completed his Ph.D. at Jawaharlal Nehru University, New Delhi, India. Dr. N.H. Prasad has more than 12 years of experience in different roles in both public and private sector enterprises, including the Ministry of Human Resource and Development, New Delhi, Government of India. He has received the prestigious “Dr. Abdul Kalam Life Time Achievement Award” and also received a “Young Faculty” award at the 2nd Academic Brilliance Awards.

Dr. N. Nalini is currently working as a Professor at the Department of Computer Science and Engineering at Nitte Meenakshi Institute of Technology, Bangalore. She received her MS from BITS, Pilani in 1999 and her Ph.D. from Visvesvaraya Technological University in 2007. She has more than 17 years of teaching and 10 years of research experience. She has several international publications to her credit and has received the “Bharath Jyothi Award” by Dr. Bhishma Narain Singh, Former Governor of Tamil Nadu and Assam, given out by the India International Friendship Society. She is a lifetime member of the ISTE, CSI, ACEEE, and IIFS.

Part I
Computing

Development of Accurate Face Recognition Process Flow for Authentication

Tripty Singh and C. Josna Mary

Abstract Rapid and tough tracking of multiple faces have many crucial applications in various fields like video surveillance and video conferencing. Hence, it gained lot of momentum from computer vision researchers. Initialization, tracking and display are the three important steps in real-time tracking of multiple faces in high-resolution videos. Out of which, tracking is quite intensive in computation. Object detection and tracking are important in many computer vision applications including activity recognition, automotive safety, and surveillance. In this paper, a computational model of real-time face tracking has been presented. The tracking problem into three parts: the region extractor is based on the integration of skin-color, motion, and silhouette features, while the face detector uses a simple, rule-based face detection algorithm, and SVM. Real-time detection and tracking implementation has been carried out and results presented.

Keywords Face detection · Skin recognition · Color · Tracking
Feature extraction · Color segmentation · Noise filtering

1 Introduction

Face recognition has been identified as a vital requirement in information security and video surveillance, identity authentication for driver license or credit card and access control. The concern about security leads the way toward face detection and tracking. Both of them have many crucial applications in various fields like teleconferencing, facial gesture recognition, robotics, telecommunication, HCI (human–computer interactions). Human detection techniques have been classified into

---

T. Singh · C. Josna Mary (✉)
Department of CSE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bangalore 560035, India
e-mail: josnam656@gmail.com

T. Singh
e-mail: tripty_smart@yahoo.co.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_1

subwindow-based approaches and part-based approaches. Subwindow-based approach has the features covariance matrices, histograms of oriented gradients (HOG) and multilevel versions of HOG. In part-based approach, the body parts are segmented and it is separately detected and the results are combined at the end. The detection algorithm includes two basic processing steps: feature extraction and detection. Viola and Jones [1] have designed a face detector which uses cascade of simple classifiers based on the Haar wavelet features. Many enhanced face detection and feature extraction algorithms are inspired or adapted from this approach.

Nowadays great interest on video processing is arising. An increasing interest on video processing applications has been observed recently. The goal of the work is to design a method for the tracking of people in this world that includes many techniques for the analysis of color, movement, and faces. The expected tasks from visual-appearance-based interface can be: (1) Detection and localization of face (2) Accurate facial feature detection and tracking (3) Facial expression analysis (4) Recognition of mental states from sequences arises from face expressions and head movements. Each listed task has its own limitations and challenges. Here, the study aims on the task of detecting face and recognition. The system for detecting face can generally be represented in Fig. 1. For implementing the face detection by video-based and tracking system, one of the parameter, color (facial color) has been identified. The color of the skin is considered as more advantageous in tracking feature because of its fast processing capability than any other facial features. Under the conditions for lighting, it is observed that the facial color is orientation invariant. The skin color of face is depends on the YUV color space [2]. Value of noncritical illuminance is eliminated which has been employed. The non-consideration of noncritical illuminance value has been found to be not affecting the successful detection of facial skin color. The movement analysis subsystem employs background subtraction which selectively eliminates the background which has moving objects and their shadows.

Skin Detector approach has been adopted for color analysis. Face detection has been incorporated under the face analysis section. The feature extraction and feature evaluation system has utilized the rectangular features of Haar wavelets which resulted in fast computation. Adaboost is a boosting strategy which has been

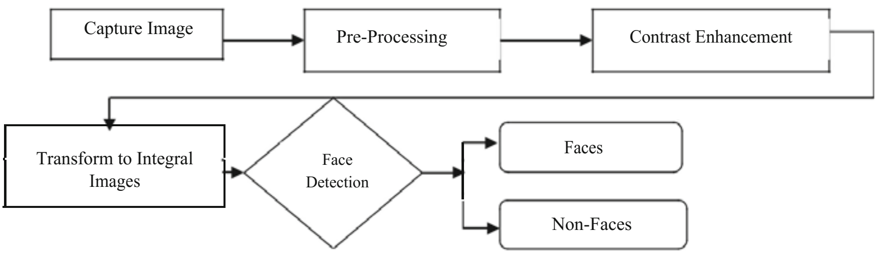


Fig. 1 Face detection system diagram

employed for the training of the detectors. For the proposed method in face authentication technique, the requirements which satisfies the authentication are:

- Invariable to the face size and face slant
- Invariable to lighting variations
- New subject can be integrated into the system without considering others features

A. **Brief Overview of Video Surveillance:** Using video analytics is the latest development in video surveillance. DVRs can store digital video and can analyze their captured images. Therefore, they can automatize the tasks related to investigation purposes. For developing Video Surveillance and Monitoring (VSAM) [3], the Defense Advanced Research Projects Agency (DARPA) Information System Office starts a program in the year 1997 which was a 3-year program. The main goal was to establish a technology based on video for using in civic and battleground areas. The work here briefly covered the basic problems in tracking face, detection, and in case of multi-camera system.

B. **Related Work:** There are a lot of methods proposed to improve the face detection over last decades which are mainly classified into two: (a) Knowledge-based method (b) Feature invariant approaches. In knowledge-based method, the main goal is to find the invariable face features in a complex environment. Hence localizing the face position. Determining whether a face is appeared in an image or not by studying the relationships among the features [4]. The invariant method approaches to invariant features, positions, brightness, etc. for detecting the human faces.

Different approaches have been proposed for segmentation of moving objects and features, like frame difference, double frame difference, and background suppression. Background suppression has been the most extensively applied technique in case of lack of a priori knowledge about the objective. Jain and Bolle's work has demonstrated [5] good achievement in video sequence analysis. Processes mentioned in the above-mentioned work has selected updating of background [6], steps for verification in order to include ghosts into the background model and usage of HSV color space in order to deal shadows. The method proposed by Jones and Viola [5] exceeds the other systems. The processing speed is high for the system proposed by Viola and Jones. Rectangular features which is a type of Haar Wavelets, uses deluge of filters which eliminates the images of non-face. The strategy for boosting is taken as asymmetrical adaboost for the training of detectors. In addition, there has been a module for color analysis which is rooted upon a standard skin detector algorithm and it has been found that by using color images there is an improvement in the efficiency of the system. Also, there has been a proposal for multiple-person tracking as seen in the work of Li and Chellappa [7]. The multi-person tracking system has two modules—1. Face detector module 2. Region extractor module. The face detector module uses a face detection algorithm, SVM, etc. The region extractor uses motion, silhouette features, integration of skin color, etc. The face detector module is simple than region extractor module. This system has been found to be less robust to in real world.

2 Background Study

A. Evolution in Face Detection: The work for face detection begun as the commencement of 1970s. In the beginning, simple approaches were used. Due to various thinking about plain background, passport size frontal face scenario, the work becomes rigid. Hence, the development in research area remained dormant since 1990s. The work becomes a reality after the evolution of face recognition and video coding systems. Decades back a research on face detection had taken place. Segmentation ideas were presented on that by using motion, color, etc. For face detection from large crowded situation neural networks and statistics are also used. The technique of face detection is categorized into two since the technique requires a priori information about the face [12, 13, 14]. These categories are differentiated by various methods used for handling the knowledge about face. Feature-based and image-based techniques are the two major detection methods used. Feature-based method involves low-level analysis and feature analysis; whereas in image-based analysis involves liner subspace methods and neural networks. Statistical approaches also come under image-based analysis.

B. Knowledge base for Face Recognition: The first technique uses knowledge of face and it follows detection methodology. From knowledge base analysis the low-level features are obtained. The color of the face skin and its geometry is analyzed at various levels. The next class has image-based representations which have been classified directly into group of face by using some training algorithms. It uses mapping schemes and training schemes for incorporating knowledge into the system. On compared to feature-based approach, image-based techniques have better performance. Mainly, there exists three methods for localization of facial features: appearance, geometry, and structure. Preprocessing is the first stage in almost all localization. It uses vertical and horizontal grey level or edge field projections. Coarse-to-fine localization is another approach used for reducing the computational load [8]. Sometimes uses color-based Segmentation schemes for identifying face first and then for lips. The important method is to extend the whole face detection methods in order to search for small features at a high resolution. Feris et al. [9] uses Gabor wavelet networks (GWNs) for identifying the region of face approximately. It then uses smaller GWNs for looking into almost eight individual features, such as corners of eyes, nostrils, mouth corners, etc. The goal of the appearance-based method is the representation of facial features.

Geometric-based methods have been found to have more information about position of face, searching of landmark by angles, distances, areas, etc. [15, 16]. There are many methods for the improvement of detection of face that are classified mainly into two: (a) Knowledge-based method (b) Feature invariant approaches. In order to describe the relations from features of a face and the face which is detected, a statistical model was built. (c) Matching of Template: The template employed with features of face has a better matching operation.

C. Face Identification: The face detection and recognition have a wide variety of approaches, out of which it is a great difficult task to concentrate on a general case in much detail. Mostly, there are three stages. The first stage is known as image processing. In this stage, somehow the input images are transformed and extract only the information which is important. Thus, it is more accurate. The classification stage needs the information which is obtained from the extraction done by processing stage. The major two tasks are lighting correction and filtering. The function of lighting task is to reduce the variation of intensity in lighting and orientation. The task of filtering is to obtain the needed information at some particular frequencies and directions. The second stage is known as image search. For the detailed evaluation in further this stage chooses only the needed portion of the image which are transformed. The next two stages is known as pattern classification. In this stage, valuation is done for the outputs obtained from first two stages. It then classifies it as background or target, face, or non face. The face detectionprocess flow is as shown in Fig. 2. To identify a person, facial images are best in biometric. This is more relevant in application areas, where secret or underground works are going on. The face recognition can be applicable from static to dynamic and identification of face from the crowded background such as airport. The face recognition is based on the following facts: (a) Features of eyes, eyebrows, lips, nose, chin, etc. (b) By comprehensive analysis, the face can be shown as the combination of the weighted canonical faces [10]. A question will arise when the face become enough for analyzing the tasks of real-time, real-world tracking, etc. A related multiple-person tracking proposed here has two modules such as face detector module [2] and region extractor module.

A. ***Method Description:*** The basic processing steps in a face recognition scheme are as follows:

1. Collect and capture the frame for input.
2. Observe the faces in the input frame.
3. Analyze whether the face observed is existing or not.

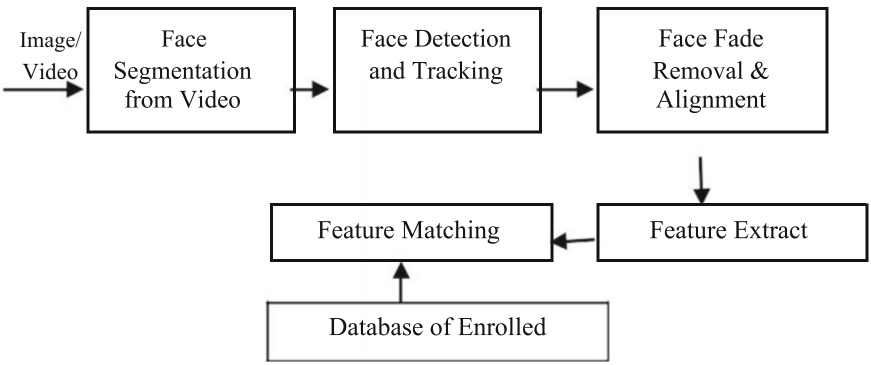


Fig. 2 Face recognition process flow

4. Template extraction is done and makes the histogram of face.
5. Determine Particle initialization with current spot and mass as zero.
6. Calculate the similarity coefficient.
7. Assign weight for each particle.
8. Sort the particle as per the weight.
9. Sketch the face surface with a rectangle with maximum weight and finally demonstrate on the screen.

The proposed method has been aimed at tracking and thus recognizing multiple human faces. The algorithm has the camera which acquires the image and then the skin colored blobs has been detected. The object hypotheses that has been maintained along with detected blobs are then associated.

3 System Study

The proposed system includes detection of foreground, segmentation of blob, tracking, etc. The diagram for proposed system is shown in Fig. 4. System has been designed with major subsystems such as Movement, Color, and Face Analysis [6].

A. Movement Analysis Module: This analysis method includes subsystems such as:

(i) ***Background Suppression, Shadow Detection***: In order to obtain the foreground images, a system known as background subtraction system has been employed. This can be extracted from the difference between current image and reference images. Shadow detection has been carried out after background suppression. The purpose of the shadow detection is to remove the unwanted sounds, noises, motions which is followed from some morphological operations. And finally foreground mask images will obtain.

(ii) ***Blob Segmentation, Noise Filtering***: Due to unwanted noises and sounds, each objects may have more than one blobs. From the foreground images blobs are segmented. The technique used for obtaining the whole object from the segmented parts is known as Blob merge. A blob having small area is considered as noise.

(iii) ***Blob Analysis and Tracking***: Using eight connectivity, blobs have been identified by the system. The average speed is computed with respect to frame differences.

(iv) ***Background Update:*** The computation of a background model is through the combination of array of previously obtained frames and the background.

B. Color Analysis Module: Under this, the actual image is get converted to its chromatic colors first. After the conversion, it is then compared with the skin color model. The skin color occurs in very small portion of the chromatic color space. Therefore, the process of matching is very fast (Fig. 3).

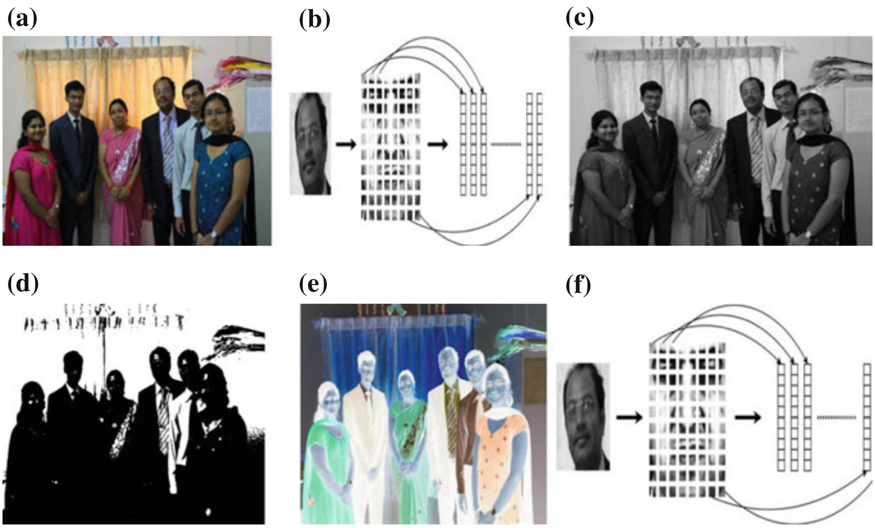


Fig. 3 **a** Original image captured, **b** Future analysis in spatio domain, **c** 32-grey level mapping of original image captured, **d** 2-grey level mapping of original image captured, **e** (255-n) grey level mapping of original image captured, **f** (255-n) grey level mapping of original image captured

(i) ***Identifying the region of skin:*** All images especially RGB images are transformed into 256 greyscale images with respect to the color distribution by multilayer perceptron. A low density point describes that the possibility to be a skin point is low, if not, the possibility is high. This greyscale image is to be segmented by choosing a threshold that result in binary image.
(ii) ***Skin Expansion:*** A 5 × 5 morphological opening technique is applied in order to eliminate some noise signals from selected points. After that dilation process of 3 × 3 square structuring is done.
(iii) ***Detection of skin blob:*** Labeling based on the region is performed here in order to compute the blobs of skin. Some portions of the blobs are noise.
(iv) ***Verification of face:*** This procedure includes:

(a) Choosing the portion of face. (b) Cosine transformation and choosing low frequency coefficient. (c) Dividing the face area into various subareas. (d) Choosing negative samples. (e) SVM training. The correlation property among the samples must be small for choosing negative sample. The images of such samples are produced in random way which contains vast information about various patterns.

C. Subsystem for Face Analysis: This approach has been the section neural network is developed based on the function related to radial.
(i) ***Detection of Face***: For detecting the face the system has a detector that has a group of filters that rejects the non-faces. It permits the faces to move on to

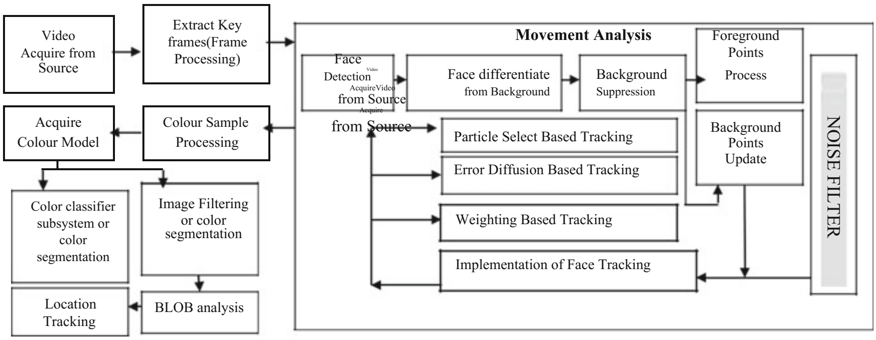


Fig. 4 Overall system block diagram

the next stage of cascade. The areas of hand and face are detected and tracked using propagated pixel hypothesis algorithm. This algorithm has the property to alter their shapes, to move in complex paths etc.

(ii) ***Overlapping of detection***: From the face detection process a window known as face windows are obtained. It is then processed further for estimating the size and orientation of finally detected face. This method is for removing the incorrect detections occurred and grouping the right detections.

(iii) ***Tracking Face:*** This approach is for removing the incorrect detections. This task is applied for inter frame operation. The detections belonging to sequence of frames is considered as same face, which uses same heuristic for overlapping of detection [3] (Fig. 4).

4 System Design

The major steps covered under the system design have been found to be:

(i) Preparing input/output specifications.
(ii) Detail about the plan of implementation.
(iii) Determine the design before implementing the system.

Step 1 Detection of Object: In this step the objects are detected. The process of detection is then applied to those cameras kept in those surroundings.

Step 2 Dimensional Object Tracking: From the above step we got the objects which are detected. These are tracked in all cameras on the basis of appearance model of objects [11].

Step 3 3 Dimensional Tracking of Object: Objects in two dimensions is joined together for locating and tracing objects in a three-dimensional system. The result obtained in this method is clubbed within the same objects.

Step 4 3 Dimensional Detection of Head: In this step, the orientation of head is detected. It uses the technique known as Head Detection Technique which was detailed early. First, a three-dimensional track is set up. Corresponding to two-sdimensional view the head is detected at first and its three-dimensional position is been triangulated.

Step 5 Assignment of Active Camera: In this step, it detects the camera that can be used for doing the task which is to done in this step.

Step 6 Camera Control based on three-dimensional positions: A three-dimensional orientation of the head and a PTZ camera is associated to object. It makes the camera which is active to be focus on the location of head. There exist many methods for zooming parameters and for controlling the tilt. An example is given below. The zoom parameter is directly proportional to the object distance from the camera and is indirectly proportional to the objects speed.

Step 7 Camera Control Based on Detection of Face: After detecting the face the camera is positioned on to the center of the face and increases the zoom level. The coordinates of the corresponding image is transformed into its specific pan/slide angles by giving the intrinsic calibration parameters of camera and the current zoom level. A damping factor is set up for eliminating the instability in feedback.

5 System Implementation

ASP.NET is the programming platform used for the multi face tracking system implementation. This platform is better for the implementation of the system since it is able to do the addition of libraries and for drawing forms fastly. System uses the real-world sequence videos which are of 320 $\times$ 240 size of pixel. Those videos are taken with a digital camera of 29.4 frames per second arrival rate and the compression is done in .avi resulting images of low quality. Those videos having more than four faces occur on at least 30 frames each face. Athlon 1.2 GHz computer was used. The rates of frame were 8.5 frames per second. It generates exact output for the inputs given. The accuracy of the expected result is verified. All the code units are to be verified before giving to the software. Testing procedure was done and successful tests produce the correct results. The position and location of eyes, area of face, are to be observed. The feature vector for the face can be obtained from the images of face,

$$P[q, r] = Q(s[q, r], t[q, r]) \tag{1}$$

where $P[q, r]$ is image; Q is the reflectance map; $s[q, r]$ and $t[x, y]$ are shape gradients. With an assumption of a Lambertian area and light source the formula can be equated as given below:

Fig. 5 Results on video for the system. The face detection windows are shown. All of them were detected

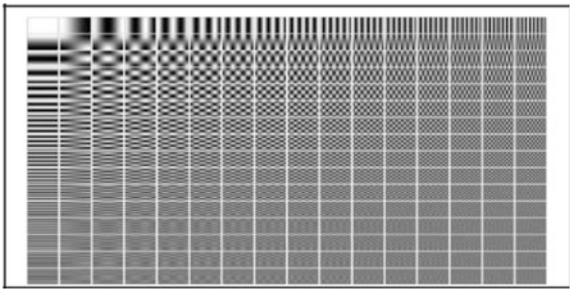


Fig. 6 Convolution kernels

$$\rho \cos\theta \,\text{or}\, I = \frac{\rho m M^s n N s}{m + n + M + N_2} \tag{2}$$

The Gradient Measure has 2 × 2 kernels of pairs are represented in the figure below. The other kernel is rotated by 90° (Figs. 5 and 6).

Those kernels are architecture in such a way that it responds to those edges which are running at 45° proportional to grid of pixel. The kernels can be applicable to those input images separately. It is applicable for obtaining the unique gradient measurement in each and every (say *Gx* & *Gy*). The gradient measure can be computed by the combination of these two which can be shown as:

$$G = G_x + G_y$$

Here, the determining method for Discrete Cosine Transformation for an image is presented briefly [7]. The interpretation of DCT for 2 dimensional images are shown below, where, $x(m, n)$ are the images and $N \times N$ is the dimension.

$$C(k, l) = 0 < k, l \leq N - 1$$

The low sub band (Y_{LL}) for the image is shown as:

$$Y_{LL}(p,\ q) = 1/4\{y(2p, 2q) + y(2p+1, 2n) + (2p, 2n+1) + y(2p+1, 2q+1)\}, 0 \le p, q \le Q/2 - 1 \tag{3}$$

Let S_{LL} (m, n), $0 \le m$, $1 \le N/2-1$ be the 2D DCT of $Y_{LL}(p, q)$. Approximate sub band for DCT is shown as:

$$C(k,l) = \begin{pmatrix} 4\cos\left(\frac{\pi k}{2N}\right)\cos\left(\frac{\pi l}{2N}\right)C_{LL}(k,l), & k,l = 0,1,\ldots\ldots,\frac{N}{2}-1 \\ 0, & \text{otherwise} \end{pmatrix} \tag{4}$$

From the DCT definition the sub band DCT is multiplied

$$\text{with } 4\cos\left(\frac{\pi k}{2N}\right)\cos\left(\frac{\pi l}{2N}\right)$$

1D DCT is computed for the rows and then for the columns are computed. This is computed by differentiating the Eq. 1 into column and row.

6 Performance Estimation

Surveillance system has a great effort in associating and describing the explanations for the measuring the performance. Hence, performance measurement is a great task [2, 17, 18]. Surveillance scheme has errors of two types:

False Positive: The error which occurs due to the false detection by a system or a false recognition which does not exists in that scenario. An example is given here. Consider, a system which monitors a guarded area may detect the physical movement in a particular area even if there is no physical object but there has lighting variations.

False Negative: The error occurs when a system fails to identify an object which is to actually identify. For example, a system which checks the guarded area is failed to detect the peoples wearing a dress with the same color and background of the environment given. Different methods for the detecting and tracking of faces are presented. The main aim is to capture the best close up images of the people. Its quality depends on the underlying components accuracy.

7 Results

The performance result is obtained for 25 subjects with a 160 × 120 resolution camera. For the detection of face area the movement information is also needed because humans cannot sit for a long while. The face area varies only when the

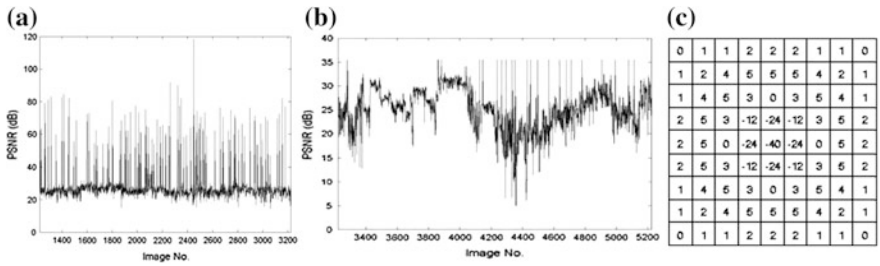


Fig. 7 **a** PSNR plot before quantization and after quantization, **b** RMSE plot before and after Q, **c** Laplace of Gaussian function with Gaussian = 1.4

Table 1 Results of recognition performance

Sr no	Database	Image measure (%)	Gradient measure (%)	Illumination-Invariant measure (%)	PCA (%)	LDA (%)	PSNR (DB)	PSNR grey	RMSE double
1	Face 1	88.3	85.3	85.3	83.3	83.3	34.64	34.83	34.05
2	Face 2	86.9	87.9	84.3	81.3	81.3	32.43	32.82	29.72
3	Face 3	87.3	86.5	82.5	87.9	83.5	34.64	34.33	34.95
4	Face 4	87.5	88.2	85.3	86.5	84.3	32.43	29.57	29.72
5	Face 5	88.2	88.7	84.3	87.9	82.5	34.33	32.70	34.37
6	Face 6	87.9	86.8	82.5	86.5	85.3	34.33	34.95	34.37

head movement varies. By this method, multiple faces got tracked which are not sensitive to the face size and light. The Discrete approximation to LoG function with Gaussian = 1.4 (Fig. 7 and Table 1).

A. **Experiments at Various Levels of Compression:** Many experiments done for the performance evaluation of DCT algorithms. On DCT compression the quantization effect can be viewed when the halving or doubling of images taken place. Low PSNR values were observed in the case of the faces seen in the video dataset especially for some particular samples. This low PSNR may contribute to the classification error. For the testing, from 10 images of face we gathered the face features which varies in size, shape, and orientation. The mean value of confidence is calculated for validating the identity asserts. Validating includes accepting or rejecting the identity claim. Here, if the mean value of confidence is greater than 0.9 thresholds, the identity assert is validated to be accepted. Otherwise, the system rejects the inaccurate asserts. This value can also be used for evaluating the similarity between the subjects. System performance remains constant to the face size. There were no effects with respect to the lighting. The techniques used here are less expensive. The testing of nearly 10 face images requires more than 30 ms on Pentium machine which is comparatively less for the particular model.

8 Conclusion

The work mentioned represents an endeavor for acknowledging toward the proposed system. This system has better performance on comparing with other similar systems in robustness, real-time tracking, etc. The goal is to improve the speed of the system in case of face detection and recognition of blob especially in color and movement. For responding to an action which happened in real-time real-time analysis of video is used which helps in obtaining data at high resolution. Such systems can be together integrated with the other sensing devices and the data about the space in which the system operates. This provides the alertness about the situations highly. Future work can be implementing on skin segmentation (clustering, skin model), object tracking (estimating objects position), and human body detection.

References

1. P. Viola, M. Jones, "Robust real-time face detection", IJCV,57(2):pp. 137–154, 2004.
2. V.K. Narendra Kumar, B. Srinivasan, "Security Based on Real Time Tracking of Multiple Human Faces Identification", I.J. Intelligent Systems and Applications, pp. 85–92, 2013.
3. M. Pateraki, H. Kondaxakis, M. Pateraki, H. Baltzakis, P. Kondaxakis, P. Trahanias "Tracking of facial features to support human-robot interaction", In Proc. IEEE International Conference on Robotics and Automation (ICRA'09), pp. 3755–3760, May 2009.
4. Yang, M.H., Kriegman, D., Ahuja, "Detecting faces in images: a survey", IEEE Trans. Pattern Anal Mach. Intell. 24(1), 34–58 (2002).
5. A. K. Jain, R. Bolle, "Biometric personal identification in networked society", Norwell, MA: Kluwer, 1999.
6. Y.L. Tian and A. Hampapur, "Robust salient motion detection with complex background for real-time video surveillance", In Proc. of IEEE Computer Society Workshop on Motion and video Computing, January 2005.
7. Baoxin Li and Rama Chellappa, "A generic approach to simultaneous tracking and verification in video", IEEE Trans. on Image Processing, no. 5, vol. 11, pp. 530–544, May 2002.
8. Antonioni, G., V. Popovici, J. Thiran, "Independent Component Analysis and Support Vector Machine for Face Feature Extraction", Audio and Video Based Biometric Person Authentication, pp. 111–118, 2003.
9. Shih, F.Y., C.F. Chuang, "Automatic extraction of head and face boundaries and facial features", Information Sciences, vol. 158, no. 1, 2004.
10. Jain, A.K. Hsu, R.L., Abdel-Mottaleb, "Face detection in color images", IEEE Trans. Pattern Anal. Mach. Intell. 24(5), pp. 696–706, 2002.
11. Ming-Hsuan Yang, Narendra Ahuja, "A survey on face detection methods", IEEE Trans. On pattern analysis and machine Intelligence (PAMI), vol. 24, no. 1, pp. 34–58, 2002.
12. http://www.cs.cmu.edu/~vsam/.
13. C. Hesher, A. Srivastava, G. Erlebacher, "A novel technique for face recognition using range images", in the proceedings of Seventh International Symposium on Signal Processing and its Application, 2003.
14. http://www.ics.forth.gr/pateraki/handfacetracking.html.

15. T. Singh, “Novel restoration process for degraded image”, in Proceedings – 2014 5th International Conference on Signal and Image Processing, ICSIP 2014, Bangalore, Karnataka, 2014.
16. Z. Zeng, M. Pantic, G. Roisman, T. Huang, “A survey of affect recognition methods: Audio visual and spontaneous face recognitions”, IEEE Trans. Pattern Anal. Mach. Intell., no. 1, vol. 31, pp. 39–58, Jan 2009.
17. X. Zhang, Z. Zhang, L. Yin, D. Hipp, P. Gerhardstein, “Perception driven 3D facial expression analysis based on reverse correlation and normal component”, Affective Computing and Intelligent Interaction (ACII) 2015 International Conference on, pp. 616–622, 2015, ISSN 2156-8111.
18. I. Siegert, R. Bock, A. Wendemuth, B. Vlasenko, K. Ohnemus, “Overlapping speech utterance duration and affective content in HHI and HCI – An comparison”, Cognitive Infocommunications (CogInfoCom) 2015 6th IEEE International Conference on, pp. 83–88, 2015.

Simulation and Performance Evaluation of WSN-Based Architecture for Animal Health Monitoring Using NS2

Ankit Bhavsar and Harshal Arolkar

Abstract The network simulator tools provide a facility to check the characteristics of network architecture. The NS-2 network simulators is one of the widely used tools by researchers to investigate the characteristics of wireless sensor network-based network architecture. Agriculture and animal husbandry are the main source of income in rural areas of India. Pet animals like cow, buffalo, sheep, goat, and others play a significant role in the life of rural people. Pet animals are used and nurtured as they are considered to be source of income. Weak health and improper treatment of animal directly affects the trade of animal husbandry and agriculture which is backbone of village life. Due to problems faced by animal owner special attention needs to be given on animal health monitoring. This paper presents a preliminary performance investigation of the WSN-based architecture model for animal health monitoring. We analyze the performance of simulation model by using metrics, such as packets delivery statistics, network throughput, and end-to-end delay.

Keywords Wireless Sensor Network (WSN) · Transmission Control Protocol (TCP) · Simulator · NS-2 · Sensor · C++ · OTCL · PAN

1 Introduction

The sensor node consists of sensing unit, computation unit, and communication unit. All these components provide capability to monitor and react to a particular environment. The interconnection of many such sensor nodes form a network called

A. Bhavsar (✉)
GLS (I & R.K. Desai) Institute of Computer Application, GLS University, Ahmedabad, India
e-mail: ankit@glsica.org

H. Arolkar
GLS (S.R. Parikh) Institute of Computer Technology, GLS University, Ahmedabad, India
e-mail: harshal@glsict.org

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_2

wireless sensor network (WSN) [15]. The WSN is usually used for monitoring, sensing, and responding to activities happening at a remote location [4–6, 8, 16].

In [3] we have proposed architecture for animal health monitoring using WSN for rural area of Gujarat. In this paper, we have simulated the said architecture in NS2. The network simulation model consists of three personal area networks (PAN). Each PAN has three sensor nodes and one cluster tree-based PAN coordinator (CTPAN coordinator) for managing the personal area network. The sensor node sense and capture the event data and transmits it to CTPAN coordinator. The data, received by the CTPAN coordinator, will be forwarded to the main PAN Coordinator for further transmission.

The paper is divided in six sections. Section 2, discusses the related work. Section 3 shows the proposed architecture for animal health monitoring. Section 4 explains simulation scenario. Section 5 analyzes the simulation result followed by the conclusion in Sect. 6.

2 Related Work

This section shows some of the NS-2-based simulation models for WSN [1, 4, 12–14, 16, 22–24, 26, 27].

Ali Al-Dahoud et al. [2] presented simulation of monitoring system for a set of photovoltaic (PV) panels. They created simulation interface for PV panels. They proposed new solution based on the WSN sensor node for the identifying fault in solar panel located at Algeria.

Kavi K. Khedo et al. [11] proposed the use of wireless sensor network for air pollution monitoring in Mauritius. They proposed an innovative system named "Wireless Sensor Network Air Pollution Monitoring System (WAPMS)" to monitor air pollution in Mauritius. They deployed huge number of wireless sensors nodes around the island. The proposed system makes use of an Air Quality Index (AQI) which is presently not available in Mauritius. They also had designed and implemented a new data aggregation algorithm named "Recursive Converging Quartiles (RCQ)." The algorithm is used to merge data to eliminate duplicates and filter out invalid readings. By implementing this algorithm, they reduced the amount of data to be transmitted to the sink and thus saving energy.

Kamlendu Pandey et al. [10] presented complete simulation model for moisture deficit monitoring and controlling the irrigated agriculture using NS-2. They used the cluster-based simulation approach. They dealt with soil moisture data, its routing, aggregation, processing, and relay to the sink. The simulation model gives the real situation of soil moisture; which helps in irrigation scheduling.

Hafsah Nirwana et al. [8] provided optimal communication path modeling for early detection of landslide using WSN. They used network simulator 2(NS-2) for simulating two different scenarios. Based on the simulation result analysis, they compared throughput, delay average, packet loss for both scenarios.

Rajesh Banala et al. [19] implemented real-time surveillance of the home security with intelligent remote monitoring based on the ZigBee technology. The system can send abnormal images and messages through MMS and SMS. For that they created simulation model and checked best routing protocol for getting highest performance. They compared AODV to DSDV protocol through simulation model.

3 Proposed Architecture

The WSN-based Animal Health Monitoring will be implemented in rural area of Gujarat. The proposed architecture shows the village scenario having three animal sheds. We propose that each animal shade works as Wireless Personal Area Network (WPAN). Each WPAN has one Cluster Tree-based PAN coordinator (CTPAN coordinator) node and multiple sensor nodes. Each sensor node senses the event and transmits the event data to concerned CTPAN coordinator. The CTPAN coordinator is responsible for starting network, configuring all sensor nodes and transmitting/receiving data to/from nodes. Further, it will transmit data to PAN coordinator which sends transmission to Gram panchayat server. Figure 1 shows the topology of the architecture proposed.

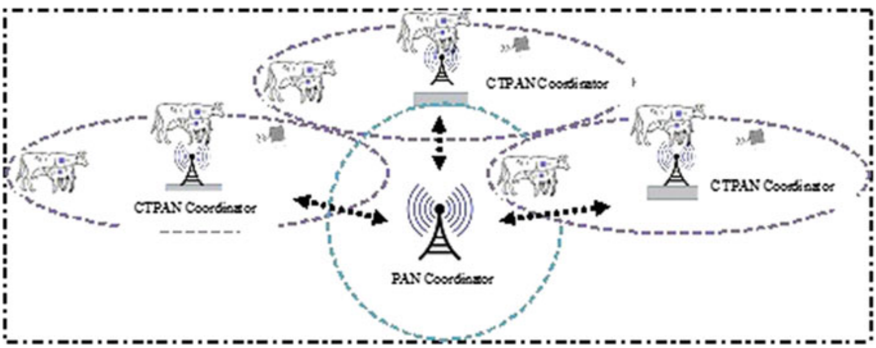


Fig. 1 Proposed architecture of three PAN-based animal shades

4 Simulation Scenario

Many open sources network simulation tools are available today. One of the best tool for testing the network system prototype is Network Simulator (Version 2), popularly known as NS2. The NS2 is event driven simulation tool which is useful in studying the dynamic nature of communication networks. It provides support to the simulation of the wired, wireless, and wireless sensor network's functions, protocols and standards. The NS2 provides the support to the simulate TCP, UDP, and other routing protocols. NS2 has several advantages like complex scenario can be easily tested, result can be quickly obtained, and large number of protocol's support [5–7, 9, 15, 17, 18, 20, 21].

The WSN-based animal health monitoring architecture consists of application layer, transport layer, and low-level network management. We propose to implement IEEE 802.15.4 at MAC layer and Transmission Control Protocol (TCP) as connection agent at the transport layer. The purpose behind selecting TCP is that it is a connection-oriented reliable transport protocol. The TCP achieves its reliability by three different phases called connection setup, data transmission, and connection termination [25]. Figure 2 shows the initial simulation layout of the three PAN-based animal sheds in NS2. Here, the topology shown in Fig. 1 has been simulated.

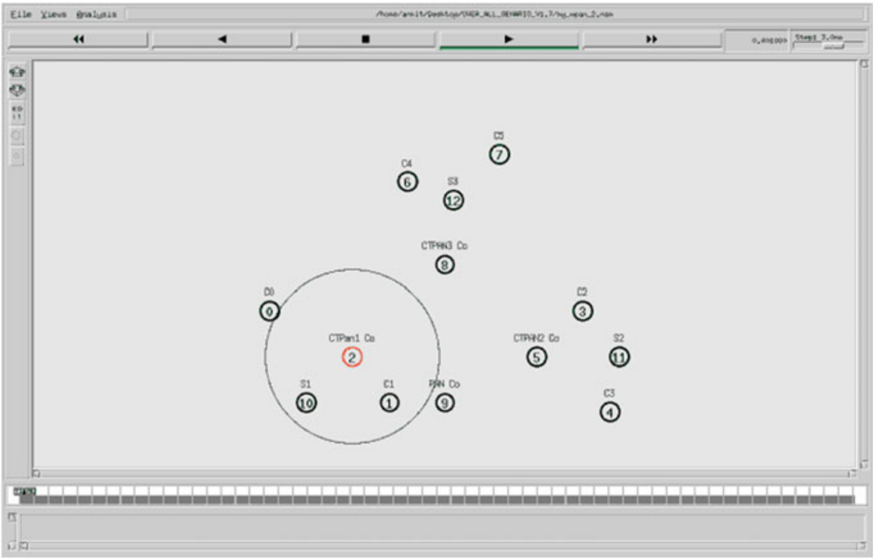


Fig. 2 Simulation layout of three PAN based on animal shades

Table 1 Simulation parameters and its values

Sr. No.	Simulation parameters	Values
1.	NS Version	NS 2.35
2.	Channel	Wireless Channel
3.	Number of Nodes	13
4.	Type of the MAC	802.15.4
5.	Antenna Model	Omni Directional
6.	Type of Routing Protocol	AODV
7.	Internet Protocol Type	TCP
8.	Package Size	50
9.	Area of Simulation	400 × 400
10.	Propagation	Two Ray Ground
11.	Interface Queue	Drop Tail
12.	Simulation Time	70 ms
13.	Traffic Type	cbr
14.	Max Packet Size	150

WPAN has been simulated to establish TCP connection that carries Constant Bit Rate (CBR) traffic type. These CBR has been selected because it is more suitable for sending periodic data sets. The simulation setup we require to set values of several simulation parameters. Table 1 shows the list of simulation parameters used along with it's values.

4.1 Simulation Process

The generic steps followed in the simulation model for the data packet transmission using wireless network are mentioned below.

1. Initializing PAN coordinator
2. Initialize PAN by sending Beacons, PAN Id and channel frequency to neighboring devices.
3. Neighboring device sends acknowledgements for Beacons to PAN coordinator thus creating a PAN.
4. Periodically PAN coordinator sends beacon to sensor nodes for allocating channel for packets transmission.
5. Channel is allocated to the neighboring sensor node that wants to transmit data packets.
6. PAN coordinator receives data packets.
7. PAN coordinator forwards the data packet to the nearest router.
8. Router may then forward the data packet to local server.

This section shows some of the screens generated within the simulation. Figure 3 show that all three personal area networks configured their sensor nodes with CTPAN coordinators.

Fig. 3 Three PAN configured under their CTPAN coordinator

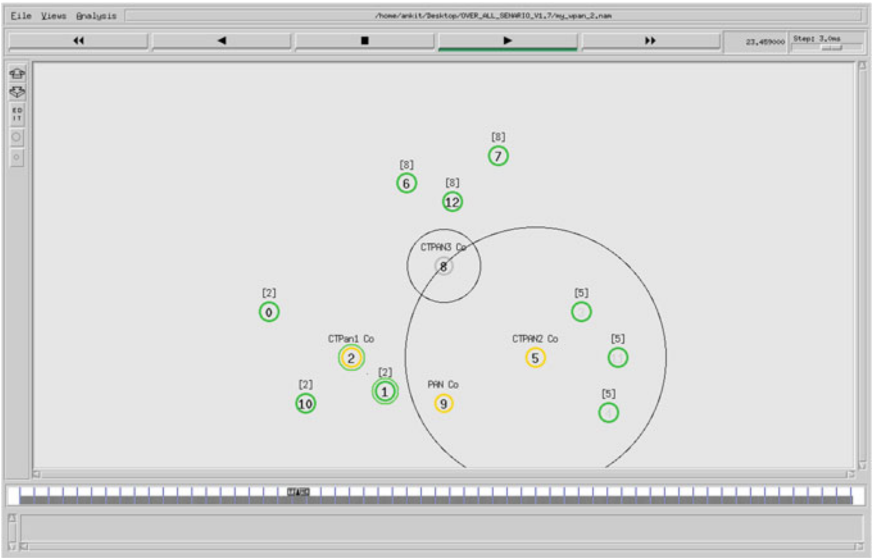


Fig. 4 Packets transmissions from node 1 to CTPAN coordinator node 2 with movement

Figure 4 shows packets transmission from node 1 to CTPAN coordinator node 2. During transmission, node 1 starts moving within the personal area network. Figure 5 shows packets transmission from node 11 to CTPAN coordinator node 5. It also shows the new position of node 1.

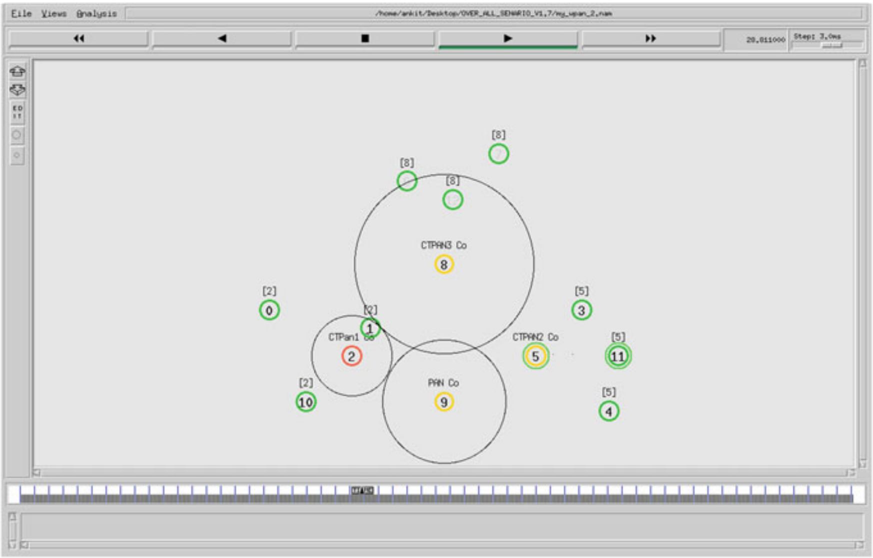


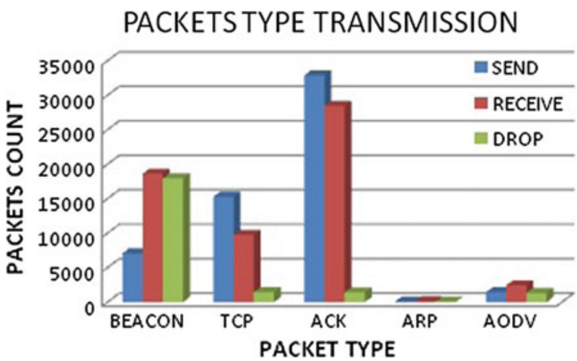
Fig. 5 Packets transmissions from node 11 to CTPAN coordinator node 5 and new Location of node 1

5 Simulation Result

The simulation was analyzed to check throughput of the network, number of data packets sent, received, and dropped by each of participating nodes in the simulation. We have also analyzed the total number of packets sent, received, and dropped within the entire network. The final analysis was done to identify the end to end delay of data packets.

Figure 6 shows the graphical output of total number of packets sent, received, and dropped based on the type of packet.

Fig. 6 Shows the total number of packet sent, received and dropped by packet type during simulation



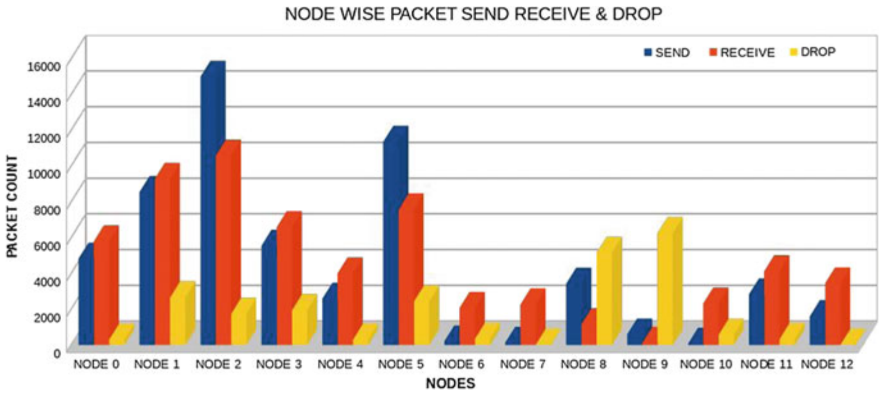


Fig. 7 shows the total number of packets sent, received and dropped by each node

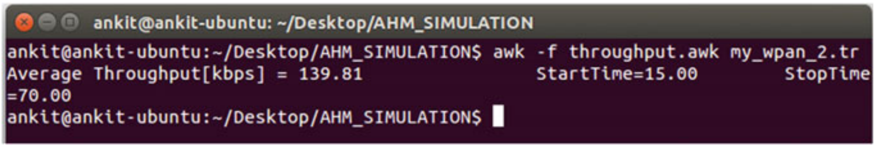


Fig. 8 Network through put measured in Kbps

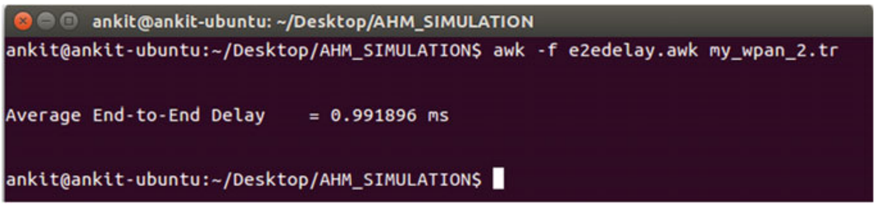


Fig. 9 Average end-to-end delay during simulation

As can be observed from the figure the, 7089 Beacons are sent while 18,636 are received and 17,953 are dropped. Similarly 15,289 TCP packets are sent while 9745 are received and 1422 are dropped. Figure 7 shows the total number of packets sent, received, and dropped by each node.

Finally Fig. 8 shows the throughput of the entire networks while Fig. 9 shows the average end-to-end delay during simulation.

6 Conclusion

The simulator result show that WSN-based architecture for animal health monitoring. It is being concluded that IEEE 802.15.4 standard is suitable for the proposed architecture. The simulation analysis shows that the packet delivery ratio is high enough and end-to-end delay is possible less for the proposed architecture. The simulation gives encouraging result for WSN-based architecture for animal health monitoring.

References

1. Adnan M. Abu-Mahfouz, Gerhard P. Hancke, sherrin J. Isaac.: Positioning system in Wireless Sensor Network Using NS-2. In: Software Engineering, pp. 91–100, ISSN: 2162-934X e-ISSN: 2162-8408 (2012)
2. Ali Al-Dahoud, Mahamed Fezari, Fatma Zohra Belhouchet.: Remote Monitoring System Using WSN for Solar Power Panels. In: First International conference on System Informatics, Modeling and Simulation, IEEE, pp. 120–125, ISBN 978-0-7695-5198-2 (2014)
3. Ankit Bhavsar, Dr. Harshal A. Arolkar.: Wireless Sensor Network: A Possible Solution for Animal Health Issues In Rural Area Of Gujarat, International Journal of Enterprise Computing and Business System (IJECBS), Vol 2, Issue 2, 2012, ISSN - 2230–8849
4. B. Manimekala, M. kayalvizhi.: A Data Transfer in Wireless Sensor Network Using AODV Protocol. In: International Journal of Computer Science, Vol. 9, Issue 1, No.1, pp. 191–200, ISSN - 1694–0814 (2012)
5. Carlos F. Garcia – Hernandez, Pablo H. Ibarguengoytia - Gonzalez, Joaquin Garcia - Hernandez, Jesus A. Perez – Diaz.: Wireless Sensor Network and Application: A Survey. In: IJCSNS, Vol 7, No. 3, pp. 261–273, ISSN: 1738-7906 (2007)
6. E. Egealopez, J. Vales-Alonso, A. S. Martinez - Sala, P. Pavon-Marino, J. Garcia-Haro.: Simulation Tools for Wireless Sensor Networks. In: SPECTS, pp. 2–9 (2005)
7. Fei Yu.: A Survey of Wireless Sensor Network Simulators Tools. In: Sensors. http://www.cse.wustl.edu/%7Ejain/cse567-11/ftp/sensor/index.html
8. Hafsah Nirwana, Zahir Zainuddin, Zaryanti Z., M. Bakri M, Muh. Tola, Nharun, Salman M, Andani.: Communication System Modeling Of Wireless Sensor Network For Early Landslide Detection. In: International Journal of Engineering Research and Application, Vol. 3, Issue 4, pp. 2412–2417, ISSN - 2248–9622 (2013)
9. Harsh Sundani, Haoyue Li, Vijay K. Devabhaktuni, Mansoor Alam, Parbir Bhattacharya.: Wireless Sensor Network Simulators A Survey and Comparisons. In: International Journal of Computer Networks (IJCN), Vol 2, Issue 5, pp. 249–265, ISSN - 1985–4129 (2012)
10. Kamlendu Pandey, S.V. Patel.: Design and Development of A Simulation Frame Work For WSN Based Moisture Deficit Monitoring And Control In Irrigated Agriculture. In: International Journal of Computer Science and Network (IJCSN), Vol. 3, No. 1, pp. 21–28, ISSN - 2234–8018
11. Kavi K. Khedo, Rajiv Perseedoss, Avinash Mungur.: A WirelessSensor Network Air Pollution Monitoring System. In: International Journal of Wireless & Mobile Network, Vol. 2. No. 2, ISSN online: 1741–1092 (2010)
12. Majid Alshamnari.: Simulation and Investigation of S-MAC Using the Network Simulator NS2. In: International Journal of Business and Social science, Vol 5, No 8(1), pp. 35–37, ISSN 2219–1933 (Print), 2219–6021 (Online) (2014)

13. Nandhini B, Srie Vidhya Janani E.: Energy Efficient Data Collection in Wireless Sensor Network. In: ICTACT Journal on Communication Technology, Vol. 4, Issue 3, pp. 796–801, ISSN - 2229–6948 (2013)
14. Nayana Hegde, Dr. Sunil Kumar S. Manvi.: Simulation of Wireless Sensor Network Security Model Using NS – 2. In: International Journal of Latest Trends in Engineering and Technology, Vol. 4 Issue 1, pp. 113–119, ISSN - 2278–621X (2014)
15. Neha Singh, Prof. Rajeshwarlal Dau, Vinita Mathur.: Network Simulator NS2 - 2.35. In: International Journal of Advanced Research in Computing Science and Software Engineering, Vol. 2, Issue 5, pp. 224–228, ISSN - 2277 128X (2012)
16. Parth M. Dave, Purvang D Dalal.: Simulation & Performance Evolution of Routing Protocols in Wireless Sensor Network. In: International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, pp. 1405–1412, ISSN - 2278–1021(2013)
17. Paulo A.C.S Neves, Joel F.P. Fonseca, Joel J.P.C. Rodrigues.: Simulation Tools for Wireless Sensor Networks in Medicine: A Comparative Study. In: Proceedings of the First International Conference on Biomedical Electronics and Devices, BIODEVICES (2008)
18. R.G. Ingalls.: Introuduction to simulation. In: WSC-02 Proceeding of the 34th conference on winter simulation, ISBN:0-7803-7615-3 (2002)
19. Rajesh Banala, D. Upender.: Remote Home Security System Based on Wireless Sensor Network Using NS2. In: International Journal of Computer Science and Electronic Engineering, Vol. 2, Issue 2, ISSN - 0975–5664 (2012)
20. Sachi Pandey, Vibhore Tyagi.: Performance Analysis of Wired and Wireless Network Using NS2 Simulator. In: International Journal of Computer Applications, Vol. 72, No.21, pp. 1–8, ISSN - 0975–8887 (2013)
21. Sirwan A. Mohammed, Dr. Sattar B. Sadkhan.: Design of Wireless Network Based On NS2. In: Journal of Global Research in Computer Science, Vol. 3, No, 12, pp. 7–16, ISSN - 2229-371X(2012)
22. Snehal Sarangi, Biju Thankchan.: A Novel Routing Algorithm for Wireless Sensor Network Using Particle Swarm Optimization. In: IOSR Journal of Computer Engineering, Vol. 4 Issue 1, pp. 26–30, ISSN - 2278–0661 (2012)
23. T. Gomathi, V. Manju, N. Anuradha.: AN efficient Blind Signature Authentication for Wireless Sensor Networks Using HECC. In: International Journal of Innovation and Scientific Research, Vol. 10, No. 1, pp. 6–10 (2014)
24. Taipei, Taiwan.: Performance Evaluation of IEEE 802.15.4 Ad Hoc Wireless Sensor Network: Simulation Approach. In: IEEE Conference on Systems, Man and Cybernetics, pp. 1443–1448 (2011)
25. Teerawat Issariyakul, Ekram Hossain.: Introduction to Network Simulation NS2. In: Springer, ISBN -978 - 0 - 387 - 71759 - 3 (2009)
26. Umesh B.N, Dr. G Vasanth, Dr. Siddaraju.: Energy Efficient Routing of Wireless Sensor Network Using Virtual Backbone and Life Time Maximization of Node. In: International Journal of wireless & Mobile Networks, Vol. 5, No. 1, pp. 108–118, ISSN (Online) 0975–3834 (2013)
27. Zhiwei Zho, Xinming Zhang, Peng Sun, Pengxi Liu.: A Transmission Power Control MAC Protocol for Wireless Sensor Networks. In: Proceeding of the Sixth International Conference on Networking, IEEE (2007)

Random Route Adoption Algorithm to Provide Source Location Privacy in Wireless Sensor Network

N. Abhilash and S. Kuzhalvaimozhi

Abstract Since wireless sensors are being deployed in hazardous conditions, emphasis on source location privacy has become elementary. This paper proposes Random Route Adoption (RAA) algorithm which is built on the idea of random walk combined with Genetic algorithm. This algorithm intends to keep all the goodness of random walk and have a greater effect in providing source location privacy with the help of genetic algorithm. As a result, there will be an increase in the safety period of the routing protocol and hence counters the effects of hop by hop traces and other adversarial traffic analysis practices.

Keywords Forwarding probability · Genetic algorithm

1 Introduction

Wireless sensors are the revolutionary devices which can perform various tasks and tune-ups persistently and prissily with minimum maintenance. These sensors are strategically deployed in a geographical area to form a network of devices which can do various tracking and monitoring [1, 2]. Tacking an object is easy but safekeeping this information from the advisories is quite challenging. Applications such as doctors monitoring soldier's health in a battle field, monitoring endangered animals in its habitat, sensitive event data has to be transmitted along the network. Advisories can backtrack to its origin and cause serious trouble to the object, hence preserving privacy of the source node is very vital.

N. Abhilash (✉)
Computer Science and Engineering, National Institute of Engineering,
Mysore, India
e-mail: nabhilash619@gmail.com

S. Kuzhalvaimozhi
Information Science and Engineering, National Institute of Engineering,
Mysore, India
e-mail: kuzhali_mozhi@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_3

There are many methods [3] to provide source location privacy (SLP) one of which is the random walk. It is a method where sensor nodes takes random path whenever it has to transmit event data to the sink. This is one of the early method which aims to give enough time for the monitored object from the advisory to dislocate from the source node. This period of time is known as safety period, which is one of the crucial security metrics. In prior to data transmission, every node has a probability of receiving the data from its neighbors. The probability of forwarding the data to its specific neighbor is known as forwarding probability [4]. In random walk, the forwarding probability of all nodes are same., i.e., a node may forward a packet to a node which already received earlier. Hence, pure random walk is found not effective in providing SLP. Many enhancements are made to this random walk to provide a better privacy and paper is also an effort in doing so, by integrating random walk with the genetic algorithm [5].

Genetic algorithm is used to look for good and robust solution which is specified in fitness criteria. The reason for choosing genetic algorithm is that, it works on chromosome which are the encoded version of parameters which gives a potential solution, it performs parallel search from the population and it uses simple fitness function which has no artificial over engineered mathematics.

2 Related Work

All the related work mentioned below are based on the approach of random walk and its enhancements. One of the basic approach is the Pure Random Walk [6]. Here, the packets are forwarded using random paths with no major constraints. The downside of this method is that all the nodes in the network will be having same forwarding probability. Hence, there is a chance of forming loops in the network path and might ease up the job for an advisory who tries to back trace the packets origin.

An enhancement to the above approach is Phantom Routing scheme [7] which is the combination of random walk and flooding. This scheme starts with a random walk up to a specified number of hops and as it reaches the hop limit, the packets are flooded till it reaches the sink node. This scheme is still considered as energy consuming and increases the network traffic.

Phantom Single Path Routing [8, 9] is one of the method to provide SLP, where the node surrounded by the neighbors are divided into sectors and all nodes find out the distance between itself to the sink and its neighbors to the sink so that a forwarding node knows that the neighbor is either far or close to the sink. A node randomly selects any one of the sector and walks till it reaches the hop limit. After this, it starts using a single path to reach the sink node. This method is said to be effective but limiting the random walk period will reduce the safety period for the source.

Directed Random Walk [10] is somewhat similar to the above method but it only selects the node with less hop count to forward the packets. The nodes with less hop count to the sink are considered as parents and every time a different parent is selected to reach the sink. [11] has also contributed in providing SLP where both sink and source randomly walks to meet a receptor node. This receptor node uses this preexisted path by the sink to forward packets to the sink.

Randomly Selected Intermediate node [12] also selects an intermediate node which must be at least some distance from both sink and source. The downside of this method is that it consumes lot of energy and in selection of intermediate node which can become either too close or too far from the source [13]. Overcomes this problem where the intermediate node will not be too far or too close to the source and gives a balanced power consumption. Here, an intermediate node is selected from the area R−r around the sink as shown in the Fig. 1a. The area R−r is known as the sink's toroidal region. Any advisory will be having a tough time in monitoring the entire toroidal region there by providing location privacy.

Similar to the above method [14] focus on the balance between energy consumption and security for the destination node. This method is divided into two phases, where in the first phase is a walk to a randomly selected intermediate node with in the toroidal region R−r surrounding the source as shown in the Fig. 1b. In

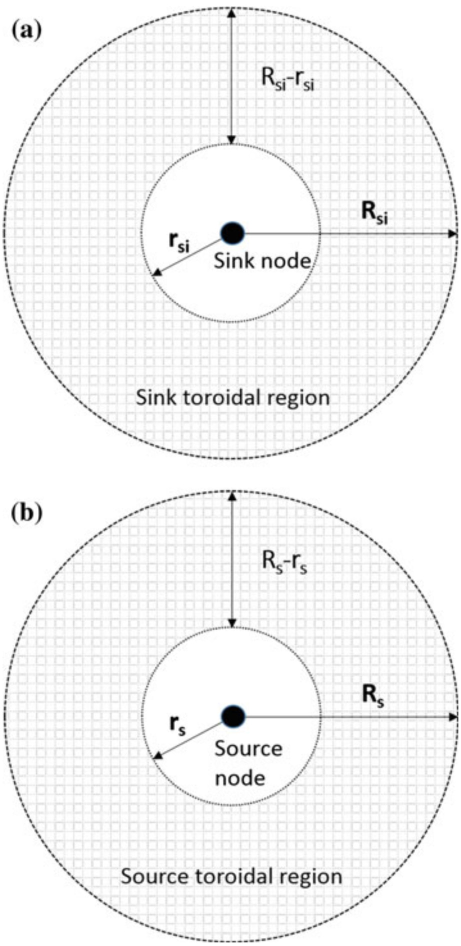


Fig. 1 **a** Sink toroidal region. **b** Source toroidal region

second phase, the packets are forwarded from that intermediate node to the sink using a shortest path along with fake packets to confuse the advisories.

There are still other different methods to provide SLP like using dummy packets, creating delay during transmission, location anonymization, etc. but our focus is to enhance the effectiveness of these random walks and to avoid cycles.

3 Random Route Adoption Algorithm

This paper proposes a Random Route Adoption algorithm which is implemented based on the random walk approach combined with genetic algorithm. Here the base station as a primary actor initiates the system by broadcasting the beacon messages to the sensors in communication range and each receiving sensors updates their routing table and forwards this message down the line of sensors so that all possible paths and their hop counts to reach the base station can be known.

Whenever an event occurs, the sensors senses the change in environment and forwards this event data to the base station with a random path of hop count i and consider this path as generation zero. As said RRA algorithm aims to increase the effect of random walk to provide SLP. This is done by altering the forwarding probability of the neighboring nodes and to do so, we are going to invoke genetic algorithm.

Genetic algorithm generates various solution for many optimization and search problems by impersonating the process of natural selection. It involves the steps such as generating initial population, evaluation, or fitness calculation and generating new population. The flow chart of RRA algorithm is shown in Fig. 3.

A. *Generating initial population: genes and chromosome*

In genetic algorithm, a gene is a locus or a block which constitutes a chromosome. Here a single sensor node is considered as a gene. Therefore a chromosome is a sequence of Genes (nodes) which is a possible solution for our problem. The starting gene of the chromosome is always the source node where an event is recorded. This gene is followed by the neighboring nodes and has the base station as the last node. The maximum length of chromosome cannot be greater than the total number of nodes in the network.

The initial population is the pool of chromosome which corresponds to all the possible path from the source node to the base station.

The gene elements will be having a binary value 0 or 1. 0 indicates that the gene has not been contributed in generation zero and vice versa which is shown in the Fig. 2 (Fig. 3).

Fig. 2 Structure of chromosome and its genes

12	15	27	29	33	58	70	77	82	99
1	0	1	0	0	1	0	0	0	1

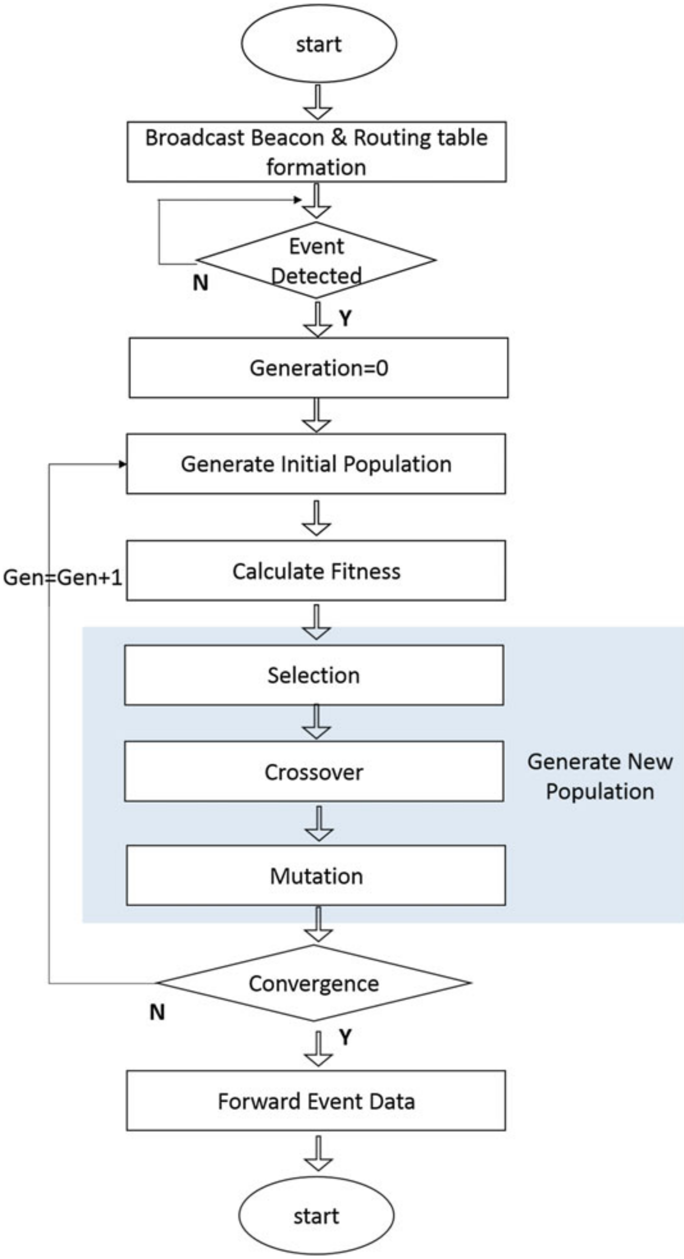


Fig. 3 Flowchart of random route adoption algorithm

B. *Fitness function*

Here, the fitness of each chromosome present in the initial population is calculated. The goal is to take a random path from source to base station without any cycles and to use most random paths to improve safety period for the source.

The input to the fitness function are the genes of the chromosome and the output is an integer value for each chromosome as shown in Eq. (1).

$$f_n = \sum_{i=1}^{i\max} \left(\frac{N_i}{TN}\right)^{-1} \times i \tag{1}$$

where

Ni is the number of nodes with gene value 1 with i hops
TN is the total nodes in the sensor network
f_n is the fitness value for each chromosome

C. *Generating new population*

New population of chromosomes can be created by performing selection, crossover and mutation.

Selection
Herbert spencer said and I quote "survival of the fittest" in the evolutionary theory of natural selection. It means that the one who is fit can survive and will be selected by the nature. Likewise, the fitness value of all the chromosomes are sorted and one

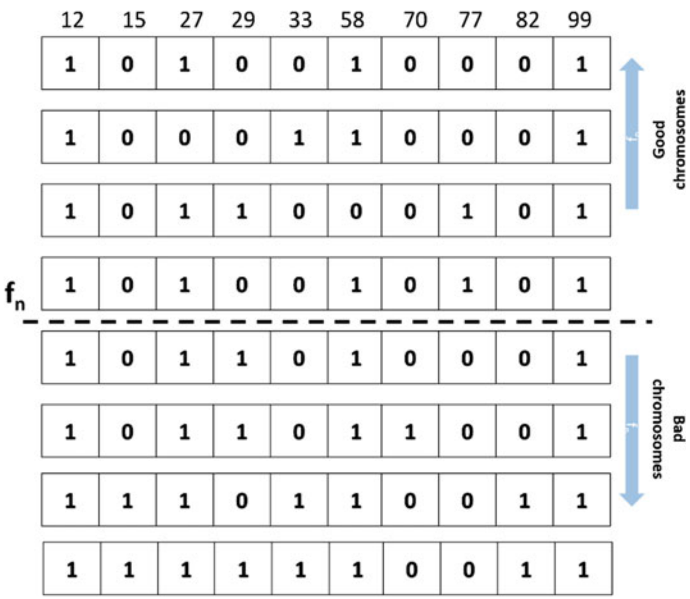


Fig. 4 Selection process

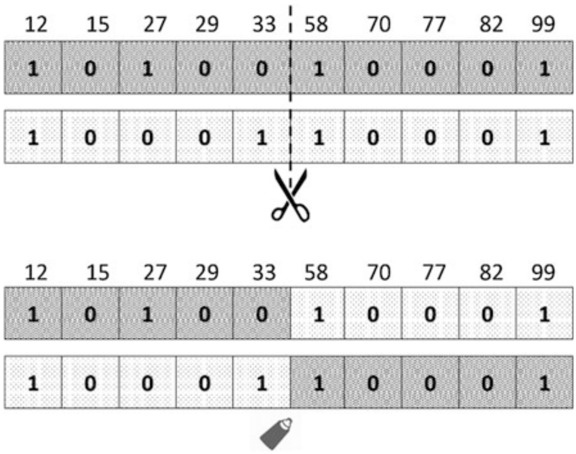


Fig. 5 Crossover of two chromosome

with the good fitness values are selected and produced to mating pool and the remaining are discarded as shown in Fig. 4.

Crossover
Here, good chromosomes are combined with each other to yield even better chromosome. As shown in Fig. 5, two chromosomes are chosen from the mating pool and a single crossover point is selected between each. The either side of crossover point of two chromosome are exchanged with each other and concatenated to produce two new offspring.

Mutation
It is done to create a new attribute which may not be found in the parent chromosome. It is done just by tossing a gene randomly as shown in the Fig. 6.

By the end of the iteration, the chromosome with the best fitness value is taken as the solution and if the solution is not satisfactory then the entire process is repeated. So every time RRA provides a new random path with varying forwarding probability which will buy enough time to the subject to move away from the source node.

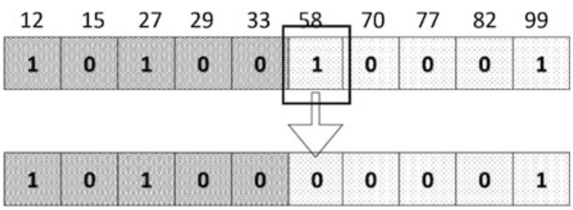


Fig. 6 Mutation process

4 Conclusion

Since sensor nodes are deployed in critical conditions, having various monitoring and tracking applications, context privacy becomes as fundamental as content privacy. Random walk with its uniform forwarding probability of neighbor nodes in considered less effective in providing SLP. Hence combining with genetic algorithm is going to alter the forwarding probability of neighbors and avoids forming cycles. Since the event data in the source randomly walks all the way to the base station, there will be a definite increase in the safety period where, the monitored object will have enough time to move from the source node before any advisory traces back.

References

1. V. P. Sadaphal; B. N. Jain. "Tracking target using sensor networks: target detection and route activation under energy constraints" *Communication Systems Software and Middleware and Workshops*, 2008. COMSWARE 2008. 3rd International Conference on 2008.
2. A. Joshi; I. Naga Vishnu Kanth; N. Samdaria; S. Bagla; P. Ranjan." GPS-less animal tracking system" *Wireless Communication and Sensor Networks*, 2008. WCSN 2008. Fourth International Conference on 2008.
3. M. Conti; J. Willemsen; B. Crispo. "Providing Source Location Privacy in Wireless Sensor Networks: A Survey". *IEEE Communications Surveys & Tutorials* on the year 2013.
4. A. Ali; R. A. Rashid; S. H. F. Arriffian; N. Fisal. "Optimal forwarding probability for real-time routing in wireless sensor network". *Telecommunications and Malaysia International Conference on Communications,* 2007. ICT-MICC 2007. IEEE International Conference on 2007.
5. L. Guo; Q. Tang. "An Improved Routing Protocol in WSN with Hybrid Genetic Algorithm*" Networks Security Wireless Communications and Trusted Computing (NSWCTC),* 2010 Second International Conference on 2010.
6. C. Ozturk, Y. Zhang, W. Trappe, and M. Ott, "Source-location privacy for networks of energy-constrained sensors," *in Software Technologies for Future Embedded and Ubiquitous Systems,* 2004. *Proceedings. Second IEEE Workshop on,* ser. WSTFEUS' 2004.
7. C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in Energy-constrained sensor network routing," *in Proc. 2nd ACM workshop on Security of Ad hoc and Sensor Networks,* ser. SASN'04, ACM. New York, NY, USA: ACM, 10 2004.
8. P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source location privacy in sensor network routing*," in Distributed Computing systems, 2005. Proceedings. 25th IEEE International Conference* on 2005.
9. P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source location Privacy in sensor network routing," *in Proc. 25th IEEE International Conference on Distributed Computing Systems,* ser. ICDCS 2005.
10. J. Yao and G. Wen, "Preserving source-location privacy in energy constrained wireless sensor networks," *in Distributed Computing Systems Workshops,* 2008. *28th International Conference* on, ser. ICDCS' 2008.
11. Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks*," in 20th International Parallel and Distributed Processing Symposium,* ser. IPDPS 2006.

12. Y. Li, L. Lightfoot, and J. Ren, "Routing-based source-location privacy protection in wireless sensor networks," *in Electro/Information Technology, 2009. IEEE International Conference* on, ser. EIT'09, IEEE. Piscataway, NJ, USA: IEEE, 6 2009.
13. L. Lightfoot, Y. Li, and J. Ren, "Preserving source-location privacy in wireless sensor network using star routing," *in 2010 IEEE Global Telecommunications Conference, ser. GLOBECOM 2010, IEEE. Piscataway, USA: IEEE communications society,* 12 2010.
14. L. Lightfoot; Jian Ren. "R-STaR destination-location privacy schemes in wireless sensor networks*". Communications (ICC), 2015 IEEE International Conference* on 2015.

Improving Energy Efficiency and Reliability of Wireless Sensor Networks with MC-AODV Routing Protocol

K. Imrana Banu, K. Rabiya Bashrin, K. Ummay Kulsum and Mohammed Ummar

Abstract The most important issues in wireless sensor networks are a powerful usage of resources, as the sensors in the system relay on these restricted resources, for example, battery lifetime, CPU, memory etc. In this way, the imperative issue in such sort of systems is energy effective and reliable. So to diminish the power consumption and to enhance the network performance, we upgrade a broadly utilized reactive routing protocol known as ad hoc on-demand distance routing protocol (AODV). For the most part to pick a best way AODV takes hop count metric, which is not suitable for WSNs in light of the fact that the nodes set in the same shortest way will drop their power at regular intervals. So the fundamental point of this paper is take a gander at hop count, as well as consider the node lifetime and remaining power while picking the best optimal path (Che-Aron in IJCSNS International Journal of Computer Science and Network Security 6(2), 2010) [1]. The upgraded AODV steering convention known as Energy routing protocol known as Energy constraint AODV (EC-AODV) routing protocol, which builds overall system lifetime by controlling the remaining power of the considerable number of nodes in the system. In this way, in view of the simulation results we demonstrate that the proposed system is more effective than the current one.

K. Imrana Banu (✉)
Computer Science and Engineering, Mother Theresa College of Engineering and Technology, Palamaner, Chittoor, Andhra Pradesh, India
e-mail: asirmu29@gmail.com

K. Rabiya Bashrin
Computer Science and Engineering, Sri Venkateswara University, Tirupathi, Andhra Pradesh, India
e-mail: krabiyabasrin@gmail.com

K. Ummay Kulsum
Department of CSE, RGUKT IIIT RK valley, Idupulapaya, Kadapa 516330, India
e-mail: asirmu@gmail.com

M. Ummar
Department of IT, Sri Venkateswara College of Engineering and Technology, Chittoor 517001, India
e-mail: mohammedummar1@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_4

Keywords WSNs · Energy factor · Node lifetime · AODV · EC-AODV

1 Introduction

A wireless sensor networks contain sensor nodes prepared for gathering information from the surroundings and communicating with each other by means of wireless transceivers. This assembled information is conveyed to one or more sink nodes by means for multi-hop communicate. For the most part, every sensor node comprises of batteries, different sort of memory, and remote handsets. In light of various application regions, for example, military, forest, agribusiness etc., different routing protocols has been proposed with different metric. By of working and application, routing protocols are arranged into proactive and reactive routing protocols. In dynamically varying areas, on-demand driven routing protocols extent a better performance, for example, Dynamic source routing and ad hoc on demand routing protocols [2].

In the event that a source node does not have a valid route in its routing table, yet at the same time willing to send a data packet, then it starts the route discovery procedure as appeared in Fig. 1.

Source node telecast the route request packet to all its neighbor nodes. On the off chance that the intermediate node which gets first route request (RREQ) packets, finds a valid path to the destination in its routing table then it setup a reverse path in

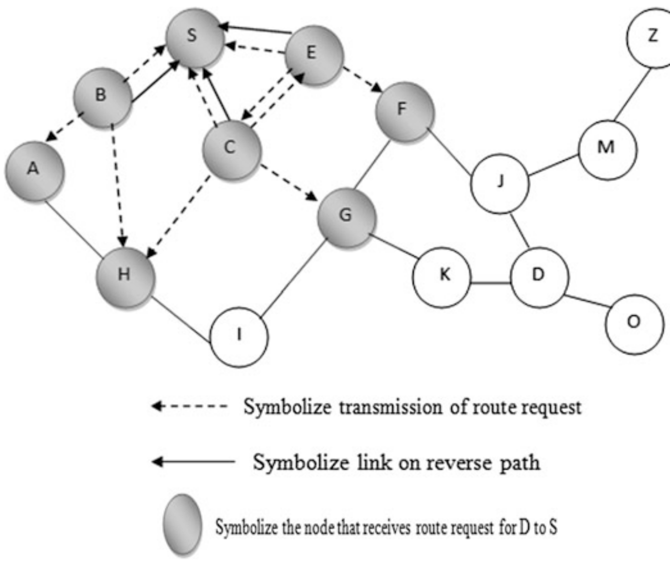


Fig. 1 Route discovery in AODV

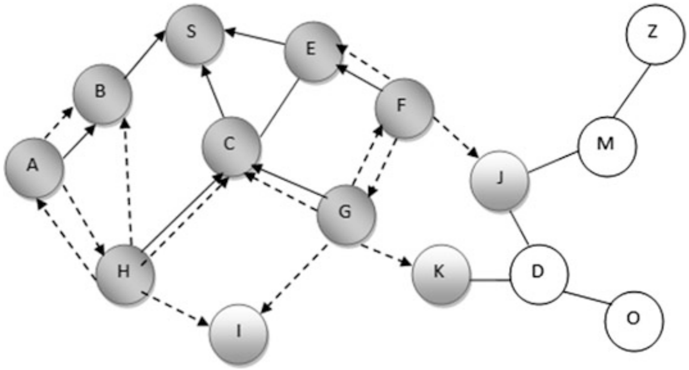


Fig. 2 Represent RREQ forwarding

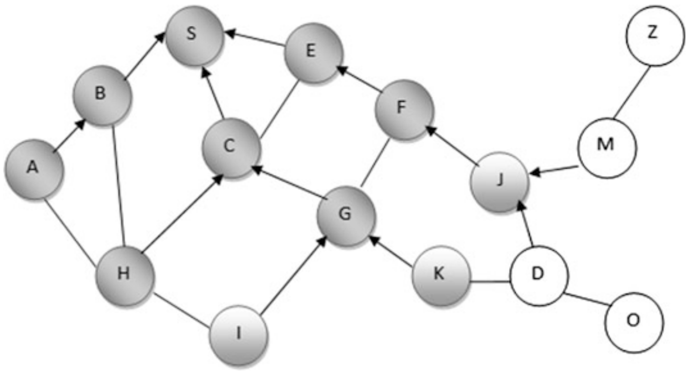


Fig. 3 Set-up a reverse path in AODV

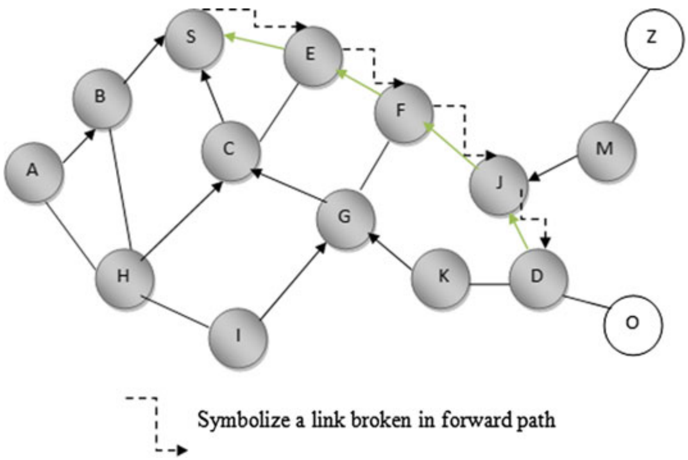


Fig. 4 Forward path set-up in AODV

the heading to source and sends the route replay (RREP) packets as appeared in Figs. 2 and 3 else the RREQ is broadcasted every one of its neighbors.

From G and H, node C gets RREQ yet it would not send once again, since node C has already send the route request once.

Node D is the aimed destination of RREQ so it does not send the route request.

At the point, when a RREP travels in the direction to the sources a forward way is setup to the destination as appeared in beneath Fig. 4. Generally, any type of conventional routing protocol for example, DSR, AODV picks a shortest path, to forward the information without considering the power of the nodes [3, 4]. So this paper upgrades the AODV routing protocol by adding a number of parameters that adjusts the power of the sensor node inside of the network. Accordingly, the general system lifetime is expanded by controlling the remaining power of the nodes

2 Energy Efficient Methodologies

Energy Efficient methodologies as the sensor nodes are distributed thickly in a field either near to or inside the process unique multi-hop wireless routing protocols between the sink nodes more over sensor nodes are required. Conventional ad hoc routing techniques do not normally meet the necessities of the Wireless sensor networks.

Energy-efficient routes can be discovered in view of the available power (PA) in the sensor nodes or the power required for sending in the connections along the routes. By one of the following methodologies energy-efficient route is chosen.

Maximum available power (PA) path: The path that has high aggregate PA is favored. By adding the PAs of each and every node through the path the aggregate PA is calculated. It is critical not to conceive paths determined by continuing paths that can associate the wireless sensor node to the sink right now path.

Minimum Energy (ME) path: The ME is the path that takes least power to transmit the information among the sink and the other nodes.

Minimum hop (MH) path: The path that makes the MH to reach the sink is favored. Minimum Energy (ME) chooses the same path presently as the Minimum Hop when the same measure of energy is utilized on each and every connection. In this manner, when hubs telecast with same energy level with no power control, MH is then identical to Minimum Energy.

Maximum minimum available power node path: The path through which the low PA is greater than the low PAs of alternate paths is favored. This blocks the risk of spending a wireless sensor node with minimum PA much sooner than the others in light of the fact that they are on path with hubs that have maximum available power.

3 Proposed EC-AODV Routing Protocol

In this proposed system, if the remaining power is not exactly the threshold value then that node does not include in route discovery and sending the information packets else the AODV protocol works at two stages.

1. Route discovery
2. Route maintenance

Route discovery
In this stage if the remaining power is not exactly the threshold then that nodes don't include in route discovery and data forwarding traffic. At the point when a source node need to send some information packets to the specified destination node, then it first finds in its routing table whether it have a valid route to the destination. It will forward the information packets on the specified path else by utilizing the route discovery process it will locate a valid path to the specified destination. Presently to locate the optimal path from source to destination the EC-AODV comes into act. Source node sends the route request (RREQ) packets that comprises of destination IP address alongside newly included field that is Energy factor (EF) and Node lifetime (NLT). In the wake of getting the RREQ packets, the intermediate node throws out the RREQ packets if its remaining energy is less than the value of threshold. Conversely, they likewise measure the EF and NTL values and upgrade their routing table. At that point, it checks if the remaining power is not exactly the threshold value along with the highest EF and NFL then it sends the RREP packet reverse path to source. At long last the best optimal path having low-hop count with highest nodes lifetime and energy factor is chosen [5].

Route maintenance
When a link is failed between source to destination then the route maintenance is used. A link can be failed due to many causes such as network congestion, mobility of nodes, energy consumption, and so on. The node sends the route error (RERR) packet to the source node, if its neighbor node does not respond to its HELLO message. The nodes update it routing table when it finds the RERR packet, at last when it reaches the source node, it makes a route discovery for a new route.

4 Evaluate Routing Parameters

The lifetime of the whole network is increased furthermore the power consumption is diminished over the system, by adding the power metrics to AODV routing protocol they are node lifetime and energy factor. The EF can be calculated by following procedure:

To begin with, figure the amount of energy taken by the nodes for transmitting and accepting the packets and after that calculate initial and current energy.

$$Energy\ factor = \frac{Remaining\ Energy(ER)}{Initial\ Energy(EI)}$$

where initial energy is the energy given to the nodes when they are created and remaining energy can be computed as

$$\mathrm{RE} = \mathrm{Current\,Energy\,(Ec)} - (\mathrm{Transmission\,Energy\,(Et)} + \mathrm{Receiving\,Energy\,(Er)})$$

where,

$$Et = \frac{Transmission\,Power * 8 * Packet\,size}{Bandwidth}$$

$$Er = \frac{Receiving\,Power * 8 * Packet\,size}{Bandwidth}$$

Next metric is Node Lifetime (NLT) that can be computed as

$$\mathrm{NLT} = \frac{\mathrm{Receiving\,Energy\,(Er)}}{\mathrm{Draining\,rate\,(DR)}}$$

where Draining Rate is the energy consumed at a specific node. Except the destination node all the remaining nodes in AODV needs to process their life time. Draining rate (Dr) is

$$\frac{\mathrm{Previous\,Energy\,(Ep)} - \mathrm{Current\,Energy\,(Ec)}}{\mathrm{Current\,Time\,(Tc)} - \mathrm{Previous\,Time\,(Pc)}}$$

EC-AODV can be done by doing a few changes to AODV's route request, route reply and routing table. To get to NLT and EF to extra fields are included into route request and route reply packets. At that point ECAODV comprises of source IP address, Destination address, Destination Sequence number, Source Sequence number, hop count, time to live, energy factor, node life time [6].

The EF and NLT are computed and added to route request packet during route discovery process then the converse path is made by the destination node or intermediate node in view of the hop count, EF, and NCL and after that upgrades its routing table.

5 Result

The throughput, packet deliver ratio, average end-to-end delay, average energy consumption, and routing overhead can be calculated by NS2. This shows the performance variation between AODV and EC-AODV.

Table 1 Comparision between AODV and EC-AODV

Metrics	AODV	EC-AODV
Throughput (bps)	688.69	994.74
Avg. end to end delay	559.86	75.332
Packet delivery ratio	0.9764	0.9998
Dropped packet ratio	1.97	0.26
Avg. energy consumption	309.94	300.78

UDP is the transport layer protocol attached at the source nodes. CBR is the constant bit rate attached at both the sender and the receiver side to create the traffic. The simulation time is fixed throughout the simulation and the data rate is varied from 0.1 to 0.6 Mbps for 802.11 Wireless LAN. The simulation results are explained below for both the existing and proposed systems.

Packet Delivery Fraction
This is the ratio between the total num of packets got at the receiver side over the total num of packets transmitted by the sender [7]. In AODV the PDF is 1.79% low, compared to EC-AODV it shown in Table 1.

Throughput
Throughput is the quantity of information send in a period of the time. As said above due to zero received packets for existing system there is no throughput. In EC-AODV, the throughput is 30.6% better compared to AODV and is shown in Table 1.

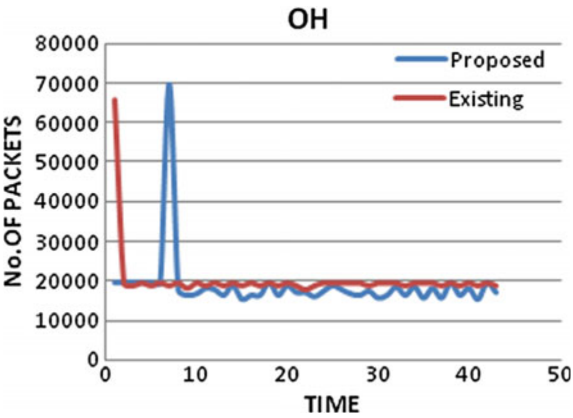
End-to-End Delay
This is defined as the mean time carried by the data packet to receive at the receiver side. In EC-AODVthe e2e delay is 48.4% low compared to AODV and is shown in table.

Dropped Packet ratio
The ratio of difference between the number of packet send and number of packet received at the destination is called as Dropped Packet ratio. In EC-AODV the Dropped Packet ratio is decreased by 1.7% compared to AODV and is shown in Table 1.

Average Energy Consumption
The sum of Energy consumed by all in nodes in the given path by the total number of hop counts gives average Energy Consumption for that path.

Over Head
Number of routing packets required for communication is referred as routing overhead. Here over head for existing system is increased compared to proposed system and is shown in Graph 1.

Graph 1 Overhead with variation of time



6 Conclusion

The proposed System expects to give the route, which has a higher energy factor from the source to the destination. WSN can endure routing break issue during packet transmission because of energy expiration. Energy efficiency of packet transmission can be enhanced by choosing an optimal path, which has effective power resource. In ECAODV, three parameters, for example, node lifetime, Energy factor furthermore, hop count include are considered routing choice. EC-AODV accomplishes a superior energy consumed and performance in QoS parameters than AODV. Be that as it may, EC-AODV is mimicked with node 30 and it does not work admirably in larger network. ECAODV still should be improved to work admirably in vast network.

References

1. Zamree Che-Aron, Wajdi Al-Khateeb, and Farhat Anwar Department of Electrical and Computer Engineering International Islamic University Malaysia Kuala Lumpur, Malaysia 50728 The “Enhanced Fault-Tolerance Mechanism of AODV Routing Protocol for Wireless Sensor Network” IJCSNS International Journal of Computer Science and Network Security, Vol. 10 No. 6, June 2010.
2. Radhika D. Joshi, Priti P. Rege, “Energy Aware Routing in Ad Hoc Networks” 6th WSEAS International conference on Circuits, System, Electronic, Control & Signal Processing, Cairo, Egypt, Dec 29–31, 2007.
3. Marina, M. K., Das S, R.; “On Demand multi-path Distance Vector Routing in Ad hoc network”, IEEE International Conference on Network Protocols, 2001.
4. Seema Verma, Pinki Nayak and Rekha Agarwal, INDIA “Energy Efficient Routing in Mobile Adhoc Networks based on AODV Protocol” IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

5. Md. Mainul Islam Mamun, Tarek Hasan-Al-Mahmud, Sumon Kumar Debnath, Md. Zahidul Islam, “Analyzing the Low Power Wireless Links for Wireless Sensor Networks”, Journal of Elecommunications, Volume 1, Issue 1, February 2010.
6. Ni Waris Chanei & Sakuna Charoenpanyasak Center of Excellent in Wireless Sensor Networks, Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Hatyai, Songkhla, Thailand. “Enhanced Mobile Ad hoc Routing Protocol Using Cross layer Design in Wireless Sensor Networks”.
7. Jerry Zhao, Ramesh Govindan, “Understanding packet delivery performance in dense wireless sensor networks”, Proceedings of the 1st international conference on Embedded networked sensor systems, November 05–07, 2003, Los Angeles, California, USA.

Aspect Based Sentiment Analysis Using R Programming

K.V. Akhil Kumar, G.V. Manikanth Sai, Nisha P. Shetty, Chetana Pujari and Aiswarya Bhat

Abstract The customers, when they buy the products online from e-commerce websites, often tend to rate these products and give reviews on that product. This rating/review system often helps the other potential customers to decide whether to purchase that product or not. However, reading all of the available reviews on a particular product, often make the customer invest a lot of time in this process as abundance of places such as blogs, review sites, etc. contain reviews. The process of sentiment analysis aims at reducing this time of the customer by displaying the data in a compact format in the form of means, analysis score, or simply histograms. The sentiment analysis procedure shown in this paper can be extended to the reviews of products in different domains. The experimental results have shown that this method exhibits better performance.

Keywords POS tagging · Sentiment analysis · Naïve bayes · R tool
SVM · e1071

K.V. Akhil Kumar (✉) · G.V. Manikanth Sai · N.P. Shetty · C. Pujari · A. Bhat
Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal University, Manipal, India
e-mail: kv.akhil95@gmail.com

G.V. Manikanth Sai
e-mail: saimanikant@gmail.com

N.P. Shetty
e-mail: nisha.pshetty@manipal.edu

C. Pujari
e-mail: chetana.pujari@manipal.edu

A. Bhat
e-mail: aiswarya.bhat@manipal.edu

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_5

1 Introduction

Nowadays, people prefer online shopping of products from the various e-commerce websites because this helps them to save time and offers them a wider range of selection at their convenience. Focusing our selection on the basis of the consumer reviews of other customers helps us to save time and filter the products based on the reviews. But most of the reviews often contain very less details about that particular product and have more of other sentences which are not useful to the potential buyer. Hence, we need to extract only that information required by the customer and trim out other undesired information, so that they can be displayed on compact devices such as mobile phones which are often handy among people. So, in this work sentiment analysis is projected for this purpose.

Sentiment analysis is one of the stages of opinion mining. In sentiment analysis, we classify the particular word into positive, negative, or neutral in order to predict the emotion of the speaker or reviewer toward the product. These reviews are given based on the customers' judgment of the product and experiences with it after using it. So, basically, the sentiment analysis of the reviews of a set of customers are being used by one customer to analyze the product and make up his mind whether to buy the product or not.

Opinion mining deals with natural language processing in order to identify the important keywords in the given sentences. Opinion mining consists of three stages —Opinion Retrieval, Opinion Classification, and Opinion Summarization. Opinion retrieval aims to extract the keywords containing the opinions or comments concerned to a particular subject of the user's interest. Opinion retrieval is followed by opinion classification which deals with classifying the extracted keywords into positive, negative, or neutral based on an existing dictionary. This can also be referred to as polarization of the words. The next stage, opinion summarization is the process of reproducing summaries from the extracted polarized keywords.

This paper aims at gathering important opinion words from the product reviews given by the existing customers taken from the Amazon review dataset, finding their orientation, i.e.- positive, negative, or neutral and finally outline the views of public to a potential user which enhances his decision-making process on whether to choose the product or not. All this has been done using R Programming language.

2 Related Work

Min Wang and Hanxiao Shi [1] have proposed an approach to realize polarity analysis of new words, and implement quantitative computation of sentiment words and automatic expansion of polarity lexicon. Their experimental results showed the feasibility and effectiveness of their approach. Their future work includes making fine-grained sentiment analysis possible from the attribute-level with an automatically built polarity lexicon.

Basant Agarwal et al. [2] worked on sentiment-rich phrases that were obtained using POS-based rules and dependency relations that were capable of extracting contextual and syntactic information from the document. Their experiments prove that by using POS patterns for the phrases, performance of sentiment analysis can be improved. In future, they would like to explore more patterns using POS Tagging to get better results.

Rui Yao and Jianhua Chen [3] have applied sentiment analysis and machine -earning concepts to study the relationship between the online reviews for a movie and the box office collection of the movie. It takes into account only positive and negative reviews leaving behind the neutral ones. Further experiments with larger datasets can be carried out to train and test the model. The comparisons among different movies can also be considered.

Kai Gao et al. [4] conducted the SVM-based algorithm to do the alternative structural formulation of the SVM optimization problem for classification. Two different datasets which are, microblogging and e-commerce were used to evaluate the performance. Experiments prove that the proposed approach, which includes feature extraction & selection and SVM, is effective in microblogging multiclass sentiment classification and e-commerce sentiment classification.

In their paper, **Pooja Kherwa et al.** [5] gather opinions and review data from e-commerce websites, social networks, popular portals, and blogs to find out what exactly people are talking about and the sentiment they are expressing. The Scoring Algorithm, which they have used, scans every line of data and gives a summary and a graphical representation if required. The efficiency of this algorithm can be improved upon if a self-learning system can be implemented.

Giovanni Acampora and Georgina Cosma [6] introduce an innovative framework consisting of methods to analyze efficiently sentiments of the customer reviews and compute their corresponding numerical data so that companies can plan their future projects. The dimension and imprecision ratings of data are calculated. As a conclusion, they propose a system to reduce the uncertainty between the reviews to validate the reviews to get the useful reviews and produce a more accurate system.

In their work, **Siddharth Aravindan and Asif Ekbal** [7] put forward a system that obtains the product features automatically from the reviews and divide it into positive and negative. It does feature extraction followed by polarity classification using association rules and supervised machine learning. Their future work is to investigate some more features for opinion mining, and to make use of classifiers that would enhance their work.

Yan Wan et al. [8] conducted a fine-grain sentiment analysis to get better results of the customer reviews and use general methods to crawl reviews and find implicit features based on POS rules. They believe that it can help producers make improvements clear and discover niche market and can also help the consumer understand the advantages as well as the disadvantages of the target product and hence make a wise selection.

3 Methodology

The modules of the system designed are illustrated in Fig. 1 and explained in the subsequent sections.

3.1 POS Tagging

Part-of-speech Tagging (POS Tagging) [9] is the process of attaching every word of a file (corpus) with its corresponding part of speech, based on its definition and relation with the adjacent phrases and words. The outcome of this process, is all the words along with their equivalent POS tag from which the words can be identified as nouns, adjectives, pronouns, verbs, etc.

This process is essential because, in the reviews, the product features are often described in the form of nouns or noun phrases and the sentiment regarding those nouns is in the form of adjectives. Therefore, extracting the noun with its corresponding adjective allows us to identify a feature of the product and the customer's emotion toward it.

The process of POS tagging involves converting each word into Unicode Transformation Format (UTF-8) so as to encode all the character vectors into 8 bit

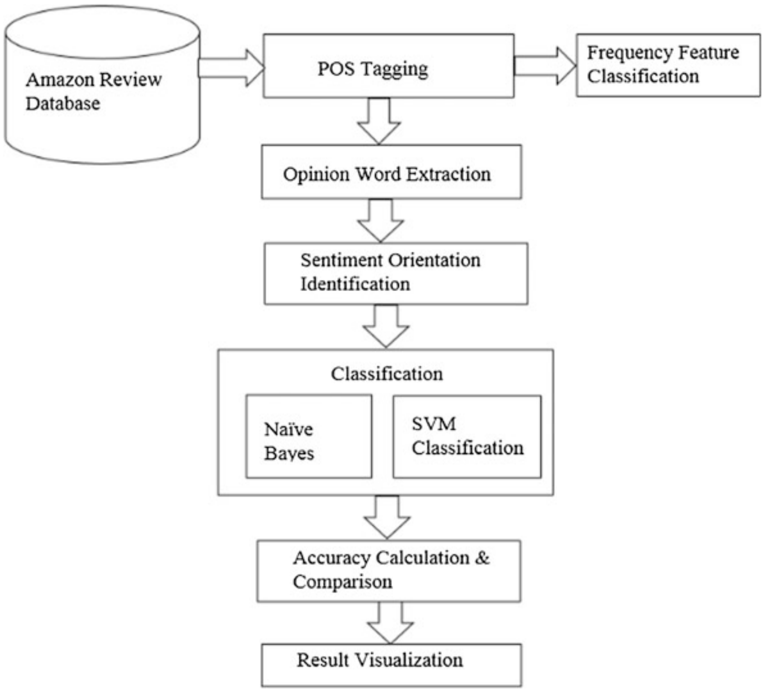


Fig. 1 System architecture

code units in order to avoid complications of byte order marks. This is followed by tokenization of all the individual characters to convert them into individual tokens. The next step is to remove the stop words which are the most common words used in a language. The program has been written in such a way that the user can add his desired words into the list of existing stop words. The final step is to apply sentence token annotations and word token annotations to the customer reviews.

From the analysis of the dataset of a particular model of a digital camera, one of the review obtained was-

the/DT macro-mode/NN is/VBZ exceptional/JJ and/VBP the/DT pictures/NNS are/VBP clear/JJ

The nouns, i.e., *macro-mode* and *pictures* and identified and the POS Tag NN is appended to them which signifies that they are nouns. The adjectives, i.e., *exceptional* and *clear* are identified and the tag JJ is appended to them signifying that they are adjectives. The extracted nouns can be used for frequent feature detection, whereas the adjectives can be used to identify the polarity of the review.

3.2 Frequent Feature Identification

As a result of the above discussion, the features of a product are described in the form of nouns. Therefore, once, the nouns and adjectives are generated, the nouns along with their respective adjectives can be used to find the most frequently repeating positive (Table 1) and negative feature (Table 2) which the customers have reviewed. From the analysis of the entire dataset of the reviews the following table of the frequent features list can be obtained.

3.3 Opinion Word Extraction

The process of POS tagging is followed by opinion word extraction which is the process of extracting all types of adjectives (i.e., comparative & superlative) in order to find the customer's emotion toward the product, i.e., positive or negative.

Table 1 Positive features

Positive features	Frequency (No. of times)
Battery	19
Picture quality	11
Zoom	11

Table 2 Negative features

Negative features	Frequency (No. of times)
Lens	5
Auto	4
Price	4

Whether the word is positive or negative can be determined by comparing each adjective extracted to a dictionary consisting of the list of positive and negative words. For example in the review:

awesome camera with great print quality but bad zoom.

The words *awesome* and *great* can easily be identified as a word having positive sentiment and therefore are counted as positive words whereas *bad* can be identified as a word having negative sentiment and hence is counted as a negative word.

3.4 Sentiment Orientation Identification

The next step is to calculate the sentiment score of each review. The sentiment score helps us to classify the total score of each review and therefore the positive, negative, and neutral reviews can be identified. A score of +1 is assigned to a positive word whereas −1 is assigned to a negative word. The total review score can be calculated by summing up the individual scores of all the adjectives in a review. In this paper, the reviews that have score greater than 0 are classified as positive, reviews having a score of less than 0 are classified as negative and a score of 0 makes the review a neutral review. For example in the following reviews:

excellent compact digital camera!

In the above sentence, the words *excellent* and *compact* are given the score of +1 each giving it an overall score of +2, therefore making it a positive review.

the main drawback of this camera is its lens.

In this sentence, the word *drawback* gives it a score −1 making the total sentence score −1, therefore making it a negative review.

amazing camera but comes at an exorbitant price

In the above sentence, the word *amazing* is given a score of +1whereas the word *exorbitant* is given a score of −1, therefore giving a total score of 0 and makes it a neutral review.

3.5 Classification

Classification should be applied on the data so that the can analyze existing data can be analyzed to predict the future trends of the data. There are many classification algorithms which make this job easy. Two classification algorithms, Naïve Bayes Classification [10] and SVM Classification [11] in R Programming language are used in this paper. For applying classification algorithms, the entire dataset is divided into training set and testing set. The training set is used to predict the results of classification on the dataset while the testing set is used to validate the results. In our experiments, we have used 33% of the dataset as testing set and the remaining as the testing set.

A. Naïve Bayes

Naïve Bayes method works on the lines of Bayes Theorem of probability to predict the class of the data. A Naïve Bayes classifier [12] estimates that the presence of one feature in a class is not related to the presence of any other feature. It is a highly scalable problem, requiring a number of parameters linear in the number of variables in a learning problem. Naïve Bayes is used so that it can give easy, fast, and accurate results compared to other classification algorithms.

B. SVM Classification

Support Vector Classification [13] is a machine-learning algorithm mainly used for classification and regression analysis. The goal of this algorithm is to find a decision boundary between the two classes that is located at the maximum distance from any point in the training data. SVM [14] is mainly used because by introducing a kernel, flexibility in threshold can be gained.

3.6 AccuracyCalculation and Comparison

Once the results of both the algorithms are acquired, the acquired results have to be compared in order to check which algorithm gives a better performance. From our analysis, we come to a conclusion that SVM Classification yields better results compared to Naïve Bayes as SVM has a higher accuracy compared to Naïve Bayes.

3.7 Result Visualization

The final result of classification can be represented in any format like bar graphs, histograms, trees, and tables. A histogram is usedcto show results of sentiment classification. Receiver Operating Characteristic Plot (ROC Plot) to show the results of our analysis. An ROC Plot is a graphical plot that depicts the results of our classification. In this, the true positive rate is plotted against the false positive rate.

4 Experiment Results

A. About the Dataset

The dataset used for this experiment is the review dataset of a particular model of popular digital camera from Amazon [15] collected over a couple of years.

This collected data has been segregated into positive and negative reviews using R Programming modules.

B. Experimental Result Analysis

The entire dataset has been given a score based on their adjectives as explained in the previous sections of the paper. Each review is given a score ranging between −2 and 6 based on the positive or negative polarity of the words. In the graph below (Fig. 2), frequency of the words is taken on the x-axis and the score of the review denoted by analysis$score is plotted on y-axis. The length of each bar of the histogram denotes the frequency of reviews of the particular score in the database.

The ROC Curve (Fig. 3) plots hit rate or true positive rate on x-axis against false alarm rate or false positive rate. The hit rate shows the part of predictions that have been identified correctly. The false alarm rate refers the expectancy of false ratio. The area under the curve shows the accuracyof the predictions.

Fig. 2 Histogram of analysis scores

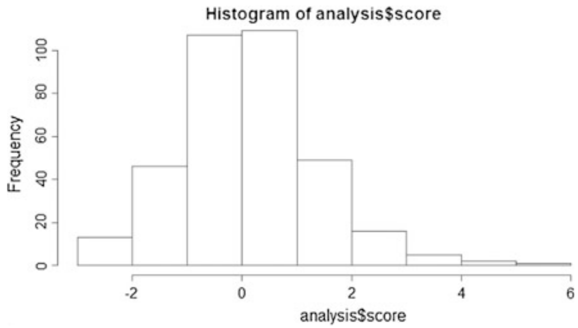


Fig. 3 ROC plot

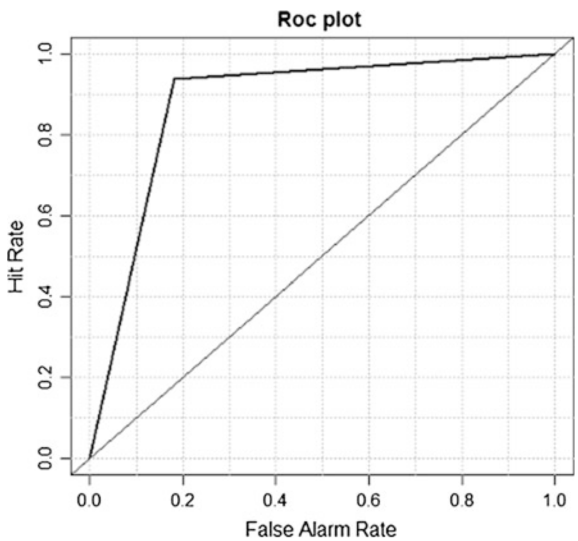


Table 3 Result comparison summary

Classification	Accuracy	Hit precision	Hit recall	Miss precision	Miss recall
SVM	0.8347	0.90	0.79411	0.7704	0.8867
Naïve Bayes	0.8014	0.8791	0.7698	0.7432	0.8645

The given table (Table 3) shows a comparison of the experimental results of SVM and Naïve Bayes classifications using Precision and Recall functions done using modules of R. Precision the fraction of extracted instances which are correct, whereas Recall is the fraction of correct instances that have been extracted. By comparing the accuracy values, it is clear that SVM has a higher accuracy and hence is a better classification method compared to Naïve Bayes.

5 Future Work

As part of our future work, the plan is to enhance the existing techniques and refine them in order to extend the process of Sentiment Analysis to include neutral sentiments along with positive and negative sentiments in order to get better results using multiclass Naïve Bayes and multiclass SVM Classification. We also plan to include several other classification algorithms in order to obtain better results. Implied Sentiments which are not expressed in words are also a means to get user views on a product. In future, we plan to understand and classify these implied sentiments accurately.

References

1. Min Wang and Hanxiao Shi "Research on Sentiment Analysis Technology and Polarity Computation of Sentiment words", Hangzhou, China, pages 331–334, 2010
2. Basant Agarwal, Vijay Kumar Sharma and Namita Mittal "Sentiment Classification of Review Documents using Phrase Patterns", In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 1577–1580
3. Rui Yao and Jianhua Chen "Predicting Movie Sales Revenue using Online Reviews" In 396 2013 IEEE International Conference on Granular Computing (GrC), pages 396–401
4. Kai Gao, Shu Su and Jiu-shuo Wang "A Sentiment Analysis Hybrid Approach for Microblogging and E-Commerce Corpus" In 7th International Conference on Modelling, Identification and Control (ICMIC 2015)
5. Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande and Prashast Kumar "An approach towards comprehensive sentimental data analysis and opinion mining" In 2014 IEEE International Advance Computing Conference (IACC), pages 606–612
6. Giovanni Acampora and Georgina Cosma, "A Hybrid Computational Intelligence Approach for Efficiently Evaluating Customer Sentiments in E-Commerce Reviews" Nottingham, United Kingdom
7. Siddharth Aravindan and Asif Ekbal "Feature Extraction and Opinion Mining in Online Product Reviews", Patna, India pages 94–99

8. Yan Wan, Hongzhurui Nie, Tianguang Lan and Zhaohui Wang "Fine-grained Sentiment Analysis Of Online Reviews" In 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) pages 1406–1411
9. Ming Sun and Jerome R. Bellegarda "IMPROVED POS TAGGING FOR TEXT–TO–SPEECH SYNTHESIS", ICASSP 2011, pages 5384–5387
10. For Naïve Bayes Classification: http://joshwalters.com/2012/11/27/naive-bayes-classification-in-r.html
11. Nagamma P, Pruthvi H. R, Nisha K. K and Shwetha N H "An Improved Sentiment Analysis Of Online Movie Reviews Based On Clustering For Box-Office Prediction" In International Conference on Computing, Communication and Automation (ICCCA2015), pages 933–937
12. Astha Gupta, R. Rajput, R. Gupta and M. Arora "Hybrid model to improve time complexity of words search in POS Tagging", Delhi, India, pages 1–6
13. Luis Gerardo Mojica and Vincent Ng "Fine-Grained Opinion Extraction with Markov Logic Networks", In 2015 IEEE 14th International Conference on Machine Learning and Applications, pages 271–276
14. For SVM Classification: https://en.wikipedia.org/wiki/Support_vector_machine
15. GuoQiang "An Effective Algorithm for Improving the Performance of Naive Bayes for Text Classification". Second International Conference on Computer Research and Development, 2010.

Frequency Domain Technique to Remove Herringbone Artifact from Magnetic Resonance Images of Brain and Morphological Segmentation for Detection of Tumor

T.D. Vishnumurthy, Vaibhav A. Meshram, H.S. Mohana and Pramod Kammar

Abstract Artifacts are frequently encountered in Magnetic Resonance Imaging. The presence of artifacts can degrade the image quality, may confuse the pathology and reduce the accuracy of image segmentation. Here, a frequency domain technique is proposed and implemented to remove Herringbone artifact in Magnetic Resonance images of brain which appear as horizontal or vertical stripes spread over the image domain. The method uses Fast Fourier transform and Canny edge detector to remove herringbone artifact. The artifact removed image is further segmented using Morphological techniques to extract the tumor part of the brain image. The results obtained for different image samples shows that the artifact removal improves the signal-to-noise ratio and accuracy of segmentation.

Keywords MR image · Herringbone artifact · Fourier transform Canny edge detector · Segmentation · Morphology · Tumor

1 Introduction

Magnetic Resonance Imaging (MRI) is a highly influential medical imaging method in modern radiology to examine the anatomy and physiology of the human brain in both health and disease. Several artifacts are frequently occurred during the

T.D. Vishnumurthy (✉) · V.A. Meshram
Department of E&CE, Jain University, Bengaluru, India
e-mail: tdv_vishnumurthy@yahoo.co.in

H.S. Mohana · P. Kammar
Department of E&IE, Malnad College of Engineering, Hassan, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_6

acquisition of MR image. Artifacts present in the image may confuse the pathology or just reduces the quality of medical investigations. The artifact ought to be removed or minimized to identify any irregularity in the brain such as tumor or lesion. Artifacts are the problems created during the acquisition of MR image and these affect the image enhancement and accuracy of image segmentation. The segmentation of MR images is challenged by an artifact. In this paper, the herringbone artifact which itself appears as a fabric or herring bone throughout the image is considered for removal. The presence of this artifact misleads the segmentation of brain images and also causes misclassification of tissues [1].

The spatial domain and frequency domain are the two techniques used to reduce the noise or artifact. In frequency domain technique, the characteristics of the artifact image are expressed as spectral components using Fourier Transform. Fourier Transform uses traditional filtering techniques, such as low- or high-pass filters to reduce the spectral components associated with artifact. This technique performs well in certain conditions, however, if the artifact is mixed with image details, then the image details are also removed. This technique smoothens the overall image but the quality of image is reduced further [2–4].

In medical applications, brain MRI segmentation is most required task because the entire analysis of the image is dependent on the outcome of the segmentation process. Hence, the different processing steps rely on accurate segmentation of anatomical regions. MR Image segmentation is frequently applied for measuring and visualizing different structures of brain. There have been different segmentation techniques for accuracy and degree of complexity. Segmentation algorithms are classified based on segmentation techniques, such as contour based, feature thresholding, region based, template matching, clustering and so on. These techniques have advantages and limitations of their own in terms of suitability, applicability, computational cost, and performance [5–9]. All these qualities can not be achieved by implementing a segmentation algorithm alone. Segmentation of brain MR Image by thresholding-or edge-based segmentation, often fail in the presence of artifact. Therefore artifact should be removed in prior to segmentation. Figure 1 describes that the artifact affects the segmentation process. In Fig. 1, given input brain MR Image is processed in the presence of herringbone artifact. Segmentation by thresholding is shown in Fig. 1b. The outcome of the segmentation using morphological operation is shown in Fig. 1c, which shows the detection of true positive pixels (green arrow) and false positive pixels (red arrow). Therefore, it is often required to remove the herringbone artifact before performing segmentation.

Morphological operations also called Mathematical morphology are used to segment the brain MR images. Morphology is the non linear branch of signal processing that deals with the application of the concepts of the set theory to image analysis. It refers to the study of structures and shape from a general scientific perspective. Morphological filters are basically nonlinear transformations that modify the geometric features of the images. The filter or operators transform the original image into another image through the process of iteration with another image of a certain size and shape referred to as structuring element [10].

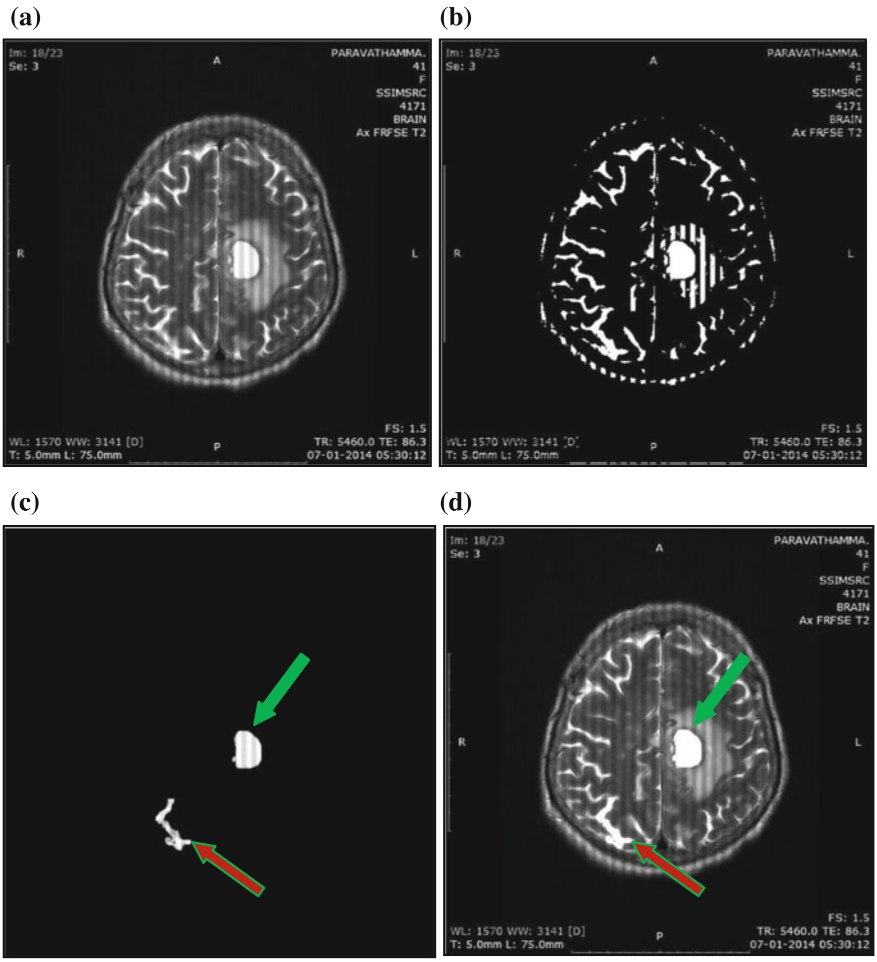


Fig. 1 Segmentation in the presence of artifact. **a** Original brain MR image with herringbone artifact. **b** Segmentation using thresholding. **c** Detection of brain tumor. **d** Tumor extracted image

2 Methodology

In the proposed algorithm, input is a brain MR image affected with herringbone artifact, the first step is to preprocess the image by removing herringbone artifact, and the second step is the segmentation by thresholding and third step to detect the brain tumor by morphological operations as indicated in Fig. 2.

A. Preprocessing to remove Herringbone artifact

The preprocessing of MR brain images is the primary step to perform image enhancement by removing artifacts present in MR Images prior to segmentation

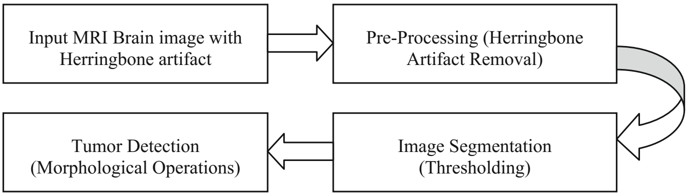


Fig. 2 Systematic overview of the proposed methodology

process. The preprocessing of the brain images starts with the removing of herringbone artifact. Following are the algorithmic steps for removal of herringbone artifact.

1. Read MRI brain image of herringbone artifact
2. Obtain the spectrum of the image using Fast Fourier Transform
3. Use canny edge detector to create filter. It is done by using Gaussian function of width sigma calculated using Eq. (1)

$$\sigma = \sqrt{\frac{\sum_{u,v}[F(u,v) - F_{\text{mean}}]^2}{MxN}} \tag{1}$$

   where, $F_{\text{mean}} = \frac{1}{MxN}\sum_{u,v} F(u,v)\, \forall\ u$ and v
4. Apply canny edge detector using suitable threshold and calculated sigma value.
5. Fill the holes in resulted image.
6. Previously generated image is complemented to restore the dc component values located at the center because it has vital information of image.
7. Resulting filter acts as a notch filter and multiply this with the spectrum (LogF) of the original artifact image
8. Restore the image using Inverse FFT

B. Segmentation

After removing herringbone artifact in the preprocessing stage, the next step is the segmentation of the MRI brain image which separates the image from its background. A global thresholding technique is applied for segmentation of the MR image and steps involved are as follows:

1. Choose a suitable threshold for the preprocessed image.
2. Convert intensity image into binary using threshold given by the Eq. (2).

$$\mathrm{g(r,c)} = \begin{cases} 1\ f(r,c) \geq T \\ 0\ \ \text{otherwise} \end{cases} \tag{2}$$

3. The pixel value which is greater than the threshold is treated as foreground or else background.

C. Tumor Detection

After segmentation, various tumor detection methods are processed on the MR brain image to detect the tumor region. The fundamental reason to apply different methods is to show only tumor region of the image which has more area and more intensity [9, 11–14]. One such tumor detection operation is mathematical morphology and windowing method [14]. Mathematical morphology is a set theoretical methodology for image analysis. Images are considered as sets of points where set operations can be performed. It extracts image components that are helpful in the depiction and description of region in terms of shape, like convex hull, skeletons, and boundaries. The dilation and erosion are most commonly used morphological operations in terms of intersection/union of an image with a translated shape known as structuring element [15]. The following are the steps involved in tumor detection:

1. Erosion is applied on the segmented image with structuring element 'disk' selected with suitable value for the 'radius' using the Eq. (3).

$$\mathrm{A} \ominus \mathrm{B} = \left\{\mathrm{z} \middle| (\mathrm{B})_{\mathrm{z}} \subseteq \mathrm{A}\right\} \tag{3}$$

   where, A is Segmented MR brain image, B is the structuring element.
2. Dilation is applied on the resulting image of step 1 with same structuring element using the Eq. (4).

$$\mathrm{A} \oplus \mathrm{B} = \left\{\mathrm{z} \middle| (\bar{B})_{\mathrm{z}} \cap A \neq \phi\right\} \tag{4}$$

   where, ϕ is empty set.
3. Create the mask of brain tumor to extract only tumor part which has more number of pixels than surrounding tissues.
4. The created tumor mask is applied on the dilated image to detect tumor part.

3 Experimental Results & Discussion

The image database is created by collecting MR images of brain from open source and also from radiologists. The proposed method is performed in the MATLAB platform. Figure 3 shows the experimental results of the preprocessing steps for artifact removal.

Artifact removed image is now used for segmentation process. Figure 4 shows the results of segmentation process. Segmentation is done by global thresholding. The preprocessed image is converted to binary image.

Figure 5 shows the outcome of the morphological operations. Erosion operation shrinks the segmented image where as dilation operation expands the eroded image. Mask is created to extract the brain tumor based on the connected components. If the object connected with large numbers of pixels then this particular object is made

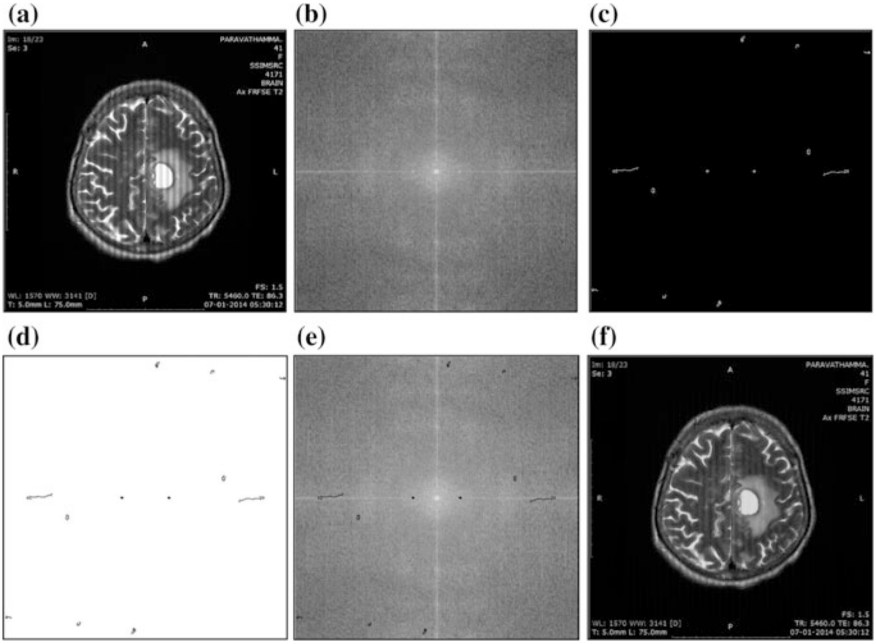


Fig. 3 Artifact removal process. **a** Input image with herringbone artifact. **b** Spectrum of input image in Log scale. **c** Result of canny edge detector. **d** Inversion of canny edge detector which is multiplied with the image spectrum of (b). **e** Resulted image after multiplication. **f** Artifact removed image

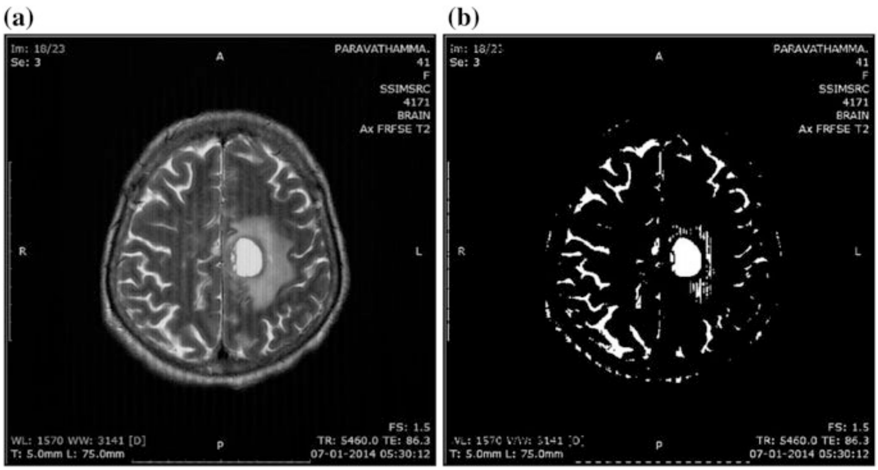


Fig. 4 Segmentation of MR image: **a** Artifact removed image. **b** Segmented image

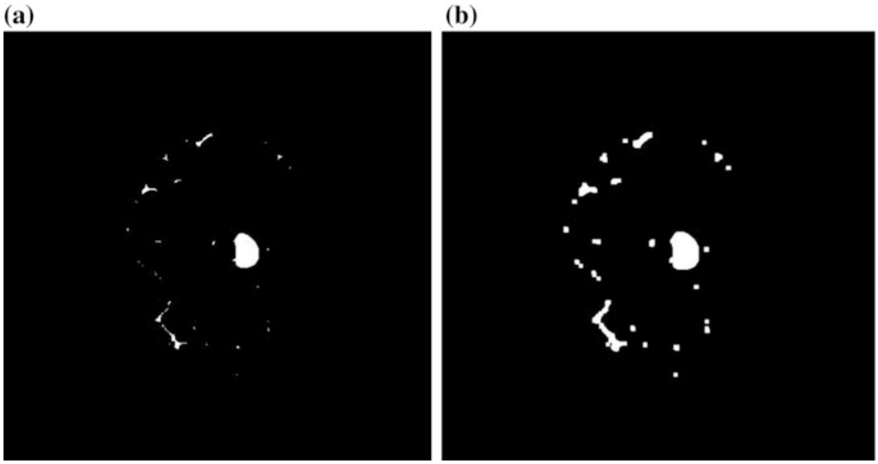


Fig. 5 Morphological operations: **a** Eroded image. **b** Dilated image

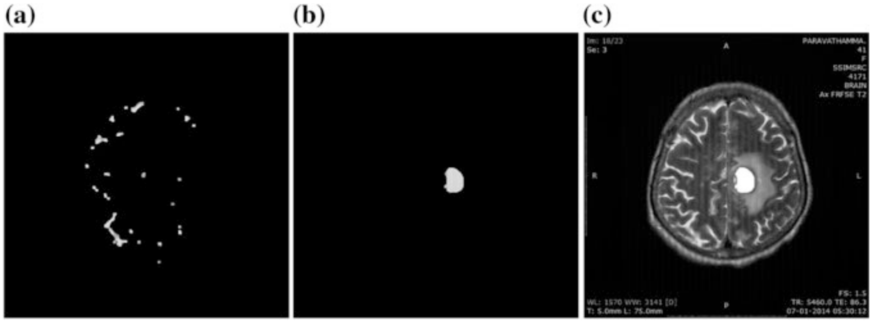


Fig. 6 Tumor detection: **a** Tumor mask. **b** Detected tumor. **c** Extracted tumor

zero in order to create the mask. Using this created mask tumor part is extracted. The results are shown in Fig. 6.

The tumor area detected by the proposed automated technique is collated with original tumor image which is segmented manually by the pathologist to check the accuracy of the segmentation. Original tumor image without artifact is manually segmented and used as a ground truth image. Accuracy of segmentation is calculated as the ratio of total area of segmented tumor to the ground truth tumor area. The number of pixels in the tumor region gives the total tumor area. The signal-to-noise ratio (SNR) is calculated using Eq. (5).

$$\mathrm{SNR} = 10\log 10\left(\frac{\mathrm{SP}}{\mathrm{NP}}\right) \tag{5}$$

where,

$$SP = \sum_{r=0}^{M-1} \sum_{c=0}^{N-1} g(r,c)^2 \tag{6}$$

and

$$NP = \frac{1}{MxN} \sum_{r=0}^{M-1} \sum_{c=0}^{N-1} [g(r,c) - g'(r,c)]^2 \tag{7}$$

g(r, c) is ground truth image and g′ (r, c) is segmented tumor image.

The sample images for testing are shown in Fig. 7.

The Table 1 depicts the tumor area, accuracy of segmentation and SNR values for sample images before and after removal of artifact.

Table 1 shows the improvisation in segmentation accuracy and signal to noise ratio. The sample input image Ima1 is over segmented before artifact removal hence its accuracy is more than hundred percent.

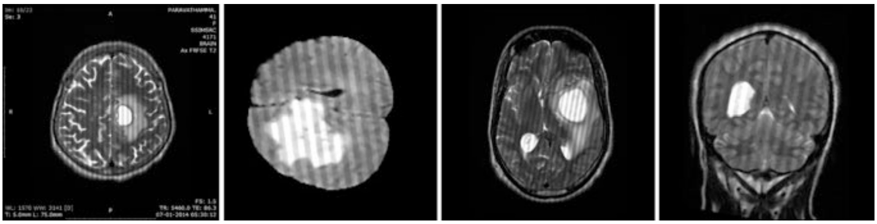


Fig. 7 Sample images (left-right) Ima1, Ima2, Ima3 and Ima4

Table 1 Segmentation assessment before and after removal of artifact with ground truth result

Sample images	Tumor area (pixels)			Accuracy (in %)		SNR (in dB)	
	Before artifact removal	After artifact removal	Actual ground truth	Before artifact removal	After artifact removal	Before artifact removal	After artifact removal
Ima1	2052	1255	1518	135.18	82.67	56.10206	76.63466
Ima2	3088	3317	3395	90.96	97.70	48.79580	50.64378
Ima3	5291	6016	6847	77.27	87.86	64.32057	68.49466
Ima4	899	942	1206	74.54	78.11	62.77677	64.47497

4 Conclusion

The proposed method removes the system related herringbone artifact from magnetic resonance images of brain and performs segmentation and detection of tumor. It is observed that the frequency domain technique preserves the edges and other image details, hence this technique is implemented to remove herringbone artifact in prior to traditional segmentation by global thresholding. Global thresholding is applied for segmentation and morphological operations such as erosion and dilation are applied with 'disk' as structuring element to detect the tumor. The technique is robust in terms of shape, size, and tumor intensity and performs well in enhancing the image by removing the herringbone artifact and detecting the brain tumor. It also shows that segmentation accuracy and signal-to-noise ratio are improved after artifact removal.

Refernces

1. Alfred Stadler, Wolfgang Schima, Ahmed Ba-Ssalamah, Joachim Kettenbach, Edith Eisenhuber, "Artifacts in body MR imaging: their appearance and how to eliminate them", European Radiology, 17, 1242–1255, 2007.
2. Chi Chang-yan, Zhang Ji-xian, Liu Zheng-jun, "Study On Methods of Noise Reduction in a striped Image", The international Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol 37, Part B6b, pp. 213–216, Beijing 2008.
3. Shu-wen W Chen, Jean-Luc Pellequer, "DeStripe: frequency-based algorithm for removing stripe noises from AFM images". Springer, BMC Structural Biology, 11:7, pp. 1–10, 2011.
4. Zuoguan Wang, Yutian Fu, "Frequency-domain Regularized Deconvolution for Images with Stripe Noise," ICIG Proc. 4th International Conference, Image and Graphics, 110–115, IEEE, 2007.
5. M. Usman Akram; Ahmad Usman, "Computer aided system for brain tumor detection and segmentation", Computer Networks and Information Technology (ICCNIT), International Conference, pp-299–302, 2011.
6. Orlando J. Tobias, Member, IEEE, and Rui Scara, Member IEEE "Image Segmentation by Histogram Thresholding Using Fuzzy sets" IEEE Transactions on Image Processing, Vol. 11, No. 12, December 2002.
7. Alan P. Mangan, Ross T. Whitaker. Surface "Segmentation Using Morphological Watersheds", IEEE Visualization '98: Late Breaking Topics, pp. 2932.
8. D. L. Page, A. F. Koschan, M. A. Abidi. "Perception-based 3D Triangle Mesh Segmentation Using Fast Marching Watersheds". Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, Vol. II, pp. 27–32, June 2003.
9. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," IEEE Transactions on Medical Imaging, vol. 20, no. 1, pp. 45–57, 2001.
10. K. Parvati, B. S. Prakasa Rao, and M. Mariya Das, "Image Segmentation Using Gray-Scale Morphology and Marker-Controlled Watershed Transformation", Hindawi Publishing Corporation, Discrete Dynamics in Nature and Society Volume 2008, Article ID 384346, doi:10.1155/2008/384346.
11. L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms". IEEE Transactions on Image Processing, Vol. 2, No. 2, pp. 176–201, 1993.

12. Clark, M.e., Hall, L.O., Goldgof, D.B., Velthuizen, R., Murtagh, F.R, Silbiger, M.S., “Automatic tumor-segmentation using knowledge-based techniques”, IEEE Transactions on Medical Imaging, vol. 117, 1998, pp. 187–201.
13. Cuadra, M.B., Gomez, J., Hagmann, P., Polio, e., Villemure, J.- G., Dawant, 8.M., Thiran, J.-Ph., “Atlas-based segmentation of pathological brains using a model of tumor growth”. In: Dohi, T., Kikinis, R (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI 2002, Springer, 2002, pp. 380–387.
14. R. Jones and I. Svalbe, “Algorithms for the decomposition of Grayscale Morphological Operation,” IEEE Transactions on. Pattern Analysis Machine Intelligence. vol. 16, pp. 581–588, 1994.
15. Robert m. Haralick, R. Sternberg, and xin hua zhuang, “Image Analysis Using Mathematical Morphology”, IEEE transactions on pattern analysis and machine intelligence, vol. Pami-9, no. 4, pp-532–550, July 1987.

Face Recognition Through Symbolic Data Modeling of Local Directional Gradient

Shanmukhappa Angadi and Vishwanath Kagawade

Abstract Face recognition systems are vulnerable to changes in expression, light, and occlusion factors. This paper presents a robust face recognition algorithm that effectively deals with these variations. The algorithm is based on the local directional gradient features that exploit the edge information in multiple directions, modeled as symbolic data object. First, face parts are detected and cropped from the face images, then the cropped images are resized to 64p $\times$ 64p grayscale images. Further, local directional gradient-based features are computed based on co-relation between pixel elements in multiple directions. The extracted directional gradient features (GC) are represented as a symbolic data object. For classification, a new symbolic similarity measure based on content is devised and employed. The experimental results on AR database show that proposed algorithm obtains an average recognition accuracy of 93.50% in the presence of illumination variations, expression changes, and occlusion variations.

Keywords Directional gradient · Face recognition · Symbolic data objects
Similarity measure · Viola-Jones · Binary pattern · Spatial convolution

1 Introduction

Face is one of the widely used universal biometric features of human beings. Due to its passive nature and ease of image acquisition, several face biometric systems/techniques have been developed and proposed by a large community of

S. Angadi
Department of Computer Science and Engineering, Centre for Post Graduate Studies, VTU, Belagavi 590018, India
e-mail: vinay_angadi@yahoo.com

V. Kagawade (✉)
Department of Computer Applications, Basaveshwar Engineering College, Bagalkot 587103, India
e-mail: vishwanath.1312@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_7

researchers. Such biometric systems have advantage over other biometrics systems requiring subjects' cooperation [1], such as iris, finger print, and palm print, etc. In general, most of face biometric recognition systems can be classified into feature and appearance-based systems depending upon the features considered for recognition [2, 3]. In feature-based face recognition systems, features of mouth, nose, eyes are detected and the relation between these parts are established. The characteristics and relations between the detected parts are used as feature descriptors. In contrast, appearance-based face recognition systems use pixel intensity values of face images for recognition.

In feature-based recognition systems, the shape and relation between facial parts are considered as feature descriptors. These systems utilize prior knowledge of human face images and most of them are based on graph matching or dynamic links. Many graph construction-based algorithms are presented which use the feature points of an image as vertices, and their corresponding feature descriptors as labels [4]. The graph constructed for face image can be used as a whole [4] or it can be used to divide the facial parts into two or more subgraphs which are further used to extract features for recognition and classification [5]. For example, the distance between facial parts and their relative positions are derived and stored in a feature space matching. During the matching process, these features are used as feature descriptors. But such representation requires reliable and accurate face detection and tracking methods, which are difficult in many situations [6].

In appearance-based recognition systems face features are extracted from whole image or some parts of the face image. More recently, Principal Component Analysis (PCA), Independent Component Analysis (ICA), Local binary pattern (LBP) [7], the local directional pattern (LDP)-based linear methods have been proposed [1]. The PCA retains most important features about original samples during the projection of samples to a feature space in order to reduce the space complexity. The ICA uses second-order and high-order statistics of input face image and projects extracted face features onto the basis vectors which are statistically independent. The LDA captures class-related information and finds a set of feature vectors that minimize the difference within a class while maximizing the difference between classes. In order to take the advantage of PCA and LDA, the KPCA plus LDA [7] method is proposed; however, it considered only lower order statistics and neglected the higher order statistical relations such as relation between three or more pixels for classification. In order to take advantage of higher statistics for classification task LBP method has been proposed. Even though, the LBP descriptor is more robust and efficient compared to other appearance-based face recognition systems to uniform illumination changes, however, it is sensitive to the presence of noise and nonuniform illumination changes in multiple directions. The gradient map gives more information about the edges or corners of an image at multiple directions [8, 9]. The gradient of adequate strength is considered important in a given task of classification [10]. The gradient-based LDP and variation of the LDP methods has been developed to handle issues of LBP features. LDP feature is obtained by computing edge information in multiple directions at each element of the image plane [3]. In contrast to linear methods many nonlinear methods such as kernel PCA have been proposed [6]. Even though much work has been done on face recognition, higher accuracy with minimal set of face features still remains a great challenge [3, 11]. However many appearance-based methods are developed which use minimum set of features for recognition but the performance of such methods is based on the type of distance/similarity measure used for classification and are computationally expensive [5]. Hence, there is a need of development of robust and efficient appearance-based system that uses minimum set of face features for person identification.

The proposed work describes an appearance-based robust and efficient person identification system which uses minimal set of features. The work also explores the advantages of symbolic similarity measures through symbolic modeling of face features [6, 12–15] during recognition and classification process. The following are some of the major contributions made in this work:

- A new directional gradient magnitude feature of a face image, using symbolic data modeling and analysis approach for face recognition.
- Issues related to changes in illumination, expression and presence of occlusion during face recognition.

A new appearance-based symbolic data modeling approach to face recognitions is proposed in this work. Initially, face parts are detected from the face images using Viola-Jones algorithm. Later, the cropped face images are resized to 64×64 pixel size and converted into grayscale images. Further from the training samples the sum of cardinal gradient magnitude (GC) in multiple directions is calculated for each class of face images and represented as a symbolic data object. Similarly for the test image, multidirectional gradient feature is calculated and represented as symbolic data object. The gradient value in multiple directions gives the valuable information about the edges or corners of the face images and plays significant role in the task of face classification. In recent works the local directional gradient features are calculated along 0° to 90° or 0° to 180° or 0° to 315° [16], however there exists many other directional gradient values in an image plane. The proposed work explores new way for the computation of local gradient features in multiple directions in the empirically chosen range. The work also proposes new content-based symbolic similarity analysis methodology for recognition of test object. The symbolic object in the trained knowledge base which yields maximum similarity is considered as identified person face class. The proposed method is tested on freely available AR [17] database that consists of face images which vary in expression, illumination, and have occlusions. The methodology outperforms and produces an average recognition accuracy rate of 93.50%.

The paper is organized into six sections. Section 2 presents an overview of the proposed approach, the technique used for feature extraction and symbolic representation is described in Sects. 3 and 4 describes the symbolic similarity analysis used for classification, further experimental results and discussions are illustrated in Sect. 5. Section 6 concludes the work and enlists the future directions.

2 Symbolic Data Modeling Approach to Face Recognition

A new approach to face recognition using symbolic modeling of multidirectional gradient features is proposed in this work. Each person class of face images is represented as a symbolic face (object) i.e. face features of images belonging to a person(class) are represented by a face object of quantitative interval-valued variables and a vector of symbolic objects. Further, a new symbolic similarity measure is devised and used for classification of test image which is also represented as a symbolic object. The different steps involved in the proposed system are illustrated in Fig. 1.

Face detection and preprocessing: In this phase, all the training and testing face samples are cropped by detecting the face using Viola-Jones algorithm [18]. All the cropped face images are down sampled to 64×64 pixels and converted into grayscale images.

Symbolic Modeling and Generation of symbolic knowledge base: This phase is carried out in two stages namely feature extraction and symbolic data modeling of face features as symbolic data objects. In feature extraction stage the sum of cardinal gradient vector magnitude is computed from each class of face images and represented as symbolic data object which is used as symbolic knowledge bases for the development of symbolic data modeling approach to face recognition.

Similarity Analysis: Using content-based similarity measure, symbolic similarity between symbolic model of training samples and symbolic data model of test face images are computed. Further, index of the most similar symbolic data object is found, i.e., symbolic data object which represents the maximum similarity is found and the result is output.

The overall process of face recognition through symbolic data modeling can be described as follows:

The face database consists of N number of classes (persons) and each class have M number of samples (each person having M number of sample images). Then, the proposed methodology involves cropping of detected face part from the original face image, extraction of face features and representation of these face features as

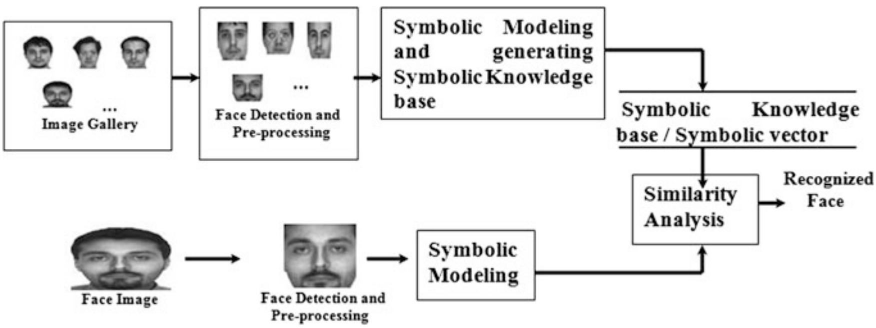


Fig. 1 Overview of symbolic data modeling approach to face recognition

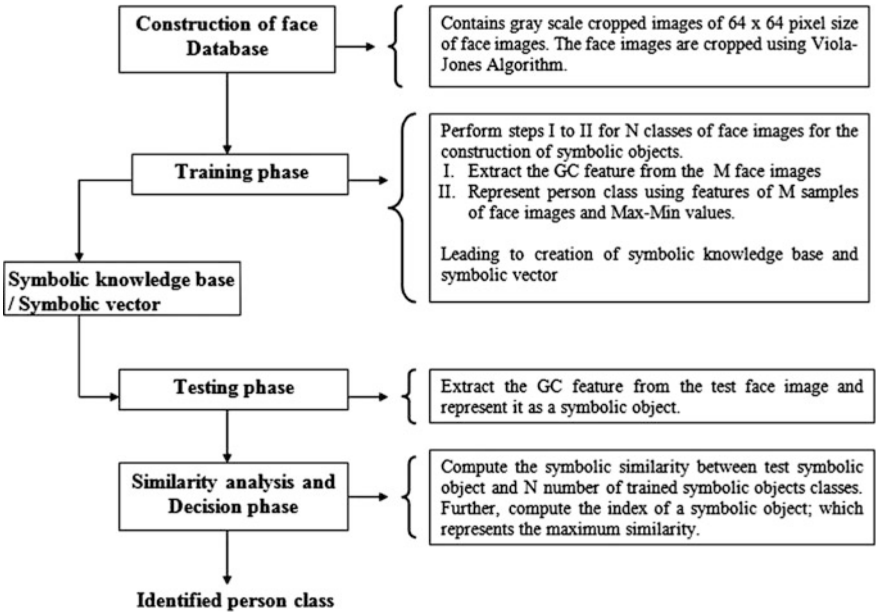


Fig. 2 Phases of symbolic data modeling approach to face recognition

symbolic knowledge base and symbolic vector. Further symbolic data analysis of trained face objects (symbolic knowledge base and symbolic vector) with features of test image + represented as symbolic object is computed. Further, the test object which is most similar to the person class object that belongs to the symbolic knowledge base is identified and output as the recognized face class. The different phases involved in the proposed model are shown in Fig. 2.

3 Symbolic Modeling

This section describes symbolic modeling process, which includes feature extraction and representation of face features as Symbolic objects.

3.1 Features for Symbolic Representation

In the work, directional gradient magnitude features of a grayscale face image are considered for face recognition. The work proposes a method that uses local correlation between neighbouring pixels (elements) in an image block for the calculation of sum of cardinal gradient vector magnitude. The method utilizes

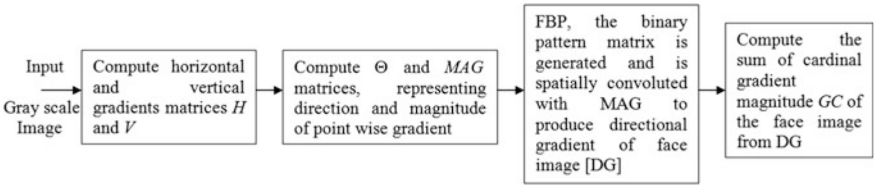


Fig. 3 Steps in computation of GC

second-order statistics, such as spatial and orientational correlation in horizontal and vertical directions. The major steps involved in the process are; computing image gradient, magnitude, edge orientation, and calculating sum of cardinal gradient vector magnitude, i.e., GC. The process involved in the computation is shown in Fig. 3.

Let $\Psi(x, y)$ be a function that represents gray scale image of size $n \times n$. Then, the gradient of a function of two variables $\Psi(x, y)$ in horizontal directions can be defined as:

$$h_{xy} = \frac{\partial f}{\partial x} \tag{1}$$

Equation (1) represents the difference in x values (horizontal direction).

In general, gradient change in horizontal direction at xyth pixel indicates change in intensity in x direction. The horizontal gradient change for all pixels of the image can be represented as in Eq. (2):

$$H = [h_{xy}]\,\forall\, \mathrm{x} = 1\,\mathrm{to}\,\mathrm{n} - 1\,\mathrm{and}\,\forall\,\mathrm{y} = 1\,\mathrm{to}\,\mathrm{n} - 1 \tag{2}$$

The gradient of a function of two variables $\Psi(x, y)$ in vertical directions indicates change in intensity of y variable and can be defined as:

$$v_{xy} = \frac{\partial f}{\partial y} \tag{3}$$

Equation (3) represents the difference in y values (vertical direction).

In general, gradient change in vertical direction at xyth pixel indicates change in intensity in y direction. The vertical gradient change for all pixels of the image can be represented as in Eq. (4):

$$V = [v_{xy}]\,\forall\, \mathrm{x} = 1\,\mathrm{to}\,\mathrm{n} - 1\,\mathrm{and}\,\forall\,\mathrm{y} = 1\,\mathrm{to}\,\mathrm{n} - 1 \tag{4}$$

Now, using the Eqs. (1) and (3) cardinal gradient magnitude *mag* at xyth pixel can be computed as:

$$mag_{xy} = \sqrt{v_{xy}^2 + h_{xy}^2}\,\forall\,\mathrm{x} = 1\ \mathrm{to\,n} - 1\ \mathrm{and}\,\forall\,\mathrm{y} = 1\ \mathrm{to\,n} - 1 \tag{5}$$

In general, cardinal gradient matrix of image $\Psi(x, y)$ can be represented by *MAG* and is computed as in the Eq. (6)

$$MAG = [mag_{xy}]\,\forall\,\mathrm{x} = 1\ \mathrm{to\,n} - 1\ \mathrm{and}\,\forall\,\mathrm{y} = 1\ \mathrm{to\,n} - 1 \tag{6}$$

Further, using the Eqs. (1) and (3) gradient orientation θxy at xyth pixel can be computed as:

$$\theta_{xy} = \tan^{-}\left(h_{xy}, v_{xy}\right)\,\forall\mathrm{x} = 1\ \mathrm{to\,n} - 1\ \mathrm{and}\,\forall\,\mathrm{y} = 1\ \mathrm{to\,n} - 1 \tag{7}$$

In general, gradient orientation matrix Θ of image Ψ can be represented as

$$\Theta = [\theta_{xy}]\,\forall\,\mathrm{x} = 1\ \mathrm{to\,n} - 1\ \mathrm{and}\,\forall\,\mathrm{y} = 1\ \mathrm{to\,n} - 1 \tag{8}$$

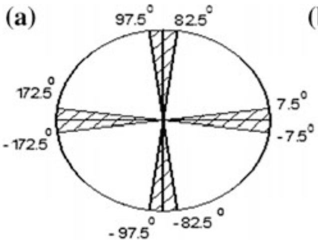
Further using the gradient orientation matrix Θ; the binary pattern matrix at xyth pixel in kth direction can be computed as described in Eq. (9)

$$bpxy^{(k)} = \begin{cases} 1\ \mathrm{if}\theta xy > \ = \alpha\ \ \mathrm{and}\ \ \theta xy\ < \beta \\ \qquad\qquad 0\ \mathrm{otherwise} \end{cases} \forall\mathrm{x} = 1\mathrm{to\,n} - 1\mathrm{and}\forall\mathrm{y} = 1\mathrm{to\,n} - 1 \tag{9}$$

In the Eq. (9) α and β represents empirically chosen specified range of directional values of a image block, k = 1 to 4 represents the four image blocks in four directions namely left, top, right and bottom directions of a plane (Fig. 4).

The content of the face image block may vary due to illumination, expression, and partial occlusion in four directions (i.e., left, right, top, and bottom of an image plane), i.e., the presence of edges and corners shows higher response values in some particular directions [3]. From the literature survey it is also found that, for the calculation of local gradient of an image the only considered gradient directions are in the range of 0° to 90° or 0° to 180° or 0° to 315° [16]; however, there exist many other ways to calculate multidirectional gradients of an image to accumulate

(a)



(b)

$$BP^k = \begin{pmatrix} BP^1 \\ BP^2 \\ BP^3 \\ BP^4 \end{pmatrix} = \begin{pmatrix} ((\theta >= \alpha_1)\ \mathrm{and}\ (\theta < \beta_1)) \\ ((\theta >= \alpha_2)\ \mathrm{and}\ (\theta < \beta_2)) \\ ((\theta >= \alpha_3)\ \mathrm{and}\ (\theta < \beta_3)) \\ ((\theta >= \alpha_4)\ \mathrm{and}\ (\theta < \beta_4)) \end{pmatrix}$$

Fig. 4 Directional computation of GC more gradient information which is essential for the development of robust and efficient face recognition system. Hence in the proposed model in order to compute total change in all four directions of a face image block, the range of θ (0, 2 π) is divided into four equal blocks. Figure 4a, b illustrate the range of values of α and β in four directions and are listed in Eq. (10)

$$\left.\begin{array}{l}\alpha = \{-7.5^\circ, 82.5^\circ, 172.5^\circ, -97.5^\circ\} \\ \beta = \{7.5^\circ, 97.5^\circ, -172.5^\circ, -82.5^\circ\}\end{array}\right\} \tag{10}$$

In this work of face recognition after thorough experiments the range of θ (0, 2 π) is divided into k directions using empirically chosen range of values for α and β. It is found that four equal division of θ (0, 2 π) with the given range of α and β as shown in Fig. 4 a, b gives the better performance. Hence in the proposed work four equal division of 15° degree in the four directions as shown in the Fig. 4a, b is employed.

In general, prominent binary pattern matrix in kth direction can be computed as in Eq. (11).

$$BP^{(k)} = [bp_{xy}^{(k)}]\,\forall\,\mathrm{x = 1\,to\,n - 1, y = 1\,to\,n - 1\,and\,k = 1\,to\,4} \tag{11}$$

Now the full binary pattern matrix in all four directions of a image plane can be computed by logically OR ing in the four matrices BP^1, BP^2, BP^3 *and* BP^4 as in Eq. (12).

$$FBP = BP^1 \vee BP^2 \vee BP^3 \vee BP^4 \tag{12}$$

where $\vee$ represent logical OR operation.

Further spatial convolution of *FBP* and *MAG* can be computed and represent the directional gradient matrix *DG* as in Eq. (13).

$$DG = (FBP. * MAG) \tag{13}$$

In general, the sum of cardinal gradient vector magnitude (GC) of a face image Ψ in four directions can be computed as in Eq. (14).

$$GC(\Psi) = \sum_{x=1}^{n-1}\sum_{y=1}^{n-1} DG \tag{14}$$

For each element of image block, the proposed work considered all the neighbouring elements and computed directional gradients using the Eq. (14). The main purpose of calculating GC is to obtain total edge weight of different edge information by considering gradients in appropriate directions. More edge information reflects higher variation [3, 8, 10] of the face image. So, the calculation of sum of cardinal gradient magnitudes is supposed to play an important role in the task of face image classification in the presence of variation in expression, illumination, and occlusion. Such variations of face images are found in freely available AR database. The proposed work uses GC as a feature descriptor and is employed in symbolic data modeling of face image classes which is described in Sect. 3.2.

3.2 Symbolic Data Modeling of Face Image

Let the face image database consists of N number of face image classes and each class of face image consists of M number of face images that vary in expression, occlusion, and illuminations (Images are drawn from AR database).

Let $\Omega = \{\Psi_1, \Psi_2, \Psi_3 \ldots\ \Psi_M\}$ be a set that represents M images of a face class, which are first-order objects. In general, all face image classes can be represented as

$$\mathrm{U} = \{\Omega_{\mathrm{i}}\}\ \forall\, \mathrm{i} = 1\ \mathrm{to}\ \mathrm{N} \tag{15}$$

The vector representing the symbolic data model of face images belonging to the ith class $\mathrm{SV}^{(\mathrm{i})}$ is represented in Eq. (16)

$$\mathrm{SV}^{(\mathrm{i})} = \{GC(\Psi_1^i), GC(\Psi_2^i), \ldots, GC(\Psi_M^i)\} \forall\, \mathrm{i} = 1\ \mathrm{to}\ \mathrm{N} \tag{16}$$

Further, the symbolic interval data representation of ith class using maximum and minimum value of $SV^{(\mathrm{i})}$ is computed as:

$$\mathrm{GC} - \mathrm{MM}^{(i)} = \{max(SV^{(\mathrm{i})}), min(SV^{(\mathrm{i})})\}\ \mathrm{where}\ \forall\, \mathrm{i} = 1\ \mathrm{to}\ \mathrm{N} \tag{17}$$

Now using the Eq. (17) symbolic knowledge base of N classes is represented as follows:

$$SKB = \{\mathrm{GC} - \mathrm{MM}^{(i)}\}\ \forall\, \mathrm{i} = 1\ \mathrm{to}\ \mathrm{N} \tag{18}$$

The matrix *SKB* represent symbolic knowledge base of the all face classes. The symbolic similarity analysis can be done by considering *SKB, SV* of trained objects and test symbolic object.

4 Symbolic Similarity Analysis

The symbolic features can be classified into quantitative, qualitative, and structured variables depending upon nature of data they represent. The similarity analysis of such components can be carried out employing position, span, and content [13] measures. The choice of similarity measure by position, span, and content depends upon type of symbolic features considered for representing the object. In this work, each symbolic object is represented using quantitative interval data variable representing the person face image (Eq. 17) and symbolic vector class (Eq. 16). Hence best choice of symbolic similarity measurement is content similarity. The symbolic similarity measure used in this work is a modified form the content-based symbolic similarity measurement proposed in the earlier works namely [13]. The procedure used in this work for employing content-based symbolic similarity measurement is described as follows:

Let

Tobj = {GC} be a GC feature extracted from the test face sample Ψ.

$SO^{(i)}$ be the ith symbolic object in the knowledge base representation $SKB/\mathrm{SV}^{(\mathrm{i})}$.

Then, the similarity measure *S* between two symbolic objects *Tobj* and an object in symbolic knowledge base represented by SKB/SV^{i}.

Now consider: $SO^{(i)} \in SKB/SV^{(i)}$ and *Tobj* are two qualitative interval type feature vectors, $\forall\, i = 1 \text{ to } N$.

Let

al = lower limit of interval $GC\text{-}MM^{(i)}$(min value), belonging to $SO^{(i)}$
au = upper limit of interval $GC\text{-}MM^{(i)}$(max value), belonging to $SO^{(i)}$
bl = lower limit of interval *Tobj* (min value)
bu = upper limit of interval *Tobj* (max value)

In test object *Tobj* both upper bound and lower bound are equal, i.e., $bl == bu$

Let *inters* = number of common elements between $SV^{(i)}$ and *Tobj*

l_s = span length of $GC\text{-}MM^{(i)}$ and *Tobj*

$= |\, max\ (au,\ bu) - al|$

The $M + 2$ symbolic features of *SO* can be represented as:

$$SO^{(i)} = \{SV^{(i)}, GC - MM^{(i)}\} \tag{19}$$

where $SV^{(i)}$ is $1 \times M$ vector of $GC(\Psi_j{}^{i})$ values, $\forall\, \mathrm{i} = 1 \text{ to } \mathrm{N} \text{ and } \forall\, \mathrm{j} = 1 \text{ to } \mathrm{M}$.

Then qualitative interval type similarity component due to content between $SO^{(i)}$ and *Tobj* features can be defined as:

$$S^{(i)}\left(SO^{(i)}, Tobj\right) = inters^{(i)}/l_s \tag{20}$$

where $inters^{(i)}$ is computed using overlap measure method. Overlap measure assigns a similarity of 1 if the values are approximately equal otherwise assigns similarity as 0. Hence, Overlap measure between $SO^{(i)}$ and *Tobj* at *ith* class can be computed as:

$$\mathrm{inters}^{(i)} = \sum_{j=1}^{M} CMP(SV_j^i, Tobj)\,\forall\,\mathrm{i} = 1 \text{ to } \mathrm{N} \tag{21}$$

where $CMP(SV_j^i, Tobj) = \begin{cases} 1 & \text{if } GC(\Psi_j^i) \cong GC(Tobj) \\ 0 & \text{otherwise} \end{cases}$ and $\forall\,\mathrm{i} = 1 \text{ to } \mathrm{N}$ and $\forall\,\mathrm{j} = 1 \text{ to } \mathrm{M}$.

Further, the net similarity of test object *Tobj* with all the N face classes can be represented as a 1 × N vector.

$$SIM = \left[S^{(i)}\right], \forall \, \mathrm{i} = 1 \, \mathrm{to} \, \mathrm{N} \tag{22}$$

The find the identify of person class, class with max similarity is found in Eq. (23)

$$\mathrm{ID\ \text{-}\ person\ \text{-}\ class} = \max(SIM) \tag{23}$$

where max(*SIM*) returns the index of the identified person class.

The results obtained from the proposed symbolic data modeling for face recognition are discussed in the results and performance analysis section.

5 Results and Performance Analysis

The experiments are conducted using the AR database. The AR database consists of 120 classes of face images, having 26 samples per class. The face images vary in illumination, expression and have partial occlusion and are taken in two sessions. The AR database has been used to evaluate the efficiency and robustness of the proposed algorithm with reference to different variations. The results of experiments show that, the performance of the proposed algorithm is superior and the recognition speed is fast because face symbolic objects are represented with minimum number of features. Hence, there is a minimum computation involved in comparison between test feature vector and trained features vectors represented as symbolic data objects. In the initial stage, face parts are detected using Viola-Jones algorithm [18] and complete face part is cropped. The cropped images are resized into 64 × 64 pixel size and are converted into grayscale image. Some of the images employed for training and testing are shown in Fig. 5.

Fig. 5 Samples of original (**a**), and corresponding cropped images (**b**) of AR database

Table 1 Composition of AR database and recognition results

Labels	Total no. samples	No. of samples recognized correctly	No. of samples misclassified	Accuracy in %
S1	480	440	40	91.67
S2	420	400	20	95.24
S3	(73 + 25) = 98	(70 + 25) = 95	(3 + 0) = 3	96.94
S4	244	227	17	93.03
S5	56	55	01	98.21
S1 + S2 + S4 + S5	1200	1122	78	93.50

In order to verify the effectiveness of the proposed methodology, 1200 images of 120 persons from AR database is considered to build the image gallery. 1200 face images are partitioned into five unequal subsets, S1 through S5. The subset S1 contains some of images that vary in left light, right light, and both side lights, S2 contains some of the face images that vary in expressions such as neutral expression, smile, anger, and scream. S3 contains face images of persons wearing scarf and sunglass. S4 contains face images of persons wearing scarf in uniform light and in varying light conditions. S5 contains face images of persons wearing sunglass in uniform light and in varying light conditions. The results obtained by experimentation on S1–S5 set are reported in Table 1.

It is observed that symbolic similarity measurement method with proposed symbolic data modeling technique for face recognition system achieves higher average recognition rate of 93.50% (Table 1).

The implementation of proposed model is carried out in Matlab R2013b platform on PC with Intel®B960 processor, 2 GB DDR3 memory. The total time taken by the proposed methodology to construct symbolic knowledge base and symbolic vector is 83.489 s. The average execution time for recognition is 0.108 s per test face images is obtained.

6 Conclusion

In this paper, a new approach for face recognition that uses directional gradients features between neighbouring pixels is presented. For each face image a sum of cardinal gradient is calculated by integrating more optimal number of directional gradients. The proposed work accumulates more edge information of an image, which achieves more efficient and robust image similarity within the class and differences amongst different classes of face images. The method employs symbolic modeling to represent the face image classes and this has minimized the number of features required for representation. The experiments show that the work is efficient and robust and accurately handles.

References

1. Amrit Kumar Agrawal, Yogendra Narain Singh, 2015, Evaluation of Face Recognition Methods in Unconstrained Environments, Procedia Computer Science 48, 644–651, www.sciencedirect.com.
2. Amr Almaddah, Sadi Vural, Yasushi Mae, 2014, Face relighting using discriminative 2D spherical spaces for face recognition, Machine Vision and Applications 25:845–857 DOI 10.1007/s00138-013-0584-z, Springer-Verlag Berlin Heidelberg 2013.
3. Taskeed Jabid, Hasanul Kabir Md., Oksam Chae, 2010, Robust Facial Expression Recognition Based on Local Directional Pattern, ETRI Journal, No. 5, October 2010, Volume 32.
4. Gayathri Mahalingam, Chandra Kambhamettu, 2010, Age Invariant Face Recognition Using Graph Matching, 978-1-4244-7580-3/10/$26.00.
5. Jiwen Lu, Venice Erin Liong, Gang Wang, Pierre Moulin, 2015, Joint Feature Learning for Face Recognition, IEEE Trans. on information forensics and security, 2015, Vol. 10, No. 7.
6. Hiremath P.S., Prabhakar C.J., 2008, Face Recognition Using Symbolic KPCA Plus Symbolic LDA in the Framework of Symbolic Data Analysis: Symbolic Kernel Fisher Discriminant Method, ACIVS, LNCS 5259, pp. 982–993, 2008, Springer-Verlag Berlin Heidelberg.
7. Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, Jianzhuang Liu, 2010, Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor, IEEE Transactions on Image Processing, Vol. 19, No. 2.
8. Ping Yang, Xin Tong, Xiaozhen Zheng, Jianhua Zheng, Yun He, 2009, A Gradient-Based Adaptive Interpolation Filter for Multiple View Synthesis, PCM 2009, LNCS 5879, pp. 551–560, 2009. Springer-Verlag Berlin Heidelberg.
9. Liang Yunjuan, Feng Hongyu, Zhang Lijun, Miao Qinglin, 2012, Gradient Direction Based Human Face Positioning Algorithm Applied in Complex Background, Advances in Technology and Management, 2012, AISC 165, pp. 385–391.
10. Shys-Fan Yang-Mao, Yung-Fu Chen, Yung-Kuan Chan, Meng-Hsin Tsai, Yen-Ping Chu, 2008, Gradient Direction Edge Enhancement Based Nucleus and Cytoplasm Contour Detector of Cervical Smear Images, ICMB, LNCS 4901, pp. 290–297.
11. Hiremath P. S., Prabhakar C. J., 2005, Face Recognition Technique Using Symbolic PCA Method, Springer-Verlag Berlin Heidelberg, LNCS 3776, pp. 266–271.
12. Chidananda Gowda K., Edwin Diday, 1991, Symbolic clusters using a new dissimilarity measure, Pattern Recognition, 24(6): 567–578.
13. Chidananda Gowda K., Edwin Diday, 1992, Symbolic clustering using a new similarity measure, IEEE Transactions on Systems, Man, and Cybernetics, 22(2): 368–378.
14. Nagabhushan P., Angadi S.A., Anami B.S., 2006, A Fuzzy Symbolic Inference System for Postal Address Component Extraction and Labelling, Springer-Verlag Berlin Heidelberg LNAI 4223, pp. 937–946.
15. Hiremath P.S., Prabhakar C.J., 2006, Face Recognition Technique Using Symbolic Linear Discriminant Analysis Method, Springer-Verlag Berlin Heidelberg, P. Kalra and S. Peleg (Eds.): ICVGIP, LNCS 4338, pp. 641–649.
16. Zhaokui Li, Yan Wang, Chunlong Fan, Jinrong He, 2015, Image preprocessing method based on local approximation gradient with application to face recognition, Pattern Anal Applic, Springer, DOI 10.1007/s10044-015-0470-6.
17. Martinez A. M, Benavente R, 1998, The AR database, CVC technical report.
18. Viola P., Jones M, Robust real-time face detection, International Journal of Computer Vision, 2004, 57(2):137–154.

Multidimensional Image Indexing Using SR-Tree Algorithm for Content-Based Image Retrieval System

Janakiraman Vanitha and Muthukrishnan Senthilmurugan

Abstract Indexing is used to reduce the time required for query operation. It will minimize the time of average case and also the worst case. It also support dynamic insertion and deletion. For applying this technique to the huge databases, we need to efficiently create multi dimensional index structures, supporting nearest neighbor query. The feature extraction of the image and multidimensional indexing is used to perform content-based Image retrieval system. The SS-tree which stands for Sphere Sphere Tree occupy much larger volume than R*-Tree (Rectangle tree) and K-D-B Tree which reduces search efficiency. SR-Tree which stands for Sphere Rectangle Tree combines the bounding sphere and rectangle. A 'region' of the SR-Tree is small volume and a short diameter by the region intersection with bounding sphere and bounding rectangle which enhances disjoint among the regions and improves performance on the nearest neighbor queries especially for non uniform data. Here, the total process time and number of time disk access the leaf and internal node measures the index structure performance. The performance test results show SR-Tree performed most efficiently among other similarity indexing structure. The Proposed CBIR system is using SR-Tree algorithm after the process of extracting color and spatial feature and stored the values of Hue, Saturation, Value, Color Histogram, and Color cooccurrence matrix with the images.

Keywords Content-Based Image Retrieval (CBIR) · Feature extraction
Multidimensional indexing · Sphere Rectangle Tree (SR-Tree)

J. Vanitha (✉)
Bharathiyar University, Coimbatore, India
e-mail: vani77_muthu@yahoo.co.in

J. Vanitha
Department of MCA, EGS Pillay Engineering College, Nagapattinam, India

M. Senthilmurugan
AVC College of Engineering, Mayiladuthurai, India
e-mail: smsenthil@hotmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_8

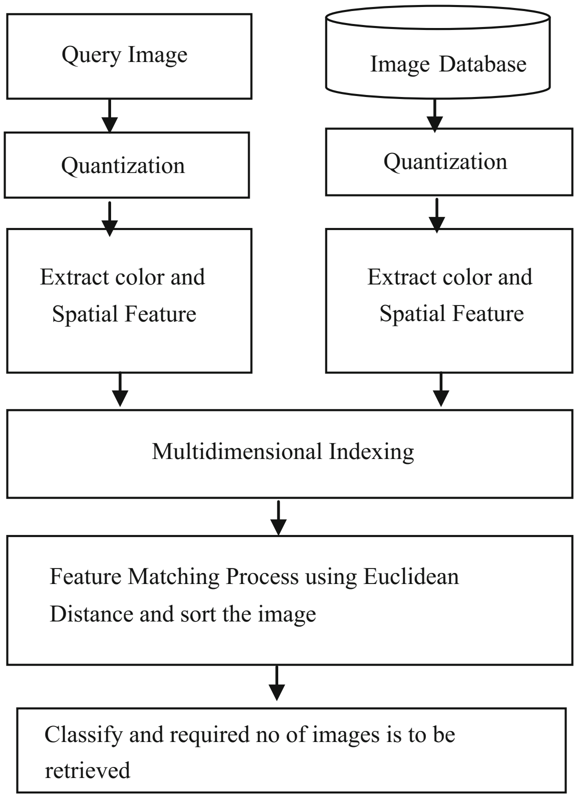
1 Introduction

The Content-Based Image Retrieval (CBIR) system (Fig. 1) is the efficient image retrieval system which uses the image contents such as shape, texture, color, and spatial information. Three stages of Image retrieval are,

1. Feature extraction from image database and query image.
2. Feature Matching Process between querying image and image database and images are indexed in database.
3. Classify the database images in database and fix number of similar images are to be retrieved.

The feature vectors of Content-Based Image Retrieval are indexed with image for improving the search efficiency. The multidimensional index structures supporting nearest neighbor query is used to apply in large databases. Similarity indexing is differentiating from other type of database indexing. These are

Fig. 1 CBIR architecture



(i) The images being indexed are represented by high-dimensional feature vectors usually 5–100 or higher. But other indexing algorithms will not process data of high dimensions.
(ii) Query for 'feature vectors' is done by one or more similarity or dissimilarity measure.
(iii) The types of query are performed different in similarity indexing than the other database indexing.

The SS-Tree is one of the best similarity indexing structure compare with R*-Tree and with the K-D-B-Tree. But it occupies much larger volume. The drawback overcomes with SR-Tree which incorporates bounding sphere and bounding rectangle. Total CPU time and count of disk accesses leaf and internal node measures the index structure performance [11]. The Performance test shows SR-Tree out performs both the R*-Tree and SS-Tree. Here especially for the nonuniform data the SR-Tree improves the performance of the nearest neighbor queries.

The remaining of the paper had been arranged as stated. Section 2, review the literature. In Sect. 3, the methods of feature extraction are to be summarized. In Sect. 4, the purpose of multidimensional indexing is explained. Feature matching process is detailed in Sect. 5. Section 6 proposes the conclusions and future potential research directions. The Sect. 7, provide the list of references.

2 Literature Review

The Similarity indexing mainly applied in visual information system. It includes image and video. The similarity query on feature vectors could find images have similar color, texture, shape, and spatial information [1, 2]. The feature vectors are pulled out from image, for example, color, hue, saturation, value, texture etc., and indexing with images. The k-d tree was the first multi dimensional indexing structure for nearest neighbor queries [3]. The 'K-D-B Tree' [4] is multidimensional indexed structure which follows the root to a leaf search path, and hence, the time of search is logarithmic to the size of database. To index multidimensional information such as the geographical coordinates, polygons, or rectangles the R-Trees [5] tree data structures are used. To index spatial information the variant of R-trees, and the R*-Trees [6] are used which has a insignificant higher construction cost than R-Trees because the data may need to be reinserted. However, the resulting tree will usually have enhanced query performance. Similar to the standard R-Trees both point and spatial data can be stored in it. The TV-Tree [7] enhances the R*-Tree performance by reduction of dimension based on its significance. The optimized version of R-Tree which is VAMSplit R-tree [8] is better than R*-Tree. The forced reinsertion mechanism of the R*-Tree is modified by the SS-Tree [9]. SS-Tree re-inserts entries unless re-insertion has been made on the same leaf or node. It supports the dynamic restructuring of the tree structure. The X-Tree [10] is a modification of the R*-Tree by using the overlap-free split and super node mechanism. The SR-Tree [11] enhances the disjoint among regions which enhances the performance on the nearest neighbor queries. The SR-Tree is the best multidimensional indexing structure among the SS-tree, the R*-Tree, and the K-D-B-Tree.

3 Feature Extractions

The image features are extracted using color histogram and color co-occurrence matrix. Before extracting, the image is quantized. This is the method of reducing count of bins taking the color that are most identical to each one and placing them inside the same bin. The Minimum bin value is 8*8 in Matlab and the Maximum bin value is 256*256.

3.1 The Color Histogram

Color histogram defined as "a set of bins where each bin denotes the probability of pixels in the image being of a particular color". Color histograms are classified into two, (i) Global color histograms (ii) Local color histogram. The Global Color Histogram means a whole image convert to the single color histogram.

A Local Color Histogram partitions an image (Fig. 2) into fixed blocks. It then gets the color histogram of each of the blocks [12]. The 3*3 block color histogram is considered better than global color histogram from the perception of human visual [13]. The 3*3 block color histogram (Fig. 3) has two types. These are

Type 1: Equally divided and same weight given for each division.
Type 2: Unequally divided and double the weight given for center division (Fig. 4).

The Type 2 is used in the Proposed CBIR System for extracting color feature.

Fig. 2 Original image

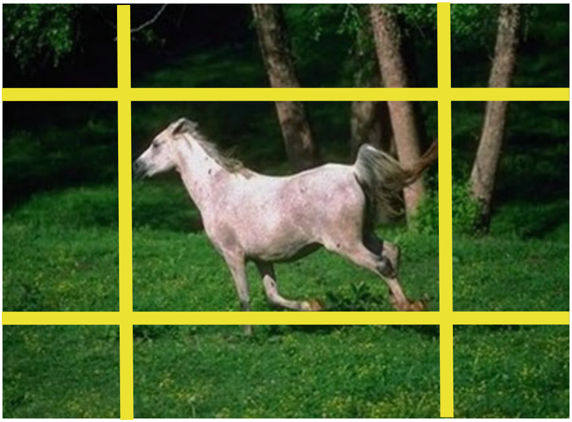


Fig. 3 Segmented image

Fig. 4 Weight for each block with proportion 1:2:1

1	2	1
2	4	2
1	2	1

3.2 Color Co-occurrence Matrix

The Color Co-occurrence Matrix (CCM) includes three dimensional matrix in which the first dimensions and second dimensions contain colors of any pair. The third dimension includes spatial distance among them. CCM is made flexible to characterize the number of color (hue) pair between adjacent pixels in the image. For each pixel in the image, here then horizontal and vertical neighbors (4-neighbors) are accounted [14]. The CCM is used to extract spatial feature in this proposed CBIR.

4 Multidimensional Indexing

The SR-Tree identifies a region by the intersection of bounding rectangle and a bounding sphere (Fig. 5). It improves the performance on nearest neighbor queries [11]. The organization of the SR-Tree has its base on R-Tree, R*-Tree, and of the SS-Tree.

A Leaf node (Fig. 6a) has the following organization.

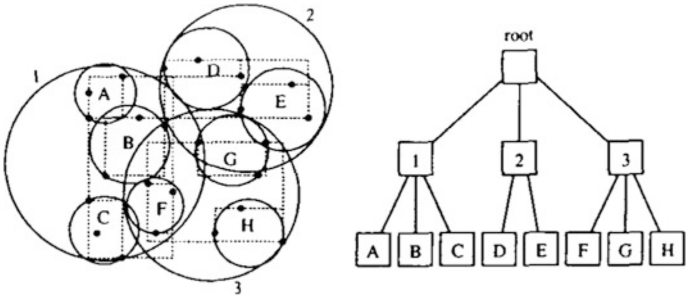


Fig. 5 The SR-tree structure

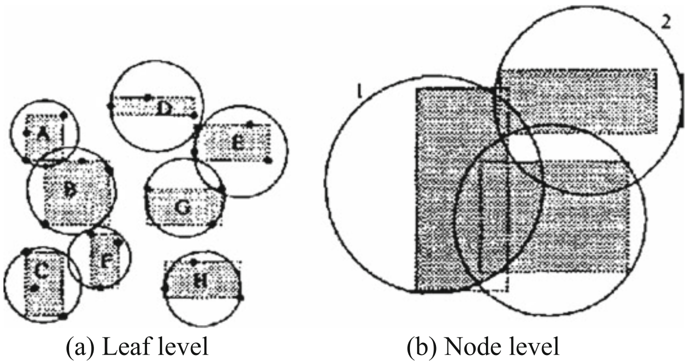


Fig. 6 Specification of regions by the intersection of a bounding sphere and a bounding rectangle

L: (E_1, E_2, …, E_n) ($m_L \leq n \leq M_L$ Where m_L and M_L are min and Max number of entries in a leaf)
E_i: (p, data) (Every entry have point p and associated attribute data)

A Node (Fig. 6b) of the SR-tree is

N: (C_1, C_2, …, $C_{n)}$ ($m_L \leq n \leq M_L$ Where m_L and M_L are min and Max number of entries in a node)
C_i: (S, R, w, child-Pointer)

Each entry corresponds to a child of the node and consists of the following four components: a bounding sphere denoted by S, a bounding rectangle denoted by R. The number of points denoted by w and a pointer to the child child-pointer. The SR-Tree insertion algorithm is based on SS-Tree [10]. The difference between SR-Tree and SS-Tree is in the method of updating the regions on top of the insertion of a new entry. The approach of updating bounding rectangle is same as R-Tree and R*-Tree. But the way of updating bounding sphere is different that of SS-Tree. Because SR-Tree determines the parent node of the bounding sphere by utilizing the children of both bounding sphere and bounding rectangles. The deletion algorithm is similar with R-Tree [5]. The nearest neighbor search is same as R*- tree and the SS-Tree [15]. This algorithm is a depth first search and has two stages. These are

(i) To make a candidate set it gather the given count of points.
(ii) It visits each leaf which contains overlapping region overlaps and revises the candidate set.

Also the traversal algorithm is same as R*- Tree and SS -Tree. The SR-Tree differs from R* Tree and SS Tree in the approach of calculating the distance from the search point to every region [11]. The performance test shows it is more efficient for fewer uniform data set. The proposed CBIR has the process of the feature vectors are stored with images using the SR –Tree algorithm for increasing similarity searching efficiency.

5 Feature Matching Process

The Feature matching process is used to find the identical images. The following steps are used for sorting the images in the image database.

Step 1: Similarity measures between image used as query and the images in database using Euclidean distance of Weighted 3*3 block color histogram ($ED_{wbch}(I_i)$).

Step 2: Similarity measures between image used as query and the images in database using Euclidean distance of Color Co-occurrence Matrix of HSV of a pixel ($ED_{ccm}(I_i)$).

Step 3: Images stored in the database are sorted using the addition of distance value from step 1 and step 2

$$(ED_{wbch}(I_i) + ED_{ccm}(I_i)).$$

Then the images in the database are classified and fixed number of images is to be retrieved.

6 Conclusions

Feature extraction can be carried out on unequally divided and double the weight given for center division of image using 3*3 block color Histogram for color feature and color co-occurrence Matrix of HSV of a pixel for spatial feature after quantize the image. The SR-Tree algorithm is used to store the values of Hue, Saturation, Value, Color Histogram, and Color co-occurrence matrix with the images. The features matching processes is used Euclidean distance and sort the images by the addition of distance value of histogram and Color co-occurrence Matrix. The future work is doing the classification of the images and the fixed count of images is to be retrieved. It also tries another indexing technique to give the better result. Therefore to provide better performance of retrieving more images are needed on the basis of these techniques.

References

1. V.N. Gudivada and V.V. Raghavan, "Content Based Image Retrieval Systems," IEEE Computer, Vol. 28, PP 18–22, Sept. 1995.
2. C. Faloutsos, R. Barber, M. Flickner, J. Hafner, et al., "Efficient and effective querying by image content," Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, Vol. 3, PP. July 1994.
3. J. H. Friedman, J. H. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," ACM Transactions on Mathematical Software, Vol. 3, pp. 209–226, Sept 1977.
4. J. T. Robinson, "The K-D-B-tree: a Search Structure for Large Multidimensional Dynamic Indexes," Proc. ACM SIGMOD, Ann Arbor, US.4, pp 10–18, Apr. 1981.
5. A. Guttman, "R-Trees: a Dynamic Index Structure for Spatial Searching," Proc. ACM SIGMOD, Boston, USA, pp. 47–57, June. 1996.
6. N. Beckmann, H. P. Kriegel, R. Schneider and B. Seeger, "The R*-tree: an Efficient and Robust Access Method for Points and Rectangles," Proc. ACM SIGMOD. Atlantic City, US. 4, pp. 322–331, May 1990.
7. K. I. Lin, H. V. Jagadish and C. Faloutsos, "The TV – Tree: An Index structure for High Dimensional Data," VLDB Journal, Vol. 3, No 4, pp. 517–542, 1994.
8. D. A. White and R. Jain, "Similarity Indexing: Algorithms and Performance," Proc. SPIE vol. 2670, San Diego, USA, pp. 62–73, Jan 1996.
9. D. A. White and R. Jain, "Similarity Indexing with the SS-tree," proc. Of the 12th Int. Conf. on Data Engineering, New Orleans, USA, pp. 516–523, Feb. 1996.
10. S. Berchtold, D. A. Keim and H. P. Kriegel, "The X Tree: An Index Structure for High Dimensional Data," Proc. Of the 22nd VLDB Conf., Bombay, India, pp. 28–39, Sep 1996.
11. Norio Katayama and Shin'ichi Satoh, "The SR-Tree: An Index Structure for High-Dimensional Nearest Neighbor Queries" Proc. ACM SIGMOD Int. Conf. on Management of Data, pp. 13–15, May 1997.
12. Wenqing Huang, Jiazhe Dai and Qiang Wu, Image Retrieval Based on Weighted Block Color Histogram and Texton Co-occurrence Matrix, International Conference on Information Technology and Management Innovation, 848–853 (2015).
13. Jun Yue, Zhenbo Li, Lu Liu, Zetian Fu, "Content based image retrieval using color and texture fused features," Mathematical and Computer Modelling 54 (2011) 1121–1127.
14. Seong-O Shim, Tae-Sun Choi, Senior Member, IEEE, "Image indexing by modified color co-occurrence matrix," IEEE Xplorer.
15. N. Roussopoulos, S. Kelley, and F. Vincent, "Nearest Neighbor Queries," Proc. AChl SIGMOD, San Jose, USA. pp. 71–79, May 1995.

A Novel Approach to Mapping for KBQA System Using Ontology

R. Mervin and A. Jaya

Abstract Knowledge Based Question Answering System (KBQA) aims to provide suitable answers for the queries posted by users using ontology. Ontology is supposed to refine the search with semantics and confer the best result to the user. Mapping of Question and Answers are mandatory in every Question Answering System. This research work proposed the framework for KBQA Mapping system, which helps in retrieving precise information from a large collection of documents. The KBQA system consists of three phases such as Question Preprocessing, Answer Evaluation and Concept Mapping of question with their answer. The question posted by the user will be evaluated to get the answer and the answer will be taken from the pool of KBQA. This research paper explores the novel approach to the mapping of Question and most relevant Answers in KBQA System.

Keywords Question answering system · Information retrieval · Natural language processing · Knowledge based question answering system Ontology · Question preprocessing · Concept mapping and ranking

1 Introduction

E-Learning is now becoming a booming technology that can confer many information resources to the end user. There are many search engines available to provide tremendous information to the user. But the user has to navigate many pages to find their relevant information. Now a day's information retrieval becomes

R. Mervin (✉)
Department of Computer Science and Engineering,
B.S. Abdur Rahman University, Chennai, Tamil Nadu, India
e-mail: mervinvijay@gmail.com

A. Jaya
Department of Computer Applications, B.S. Abdur Rahman University,
Chennai, Tamil Nadu, India
e-mail: jayavenkat2007@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_9

easiest task with the help of Question Answering System. It is a specialized form of Information Retrieval which seeks knowledge. There are two types of Question answering System such as Open Domain QA System and Closed Domain QA System. Open domain QA System deals with questions about anything, and can only rely on general ontology where as closed domain QA system deals with questions under a specific domain. KBQA system plays major role to give exact answers rather than providing irrelevant document during search. The major challenge in knowledge based question answering system development is building a huge knowledge base with objective and correct factual knowledge in the preferred domain [3]. The process of collecting useful knowledge from various sources and maintaining this information in a knowledge repository is a useful process when providing a required answer on demand with greater accuracy and efficiency.

The user query posted by the user has to be analysed both semantically and syntactically. Ontology provides semantic meaning of concepts. It also defines the relationship between the concepts. The technique of visualising these relationships with different kind of relationship called mapping. In Question Answering System every user's question has been analysed syntactically and semantically. Then the question is mapped with the answer using concept mapping. Concept mapping is the way of representing the relationship between concepts and help to develop the logical thinking.

The remaining part of the paper is organized as follows, Sect. 2 deals Literature Survey, Sect. 3 describes Framework of KBQA Mapping System and Sect. 4 gives the experimental Results. Finally Sect. 5 concludes the paper by giving a brief glimpses into the future directions of research in this area.

2 Literature Survey

Atapattu et al. [1] utilized the framework for extracting question-specific concept maps (QSCMap) and proposed a method that return answers as a concept map, which further encourages meaningful learning and knowledge organisation. Here the Question type will be converted to triple form and they have taken Descriptive and Comparison type of question to analyse. Triple Enhancer is used to find the synonyms of word to find the keyword from the domain model. Map merge method is used to automatically combines the concept maps of related topics by reading the Concept Map Extensible Language (CXL) [10] file.

Khune [2] offered a mapping framework for Multiagent ontology which is having heterogeneous data in web and assigned the semantic relation using semantic relation interpreter which improves response time. Multiagent system mapping system was developed to map the question with their answer.

Suresh Kumar [3] was proposed a methodology to extract the concept relation from the unstructured text using Navie Bayes Classifier. Hand coded dependency parsing pattern rules and binary decision tree based rules are generated to accomplish the concept mapping to the seed concept. The triple set is used to pull out the feature of data which provides the key for question and answer mapping.

Dang Truong Son [4] was proposed an experimental QA system for Vietnamese language which is converted the given question by expanding the sentence with synonyms and trimming the list of extended sentences by language modelling, thereby moving the user's question to suit the question-answer database. Part of Speech, TF-IDF methods are used to choose the most relevant answer.

Moussa [5] was developed QASYO which is a sentence level question-answering system that integrates natural language processing, ontologies and information retrieval technologies in a unified framework. They have used YAGO which is a new light-weight ontology. It uses WR table and Verb table to find out the relationship between the candidate combinations. The various techniques and methodology of question answering system and their mapping system was discussed in [6–9] by various author. These methods were studied and analysed and their drawbacks were identified.

3 Architecture Diagram

The Fig. 1 shows the Framework of KBQA Mapping System which consists of three parts such as Question Preprocessing, Answer evaluation and Concept Mapping. The Question Preprocessing step consists of Question classification, Question Analysis and Keyword Extraction. The query/question from the user will be taken as the input for question pre processing and has to be analysed based on the syntactic and semantic analyser. From the information the keyword has to be extracted to find the relevant answer which is used to map with answer from the

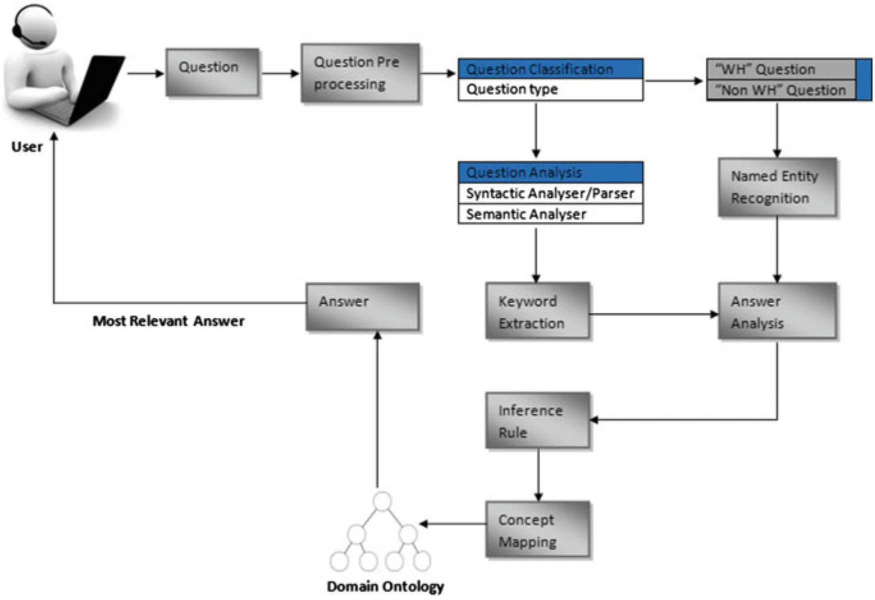


Fig. 1 Framework of KBQA mapping system

ontology. After applying the reasoning with all the answers, the most relevant answer will be given to user.

3.1 Question Preprocessing

The system utilizes, a Parts of Speech (POS) tagger which is identifies the question words (why, what, how, where, when, which) and the keywords from the user query. For example the question "Who is the father of Computer?" will be defined as follows:

Who is the father of Computer?

Who- ProNoun (WP)
Is- Verb (VBZ)
The- Determiner (DT)
Father- Noun (NN)
Of- Preposition (IN)
Computer- Noun (NN)

Once the POS tagger process gets completed, by using the question words, the parse tree has to be built with the help of Stanford Parser to understand the concept easily. Example of parse tree for question is shown in the Fig. 2

[S [WP Who] [VP [V is]] [DT the] [^NN father Computer] [IN of]]

Using the POS tagger and the Named Entity recognizer to narrow down the search of the keyword. For example, a query with the word who and when could point to person and a date. Now the next step is to find the exact field to match in-order to extract the answer that the user seeks. When question type has been successfully mapped to a top concept, only terms related to this concept will be added to the term context representation. This way we obtain the terms that made up the context of a unique definition term. The following list of words which are used to identify the answer type for the question posted by the user.

1. who/whom—PERSON
2. when—TIME/DATE
3. where/what place—LOCATION
4. what time (of day)—TIME
5. what day (of the week)—DAY
6. what/which month—MONTH
7. what age/how old—AGE

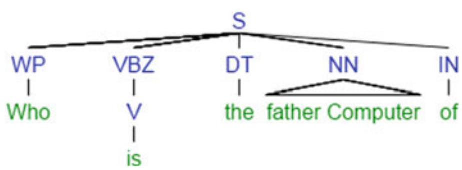


Fig. 2 Parse tree for question

8. what brand—PRODUCT
9. what—NAME
10. how far/tall/high—LENGTH
11. how large/big/small—AREA
12. how heavy—WEIGHT
13. how rich—MONEY
14. how often—FREQUENCY
15. how many—NUMBER
16. how long—LENGTH/DURATION

3.2 *Concept Mapping of Tokens*

Concept Mapping is used to represent the relationship between concepts. With the help of Ontology Construction the Domain, Subjects and their relationship is defined. The Fig. 3 shows the Domain Ontology Construction for the Question Type. The keyword which is driven from the user question has to be mapped with the help of concept mapping.

```
//Algorithm for Ontology Construction
Procedure OntologyConsturction()
Begin
Let Q= {q1,q2,q3,...qn} be the generalised Domain of Question
A = {a1,a2,a3, ...an } be the Answer
r be the Relationship between Q and A
Step 1: Initialise the Domain Q
Step 2: Identify the Subject-Predicate-Object of Q and A
Step 3: Define the Class, Individuals, Object and Data Property
Step 4: Assign the values using Object and Data Property.
Step 5: Call ConceptMapping
End
```

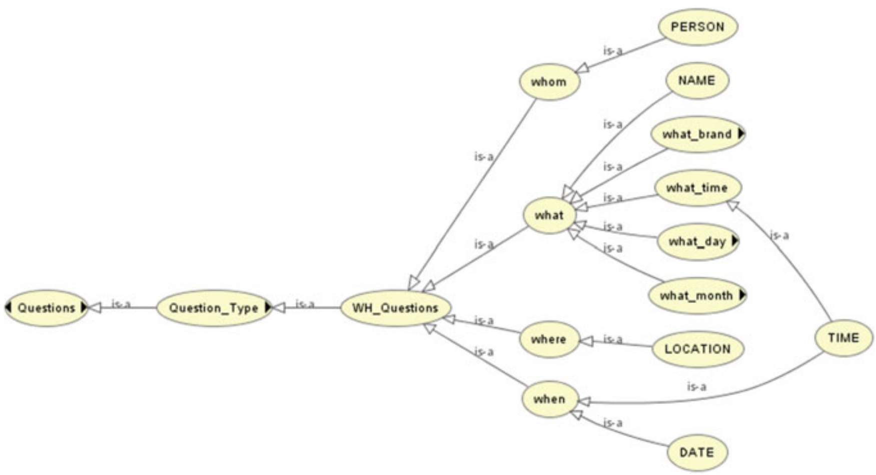


Fig. 3 Domain ontology construction

```
//Algorithm for Concept Mapping
Procedure ConceptMapping
Begin
Step 1: Get the Input Query from the User
Step 2: Find the Keyword from the User
Step 3: If (the Keyword exists) Then
Step 4:         If A ⊆ Q Then
Step 5:                 r = A ∩ Q.
Step 6:                 Check the Ontology for
matching the Keyword.
Step 7:         End if
Step 8: Else
Step 9:         Return
Step 10: Endif
End
```

Figure 4 shows the mapping Question with the Answers. In KBQA, there is need to map the questions with the exact answer. The user may confer many answers for the particular question. The mapping is done with the help of LP-LSA.

Initially the keyword has to taken from the search word. The stop word which is used commonly has to be removed from the search word. Some of the stop words are {a, an, has, is, are, and, for, to, the}.

Let A be the set of Answers $A = \{a_1, a_2, a_3, \ldots a_n\}$, Q be the set of Questions $Q = \{q_1, q_2, q_3, \ldots q_n\}$ and K is the set of Keywords $K = \{k_1, k_2, k_3, \ldots k_n\}$ Suppose that $q = f(a)$ is a function and $k > 1$; then $q = kf(a)$ is obtained by mapping the question and their the answers with the help of keyword k. Now the keyword has to be searched in the given Document. Based on the keyword in each answer the count matrix has to be created. Index is treated as row and the title in the search is treated as column. Now the weight calculation has to be done with the count matrix which is worn to modify the count. TF-IDF is used for this purpose.

$$\begin{aligned} \mathrm{TF(k)} = {} & (\text{Number of times keyword 'k' appears in a Answer})/ \\ & (\text{Total number of terms in the Answers}). \end{aligned} \tag{1}$$

$$\mathrm{IDF(k)} = \log_e(\text{Total number of Answers}/\text{Number of Answers with keyword 'k' in it}) \tag{2}$$

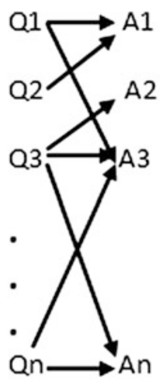


Fig. 4 Mapping of question with answers

From Eqs. (1) and (2), *TF* * *IDF* can be calculated as

$$TF * IDF(keyword) = \mathrm{TF(k)} * \mathrm{IDF(k)} \tag{3}$$

where *TF(term)* is the frequency of a term in the given document, IDF is Inverse Document Frequency. Then the co-occurrence table of counts has to be calculated $C = (n(a_{i,}k_j))_{ij}$ using the keyword K. Further assume that with each observation $(a_{i,}k_j)$ an unobserved class variable $x_y \in \{x_1, \ldots x_y\}$ is associated. A class x_y can be regarded as a multiple answers for a single question.

4 Experimental Result

The Framework of KBQA Mapping System is implemented using Java as front end and MySQL as backend. The domain ontology is constructed with the help of Protege. Figure 5 shows the ontology construction for keywords. After the

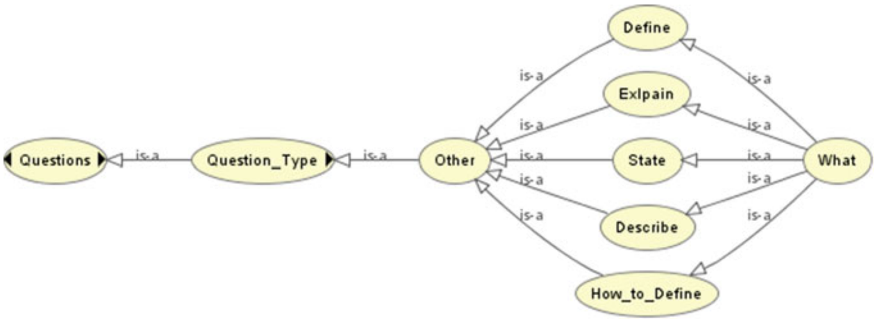
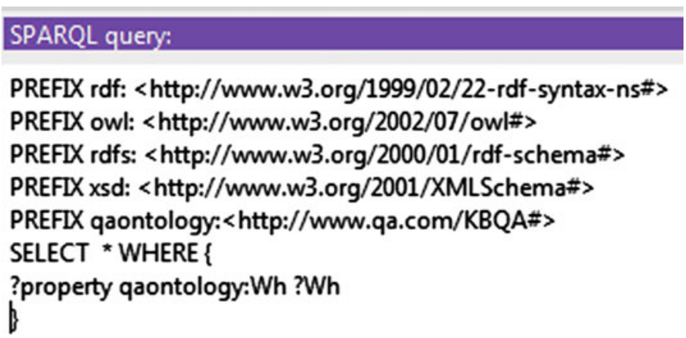


Fig. 5 Keyword ontology


SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX qaontology:<http://www.qa.com/KBQA#>
SELECT  * WHERE {
?property qaontology:Wh ?Wh
}
```

Fig. 6 SPARQL query

property	Wh
Whques	"What"@
where	"Location"@
What	"Person"@
Whques	"Where"@
Who	"Person"@
What	"Organisation"@
Whques	"When"@
Whques	"Who"@
When	"Time"@
What	"Name"@
When	"Date"@

Fig. 7 Query output

construction of the ontology SPARQL query was executed and the same query was implemented in Java. It is shown in Fig. 6 and the result is displayed in Fig. 7.

5 Conclusion

In this research the Framework of KBQA Mapping System is discussed and the experiments were done with 100 questions. The syntactic and semantic structure of the question given by the user has been analysed. Once the question is analysed the keyword, which was extracted to find the answers was taken for further analysis. The list of keywords time, person, location etc., is used to map with the answer to the user question. For 100 questions, there are 81 questions that the correct answer can be found in the top rank.

References

1. Thushari Atapattu, Katrina Falkner, and Nickolas Falkner "Educational Question Answering Motivated by Question-Specific Concept Maps" Springer International Publishing Switzerland, pp. 13–22, 2015.
2. Rupali R. Khune "Mapping of Semantic Web Ontology in User Query System" International Journal of Computer Applications (0975–8887) Volume 111 – No 14, February 2015.
3. G. Suresh Kumar and G. Zayaraj "Concept relation extraction using Naïve Bayes classifier for ontology-based question answering systems" Journal of King Saud University – Computer and Information Sciences, pp. 13–24, 2014.
4. Dang Truong Son, Dao Tien Dung "Apply A Mapping Question Approach In Building The Question Answering System For Vietnamese Language" Journal of Engineering Technology and Education and The 2012 International Conference on Green Technology and Sustainable Development (GTSD2012), pp. 380–386, 2012.
5. Abdullah M. Moussa and Rehab F. Abdel-Kader "QASYO: A Question Answering System for YAGO Ontology" International Journal of Database Theory and Application Vol. 4, No. 2, June, 2011.
6. A. H. Hafizullah, and A. Aniza, "The Spiral OWL Model? Towards Spiral Knowledge Engineering", World Knowledge Engineering, World Academy of Science, Engineering and Technology Vol: 38 0000-00-00, 2010.

7. Saravanan, Ravindran & Raman. “Improving legal information retrieval using an ontological framework.” Arti. Intell. Law, 2009, **17**(2), 101–24.
8. Chi & Chen. “Ontology and semantic rules in document dispatching”, The Electronic Library, 2009, **27**(4), 694–707.
9. Z. Li, V. Raskin, and K. Ramani, “A Methodology of Engineering Ontology Development for Information Retrieval”, International Conference on Engineering Design, ICED’07. Paris, France. 2007.
10. Canas, A.J., et al.: “CmapTools: a knowledge modeling and sharing environment”, First International Conference on Concept Mapping, Pamplona, 2004.

Author Biographies

R. Mervin Pursuing her Ph.D. in Computer Science and Engineering from B.S. Abdur Rahman University, India and she is working as an Associate Professor, with the department of Computer Science and Engineering in Saveetha Engineering College.

A. Jaya Professor in the Department of Computer Applications at B.S. Abdur Rahman University. Her research interest includes AI, CBR, Ontology, Web mining, Knowledge mining etc. She published more than 50 international journals.

An Adaptive Grouping Scheme for Avoiding Hidden Node Collision in IEEE 802.15.4 LR-WPAN

Mohammad Irfan Khan and Rakesh Rathi

Abstract IEEE 802.15.4 standard is proficient to execute with low-power for short distance in low-rate wireless personal area networks. For reducing the consumption of power in IEEE 802.15.4, slotted CSMA/CA protocol is used. However, the performance of LR-WPAN still degrades due to the hidden node problem (HNP). This paper proposed a simple and efficient grouping strategy to mitigate the HNP. The main concept is to gather the hidden information for every node, and then assign the nodes in the group on the basis of collected information. This scheme splits the superframe of IEEE 802.15.4 into numerous subperiod without overlapping each other. Simulation results shows that the given scheme not only mitigate the collision but also improves the transmission capabilities.

Keywords Superframe · Collision · Hidden Node Problem

1 Introduction

In twentieth century, wireless sensor networks (WSNs) have get attention from areas like industries and academics. With the evolution of sensor network technologies, nowadays WSNs are developed for different application domain like home automation, industrial automation, and health care. The existing applications rely on IEEE 802.15.4 specification to operate in low rate for short distance communication [1].

WSNs are comprised of several scattered nodes along a gateway which collects the information from nodes [2]. Moreover, sensor nodes are developed by different vendors. So there will be a lack of communication between them. IEEE 802.15.4 is the one candidate through which sensor nodes of different vendor can communicate.

M. Irfan Khan (✉) · R. Rathi
Department of Computer Science & Engineering,
Government Engineering College, Ajmer, Rajasthan, India
e-mail: irfankhanmic@gmail.com

R. Rathi
e-mail: rakeshrathi4@rediffmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_10

In WSNs, there will be a situation in which two different nodes that are invisible of each other can communicate with a common visible node at a particular time interval. So, in this situation, collision occurs and this collision is known as hidden node collision (HNC).

Busy tone mechanism [2] is the popular solution for avoiding the hidden node problem (HNP). This mechanism requires two channels: one is data channel and second is control channel. When the channel is idle, node can transmit data on data channel; if a channel is busy, receiving node send a busy tone on a control channel. The problem with that mechanism is that it requires an additional channel. So this mechanism is not considered to be the ideal solution for avoiding HNP.

Zigzag decoding [3] is another approach for mitigating the HNP. In this mechanism, a buffer is required on the coordinator side. Moreover, it also requires lots of processing to recover the collided data. Hence, it not considered as a good idea to address the issue of HNP.

Review of [4, 5] states that HNP is also avoid through grouping techniques. In these techniques, node forms the group on the basis of hidden relationship. When the new node wants to join the group, all the groups must be rearranged, hence, it requires lots of power for collecting the hidden information, when a new coming node wants to join the group.

Fast recovery [6] and H-NAMe [7] are other approaches for resolving the HNP. In fast recovery mechanism, nodes may turn off their transceiver for saving the energy. While remaining in the sleeping mode, there is a probability of missing some information. On the other hand, in H-NAMe, the process of gathering the hidden information could result in large overhead. Therefore these two approaches will not be adapted to resolve the HNP.

This paper proposed a simple grouping scheme based on polling process. The basic idea is that nodes in the same group could not be hidden with each other. In this process, coordinator sends a request message toward every nodes. Nodes that failed to acknowledge the request will be consider as hidden node, after that coordinator assign the groups for the nodes on the basis of collected information. The proposed scheme is described in four major phases: the first one is hidden node relationship, the second is grouping of nodes and the final phase is bandwidth allocation.

The remaining paper is organized in the following manner. Section 2 gives the overview of IEEE 802.15.4, Sect. 3 briefly explain the grouping strategy, Sect. 4 explains the simulation results and Sect. 5 concluded this paper.

2 IEEE 802.15.4 Overview

An IEEE 802.15.4 device globally operates at 2.4 GHz with data rate of 250 kb/s [8]. The transmission range lies in between 10 to 100 m. Devices under this standard should be categorized as full function device (FFD) and reduced function device (RFD). The first active FFD will become the coordinator. The Medium access control (MAC) of LR-WPANs also support two topologies, star and peer-to-peer topologies. In this article, we only discussed star topology in which communication is maintained between nodes and a central controller. IEEE 802.15.4 networks either operates in beacon-enabled mode or in non-beacon-enabled mode. In beacon-enabled mode, a central device should be considered as coordinator. Moreover, devices should transmit data by using superframe structure as shown in Fig. 1. The entire superframe is mainly partitioned into two parts that are active period and inactive period. The active period is further partitioned into contention access period (CAP) and contention free period (CFP). The entire superframe is bounded between beacon frames. Note that, CAP begins from the end of the beacon frame and stops at the start of CFP. Slotted CSMA/CA is applied in the CAP portion of the active period for medium access. However, for real time service, CFP is reserved. CFP is made up of several guaranteed time slots (GTS).

At most seven FFD may demand the coordinator to assign the GTS for real time service. Moreover, the entire active portion of superframe is partitioned into 16 equally sized time slot and GTS must occupy more than one time slot. The total duration of active period is determined by superframe duration (SD). The value of SD is controlled by system parameter that is superframe order (SO). While the time between two consecutive beacon is specified by beacon interval (BI), the value of BI is controlled by beacon order (BO) parameter.

SD and BI are calculated as follows

$$\begin{aligned} SD &= aBaseSuperFrameDuration \times 2^{SO} \\ &= aNumSuperFrameDuration \times 2^{SO} \\ &= 16 \times 60 \times 2^{S0}(symbols) \\ &= 960 \times 2^{SO}(symbols) \end{aligned}$$

$$\begin{aligned} BI &= aBaseSuperFrameDuration \times 2^{BO} \\ &= 960 \times 2^{BO}(symbols) \\ & 0 \leq SO \leq BO \leq 14 \end{aligned}$$

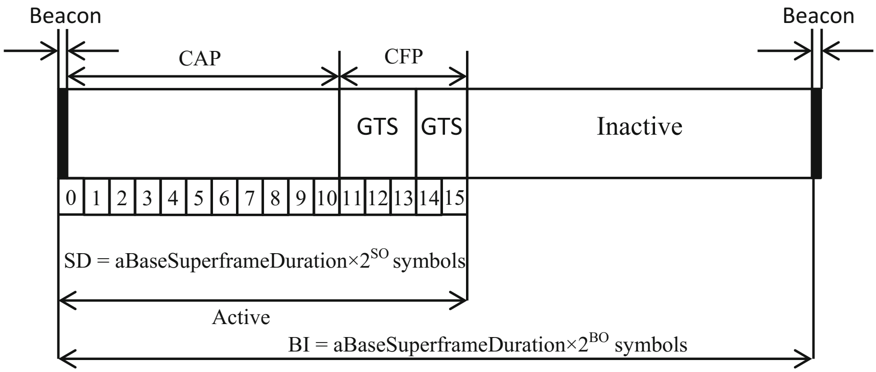


Fig. 1 Superframe structure of IEEE 802.15.4

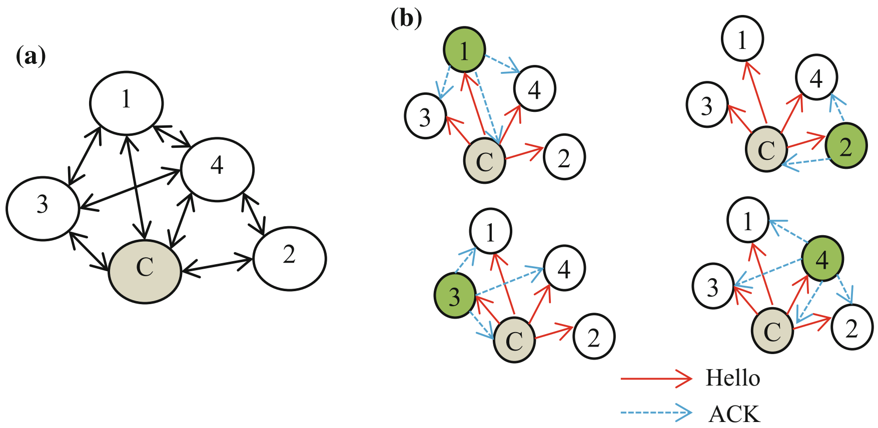
The BI may optionally include inactive period, nodes can turn off their transceiver for saving the power.

3 Grouping Scheme

The given scheme fundamentally consist of three phases: collection of hidden node information, group engagement, and group access period notification

A. Hidden Node Information Collection

Once the situation of hidden node is identified by the coordinator, it is ready to gather the hidden information among nodes; we consider four nodes and a coordinator as an example in Fig. 2a. In this figure, bidirectional link between nodes indicate that information is exchanged between them. In the given topology, hidden node situation still exists, as node 2 is not directly connected with nodes 1 and 3. Now coordinator broadcast a request message for all the nodes. Note that request



(c)

Node	Hidden nodes
1	2
2	1, 3
3	2
4	No

(d)

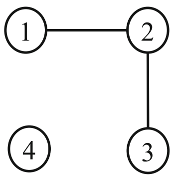


Fig. 2 Procedure of collecting hidden node information. **a** Topology, **b** Polling process, **c** Record of Hidden node, **d** Graph of hidden node

message is attached with beacon frame, to order all the nodes in the ingoing superframe to become active, and waits for the data frame with empty payload and this data frame is known as polling frame shown in Fig. 2b. Those nodes that do not acknowledge the polling frame are considered to be the hidden node by coordinator.

After the successful completion of polling process, all the nodes contain the information of their hidden nodes. Now coordinator broadcast a reporting message for all the nodes to return their hidden information with the help of polling approach. After receiving the hidden information, coordinator builds a hidden node graph. Figure 2d shows the hidden node graph

B. The Node Grouping

Once the construction of hidden node graph is completed, coordinator allocates the nodes in the group on the basis of collected hidden information. Note that directly connected node in the graph could not be resided in the same group. For better utilization of channel, grouping scheme must full fill two conditions: (1) number of nodes is balanced among groups and (2) to cover the entire network, minimal number of group must be used.

The solution of reducing the group looks like Hamiltonian subgraph problem [9]. It has been prove as NP-complete problem. In this article, we propose a simple algorithm to balance the nodes among groups.

Assume the set of nodes be *V*, except the coordinator in LR-WPAN. The given algorithm starts picking a node with largest degree from set *V* and form the group. The picked node is erased from set *V*. Now the node with second highest degree is examined from set *V* whether having an edge toward the node of that already formed the group or not. If not, it can be assign in the formed group, else algorithm marked this node and passes away. Similarly, all the remaining nodes in the topology are examined on the basis of their degree. Once all the node in set *V* is marked, given algorithm starts assigning the second group with procedure that already described. This process continues until the set *V* will become empty. The grouping algorithm is described below.

Grouping Algorithm [5]

Input: Given a hidden node graph $G = (V,E)$;

Output: Group sets $S_1, S_2, \ldots, S_g$; $(g \leq k)$

$k = 1$; // group index

While $|V| > 0$ **do**

Select node v from V of the highest degree;

Build group set $S_k = \{v\}$;

$V = V - \{v\}$

build temporary set $T = V$;

while $|T| > 0$ **do**

select a node t from T of largest degree;

$T = T - \{t\}$;

If node t has no edge towards any node in S_k **then**

$S_k = S_k + \{t\}$; // join to group k

$V = V - \{t\}$;

end if

end

k++;

end;

C. Grouping and Bandwidth Allocation Notification

The given grouping scheme modified the superframe structure and the entire active period is partitioned into three parts: contention access period (CAP), group access period (GAP), and contention free period (CFP) as shown in Fig. 3.

It is already described in IEEE 802.15.4 standard that superframe must reserve 22 UBPs for CAP. This minimal period is used by a fresh device to send the management frame for tie up. GAP is partitioned into a number of subperiod by

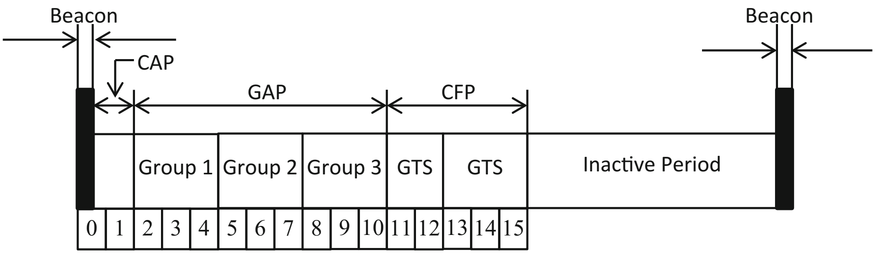


Fig. 3 Modified superframe structure

Octets: 2	1	4/10	2	Variable	Variable	Variable	Variable	2
Frame control	Sequence number	Addressing fields	Superframe Specification	GTS Fields	GIF Field	Pending Address fields	Beacon payload	FCS

Fig. 4 Format of beacon frame

Table 1 Simulation parameters

Parameters	Values
Nodes	1 coordinator, 20 static nodes
BO = SO	3
Frame length	10 & 20 bytes
Transmission range	15 m
Simulation time	300 s
CAP	22 UBPs (=440 symbols)

using slotted CSMA/CA. Note that, if a node wants to send a data and the corresponding subperiod gets expired, it must wait for the next period.

Once the bandwidth is assigned for each group, the GAP result is broadcast by coordinator through beacon frame along with an additional group information field (GIF) as shown in Fig. 4. The GIF field lies between pending address field and GTS field. The optional GIF field have information about group id along with group length. It also contain the information from where GAP slot is started and ends.

4 Experimental Results

In this section, we compare the proposed scheme with IEEE 802.15.4 standard. All the simulation related to experiment is performed on NS3 simulator. In the simulation model, we consider star topology with one coordinator and 20 static nodes. We also assume that medium is free from error and noise. The bandwidth of medium is B bytes/s. Frame arrival rate of every node should precede the Poisson distribution along a mean of λ (frames/s) and L is length of frame. Therefore, traffic load is calculated as $(N\times\lambda\times L)/B$. Table 1 shows the list of parameters that are taken in consideration while performing the simulation.

Figure 5 demonstrates the graph plot between traffic load and goodput (%). Goodput is the total access channel bandwidth measured in percentage. Figure 5a shows the experimental result calculated with respect to frame length of 10 bytes, while Fig. 5b shows the simulation result with respect to frame length of 20 bytes, both of these results are calculated with identical parameters shown in Table 1, except the frame length. These results show that grouping scheme shows the stable

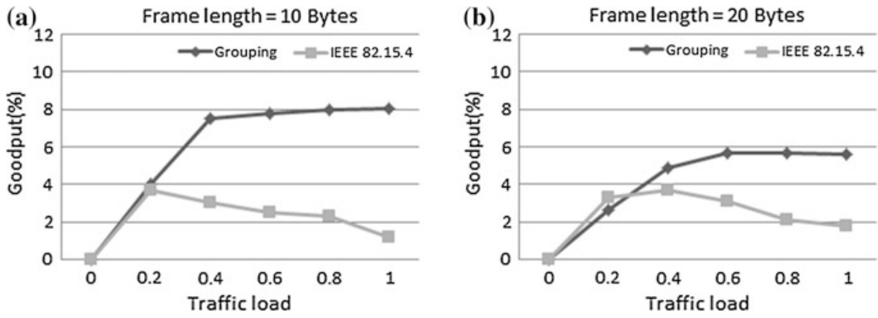


Fig. 5 Goodput versus traffic load **a** goodput when L = 10, **b** goodput when L = 20

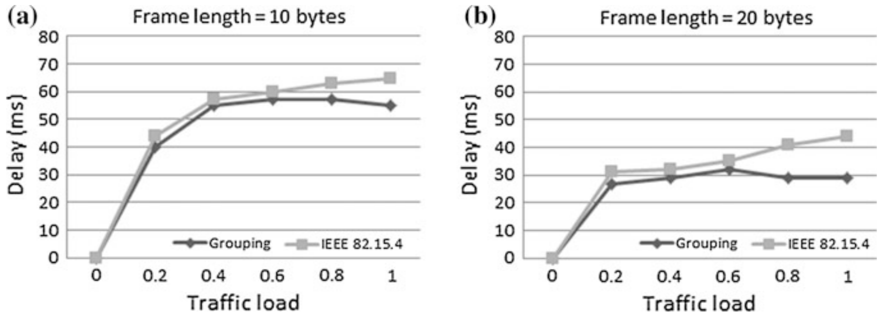


Fig. 6 Delay versus traffic load **a** Delay when L = 10, **b** Delay when L = 20

result in terms of goodput, while IEEE 802.15.4 performance degrades due to the contention.

Figure 6 shows the graph between delay and traffic load, here Fig. 6a shows simulation with respect to frame length of 10 bytes and Fig. 6b with respect to frame length of 20 bytes. These figures show that grouping scheme performed well under heavy traffic load as compared to IEEE 802.15.4. IEEE 802.15.4 performance degrades due to the repeated transmission of collided frame.

5 Conclusion

A simple grouping scheme has been introduced to mitigate the HNC. The given scheme makes use of simple grouping algorithm to construct the groups. The proposed strategy not only enhances the performance of IEEE 802.15.4 devices, but also ensures fare transmission period among them. Simulation outcome shows that given strategy improves the performance in terms of goodput and delay. Finally, we concluded that the proposed strategy enhances the performance of IEEE 802.15.4 protocol along with maintaining the Quality of Service of LR-WPANs.

References

1. Shiann-Tsong Sheu, Yun-Yen Shih and Wei-Tsong Lee, "CSMA/CF Protocol for IEEE 802.15.4 WPANs", *IEEE Trans.*, Mar. 2009.
2. Zygment J. Haas and Jing Deng, "Dual busy tone multiple access (dbtma)- a multiple access control scheme for ad hoc networks", *IEEE Trans*., June 2002.
3. Shyamnath Gollakota and Dina Katabi, "Zigzag Decoding: Combating Hidden Terminals in Wireless Networks", *ACM* 2008.
4. Lian-J. Hwang, Shiann-T. Sheu, Yun-Y. Shih and Yen-C. Cheng, "Grouping Strategy for Solving Hidden Node Problem in IEEE 802.15.4 LR-WPAN", *IEEE-WICON*, July 2005.
5. Hsueh -W. Tseng, Ai -C. Pang, Jenhui Cheng and Chin -F. Kuo, "An Adaptive Contention Control Strategy for IEEE 802.15.4-Based Wireless Sensor Networks", *IEEE Trans.* 2009.
6. Shengzhi Zhang and Sang -J. Yoo, "Fast Recovery from Hidden Node Collision for IEEE 802.15.4 LR-WPANs", *IEEE International Conference on Computer and Information Technology*, 2007.
7. Anis Koubaa, Ricardo Severino, Mario Alves and Eduardo Tovar, "Improving Quality-of-Service in Wireless Sensor Networks by Mitigating "Hidden Node Collision", *IEEE Trans.*, Aug. 2009.
8. Ju-Hyun Lee, Jae-Kark Choi and Sang-Jo Yoo, "Group Node Contention Algorithm for Avoiding Continuous Collision in LR-WPAN", *IEEE International Conference on Computer and Information Technology*, 2009.
9. M. R. Garey, D. S. Johnson, G. L. Miller and C. H. Papadimitriou, "The Complexity of Coloring Circular Arcs and Chords", *Society for Industrial and Applied Mathematics, ALG. DISC. MATH.*, June 1980.

An Approach to Classify Engineering Materials Using Machine Learning Algorithm

P.J. Antony, Prajna Manujesh and N.A. Jnanesh

Abstract This review paper explores the attempts made by the numerous authors in the field of material selection. There are ample amounts of works carried out in the field of materials engineering with data mining approaches. From the literature it is revealed that not much of the work is explored on the classification of advanced composite materials using machine learning approaches.

Keywords Material selection · Data mining approach · Advanced composite materials · Machine learning

1 Introduction

Material science is an interdisciplinary field for applying the properties of materials in various fields of building. The extent of the materials in designing is boundless. Each new item is either in light of new material configuration or change in the current material properties. The quick improvements in the field of materials science with due significance to materials building is the present days need [1]. Therefore, the materials designer ought to have careful information for the material determination to the proposed applications in enhancing new items. The best item is the result of adjusted materials and ideal outline developments. Along these lines the determination of best materials spins around the material choice systems. Expectedly the material determination is trailed by examining the material information set and handling consecutively to channel, arrange lastly the learning removed from information set can be utilized for different configuration applications [2]. Along these lines the choice and characterization of materials are vital and basic in the active field of material designing.

P.J. Antony · P. Manujesh (✉) · N.A. Jnanesh
Department of Computer Science and Engineering,
KVG College of Engineering, Sullia 574327, India
e-mail: prajnamanu@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_11

A. *Composite Materials*

Composite materials are an insignificant blend of two or more distinctive materials. One typically associates being the consistent stage (network) and the other being the broken stage (fortification). It is the flexibility of pattern consolidated with solidness of strengthening. The properties of these materials are unequivocally subject to the composite structure and these materials are broadly acknowledged materials in view of their versatility to various circumstances and the relative straightforwardness of mixing with different materials to show attractive properties [3]. The use of composites keeps on developing at amazing rates as these materials are utilized more as a part of their current markets and to end up built up in generally new markets, for example, biomedical and common structures [4]. The composites are customized to suit the particular applications making a more noteworthy favorable circumstances like; low warm conductivity and low coefficient of warm extension, high hub quality and solidness, and so forth. A key variable driving the expanded utilizations of composites over the late years is the improvement of new propelled types of cutting edge materials. This incorporates the improvement in elite sap framework and new styles of fortifications [5].

The utilization of composites has turned out to be progressively alluring distinct option for the ordinary materials for some building applications. The segments can be produced essentially because of their expanded quality, strength, consumption resistance, imperviousness to exhaustion and harm resilience attributes likewise they guarantee huge chances to assume an expanding part as interchange material to replace timber, steel, aluminum, and cement in structures even in high weight and forceful ecological circumstances [6]. Uses of composite are expanding colossally alongside the simultaneous requirement for information era. These are utilized as a part of practically every sort of designing structures, with their use going from airplanes, rocket through to water crafts, ships and seaward stages, autos, brandishing merchandise, compound preparing supplies, and common foundation, for example, spans and buildings [7, 8]. With the innovation advancements and improvements in procedures and items, the composites have ended up appealing contender for generally applications. The data to pick certain materials needs exhaustive comprehension and learning for choice parameters to fit into basic applications. Consequently determination of materials is the principal basic stages were accomplishment on materials decision is managed.

B. *Machine Learning*

Machine learning is a way to deal with configuration of a particular model that permits PCs to pick up information without being expressly modified. The material disclosure includes colossal measurable abnormality in the example of information dissemination. As needs be, various machine learning calculations will hardly take after how human might approach a learning assignment [9]. The regulated learning is the one class of machine realizing, where the class yield is refer to for all conceivable preparation and in addition test information. Further it is genuinely normal in order issues on the grounds that the goals are over and again guide the PC to take in an arrangement framework. For the most part, grouping learning is suitable for any issue in the connection where the class of obscure information will be anticipated utilizing the model from now on assembled. Sometimes, it would not be important to give set of groupings to each occurrence of the issues; rather the specialists can work out the arrangements for it and this would show as an unsupervised learning in any order connection [10, 11]. They guarantee huge advances within the materials science, and hold the broad surety for materials exploration and information revelation in defying with another material. Along these lines it would be helpful if its properties could be anticipated utilizing the past learning relating to other comparative known materials, as opposed to turning new trials or difficult computations expending time and cost [12].

2 Literature Survey

There are generous amount of work done using various algorithms of machine learning like genetic algorithms, neural network approach, support vector machines. Recently the materials scientists have begin the data mining ideas for discovering new materials for diverse applications of engineering.

A. *Genetic Algorithms Approach for the Design and Optimization*

Genetic algorithm (GA) is an inquiry heuristic that mirrors the procedure of natural selection. This heuristic is routinely used to produce valuable answers for optimization and inquiry issues. The subsequent section highlights the earlier work done by the various authors using genetic algorithm on material engineering.

Mc Crory et al. [13] observed the damage classifications in carbon fiber composites using acoustic emission method. The authors used ANN (Artificial Neural Network), UWC (Unsupervised Waveform Clustering), and MAR (Measured Amplitude Ratio) techniques, to distinguish different signal types arising in a carbon fiber panel subjected to buckling. And they lead to the characterization of the damage occurring within the panel. Further, Leo Marco et al. [14] have proposed a new and innovative data preprocessing technique that converts real-valued ultrasonic data into complex valued signals. However, Dharmadhikari sagar et al. [15] reviewed on the optimization of drive shaft using genetic algorithm and ANSYS. Composite materials along with conventional steel material for drive shaft can increase the advantages of design due to its high specific stiffness and strength. For identifying the elastic constants of composite laminates by using vibration test data, Maletta and Pagnotta [16] revealed a method which combines finite element analysis with genetic algorithm. Furthermore Badallo et al. [17] presented a comparative study on three common genetic algorithms namely; Archive-based Micro Genetic Algorithm (AMGA), Neighborhood Cultivation Genetic Algorithm (NCGA), and Non-dominate Sorting Genetic Algorithm II (NSGA-II) considering three different strategies for the initial population. The authors [17] studied with the objective of minimizing mass and maximizing the critical buckling load on a ‘T’ shaped stringers using CFRP stiffened panels.

In short, genetic algorithms are very easy to understand and virtually it does not demand any mathematical information. Especially the algorithms are better for optimization problems and also it solves the problems with multiple solutions.

B. *Machine Learning on Materials data*

During trade-off situations it would be worthwhile if its properties could be anticipated by utilizing past information relating to other comparable known materials, rather than falling back on new investigations on the other laboratory experiments. The biggest difficulty in creating such a prediction machine is the consistent definition of a material unique mark [12].

Authors Liu Ruoqian et al. [18] used machine learning approach to identify the complete space (or as much of it as possible) of microstructures that are theoretically predicted to yield the preferred mixture of properties demanded by a selected purpose. In addition Maree et al. [19] have investigated a generic machine-learning approach for the detection of known materials in using hyper spectral images. It was applied for the detection of simulated gaseous traces in thermal infrared hyperspectral images of real-world scenes. In spite of these findings, Garcia Angela et al. [20] have found the exact of scientific methods like CART and neural network. The work also focused with specimens 30% compressive anxiety strain with momentary velocity impacts to anticipate Young’s modulus. Pilania Ghanshyam et al. [21] have showed that the materials discovery process can be significantly expedited and simplified if we can learn effectively from the available information or data. Similarly Liu et al. [22] have explored the multiple data mining experiments and strategies for establishing statistical models for capturing elastic localization relationships in high contrast composites. The efficacy of different approaches for feature selection and regression were studied systematically. In short the machine learning can be used for multidisciplinary areas for the sophisticated pattern recognitions. Thus the approach can be effectively used to perform intelligent decisions in building a model for predicting unknown data.

C. *Data Mining and Knowledge Discovery on Materials Data*

The usage of information mining systems out of sight of materials science and designing is seen as a fundamental development of materials informatics. This interdisciplinary study organizes data science, software engineering, and different spaces to give more current comprehension and empower learning disclosure. The materials informatics is a stage for material scientists to decipher exploratory data through the machine learning approaches. The work composed with new representations, arrangements, and more probable human consolidations driven by the space specialists. It can similarly accelerate the assessment process and guide the improvement of new materials with different engineering properties [23]. The Data mining and knowledge discovery techniques were employed to validate their usefulness in acquiring information about the viscoelastic properties of vapor-grown carbon nanofiber (VGCNF)/vinyl ester (VE) nano composites solely from the data derived from an investigational design studied by Abuomar Osama et al. [23]. To classify microstructures into groups automatically with high accuracy DeCost et al. [24] used support vector machines effectively. The authors claim the method best suits for finding the associations with large and dissimilar micro structural image data by comparing the histograms of the visual techniques.

In addition to the above paper the similarity based engineering materials selection model was proposed and implemented by Vanajakshi [25]. The work aims to select engineering materials based on the composite materials constraints. The result reviewed from this model was sustainable for the effective decision making in advanced engineering materials design and applications. However the authors Sharma and Krishna [26] have succeeded in developing a knowledge discovery system for the selection of cost effective polymer composites for a cylindrical pipe with longitudinal filament winding in the course of data mining approach. In addition Ashby et al. [27] found a novel material selection approach using software assisted tools. It contains a database of quantitative and subjective information for an extensive variety of engineering materials: metals, polymers, composites, and regular materials. Doan et al. [28] used an unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer-composites. The work shows the different accessible challenges of AE analysis and damage detection in composites.

The Knowledge Discovery System through Data Mining procedures were being actualized by Doreswamy et al. [29] for finding the knowledge for practical polymer composites that meets the configuration necessities of a segment. The methodology utilizes the Decision Tree Classifier (DTC) with linear function. The work characterizes the fiber class into short, medium, and long fiber classes identified with the basic length of the fiber ‘l’c. A suitable test quality all things considered volume fractions, volumes, masses, expenses of long strands and of polymer network are examined through the composite principle of blend. The framework uncovered valuable information’s for the outline of designing materials. The work contains a group of modest bunch figurings for low cost polymer composite determination [26, 30].

Naïve Bayesian and C4.5 decision tree classifiers approaches used by Addina et al. [31], the work aims in classifying and selection of materials to suit some design specifications. The authors also highlight the predictive parameters and standard measures for the classification on the different class of materials. The data mining techniques were effectively implemented by Suh and Rajan [32] for the novel data integration among homogeneous materials databases. The Rajan also [33, 34] reports the task of selecting the right material for a given engineering design can appear to be overwhelming. The designers can fetch numerous materials for commercial applications. But the approach to sort the material with optimum selection requires a systematic approach based on understanding of the nature of materials science and engineering.

The above literature summarizes the data mining techniques are excellent means to predict future trends and helps in decision making in the field of materials informatics.

D. *Work Done on Materials Using Neural Network*

Artificial neural systems (ANNs) comprises set of exhibiting methodology, which can be used to endeavor issues that are troublesome for standard computers and human creatures. The ANNs has been associated with showcase troublesome strategies in various outline fields like, aviation, car, gadgets, assembling, mechanical technology, and information transfers, and so on. Over the late years the enthusiasm for the ANN exhibiting in the fields of physical metallurgy and materials science has been extended rapidly. Thus the neural systems assists critical focal points in taking care of issues that require constant encoding and translation of associations among the variables of high-dimensional space. The ANN's also offers very basic level distinctive way to deal with material demonstrating and material handling, control systems than the quantifiable systems [35].

Liu Ce et al. [36] have proposed Augmented Latent Dirichlet Allocation (ALDA) model to combine the rich set of low and mid level features captured through various aspects of material appearance under a Bayesian generative framework. The authors presented a model to identify the high level categories of materials like glass, metal, plastic, or wood, as a substitute of clearly estimating reflectance properties. Lee et al. [37] predicted the fatigue damage behavior using ANN approach. The authors report the possibility of accurate representations of the crack growth and cycle ratio out of very small experimental data. Thus the authors conclude ANN as a better approach for life estimation parameters. Dikbas et al. [38] succeeds in using ANN approach for the prediction of diffusion bonding behavior of SiCP reinforced aluminum alloy composites. Thereafter António et al. [39] proposed an artificial neural network (ANN) aims at modeling of machining circumstances in orthogonal cutting of PEEK composites. The design of experiments (DOE) approach for Polypropylene and waste ground rubber tire powder composites reported by Zhang Shu Ling et al. [40]. The approach predicts the effect of four polymer contents on the mechanical properties.

A novel way to deal with damage discovery in composite structures utilizing hyper spectral image index analysis algorithm with neural network modeling employing Weight Elimination Algorithm (WEA) was presented and discussed by Iskandarani and Mahmoud [41]. Likewise Kumar Sanjeev et al. [42] used ANN coupled with Taguchi approach for the optimization and prediction of surface roughness. The authors observed good agreement between experimental and predicted results. Meanwhile Altinkok and Necat [35] used back-propagation (BP) neural network for modeling of metal matrix composite and built a model for artificial neural network on the prediction of tensile strength, hardening behavior, and density properties of the Al_2O_3 particulate-reinforced Al–Si10 Mg composites. Incidentally Atuanya et al. [43] have predicted the mechanical properties of date palm wood fiber-recycled low density polyethylene composite using ANN technique.

In summary the neural network permits solutions to materials problems where multiple selection constraints must be satisfied simultaneously.

E. *SVM Related*

SVM finds a nonlinear decision function in the input space by mapping the data into a higher dimensional feature and separating it there by means of a maximum margin hyper plane. The reason why SVM insists on finding the maximum margin hyper plane is that it offers the best generalization ability and it allows not only the best classification performance on the training data, but also does the correct classification of the future data. Figure 1 shows how to map a data from higher dimension to lower dimension [44].

Ding et al. [45] have shown that the SVM network can make out the difference of AE sources more perfectly than using the BP neural network. Abu Omar et al. [46] used the support vector machines model to identify the desired mechanical property response resulted from a chosen untested combination. The model was also able to identify the desired mechanical property response (high storage modulus, high true ultimate strength, or high flexural modulus) resulted from a chosen untested combination of nine input factors mentioned in their study. Furthermore Sundararaghavan et al. [47] has worked on 3D microstructure grouping structure has been produced in view of support vector machines implementing a proficient measurable learning methods. Meanwhile Das et al. [48] have used some class SVM's to classify the damage signatures in composite plates. Subsequently Wang et al. [49] designed a measuring system using decision tree learning for insulating material hydrophobicity. Tang et al. [50] applied Support Vector Machine to set up a nonlinear mapping from influence factors of material performances to mechanical properties. Besides Fauvel Mathieu et al. [51] have proposed a method based on the data fusion of the morphological information and the original hyperspectral data.

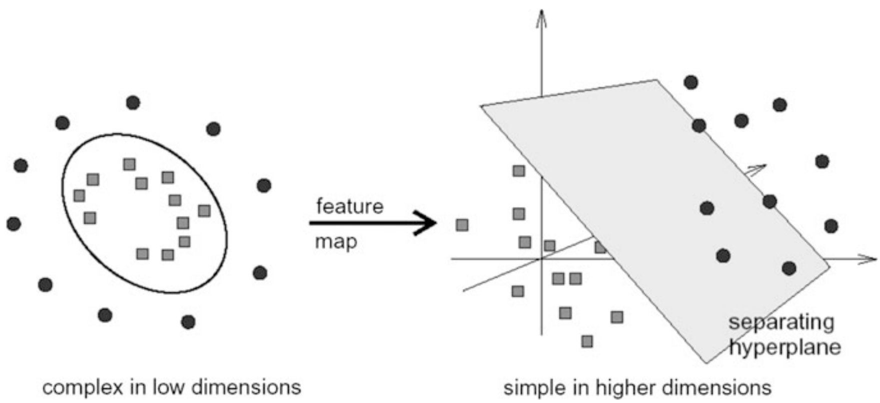


Fig. 1 Non-linear classifier

The authors claim the absolute classifications using a Support Vector Machines classifier. Finally it concludes that the support vector machine can be used effectively to yield the better results during selection of materials.

3 Objective

The main objective of the proposed research includes:

- To determine the association between data mining and machine learning systems to effectively discover the attributes that governs,specific properties of advanced composite materials from different corpora.
- To compare for associations and patterns, relating information among the different datasets of advanced composite materials, also to set up potential association between parameters that are not easily studied experimentally in a coupled manner.
- To get better material selection from the material classification process using any algorithm yields increased efficiency for optimizing materials processing techniques.

4 Methodology

Advanced composite material selection framework consists of three phases; foremost the Data acquisition in which pre processing of data will be done. The second phase continued with training of the filtered data that is known to specific classes and creating a classifier model based on the known data samples. Training is an iterative process where we need to build the best classifier model, in each iteration the built model is tested against the test data. Classification is the process of taking

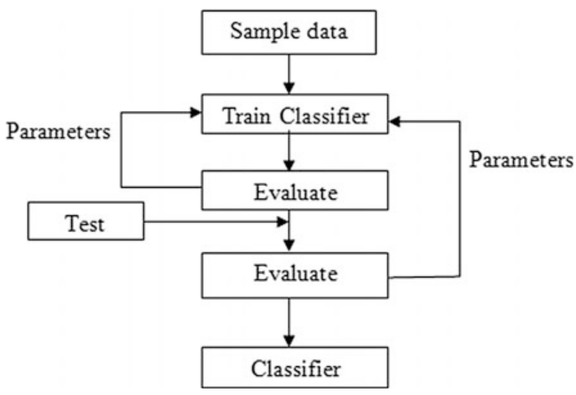


Fig. 2 Method of proposed approach

of classifier model built with a training data set and running it on the unknown test data to determine the class of data it belongs and finally to obtain the accurate classifier. Figure 2 shows the working model of the proposed approach.

5 Conclusion

The challenging task in materials science is the selection of materials from the material data set for a particular application in such a way that it should meet the design criterion. Through the published literature referred it is affirmative that no much work has been done on the composite materials extracting their properties. Earlier work explored does not contain feature classification based on complete properties like fiber type, effect of fiber orientation, fiber patters, fiber strength, mass density, specific strength, fire conflict, electrical properties, design flexibility and manufacturing economy, and degradation mechanisms, etc. Thus further work can be extended to obtain required classification of the advanced composite material systems taking wider array of physical properties into affect and also it reduces the repetitive manufacturing process for inventing new class of materials with enhanced physical attributes.

References

1. Caputo, Barbara, Eric Hayman, Mario Fritz, and Jan-Olof Eklundh. "Classifying materials in the real world." Image and Vision Computing 28, no. 1 (2010): 150–163.
2. Marsden, Will, and Stephen Warde. "Data Management for Composite Materials." (2010).
3. Campbell, F. C. "Structural composite materials: ASM international." ASM International, United States of America (2010): 50–100.
4. D. Cebon and M F Ashby, "Computer-Based Materials Selection for Mechanical Design" "Computerized and Networking of material Databases", ASTM STP 1140, Philadelhia, (1992).
5. Pilato, Louis A., and Michael J. Michno. Advanced composite materials. Springer Science & Business Media, 1994.
6. Balasubramaniam, R. Callister'S Materials Science and Engineering: Indian Adaptation (W/Cd). John Wiley & Sons, 2009.
7. Groza, Joanna R., and James F. Shackelford, eds. Materials processing handbook. CRC press, 2007.
8. Chawla, Krishan K. Composite materials: science and engineering. Springer Science & Business Media, 2012.
9. Ayodele, Taiwo Oladipupo. Types of machine learning algorithms. INTECH Open Access Publisher, 2010.
10. Domingos, Pedro. "A few useful things to know about machine learning." Communications of the ACM 55.10 (2012): 78–87.
11. Rob Schapire, Princeton University, "Machine learning algorithms for classification", http://www.cs.princeton.edu/courses/archive/fall04/cos557/Articles/schapire-slides.pdf
12. Mueller, T., A. G. Kusne, and R. Ramprasad. "Machine learning in materials science: Recent progress and emerging applications." Rev. Comput. Chem. (Accepted for publication) (2015).

13. McCrory, John P., Safaa Kh Al-Jumaili, Davide Crivelli, Matthew R. Pearson, Mark J. Eaton, Carol A. Featherston, Mario Guagliano, Karen M. Holford, and Rhys Pullin. “Damage classification in carbon fibre composites using acoustic emission: A comparison of three techniques.” Composites Part B: Engineering 68 (2015): 424–430.
14. Leo, Marco, David Looney, Tiziana D’Orazio, and Danilo P. Mandic. “Identification of defective areas in composite materials by bivariate EMD analysis of ultrasound”. Instrumentation and Measurement, IEEE Transactions on 61, no. 1 (2012): 221–232.
15. Dharmadhikari, Sagar R., Sachin G. Mahakalkar, Jayant P. Giri, and Nilesh D. Khutafale. ““Design and Analysis of Composite Drive Shaft using ANSYS and Genetic Algorithm” A Critical Review.” Int. J. Modern Eng. Res 3, no. 1 (2013): 490–496.
16. Maletta, C., and L. Pagnotta. “On the determination of mechanical properties of composite laminates using genetic algorithms.” International Journal of Mechanics and Materials in Design 1.2 (2004): 199–211.
17. Badalló, P., D. Trias, L. Marin, and J. A. Mayugo. “A comparative study of genetic algorithms for the multi-objective optimization of composite stringers under compression loads.” Composites Part B: Engineering 47 (2013): 130–136.
18. Liu, Ruoqian, Abhishek Kumar, Zhengzhang Chen, Ankit Agrawal, Veera Sundararaghavan, and Alok Choudhary. “A predictive machine learning approach for microstructure optimization and materials design.” Scientific reports 5 (2015).
19. Marée, Raphaël, Benjamin Stévens, Pierre Geurts, Yves Guern, and Philippe Mack. “A machine learning approach for material detection in hyperspectral images.” (2009): 106–111.
20. García, Ángela, Ofélia Anjos, Carla Iglesias, Helena Pereira, Javier Martínez, and Javier Taboada. “Prediction of mechanical strength of cork under compression using machine learning techniques.” Materials & Design 82 (2015): 304–311.
21. Pilania, Ghanshyam, Chenchen Wang, Xun Jiang, Sanguthevar Rajasekaran, and Ramamurthy Ramprasad. “Accelerating materials property predictions using machine learning.” Scientific reports 3 (2013).
22. Liu, Ruoqian, Yuksel C. Yabansu, Ankit Agrawal, Surya R. Kalidindi, and Alok N. Choudhary. “Machine learning approaches for elastic localization linkages in high-contrast composite materials.” Integrating Materials and Manufacturing Innovation 4, no. 1 (2015): 1–17.
23. Abuomar, Osama, Sasan Nouranian, Roger King, Jean-Luc Bouvard, Hossein Toghiani, Thomas E. Lacy, and Charles U. Pittman. “Data mining and knowledge discovery in materials science and engineering: A polymer nanocomposites case study.” Advanced Engineering Informatics 27, no. 4 (2013): 615–624.
24. DeCost, Brian L., and Elizabeth A. Holm. “A computer vision approach for automated analysis and classification of microstructural image data.” Computational Materials Science 110 (2015): 126–133.
25. M. N. Vanajakshi. “Similarity measuring approach for engineering materials selection.” International Journal of Computational Intelligence Systems 3.1 (2010): 115–122.
26. Sharma, S. C., and M. Krishna. “Knowledge Discovery System For Cost-Effective Composite Polymer Selection-Data Mining Approach.” (2005): 185–190.
27. Ashby, Michael F., and D. Cebon. “Materials selection in mechanical design.” Le Journal de Physique IV 3.C7 (1993): C7–1
28. Doan, Dinh Dong, Emmanuel Ramasso, Vincent Placet, S. Zhang, L. Boubakar, and N. Zerhouni. “An unsupervised pattern recognition approach for AE data originating from fatigue tests on polymer–composite materials.” Mechanical Systems and Signal Processing 64 (2015): 465–478.
29. Doreswamy, Hemanth KS. “Performance Evaluation of Predictive Classifiers for Knowledge Discovery from Engineering Materials Data Sets.” arXiv preprint arXiv:1209.2501 (2012).
30. Doreswamy, “Machine learning classification and selection model for engineering materials design applications”. Published in the proceedings of International Conference on Systemics: Cybernetics and Informatics, ICSCI pp. 148–150, India (2008).

31. Addina, O., S. M. Sapuanb, and M. Othmanc. “A naïve-bayes classifier and f-folds feature extraction mehtod for materials damage detection.” International Journal of Mechanical and Materials Engineering (IJMME) 2.1 (2007): 55–62.
32. Suh, C., and K. Rajan. “Invited review: Data mining and informatics for crystal chemistry: establishing measurement techniques for mapping structure–property relationships.” Materials Science and Technology 25.4 (2009): 466–471.
33. Krishna Rajan, “Materials Informatics and Data Mining for Materials Science”, a scientific report on September 24, 2008.
34. Rajan, Krishna. “Combinatorial materials sciences: Experimental strategies for accelerated knowledge discovery.” Annu. Rev. Mater. Res. 38 (2008): 299–322.
35. Altinkok, Necat. “Use of Artificial Neural Network for Prediction of Mechanical Properties of α-Al2O3 Particulate-reinforced Al–Si10 Mg Alloy Composites Prepared by using Stir Casting Process.” Journal of composite materials 40.9 (2006): 779–796.
36. Liu, Ce, Lavanya Sharan, Edward H. Adelson, and Ruth Rosenholtz. “Exploring features in a bayesian framework for material recognition.” InComputer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 239–246. IEEE, 2010.
37. Lee, Dong-Woo, Soon-Hyeok Hong, Seok-Swoo Cho, Won-Sik Jool, “A study on fatigue damage modeling using neural networks.” Journal of mechanical science and technology 19.7 (2005): 1393–1404.
38. Dikbas, mustafa taskin uğur caligulu halil. “artificial neural network (ann) approach to prediction of diffusion bonding behavior (shear strength) of sicp reinforced aluminum metal matrix composites.” journal of yaşar university 3.12 (2008): 1811–1825.
39. António, CA Conceição, J. Paulo Davim, and Vítor Lapa. “Artificial neural network based on genetic learning for machining of polyetheretherketone composite materials.” The International Journal of Advanced Manufacturing Technology 39. 11–12 (2008): 1101–1110
40. Zhang, Shu Ling, Zhen Xiu Zhang, Zhen Xiang Xin, Kaushik Pal, and Jin Kuk Kim. “Prediction of mechanical properties of polypropylene/waste ground rubber tire powder treated by bitumen composites via uniform design and artificial neural networks.” Materials & Design 31, no. 4 (2010): 1900–1905.
41. Iskandarani, Mahmoud Z. “Application of Neural Networks to Matlab Analyzed Hyperspectral Images for Characterization of Composite Structures.” Journal of Intelligent Learning Systems and Applications 5.03 (2013): 143.
42. Kumar, Sanjeev, Ajay Batish, Rupinder Singh, and T. P. Singh. “A hybrid Taguchi-artificial neural network approach to predict surface roughness during electric discharge machining of titanium alloys.” Journal of Mechanical Science and Technology 28, no. 7 (2014): 2831–2844.
43. Atuanya, Clement Uche, Chidozie Chukwuemeka Nwobi-Okoye, and Okechukwu Dominic Onukwuli. “Predicting the mechanical properties of date palm wood fibre-recycled low density polyethylene composite using artificial neural network.” International Journal of Mechanical and Materials Engineering 9.1 (2014): 1–20.
44. Antony, P. J. Computational linguistic tools and machine translation system for Kannada language. Diss. Amrita vishwa vidyapeetham, 2012.
45. Ding, Peng, Qin Li, and Xunlei Huang. “Classification of acoustic emission sources produced by carbon/epoxy composite based on support vector machine.” In IOP Conference Series: Materials Science and Engineering, vol. 87, no. 1, p. 012002. IOP Publishing, 2015.
46. AbuOmar, O., S. Nouranian, R. King, T. M. Ricks, and T. E. Lacy. “Mechanical property classification of vapor-grown carbon nanofiber/vinyl ester nanocomposites using support vector machines.” In Proceedings of the International Conference on Data Mining (DMIN), p. 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014.
47. Veeraraghavan Sundararaghavan, Nicholas Zabaras, “Classification and reconstruction of three-dimensional microstructures using support vector machines”, 2nd edition, ELESIVIER. New York, (2003).

48. Das, Santanu, Ashok N. Srivastava, and Aditi Chattopadhyay. “Classification of damage signatures in composite plates using one-class SVMs.” In Aerospace Conference, 2007 IEEE, pp. 1–19. IEEE, 2007.
49. Wang, Quan-De, Zhi-Feng Zhong, and Xian-Pei Wang. “Design and implementation of insulators material hydrophobicity measure system by support vector machine decision tree learning.” In Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, vol. 7, pp. 4328–4334. IEEE, 2005.
50. Tang, Jia-li, Qiu-ru Cai, and Yi-jun Liu. “Prediction of Material Mechanical Properties with Support Vector Machine.” In Machine Vision and Human-Machine Interface (MVHI), 2010 International Conference on, pp. 592–595. IEEE, 2010.
51. Fauvel, Mathieu, Jón Atli Benediktsson, Jocelyn Chanussot, and Johannes R. Sveinsson. “Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles.” Geoscience and Remote Sensing, IEEE Transactions on 46, no. 11 (2008): 3804–3814.

Performance Analysis and Assessment of Safety Critical Systems Using Ann

K. Sruthi, R. Krishnaprasad, Manju Nanda and J. Jayanthi

Abstract Many of the safety critical systems (SCS) require the failure analysis to ensure that the applications work properly and safely. Failure of such safety critical systems results in huge losses. Unmanned aerial vehicles (UAVs) are safety critical systems. System Safety Assessments (SSA) is employed for analyzing these systems. SSA monitors the degraded functionalities of systems. Sensor faults are the root causes for most of the system faults and need to be assessed properly. The availability of several tools for the design, implementation, and simulation of artificial neural networks (ANN) makes it easier to use in safety assessment applications. MATLAB provides ANN features to analyze the sensor faults. In this paper we propose ANN as a novel approach to ensure the system safety and analyze the failures due to the sensor faults. UAV groundspeed sensor is considered as a case study. MATLAB/Simulink tool suite is used to implement the sensor model. The model outputs are classified and the classified values of sensor are used for prediction. The predicted values are classified for verifying the future performance. Simulation results discuss the classified values, performance analysis, and predicted outputs of the sensor model.

Keywords UAV · ANN · SCS · SSA

K. Sruthi (✉)
CESP, Jawaharlal Collage of Engineering and Technology, Palakkad, Kerala, India
e-mail: 92sruthik@gmail.com

R. Krishnaprasad
CSIR-National Aerospace Laboratories, Bangalore, India
e-mail: krishsnaprasad.rajan90@gmail.com

M. Nanda · J. Jayanthi
Aerospace and Electronics Division, CSIR-National
Aerospace Laboratories, Bangalore, India
e-mail: manjun@nal.res.in

J. Jayanthi
e-mail: jayanthi@nal.res.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_12

1 Introduction

Safety-critical systems are those systems whose failure could result in loss of life, significant property damage, or damage to the environment. Many modern information systems are becoming safety-critical in a general sense because financial loss and even loss of life can result from their failure [1]. Engineering of safety-critical systems is a complex task involving many technical fields. Unmanned aerial vehicles (UAVs), also known as drones, are aircraft either controlled by 'pilots' from the ground or increasingly, autonomously following a pre-programmed mission. It is one of most common example of safety critical systems. Due to their advantages, UAVs are used in many civil and military applications such as aerial surveillance, remote sensing, cargo delivery, etc. About 80% of flight incidents concerning unmanned aerial vehicles (UAV) are due to faults affecting propulsion, flight control surfaces, or sensors [2]. Here we are considering the UAV system as a set of sensors namely groundspeed sensor, angle of attack sensor, accelerometer sensor, angle of sideslip sensor, differential pressure sensor, etc.

Safety assessment is the process for assuring the safety and reliability of critical aeronautical systems. It uses probabilistic analysis to provide precise measures about the safety requirements of a system. The basic information used as input to these techniques are failure conditions and failure rates [3]. We define prognosis to be detecting the precursors of a failure, and predicting how much time remains before a likely failure. Algorithms that use the data-driven approach to prognosis learn models directly from the data, rather than using a hand-built model based on human expertise [4]. One useful approach is the data-driven approach, also known as the data mining approach or the machine learning approach, which uses historical data to automatically learn a model of system behavior [5]. So data-driven approach along with the safety assessment helps in the safe handling of UAV systems. Being the most famous data-driven approach, ANN finds its application in this case too. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [6]. The sensor model values are given to classifier and predictor which is used for prognosis. Groundspeed sensor is implemented using MATLAB/SIMULINK model and the outputs are given to ANN algorithm. Groundspeed is the horizontal speed of the aircraft relative to ground. Groundspeed sensor gives reasonable results in calm weather conditions. It measures actual airspeed resulting in better throttle control and aircraft performance especially in windy conditions [7]. The classifier output specifies whether the sensor works satisfactorily or not. Predictor predicts the future output values.

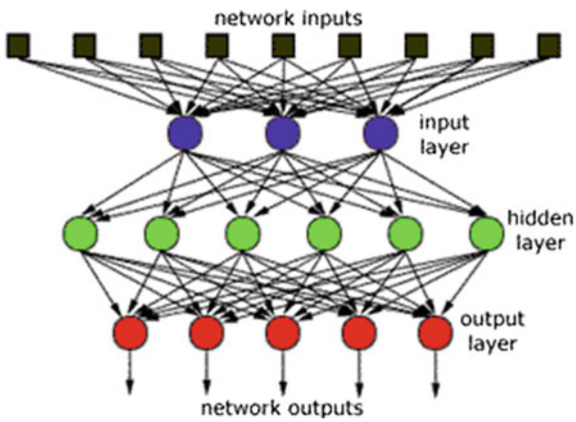
The paper organized in various sections as: Sect. 2 gives details about the ANN and tools, Sect. 3 explains block diagram and flowchart of the proposed approach, sensor model and algorithm details are given in Sect. 4 and results and conclusion are given in Sects. 5 and 6, respectively.

2 Related Work

Large variety of methods have been proposed for designing Safety analyzing systems. Some of the approaches are: pattern recognition, parameter estimation [8], observers (Luenberger observers, extended Luenberger observers, and unscented observers) [9], filters (Kalman filters, extended Kalman filters, unscented Kalman filters) [10], multiple model adaptive estimation and intelligent technology [11]. Intelligent technology consists of fault diagnosis such as expert system method, graphical method, soft computing, and distributed artificial intelligence. Graphical systems includes fault tree analysis and Bayesian network. Soft computing method consists of rough set, artificial neural network, support vector machine, fuzzy logic [12], artificial immune system, and decision tree are included. Among all these we are using ANN method. Bayesian networks are used to describe the relationship among project state impact factors, decisions to be taken and the satisfaction levels related to each decision. Approximated results are available in an efficient way while using Neural networks. An Artificial Neural Network (ANN) [13] is an information processing paradigm that is stimulated by the way biological nervous systems, such as the brain, process information. Usually, neural networks operate in two phases. The first phase is a learning phase where each of the nodes and links adjust their strength in order to match with the desired output. A learning algorithm is in charge of this process. When the learning phase is complete, the neural network (NN) is ready to recognize the incoming information and to work as a pattern recognition system (Fig. 1).

NN-based schemes benefit from factors such as; online learning algorithms, ability to model nonlinear systems and are programmable in a few lines of code [14]. There are several open source tools are available for ANN like open NN, just NN, Neural Machine, plug & Score, FANN, etc. Some of them are useful for commercial purpose while some others help in research works [15]. MATLAB is having lot of functionalities and tutorials are also available. So MATLAB is used

Fig. 1 ANN general architecture



for implementing the proposed method. The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network. The NARX model is based on the linear ARX model, and commonly used in time-series modeling.

3 Proposed Approach

The block diagram and the flowchart is explained in this section.

3.1 Block Diagram

Here the inputs are given to MATLAB SIMULINK model. These models are built according to the characteristic equation of each sensor. The outputs of sensor models are given to the neural network where the classification and prediction takes place. The predicted values are again given to the classifier which helps in evaluating efficiency of the system. The input values are decided on the basis of range of each individual values of the sensors (Fig. 2).

The proposed approach is relatively simple and precise than other safety assessment techniques. It checks the status of the system and along with that it monitors the future performance also. So, this approach helps in proper working as well as the long time life period of the system (Fig. 3).

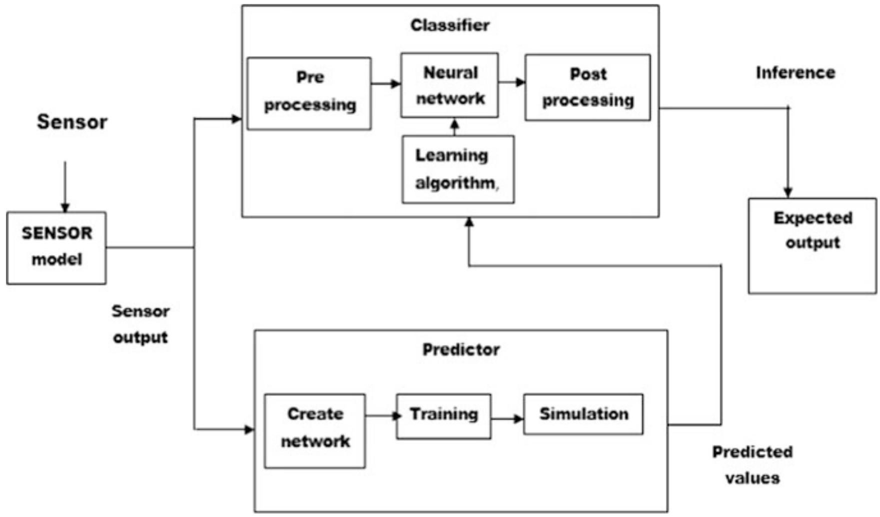


Fig. 2 Block diagram of the proposed approach

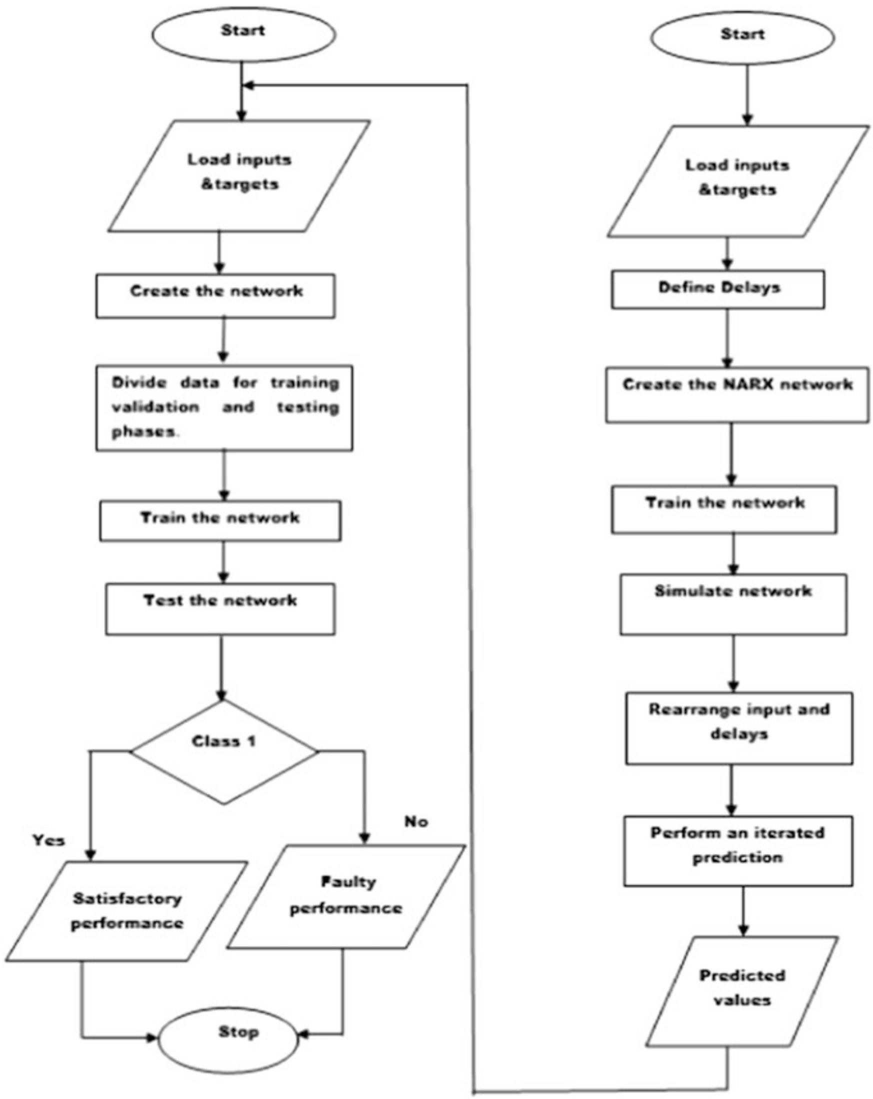


Fig. 3 Flowchart representation of system

3.2 Flowchart

The flowchart of the implemented program is given below. In the flowchart the general workflow is explained. In the case of classifier, after loading inputs and targets, networks is created, trained, and tested with real time data. The data is categorized based on the range of output values. If it belongs to class 1 the system gives satisfactory performance otherwise it indicates some failure condition. In the predictor part, the NARX network is created, trained, simulated, and used for iterated prediction. The predicted values are again given to classifier network to check the working in the future slots.

4 Implementation

In this section classification and prediction algorithms are explained briefly along with the explanation of sensor model.

4.1 Modeling of Sensor

The sensor model has been created based on the characteristic equation. We have implemented the model for groundspeed sensor and its out values are processed accordingly. All the other sensors mentioned in Sect. 1 can be implemented similarly. Consider the case of ground speed sensor. The characteristic equation is given below:

$$Vg = \sqrt{(v^2 + [Vw]^2 + 2.v.Vw\cos(\Psi - Xw))} \tag{1}$$

where, Vg = groundspeed, Vw = wind speed, v = airspeed, Xw = wind direction and Ψ = yaw value.

The database is created according to the equation and range of value of each input. The inputs are imported to the Simulink model. The simulated outputs are saved in the work space. The outputs of the Simulink model are given to the ANN algorithm (Fig. 4).

4.2 Classification Algorithm

In this case the inputs are the groundspeed monitored from the output of the Simulink model. The targets are the classes which indicate to which category the input given belongs to. Pattern recognition networks are feedforward networks that can be trained to classify inputs according to target classes. The target data for pattern recognition networks consists of vectors of all zero values except for a 1 in element *i*, where *i* is the class they are to represent. Processing functions transform user input data to a form that is easier or more efficient for a network.

Training phase: The first step in training is to divide data for training, validation, and testing. Here we are using the method of sampling for this. The

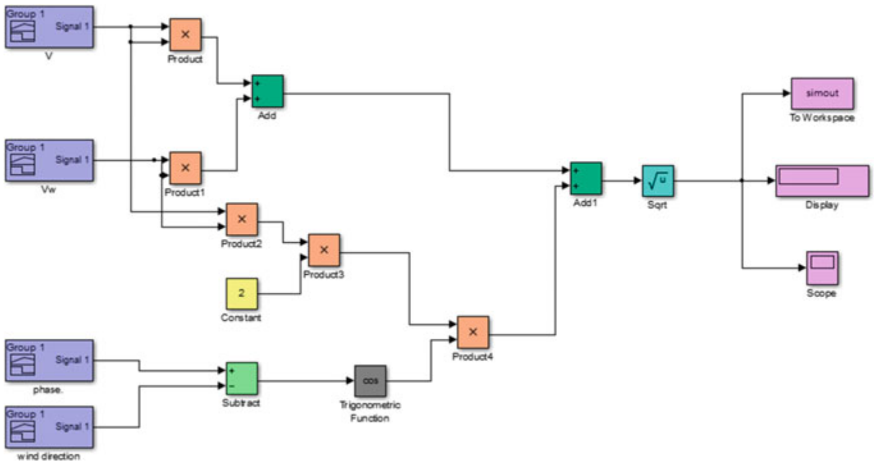


Fig. 4 Groundspeed sensor model

performance function used here is the mean square error mse—the average squared error between the networks outputs and the target outputs *t*. There are two different ways in which training can be implemented: incremental mode and batch mode. In incremental mode, the gradient is computed and the weights are updated after each input is applied to the network. In batch mode, all the inputs in the training set are applied to the network before the weights are updated. Here it is used the batch mode training with the train command which is faster and produces less errors. When training large networks and when training pattern recognition networks, 'trainscg' and 'trainrp' algorithms are good choices. Their memory requirements are relatively small, and yet they are much faster than standard gradient descent algorithms. Here 'trainscg' algorithm is used.

Perceptron concept: Perceptrons are simple single-layer binary classifiers, which divide the input space with a linear decision boundary. Perceptrons can learn to solve a narrow class of classification problems. Classification with pattern recognition gives more precise and accurate results.

4.3 Prediction Algorithm

When delay lines are used with multilayer feedback networks it can be used for prediction purposes. Some of the methods here used include NARX, NAR, etc.

Simple NARX method: Here after importing the inputs and targets, the delay values are specified. The typical workflow is to fully create the network in open loop, and only when it has been trained it is transformed to closed loop for multistep-ahead prediction. Thenby using 'narx' command the NARX network is created according to the hidden layer size given. Then the network is trained and

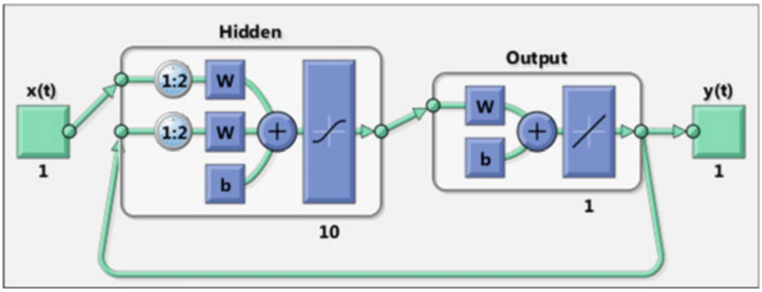


Fig. 5 Closed loop network for multistep-ahead prediction

simulated. The errors are plotted. For changing the series parallel configured network to parallel configuration 'closeloop' command is used. It helps for the multistep-ahead prediction (Fig. 5).

NARX network with training algorithm: Here the same NARX network is used along with the training steps which is explained in the classification part. The 'trainlm' algorithm is used to train the network. It is one of the fastest back propagation algorithm.

Function fitting networks: Function fitting is the process of training a neural network on a set of inputs in order to produce an associated set of target outputs. Once the neural network has fit the data, it forms a generalization of the input-output relationship and can be used to generate outputs for inputs it was not trained on. It can be used in the case of sensor values also.

NAR, Nonlinear input-output methods can also be used for the purpose. Only inputs differ in that case.

5 Results

The following images illustrate k-Means clustering applied to a set of randomly generated 2-d data points. The points are generated using a normal distribution cantered at a mean location and with a constant standard deviation. The section contains the output samples as well as the performance graphs of classification and prediction. Classifier outputs are saved in workspace. The 1 in first column indicates class 1, i.e., satisfactory performance. Here instead of exact value, approximated value is an output like 0.99 (Fig. 6).

Performance graph: The graph shows the mse (mean squared error) of training, validation, and testing phases according to epochs. Performance is measured in terms of mean squared error, and shown in log scale. It rapidly decreased as the network was trained. Network gives the best performance at epoch 45.

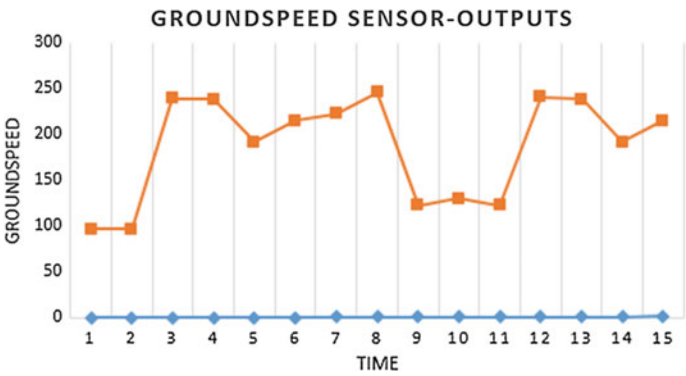


Fig. 6 Output waveform of groundspeed sensor

Error histogram: It shows an approximation of errors occurred at training, validation, and testing phases. Errors are calculated by subtracting outputs from targets.

Confusion matrix: The confusion matrix shows the percentages of correct and incorrect classifications. Correct classifications are the green squares on the matrices diagonal. Incorrect classifications form the red squares. If the network has learned to classify properly, the percentages in the red squares shows very small values, indicating few misclassifications. These performance measurements are available from the command window. These three plots, i.e., Performance plot, error histogram and confusion matrix are the outputs of classifier.

Also, Prediction can be done using the sensor values. The following graph shows the difference between the actual and predicted output. The figure illustrates the iterated prediction. The blue line is the actual output value of the sensor, and the red line is the value predicted by the NARX neural network.

The table shows the error values found in several prediction methods used. NARX with training gives better results (Table 1) (Figs. 7 and 8).

By analyzing these plots we can arrive at a conclusion about the working of sensor. If it is not working properly we can do failure preventive measures so that it helps to increase the lifetime of the system. The process is comparatively simpler while considering the Bayesian networks. Thus the classifier as well as predictor outputs together help for the safety assessment of system.

Table 1 Errors in each prediction methods

Prediction method	Errors in each time steps					
NARX with training	7.4411	6.8570	9.7674	7.9233	8.016	9.3242
Simple NARX	8.4411	9.8570	10.7674	8.9233	10.016	11.3242
NAR	15.441	19.8570	18.7674	12.923	13.016	15.3242
Nonlinear i/p-o/p	20.441	22.857	70.7674	40.9233	30.0167	27.3242

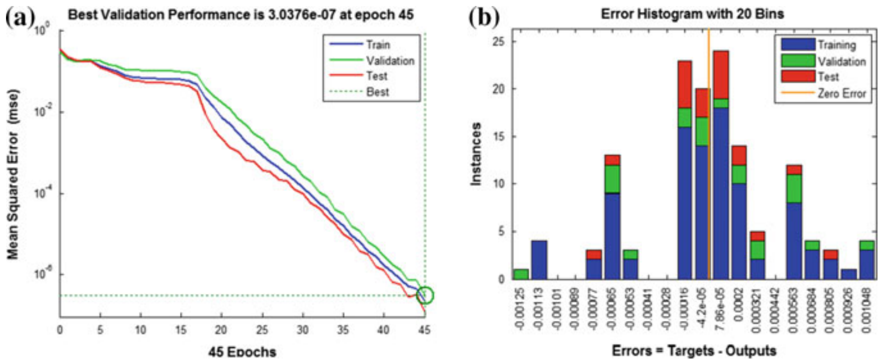


Fig. 7 **a** Performance graph, **b** Error histogram

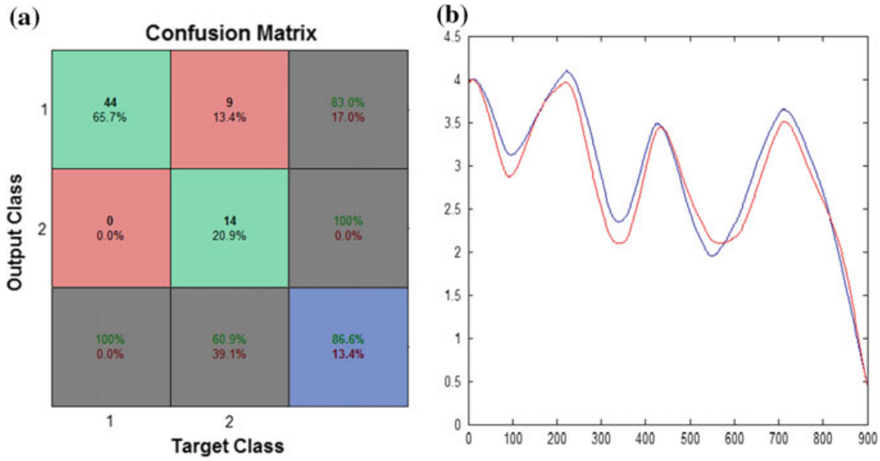


Fig. 8 **a** Confusion matrix, **b** Prediction plot

6 Conclusion

This paper explains the need of machine learning algorithms to tackle very large amount of data. Safety-critical systems are those systems whose failure could result in loss of life, significant property damage, or damage to the environment. From a software perspective, developing safety critical systems is required in significant areas such as specification, architecture, verification, and process. Due to the advantages, UAVs are used in many civil and military applications such as aerial surveillance, remote sensing, and cargo delivery. Being one of the most important example of safety critical system the safe handling of such UAV systems are very important. Safety assessment of such systems can result in efficient monitoring of their performance. Such safety systems help to improve the lifetime of UAV systems. Analyzing the safety is not just concerned about the sensor faults. Many areas under which safety analysis can be used is under research. Use of ANN networks are very helpful while detecting the faults. However, the methods which use other some other techniques like Bayesian networks, Hidden markov models, etc., are also under the research.

References

1. Inseok Hwang, Sungwan Kim, Senior, Youdan Kim, and ChzeEngSeah. "A Survey of Fault Detetion, Isolation, and Reconfiguration Methods, IEEE transactions on control systems technology", Vol. 18, pp. 636–653, (2010)
2. Diogo Leite Rebouças, Fábio Meneghetti Ugulino de Araújo, André Laurindo Maitelli "Use of artificial neural networks to fault detection and diagnosis", ABCM Symposium Series in Mechatronics—Vol. 5, pp. 666–675, 2012
3. Lennon R. Cork1, Rodney Walker1, Shane Dunn "Fault Detection, Identification and Accommodation Techniques for Unmanned Airborne Vehicle".
4. Adriano Gomes, Alexandre Mota, Augusto Sampaio, Felipe Ferri, and Julio Buzzi "Systematic Model-Based Safety Assessment via Probabilistic Model Checking"
5. Iz Al-Dein Al-Zyoud and K. Khorasani "Neural Network-based Actuator Fault Diagnosis for Attitude Control Subsystem of an Unmanned Space Vehicle"
6. Iman Sadeghzadeh1, and Youmin Zhang: "A Review on Fault-Tolerant Control for unmanned Aerial vehicle, St. Louis, Missouri", 2011
7. Bhaskar Prasad Rimal, Hyoim Shin and Eunmi Choi "Simulation of Nonlinear Identification and Control of Unmanned Aerial Vehicle: An Artificial Neural Network Approach", 2009, pp. 442–447
8. J Marzat, H Piet-Lahanier, F Damongeot, and E Walter, "Model-based fault diagnosis for aerospace systems", a survey, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2011, pp. 1–31.
9. Iman Sadeghzadeh, and Youmin Zhang, "A Review on Fault-Tolerant Control for Unmanned Aerial Vehicles (UAVs)", American Institute of Aeronautics and Astronautics, 2011, pp. 1–12.
10. Isermann, "Supervision, Fault- Detection and Fault-Diagnosis Methods- an Introduction", Elsevier science Ltd, 1997, pp. 639–652.
11. Hamid R. Berenji and Yan Wang, "Wavelet Neural Networks for Fault Diagnosis and Prognosis", IEEE International Conference on Fuzzy Systems, 2006, pp. 1334–1339.
12. Ravi Kapadia, Greg Stanley, Mark Walker, "Real World Model-based Fault Management", pp. 1–8.
13. Georg Jager, Sebastian Zug, TinoBrade, Ana-Maria Cretu, Christian Moewes: "Assessing neural networks for sensor fault detection"
14. Samy. I. Postlethwaite, D.-W. Gu, I.S. Fan "Detection of multiple sensor faults using neural networks- demonstrated on an unmanned air vehicle (UAV) model"
15. Gurcan Lokman, Guray Yilmaz "A new method for Anomaly Detection and Target Recognition", International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA, 2014. pp. 77–583, 2014

Functional Mode Analysis of Safety Critical Systems Using Causal Bayesian Networks (CBNs)

Susmitha Thiyyanoor, R. Krishnaprasad, Manju Nanda and J. Jayanthi

Abstract System Health Management (SHM) is a key technology for detecting, diagnosing, predicting, and mitigating the adverse events during the operation of safety-critical systems. Safety critical systems, specifically Unmanned Aerial Vehicles (UAV) are an important part of today's era whether it is in civilian, commercial, defense, or in military domain. Proper functioning of these sensor systems is very crucial as their faults can result in serious consequences, but they often fail in spite of extensive verification and validation efforts, which raise safety concerns. This paper discusses functional mode analysis of speed and direction sensor to perform SHM using Causal Bayesian Networks (CBN) that can tackle problems associated with system bugs and failures. Sensor parameters from UAV system in real-time is learned, modeled, and analyzed in using Bayesian network. The simulation output graphically shows the influence analysis of sensor parameters on the overall health of the UAV system as a case study. The network performance along with a comparison of actual and predicted values is displayed in the simulation section.

Keywords Safety critical systems · Unmanned aerial vehicles
Functional mode analysis · System health management
Causal bayesian networks · BayesiaLab

S. Thiyyanoor (✉)
Jawaharlal Collage of Engineering and Technology, Palakkad, Kerala, India
e-mail: susmithatm@gmail.com

R. Krishnaprasad
CSIR-National Aerospace Laboratories, Bangalore, India
e-mail: krishsnaprasad.rajan90@gmail.com

M. Nanda · J. Jayanthi
Aerospace and Electronics Division, CSIR-National Aerospace Laboratories, Bangalore, India
e-mail: manjun@nal.res.in

J. Jayanthi
e-mail: jayanthi@nal.res.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_13

1 Introduction

Safety critical systems are those systems whose failure can cause serious threat to life and property. Unmanned aerial vehicle (UAV) is an important category of safety critical systems and is commonly known as drone or an aircraft without a human pilot. The flight of UAV may be controlled autonomously by onboard computers or by the remote control of a pilot on the ground or in another vehicle. A flight control system that makes the UAV fly or run automatically is called autopilot, have a lot of functions such as guidance, control, and navigation. It act as brain of UAV, this subsystem controls UAV by generating control signals on the basis of desired target information and waypoints. Nowadays space activities including UAVs are characterized by increased constraints in terms of onboard computing power and functional complexity combined with reduction of costs and schedule. This scenario necessarily originates impacts on the onboard software with particular emphasis to the interfaces between onboard software and system/mission level requirements.

System Health Management is an important factor in system level requirement. Health management is performed on the running safety critical systems with the goal to perform diagnosis and prognosis and hence isolates the faults close to their source so that a fault in a sub-system does not lead to a general failure of the global system [1]. A SHM system continuously monitors the behavior of the software and the interfacing hardware or sensor components. Using an abstract model of the software, the SHM can detect unexpected behavior, reason about its root cause, and trigger failure repair or mitigation actions. Only recently, health management systems that monitor software have been developed. The goal of SHM system is to correctly diagnose off-nominal situations with special consideration to sensors that are incorporated in the UAV system. If any of the sensors cause failures in the active stage due to faults, it will affect the functionality of UAV. So every sensor is required to be monitored for its proper functionality which is an important SHM application [2].

Prominent SHM techniques are using Kalman filters, Bayesian Hidden Markov Model approach [3], Artificial Neural Networks (ANN), Causal Bayesian Networks, etc. [4]. Of these, Bayesian Networks can be built from human knowledge, i.e., from theory, or they can be machine-learned from data [5] and holds good for aircraft guidance, navigation, and control [6]. Also, Causal Bayesian Networks can be modeled with their node-arrow structure and due to their graphical structure, machine-learned Bayesian networks are visually interpretable, therefore promoting human learning and theory building.

The paper is organized as follows: Sect. 2 describes the Background work behind the project, Sect. 3 explains the Proposed Approach, Sect. 4 describes the implementation of the system, Sect. 5 shows the Simulation results and finally Sect. 6 gives the scope and conclusion for the proposed system.

2 Background Work

Various SHM techniques exist including Kalman filters, ANN, Causal Bayesian networks etc. Causal Bayesian networks (CBNs) represent multivariate probability distributions and are used for reasoning and learning under uncertainty [7–9]. They are often used to model systems of probabilistic nature [4]. Random variables are represented as nodes in a directed acyclic graph model, while conditional dependencies between variables are represented as graph edges. A key point is that a CBN, whose graph structure often reflects a domain's causal structure, is a compact representation of a joint probability table.

Many CBN tools exist at present like BayesNet toolbox, Hugin [10], GeNIe and SMILE [11], Netica [12], UnBBayes, OpenMarkov, Direct Graphical Models, etc. But the recent advancement among CBN tools is BayesiaLab 5.4.3 (released in 2015) which is being used in this system. BayesiaLab is a powerful desktop application (Windows/Mac/Unix) with a sophisticated graphical user interface, which provides users a comprehensive "laboratory" environment for machine learning, knowledge modeling, diagnosis, analysis, simulation, and optimization. BayesiaLab leverages the inherently graphical structure of Bayesian networks for exploring and explaining complex problems.

3 Proposed Approach

Functional Mode Analysis or System Health Management (SHM) of UAV sensors using Bayesian Networks has been proposed in this paper. This section consists of the block diagram for the proposed system and the description for the concepts of Knowledge modeling and Machine learning performed using CBN in BayesiaLab.

Probabilistic models based on directed acyclic graphs (DAG) have a long and rich tradition and their variants have appeared in many fields. Within statistics, such models are known as directed graphical models; within cognitive science and artificial intelligence, such models are known as Bayesian networks. The name honors the Rev. Thomas Bayes, whose rule for updating probabilities in the light of new evidence is the foundation of the approach. It addresses both the case of discrete probability distributions of data and the more complicated case of continuous probability distributions. In the discrete case, Bayes' theorem relates the conditional and marginal probabilities of events A and B, provided that the probability of B not equal zero:

$$P(A/B) = P(A) * \frac{P(B/A)}{P(B)} \tag{1}$$

The fact that the significant parameters that influence the sensors in a UAV autopilot also influence the proper functioning of these sensors and ultimately

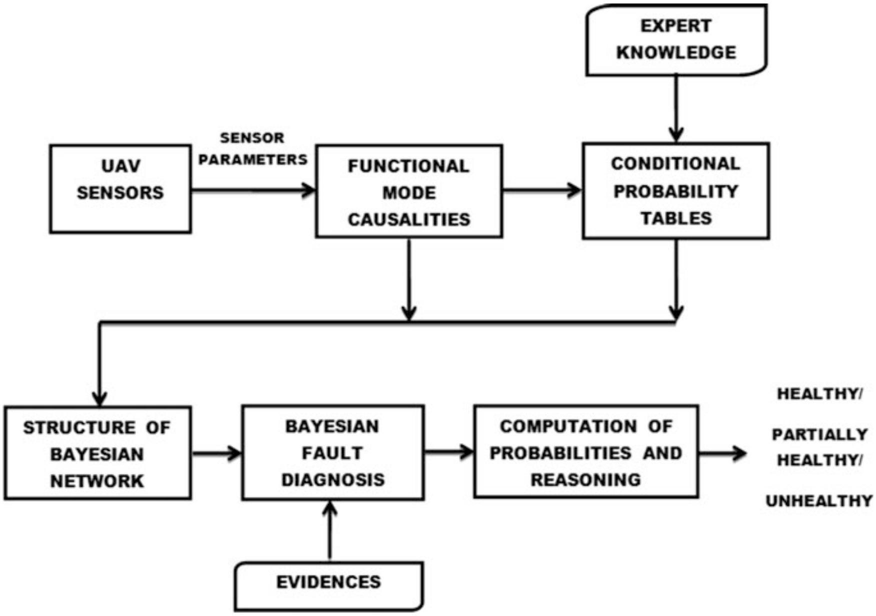


Fig. 1 Implementation of the system

determine the overall health of the UAV system is being exploited. Proposed approach consists of UAV sensor parameters and their derived functional mode causalities, corresponding conditional probability tables (derived from expert knowledge), and the hence derived Causal Bayesian Network. The Bayesian fault diagnosis is then performed on this network by providing evidences and computation of different fault probabilities. Thus, the system can be checked for its different functional modes, i.e., healthy, partially healthy, and unhealthy (Fig. 1).

The implementation of the system mainly constitutes two sections:

3.1 Knowledge Modeling and Evidential Reasoning

Reasoning in complex environments creates cognitive challenges for humans. Subject matter experts often express or model their causal understanding of a domain in the form of diagrams, in which arrows indicate causal directions. This visual representation of causes and effects (known as Knowledge Modeling) has a direct analog in the network graph, in BayesiaLab. The model can then be analyzed by acquiring proper evidences and then executing them. The steps in knowledge modeling can be briefed as below in the flow chart.

In Knowledge Modeling, the complex information is often simplified to form causal relationships between different variables involved in it. In BayesiaLab, the probabilistic relationships between variables in conditional probability tables (CPT) have to be described, which means that no functional forms are utilized.

First stage nodes represent the parameters, second stage nodes represent the healthy and unhealthy mode of the sensors, and third stage nodes represent the overall functional modes of the system. Arrows specify the conditional dependency between the nodes. Necessary knowledge is assigned as Conditional Probability Tables to the nodes of the system. Switching on to the Validation mode helps users to check the functionality modes of the system by giving suitable hard and soft evidences.

3.2 Machine Learning

The earlier work gives a concept that can be helpful for many research works but it does not facilitate real-time processing of input data from the parameters of interest to determine the functionality modes of the sensors and thus the system. Human expert knowledge is useful for identifying causal relations, but proves to be inefficient in real time. Most of the research works face the challenge of handling with the real-time data acquired from on board during active run of the system. Machine Learning comes into play at this instinct where real-time processing is a must.

Machine Learning is implied in this system to learn and then establish the predictive importance of a range of variables with regard to a target variable. The domain is the UAV Autopilot system and we wish to examine the relationships between sensor parameters and the overall functionality of the system. The real-time data from the UAV sensors is acquired as input to the Bayesian Networks as excel spread-sheets (or Comma Delimited file). The highly optimized learning algorithms that can quickly uncover structures in datasets are considered in BayesiaLab for the process of testing and learning. Naive Bayes algorithm which is a Supervised Learning approach, a causal dependency is formed between the "Class" node and the other nodes. It was found to be more useful for the sensor data because the target variable which is the functionality mode of the sensor will always have a causal dependency on every parameter that defines the sensor. The relationship between the Target node and other nodes is viewed by highlighting the Mutual Information between them which reflects the predictive importance of the parameters on the target node.

Machine learning has many benefits over knowledge modeling, of which the most important one is real-time processing capability. Also in machine learning, a detailed analysis and comparison of different parameters on the target node can be obtained which is absent in Knowledge modeling. It also helps in interpretation of real-time data in different perspectives: Target Interpretation tree, Adaptive Questionnaire, Mapping, etc.

4 Implementation

Implementation of the proposed system is shown in this section. It describes the procedures for knowledge modeling in BayesiaLab and designing of test data for machine learning. Also selection of different functional modes based on whether the parameters are in the proper nominal range or not is also explained here in this section.

4.1 *Knowledge Modeling and Evidential Reasoning*

The proposed work relies on considering the Speed and Direction Sensor and its different parameters. The parameters of interest and their nominal ranges of operation are Output Ambient Temperature (−40–150 °C), Output Current (30–85 mA), Magnetic Offset (−60–60), Output Frequency (0–40 kHz), Output Air Gap (0.75–3 mm), and Duty Cycle Variation (40–60%). Each of these parameters individually contributes to the health of the sensor and thus the overall health of the system. The effect of variation of these parameters from their nominal values in health of the sensor is described as: Healthy—if all sensors are healthy or 0 unhealthy sensors; partially healthy—if 1–3 sensors are unhealthy; Unhealthy—if 4–6 sensors are unhealthy.

Now, similar to the Speed and Direction sensor, 5 other sensors were also identified which are gyro sensor, accelerometer, angle of attack, altimeter, and differential pressure sensor. The overall functionality mode of the system (whether Healthy, Partially Healthy or Unhealthy) is determined by the functional mode analysis of each sensors. Figure 2 shows the overall modeled CBN for the proposed system.

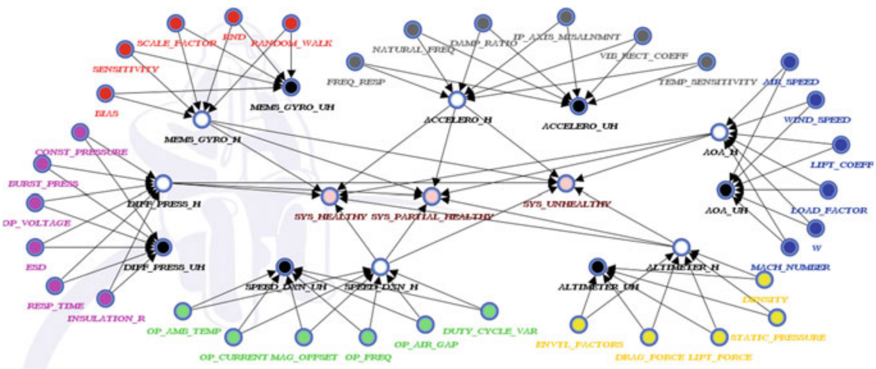


Fig. 2 Knowledge modeling

4.2 Machine Learning

On a top-level, real-time data from software and hardware sensors is learned and then presented as the nodes of the Bayesian network, which in turn performs its reasoning (i.e., updating the internal health nodes) and returns information about the health of the software. Machine learning of sensor data is explained below. Prior to that, designing of test data is to be discussed.

One-fifth (20%) of the data are chosen as test data from which the software learns the system. There are six parameters considered for the sensor.

Hence, $2^6 = 64$ combinations of data is used as Test set and thus, $64 * 5 = 320$ data combination is included in the database as input to the CBN and remaining data is said to be Learning Set.

The data set input is shown in Fig. 3. The dataset is then Machine-Learned after certain pre-processing steps like Discretization, Normalization, etc. The Learning algorithm applied is Naïve Bayes algorithm that best suits such applications where there is a single target that depends on several sub-factors. Figure 4 shows the machine learned CBN for the overall system as per Naïve Bayes algorithm.

The obtained outputs and differences are shown in the next section. In Knowledge Modeling, only just reasoning is possible, whereas in Machine learning, detailed analysis of data is possible which comes under three categories:

4.2.1 Performance Analysis

The relationship between the Target node and other nodes is viewed by highlighting the Mutual Information between them which reflects the predictive importance of

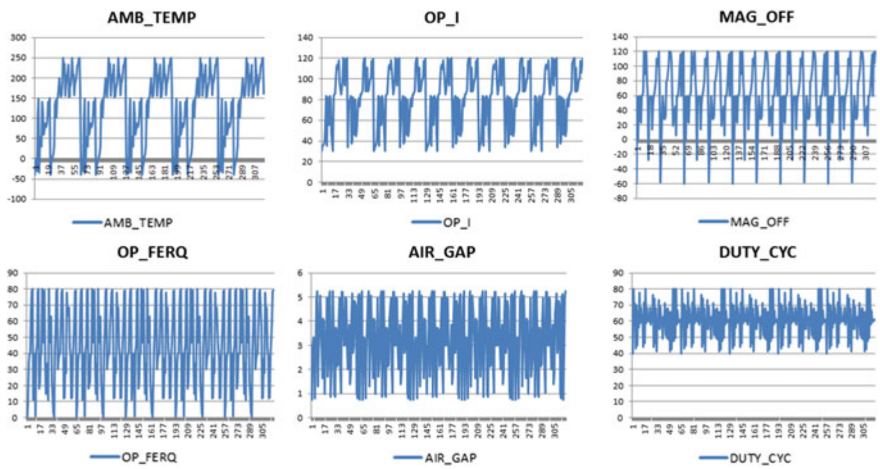


Fig. 3 Data set input for machine learning

Fig. 4 Machine learned Causal Bayesian Network

the parameters on the target node. The network performance is analyzed to know how the Naïve Bayes learning algorithm predict the states of the Class variable, Healthy or Unhealthy. Network performance on the target is shown in the Simulation Results.

To mitigate any sampling artifacts that may occur in such a one-off Test Set, we can systematically learn networks on a sequence of different subsets and then aggregate the test results. For this, we perform K-Folds Cross Validation (to iteratively select K different Learning Sets and Test Sets and then, learn the networks and test their performance) and is also shown in Simulation Results. Next step is Structural coefficient ("significance threshold" for network learning) Analysis. This analysis shows Structure/Target Precision Ratio which is a very helpful measure for making trade-offs between predictive performance versus network complexity.

4.2.2 Model Inference

The main objective of the proposed system is to derive a correlation between different parameters of the network and the target. Target Correlation, is obtained by sorting the parameters based on their Mutual Information with the target node "Class."

4.2.3 Interactive Inference

Interactive Inference is a special feature that helps the user to review the individual predictions made based on the model. The user can give evidences to check for the different functionality modes that hold for the system. Adaptive Questionnaire is an important category of Interactive inference where only individual cases are under review. End user can check any number of evidences and the probability distribution of the target node gets updated as a result. Not only the target node, but also all other nodes get updated upon setting evidence, reflecting the omnidirectional nature of inference within a Bayesian network. The process of updating can be continued until an acceptable level of certainty regarding the diagnosis is achieved.

Target Interpretation Tree is the next significant inference, and is explicitly shown in the form of a static graphical tree. The Target Interpretation Tree is induced once from all cases and then prescribes in which sequence evidence is be sought for gaining the maximum amount of information towards a diagnosis. Mapping is another inference category where the size of the nodes is proportional to the Mutual Information with the Target Node given the current set of evidence.

5 Simulation and Results

This section lists the obtained simulations and results. Simulated results for machine learning are classified under 3: (i) Performance Analysis; (ii) Model Inference; and (iii) Interactive Inference which are also described below.

5.1 *Knowledge Modeling and Evidential Reasoning*

In Knowledge modeling, the information about the health of the system is extracted from the posterior distribution, specifically from health nodes. The simulated outputs are briefed below in Fig. 5.

Figure 5 shows 3 sections: (i) the probabilistic distribution of each parameter of Speed and Direction sensor. Initially, all probabilities are normalized which shows an effect on healthy and unhealthy nodes of the sensor. (ii) Shows the effect of hard evidences on the healthy and unhealthy nodes of the sensor. (iii) Shows the effect of soft evidences on the healthy and unhealthy nodes of the sensor.

Fig. 5 Simulated output for evidential reasoning in BayesiaLab

5.2 Machine Learning

In machine learning, all sensor data, which are usually time series, must undergo a pre-processing step, where certain (scalar) features are extracted. These values are then discretized into symbolic states or normalized numeric values before presented to the Causal Bayesian model. The optimization criteria in BayesiaLab's learning algorithms are based on information theory based on which desired simulated results are obtained.

Figure 6 shows the Mutual Information between the parameters of the sensor and the target "CLASS", which shows the healthy and unhealthy modes of the sensor.

Figure 7a shows the overall performance can, which is expressed as the Total Precision, and is computed as total number of correct predictions (true positives + true negatives) divided by the total number of cases in the Test Set, i.e., (17 + 41) ÷ 64 = 90.625%.

Figure 7b shows that with different samples of data considered, the system proves good as it shows a total precision of (81 + 204) / 320 = 89.06%.

Figure 8 shows different parameters of the sensor that are sorted in the decreasing order of correlation with the target. The correlation is calculated based on the Mutual Information between different parameters and the target node. It is clear from the figure that OP_I has more correlation with the target followed OP_FREQ and so on. It is also verified that slight variations in the parameter probabilities induce effective variations in the target node too.

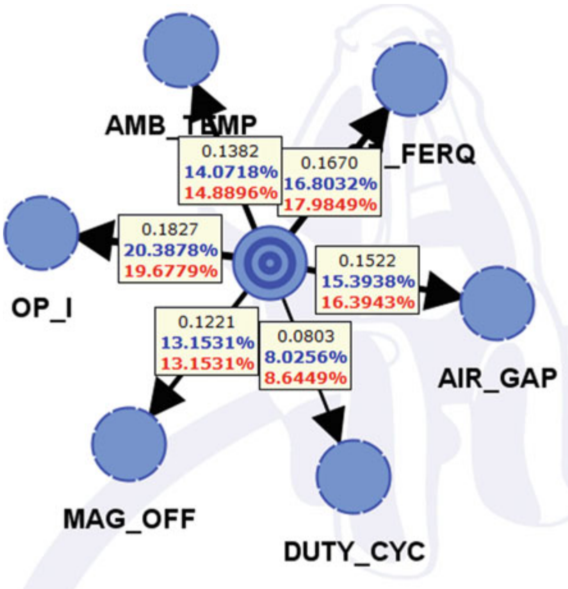


Fig. 6 Machine learning technique to obtain mutual information between CLASS variable and different parameters

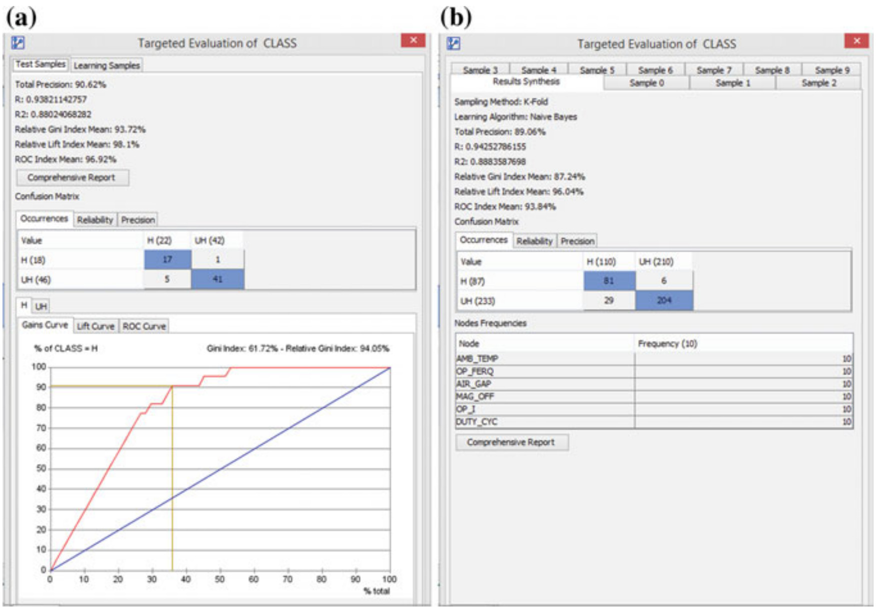


Fig. 7 **a** Network performance based on the target, **b** K-fold validation

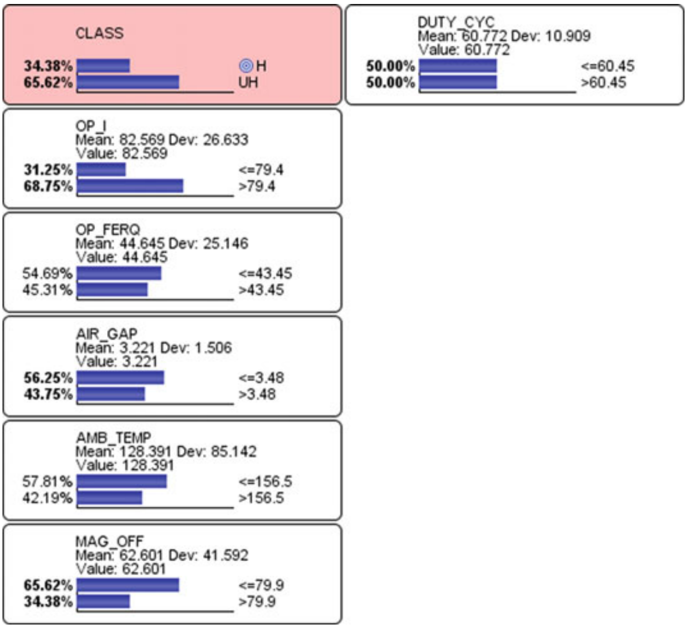


Fig. 8 Target correlation based on mutual information

Figure 9 shows the Adaptive Questionnaire section in which a third person can provide his/her evidences which are known to him/her. He/she need not know the system design. With the known evidences, even if only individual cases are under review, the system provides proper diagnostic support.

Figure 10 shows the Target Interpretation Tree which is induced from the cases shown in the monitor:

AMB_TEMP = p{ <=156.5:28.95%; >156.5:71.05%};
OP_I = <=79.4;
MAG_OFF = > 79.9;
OP_FREQ = p{ <=43.45:54.39%; >43.45:45.61%);
AIR_GAP = > 3.48;

The tree picturizes the effect of these evidences on the parameter DUTY_CYC and how its information is to be sought for gaining the maximum amount of information towards a diagnosis.

Figure 11 shows the mapping of target node with respect to other parameter nodes by applying Node analysis. The size of the nodes is determined by their Mutual Information with the Target node.

These simulation results under machine learning show the effect of the real time individual parameter values on the health of the Speed and Direction sensor. The

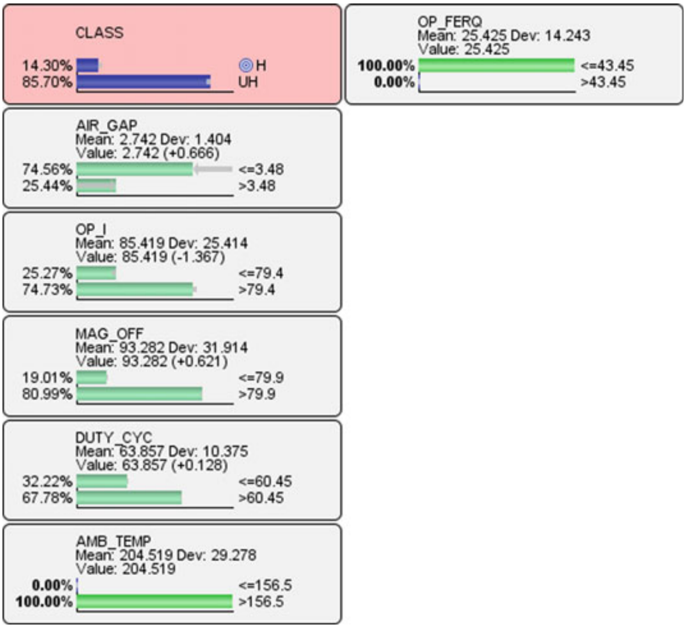


Fig. 9 Adaptive questionnaire

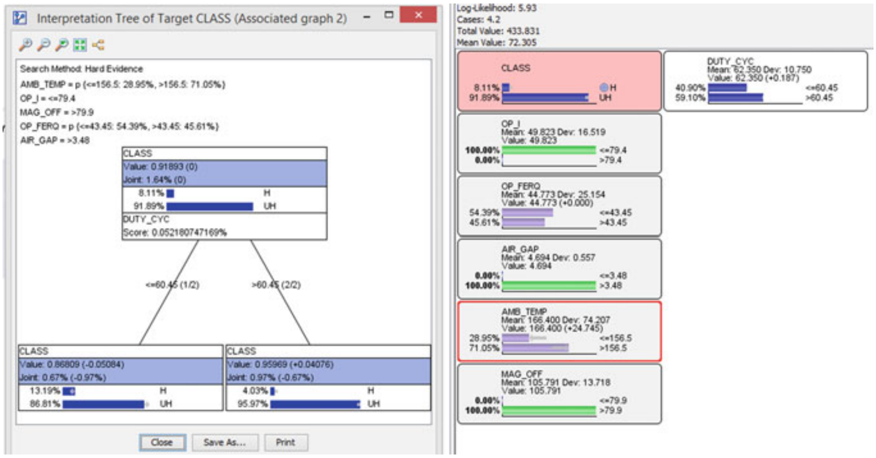


Fig. 10 Target interpretation tree

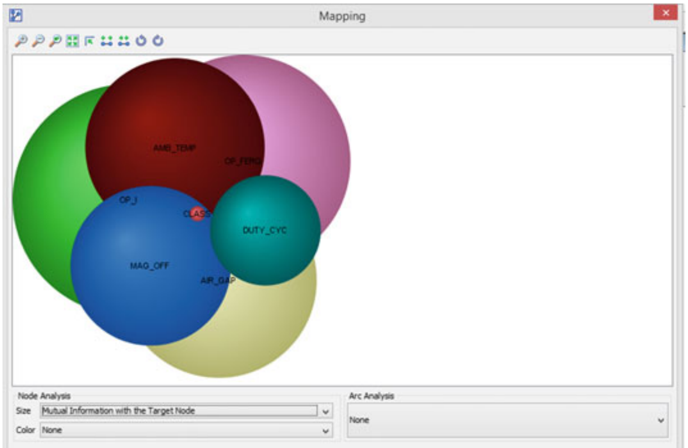


Fig. 11 Mapping

different outputs give the correlation of these parameters on the functional mode of the sensor and this correlation is being calculated using the concept of Mutual information in Bayesian Networks. The system also facilitates in interaction between an end user and the system through its features like Target Interpretation tree and Adaptive Questionnaire.

6 Conclusion & Future Work

System Health Management of safety critical systems is an important concept and is put to extensive research works for the past years. Irrespective of their complexity, many Fault Detection and Recovery techniques were tried out on UAVs. SHM is a key technology for detecting, diagnosing, predicting, and mitigating the adverse events during the operation of safety-critical software systems. Since size and complexity of software for even tiny autonomous systems increase dramatically, we think that powerful on board means for real-time fault detection and diagnosis can provide a crucial additional layer of reliable functioning. Causal Bayesian Networks, because of their numerous advantages prove to be useful in the area of SHM.

The concept of Functional Mode Analysis for SHM to be implemented using Causal Bayesian Networks has been discussed here. Knowledge Modeling and Machine Learning were tried out and irrespective of the various advantages modeling offered, machine learning was found to be more useful for practical real-time applications as they learned the system behavior on their own and facilitated for further analysis purposes.

The research work can be extended to robustly handle unexpected and un-modeled failures that can cause threat to both life and property. It can also be extended to artificially model Bayesian models when specific anomalies are found in the system.

Acknowledgements The authors thank the Director, CSIR-NAL for his support and encouragement to carry out this research work.

References

1. Holzel, N. B., Schilling, T., Neuheuser, T. and Volker Gollnick, Hamburg University of Technology, Air Transportation Systems, Hamburg, Germany, Lufthansa Technik AG, Hamburg, Germany: System Analysis Of Prognostics And Health Management Systems For Future Transport Aircraft, 28th International Congress of the Aeronautical Sciences, ICAS (2012)
2. Nagabhushan Mahadevan, Abhishek Dubey, Gabor Karsai,: Application of Software Health Management Techniques, Waikiki, Honolulu, HI, USA, SEAMS'11, May Vol. 23–24, pp. 1–10, 2011
3. Enkhjargal Dorj, Chaochao Chen, Michael Pecht, Fellow, IEEE, "A Bayesian Hidden Markov Model-Based Approach for Anomaly Detection in Electronic System", IEEE, 2013
4. Johann Schumann, Ole J. Mengshoel, Ashok N. Srivastava, Adnan Darwiche, FSE/DSP 2010 FSE/SDP Workshop,: Towards Software Health Management with Bayesian Networks
5. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publications, University of California, Los Angeles, (1988)

6. Johann Schumann, SGT, Inc. NASA Ames, Timmy Mbaya, University of Massachusetts.: Bayesian Software Health Management for Aircraft Guidance, Navigation, and Control, Boston, Ole Mengshoel, Carnegie Mellon University, NASA Ames, Moffett Field, Annual Conference of the Prognostics and Health Management Society, (2011)
7. Sylvain Verron, Jing Li, Teodor Tiplica.: Fault detection and isolation of faults in a multivariate process with Bayesian network, Informatics and Decision Systems Engineering, Arizona State University, Tempe, Journal of Process Control 20, pp. 902–911 (2010)
8. B. Bhajantri Lokesh, Basaveshwar Engineering College, Karnataka, India and N. Nalini, Bayesian Network Based Fault Tolerance in Distributed Sensor Networks, Nitte Meenakshi Institute of Technology, Karnataka, India, Journal of Telecommunications and IT, pp. 44–52
9. Lo, C.H., Wong Y.K. and Rad A.B., The Hong Kong Polytechnic University Hung Hom, Kowloon, Hong Kong: Bayesian Network for Fault Diagnosis
10. Hugin This is available at: http://www.hugin.com./
11. GeNIe This is available at: http://www.2.sis.pitt.edu/ ~ genie/.
12. Netica This is available at: http://www.norsys.com/networklibrary.html

Concept-Based Extractive Text Summarization Using Graph Modelling and Weighted Iterative Ranking

S. Chitrakala, N. Moratanch, B. Ramya, C.G. Revanth Raaj and B. Divya

Abstract Text summarization is a process of reducing the whole text document into a summary by retaining the most important information and to present it to the end user. Wikipedia which is a human-generated knowledge base is used to identify the key sentences using weighted iterative ranking algorithm which is the variation of HITS algorithm. The pre-processed input document is used in the construction of bipartite graph which maps the input sentences to the Wikipedia concepts. The bipartite graph captures the nested level of relationship between the sentences and concepts to ensure the highest level of efficiency in the extractive output summary. Weighted iterative ranking algorithm is used to retrieve top ranked sentences. The system produce summaries with good coverage, high coherency and low redundancy. The system can be deployed to summarize news articles, producing abstracts from documents, summarize web pages. The new article summarization would be helpful to mobile users.

Keywords Natural language · Wikipedia · Text summarization
Weighted iterative ranking · Bipartite graph · Concept based

S. Chitrakala (✉) · N. Moratanch · B. Ramya · C.G. Revanth Raaj · B. Divya
Department of CSE, College of Engineering, Anna University, Chennai, India
e-mail: au.chitras@gmail.com

N. Moratanch
e-mail: tancyanbil@gmail.com

B. Ramya
e-mail: ryaria18@gmail.com

C.G. Revanth Raaj
e-mail: revanthrajcgr@gmail.com

B. Divya
e-mail: divyabalamurugan95@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_14

1 Introduction

Summarization generally focuses on generating a condensed and crisp version of a document that covers the document's main topic. In recent times, text summarization plays a prominent role in providing the most important and precise information to the users. There are two approaches to automatic summarization, extractive and abstractive methods of summarization. Extractive method of summarization selects a subset of existing words, phrases or sentences in the original text to form the summary. Abstractive method of summarization build an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a human might generate.

Summaries are composed by number of sentences. So, the basic idea of arriving at the summary is to include the sentences that serve more meaning to the summary and the sentences should be present in the same order as it is given in the original document. In other words, every selected sentence is expected to be both salient and novel.

The main contribution of this work is to cast the Wikipedia-based summarization problem into a general sentence-concept *bipartite framework*, and weighted *iterative ranking algorithm* for selecting summary sentences. The summary of sentences is produced to the users. Also, the system provides incremental summarization. The one-third of the original document is produced as the summary to the end users which holds all the main idea and important sentences needed to be delivered through the document.

This paper brings the approach for extractive text summarization using weighted iterative ranking algorithm.

The paper on "Text Summarization using Wikipedia" [1] captures relationship between sentences with the help of wiki concepts modelling it as a bipartite graph. The variation proposed in CSUMMIT is that, the system considers nested level of relationship between sentences and concepts to improve the efficiency of the generated summary to an extent.

The rest of the paper is organized as follows. Section 2 details the related work in the summarization domain followed by Sect. 3 which elaborates the system architecture in detail.

2 Related Work

More recently, summarization has become a successful task and many studies have been taken on that. The approaches are majorly classified as supervised and unsupervised learning approach whereas the latter is focused more in the new summarization algorithms.

A graph-based approach LexRank [2], where the salience of the sentence is determined by the concept of Eigenvector centrality. The sentences in the document are represented as a graph and the edges between the sentences represent weighted cosine similarity values. The sentences are clustered into groups based on their similarity measures and then the sentences are ranked based on their LexRank scores similar to PageRank algorithm [3] except that the similarity graph is undirected in LexRank method. The method outperforms earlier versions of lead and centroid based approaches.

In paper [4–6] fuzzy logic approach is used for automatic text summarization which is based on the feature selection and feature extraction. The sentences are ranked based on the fuzzy logic scoring which is obtained by applying fuzzy rule based. The summary is generated by ordering the ranked sentences in the order they occur in the original document to maintain coherency.

In concept-based approach, the concepts are extracted for a piece of text from external knowledge base such HowNet [7] and Wikipedia [8]. In the methodology proposed in [7], the importance of sentences are calculated based on the concepts retrieved from HowNet instead of words. A conceptual vector model is built to obtain a rough summarization and similarity measures are calculated between the sentences to reduce redundancy in the final summary.

An algebraic-statistical method Latent Semantic Analysis (LSA) [9, 10] is used where hidden semantic structures of words and sentences and popularly used in text summarization task are extracted. It is an unsupervised approach that does not need any sort of training or external knowledge. LSA captures the context of the input document and extracts information such as words that frequently occur together and words that are commonly seen in different sentences. A high number of common words amongst the sentences indicate that the sentences are semantically related.

Dharmendra Hingu, Deep Shah and Sandeep S. Udmale proposed an extractive approach [11] for summarizing the Wikipedia articles by identifying the text features and scoring the sentences accordingly incorporating neural network model [12]. The preprocessed passage is sent to the feature extraction steps, which is based on multiple features of sentences and words. The scores obtained after the feature extraction are fed to the neural network, which produces a single value as output score, signifying the importance of the sentences. Usage of the words and sentences is not considered while assigning the weights which results in less accuracy.

Conditional random fields (CRF) [13] are used to identify and extract correct features to determine the important sentence of the given text. CRF segmentation assigns a label sequence to each token based on the training set. The goal of the proposed approach is to classify the sentences based on the patterns to segments. The main advantage of the method is that it is able to identify correct features and provides better representation of sentences and groups terms appropriately into its segments.

3 System Architecture

The system focuses to build a summarization system using graph based approach which employs Wikipedia concepts to determine the key sentences using weighted iterative ranking algorithm based on variation of HITS algorithm. Generalized bipartite graph framework with inclusion of concepts ensures coverage, the use of nested level relationship between sentence and concepts aids in better capturing of information and weighted iterative ranking algorithm promotes coherency.

A. *Pre-processing*

The original source document is given as input to the system. The input document is divided into meaningful units. The sentences are tokenized and are produced as output. The tokenized sentences are further processed for removal of stop words. The words with low semantic content are termed as stop words. These words do not contribute in identifying the important sentences in a text for example prepositions, articles, etc. These noisy terms are very common within a text and can be removed are stemmed. Stemming is a process of reducing the words with the same root or stem to a common form. This removal can be done with the help of maintaining a database of stop words. Further the sentences are stemmed. Stemming is a process of reducing the words with the same root or stem to a common form. This is done by eliminating the variable suffixes. The preprocessing is followed by other steps as shown in Fig. 1 which shows the work flow of the proposed system.

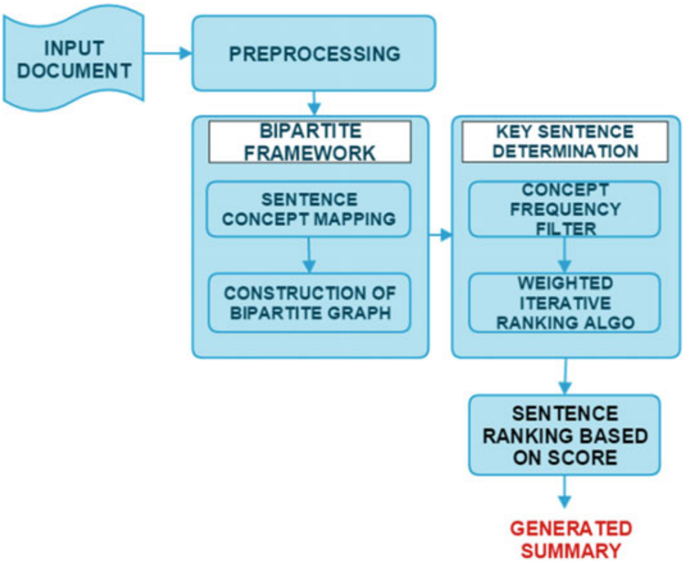


Fig. 1 Work flow of the proposed system

B. *Sentence-Concept mapping*

The pre-processed text document is given as query to the mapper. The mapper will map these queries with the corresponding Wikipedia concepts. Wikipedia article titles (concepts) are extracted from the results to the query ("hits" in Lucene terminology). The entire Wikipedia corpus is indexed using the Lucene engine. The Wikipedia dump is pre-processed to remove XML tags and other unnecessary information such as talk pages, comment section and edit history information. The cleaned Wikipedia dump is indexed using Lucene engine. The text and title fields of Wikipedia dump are indexed to make query search faster. The concepts (Wikipedia title name) is extracted for each pre-processed sentence, i.e. "query" and top concepts are retained for each query. The relationship between sentences is captured by higher degree of overlap of concepts among sentences. In most news articles, there exists overlap in concepts between sentences which conveys the relationship between two different sentences. For example, if two sentences are mapped to average number of same concepts, then they are more related to each other. The relationship can be captured with the help of bipartite graph data structure which consists of two disjoint set of nodes namely sentence node and concept nodes.

C. *Construction of Bipartite graph*

The sentence–concept mapping is represented as weighted bipartite graph- two distinct set of nodes representing sentences and concepts respectively. The bipartite graph consists of two sets of nodes (i) sentence node (ii) concept node and weighted edge defines the relationship between sentence and concept with the weight indicating the degree of correlation (The weight mentioned here refers to "Lucene hit score"). An edge exists from a sentence node to concept node if the concept is one amongst the extracted concept from Wikipedia. The graph is modelled as many-many mapping since one sentence can map to "n" number of concepts and a concept node can map to "m" number of sentences, which shows there is possibly overlap in concepts among sentences which helps in measuring the most related sentences to maintain coherency. The model also captures nested level of relations, i.e. when two sentences are related (Two sentences are related when they are mapped to higher number of same set of concepts), an edge can be simulated between a concept and a sentence to which it is not mapped initially based on transitive relationship between the corresponding related sentences and concepts that they are mapped to. This transitive dependence captures the second level of relationship between the sentence and concepts. The graph-based modelling ensures that the relationship among sentences is well captured and application of ranking algorithms to graph models are more efficient Fig. 2.

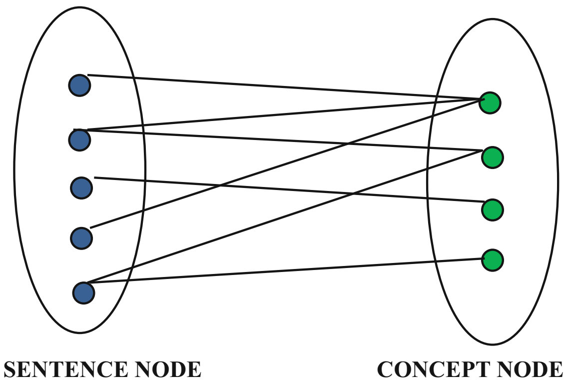


Fig. 2 Example of bipartite graph consisting of two sets of disjoint nodes U, a set of sentence nodes and V, a set of concept nodes

Algorithm1 Construction of bi-partite graph

Input: Sentence-Concept mapping C
Output: Bi-partite graph G
for each sentence s_i in source document
    add a node s_i to sentence set N
end for
for each concept c_j in concept class
    add a node c_j to concept set M
end for
for each mapping in C from s_i to c_j
    add an undirected edge between s_i and c_j nodes in graph
    $edge_i$.weight= Lucene_hit_score;
end for

D. *Key Sentence Determination*

The main task of summary generation is to select key summary sentences that form a part of the summary. The sentences are selected based on the scores associated with the sentence. An iterative ranking algorithm is proposed to calculate sentence-concept score and rank the sentence based on their score which is helpful in identifying summary sentences.

1. *Concept frequency filter-Sentence Filtering*

A simple heuristic to filter important sentences is to rank the concepts in descending order based on their frequency. The concept-frequency score is used only to eliminate those sentences that do not contribute to the summary by any means and it acts as a filtering technique rather than a ranking method. The frequency here refers to the number of sentences that maps on to a particular concept. It is considered that more sentences the concepts maps to, it becomes important and core concept of the article since may sentences correspond to that concept. The sentences that maps on to the highest ranked concepts may contribute to the final summary. This simple heuristic however does not distinguish between summary and non-summary sentences. The heuristic is used only to eliminate those sentences that maps only to a low ranked concept or no concept. The importance of a concept cannot be determined only with its frequency. The bipartite graph is updated after elimination of sentences that cannot contribute to the final summary. The importance of sentence has to be incorporated along with the concept to find summary sentences. Thus, an iterative ranking algorithm is proposed to mutually calculate sentence-concept scores which aid in selecting summary sentences. The pseudo code is presented in Algorithm 2.

Algorithm 2 Concept-frequency filter

```
Input: Bipartite-graph G
Output: Updated Bi-partite graph G after elimination of
            certain sentences
for each concept node c_j in G do
    c_j.frequency_score ← c_j .no_of_edges
end for
Sort frequency score of concept in descending order
for each sentence node s_j in G do
    if s_i .edgelist ∊ {∅} ∨ s_i .edgelist ∊ {c_l } then
        Delete sentence node s_i from graph G
    end if
end for
```

2. *Weighted iterative ranking algorithm*

The main goal of Sentence Ranker is to rank sentence nodes in bipartite graph G in descending order of their importance. The importance of a sentence is tied to the concept and vice versa. Thus, the sentence-concept score is mutually calculated. The basic idea of algorithm is based on HITS algorithm [14] which works iteratively in mutually reinforcing manner to rank web pages based on authority-hub scores. A score is associated to each concept and sentence node in graph G iteratively. The iterative update is done for K times where K is determined based on convergence property. The sentence-scores are normalized after each iterative update to prevent them from exceeding without bound. From observation and analysis, it is noted that sentence scores are steady within 5–10 iterations. The ranking is done only once for all sentence nodes in graph G and global ranking is saved permanently. The values of g_{ij} and h_{ij} in Algorithm 3 refers to the Lucene hit scores for forward mapping (Sentence to concept) and backward mapping (Concept to Sentence) respectively. Refer Algorithm 3 for pseudo code. The following equations are used in Algorithm 3.

$$s_j^{(k+1)} = \sum_{i \in Nj} g_{ij} c_i^{(k)}, \forall j \in M \tag{1}$$

$$c_i^{(k+1)} = \sum_{j \in Mi} h_{ij} s_j^{(k)}, \forall i \in N \tag{2}$$

where $s_i^{(k)}$ represents sentence score and $y_j^{(k)}$ represents the sentence score and concept score after the kth update and concept score after kth update and $x_i^{(0)}$ is initialized to $1/\sqrt{n}$. Normalise the sentence score after each iteration, so that scores doesnt exceed the bounds as given in Eq. (3)

$$\sum_{i \in N} \left(x_i^{(k)} \right)^2 = 1 \tag{3}$$

The sentence are ranked in descending order in descending order of $s_j^{(K)}$ as given in Eq. (4)

$$r = \arg\left(\text{descend}\left(x_1^{(k)} x_2^{(k)} \ldots\ldots x_n^{(k)}\right)\right) \tag{4}$$

The sentence score ranked in descending order is used to generate the final summary.

E. *Generation of Summary*

The summary is generated by selecting d leading sentences according to their rank. Let r denote set of sentences {s_{1},s_{2},s_{3}….s_{n}} where the indices denotes the rank of the sentence and d < n as in Eq. (4). The value of d is selected such that the summary generated is approximately nearer to one-fourth of the document. In practice, word-based summaries are required where word-size is set default to 50-word or 100-word summary. Since, the sentences are ranked using iterative sentence ranker, word-based summaries are produced by approximating the summary to the nearest sentence delimiter. Thus, both sentence-based and word-based summary can be generated where d can either be set as default or calculated dynamically based on the number of sentences in the article. The short summary covering major concepts in the article is presented to the user.

4 Experimental Results and Analysis

The system has been tested against the standard DUC 2002 dataset provided by National Institute of Standards and Technology (NIST) [15]. The DUC 2002 dataset consist of about 567 English NEWS articles. DUC is the most commonly and frequently used dataset for summarization task.

The most commonly used evaluation metrics in summarization domain are ROUGE metric, precision, recall and f-measure scores. The ROUGE evaluation approach depends on n-gram co-occurrence between the reference summary (i.e. ideal summary) and the machine generated extractive summary. ROUGE-N is computed as follows

Algorithm 3 Weighted Iterative Sentence Ranker

```
Input: Bipartite-graph G
Output: Sentence rank
Initialize si .score=1/√n for k=0
for k ϵ {0,...,K-1} do
    normalise ← 0
    for each concept cj in G do
        cj.score ← 0
        for each sentence si in cj.edgelist do
            cj.score=cj.score*gij +si.score
        end for
    end for
    for each sentence si in G do
        si.score ← 0
        for each concept cj in si .edgelist do
            si.score=si.score*hij + cj.score
        end for
        normalise+=square(si)
    end for
    normalise=sqrt(normalise)
    for each sentence si in G do
        si.score=si.score/normalise
    end for
end for
Rank sentences in descending order of their sentence rank
```

$$\textbf{ROUGE}-\textbf{N}=\frac{\sum_{S\in \text{reference\_summaries}}\sum_{\text{N-grams}}\text{Count}_{\text{match}}(\text{N-gram})}{\sum_{S\in \text{reference\_summaries}}\sum_{\text{N-grams}}\text{Count}\ (\text{N-gram})}$$

where n stands for the length of the n-gram, **Count$_{\text{match}}$ (N-gram)** is the maximum number of n-grams co-occurring in a machine generated summary and the ideal summary, **Count (N-gram)** is the number of N-grams in the ideal summary. Since ROUGE -1 scores are not sufficient enough to distinguish different summarizers extended set of evaluation metrics such as precision, recall, f-measure are calculated.

$$\text{Recall}=\frac{|S_{\text{ref}}\cap\ S_{\text{cand}}}{|S_{\text{ref}}|}$$

$$\text{Precision}\ =\frac{|S_{\text{ref}}\cap S_{\text{cand}}|}{|S_{\text{cand}}|}$$

Where, $S_{\text{ref}}\ \cap\ S_{\text{cand}}$ indicates the number of sentences that co-occur in both reference and candidate summaries.

The CSUMMIT summarizer system is compared against MS WORD summarizer which shows greater ROUGE- 2 scores compared to MS summarizer and the corresponding results are shown in Table 1. The visual interpretation of the result is shown in Figs. 3, 4 and 5 shows the implementation snapshot of CSUMMIT summarizer.

Table 1 Evaluation results

Summarizer	Rouge-1	Rouge-2	Precision	Recall	F measure
CSUMMIT	0.47	0.25	0.60	0.52	0.55
MS WORD	0.47	0.16	0.36	0.39	0.37

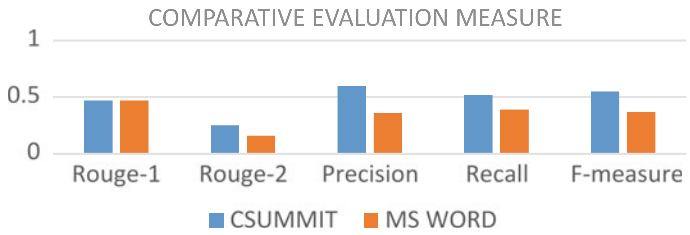


Fig. 3 Comparative analysis of evaluation metrics

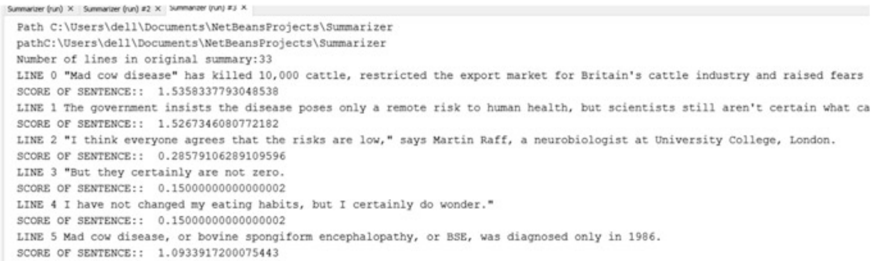


Fig. 4 Example snapshot indicating sentences along with its scores after applying weighted iterative ranking algorithm

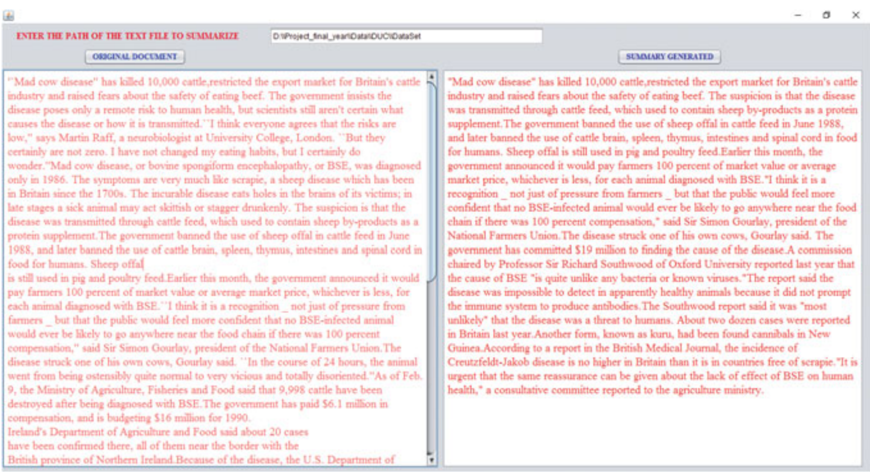


Fig. 5 Example snapshot of summarized text for the original document from DUC 2002 dataset

5 Conclusion

Wikipedia- knowledge base generated by human is employed to identify the important sentences in the given input document since it is used to identify the salient topics. This paper combines the bipartite graph framework with weighted iterative ranking algorithm to determine the key sentences. The bipartite graph is extended to capture the nested levels of relationship between sentences and concepts which ensures a higher level of efficiency in the generated summary. The weighted iterative ranking algorithm which uses weighted graph generates the extractive summary. CSUMMIT summarizer shows improved performance compared to the baseline summarizers.

The system can be extended for multiple documents and can be made domain specific to match the requirements of the user.

References

1. Sankarasubramaniam, Yogesh, Krishnan Ramanathan, and Subhankar Ghosh. "Text summarization using Wikipedia." *Information Processing & Management* 50.3 (2014): 443–461.
2. Erkan, Günes, and Dragomir R. Radev. "LexRank: Graph-based lexical centrality as salience in text summarization." *Journal of Artificial Intelligence Research* (2004): 457–479.
3. Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, Stanford, CA.
4. Farshad Kyoomarsi, Hamid Khosravi, Esfandiar Eslami and Pooya Khosravyan Dehkordy, "Optimizing Text Summarization Based on Fuzzy Logic", In proceedings of Seventh IEEE/ACIS International Conference on Computer and Information Science, IEEE, University of Shahid Bahonar Kerman, UK, 347–352, 2008.
5. Ladda Suanmali, Mohammed Salem, Binwahlan and Naomie Salim, "Sentence Features Fusion for Text summarization using Fuzzy Logic, IEEE, 142–145, 2009.
6. Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan, "Fuzzy Logic Based Method for Improving Text Summarization", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 2, No.1, 2009.
7. Meng Wang, Xiaorong Wang and Chao Xu, "An Approach to Concept Oriented Text Summarization", in Proceedings of ISCIT'05, IEEE international conference, China, 1290–1293, 2005.
8. Ramanathan, Krishnan, et al. "Document summarization using Wikipedia."*Proceedings of the First International Conference on Intelligent Human Computer Interaction*. Springer India, 2009.
9. Ozsoy, Makbule Gulcin, Ferda Nur Alpaslan, and Ilyas Cicekli. "Text summarization using latent semantic analysis." *Journal of Information Science* 37.4 (2011): 405–417.
10. Mashechkin, I. V., et al. "Automatic text summarization using latent semantic analysis." *Programming and Computer Software* 37.6 (2011): 299–305.
11. Hingu, Dharmendra, Deep Shah, and Sandeep S. Udmale. "Automatic text summarization of Wikipedia articles." *Communication, Information & Computing Technology (ICCICT), 2015 International Conference on*. IEEE, 2015.

12. Khosrow Kaikhah, “Automatic Text Summarization with Neural Networks”, in Proceedings of second international Conference on intelligent systems, IEEE, 40–44, Texas, USA, June 2004.
13. Batcha, Nowshath K., Normaziah A. Aziz, and Sharil I. Shafie. “CRF based feature extraction applied for supervised automatic text summarization.”*Procedia Technology* 11 (2013): 426–436.
14. Kleinberg, Jon M. “Authoritative sources in a hyperlinked environment.”*Journal of the ACM (JACM)* 46.5 (1999): 604–632.
15. Document Understanding Conference (DUC), 2002. http://tides.nist.gov/.

Kannada Speech Recognition Using Decision Tree Based Clustering

Sharada C. Sajjan and C. Vijaya

Abstract This paper presents the recognition of Kannada language by employing decision tree based clustering to build context dependent triphone HMMs. Language-specific knowledge is required to generate phonetic questions required for clustering triphones. The success of clustering depends mainly on correctly identifying the parameters that are correlated and are grouped. By using phonetic table of the target language and universal list of articulatory questions, language specific articulatory question list is created which can then be used to cluster triphones. This method of triphone clustering requires least human intervention. Perplexity is also measured for the database of Kannada language and the database having lower perplexity is used for testing performance of the generated triphones. It is observed that, triphones which are clustered using universal list of articulatory questions performs well compared with manually created phonetic question list.

Keywords Acoustic model · Language model · Clustering · Phonetic questions Triphones · Perplexity

1 Introduction

Most of the speech recognition systems employ statistical approach, i.e., Hidden Morkov Model (HMM) to represent speech parameters and apply Viterbi search algorithm to recognize the unknown pattern [1]. This statistical approach does not require much information about linguistic knowledge. Once the training database is available, i.e., acoustics along with its transcriptions, the system can be trained to build models. Since speech is context dependent, all recognition systems have

S.C. Sajjan (✉) · C. Vijaya
Department of Electronics and Communication Engineering,
SDM College of Engineering and Technology, Dharwad, Karnataka, India
e-mail: sharadasajjan@yahoo.com

C. Vijaya
e-mail: vijayac26@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_15

moved to triphone modeling [2]. Drawback of this approach is that this requires a large number of triphones to be trained. No single training corpus can possibly contain such a large number of units. Another problem with huge amount of triphone list is the requirement of large memory space to store. This increases the search space and the hence time of execution which makes it nonreal time. To alleviate these problems triphones, with similar acoustics have been grouped together called clustering and the phones in the same group share the same set of parameters [3]. The success of clustering mainly depends on correctly identifying the triphones that are correlated and are grouped together.

There are two methods of clustering triphones called data-driven and decision tree based. In data-driven method, triphones that produce similar speech features are clustered. The disadvantage is that the triphones which are not seen in training database cannot be recognized. Decision tree based clustering allows parameters having similar acoustics to be grouped as a single unit. It provides a solution to unseen triphones also [4]. The level of similarity is mainly determined by phonetic decision trees built by the human expert. He should have thorough knowledge about the language [5]. Building phonetic decision tree involves a set of phonetic questions to be defined by human expert. He should determine whether left or right context of a phone is a vowel or consonant and if consonant, whether its left or right is a nasal and so on [6]. Recently efforts have been made to reduce human intervention in clustering triphones leading to automatic language independent triphone training [7]. The same has been referred here for the purpose of creating phonetic question list for the target language Kannada. This method of generating phonetic questions which requires least human intervention gives better recognition rate compared with human effort in framing questions.

Database preparation which plays an important role in recognition performance is also discussed. Training and test database is prepared for Kannada language, perplexity which is a measure of corpus quality is measured and database with lower perplexity is used to test the recognition performance of the system.

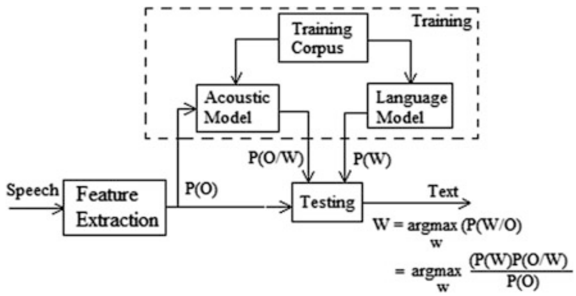
2 System Overview

Given an acoustic observation O of the test utterance, word sequence W is determined by using Bayes theorem given by

$$W = arg\max_{w}(P(W|O) = arg\max_{w}\left(\frac{P(W)P(O|W)}{P(O)}\right) \tag{1}$$

where $P(W)$ is priori probability of word string estimated by n-gram language model, $P(O/W)$ is an acoustic observation of the word string, the information of which is stored in an acoustic model, $P(O)$ is an acoustic observation of test utterance. To find unknown word sequence, maximization of numerator given in Eq. (1) is performed, since the denominator term $P(O)$ is fixed as it is of the test

Fig. 1 General architecture of continuous speech recognition



sentence. Figure 1 shows the general architecture for continuous speech recognition. It mainly consists of three modules, feature extraction, training and testing or decoding.

From the literature survey it is observed that the MFCC representation of speech provides better feature extraction tool for speech recognition applications. The recognition rate also depends on the quality of corpus used, language model and an acoustic model. Triphone based acoustic modeling is used and to cluster triphones phonetic question list for target language is required. In this paper, estimating corpus quality and preparation of phonetic question list automatically for Kannada language is discussed. This phonetic question list is later used to cluster context independent triphone HMMs to represent final models.

3 Language Model

Language model is used to predict each symbol in a sequence given its n–1 predecessors. It defines syntax of a language. The probability of a word sequence W is given by

$$\begin{aligned} P(W) &= P(w_1, w_2, \ldots, w_n) \\ &= P(w_1)P(w_2/w_1)P(w_3/w_1w_2)\ldots(w_n/w_1w_2\ldots w_{n-1}) \end{aligned} \tag{2}$$

Most of the words do not depend on large history because the language is highly structured. A simple approach is the bigram language model which uses only one step history, i.e.,

$$P(W) = P(w_1)P(w_2/w_1)P(w_3/w_2)\ldots P(w_m/w_{m-1}) \tag{3}$$

Suppose the training corpus has missing words, then the estimated probability would become zero. To overcome this, a small portion of probability distribution is reserved for out of vocabulary words.

Another way to overcome data sparseness is smoothing by interpolating unigram, bigram relative frequencies.

$$P(w_i/w_{i-1}) = \lambda_1 P(w_i) + \lambda_2 P(w_i/w_{i-1}) \tag{4}$$

where the weights satisfy $\lambda_1 + \lambda_2 = 1$.

Third method to make up for lack of training data used in many speech recognisers is backing-off, i.e., if the system does not have enough evidence for bigrams it should be backed off to unigram probability scaled by a back-off weight. When a back-off bigram is used, backed-off transitions can share a common loop-back transition.

3.1 Backed-off Bigram Language Model

Bigram probability is computed by

$$\begin{aligned} P(i,j) &= \frac{N(i,j) - D}{N(i)} & \quad & \textit{if } N(ti,j) > \; t \\ &= b(i)p(j) & & \textit{otherwise} \end{aligned} \tag{5}$$

where, $N(i, j)$ is the number of times the word j follows word i, $N(i)$ is the number of times the word i appears. D is the discount or backed-off weight generally set equal to 0.5. When the bigram count falls below the threshold t, the bigram is backed off to unigram scaled by a backed-off weight $b(i)$ [6]. In this work, backed-off bigram language model is used.

3.2 Perplexity

It is a measure of an average of how many different most probable words can follow any given word [6]. It is related to entropy to assess the actual performance of a language model. Perplexity is given by

$$\text{Perplexity} = \dot{P}(w_1, w_2, \ldots, w_m)^{-\frac{1}{m}} \tag{6}$$

where $\dot{P}(w_1, w_2, \ldots, w_m)$ is the probability estimate assigned to the word sequence $w_1, w_2, \ldots, w_m$ by a language model. Evaluation of perplexity involves scanning of training set and computation of n-grams and their probabilities. These are stored in language model files. Goodness of a language model is estimated by using it to compute perplexity on unseen test set. Better the language model then lower is its test set perplexity.

4 Acoustic Model

Acoustic variations of speech signal are represented by an acoustic model. Since speech is context dependent, context-dependent triphone HMMs are build [8]. To overcome data sparseness and to reduce the number of triphones, triphones are clustered using decision tree based method. Initially monophone HMMs are built using training acoustics and their transcriptions. Retraining is then performed using embedded Baum-Welch reestimation [9]. Next, context-independent triphone HMMs are built by cloning monophones which will generate very large number of triphones. Triphones having similar acoustics are then clustered which will place triphones having phonetic similarity in one group. This will reduce number of triphones. Phonetic decision tree based algorithm is used for clustering which will tie the states based on phonetic similarity. This requires thorough knowledge about acoustic characteristics of target language to frame phonetic questions. Figure 2 explains the procedure for generating phonetic questions for the target language. The articulatory characteristics are same for all linguistics and hence can be shared and are referred as the universal articulatory based question list. Universal question list presented in [7] is referred, samples of which are shown in Table 3 of Sect. 5. Monophones of the target language are listed and each question of the universal question list is transformed into a language specific question. Such of those universal questions which are not applicable to the target language are discarded. The final step uses the target language specific question list created in above step to cluster unclustered triphones to create clustered triphones. The triphones so obtained are retrained using the entire database. It was observed that this unsupervised training procedure outperformed the handcrafted training method in creating phonetic question list.

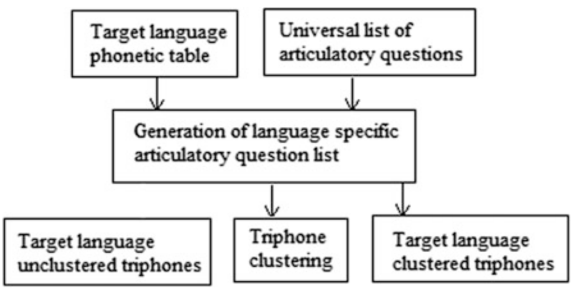


Fig. 2 Generation of phonetic questions of target language

Table 1 Measure of perplexity

Training database used	Test database used	Perplexity measure
Text1_train(story)	Text1_test(story)	95.48
Text1_train (story)	Text2_test(news)	324.34
Text2_train(news)	Text1_test(story)	296.93
Text2_test (news)	Text2_test(news)	102.31

5 Implementation

Kannada speech is recorded at sampling rate of 16 kHz with 16 bits per sample value using audacity software. Text 1 consists of 20 stories from Ramayan book as training utterances and two stories from the same book other than training data as test utterances. Text 2 database is from news having 200 sentences as train utterances and 50 sentences as test utterances. Perplexity is measured as shown in Table 1.

It is observed from the table that the perplexity is low if both train and test database are from same text source as expected since the probable of n-gram word sequences occurring in test set is more if training database is from same text source. The dataset having lower perplexity is used for testing recognition performance.

Monophone list of target language which generally includes alphabets of a language is created. Table 2 shows the sample entries of phoneme list of target language Kannada. Following steps are performed to generate clustered triphone HMMs.

1. Initially, using training transcriptions and their acoustics, three-state left–right monophone single Gaussian HMMs are build and retrained using Baum–Welch reestimation.
2. These monophone HMMs are cloned to generate untied context-dependent triphone HMMs and retrained using Baum–Welch reestimation.
3. States of triphones having same center phone having phonetic similarity are clustered. In each cluster, a typical state is chosen and all cluster members are tied to this state.
4. The number of mixture components is then increased and the models are reestimated until desired number of mixture components is reached.

Step 3 uses decision tree based clustering method explained in Sect. 5 to cluster triphone HMMs which uses target language phonetic table samples of which is shown in Table 2 and universal list of articulatory questions, sample list shown in Table 3. Target language phonetic table lists all the basic sounds of Kannada language along with their articulatory characteristics. To have less human intervention, universal list of articulatory questions given in [7] referred. Not applicable questions to the target language in the universal list are simply discarded. The triphone HMMs so generated are retrained using Baum–Welch reestimation. Then mixture splitting [6] is performed successively up to eight Gaussians to generate final models which are stored as templates. Viterbi decoding algorithm is then used for testing the unknown utterance. Table 4 shows word error rate for the two different methods of clustering triphones, one with manually created question list and other by referring to universal question list. It is observed that triphone HMMs clustered using universal question list lead to 15.26% word error rate versus a word error rate of 21.34% for the manually trained models which again depends on the judgment by human expert.

Table 2 Samples of target language phonetic table

Vowels and diphthongs									
Phone	Class	Position of jaw	Position of articulation	Vowel type	Continuance	Rounding of lips	Length of phone	Tension in cheeks	
ಈ (I)	vowel	high	front	voiced	cont	unround	long	lax	
ಇ (i)	vowel	high	front	voiced	cont	unround	short	tense	
ಊ (U)	vowel	high	back	voiced	cont	round	long	lax	
ಉ (u)	vowel	high	back	voiced	cont	round	short	tense	
ಅ (a)	vowel	mid	front	voiced	cont	unround	short	tense	
ಐ (ai)	dipht			voiced	ncont				
Consonants and semivowels									
Phone	Class	Manner of articulation	Position of articulation	Point of articulation	Voicing	Continuance	Position of tongue	Zone of articulation	Stridency
ಪ್ (p)	cons	stop	front	bilabial	unv	ncont	ncoron	ant	unst
ಬ್ (b)	cons	stop	front	labial	voiced	ncont	ncoron	ant	unst
ಸ್ (s)	cons	fricate	central	alveolar	unv	cont	coron	nant	strid
ಮ್ (m)	cons	nasal	front	bilabial	voiced	cont	ncoron	ant	unst
ಯ್ (y)	semiv	glide	central	alveopalatal	voiced	ncont	coron	nant	unst
ನ್ (n)	cons	nasal	central	alveopalatal	voiced	cont	coron	nant	unst

Table 3 Samples from universal list of articulatory questions [7]

Question name	Question definition	Question name	Question definition
Boundary	SILENCE	Unvcd_closure	CLOSURE & UNVOINCED
Vcd_Closure	CLOSURE & VOICED	Closure	CLOSURE
Front_stop	FRONT & STOP	Central_stop	CENTRAL & STOP
Back_stop	BACK & STOP	Affricate	AFFRICATE
Stop	STOP	Vcd_stop	VOICED & STOP
Unvcd_stop	UNVOICED & STOP	Nasal	NASAL
Fric	FRICATIVE, AFFRICATE	Fricative	FRICATIVE
Voiced_Fric	(FRICATIVE, AFFRICATE) & VOICED	Vowel	VOWEL

Table 4 Word error rate for different methods of creating phonetic questions

Clustering untied triphones using	% WER
Manually created phonetic questions	21.34
Phonetic questionscreated using universal list of articulatory questions	15.26

6 Conclusion

This paper presents a methodology for generating phonetic questions for Kannada language which are used for clustering triphones. Generally such question sets were constructed manually based on similarity judgments by a human expert which require depth knowledge in phonology of a language. Since the acoustic characteristics are independent of linguistics, most commonly accepted universal question list is referred and phonetic questions are generated for Kannada language. This provides word error rate of 15.26 as against 21.34 for manually created question list which depends on the judgment by an expert. Further the quality of database is also tested by the term perplexity. Database which provides lower perplexity is used for testing the performance of generated triphones.

Acknowledgements The authors would like to acknowledge IEI India, VTU Belagavi, management of SDME society and Principal of SDMCET, Dharwad for providing all the support and encouragement to carry out the research work.

References

1. Rabiner L. R., Juang B. H.: Fundamentals of Speech Recognition, Prentice Hall, 1993.
2. P.C. Woodland, J.J. Odell, V. Valtchev, S. J. Young.: Large Vocabulary Continuous Speech Recognition Using HTK. Proceedings of ICASSP, 1994.
3. S. J. Young and P. C. Woodland.: State clustering in HMM based continuous speech recognition, Computer Speech and Language. 1994.

4. S. J. Young, J. J. Odell, P. C. Woodland.: Tree Based State Tying for High Accuracy Acoustic Modelling.
5. Supphanat Kanokphara, Anja Geumann, Julie Carsson Berndsen.: Accessing Language Specific Linguistic Information for Triphone Model Generation: Feature Tables in a Speech Recognition System.
6. S. Young and others.: The HTK book (for HTK version 3.4), Cambridge University Engineering Department, March 2009.
7. Automatic language independent triphone training using a phonetic table, US7289958 B2, Oct 30, 2007.
8. Yunxin Zhao.: A Speaker Independent Continuous Speech Recognition System using Continuous Mixture Gaussian Density HMM of Phoneme- Sized Units, IEEE transactions on speech and audio processing.
9. Andrej Ljolje.: Development of an Acoustic-Phonetic Hidden Markov Model for Continuous Speech Recognition, IEEE Transactions on signal processing, Vol 39, No.1, 1991.

Automatic Hand Raise Detection by Analyzing the Edge Structures

J. Jesna, Athi S. Narayanan and Kamal Bijlani

Abstract Modern education is now concentrated on E-learning, possible of learners being in virtual classroom and geographically distributed. Interactivity is the important part of design of such learning management systems. Better designs will reduce the isolation feeling among the tutors and students because they are at different areas in E-learning. Hand raise is a typical interaction that can happen in a real classroom. So this paper aims to design an automatic system that can detect the hand raise in E-learning classroom and these will keep the synchronization of classroom much better. The proposed method detects the hand raise by analyzing the edge structures of hand. The system will first find out the region of interest of each student in the classroom and find the edges from the skin-colored region of ROI of each student. Finally by doing some morphological operations find out the hand-raised regions. Without using any classification or object detection complex methods the proposed method detects the hand raise in lesser time with 91% accuracy.

Keywords Hand raise detection · E-learning · Skin detection · Steerable gaussian filter · ROI

1 Introduction

In an E-learning scenario, the students and the lecture may be in different places, so we want to reduce the isolation feeling among them by creating a feeling like traditional classrooms. Only at this time the E-learning course should be successful.

J. Jesna (✉) · A.S. Narayanan · K. Bijlani
Amrita e-Learning Research Lab (AERL), Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Amrita University, Amritapuri, India
e-mail: jesnaj11@gmail.com

A.S. Narayanan
e-mail: mail2athi@gmail.com

K. Bijlani
e-mail: kamal@amrita.edu

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_16

When designing such E-learning classrooms there need good interaction patterns. Hand raise is a type of an interaction pattern occurs in a traditional classroom, i.e., if any student wants to ask a doubt or interrupt a lecturer for telling his or her opinions or ideas when the lecturer taking the classes, the student can immediately raise their hand and ask his or her doubts immediately. It can be useful for conducting some kind of group activities, quizzes in classrooms, etc. If we can include such type of interactions in E-learning classroom it will be helpful for getting a feel like traditional classroom.

There are so many researches related to the hand gesture recognition. Hand gesture recognition comprises of several types of gestures and by using machine learning algorithms and models they are automatically detected. But with hand raise many types of gestures can occur, i.e., the portion of palm may vary. One can raise their hand with open palm or closed palm or half closed palms or pointing any finger. Most of the hand raise detection systems concentrate only on some typical hand gestures not all gestures. Chiang et al. [1] proposes a hand raise detection system by training three types of hand gestures. They are right hand, left hand, and normal gestures. They segment the human bodies by using some foreground extraction method and by using some scale-invariant features and machine learning algorithms detect the hand raises automatically. But their method aims only three types of gestures and if the effective segmentation of student bodies fails then the method could not work properly.

Kinect is a gesture recognition device. Some of the work related to the Kinect devices. Hariharan et al. [2] proposes a method for finding hand raised student in a video by using Kinect devices. There work concentrated on the raising hand gesture of the student. If a student raises his hand the Kinect will identify the student and by using some gesture recognition algorithm they find out the gesture of the student. Tang et al. [3] also proposes a method for hand tracking and pose estimation based on Kinect. But in E-learning scenario keeping Kinect devices in all the areas is a difficult task because students and the tutors will be in different global areas.

N. B. Bo et al. [4] proposes a hand raising gesture detection framework by locating the arm position using the geometrical edge structure of hand instead of using any object detection or complex machine learning algorithms. They find out the edge structure of hand raise by using canny edge detector and some morphological operations. But the method has so many false detections due to the back ground subjects and noises because they directly apply the edge detection algorithm with the background subjects. Our motivation to do the work is depend on this area.

The proposed method detects the hand raise including all types of gestures by analyzing the structure of straight edge from the skin colored region of ROI and then filter out the hand raise by some morphological operations and the geometrical structure of hand.

2 Block Diagram for Hand Raise Detection

The proposed method consists of four parts: finding region of interest (ROI), skin color detection, edge detection, and straight line drawing and decision-making as shown in block diagram (Fig. 1). Here the non-shaded blocks represent the important steps for detection system, blue-shaded blocks represents the important algorithms and models needed for each section and finally green-shaded block represents the detection result, i.e., whether the image contains a hand raise or not.

3 Hand Raise Detection

Proposed method is explained with the help of two different images. In one image, the person has raised hand near to the face and the other one raise the hand above the face.

A. *Finding Region Of Interest*

Select the input image and find out the faces by robust Viola et al. [5] face detector as shown in Fig. 2a, b. Then select the region of interest of each person by using the parameters of detected faces. It can be calculated as:

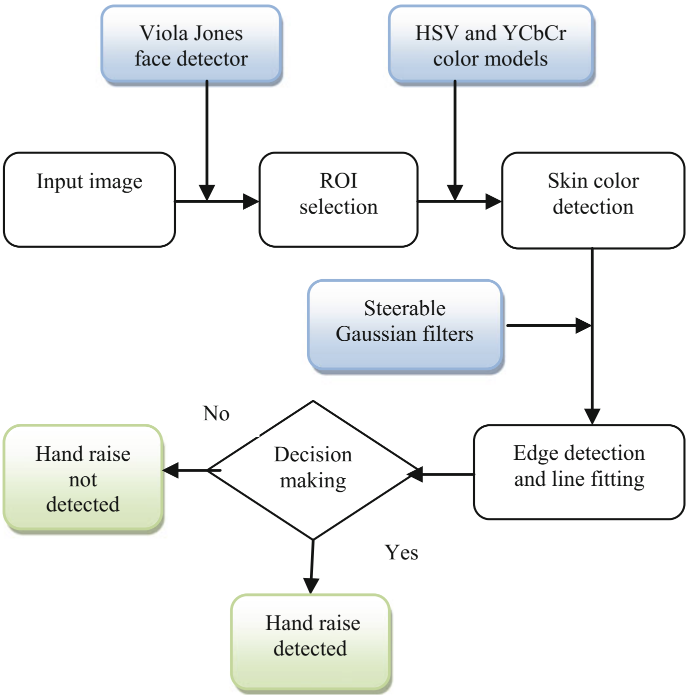


Fig. 1 Block diagram of hand raise detection system

Let W_{face} be the width of the face and H_{face} is the height of the face detected, then

$$W_{ROI} = 5 * W_{face} \quad (1)$$

$$H_{ROI} = 2 * H_{face} \quad (2)$$

where W_{ROI} and H_{ROI} be the width and height of the region of interest. The ROI selected is shown in Fig. 2c, d.

B. *Skin color detection*

From the ROI proposed system need only the skin-colored region for detecting the hand raises. Different types of color models are available for skin detection. Here combination of YCbCr and HSV color space is used for color detection [6]. The HSV color model contains hue, saturation, and value. The hue channel will defines the color itself, whereas the saturation channel defines the light combined with hue and the value determines the image brightness. The YCbCr color space includes the Y channel, which represents the luminance component, and the Cb and Cr channels describe the chrominance components. In order to find the skin color

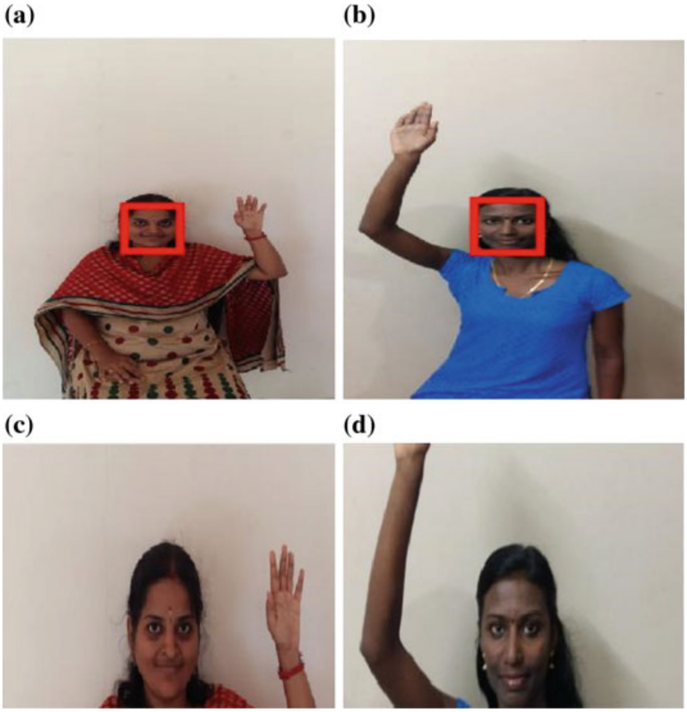


Fig. 2 ROI selection results: **a**, **b** showing results for face detection through the red rectangle. **c**, **d** showing results for ROI selection using the detected faces

regions first convert the ROI's RGB space into corresponding YCbCr and HSV spaces. Then find the skin pixels by using the range of pixels given below.

$$\begin{cases} 140 \leq Cr \leq 165 \\ 140 \leq Cb \leq 195 \, and \\ 0.01 \leq Hue \leq 0.1 \end{cases} \tag{3}$$

This method yield comparatively good results than the other color models and the output obtained is shown in the Fig. 3a, b.

C. *Edge detection and straight line fitting*

Fitting appropriate edge detection algorithm is a difficult task. For edge detection our method uses steerable Gaussian filter [7] with order 2 and 360° gradient angle. The importance of steerable filters is that it can detect the edges according to the direction we given. Our ultimate aim is to get two line segments for the raised hand region. The method proposed by N. B. Bo et al. [4] uses canny edge detector for finding the edges. Canny edge detector is not suitable for noisy images and more over it cannot segment the images according to the orientation. If we apply steerable Gaussian filter, the hand region can easily split into two regions since two regions are in different orientations as shown in Fig. 4a, b.

Then perform some morphological operations in order to filter the hand raised regions and avoid the unwanted noises from the images. From Fig. 4a, b avoid the small segmented areas and the face positioned edges. For we apply the following operations:

- Apply morphological cleaning and skeletonization to make the pixel width as one. Then perform image closing with structuring elements as shown in Fig. 4c, d. In these steps some of the large segments at face positions are split into small segments.
- Then remove the branching points to get perfect splitting of hand regions. The output obtained after this step is shown in Fig. 4e, f.

(a) (b)

Fig. 3 Skin detection results: **a** result for skin detection in hand raised near to the face and **b** results for skin detection in hand raised above the head

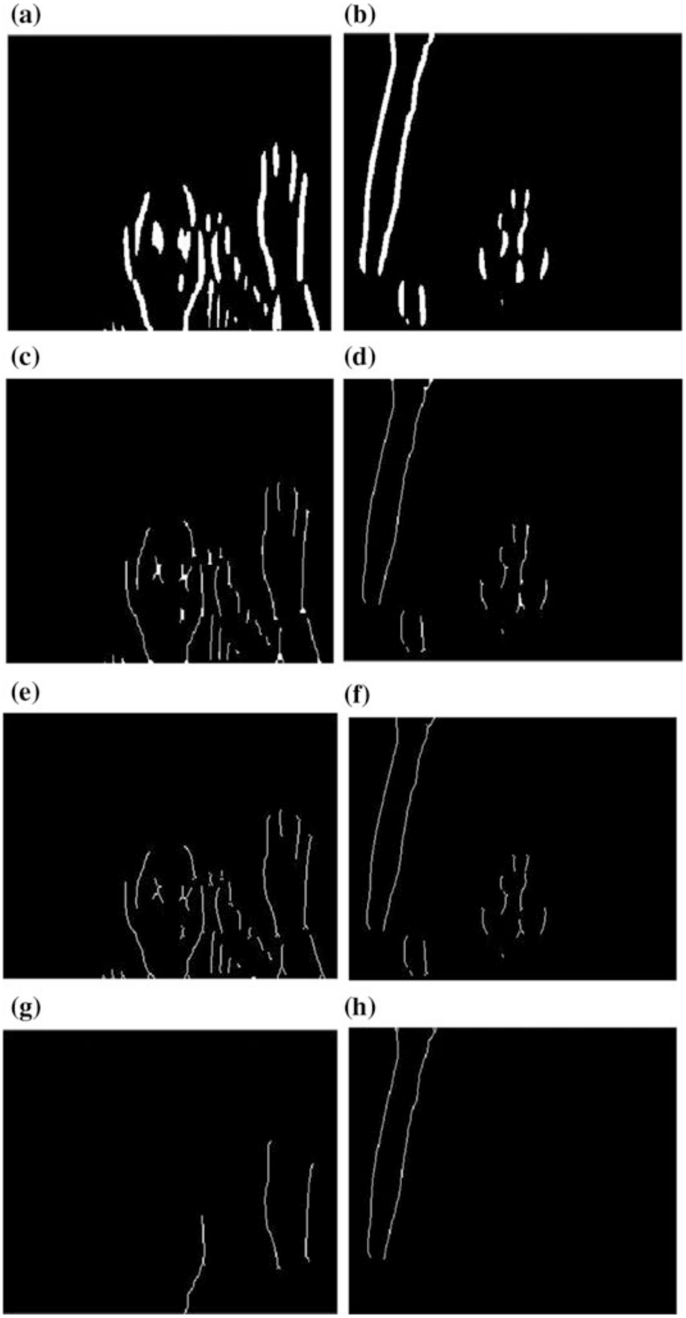


Fig. 4 Result after applying steerable Guassian filter in both images: In **a** we are aim to get the palm position of the raised hand and in **b** we aims to get the hand region of the raised hand. **c** and **d** shows result after cleaning and skeletonization and image closing morphological operations. **e** and **f** shows result after branching points avoided, **g** and **h** shows result after avoiding small segmented areas. **i** and **j** showing line fitting results using *red lines*

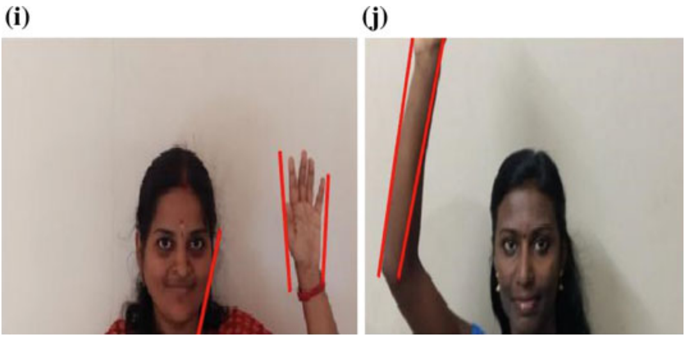


Fig. 4 (continued)

- Finally remove the edge segments with area less than 50 as shown in Fig. 4g, h. This step will avoid the unwanted segments from the image so that the final segment includes one of the following:

  a. Two segments related to hand region or
  b. Three or four segments related to hand and face region.

After the above steps any simple line fitting algorithm can be used for drawing the lines through the final edge segments [8]. The detected lines are shown in Fig. 4i, j.

D. Decision-making

Finally the processed image contains maximum of four lines including the hand-raised portions and face region. From this our work will filter only two lines that indicate the hand raise. These two lines are nearly slopped and distance between them is very low. Using these criteria the decision will be made. The condition we applied is:

- Find two nearly slopped lines such that their slope difference is less than 25 and
- Find the smallest distance between these two line segments. If the smallest distance is between 25 and 55 then these two lines represent the portion of hand raise. Otherwise hand raise is not detected in the mage.

The hand raise detected using this criterion is as shown in Fig. 5a, b.

4 Results and Analysis

Our method is tested with the images containing six different students and analyzes the work using different poses of hand raises including left hand, right hand with their different orientations and gestures. The detailed description of the test results is shown in Table 1. NH, LH, RH stands for number of non-raised hands, left hand

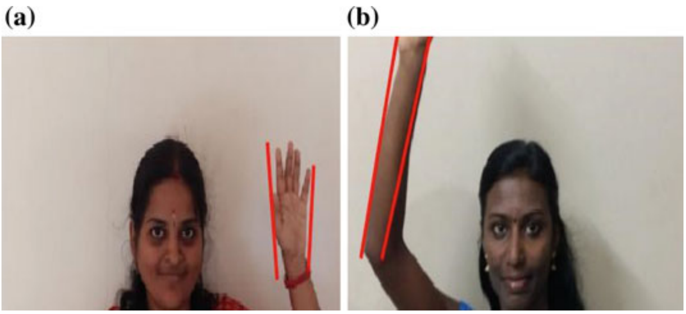


Fig. 5 Final results for hand raise detection. In **a** the raised hand is detected by avoiding the line segments of face by using criteria defined in the decision-making step and in **b** only need to check the criterion that whether the two segments indicating a hand raise or not

Table 1 Detection results by analyzing each student's hand raise gestures

	Total # of images	NH	LH	RH	TP	FN	TN	FP
Student 1	8	3	2	3	5	0	3	0
Student 2	6	0	0	6	6	0	0	0
Student 3	21	4	10	7	16	1	2	2
Student 4	14	2	6	6	11	1	2	0
Student 5	9	2	2	5	6	1	2	0
Student 6	22	2	6	14	18	2	2	0
Total	80	13	26	41	62	5	11	2

and right hand. Similarly TP, FN, TN, FP represents number of true positives, false negatives, true negatives, and false positives, respectively.

Table 1 shows that regardless of the gestures and orientations, the proposed method gives good results for raised hand detections. The proposed method is tested with total 80 images from different students which include 67 hand raises and 13 non-hand raised images. From total 67 hand raise images there are total 62 true positive images, i.e., the hand raises detected as correctly and five false negative images indicating does not handle hand raise correctly. The method is tested with non-hand raised images also. From 13 non-hand raised images 11 of them gives correct output and two of them interpreted as wrong. The accuracy of the system can be calculated as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} * 100 \tag{4}$$

The accuracy obtained is 91.25%.

Some of the detection results are shown in Fig. 6. These example figures include hand raises with normal gestures to complex gestures. It includes images with student wearing watches and background object such as chairs, etc. The proposed detection framework provides correct results in all such cases. False detections are

Fig. 6 Result for hand raise detection in different cases. The *red line* drawn shows the detected hand raise poses. Here each row contains three examples, each of them shows the detection result corresponds to its original image

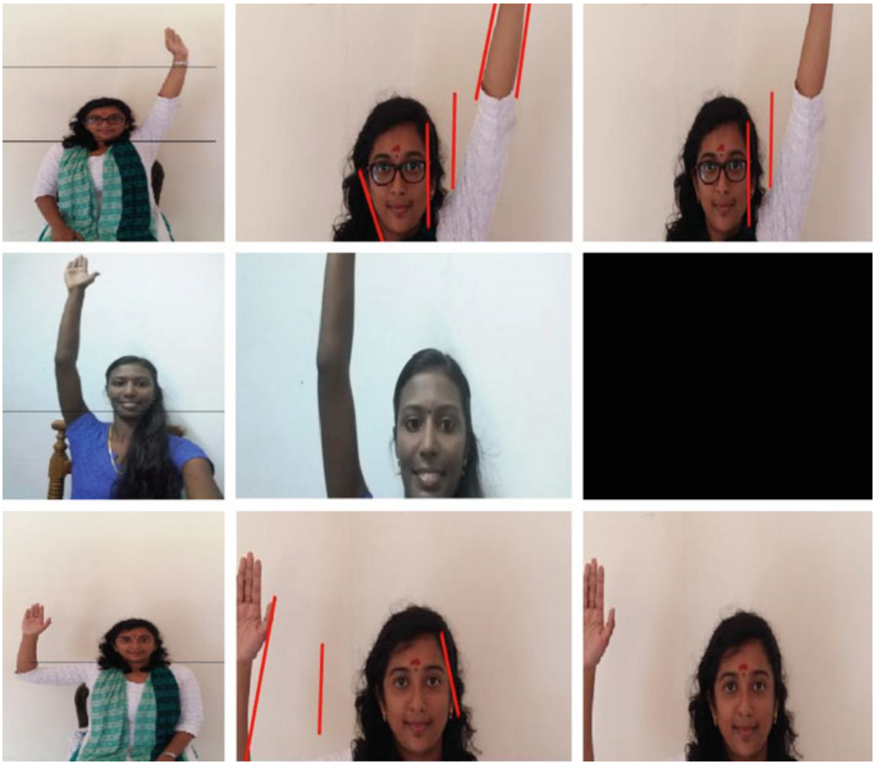


Fig. 7 Each row indicates false detections. *First row* shows false detection due to failure of edge detection. Here the system detects the hand raise by locating the false lines in the ROI as shown in *top right* image. In *second row* image the skin color detection fails due to bad light condition. *Third row* shows example of student who raised her hand outside the ROI

due to: some bad lighting condition, failure of correct edge detection and in one image the student raised her hand outside the ROI defined. Due to some very bad light conditions face detection results and skin color detection are get faulted. Some of the failure cases are shown in Fig. 7. Instead of the above failure cases the system's performance is exceptionally good.

5 Conclusion

This paper proposes an efficient method to design an automated hand raise detection system without using any complex gesture training machine learning algorithm or calculations. We evaluated the method by analyzing six different students with differing hand raise poses and lighting conditions. From these we found that the method has 91% accuracy. The method detects the hand raise including almost all types of gestures because the system concentrates only the segments of raised hand region. Skin color segments helps to locate the hand-raised region and steerable Gaussian filters help segmenting the raised hand region into two area and these will keep easier in detecting the hand raise as two lines. Since our method does not require any back ground modeling or complex calculations it can detect the hand raise in less than 5 s so that it can be implemented in real-time video sequences. Therefore our future work concentrated on evaluating the performance of the method in video sequences with more people and in a highly cluttered environment.

References

1. Chiang, Cheng-Chieh, Cheng-Chuan Tsai, and Greg C. Lee. "Vision-based Raising Hand Detection in Classroom." In *MVA*, 2013, pp. 61–64.
2. Hariharan, Balaji, S. Padmini, and Uma Gopalakrishnan. "Gesture recognition using Kinect in a virtual classroom environment." *Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on*. IEEE, 2014, pp. 118–124.
3. Tang, Cheng, Yongsheng Ou, Guolai Jiang, Qunqun Xie, and Yangsheng Xu. "Hand tracking and pose recognition via depth and color information." In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, IEEE, 2012, pp. 1104–1109.
4. Bo, Nyan Bo, et al. "Detection of a hand-raising gesture by locating the arm." *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*. IEEE, 2011, pp. 976–980.
5. Jones, Michael J., and Paul Viola. "Robust real-time object detection." *Workshop on statistical and computational theories of vision*. Vol. 266. 2001.
6. Mustafa UCAK (2007). *Face and eye detection* [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/13716-face-eye-
7. Freeman, William T., and Edward H. Adelson. "The design and use of steerable filters." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 9 (1991), pp. 891–906.
8. [Online] Available: "http://in.mathworks.com/matlabcentral/answers/89335-how-do-i-make-a-best-fit-line-along-with-getting-r-2-on-matlab"

Performance Evaluation and Improvement of R-CTP for Enhanced Packet Transmission and Latency

Jackson Preetham Machado and Pranesh V. Kallapur

Abstract Due to the stringent resource constraints and highly application-specific nature of Wireless Sensor Networks designing efficient routing protocol is a big challenge. The Collection Tree protocol based on Trickle algorithm was proposed as routing protocol for WSN. Further, several improved variants of CTP were proposed to enhance routing efficiency and R-CTP is one among them. In this paper, performance evaluation of R-CTP is presented using Castalia Simulator against chosen performance parameters. Further, this paper proposes some improvements to R-CTP to enhance performance in terms of number of transmitted packets and latency parameters. Simulation results presented in this paper show that, proposed improvements to R-CTP yields improvement in overall performance, 12% increase in transmitted packets, and 2% improvement in Latency respectively. The paper is concluded with mentioning of future directions for research.

Keywords Collection tree protocol (CTP) · Rainbow collection tree protocol (R-CTP) · Wireless sensor networks (WSNs)

1 Introduction

The Wireless Sensor Networks (WSNs) support sensing the environments, processing the data, and collecting the data from the large number of nodes. The WSN finds wide range of applications in industry, defence, and home automation. The WSN is characterized by stringent resource constraints. Such stringent resource constraints make design and implementation of WSN very difficult. Routing has been an important issue to be addressed in WSN. The aim of the routing is to deliver the packets in multi-hop communication to the destination while increasing

J.P. Machado (✉) · P.V. Kallapur
Department of ISE, NMAM Institute of Technology, Nitte, Udupi, Karnataka, India
e-mail: jacksonpreetham@gmail.com

P.V. Kallapur
e-mail: praneshvkallapur@yahoo.co.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_17

network performance. WSN topology can be classified mainly as flat and hierarchical. Many routing protocols have been proposed for addressing routing issue in both types of topologies. Collection Tree Protocol (CTP) has been one among the routing protocols. CTP uses the Trickle algorithm to make best use of the cost and more flexible. CTP is based on the spanning tree concept supporting many-to-one routing. The available results for CTP showed that CTP has been, so far, performing better than any other contemporary protocol during its proposal. CTP achieves mainly four goals: robustness, reliability, hardware independency, and efficiency.

The CTP makes use of Four-Bit Link Estimator to find the link to the parent node. The data is transferred to the next node based on a parameter called Expected Transmission Count (ETX). ETX measures the link quality between pair of nodes. The path for data transmission between source and destination will decided be based on the ETX value calculated for that path. Several improved versions of CTP were proposed for enhancing the performance of CTP among which R-CTP is also the one. The work presented in this paper aims to evaluate the R-CTP performance and improve the same suitably.

In this regard, this paper is structured as follows. Section 2 discusses about the related work of CTP and its variants. The Sect. 3 presents the simulation results of R-CTP and evaluates its performance, Sect. 4 propose some improvements to existing R-CTP, Sect. 5 presents simulation results for proposed improved version of R-CTP, and Sect. 6 presents the analysis of simulation results. The paper is concluded in Sect. 7 along with mentioning of directions for future research work.

2 Related Works

Routing has been an issue of paramount important, be it in a wired or wireless networks with limited or unlimited power. There had been many works in the past related to the routing in networks. Routing in WSN also has seen many such research proposals. The R-CTP is the Enhanced version of Collection Tree Protocol which is described in [1]. The CTP suffers from the poor performance and deployment issues as indicated in reports showing around 2–68% wide range of performance variations [4]. CTP Neo [4] has been proposed to with two mechanisms, one is data path validation and four-bit link estimator which is discussed is in [4]. CTP-TICN here has done some changes in link estimation calculation; it provides load balancing and it uses EETX instead of ETX which is discussed in [5]. POCTP is based on the definition of Pareto optimal route that it has been evaluated by using hierarchical Petri Net modelling technology which is presented in [6]. ICTP is based with long path good link quality and short path with weak link and it decrease the reliability in one side and avoids congestion and improves the reliability discussed in [7]. BCTP enhances CTP by enabling the nodes to balance the traffic to reduce the energy [8]. BCTP is balanced version of CTP; it avoids the traffic by enabling the network and this is discussed in the paper [8]. WSN are mainly application-specific, an attempt to propose the application-specific protocol architectures for communication in WSN had been presented in [9, 10]. Negotiation-based approaches for data aggregation are proposed in [11] whereas Directed Diffusion, a particular data aggregation method, is presented in [12]. Energy metrics based, chain-based, and tree-based data aggregation approaches are discussed in [13–15] respectively. Similar works were proposed in [16, 17]. Additionally, the performance comparison of LEACH, CTP, and ECTP using Castalia Simulator is presented in [18]. Some improvements to CTP have been proposed in [19, 20]

However, a detailed performance evaluation of R-CTP with CTP and other related variants has not been presented, so far. Additionally, as indicated in Sect. 3, there have been some disadvantages with R-CTP performance.

Thus, motivation of this paper happens to evaluate the performance of R-CTP with CTP and other related variants of CTP and improve some of the disadvantages of R-CTP using Castalia Simulator.

3 Performance Analysis of R-CTP, CTP and Other CTP Variants

The simulations of improved R-CTP, original R-CTP, CTP and other variants of the CTP are carried out using the Castalia simulator [21] for WSNs and BANs. Castalia [21] provides a generic platform "first order validation of an algorithm before moving to an implement on a specific platform" [21]. The Castalia simulator runs on the top of the OMNeT++ [21]. The proposed objectives are implemented and evaluated its performance. We have used Castalia 3.2 version for the simulation which is built upon the OMNeT++ 4.4.1 version.

The simulation has been run for 300 s for 100 static and homogenous nodes. The simulation is run for several steps to get clear results. Packet drop analyzed for last 10 packets sent to the node in the past. Remaining other factors are unaltered. These various parameter results are taken for topologies of various size ranging from 10 to 100 nodes. The performance evaluation considered following major parameters: energy Consumed in joules (J), application level latency in milliseconds (ms), number of transmitted packets, number of duplicates packets (retransmissions), number of transmitted and received beacons, Rx Breakdown, and data delivery ratio and CTP data.

The authors of this paper have implemented the R-CTP using Castalia Simulator for the purpose of evaluating its performance with CTP and other CTP variants like E-CTP and the ones proposed in [19, 20] (Fig. 1).

By collecting the data from the simulator we plotted the data in the graphs and analyzed the performance of R-CTP with the other CTP variants

A. *Consumed Energy*: In the R-CTP, when the node size is very less the energy consumed is very less. As the number of the node increases, that energy

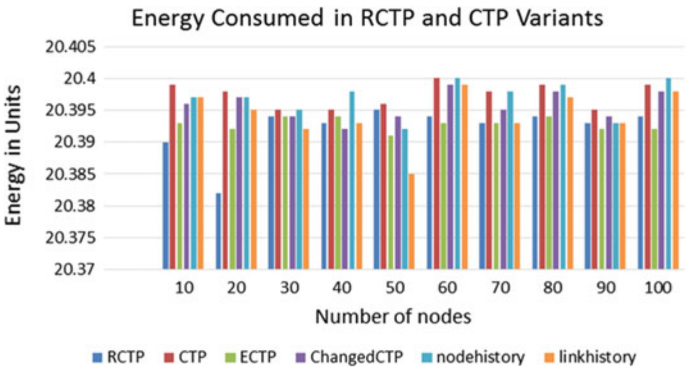


Fig. 1 Energy consumed in the R-CTP and CTP variants

consumed graph will become stable. When compared to the other variants, the R-CTP has less energy consumption, which is shown in the Fig. 2. Due to the less computation of the nodes AETX, it does not take the more energy consumption.

B. *Transmitted Packet*: The number of data and control packets by the R-CTP is comparatively very high compared to the other. The ECTP has the next highest values of transferring the packets. Because of finding the ATEX, it will transfer more packets and reduce the overhead of transferring. The transmitted packets in R-CTP and other CTP variants are shown in the Fig. 3.

C. *TX and RX Beacons*: In the TX Beacons the ECTP [2] has the more TX beacons. While compared to other variants, the ECTP will increase the TX beacons rate and R-CTP is also in the same proportional but it has very less TX beacons with respect to the ECTP. In the RX beacons, ECTP has more stable when the nodes are very less, it gradually decreases with increase in the numbers. R-CTP is initially in small numbers, and it increases with increase in the number of nodes. TX and RX beacons are shown in the Figs. 3 and 4.

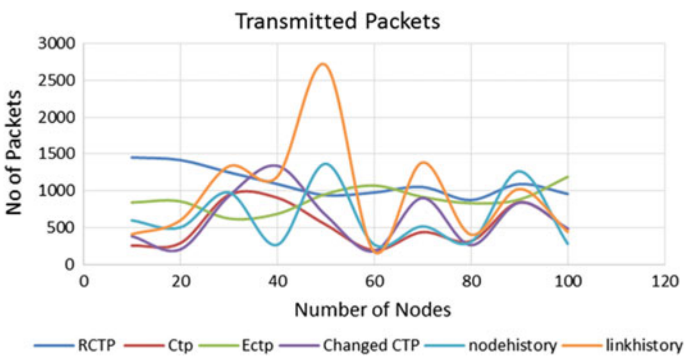


Fig. 2 Transmitted packets in the R-CTP and CTP variants

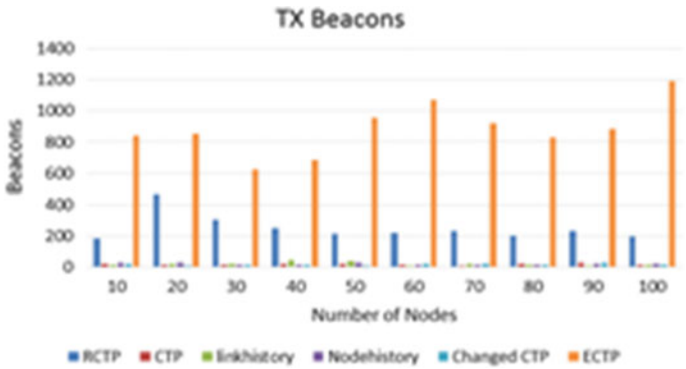


Fig. 3 TX beacons in the R-CTP and CTP variants

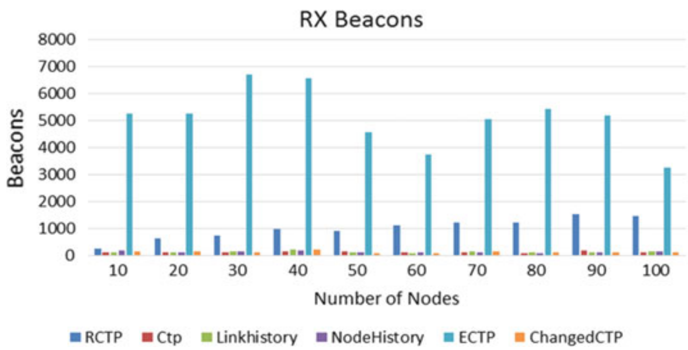


Fig. 4 RX beacons in the R-CTP and CTP variants

D. *Data Delivery Ratio*: The data delivery ratio in the R-CTP very less in the node size is very less, as the number of the nodes increases the data delivery ratio also increases. Figure 5 shows that the data delivery ratio in R-CTP and CTP variants. This is due to the new way of finding the parent selection that will perform the more data delivery ratio.
E. *Duplicates*: The duplicate packets in the R-CTP is very less. With compared to other R-CTP has the least in the observed graph. Because as the number increases the value of link metric might be same and there may be chances of the receiving multiple packets, these leads to confusion in the network. The comparison shown in Fig. 6.
F. *Latency*: From Fig. 7, the Application level latency is the less in the case of R-CTP during the number of nodes are less. The latency is high in the all the variants where the (200, inf). In that situation the ECTP as the less compared to the R-CTP. When the nodes size is below 200, R-CTP has less application level latency. The latency is less in the case of R-CTP due there is no much calculations in the finding the AETX on the R-CTP.

Fig. 5 Data delivery ration in R-CTP and CTP variants

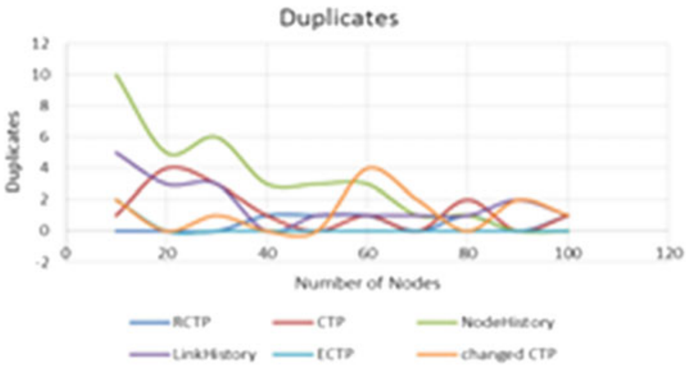


Fig. 6 Duplicate packets in the R-CTP and CTP variants

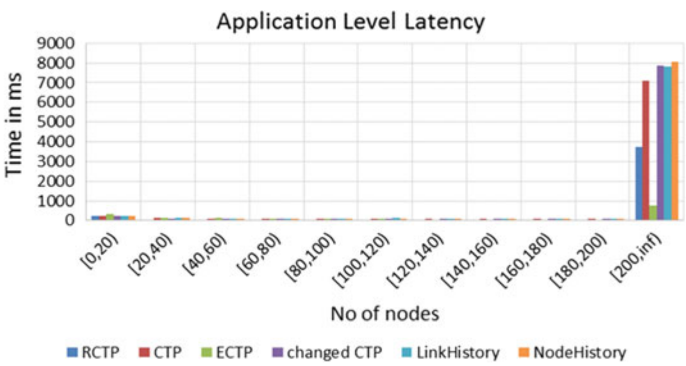


Fig. 7 Latency in R-CTP and CTP variants

4 Proposed Work

By evaluating the above graphs we came to know that it can be enhanced further by improving some of its disadvantages mentioned in Sect. 3. The observed disadvantages and proposed solutions to overcome the same have been presented as given below:

A. *Removing bad entries from the 1-hop distance*: When a node gets a routing frame, it should update the routing table for efficient transmission of packets with the latest routing address. When the values of AETX changes frequently, the R-CTP must send the broadcast frame as soon as to notify the other nodes. The parent fields acts like a substitute for the single-hop destination of a data packet: if the child's AETX is below the parent it detects automatically by parent. When this hears the child advertise an AETX below of its parent, then it should schedule routing frame for the future transmission.

In this routing table of R-CTP it stores the some address such as invalid address, current node address, address of neighbour's node which leads to the lots of confusion during the transmission of data packet to the nearest AETX node. By this problem, we are eliminating the bad entries in the routing table, i.e., unnecessary address from the table. So that we can find the address very efficiently. In this we are storing the only current address of nearest AETX of a node. By updating this we reduced the routing time of the packets to the nearest node with best AETX. By using this, we can reduce the energy consumption and the latency in the R-CTP.

B. *Improved parent selection method*: In R-CTP, the link quality metrics is calculated by AETX. Here we just changed with the parent selection in R-CTP [1]. The parent selection procedure is repeated till it finds inconsistency in the network when a node receives a beacon to reconsider the topology. It may be neighbour comes out from congestion mode or parent may be unreachable.

Here we are improving the parent selection of existing method. In this, we are interested to find the less values of the parent. Here we are calculating the value of the actual link parent and link of the parent that is AETX. After this we are comparing both values which the parent value will assign. As the link values are smaller, the distance between sender and parent is smaller. Hence it is get the nearest node for forwarding the data packet. Due to modification of the parent selection, we are getting nearer parent and increasing in the number of packets.

C. *Improved method for loop avoiding in the Link Estimator*: The accurate link quality estimation is very important task in the routing. The poor link estimate may cause a 200 or more percentage of failure in the communication or slowdown the network. The link estimation is more of a headache in the development to find the exact link that can communicate efficiently. The four-bit link estimator provides the four bit of information: 1 bit from the physical layer, 1 bit from link layer and 2 bits from the network layer.

In the physical layer we can find the channel quality during a packet. The link layer, we can measures whether the packet is delivered or acknowledged. From the network layer we can check the links are more important for the higher layer performance.

In the link estimator, we are removing the loop that cause for the increasing in the latency. By removing the unnecessary loops which to find the link estimation of next nodes. By optimizing the links of the link estimator we can reduce the energy consumption in the system that can take long time and power for the finding the link. Thus we can find the link for the next nearest node; because of this we are good to reduce the latency and energy factor also.

D. *Improved EETX for the new Beacon estimation*: A collection tree protocol builds and then maintains minimum cost trees to nodes which make them as tree roots. This protocol is address-free: if there are multiple base stations, it sends one with minimum cost without knowing that it is the address. This protocol basically broadcast the control beacons at a regular fashion, i.e., at fixed interval. When the cost drops the beacon interval is reset by the CTP. This is not mandatory for the correctness.

The improvement of EETX, here we find the nearest node values of EETX. The EETX is expected number of transmissions. Here we are use random numbers. The random number is given to the new beacon estimation for the new node which is to interfere. From this, the fast response from the beacon can assign the new value of EETX. On finding the new value of EETX for the new beacon, we can get more number of TX and RX beacons.

5 Simulation Results of Proposed Improvements

The simulation is carried out using the Castalia Simulator, the parameter are kept same as the existing simulation as described in the Sect. 3. Here the authors are defined the some changes to protocol that are described in the Sect. 4 and executed the simulation.

The simulation has been run for 300 s for 100 static and homogenous nodes. The simulation is run for several steps to get clear results. Packet drop analyzed for last 10 packets sent to the node in the past. Remaining other factors are unaltered. These various parameter results are taken for topologies of various sizes ranging from 10 to 100 nodes. The performance evaluation considered following major parameters: energy Consumed in joules (J), application level latency in milliseconds (ms), number of transmitted packets, number of duplicates packets (retransmissions), number of transmitted and received beacons, Rx Breakdown, and data delivery ratio, and CTP data.

6 Evaluation of the Improved Results

A. *Energy Consumed in Improved R-CTP and other CTP Variants*: The total network life time can be calculated till the network is fully functional, i.e., total time till which all the nodes in the network works or until the first node runs out of power. Estimating the energy consumption is a vital key in the network; so it is essential in each node. Figure 8 shows the energy consumed by the improved version of R-CTP and R-CTP and other CTP variants. The energy consumed by each node in the system is calculated in the variant of improved R-CTP is ranging from 20.393–20.388. This shows the gradual decrease in the energy content decrease in the nodes as the number of nodes increases gradually. This is due to the removing of the bad entries in the 1 hop distance and loop avoidance in the link estimator, so there is no more need of extra calculation in finding the address of the nearest nodes.

B. *Transmitted Packets*: The transmitted packets refers to the both data packets and the control packets. The packets refer here the amount of traffic on each node in the network it relays. Figure 9 shows the comparison improved R-CTP and other CTP variants. The comparison shows the large number of transmission of packets, which is about 12% more transmits the packets compare to the original R-CTP. The increasing in the number of transmitting packets is due to the improved parent selection method. Here we described that it finds the good neighbour with least cost of parent. As the number of link metric is less the parent is nearer to the sender.

C. *Improved R-CTP RX & TX Beacons*: The transmission and receiver of the beacons are used to hear the in the application layer to identify the localization. On improving the R-CTP by above factors, we got the massive increase in the RX beacons. Figure 10 shows the TX beacons initially start with low value to the less number of nodes. As the number of nodes increasing gradually, the improved version of R-CTP increase the number of TX beacons heavily. In the

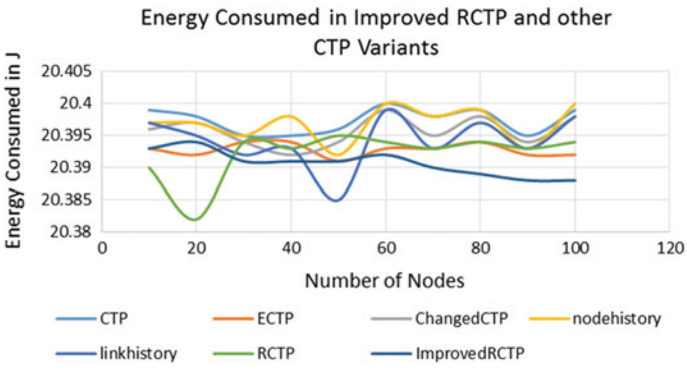


Fig. 8 Energy consumed in improved R-CTP and R-CTP and other variants

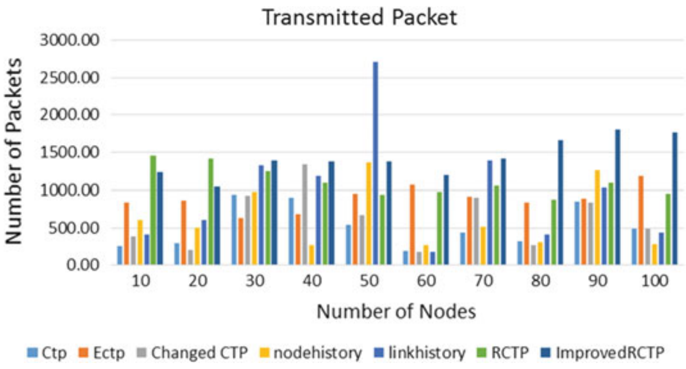


Fig. 9 Transmitted packets in the improved and other CTP variants

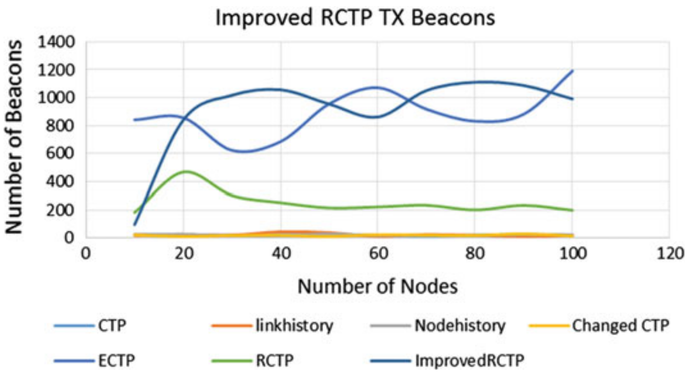


Fig. 10 Improved RTCP TX beacons

other case Fig. 11. RX beacons it also increase with straight line with the number of nodes increase frequently. This is due to the new method for the new beacon estimation; here we are calculating the new EETX value. These values will find the new beacons that will find when the new node interface.

D. *Data Delivery Ratio*: The data delivery ratio (DDR) parameter reports the number of packets transfer of the packets effectively in the network. DDR is a ratio of successful data packets received to the attempted packets transmitted. Figure 12 shows the DDR with different protocols. The improved R-CTP gradually increases in the packets as the number of nodes increases. In the case of R-CTP the number of DDR is more compared to the improved version. The improved is not as bad as original but it is the equal proportional in the DDR content.

E. *Duplicates (Retransmission)*: The duplicate packet means the twice or more than a same packet received at the receiving host. It is not good that it gets more duplicate packets. As it gets more duplicate, the application might think it as fresh packet and which may feel in the confusion. Figure 13 shows the

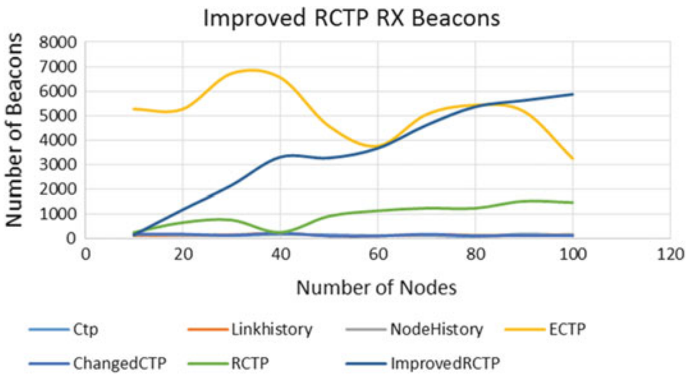


Fig. 11 Improved R-CTP RX beacons

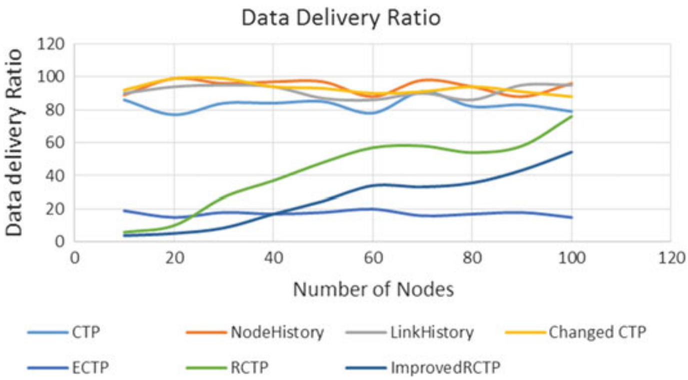


Fig. 12 Data delivery ratio

duplicates that have little more duplicate packets on improved version compared to the R-CTP. When we optimizing the parent selection it may be confusing due to the selection of parents when the value of the link metric is close and the number of nodes are more. This might may leads to increase in the duplicates.

F. *Application Level Latency*: The latency of the application level refers to the amount of time taken to the packet to reach from the source to the destination. It is very critical impact that when a node is travelling from one node to another in the alarm-based application. In many cases, we trade off in minimizing the source sink delay and energy simultaneously. Figure 14 shows the application level latency of the different protocols, where in the improved version of R-CTP is 2% less of R-CTP. Due to the loop avoiding in the link estimator, it will not find the more time and less computation, so the latency will be reduced. The protocol is very efficient when compared to the other variants CTP.

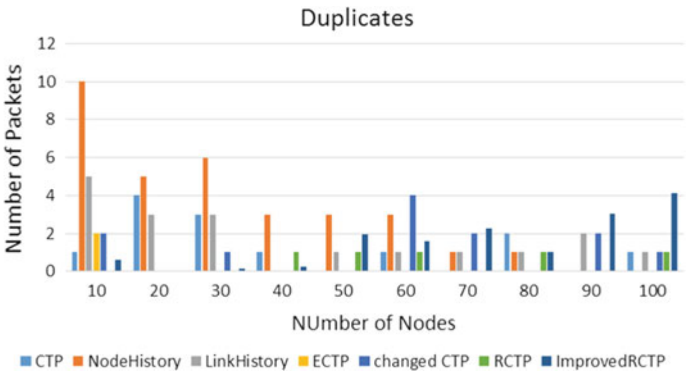


Fig. 13 Duplicates

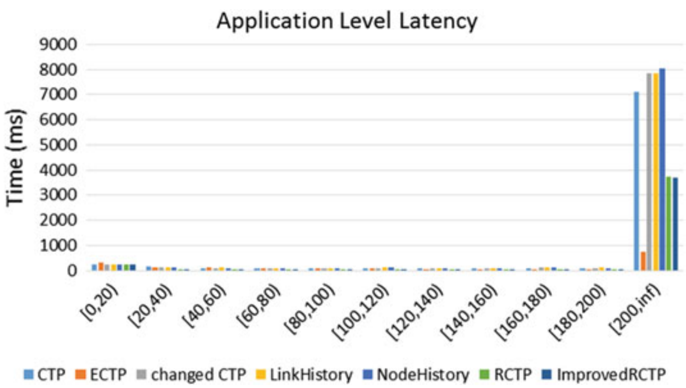


Fig. 14 Application level latency

G. *Packet Breakdown*: The received packets breakdown can be influenced by variety of factors that is interference of other nodes, sensitivity of the packet, and environment factors. This metric refers to the number of packets that are failed or received by each node respectively with and without interference, failed below sensitivity. Figure 15 shows the RX packet breakdown. The breakdown is shows in the improved version of R-CTP.

To summarize this, the proposed improved variants of R-CTP to improve the number of transmitted packets by 12% and the latency is reduced by the 2%. The energy consumption factor is reduced by some factors is shown in the graph. And additionally we proposed to improve the other factors such as energy consumption, TX and RX factors, breakdown situations and number of received packets.

And we respect the overall performance which here we observed of the different parameters of R-CTP with the CTP and original R-CTP. And here our main goal is to improve the data aggregation as well as the routing fields, which can be achieved effectively and efficiently.

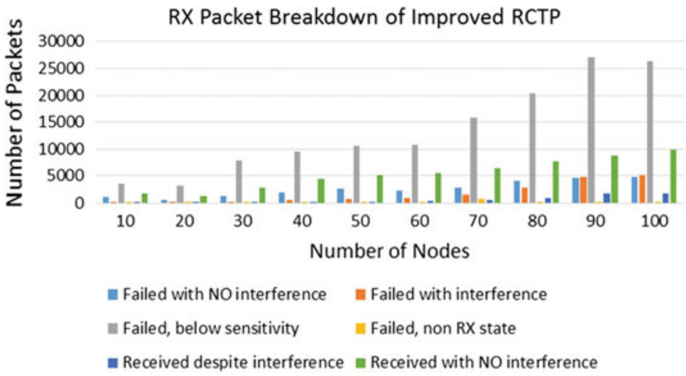


Fig. 15 RX packet breakdown of improved R-CTP

7 Conclusion and Future Work

The data aggregation here assumes with great importance to WSNs. There have been solutions has provided in the past and he latest one has the efficient. The R-CTP is mainly focused on the routing layer of the CTP which provides effectively on the routing to achieve best performance. Now the R-CTP has a very good protocol which we are compared here with the CTP and its different variants. We further identified the difficulties in the R-CTP and provides the best solutions for the efficient transfer of the data. While comparing the other protocols we found the improved version of R-CTP is improved 12% more transfer of the packets, reduce in the latency by 2% and huge increase in the TX and RX packets and rest also assured.

Further future research work can be enhanced by the various factors. This mainly can be dealt with dynamic load balance for the link quality, so that it may reduce the link quality headaches; and in the dynamic adaptively problem solving for trickle algorithm and adaptive beaconing.

References

1. Fariborz Entezami, Martin Tunicliffe, and Christos Politis, "*RCTP: An Enhanced Routing Protocol Based on Collection Tree Protocol*," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 363107, 7 pages, 2015. doi:10.1155/2015/363107
2. Jingjing Zhang, Zhenqi Yang, Baowei Wang, Huiyu Sun and Xingming Sun, *E-CTP: An Energy-balanced Collection Tree Protocol for Power Constrained Wireless Sensor Networks*, International Journal of Grid and Distributed Computing. Vol.7, No.2 (2014), pp. 115–126
3. Ugo Colesanti and Silvia Santini, *The Collection Tree Protocol for the Castalia Wireless Sensor Networks Simulator,* Technical Report 729, Department of Computer Science, ETH Zurich, Zurich, Switzerland, June 2011

4. O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis, "Collection tree protocol," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys'09)*, pp. 1–14, ACM, New York, NY, USA, November 2009.
5. Y. Song, Y. Chai, F. Ye, and W. Xu, "A novel TinyOS 2.x routing protocol with load balance named CTP-TICN," in *Knowledge Engineering and Management*, vol. 123 of *Advances in Intelligent and Soft Computing*, pp. 3–9, Springer, Berlin, Germany, 2011.
6. Y.-Z. Wu, D.-P. Quan, and H.-G. Han, "Pareto optimal collection tree protocol for industrial monitoring wsns," in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps'11)*, pp. 508–512, IEEE, Houston, Tex, USA, December 2011.
7. Y. Li, H. Chen, R. He, R. Xie, and S. Zou, "ICTP: an improved data collection protocol based On CTP," in *Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP'10)*, pp. 1–5, October 2010.
8. J. Zhao, L. Wang, W. Yue, Z. Qin, and M. Zhu, "Load migrating for the hot spots in wireless sensor networks using CTP," in *Proceedings of the 7th International Conference on Mobile Adhoc and Sensor Networks (MSN'11)*, pp. 167–173, IEEE, Beijing, China, December 2011.
9. W. R. Hein Zelman, "Application-specific protocol architectures for wireless networks", PhD Thesis, Massachusetts Institute of Technology, June 2000.
10. W.R. Hein Zelman, A.P. Chandrakasan, H. Balakrishnan, "An application-specific protocol architecture for wireless micro sensor networks, IEEE Trans Wireless Communications", October 2002, pp. 660–670.
11. J. Kulik, W.R. Heinzelman, H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks, Wireless Networks", vol. 8, March 2002, pp. 169–185.
12. C. Intanagonwiwat, R. Govindan, D. Estrin, "Directed Diffusion: A Scalable and robust communication paradigm for sensor networks", Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCOM '00), August 2000.
13. S. Lindsey, C. Raghavendra, K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics", IEEE Trans. Parallel and Distributed Systems, vol. 13, no. 9, September 2002, pp. 924–935.
14. K. Du, J. Wu, D. Zhou, "Chain-based protocols for data broadcasting and gathering in sensor networks," International Parallel and Distributed Processing Symposium, April 2003.
15. M. Ding, X. Cheng, G. Xue, "Aggregation tree construction in sensor networks," 2003 IEEE 58th Vehicular Technology Conference, vol.4, no.4, October 2003, pp. 2168–2172.
16. H. O. Tan, I. Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," SIGMOD Record, vol. 32, no. 4, December 2003, pp. 66–71.
17. K. Vaidhyanathan, S. Sur, S. Narravula, P. Sinha, "Data aggregation techniques sensor networks," Technical Report, OSU-CISRC-11/04- TR60, Ohio State University, 2004.
18. Pranesh V Kallapur, Vibha Prabhu, "Performance Analysis of LEACH, CTP, ECTP Using Castalia
19. Pranesh V Kallapur, Vibha Prabhu, "Improved Variant of CTP for Enhanced Data Delivery Ratio and Reduced Computation Overhead", IEEE International Conference On Applied and Theoritical Computing & Communication Technology(ICATCCT-2015), 29–31 October, 2015, India
20. Pranesh V Kallapur, Vibha Prabhu, "Enhanced Variants of CTP for Improved Data Delivery Ratio and Energy Consumption", IEEE International Conference On Applied and Theoretical Computing & Communication Technology (ICATCCT-2015), 29–31 October, 2015, India
21. Castalia User's Manual http://castalia.npc.nicta.com.au/documentation.php
22. OMNeT ++ User Manual (Version 4.4.1). www.omnetpp.org/doc/omnetpp33/manual/usman.html.
23. Ugo Colesanti and Silvia Santini, *A performance evaluation of the collection tree protocol based on its implementation for the Castalia wireless sensor networks simulator.* Technical Report 681, Department of Computer Science, ETH Zurich, Zurich, Switzerland, August 2010.

24. B. Krishnamachari, D. Estrin, S. Wicker, "The impact of data aggregation in wireless sensor networks," Proc. 22nd International Conference on Distributed Computing Systems Workshops, July 2002, pp. 575–78.
25. TinyOS Wiki, "Tep119-collection," http://www.tinyos.net/tinyos-2.x/doc/txt/tep119.txt.
26. T. A. Ramrekha and C. Politis, "A hybrid adaptive routing protocol for extreme emergency ad hoc communication," in *Proceedings of the 19th International Conference on Computer Communications and Networks (ICCCN)*

Signal Processing of Motor Imagery EEG Waves Using Empirical Mode Decomposition

Ajithkumar Sreekumar, M. Uttara Kumari, Krishna Chaithanya Vastare, Suraj Madenur Sreenivasa and N. Apoorva

Abstract Electroencephalogram (EEG) is the most convenient method for recording the electrical activities of the brain, for Brain Computer Interface (BCI) applications. This EEG data is notoriously noisy. A variety of frequency estimation techniques are used in feature extraction. This is possible due to the presence of information of interest in frequency bands which are well defined. The application of EMD (Empirical Mode Decomposition) on the recorded EEG waves of subjects', renders time-frequency data depicting instantaneous frequencies. EMD is chosen to obtain Hilbert–Huang Transform (HHT) of the data which is chosen over Fourier Transform (FT) owing to the nonstationarity, closely spaced frequency bands of interest and low SNR of the recorded data. HHT of the data can be used to obtain a feature or signature, which can be used as a command signal for various BCI applications.

Keywords EEG waves · Empirical mode decomposition · Fourier analysis Hilbert huang transform

A. Sreekumar (✉) · M. Uttara Kumari · K.C. Vastare · S.M. Sreenivasa
Department of Electronics and Communication Engineering,
R.V. College of Engineering, Bengaluru 560059, India
e-mail: akajithkumar351@gmail.com

M. Uttara Kumari
e-mail: uttarakumari@rvce.edu.in

K.C. Vastare
e-mail: krishnachaithanya.v@gmail.com

S.M. Sreenivasa
e-mail: surajathreya@gmail.com

N. Apoorva
Department of Information Science and Engineering,
R.V. College of Engineering, Bengaluru 560059, India
e-mail: apoorvanagaraj95@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_18

1 Introduction

The cornerstone of BCI is the exclusive use of brain activity in computer-aided control. Neuroprosthetics and bioengineering are the major fields of BCI applications. EEG for brain activity recording is used widely due to its noninvasive nature, affordability and operation in real-time [1]. Motor imagery BCI being the imagination of motor action without actual physical movement, has clear practical significance [2]. This requires extended training periods, is challenging to analyse and has limited BCI channel capacity.

EEG research carried over more than seven decades has resulted in the introduction of an abundant class of quantitative feature extraction from EEG signals. Similar to any other signal, an elaborate mathematical model of the EEG signal is very promising [3]. Physiological findings and mathematical models correlating the EEG to electrical activities of a single nerve cell remain challenging and no mathematical model of EEG has yet achieved the aforementioned goal of modelling the wide varieties and dynamics of EEG. Autoregressive modelling of tiny EEG segments is successful to a certain extent [4]. The nonstationarity of EEG waves leads to the fact that mathematical models using static stationary equations are not suitable. Hence dynamic mathematical models have to be developed leading to increased complexity in the models.

1.1 Nonstationarity in EEG Waves

Transient events manifest themselves as the nonstationarity phenomenon of the EEG waves, such as alteration of homogeneous segments with dissimilar statistical features, spikes, sharp waves and spike-wave discharges. Visual inspection is sufficient for the identification of the transient behaviour owing to its specific patterns, but the identification of relatively homogeneous intervals requires theoretical basis [5].

The EEG data is converted to digital form to carry out the computer-aided analysis. The most of the prominent EEG components are present in the frequency range of 1 Hz to 30 Hz, hence the digitising rate is the range of 60 Hz–150 Hz [5]. Intervals which are less than 0.5–1 s need not be checked for stationarity if about, 50–100 samples are required for statistical characterization.

1.2 EEG Waves Features

The high frequency waves in the EEG data have been digitally removed by passing them through a Backman filter, low-pass filter with −67 dB at a 32 Hz cut-off. Later this filtered data is sampled at the rate of 64 Hz for feature extraction.

Table 1 Features of EEG waves

Parameter for extraction	Features	Symbols/Band
Time domain extraction	Statistical mean	X
	Sum of absolute amplitudes	$\lvert X\rvert$
	Sum of squared amplitudes	X^2
	Standard deviation	X_{cr}
	3rd moment of mean amplitude	m^3
	4th moment about the mean amplitude	m^4
	Skewness coefficient	g_1
	Kurtosis coefficient	g_2
Frequency domain spectral amplitude	$\lvert A\rvert$	0–32 Hz
	$\lvert\beta\rvert$	13–32 Hz
	$\lvert\alpha\rvert$	8–13 Hz
	$\lvert\theta\rvert$	4–8 Hz
	$\lvert\delta\rvert$	0–4 Hz
Frequency domain relative spectral amplitude	Alpha band	$\alpha\%$
	Alpha band	$\alpha\%$
	Beta band	$\beta\ \%$
	Theta band	$\theta\ \%$
	Delta band	$\delta\%$
Mean of frequency bands	Entire EEG spectrum	F_M
	Alpha band	F_α
	Beta band	F_β
	Theta band	F_θ
	Delta band	F_δ
Statistical analysis of frequency spectrum	Squared amplitudes	A^2
	Standard deviation	F_{SD}

Time, frequency and statistical tools are the approaches considered for extracting EEG waves features. From the sampled data 24 EEG signals are extracted, each with a window of 4 s. With this data the features listed in Table 1, can be extracted.

Fourier transform has been applied with a Hanning window on the EEG data to calculate the rest of the features. The following features comprising of the EEG waves spectral amplitudes are obtained. Using the same Fourier analysis, the relative spectral amplitudes of EEG, the mean frequencies, sum of squared amplitudes are thus obtained.

1.3 EEG Waves Feature Selection

Three standards are used to resolve the EEG features corresponding to behavioural alertness level. First, the EEG features having higher average rank across eight procedures of data processing are mainly selected. The level of alertness is primarily manifested as the features with higher average rank numbers. Second, the EEG features with higher individual inconsistency are not selected as they have constrained use to the estimation of the level of alertness among subjects. The individual inconsistency is estimated using a measure called as Reliability Index (RI), which is given in Eq. 1.

$$\mathrm{RI} = |\mathrm{N} - \mathrm{P}|/(\mathrm{N} + \mathrm{P}) \quad (1)$$

where N/P is the corresponding addition of the ranks of negative/positive correlation coefficients across eight processing procedures or across different subjects. RI has a value in the range of 0–1. Features having higher RI values have lesser individual inconsistency with respect to the level of alertness. In the last standard, the interrelationships of EEG features are analysed to determine the redundancy of the EEG features in estimation of alertness. The addition of the ranks of negative/positive calculated in the previous standard is used here. Larger inter-correlation is obtained for higher values of overall sum of ranks. Finally, the EEG features having low individual inconsistency, lesser inter-correlation, larger correlation with alertness level are selected as consistent and effective factors for estimation of alertness level [5].

2 Experimentation

The proposed work is to find out the differences between the brain signals of different individuals, and to understand what makes them unique. The experimental setup for recording of EEG waves is as shown in Fig. 1. This uniqueness can be used as a password to a website/security system making it more robust [6].

So far, through the literature survey, the various methods currently existing for sensing EEG signals have been understood. In particular, techniques to record signals which are produced for a particular action. A protocol to capture brain signals has been developed, which:

Fig. 1 Recording EEG waves of the subjects using enobio kit

Table 2 Specifications of the enobio device

Parameter	Value
Total number of channels	8
Number of EEG records	18277
Number of packets lost	0
EEG sampling rate	500 samples/s
EEG units	nV

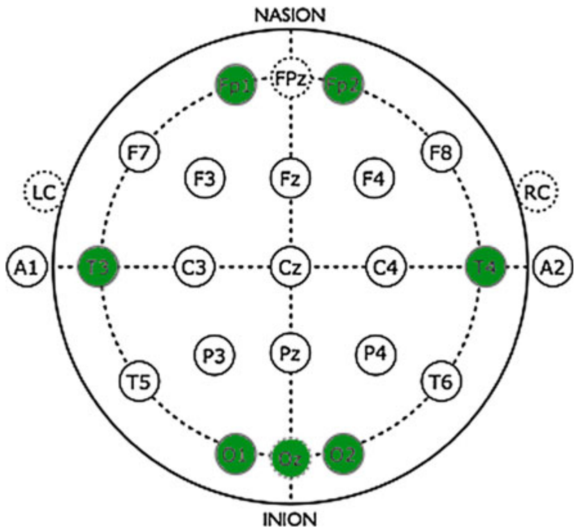


Fig. 2 Placement of the electrodes

1. Motor Imagery: The thought for the muscular movement is aided by a visual stimulus.
2. SSVEP: A protocol which uses images, which are flashed at the rate of 1 Hz, and provides visual stimulus to the subject.

The signals were captured using a device called Enobio. Enobio is a wireless technology which comes in three versions, 8, 16 and 32 channel electrode systems, used for collecting EEG signals. For the data collection, eightchannel electrode system was used. The complete device specification of the device is depicted in Table 2. The placement of electrodes on the subject is as shown in Fig. 2. In the Fig. 2, Green Dots indicate the placed EEG electrodes. Then, these signals were processed to retrieve information and points of interest. After filtering the signals, features must be extracted from them, and finally classified.

The placement of the electrodes mainly depends on the type of signal being analysed. The occipital lobe is responsible for the decoding of visual stimuli. Steady State Visual Evoked Potential (SSVEP) are usually recorded from occipital scalp for this reason. The channels which were used to collect VEP are Fp1, Fp2, O1, O2, Oz, T3 and T4 taking P3 as reference electrode. For Motor Imagery protocol, T3 and T4 are replaced by C3 and C4. The experimentation details are as shown in Table 3.

Table 3 Experimentation details

Parameter	Value
Number of subjects	8
Age	20–30
Number of trails on each subject	50
Duration of the protocol	30 s

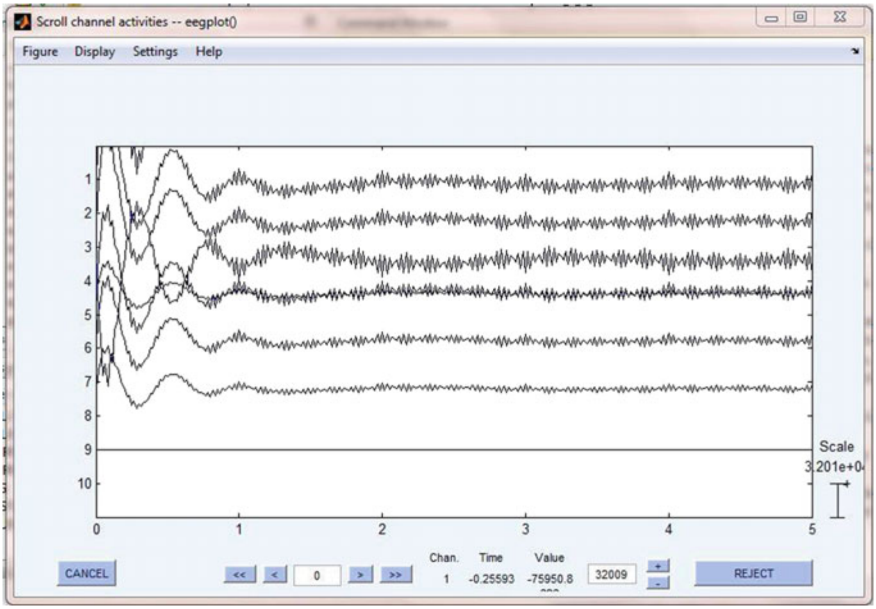


Fig. 3 Time domain representation of the EEG waves of a subject

2.1 Time Domain EEG Waves of the Subjects

The EEG waves are recorded in time domain as shown in Fig. 3. The time domain EEG waves are stored in easy files. These files are processed using EEGLAB toolbox in MATLAB [7].

2.2 Power Spectral Density

The time domain signals are first filtered to remove various noises such as power line 50 Hz noise. Later EEGLAB provides tools to eliminate the mechanical disturbances from the eye movement and the blink rate. This processed time domain signal is then converted to its PSD format as depicted in Fig. 4.

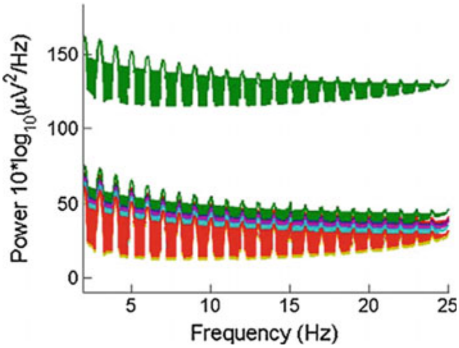


Fig. 4 PSD of the EEG waves of a subject

3 EEG Signal Processing

EEG Signal processing techniques comprise of time domain analysis, for digital filtering and independent component analysis [8], frequency domain analysis, for analyzing the spectrum, and time-frequency domain analysis.

3.1 *Fourier Analysis for EEG Waves*

EEG data is nonstationary, nonlinear and aperiodic by nature and Fourier Analysis (FA) works best with stationary, linear and periodic signals. The local nonlinearity of the data results in considerable spreading with the application of Fourier analysis, as the basis functions of Fourier analysis are global. The dispersion is even higher for data which significantly digress from the sinusoidal form. Thus the use of Fourier analysis is not preferable for EEG.

3.2 *Time-Frequency Domain Analysis for EEG Signals*

The progression of a signal in frequency as well as time domain is described by the time–frequency domain analysis. There are linear and nonlinear methods to carry out this analysis. The linear methods being Short-time Fourier transform (STFT) and Wavelet transform. The lone nonlinear method available is the Hilbert–Huang Transform (HHT).

The nonlinear nature of EEG waves makes HHT, the logical choice.

3.3 Hilbert Huang Transform

The instantaneous frequency data is obtained by the decomposition of the signals into Intrinsic Mode Functions (IMF) along with a general tendency of the signal and the HHT is way to do this. HHT is very suitable for data which is nonlinear and non-stationary [9]. Unlike Fourier Transform (FT), HHT is not a theoretical tool, but it is like an empirical or algorithmic approach that is very data-driven. Hilbert spectral analysis after the EMD (Empirical Mode Decomposition) are the two parts of HHT.

EMD is very efficient due to its adaptive nature. EMD is applicable to nonlinear and nonstationary processes as the method is based on local behaviour of data in the form of time series. Intrinsic Mode Functions (IMF) are the result of EMD which help in representing the data in time–frequency domain.

3.4 Empirical Mode Decomposition

The Hilbert spectral analysis is applied to IMF which is obtained by applying EMD to the data set.

IMF is more general as it not a simple harmonic function as in FT, but is a simple oscillatory mode. In FT the simple harmonic components have fixed frequency and amplitude, whereas in HHT the IMF has adaptive amplitude frequency which is dependent on local characteristics of the data [10].

Sifting is the process used to obtain the IMF. The process is carried as follows:

1. The cubic spline line connecting all the local maxima forms the upper envelope [7].
2. Lower envelope is formed by applying step 1 to the local minima [7].

All the data is sandwiched between the upper and lower envelopes. Their average is a1. The first component c1 is obtained by subtracting a1 from the data:

Ideally, c1 must obey the constraints imposed on IMF. This is taken care by the method used to construct c1 in the preceding step making it symmetric about 0 with maxima > 0 and minima < 0. After the initial round of sifting, a trough may become a local minimum.

The extrema obtained in the new rounds help in revealing the correct modes, which vanished in the prior rounds of examination. In the next rounds of sifting the component c1 is treated as data [10].

$$X(t) - a_1 = c_1 \tag{2}$$

With n times of recursive sifting, c_1 converts to an IMF, that is

$$c_1 - a_{11} = c_{11} \tag{3}$$

Table 4 Comparison between Fourier transform and Hilbert–Huang transform

Transform	FT	HHT using EMD
Operation	Convolution: uncertainty and global	Differentiation: certainty and local
Representation	Energy vs frequency	Energy vs time and frequency
Linearity condition for the data	Linear	Nonlinear
Stationarity condition for the data	Applicable for stationary data	Applicable for even non stationary data
Extraction of features	Not possible	Possible
Core concept	Derived from theory	Developed as an empirical method

Then, the component c_{1n} is chosen as the 1st IMF component of the data set:

$$c_{1(n-1)} - a_{1n} = c_{1n} \tag{4}$$

$$c_1 = c_{1n} \tag{5}$$

The above algorithm results in the decomposed EEG data providing the instantaneous frequency of the subject. This information can be utilised for various applications. Table 4 shows the comparison between Fast Fourier Transform (FFT) and Hilbert–Huang Transform (HHT). The results section deals with the consequences of these differences between the methods, on the suitability of the method to be chosen for data obtained from a particular protocol.

4 Results

The simulation results in this section are obtained using EEGLAB plugin [6] for MATLAB v15a. This section will summarise the results obtained for the two protocols, SSVEP and Motor Imagery. In the SSVEP protocol the images are flashed at a regular rate of 1 Hz. The FFT of a percentage of the data collected with SSVEP protocol for a particular subject is shown in Fig. 5. The two parts of Fig. 5 are the Event-Related Spectral Decomposition (ESRP) in dB and Inter-Trial Coherence (ITC) phase in rad. The Event-Related Potential (ERP) is also plotted. The events here refer to the flashing of images on the screen. The plots in Fig. 5 reveals that FFT is suitable for analysing the data obtained through SSVEP protocol as a regular pattern is observed in the FFT as well as the ERP's.

Figure 6 shows the FFT of the data obtained through motor imagery protocol. There is no regular pattern in ESRP, ITC phase or ERP's. This is because the data obtained through motor imagery is aperiodic data. Therefore, HHT is suitable. This is justified in Figs. 7 and 8. The Channel 12 corresponds to C3 and channel 13 corresponds to C4 in 10-20 International electrode placement system. These are the

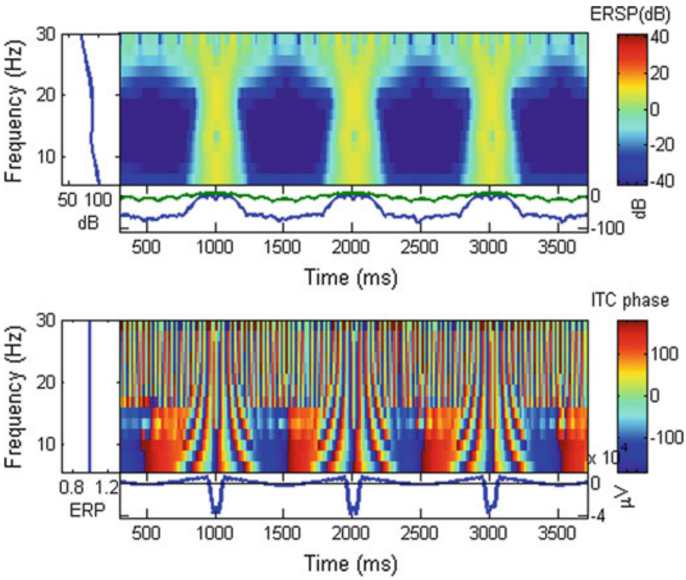


Fig. 5 FFT of steady state visual evoked potential data

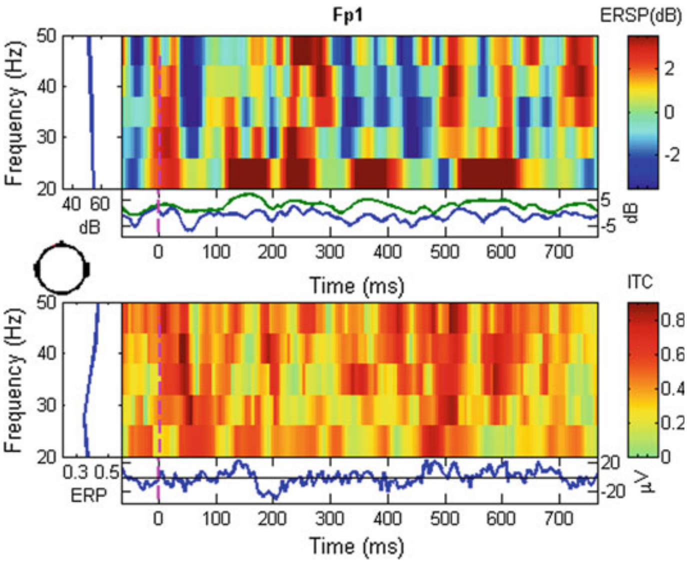


Fig. 6 FFT of motor imagery EEG waves data

two channels in central lobe where the activity increases during motor imagery. Figures 7 and 8 correspond to the HHT of the data collected using motor imagery protocol. The motor imagery action performed in this case is the subject thinking

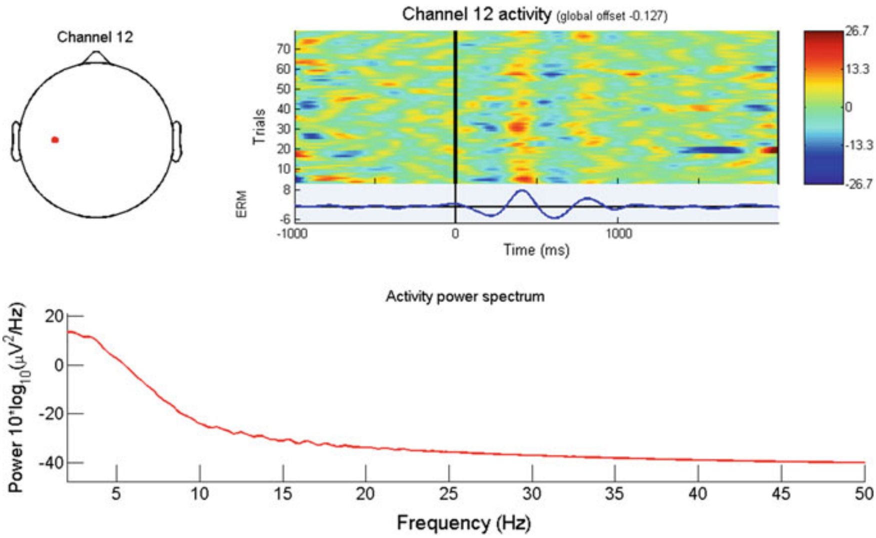


Fig. 7 Event related synchronisation (ERS) of C3

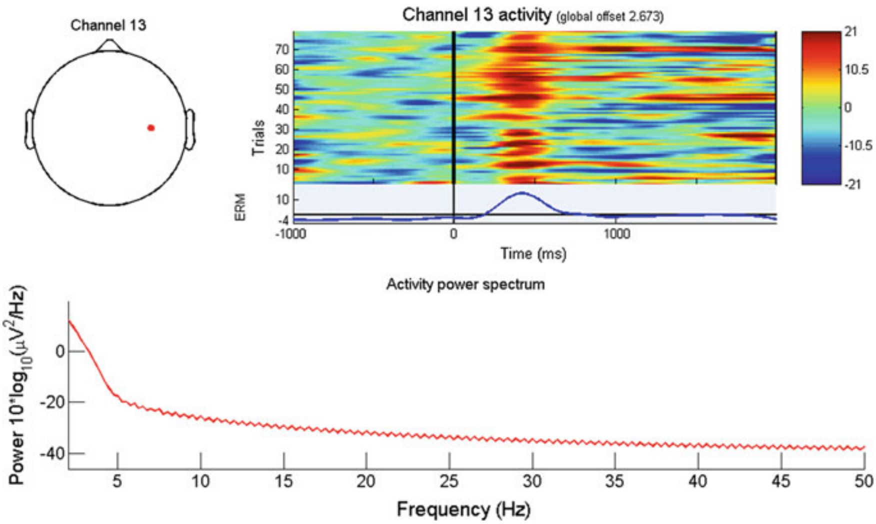


Fig. 8 Event related desynchronisation of C4

about moving right hand. The activity spectrum clearly shows the Event Related Synchronisation (ERS) of C3 in Fig. 7 and Event-Related Desynchronisation (ERD) of C4.

5 Conclusion

Development of any BCI application requires the features of EEG waves of the subject. The various EEG features have been enlisted in the proposed research work with their significance. The EEG data for the subjects have been extracted from the proposed protocol and the Enobio electrodes have been used to record the EEG waves of the subject. Later it has been proved that Fourier analysis is not entirely sufficient for the EEG waves feature extraction as the data is non-stationary, non-linear, and not periodic. Hilbert–Huang Transform is applied to the data to obtain the time–frequency joint and to extract the features. From the results it is evident that, for protocols such as motor imagery, HHT is the best suited protocol for feature extraction owing to its application on aperiodic data. FFT serves to give satisfactory results when applied to protocols like SSVEP owing to the periodic data generated by these protocols.

HHT of the data is obtained using Empirical Mode Decomposition algorithm which is a pragmatic approach to analyse non-stationary data sets, such as EEG waves data. EMD uses a heuristic approach for signal decomposition, which reduces the signal for a given locality with no restriction on conditions such as, sparseness, independence or orthogonality.

6 Future Work

The algorithm and the protocol used have been experimented with an 8-channel EEG electrode system to analyse motor imagery EEG waves, as compared to the 59 channel EEG electrode system used in [1]. Thus the proposed system is more economical and effective. The EEG waves corpus developed can be utilised for various embedded system applications such as, for human identification/security [8], controlling robots, etc. Motor imagery EEG waves processing can also be used to develop prosthetic arms/legs for physically challenged people. With the increase in number of EEG electrodes, the entire brain of a patient can be mapped, which would result in faster and efficient diagnosis.

Specifically, by improving the protocol and the correct placement of electrodes, diseases such as Alzheimer's disease, Dementia, Parkinson Diseases and many more can be diagnosed at their early stages. Also, diseases such as anaemia can also be detected using EEG waves by correlating it with the ECG waves. Since Brain waves control all the organs and systems of the human body, comprehending them can result in understanding the human body itself. Therefore, with a better understanding of the human body, the lives of the people can be enriched.

References

1. C. Park, D. Looney, N. Rehman, A. Ahrabian and D. P. Mandic, "Classification of motor imagery BCI using multivariate empirical mode decomposition", IEEETrans. Neural Syst. Rehabil. Eng., vol. 21, no. 1, pp. 10–22, 2013.
2. S. Cososchi, R. Strungaru, A. Ungureanu and M. Ungureanu, "EEG feature extraction for motor imagery", Proc. of IEEE Engineering in Medicine and Biology Society, pp. 1142–1146, 2006.
3. S. H. Hsu, T. R. Mullen, T. P. Jung, G. Cauwenberghs, "Real-Time Adaptive EEG Source Separation Using Online Recursive Independent Component Analysis", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol 24, Issue 3, 2016.
4. Bradley J. Edelman, Bryan Baxter, and Bin He, "EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks", IEEE Transactions On Biomedical Engineering, Vol. 63, No. 1, January 2016.
5. E. Maiorana, D. La Rocca, P. Campisi, "On the Permanence of EEG Signals for Biometric Recognition", IEEE Transactions on Information Forensics and Security, Vol 11, Issue 1, 2016.
6. A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis", Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, Mar. 2004.
7. Yuan Zou, V. Nathan, R. Jafari, "Automatic Identification of Artifact-Related Independent Components for Artifact Removal in EEG Recordings", IEEE Journal of Biomedical and Health Informatics, vol 20, 2016.
8. Pritham Gajakumar Shah, Krishna Chaithanya Vastare, Ajithkumar Srikumar, Suraj Mademur Sreenivasa, Adarsh Puvvadi Ram Mohan Kumar, Karthik Rajashekhar Kodada, "Development of a novel EEG wave controlled security system", 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), vol 3, pp. 116–120, 2015.
9. Su Yang, Farzin Deravi, "Novel HHT-Based Features for Biometric Identification Using EEG signals", 2014 22nd International Conference on Pattern Recognition, pp. 1922– 1927, IEEE, 2014.
10. M. Lambert, A. Engroff, M. Dyer, B. Byer, and E. M. Decomposition, "Hilbert–Huang transform", in Wikipedia, Wikimedia Foundation, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Hilbert%E2%80%93Huang_transform.

Mining Infrequent Itemsets from Text by Applying Recursive Learning and Fuzzy Logic

Prajakta V. Koti and G.A. Patil

Abstract Nowadays mining of infrequent itemsets, i.e., unrelated data items are gaining significance in encountering some powerful information and their importance in many applications like fraud detection, studying new emerging diseases, etc. Today text is growing from many fields and mining such text data can benefit in valid and purposeful data. This paper focuses on mining infrequent itemsets from text by using recursive learning and applying fuzzy logic. Recursive learning is used to set some suitable threshold value thereby unwanted itemsets are pruned and required itemsets are retained. M-hierarchical tree is formed based on infrequent itemsets and it generates different clusters of low support value infrequent itemsets. Finally fuzzy logic is applied to clusters to extract lowest value infrequent itemsets. These in turn help in increasing sales of rare items in shop, studying rare patterns of new emerging diseases, etc.

Keywords Text mining · Infrequent itemsets · Fuzzy logic

1 Introduction

Mining data is widely performed to extract important and beneficial information. Mining can be performed to extract frequent or infrequent items. Based on threshold value, the itemsets above threshold value are frequent and below are infrequent. Frequent itemset mining has wide variety of applications like market-basket analysis, education, etc. In recent years infrequent itemset mining is found to be significant in fraud detection, bioinformatics, etc.

P.V. Koti (✉) · G.A. Patil
Department of Computer Science & Engineering, D. Y. Patil College of Engineering & Technology, Kolhapur, India
e-mail: prajaktakoti@gmail.com

G.A. Patil
e-mail: gasunikita@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_19

Infrequent itemsets usually were unconcerned and excluded due to their unimportance. All infrequent itemsets are not useful but some infrequent itemsets may be found to be useful. For example, in book shop the owner identifies buying patterns of engineering students like many students are buying java book, few students are buying java book and ajax book and few students buying java book and story book, etc. The owner may notice ajax book is related to study so he may set some discount on ajax book, so the students buying java book may get attracted to discount and buy ajax book. Such rare pattern can be proving to be beneficial.

These days text is increasing from many applications like social media, government data, medical information, etc. Such text as it is cannot be used but by applying some data mining techniques these textual data can be used to extract valuable information.

So a system is proposed to mine infrequent itemsets from textual data by using recursive learning and fuzzy logic. Support value is calculated for input words. Recursive learning is perfomed to extract infrequent itemsets below threshold value and discard itemsets above threshold value. M-hierarchical tree is formed based on infrequent itemsets which generates different clusters with low support value infrequent itemsets. Finally fuzzy logic is applied to set new rules based on clusters of support value of itemsets. The range will be split into five different values. Two lowest values among the five values are considered. The final output is infrequent itemsets with lowest support value.

The intention to develop such system was developing an efficient method which will contribute to enhance mining of infrequent items from text by applying recursive learning and fuzzy logic approaches. This system may benefit in increasing sales of infrequent items, in medical field to prevent new diseases based on rare patterns etc.

2 Related Work

Luca Cagliero and Paolo Garza [1] have focused on mining infrequent itemsets from numerical datasets with use of fp tree. They modified algorithm IWI-Mining, Minimal IWI in such a way that they considered prefix structure storation of patterns, hence there is no need to rebuild fp tree. So it saved memory and time.

Suzan Wedyan [2] has contributed to enhance the associative classification algorithm with modifications and new concepts related to existing algorithms. Rules generation based on confidence and support. They compared algorithm considering rule learning, rule sorting, rule pruning, classifier building, and classifier prediction.

Soumadip Ghosh et al. [3] improved frequent itemset mining with use of genetic algorithm. Here authors made use of genetic algorithm to find frequent itemsets from dataset and they tested it on different large datasets which resulted correct and appropriate with respect to global search and time.

Xiaomeng Wang et al. [4] have proposed a method for finding frequent itemsets easily based on recursive procedure. They proposed recursive algorithm based on concept of step-by-step elimination of items from transaction database and main advantage of this algorithm was that it worked without any complex data structures. The limitation of this concept is step-by-step elimination is carried by using Fuzzy classification theory which actually neglected the low ranked frequent itemsets.

Paolo Ciaccia et al. [5] have developed a new method M- tree to search in large datasets. Their proposed method was for organizing and searching large datasets. Also the algorithms for insertion and split managements are proposed which always balance the M-Tree and it allowed efficiently to find out the nearest neighbor similarity measure.

Jiong Guo et al. [6] contributed to generate clusters based on distance. Their method is based on simple search tree strategy. Two reduction rules are presented for forming clusters. Two elements with same distance are added in same cluster and other pairs which will not appear are discarded.

Fernando Crespoa and Richard Weberb [7] elaborate a methodology for dynamic data mining based on fuzzy clustering, which allows updates of the underlying classifier. They applied given methodology for dynamic customer segmentation and dynamic traffic state identification using real-world data. Authors state that the given implementation is well suited for customer segmentation and traffic management.

Ladda Suanmali et al. [8] have proposed a method for improving text summarization based on numerical value. They have considered different features of sentences for summarization. Fuzzy logic is used to assign numeric values to sentences to improve quality.

3 Methods

Figure 1 depicts the system used to mine infrequent itemsets from given set of documents. Preprocessing is performed on set of documents to generate meaningful words. Then powerset of preprocessed words is generated. Support count for those itemsets is calculated. Recursive process is used to discard frequent itemsets and retain infrequent itemsets. M-hierarchical tree is formed which generates clusters with low support value itemsets. To this clusters fuzzy logic is applied to extract infrequent itemsets with lowest support value.

A. *Prepcosessing*

Preprocessing is performed to generate meaningful words. Given input consists of words such as the, is, a etc. These are not useful and meaningless so they are discarded. Following two techniques are used to discard unwanted words:

- Stemming: In this part the porter algorithm is used to eliminate suffixes like words ending with ing, ness etc.
- Stop word Removal: In this part, the bag of 525 words is used which consist of stop words stored in array. The input files provided, the words present in that file

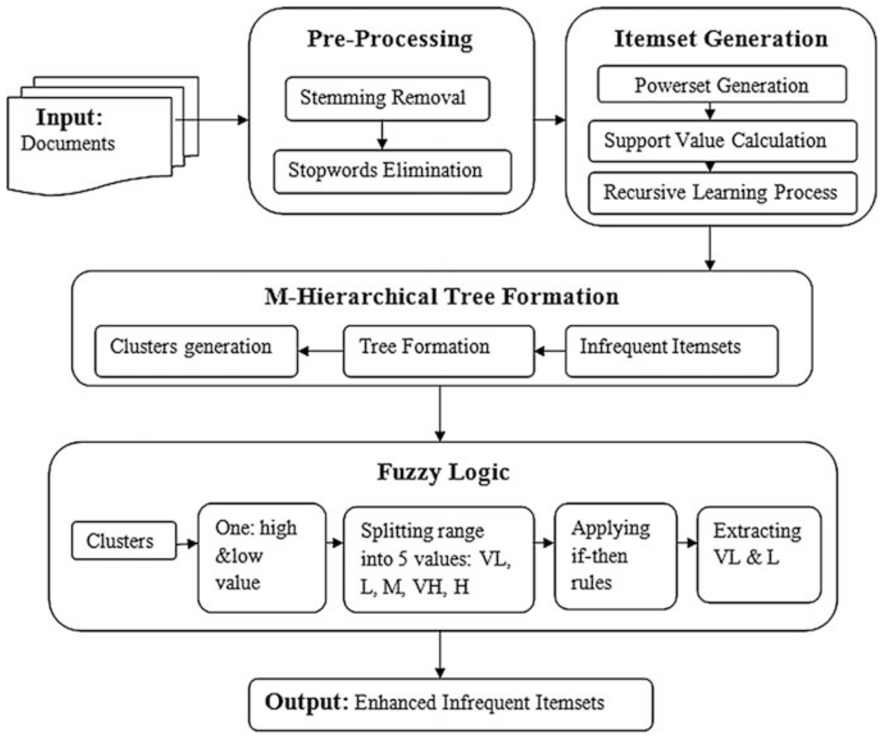


Fig. 1 System architecture

are checked with bag of words, if any stop word is present then that word is deleted.

B. *Itemset Generation*

Infrequent itemsets are generated by applying powerset algorithm and recursive learning process.

- Powerset Generation

Poweset is considered as it generates all possible combinations of words. The powerset of preprocessed words is generated by using following equation:

$$2^I - 1 \tag{1}$$

where I is set of words. It generates all subsets of words except the empty set. The subset generated is represented by an array of Boolean values of the same size of set, where 1 indicates presence of word in that set and 0 indicates the absence of that word in the set.

- Support Calculation

$$\text{Support} = \frac{\text{Occurence}}{\text{Total Support}} \quad (2)$$

The support for itemsets is calculated by applying above Eq. (2). Where occurrence is the item present in the transaction and total support is the number of itemsets in the transaction. Here the support count for each item is calculated and stored in array.

- Recursive Learning & Pruning

Threshold value in between 0 and 1 is supplied. Decision tree of itemsets is generated. Itemsets below threshold value are added to left side of tree and rest to the right. The right side of tree is discarded as they are frequent itemsets. The left node of all nodes into left subtree is considered. These left node itemsets are infrequent itemsets and are stored. Counter is maintained to get number of infrequent itemsets found.

C. *M-hierarchical Tree Formation*

M-hierarchical tree is formed by considering infrequent itemsets generated by itemset generation module. The root node is formed by selecting first itemset from these generated itemsets with support value. The insertion of infrequent itemsets in tree is based on comparing support values of itemsets. If the itemset has less support value than root node's support value, that itemset will be added to left side of tree. If the itemset support value is more than root node's support value then that itemset will be added into right side of the tree. Recursively the itemsets will be added into left and right subtree until all infrequent itemsets are added into tree. In this way, the tree will be generated. Preorder traversal is performed on generated tree where first root node is visited then left node and finally right node is visited. After traversal of tree finally clusters are generated consisting of itemsets with low support values. Numbers of clusters are not predefined. The cluster formation varies based on supplied threshold value. Each cluster consists of different itemsets with same support value.

Following example can be considered for generating m-hierarchical tree.

Table 1 shows the different infrequent itemsets with their support value that are used in generating tree.

Figure 2 represents the different levels of tree generated. The first itemset is considered as root node. The insertion of next node in tree is based on comparing value of support value of itemset. If itemset has less support value than root node then it is added to the left of tree else to the right of tree.

D. *Fuzzy Logic*

Cluster consisting of itemsets with support value generated by m-hierarchical is considered. Among all the clusters values one highest and one lowest support values are set. Then these values are split into range of five membership values as:

Table 1 Itemsets with support value

Itemsets	Support value
strong	0.251
cash	0.125
comp	0.039
earn	0.016
conti	0.063
pric	0.267
xone	0.251
out	0.031
pct	0.335
improv	0.0080
cit	0.0040
invest	0.0075

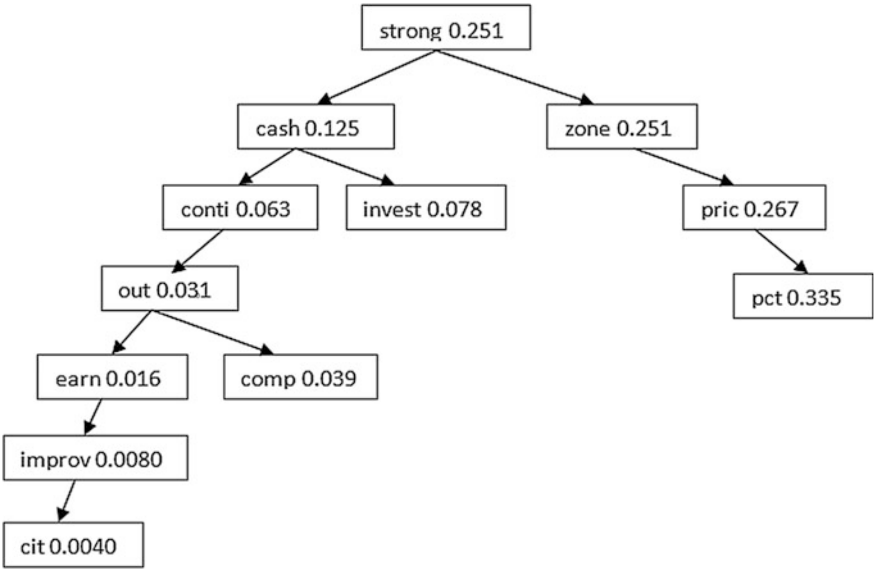


Fig. 2 M-hierarchical tree generation

very low, low, medium, high, and very high. By applying if-then rules itemsets with very low and low membership value are extracted. This way the rare itemsets with lowest support value are mined. For example we have clusters with different support values as: {c1 0.063}, {c2 0.016}, {c3 0.040}, {c4 0.023}, {c5 0.025}. One high and one low value is considered, i.e., 0.063 and 0.016. This range is split into five membership values as very low {0.016-0.025}, low {0.026-0.0344}, medium {0.345-0.0438}, very high {0.0439-0.0532}, high {0.0533-0.063}. By applying if-then rules, itemsets with very low and low are extracted as they are itemsets with very low support values.

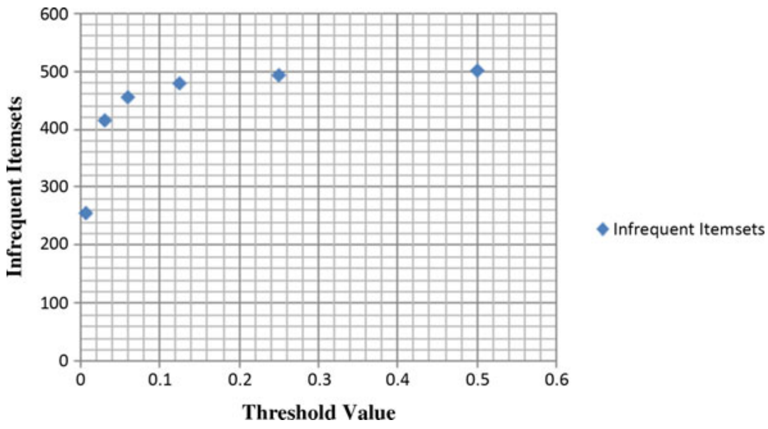
4 Evaluation of Results

The reuters dataset is used which consists of news. Each file consists of news which has date, title and body part. The body part from each file is considered as it contains news description onto which all methods are applied for getting output infrequent itemsets.

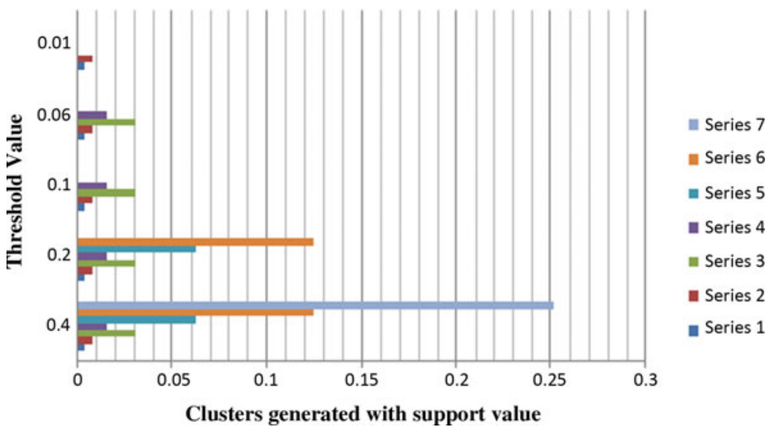
Graph 1 depicts infrequent itemsets generated with different threshold values.

Graph 2 depicts the clusters generated with different support values.

Table 2 shows the accuracy of infrequent itemsets generated for different threshold values. Maximum accuracy is achieved for 0.2 threshold value.



Graph 1 Number of infrequent itemsets generated



Graph 2 Clusters generated with support values

Table 2 Accuracy

Threshold value	Accuracy (%)
0.2	90
0.3	87
0.4	81

5 Conclusion

In this paper, we have proposed a system to mine infrequent itemsets from textual data with use of recursive learning process and applying fuzzy logic. The experimental results illustrate the number of infrequent itemsets generated for different threshold values. The clusters are generated with low support value and are varied for different threshold value. So it leads in extraction of accurate and rare infrequent itemsets.

References

1. Luca Cagliero and Paolo Garza “Infrequent Weighted Itemset Mining using Frequent Pattern Growth”, IEEE Transactions on Knowledge and Data Engineering, pp. 1–14, 2014.
2. Suzan Wedyan “Review and Comparison of Associative Classification Data Mining Approaches” World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:8, No:1, 2014.
3. Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar, Partha Pratim Sarkar “Mining Frequent Itemsets Using Genetic Algorithm” International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.4, October 2010
4. Xiaomeng Wang, Christian Borgelt and Rudolf Kruse “Fuzzy Frequent Item Set Mining based on Recursive Elimination”.
5. Paolo Ciaccia Marco Patella Pave1 Zezula “M-tree: An Efficient Access Method for Similarity Search in Metric Spaces”.
6. Jiong Guo, Sepp Hartungy, Christian Komusiewicz and Rolf Niedermeier “Exact Algorithms and Experiments for Hierarchical Tree Clustering”.
7. Fernando Crespoa, Richard Weberb; “A methodology for dynamic data mining based on fuzzy clustering”.
8. Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan “Fuzzy Logic Based Method for Improving Text Summarization”.

Mitigating Spam Emails Menace Using Hybrid Spam Filtering Approach

Stanlee Nagaroor and G.A. Patil

Abstract Spam is a known problem to email users. Sending spam emails is one of the easiest form of advertising and hence a useful medium of communication like email is abused by spammer to send junk email. Most antispam solution focus only on analysing text of the body of email messages for detecting spam emails. We have developed a spam filter that separates spam from non-spam (ham) emails by analysing header, URL, body, and attachments. Header and URL are checked against rules and text of body and attachments are checked by Bayesian classifier and Apriori algorithm. Only (*.rtf, *.txt, *.docx, *.doc,*.pdf) attachment files are examined. The experimental results reveal that checking attachments of emails played significant role in spam detection and hence attachment checks should be extended to more file types for better spam detection.

Keywords Email · Spam · Bayesian · Apriori

1 Introduction

Spam email can be defined as unsolicited or bulk mail sent for the purpose of advertising product or for other malicious intent. According to commtouch [1] Internet threats trend report, 97.4 billion spam emails were sent each day during the first quarter of 2013. Most of the spam emails are designed to solicit money from the recipients and to achieve this they offer products that claims to miraculously cure health problems like diabetes, obesity, hair fall, etc. Spammers also know that humans are greedy by nature and hence they come up with get-rich schemes which claims to make you rich in no time. Spam emails also contain sexual content that

S. Nagaroor (✉) · G.A. Patil
Department of Computer Science & Engineering, D. Y. Patil College of Engineering & Technology, Kolhapur, India
e-mail: stanlee.n@gmail.com

G.A. Patil
e-mail: gasunikita@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_20

involve pornography or promote some drugs or products that could enhance sexual performance. Some of the countries have enacted laws like the CAN-SPAM act, The CAN-SPAM Act is a law that sets the rules for sending of commercial email and has provision of penalties for violations. However all countries do not have an antispam law. Even if such law exists it is not implemented effectively. According to spamhaus [2], United States, China, and Russia are top three generator of spam emails.

2 Related Work

There are many approaches to deal with the problem of spam and different researchers have come up different solutions and different algorithms to detect spam. Most of the work focuses on analysing the content (body of email) message. These include using algorithm like naïve Bayes [3, 4] for classifying email messages into spam and non-spam emails. Some of the algorithms that are commonly used for classification of emails are SVM [5], RF [6], K means cluster [7], K nearest neighbour [8].

Some researchers focused on extraction of features from header [9, 10, 11] to find if it can be used for spam detection. Researchers have also focused on using digest [12, 13] of emails to detect spam emails. Structural features [14] of emails are also used for detection of spam emails. A system was also developed for detecting phishing URL [15] based on Lexical-based features, Keyword-based features, Reputation-based features, and Search Engine based features. Other techniques use block list like Domain Block List (DBL) the block list of the spamhaus [16] project is used by several organisations and email service providers. To best of my knowledge, researchers have written too little on attachments spam and techniques for detection of attachment spam is not made public by those software companies that claim to detect attachment spam.

3 Methodology

Spammers want readers to respond to their mails, hence most of the spammers provide URL in the message body that direct to spam website. Hence examining URL is important in spam detection. Also spammers use fake headers to hide their identity or to pretend to be someone else. Using black list or white list against sender email address will not give correct results, hence other header fields should also be examined. The proposed spam filter will examine other fields in headers too.

Emails from mail server are read using IMAP protocol. The spam filter examines headers and URLs of emails for classification. Only some emails are classified by header and URL checks alone. Emails which cannot be classified by header or URL checks undergo body and attachment analysis where Bayesian classifier is used for classification. The emails which are classified as non-spam (ham) by Bayesian classifier is further examined for the presence of spam-associated words (generated by Apriori algorithm) which will help to detect spam emails not detected by

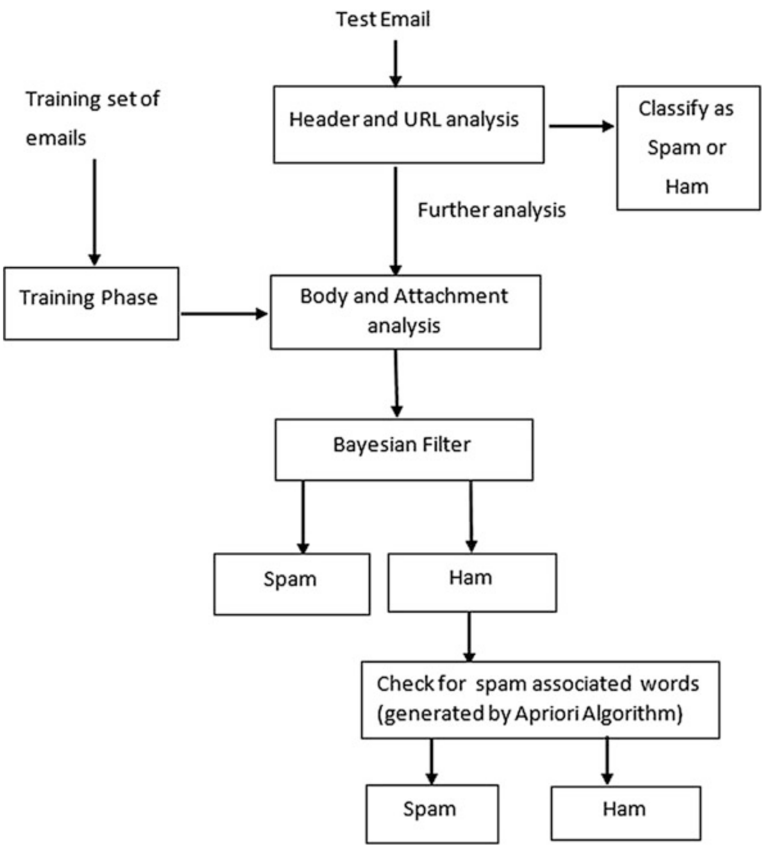


Fig. 1 Architecture of spam filter

Bayesian classifier. Emails that are classified as ham by Bayesian classifier and if it does not have spam associated words it will be considered as ham (Fig. 1).

We can examine body and attachments only after training the spam filter using training dataset of spam and ham emails. The knowledge gained by the filter during the training phase will be used for analysis of body and attachment in the testing phase.

Section 3.1 gives description of training of our spam filter while Sect. 3.2 describes the process of detection of spam emails by analysing different parts of email messages.

3.1 Training of Spam Filter

Bayesian classifier must be trained to distinguish spam emails from non-spam (ham) emails. To train the classifier we must provide two datasets (spam dataset and ham dataset). Bayesian classifier scans through the training sets of ham and spam emails and takes note of the words occurring in the datasets and also the proportion in which these words appear in dataset. This knowledge helps to assign proper probabilities to words.

Each email in spam or ham dataset goes through following steps:

I. Removal of html tags.
II. Removal of stop words.
III. Stemming of words.
IV. Storing words in database and calculating probability.

To reduce computation cost and unnecessary space overhead, stop words and html tags are removed from the emails. To correctly take in account the occurrence of a particular word in spam and ham message the words are stemmed, i.e., words like waiting, waited are reduced to their root form wait.

The details of each steps in training are given below

I. Removal of HTML tags and entities

Regular expression are used to replace or remove HTML tags. HTML tags like <script></script>, <style></style> are completely removed along with the text that falls between the opening and closing tags. Some tags are replaced by appropriate text if a particular pattern matches, for example,
 tag is replaced with "\n". Regular expressions are also written to handle entities, special symbols, for example, < is replaced with '<'.

II. Removal of stop words

Stop words are those words that do not help in classification and hence are removed. Examples of stop words include articles, conjunction, pronouns, and preposition. If the email contains stop word it is removed from the email as a part of pre-processing step. Removal of these words reduce computational and space overhead.

III. Stemming of words

After removal of HTML tags and stop word from the text, the words that contains only numbers are removed at this stage. Also words are reduced to their root form using Porter stemming algorithm.

IV. Storing words in database

After the messages in spam dataset passes through above three steps. The remaining words (tokens) are stored in spam table of database. If a single email contains duplicate words, it is removed before storing it in spam table of database. By removing duplicates the frequency corresponding to a word will indicate how many spam emails contains that word for example, if word 'free' occurs four times in spam email one and three times in spam email two. The spam table will contain entry <free,2> instead of <free,7> indicating 'free' word is seen in two spam mails.

Similarly, emails from ham dataset is processed by removing HTML tags, removing stop words and applying stemming to the word. The remaining words are stored in ham table of database. If a single email contains duplicate words it is removed before storing it in ham table of database.

Probability that a word is spam is calculated using formula,

$$P\left(\frac{S}{W}\right) = \frac{P\left(\frac{W}{S}\right) . p(S)}{P\left(\frac{W}{S}\right)p(S) + P\left(\frac{W}{H}\right)p(H)} \tag{1}$$

where,

1) *P(S/W)* is the probability that a message is a spam, knowing that the word W is in it.
2) *P(S)* is the overall probability that any given message is spam.
3) *P(W/S)* is the probability that the word W appears in spam messages.
4) *P(H)* is the overall probability that any given message is not spam (ham).
5) *P(W/H)* is the probability that the word W appears in ham messages.

There is no prior reason for an incoming mail to be considered as spam or ham so we assign equal probabilities for *P(S)* and *P(H)*, hence *P(S)* = *P(H)* = 0.5. The above Eq. (1) becomes

$$P\left(\frac{S}{W}\right) = \frac{p\left(\frac{W}{S}\right)}{P\left(\frac{W}{S}\right) + P\left(\frac{W}{H}\right)} \tag{2}$$

Apriori Algorithm is used to generate sets of spam associated words from training data set of spam emails. To apply this algorithm to find association between spam words, we have to create our transaction database from training dataset of spam emails by selecting spam words from each training email (all spam words in single mail will be considered as a single transaction). Transaction database (T) will contain several such transaction. Then we will use Apriori algorithm to generate sets of spam associated words like {win, lottery}. This indicates that spam words won and lottery are closely associated (or related).

Attachments will be read and text from the attachments will be analysed using Bayesian and Apriori algorithm for classifying emails.

3.2 Detection of Spam Emails

Following checks are performed on headers of emails to detect spam emails:

1. SPF checks: Sender Policy Framework (SPF) is an email validation technique used to detect email spoofing by providing a mechanism to check that incoming email from a domain comes from a host authorised by that domain's administrators. If an email claims to come from a certain domain but the IP address of sender is not in SPF record published by domain's administrator then mark that email as spam.

2. MX record checks on the ‘from’ field of the message: The reason for performing MX record check is that some spammers use nonexistent domain names in ‘from’ field of header. To check whether the domain in the ‘from’ field has a mail exchange record we send a query to DNS server asking for MX records for the domain in the ‘from’ field. If we get MX record details from DNS server then domain name in ‘from’ field exists otherwise it means that nonexistent domain is used in ‘from’ field and hence the message is marked as spam.
3. Check if an email is a real reply email: If the subject contains ‘Re:’ check if the ‘In-Reply-To’ header is present. In a genuine reply message ‘In-Reply-To’ header contains message id of email for which this message is a reply. Spammers sometimes use ‘Re:’ in their subject but it is a fake reply message. Hence the ‘In-reply-to’ field is not present in the header of spam email or it does not contain value in it.

In such cases mark that email as spam.

4. Detect spam pattern in subject. Regular expressions are used to detect presence of spam like pattern in subject.
5. If an email passes SPF checks and domain in ‘from’ field and ‘return-path’ matches, then the domain is compared against those in white list. If the domain is present in white list, mark email as ham.
6. Check headers against black list: check address in ‘received’ fields and ‘from’ field against black list. If the domain is present in black list mark that email as spam.

If content type of email body is html then first an Html Document object is created and then descendants ‘form’ nodes, ‘anchor’ nodes and ‘area’ nodes are extracted. If URL is present in form’s action attribute or anchor’s href or area’s href attribute, then these URL’s are extracted and following checks are done on URL.

I. Domain name from the URL is compared against black list. If the domain name is present in the black list mark email as spam.
II. Check if text and href of anchor contains URL and whether these URL’s point to different domains.

For example, <ahref=http://www.scamsite.com>www.paypal.com. Spammers use such technique to direct user’s to phishing site. If the URL in href and text point to different domain’s mark that email as spam.

The text from body of email will be extracted. Attachments of email are read as sequence of bytes, then the bytes are decoded and formatting information is removed to get data in plain text. Different file types have different formatting info. Only (*.rtf, *.txt, *.docx, *.pdf) attachment files are examined. Attachments are

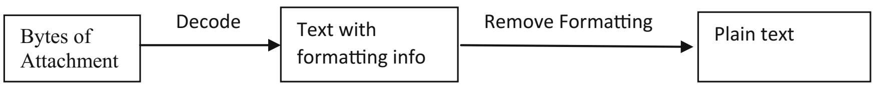


Fig. 2 Conversion of bytes in attachment to plain text

read as sequence of bytes and it is decoded to get text with formatting information and latter the formatting information is removed to get plain text as shown in Fig. 2.

The text from body or attachment will be preprocessed (tokenization, stop word removal, html tags removal, and stemming). The description of preprocessing steps are given in Sect. 3.1.

Spamicity of words in email will be calculated by formula (2).

Probability (p) that an Email (E) containing words $\{W_1, W_2, W_3, \ldots W_n\}$ (in body or attachment) is spam is given by formula (2).

$$P = \frac{P\left(\frac{S}{W1}\right)P\left(\frac{S}{W2}\right)P\left(\frac{S}{W3}\right)\ldots P\left(\frac{S}{Wn}\right)}{P\left(\frac{S}{W1}\right)P\left(\frac{S}{W2}\right)\ldots P\left(\frac{S}{Wn}\right)\ldots + \left(1 - P\left(\frac{S}{W1}\right)\right)\left(1 - P\left(\frac{S}{W2}\right)\right)\ldots\left(1 - P\left(\frac{S}{Wn}\right)\right)} \tag{3}$$

where,

$P\left(\frac{S}{W1}\right), P\left(\frac{S}{W2}\right), P\left(\frac{S}{W3}\right), \ldots, P\left(\frac{S}{Wn}\right)$, are spamicity of words $\mathrm{W}_1, W_2, W_3, \ldots W_n$.

Emails having value of p above a threshold are considered as spam. Rest of the emails are considered ham by Bayesian classifier.

The emails that are considered as ham by Bayesian classifier are further examined for the presence of spam words association. The spam words from emails are retrieved and checked against spam association rules (or sets of spam associated words, generated by Apriori algorithm). If association is present then email is marked as spam else it is marked as ham.

4 Experimental Results

We use the following parameters to measure performance of spam filter:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

where,

TP = True positive, i.e., spam emails correctly identified as spam
TN = True Negative, i.e., ham emails correctly identified as ham
FN = False Negative, i.e., spam emails that are not correctly identified
FP = False Positive, i.e., ham emails that are not correctly identified

Total Emails in dataset used for Testing: 250, Ham emails in dataset: 200, Spam emails in dataset: 50. The results given by our system is shown in Table 1.

Table 1 Correctly classified and incorrectly classified emails

	Correctly classified	Incorrectly classified
Ham	193 (TN)	7 (FP)
Spam	45 (TP)	5 (FN)

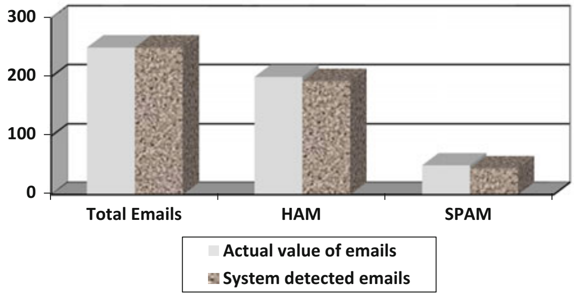


Fig. 3 Comparison of actual value and system generated result

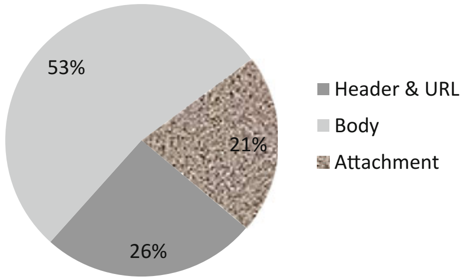


Fig. 4 Distribution of spam emails detected by analysing different parts of email messages

$$\text{Accuracy} = (193 + 45)/250 = 0.952 = 95.2\%$$
$$\text{Precision} = (45)/(45 + 7) = 0.865 = 86.5\%$$
$$\text{Recall} = 45/(45 + 5) = 0.9 = 90\%$$

The column chart (Fig. 3) shows the number of actual ham and spam emails in dataset and ham and spam emails correctly detected by the system.

The pie chart shown in Fig. 4 shows how each part of email messages helped in detection of spam emails. This diagram shows that 53% of spam emails were detected by analysing the body, 21% of the spam emails were detected by analysing attachments and remaining percent of spam emails were detected by header and URL checks.

5 Conclusion and Future Work

Spam emails are unwanted emails and needs to be weeded out if we have to reinstate confidence of the people on this useful medium of communication. This paper discussed the different approaches used by researchers to deal with the problem of spam. This paper gives details on our approach to solve the problem by performing checks on header and URL using rules and analysing the text of body and attachments using Bayesian classifier and Apriori algorithm. Our future work will be on extending our work of analysing attachments on other file types and checking their impact on overall spam detection of spam filter.

References

1. "Internet threats trend report", Commtouch, April 2013.
2. https://www.spamhaus.org/statistics/countries, March 2016.
3. G. Bhagyashri and H. Pratap "Auto emails classification using Bayesian filter", International Journal of Advanced technology & Engineering Research, vol 3, Issue 4, July 2013.
4. Smera Rockey and Rekha Sunny "A hybrid spam filtering technique using Bayesian spam filters and aritificial Immunity Spam Filters" International Journal of engineering research and technology", vol 3, Issue 5, May 2014.
5. Sushama Chouhan, "Behavior Analysis of SVM Based Spam Filtering Using Various Kernel Functions and Data Representations", IJERT, vol. 2, Issue 9, September – 2013.
6. B.Gaikwad and P. Halkarnikar, "Random Forest Technique for E-mail Classification", International Journal of Scientific & Engineering Research, vol 5, Issue 3, March-2014
7. Nadir Omer, Othman Ibrahim and Waheeb "An improved spam email classification mechanism using K means clustering", vol 60, No 3, February 2014.
8. Ali Aski "A Proposed Algorithm for Spam Filtering Emails by hash table approach", International Research Journal of Applied and Basic Sciences vol.4 (9), 2436–2441, 2013.
9. Omar Al-Jarrah, Ismail Khater and Basheer Al-Duwairi "Identifying Potentially Useful Email Header Features for Email Spam Filtering", ICDS 2012: The Sixth International Conference on Digital Society.
10. Alberto Treviño, J. J. Ekstrom "Spam Filtering Through Header Relay Detection", Brigham Young University.
11. Fernando Sanchez,Zhenhai Duan,Yingfei Dong "Understanding Forgery Properties of Spam Delivery Paths".
12. E. Damini,S. De Capitani di Vimercati and P.Samarati "An Open Digest-based Technique for Spam Detection", DTI - University di Milano - 26013 Crema, Italy.
13. E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. "An open digest-based technique for spam detection". In Proceedings of The 2004 International Workshop on Security in Parallel and Distributed Systems, San Francisco, CA, USA, September 2004.
14. Sarju,Riju Thomas and Emilin shyni "Spam email detection using structural featues", International Journal of computer Applications,vol 89, issue 3, March 2014.
15. Ram B. Basnet, Andrew H. Sung, Quingzhong Liu "Learning to detect phishing URL" IJRET: International Journal of Research in Engineering and Technology.
16. https://www.spamhaus.org/
17. G.SenthilKumar, S.Bhaskar and M.Rajendran "Online Message categorization using Apriori algorithm", International journal of computer trends and technology, June 2011.

Novel HBM Spacing Policy for Fuzzy Based Forward Collision Avoidance System in Vehicle

S. Paul Sathiyan, S. Suresh Kumar and A. Immanuel Selvakumar

Abstract Use of fuzzy controller offers the advantage of assimilating the human intelligence for nonlinear complex system like autonomous vehicle. In order to prevent rear end collision of vehicles travelling in the same direction, Forward collision avoidance system is introduced which uses automated throttle and brake adjustment. Occupant comfort and safety are used as performance measures for assessing the performance of the system. The performance of the system depends on the spacing policy which describes about the distance to be maintained between the vehicles (lead and host) travelling in the same lane and direction of travel with respect to the velocity of travel. This research work proposes a novel Human Behavioural Model to determine the spacing between the vehicles and a heuristic-based Fuzzy controller since it does not require a detailed mathematical model of the system. Test result shows that the controller with the proposed method has shown an improved performance over the other existing spacing policies.

Keywords Forward collisioin avoidance system; spacing policy
Fuzzy logic · Vehicle occupant comfort and safety

1 Introduction

According to the investigation conducted during 2012–2015 by National Transportation Safety Board (NTSB) USA, 28 fatalities and 90 injuries are reported in nine rear-end collisions. Accidents on the roads can be reduced to a larger extend when drivers are provided assistance with intelligent systems during driving [1, 2]. In the recent decades, Driver Assistance System (DAS) is gaining research focus. This system can be classified into passive assistance system and active assistance system.

S. Paul Sathiyan (✉) · A. Immanuel Selvakumar
EEE, Karunya University, Coimbatore, India
e-mail: sathiyanme@yahoo.co.in

S. Suresh Kumar
ECE, Dr. N.G.P. Institute of Technology, Coimbatore, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_21

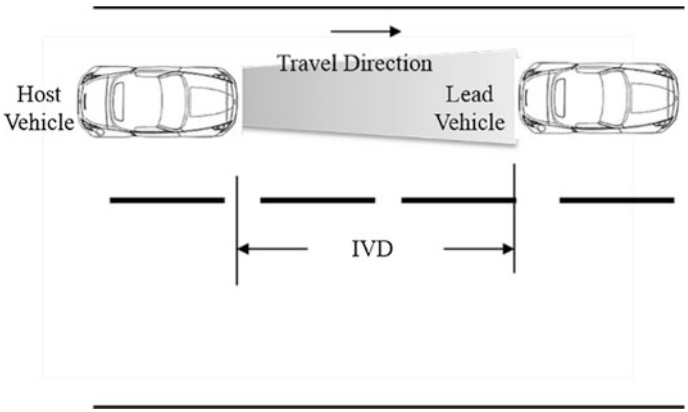


Fig. 1 Vehicle following scenario

Passive assistance system provides audio /visual warning of a collision risk. Active safety system uses autonomous braking and throttling for collision risk reduction. The risk of accidents can be reduced by 14% when vehicles are fitted with autonomous braking systems [2]. Congestion on a highway can be greatly avoided when 25% of the vehicles in highway are equipped with DAS. Forward Collision Avoidance System (FCAS)is an example for active safety system and it is developed by integrating throttle control and braking control. FCAS uses a LASER, RADAR or camera-based vision system for measuring the distance between two vehicles (lead and host) travelling in the same direction and lane as shown in Fig. 1.

FCAS fitted in the host vehicle, senses the Inter-Vehicle Distance (IVD) and adjusts the throttle and brake of the vehicle automatically and reduces the potential of rear-end collision. Multiple input multiple output fuzzy controller is used for as vehicle controller (Fig. 2). The actual velocity of the host vehicle (H_V_{vel}) and the actual distance measured between the two vehicles (M_{dist}) using sensor is used to compute the target velocity (T_{vel}) and desired distance (D_{dist}). From these parameters the velocity (V_e) and distance errors (D_e) are computed. These values are used by the fuzzy controller to compute the throttle/brake value.

2 Spacing Policies

Spacing policy has an influence on vehicle safety while following a lead vehicle, highway traffic flow, and user acceptance of the driver assistance system [3]. It describes about the distance to be maintained between the vehicles (lead and host) travelling in the same lane of travel with respect to the host vehicle velocity. Excessive spacing may lead to “cut ins” of vehicles from other lane, which make the worthiness of the controller to be questioned. Driver feels unsafe when the

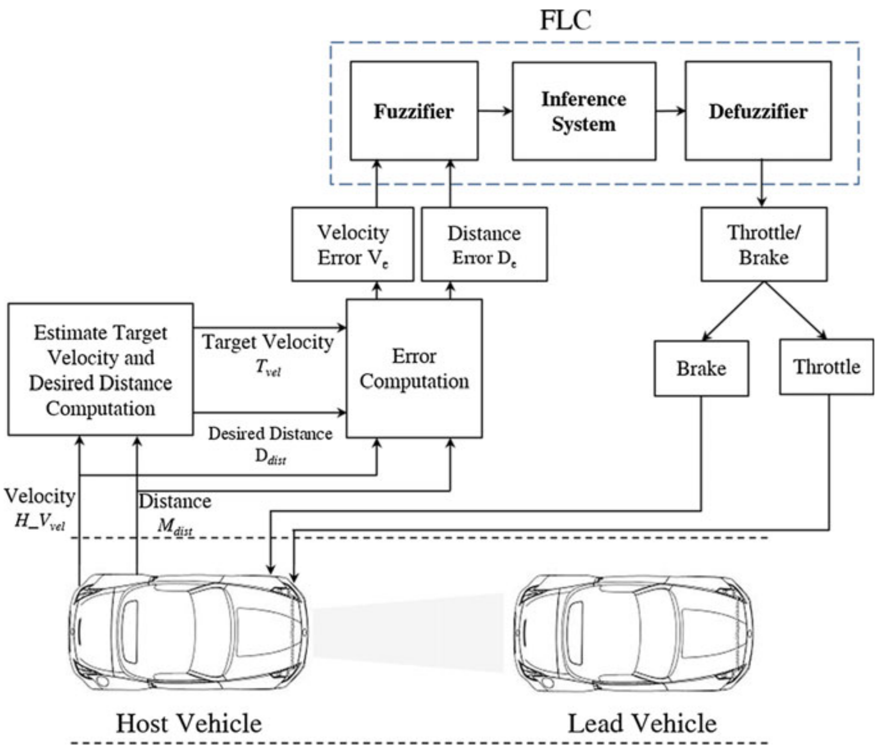


Fig. 2 MIMO FLC framework for VLC

spacing is moderate. The preliminary step in developing active driver assistance system is to determine the safer inter vehicle distance that needs to be maintained between the vehicles travelling in the same lane. The following equations help in determining these two errors:

$$V_e = T_{\text{vel}} - H\_V_{\text{vel}} \tag{1}$$

$$D_e = D_{\text{dist}} - M_{\text{dist}} \tag{2}$$

where H_V_{vel} is the host vehicle velocity and M_{dist} is the actual distance measured between the vehicles using sensors. In order to determine the target velocity T_{vel} and desired distance D_{dist} (Fig. 2), literatures provide two methods to generate the same. Constant time gap policy [4, 5] and variable time gap policy [6–8].

A. *Constant time gap policy (CTG)*

CTG (Fig. 3) is the most commonly used spacing policy by the researchers and automotive manufacturers. According to CTG spacing policy, the spacing error is given by

$$\delta_i = h * H_{V_{\text{vel}}} + L - M_{\text{dist}} \tag{3}$$

where h is the Time Head Way (THW) in seconds which is defined as the difference between the time when the front of a vehicle arrives at a point on the highway and the time the front of the next vehicle arrives at the same point (in seconds), L length of the vehicle in metres. Different time gaps in seconds with respect to operating zones were recommended in [9]. For highway the author recommends the use of 1.50–2.19 (s) as time gap and for rural 1.66–3.21 (s). Hence the driver should possess the knowledge about THW setting according to the operating zone. Improper setting will also lead to undesirable effect. Driver's intervention is required for changing THW whenever the vehicle crosses operating zone during travel. CTG assumes that both the lead and the host vehicle have identical and constant decelerations during braking manoeuver. This assumption produces solutions with high jerks and consequently, low comfort [10]. Further this model could produce undesirable large inter-distances reducing the lane capacity at lower speed. Spacing distance is considered linear throughout the entire speed of operation, whereas the vehicle braking characteristics may not be linear at all speeds.

B. *Variable time gap policy (VTG)*

The inter-vehicle spacing would be a nonlinear function (Fig. 3) of the host vehicle velocity. Equation (4) defines the VTG spacing policy as

$$S(\dot{x}_i) = \frac{1}{\rho_m \quad \left(1 - \frac{\dot{X}_i}{V_f}\right)} \tag{4}$$

where ρ_m is a density parameter and V_f is a speed parameter. Value of $\rho_m = 1/L$, $L = 5$ and $V_f = 75$ mph are used. One can interpret from Fig. 3 that spacing increases with velocity in the case of the VTG policy, but not proportionally. This

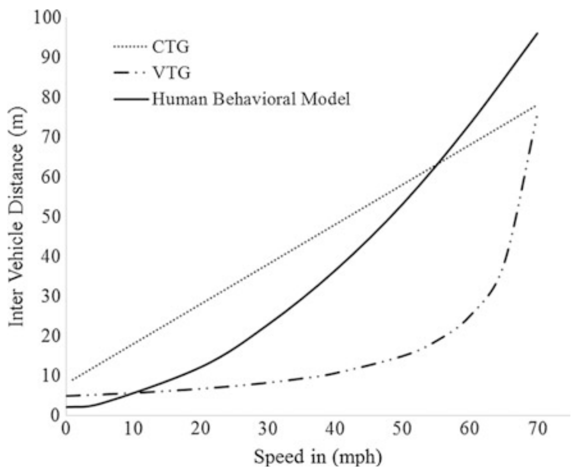


Fig. 3 Inter-vehicle spacing policies

Table 1 Human behaviour model spacing policy

Velocity (m/s)	Perceiving distance (m)	Braking distance (m)	Overall stopping distance (m)
8.94	6	6	12
13.4	9	14	23
17.9	12	24	36
22.4	15	38	53
26.8	18	55	73
31.3	21	75	96

spacing policy allows a higher traffic capacity trading off safety. Spacing distance is less which alarms safety from driver's perspective. Though the severity of collision is less at lower speeds, still it leads to rear-end collision.

C. *Proposed Human Behavioural Model Spacing Policy*

The proposed Human Behavioral Model (BHM) for inter vehicle spacing is designed to achieve both safety and traffic capacity in all operating zones (urban rural and highway). It uses the "two second rule" technique. Especially at urban operating zones (city driving), when the speed of operation is less, the proposed spacing policy allows a lower spacing than the counterpart, thus ensuring more number of vehicles. Meanwhile, maintaining sufficient distance between the vehicles so that the driver can react within two seconds under any hazardous situation. The relation between the spacing distance and speed of travel is given as numerical in Table 1 and plotted graphically in Fig. 3. The proposed controller allows to have a higher IVD at higher speeds than the other two counter parts.

3 Fuzzy Controller for FCAS

Heuristic approach based Fuzzy Logic controller (FLC) is used as the controller in this research work. In [11] a FLC with two inputs (Distance Error and Velocity Error), two outputs (Brake command and Velocity set point ratio), and seven triangular membership functions for both the inputs and seven singleton membership functions for the output has been discussed. The author had used 42 rules in the rule base. Same set of input variables with five triangular membership functions for the input and the output is being discussed in [12]. The author had used brake and throttle (Gas) as output variables. The rule base was populated with 25 rules. 81 rule fuzzy method is presented in [13] to determine a suitable throttle and brake control for ACC systems. Several research articles had shown that the FLC had been developed with triangular membership function with Distance Error and Velocity Error as input and throttle/brake command as output. Hence, FLC controller is employed in this research with the spacing policy shown in Sect. 2. Fuzzy controller with the following rule

Table 2 Fuzzy rule table

Distance error	Velocity error	Throttle/Brake
NVL	NL	PL
PS	NM	NL
PVL	NL	PVL
PL	PS	PVL
NM	NS	NS
NVL	PM	NL
NM	PS	PL
NM	NVL	PM
NVL	Z	NL
PVL	PL	PL
PVL	PM	NM
PL	NS	PS
NVL	NL	PM
NVL	Z	PM
PM	PM	PS
PL	NS	Z

table (Table 2) is used for testing the performance. Nine membership functions are used to define the input (Distance and Velocity Errors) and output ranges (throttle/brake). The input V_e and D_e ranges between ± 38 m/s and ± 96 m respectively. Output variable *t/b* varies between ± 1.

4 Test Scenarios and Simulation Result

Computer-based prediction of performance and range allows to quickly experimenting various aspects of the vehicle. For modelling and simulation of dynamic systems and process MATLAB/Simulink provides a powerful environment. All wheel-drive vehicle model available in Matlab/Simulink is used in [14]. Longitudinal model of the vehicle is used in [15] to study about the vehicle longitudinal dynamics control for improved vehicle safety. Padash in [16] used the MATLAB/Simulink model to develop and study about the fuzzy based adaptive cruise control for vehicle. Hence this research uses a vehicle model available in MATLAB/Simulink for designing and analysing the proposed work.

A. *Test Scenarios*

Since driving pattern represents the driving behaviour of a location, they are used as one of the test scenario along with manoeuvring scenario. Modem.urban8 driving cycle has a duration of 250 s, maximum and average velocities of 53.5 and 15.879 km/h. As discussed earlier the occupant comfort and safety is considered as performance measures. These test scenarios are shown in Figs. 4, 5, 6.

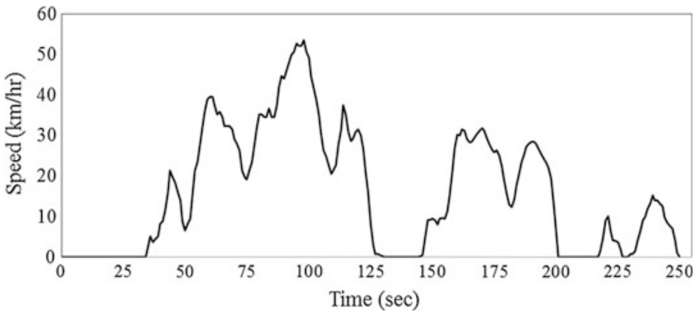


Fig. 4 Modem.Urban8 driving pattern

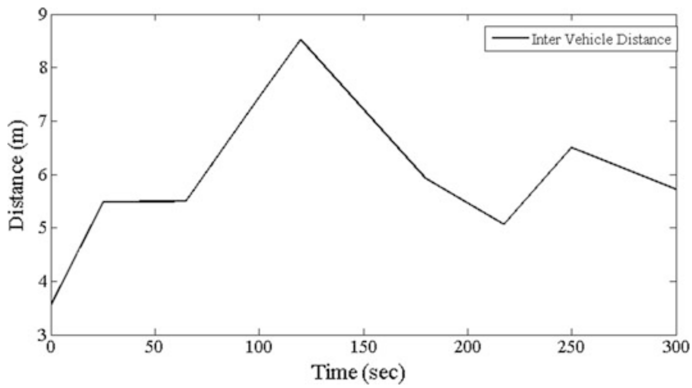


Fig. 5 IVD pattern 1

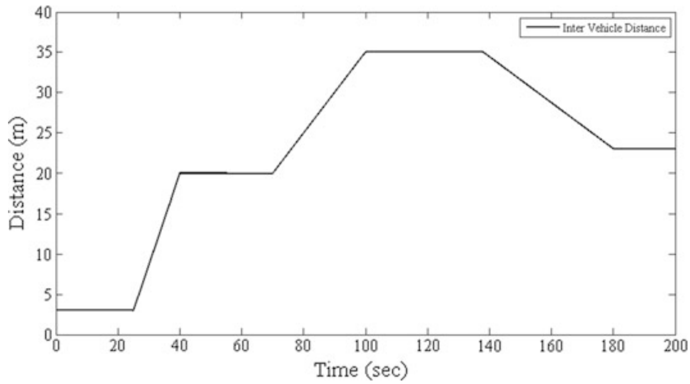


Fig. 6 IVD pattern 2

Table 3 Performance comparison

		Modem.urban8 driving pattern			IVD pattern 1			IVD pattern 2		
Spacing policy		CTG	VTG	Proposed HBM	CTG	VTG	Proposed HBM	CTG	VTG	Proposed HBM
Max acceleration	LV	1.833	1	1.833	1.818	1	1.818	1	1	1
	HV	37.155	37.155	1.903	37.155	37.155	2.246	37.155	37.155	1.495
Max jerk	LV	1.611	1	1.611	1.818	1	1.818	1	1	1
	HV	36.716	36.716	1.354	36.716	36.716	2.246	36.716	36.716	1.495
Max PD		373.657	567.731	0.472	371.657	565.731	0.596	353.943	546.761	0.661
Max spacing		1	1	0.133	1	1	1.780	1	1	1
MSVE		0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001
MSDE		0.003	0.003	0.002	0.003	0.003	0.003	0.003	0.003	0.003
MSA		22.391	27.989	0.367	18.659	27.989	0.016	27.989	27.989	0.060
MSJ		3.814	4.767	0.161	3.178	4.767	0.027	4.767	4.767	0.024
TCFI (MSVE + MSDE + MSA + MSJ)		26.210	32.761	0.532	21.842	32.761	0.049	32.761	32.761	0.088

Occupant safety performance measure can be elaborated as the maximum safer distance that needs to be maintained between the vehicles. The value of maximum penetrating distance (Max PD, Table 3) should be as low as possible. In the same time the spacing should be low (Max Spacing) so that "Cut ins" shall be avoided. The comfort performance measure can be validated by measuring the acceleration, deceleration and jerk values measured. These values must be as small as possible. While following a lead vehicle, the values of Mean Square Distance Error (MSDE) and Mean Square Velocity Error (MSVE) are expected to be minimum. The overall performance of the controller is computed by finding the Total Cost Function Index (TCFI) value which is the addition of MSDE and MSVE along with Mean Square Acceleration (MSA) and Mean Square Jerk (MSJ).

Test result show that for Modem.urban8 driving pattern 97% improvement on TCFI was recorded along with 94% reduction in Max Acceleration when the Proposed HBM is compared with the counterparts. There was not much increase of spacing found. In case of IVD pattern 1, TCFI has been improved by 99% over CTG and VTG spacing policies. It is the same case with IVD Pattern 2. Maximum spacing between the two vehicles are reduced to 1.56 and 1 metres from 1.78 m for both IVD patterns (12% and 43% reduction in spacing without violating the spacing policy at a particular instant). This would increase the lane capacity. In common the acceleration (Max Acceleration, MSA) is found to be higher for CTG and VTG.

5 Conclusion

A new spacing policy is proposed based on the human behaviour while approaching a lead vehicle moving in different speeds is presented. Controller for FCWS is tested for different spacing policies. Since FCAS uses a controller designed using human reasoning and also a spacing policy designed after the human behaviour, the controller seems to behave more efficiently than the other spacing policies, with respect to the fuzzy controller. The proposed spacing policy gave an improvement above 97% (TCFI) for all the test conditions used. Capacity of the lane is also increased while avoiding rear-end collision.

References

1. M. Ruikar, "National statistics of road traffic accidents in India," *J. Orthop. Traumatol. Rehabil.*, vol. 6, no. 1, pp. 1–6, 2013.
2. S. K. Gehrig and F. J. Stein, "Collision avoidance for vehicle-following systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 233–244, 2007.
3. J. Wang and R. Rajamani, "Should adaptive cruise-control systems be designed to maintain a constant time gap between vehicles?," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1480–1490, 2004.
4. "3. Overall Evaluation of the Constant Time-Gap Spacing Policy," no. 4, pp. 41–42.

5. K. Santhanakrishnan and R. Rajamani, "On spacing policies for highway vehicle automation," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 198–204, 2003.
6. D. Yanakiev and L. Kanellakopoulos, "Nonlinear spacing policies for automated heavy-duty vehicles," *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1365–1377, 1998.
7. D. Yanakiev and I. Kanellakopoulos, "Variable time headway for string stability of automated heavy-dutyvehicles," *Proc. 1995 34th IEEE Conf. Decis. Control*, vol. 4, no. December, 1995.
8. T. Vtg, H. Capacity, A. T. The, C. Of, and T. Vtg, "4. Addressing the Trade-Off Between Safety and Traffic Flow," pp. 43–57.
9. J. Trnros, L. Nilsson, J. Ostlund, and A. Kircher, "Effects of ACC on Driver Behaviour, Workload AND Acceptance in Relation to Minimum Time Headway," in *9th World Congress on Intelligent Transport Systems*, 2002, vol. 1.
10. J.-J. Martinez and C. Canudas-de-Wit, "A Safe Longitudinal Control for Adaptive Cruise Control and Stop-and-Go Scenarios," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 2, pp. 246–258, 2007.
11. N. Benalie, W. Pananurak, S. Thanok, and M. Parnichkun, "Improvement of adaptive cruise control system based on speed characteristics and time headway," *2009 IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS 2009*, pp. 2403–2408, 2009.
12. C. C. Tsai, S. M. Hsieh, and C. T. Chen, "Fuzzy longitudinal controller design and experimentation for adaptive cruise control and stop&Go," *J. Intell. Robot. Syst. Theory Appl.*, vol. 59, no. 2, pp. 167–189, 2010.
13. H. Holzmann, C. Halfmann, S. Germann, M. Würtenberger, and R. Isermann, "Longitudinal and lateral control and supervision of autonomous intelligent vehicles," *Control Eng. Pract.*, vol. 5, no. 11, pp. 1599–1605, 1997.
14. D. Zhao, Z. Hu, Z. Xia, C. Alippi, Y. Zhu, and D. Wang, "Full-range adaptive cruise control based on supervised adaptive dynamic programming," *Neurocomputing*, vol. 125, pp. 57–67, 2014.
15. H. a. Hamersma and P. S. Els, "Longitudinal vehicle dynamics control for improved vehicle safety," *J. Terramechanics*, vol. 54, pp. 19–36, 2014.
16. M. Padash, M. Tavassoli, A. Khoei, and K. Hadidi, "A Sophisticated Algorithm for Using Fuzzy Logic Controllers in Adaptive Cruise Controller Systems," vol. 1, no. 4, pp. 134–141, 2013.

Comparative Analysis of New Hybrid Approach for Overlay Construction in P2P Live Streaming

Kunwar Pal, M.C. Govil and Mushtaq Ahmed

Abstract There has been a surging increase in the demand for media streaming over the internet during the last few decades. Live streaming comprises of a majority portion of streaming video over the Internet. Peer-to-peer network plays a very significant role by controlling the overhead at the server side. Also, the peer-to-peer network is easily scalable due to which it encourages file sharing and live video streaming. Overlay construction comprises a major part of peer-to-peer communication. Different overlay approaches like tree and mesh have already been discussed. In this paper, we perform a comparative analysis of the New Hybrid Approach for Overlay Construction in P2P Live Streaming. This approach is a combination of the tree and mesh topology. The simulation results presented in this paper verifies that our approach is better than the DenaCast approach in respect of end to end delay, start-up delay, and frame loss.

Keywords P2P · Peer to peer network · Live video streaming
Resource utilization · Content delivery network · ISP traffic

1 Introduction

According to Cisco, 90% of congestion over the Internet would be due to media transmission till the end of 2019 [N]. There has been an enormous increase in the demand for video by Internet users. It may be in the field of education for educational purposes (e.g., online lectures), for entertainment (e.g., to watch sports

K. Pal (✉) · M.C. Govil · M. Ahmed
Department of Computer Science and Engineering, Malaviya National Institute of Technology, Jaipur, India
e-mail: 2013rcp9568@mnit.ac.in

M.C. Govil
e-mail: govilmc@gmail.com

M. Ahmed
e-mail: mahmed.cse@mnit.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_22

events), or for communication (e.g., video chat) [1]. Media streaming can be classified in two ways: Video on Demand and live video streaming. In the video on demand user can enjoy the video at any time no real-time constraints are applied to it. While in live video streaming user can watch the video only at a particular broadcast time [2]. The video is displayed at a particular broadcast time in live video streaming. Due to the rapid increase in demand for video streaming the cost and complexity of the traditional client-server architecture at server side is increasing. In the client-server architecture, the server is the only source to upload the media content [3].

Peer-to-peer network is an efficient approach to handling the issue of the rising cost and complexity on the server side. Peer-to-peer network is a distributed network in which the task of uploading the content is jointly handled by the server and the end device instead of just being dealt with by the server alone. Peer-to-peer network is easily scalable, and the peer can use their upload bandwidth so that there is more utilization of the peer bandwidth. In this way, the requirement for a new user can easily be fulfilled without any extra load on the server. CoolStreaming is the landmark implementation of live video streaming in peer-to-peer network. Author [4] has discussed the different implementation issues related to DoNet/CoolStreaming in large scale and the various design issues. CoolStreaming uses a swarm-based approach for content delivery and the gossiping protocol for overlay construction. There have been many improvements done in this approach, and the same have been studied. Author [5] has found that ‘churn’ is a crucial factor of media streaming from recently generated video events.

The researchers have also discovered that different distribution of resources also has a major effect on live video streaming. The basic idea about CoolStreaming, its design and implementation have been described by Susu Xie [6]. The main focus of this work is on providing a non-optimal real working system. The peer selection procedure that has been used is randomized and has been demonstrated by presenting some real traces. Buffering techniques are used for resolving the problems of heterogeneity and dynamics. The hybrid push-pull scheme is used for scheduling purpose. Peer-to-peer network is a very challenging approach, and some of the major issues related to it are discussed by B.L.B.Li [7]. The performance of the network is affected by different factors like inherent stability, autonomous nature of the Internet and no service guarantee. Heterogeneity, the uploading capacity of end devices, firewall, and NAT (Network Address Translation) also has a significant impact on the performance of the network. B.Li [8] has presented some other challenges like cross-ISP traffic that are of major concern and has also discussed a new ISP-friendly approach to resolving the problem of cross-ISP traffic.

Peer-to-peer network comprises of different components: Overlay Construction, Scheduling scheme, and Selfish peer. In this paper, we would discuss all the issues regarding the overlay construction. Our prime concern is that how the end devices are connected to each other? The peer should be connected in such a way that they can get the appropriate data without missing the frame deadline.

2 Overlay

Traditionally, the overlay construction can be further classified in two ways: Tree overlay and Mesh overlay.

1. *Tree Overlay*

Transfer of data between the peers is the responsibility of the network layer. The basic approach used for the transmission of the media through the network is IP Multicast. IP multicast is a very famous, traditional, and efficient approach used for the transmission of data to more than one user at an instance of time. However, this approach has some issues like congestion, scalability, network address translation, error and flow control that affect the performance of the network [9]. Application layer multicast is a recently emerging solution for efficient multicasting. Application layer multicast is an approach that can solve the issue of IP multicast. The end device can further forward the data to other devices in an application layer multicast whereas in an IP multicast the end device can only see the video, no forwarding can be done by the end user. The overcast network is scalable, and it can handle the congestion and system failure issues.

For multicasting tree structure is probably the simplest overlay. Tree architecture can easily be maintained and is less complex in comparison to the other different architectures because a parent–child relationship is established between the peers in the tree architecture. Due to this parent–child relationship between the peers; the overlay is loop free, replication of data is less, and the overlay is easy to maintain. The video server is available at the highest level (level 0) and further the peers are distributed at different levels.

The nodes at the top level work as a parent for the peer that is at the lower level and the peer at lower level works as a child. The peers that are at the leaf work only as a child, and they cannot upload the content to the other peers. In a tree or a multi-tree, the video contents are pushed along the distinct path from parent to child. The Push scheduling scheme is used in tree overlay for media transmission. NICE and ESM are some of the examples of tree-based overlay [1].

Tree overlay is a delicate structure. In this structure, the child peer is fully dependent on the parent peer for video transmission as the parent peer transfers the video content to all available children in its child table, so the upload bandwidth of the parent peer is distributed equally among all its child peers. If the link between a parent and child breaks due to some error in the network, then the child peer would suffer, and it would lose the video frames. Also, in case the parent peer leaves the network, all the children peers in the entire structure would lose their continuity. Not only the immediate children of the parent peer but also their child's child, they would also lose the video frames. The child peer being dynamic does not have much effect on the network, but if the parent peer is dynamic; then it would affect the performance of the overall network. The complexity of the tree structure increases if more parents are dynamic and hence the cost of maintaining the parent–

child relationship between the peers would increase. So the frequency of the peers leaving the network is a major parameter in tree overlay.

2. *Mesh Overlay*

Mesh overlay is another solution for overlay construction problem. Every peer in the mesh overlay is connected to every other peer available in the overlay. Push based scheduling scheme is not used in mesh overlay because here many to many links are available between the peers. Hence, it is possible that the peer can get same data from different peers. So, replication of data is an issue in push scheduling scheme in the mesh overlay. The dynamic nature of the peer does not have much effect on mesh overlay. If some peer is leaving the network, then the affected peer can find the video from the other connected peers because the same video is available with many other peers and in the given overlay a peer can connect to multiple peers. Thus, in this case, the performance of the whole network would not be affected. Peers who are lower in the hierarchy can also upload the video content to other connected peers. Bandwidth utilization of peer in mesh overlay is more as compared to that in tree overlay [10]. Chainsaw, Bullet, CoolStreaming, Anysee are some of the examples of mesh overlay [11]. There is a frequent exchange of notifications in the mesh overlay model, so the major disadvantage of this model is that it suffers from efficiency latency tradeoff.

Content bottleneck and bandwidth bottleneck are the main issues in live video streaming in peer-to-peer network. Mesh-based peer-to-peer live video streaming is described in PRIME [12]. The approach uses swarming content delivery scheme for data transmission between peers. Authors provide the solution to both the problems discussed above. Efficient pattern delivery scheme is used for solving the issue of the content bottleneck. Also, for solving the problem of bandwidth bottleneck the author use the bandwidth degree condition scheme [13]. Data in mesh overlay has to be pulled from the other peers, so generally pull-based scheduling scheme is applied in mesh overlay as this scheme is simple and robust. Pull-based approach provides a near to optimal solution to the scheduling problem in terms of throughput and bandwidth utilization; without any bandwidth measurement and intelligent scheduling scheme. It uses specifically defined parameter settings and considerable raw streaming where server bandwidth is more than the raw streaming rate which is a mandatory condition for uninterrupted live video streaming. This approach also follows some assumptions like the server bandwidth should be at least three times of the raw streaming rate and the peers which form a group in the network should be less than 10,000 [14]. To remove the problems of overhead and delay the author provides a new push-pull hybrid approach. This approach provides an optimal solution with low overhead and minimum delay as verified by the simulation results.

TURINstream a novel peer-to-peer video streaming approach is designed to solve the issue of quality of service (QoS) [15]. In this approach; the peers are organized in a cluster, and separate overlays are created for control and media packets. A distributed algorithm is used for creating the clusters, which fully utilize the upload capacity of the participating nodes. TURINstream prototype is developed and simulated on PlanetLab. For testing this approach, a scenario is created using flash crowd, dynamic nature of peer and limited upload bandwidth.

3. *Hybrid Overlay*

Usually, overlay creation is categorized into, tree overlay and mesh overlay. Tree overlay is easy to handle, have low control overhead and also have less transmission delay as compared to the mesh overlay. Mesh overlay has more bandwidth utilization, but it is more reliable as compared to tree overlay. However, some authors have combined both the overlays and have provided a hybrid overlay structure for peer-to-peer network. This approach combines the property of both the approaches and provides better overlay construction design. The Author uses two overlay one for control tree and the second for data mesh. Control tree is used for transferring the control packets between the peers while the actual media packets are transferred through the data mesh. For creating control tree and data mesh, geographical location of the peers and layered peer selection mechanism are used respectively. The efficiency of the peer should be checked periodically; and if the peer is not efficient, remove it from the overlay [16].

mTreebone is an approach in which a tree is created using stable peers; stable peers are those that stay for more time in the network or that are less dynamic in nature. The backbone of the mTreebone is a tree created using the stable peers. The stability of a peer is calculated using an optimal age threshold value. If a peer has more age than the threshold then it is considered as a stable peer; otherwise, it is an unstable peer. For creating the mesh between peers, a stable peer is further connected with unstable peers. For data transmission push-pull scheme is used. With the help of partitioning and scheduling scheme, this approach provides minimum control overhead and transmission delay [17].

Another approach HyPO is described by Haesun Byun [11]. First a tree overlay is created in HyPO approach using the similar bandwidth of a peer in a particular geographical place and then mesh overlay is created further. The depth of the tree is optimized using an even distribution of peers in the network. In comparison with the other approaches, the transmission time is minimized; because the depth of the tree is minimized. This approach also reduces the control overhead. The main problem with this approach is that there is no certain criterion for the tree and mesh peers; also, the upload bandwidth of the stable peers is not fully utilized.

The two layer architecture is used for overlay in the ToMo approach. A mesh is created at the first layer while a tree is created at the second layer. When the number of peers is less, all the peers are in the mesh overlay. If more peers are added to the overlay, then the tree overlay is created in such a manner that the tree layer is one less than the mesh layer. Push-based approach is used for data transmission by the author to minimize the latency. The source peer maintains the depth of the structure. Each peer periodically sends the control information to source peer, due to which complexity at the source peer is very high [1].

ISP (Internet Service Provider) is also a crucial factor in peer-to-peer network that is discussed by Huey-Ing Liu. For creating the overlay, the author has considered both the peer as well as the ISP. First of all the subnets are created using ISP, and then these subnets are further distributed in the form of a cluster. The bandwidth is used as a parameter for creating the cluster and peers of the same bandwidth lie in the same cluster. Tree overlay is formed between the source peer of each cluster, and mesh overlay is created between the peers of the same cluster. The connections between different clusters are only maintained through the tree overlay. Inter-cluster communication is prohibited for clusters at different layers, so there is a limited utilization of upload bandwidth of the peers [18].

Thinh Nguyen Kim combined CDN (Content Distributed Network) and peer-to-peer network, and provided a group-based CDN-P2P hybrid architecture (GCP2P) [19]. Leveraging the property of both the approaches, GCP2P provides reliability, scalability; interrupt latency, and less control overhead. The peer that is physically near to CDN server is selected as a super peer. Peer selection in GCP2P is content/location based, and the super peer is selected on the basis of location. On the basis of area and peer's channel preferences, physical groups are formed and named as sub-overlay. Through simulation results, it has been proved that the hybrid approach provides less start-up delay as compared to CDN and P2P individually.

From the above discussion, it is apparent that there is problem of maximum bandwidth utilization in all approaches. To solve this problem of upload bandwidth utilization the author has described a Hybrid Live P2P Streaming Protocol (HLPSP) [10]. HLPSP is a novel approach which is created by the combination of tree and mesh overlay. Groups are created using upload bandwidth of the peers. The peer who has higher upload bandwidth is higher in the hierarchy and would work as a source for peers who are lower in the hierarchy. The source peer is at level 0 with highest upload bandwidth. A comparison of this approach is done with the Denacast approach (an enhanced version of cool streaming). The simulation results show that HLPSP provides less start-up delay, less control overhead, and better data flexibility. The author considers the uploading bandwidth as a prime factor for overlay construction, but other factors like geographical locations, dynamic nature of peer are not considered. These factors also affect the performance of the network.

3 New Hybrid Approach

The new hybrid approach provides a solution to the issues that have been discussed above. In our previous work [N] we have defined a new hybrid approach. For creating an overlay different parameters like upload bandwidth, age and geo-location are used; and peers are categorized according to ISP. An overlay is created at various levels and for calculating the appropriate level remaining upload bandwidth of the peer is used. Peers are distributed in two ways: stable node and unstable node. The stability of a peer is calculated using age and upload bandwidth capacity. Here an optimal threshold value 'th2' is chosen for stability and for new peer 'th2' it is defined as the average age of available peers in the network, which is the duration of its last session for an old peer. As stable nodes can greatly boost the streaming quality and also effectively reduce the control overhead they are always preferred to create a tree. The nodes which are chosen for mesh are those whose peers have maximum age or maximum upload bandwidth. In this approach, the mesh is not fully mesh therefore it overcomes the disadvantages of full mesh approach, and only the stable nodes are chosen as parents which helps in increasing the reliability of this architecture, also higher bandwidth peer near to the source helps in less start-up delay. So this approach has advantages of both trees as well as mesh architecture.

When a new peer *P* enters the network, it sends a request to the tracer. The format of the request tuple is < Upload Bandwidth, Geo Location, Content, Age >. From content id, the tracer finds out that which video the peer wants to watch, and currently how many peers are watching this video. Then using the Geo-location parameter the tracer discovers the nearest server for the client peer. The tracer stores this peer in the tracer table at appropriate *level[P]*. Tracer finds-out *list[P]* of recommended parent peers for the client P. For calculating the *list[P]* tracer uses algorithm 4[N]. Only those peers are added to the *list[P]* which are stable and have sufficient upload bandwidth at the same level or at lower level in the hierarchy. If *list[P]* is still not full, then unstable peer is also added to the *list[P]*. Tracer sends this *list[P]* to peer P. Peer P receives the list and sends a connection request to the best available peer. If a connection is established, then the data transfer takes place according to the scheduling scheme. Otherwise, the peer sends the connection request to other peers available in the *list[P]*. Parameter (Max_P_N, P_N, th1, th2, level[pk], th3, age[pi], list, Bu[pk]) have the same meaning as described in our previous work [20].

Overlay construction in new hybrid overlay approach is a combination of tree and mesh overlay; it provides the benefits of both the approaches. Tree overlay is created using stable peers only, and stability is calculated according to the age of the peer, so this approach is more reliable. Tree overlay is the backbone of this approach, and high upload bandwidth peers lie higher in the hierarchy. So, when a new peer is added to the network, there is no need to establish the relationship between the peers like in mesh. There is a predefined parent- child relationship between the peers, so the start-up delay is minimized. The congestion over the network is minimized because the overlay is not entirely mesh. Bandwidth utilization of stable peer is more, and peers who are lower in the hierarchy can also use their upload bandwidth, so the overall bandwidth utilization of network is increased.

In this paper, we simulate the approach and provide the results that support our approach. Different parameters like start-up delay, end to end delay, playback delay, frame loss ratio and packet drop due to destination unreachable are used to verify the result.

4 Simulation and Result

1. *Simulation Setup*

For implementing New Hybrid Overlay approach, we use Oversim simulator. Oversim is a peer-to-peer network simulator and open source overlay for OMNET ++ simulation environment [OS]. Discrete event simulation (DES) is used for processing and exchanging the network message in OverSim [21]. Oversim is distributed into modules. NED (Network Description) is a topology description language which is used for defining the module; while C ++ language is used for doing processing between the modules. Different parameters, which are used for simulation, are given in Table 1.

2. *Simulation Results*

Start-up delay is considered a crucial factor for live video streaming. The new hybrid approach has a lower start-up delay as compared to the Denacast approach. Due to the stable structure of the overlay (parent–child relationship), the time required for new peer to receive the desired frame is minimized so, start-up delay is also minimized. Initially, when the number of peers in the network is less, then both approaches behave approximately the same. The start-up delay in new hybrid approach gradually increases as compared to Denacast with an increase in the number of peers in the overlay. The stable structure plays a crucial role in the overlay. If peers are unstable, it becomes difficult to maintain a parent–child relationship as frequently finding parents directly affects the start-up delay of peer (Fig. 1).

As shown in the result Figure [N], as the number of peers in the network increase the start-up delay in the new hybrid approach also increases, due to the stable structure. In Denacast new peer finds out the path up to the server, which takes too much time and start-up delay increases (Fig. 2).

Overlay construction also affects the end-to-end delay between the peers. If the relationship between the peers is predefined and the peers are stable, then the end to

Table 1 Simulation parameters and values

S. No	Parameter	Value
1.	Simulation duration	200 s
2.	Average video bit rate	512 Kbps
3.	Video Codec	MPEG4 Part I
4.	Buffer map exchange period	1 s
5.	Chunk size	5 Frames
6.	Source number	3
7.	Number of runs	10
8.	Maximum number of levels	6
9.	Average chunk length	130 Kb
10.	Neighbour notification period	2 s

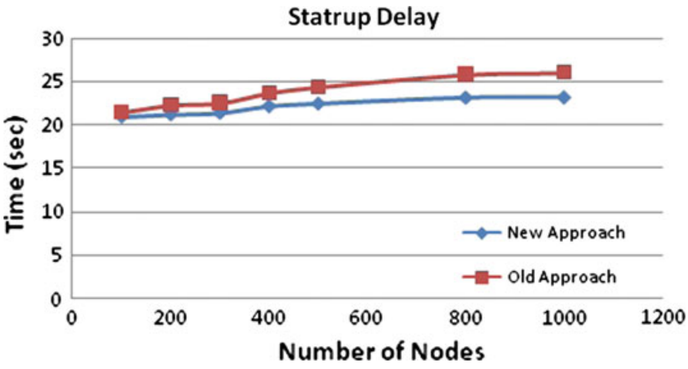


Fig. 1 Average start-up delay

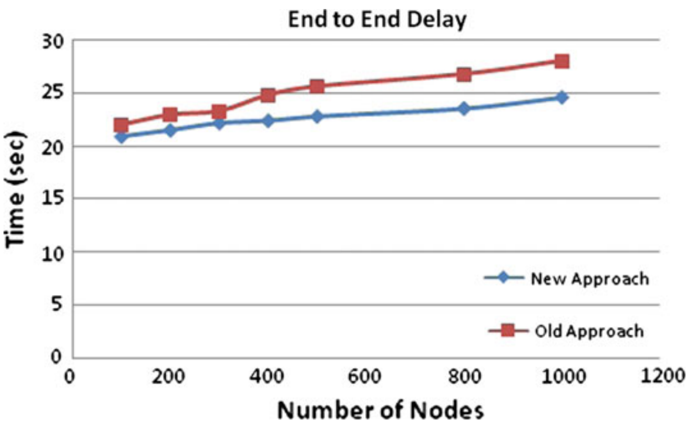


Fig. 2 Average end to end delay

end delay between the peers is minimized. The peers who are more stable and have more upload bandwidth are above in hierarchy while constructing overlay in the new hybrid approach. This affects the overall end to end delay between the peers. The figure [N] represents the end to end delay between peers in both the approaches. The number of peers in the network is less hence difference between both the approaches is minimized, but it increases as the number of the peers in overlay increases.

The Frame loss ratio in Denacast approach is more as compared to the new hybrid approach. With a rise in the number of peers, there is also an increase in the number of frames. In new hybrid overlay approach, peers (especially parent's peers) are stable so frame loss is minimized and frame loss ratio also decreases; while in the Denacast approach it increases with an increase in the number of peers (Fig. 3).

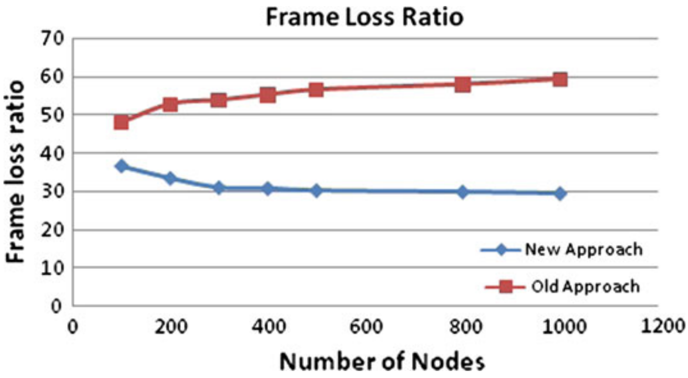


Fig. 3 Average frame loss ratio

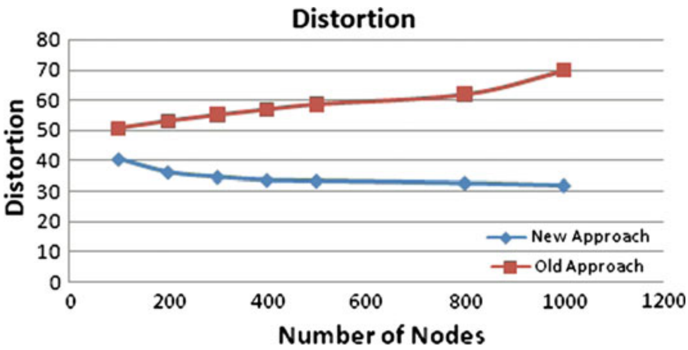


Fig. 4 Average distortion

With an increase in the number of peers, distortion of video increases in the Denacast approach; while it decreases in the hybrid overlay approach. But in a peer-to-peer network, it increases with an increase in the frame loss ratio. The dynamic nature of peers affects the video quality; if parent peers are dynamic it will affect the video streaming at the child peer. The more a dynamic peer is higher in the structure; the more would be an effect on its dependent children. In the new hybrid overlay approach, peers that are above in the hierarchy are stable, so there is a smooth continuity of video at the child peer. Hence, distortion in the new hybrid overlay approach decreases whereas; it increases in the Denacast approach. Figure [N], also shows that as the number of peers in the overlay increases, due to stable peer; distortion of video increases in the Denacast approach whereas it decreases in the new hybrid approach (Figs. 4 and 5).

Playback delay of hybrid overlay approach is more in comparison to Denacast approach. However, it approximately remains the same with an increase in the number of peers. So the overall performance of the network increases if we follow the new hybrid overlay approach.

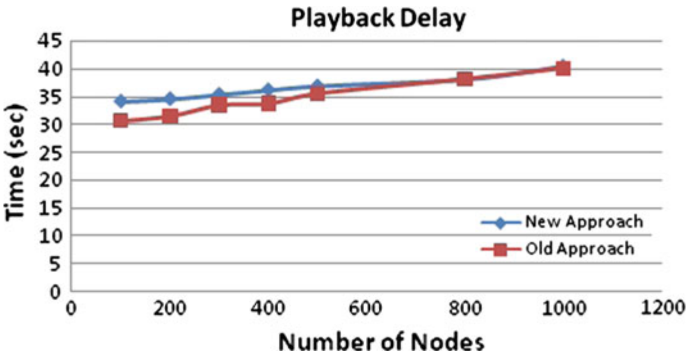


Fig. 5 Average play back delay

5 Conclusion

In this paper, we discussed different approaches to overlay construction in live video streaming for peer-to-peer network.

Overlay construction in peer-to-peer network is a challenging issue. We provide a new overlay design approach (New hybrid overlay approach) which is a combination of both tree and mesh overlay. This approach inherits properties of both overlay designs. This new approach is easily scalable and more reliable. We have described a simulation of this approach in this paper. A comparative analysis of new hybrid overlay approach with Denacast shows that parameters like the end to end delay, start-up delay; frame loss ratio, total distortion, and playback delay support this new approach. The performance of both approaches is approximately same for less number of peers but when the number of peer increases network performance of new hybrid overlay construction approach is better as compared to Denacast approach. These parameters define the performance of the network. Hence, in a peer to peer network the overall performance of a network in live video streaming increases.

References

1. S. Awiphan, Z. Su, and J. Katto, “ToMo: A two-layer mesh/tree structure for live streaming in P2P overlay network,” 2010 7th IEEE Consum. Commun. Netw. Conf. CCNC 2010, 2010.
2. J. Jannotti, D. K. Gifford, K. L. Johnson, M. F. Kaashoek, and J. W. O’Toole Jr., “Overcast: reliable multicasting with on overlay network,” OSDI’00 Proc. 4th Conf. Symp. Oper. Syst. Des. Implement., p. 14, 2000.
3. V. Venkataraman, P. Francis, and J. Calandrino, “Chunkyspread: Multi-tree Unstructured Peer-to-Peer Multicast,” Proc. 14th IEEE Int. Conf. Netw. Protoc. – ICNP’06, pp. 2–11, 2006.

4. X. Zhang, J. Liu, B. Li, and T. S. P. Yum, “CoolStreaming/DONet: A data-driven overlay network for efficient live media streaming,” Proc. IEEE Infocom, vol. 3, no. C, pp. 13–17, 2005.
5. B. Li, S. S. Xie, G. Y. Keung, J. C. Liu, I. Stoica, H. Zhang, and X. Y. Zhang, “An empirical study of the coolstreaming plus system,” Ieee J. Sel. Areas Commun., vol. 25, no. 9, pp. 1627–1639, 2007.
6. S. Xie, B. Li, S. Member, G. Y. Keung, X. Zhang, and S. Member, “Coolstreaming : Design, Theory,and Practice,” vol. 9, no. 8, pp. 1661–1671, 2007.
7. B. L. B. Li and H. Y. H. Yin, “Peer-to-peer live video streaming on the internet: issues, existing approaches, and challenges [Peer-to-Peer Multimedia Streaming],” IEEE Commun. Mag., vol. 45, no. June, pp. 94–99, 2007.
8. B. Li, S. Xie, Y. Qu, G. Y. Keung, C. Lin, J. Liu, and X. Zhang, “Inside the new Coolstreaming: Principles, measurements and performance implications,” Proc. – IEEE INFOCOM, pp. 1705–1713, 2008.
9. Y. H. Chu, S. G. Rao, S. Seshan, and H. Zhang, “A case for end system multicast,” IEEE J. Sel. Areas Commun., vol. 20, pp. 1456–1471, 2002.
10. C. Hammami, I. Jemili, A. Gazdar, A. Belghith, and M. Mosbah, “Hybrid live P2P streaming protocol,” Procedia Comput. Sci., vol. 32, pp. 158–165, 2014.
11. H. B. H. Byun and M. L. M. Lee, “HyPO: A Peer-to-Peer based hybrid overlay structure,” 2009 11th Int. Conf. Adv. Commun. Technol., vol. 01, pp. 840–844, 2009.
12. N. Magharei and R. Rejaie, “Understanding Mesh-based Peer-to-Peer Streaming,” 2006.
13. N. Magharei and R. Rejaie, “PRIME : Peer-to-Peer Receiver-drIven MEsh-based Streaming,” pp. 1415–1423, 2007.
14. M. Zhang and Q. Zhang, “Understanding the Power of Pull- based Streaming Protocol : Can We Do Better ? Presented by Rabin Karki Background,” vol. 25, no. 9, pp. 1678–1694, 2007.
15. A. Magnetto, R. Gaeta, M. Grangetto, and M. Sereno, “TURINstream: A totally push, robust, and efficient P2P video streaming architecture,” IEEE Trans. Multimed., vol. 12, no. 8, pp. 901–914, 2010.
16. Q. Huang, H. Jin, and X. Liao, “P2P live streaming with tree-mesh based hybrid overlay,” Proc. Int. Conf. Parallel Process. Work., no. 60433040, 2007.
17. B. C. Canada, “mTreebone : A Hybrid Tree/Mesh Overlay for Application-Layer Live Video,” Distrib. Comput., 2007.
18. H. I. Liu and I. F. Wu, “MeTree: A contribution and locality-aware P2P live streaming architecture,” Proc. – Int. Conf. Adv. Inf. Netw. Appl. AINA, pp. 1136–1143, 2010.
19. T. N. Kim, S. Jeon, and Y. Kim, “A CDN-P2P hybrid architecture with content/location awareness for live streaming service networks,” Proc. Int. Symp. Consum. Electron. ISCE, pp. 438–441, 2011.
20. P. Kunwar. Govil M.C, Ahmed Mushtaq, “A New Hybrid Approach for Overlay Construction in P2P Live Streaming,” in Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference, Kochi, INDIA, 2015, pp. 431–437.
21. I. Baumgart, B. Heep, and S. Krause, “OverSim: A Flexible Overlay Network Simulation Framework,” 2007 IEEE Glob. Internet Symp., pp. 79–84, 2007.

Ant-AODV-VANET: A Bio-Inspired Routing Protocol for VANET

Atreyee Datta, Chandreyee Chowdhury and Sarmistha Neogy

Abstract VehicularAdhoc Network (VANET) is one of the interesting and challenging research areas in mobile communication domain. For the past few decades, researchers are trying to develop appropriate routing protocols for VANET. Among the different types of bio-inspired algorithms, Ant Colony Optimization (ACO) is found to be most recent and efficient technique in the context of VANET. Due to numerous use of VANET in transportation systems, it has emerged as a serious issue for the researchers to upgrade existing algorithms for accepting the challenges of VANET. Moreover VANET is a promising approach for future intelligent transportation system (ITS). In this paper, we propose a new routing protocol named Ant-AODV-VANET, which incorporates movement information of vehicles into the route discovery process. A suitable parameter, referred to as optimum value (Op_{value}) is found out. Finally, we evaluate the accuracy of Op_{value} using the proposed routing protocol.

Keywords VANET · Optimization · ACO · Accuracy · AODV

1 Introduction

Wireless mobile communication is becoming very popular due to the flexibility that makes it suitable for different applications. Vehicular Adhoc Network (VANET) [1, 2, 3] is an application field of Mobile AdhocNetwork (MANET) [4, 5]. VANET is meant to be an autonomous, self-configurable, decentralized network for sharing and exchanging information. It provides different types of facilities, including traffic

A. Datta · C. Chowdhury · S. Neogy (✉)
Jadavpur University, Kolkata, India
e-mail: sarmisthaneogy@yahoo.com

A. Datta
e-mail: atreyee_chau@yahoo.co.in

C. Chowdhury
e-mail: chandreyee.chowdhury@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_23

management, road safety, among others. Due to high mobility of vehicles, routing is difficult to be implemented in VANET. The movement of nodes (vehicle) in VANET is restricted by road course, encompassing traffic, traffic regulation, and also driver of the vehicle. In this scenario, there are two types of communications, one is Vehicle to Vehicle (V2V) [6] and other is Vehicle to Infrastructure (V2I) [6].

In many engineering fields, several problems can be solved by the concept of Ant Colony Optimization (ACO) [7] process. Real ants are social insects and their movement is natural. The main objective of our present work is to evaluate the optimum path between a source and destination vehicle, using the concept of pheromone concentration which is influenced by ACO. Here in VANET scenario, pheromone parameter is considered to be equivalent to Link Quality (LQ) of the adjacent vehicles. The value of Total Weight of the Route (TWR) [8] is calculated by utilizing the vehicles' movement information (e.g., position, direction, acceleration) and LQ value of two adjacent nodes.

Optimization is a methodology or a process of making something as fully stable, perfect or effective as possible. In computing, optimization is the method of modifying a system to make some characteristics of it work more efficiently using the fewer resources. Different types of optimization techniques are already implemented. Among these, we focus on bio-Inspired Optimization [9] techniques. Bio-inspired computing is a field of study that loosely associates together subfields related to the topics of emergence and social behavior. It is often closely related to the field of artificial intelligence (AI). It is the discipline that deals with natural and artificial systems. Swarm Intelligence (SI) [9] is a part of Bio-inspired computing.

1.1 Ant Colony Optimization (ACO)

Ant Colony Optimization [7] is a very renowned optimization technique which is inspired by the behavior of social insects (real ants) while they search for source of food. Ants communicate with each other through pheromone trails left in the path through which they pass. The first ants that leave the nest go out looking for food wandering randomly, and deposit pheromone along the path they travel. When an ant finds the food it returns to the nest depositing pheromone on the path. Other following ants occasionally will find the pheromone trail and they, in turn, tend to choose the route with the highest pheromone concentration [7].

In ACO, the visibility of pheromone quantity $\eta_{i,j} = \frac{1}{d_{i,j}}$ and transition probability [7] from town i to town j is given by $P_{i,j}$:

$$P_{i,j} = \frac{\tau_{i,j}^{\alpha} \eta_{i,j}^{\beta}}{\sum_{j=1}^{n} \tau_{i,j}^{\alpha} \eta_{i,j}^{\beta}} \tag{1}$$

Here, n is the total number of towns, $d_{i,j}$ is the Euclidean Distance [7] between two towns. α is a parameter to control the influence of $\tau_{i,j}$ and β is a parameter to control the influence of η_{ij} .The rest of this paper is organized as follows: In Sect. 2, we discuss related works about routing protocols. The main idea of the Ant-AODV-VANET routing protocol will be showed in Sect. 3. This section describes the working procedure of the proposed routing protocol. Section 4 presents the analysis of the accuracy of measurement of the parameters considered for the proposed protocol. Finally, we draw the conclusion in Sect. 5.

2 Related Work

Routing is the process of choosing the best path from source to destination in the network. For high mobility of nodes of VANET, existing MANET routing protocols are not compatible with VANET. Routing protocols of VANET are broadly divided into two types: *Table-Driven Protocols* and *Source Initiated on-Demand Protocols* [10]. Now we consider some routing protocols of VANET. These are

(a) *Vehicular-Assisted Data Delivery (VADD)*: The routing procedure of VADD [11] is based on the idea of carry and forward concept by using predictable vehicle mobility. In this protocol, large delay occurs due to change of topology and traffic density.
(b) *Geographical Opportunistic Routing (GeOpps)* [11]: It is based on the idea of utilizing the navigation system suggested routes of vehicles for selecting the forward node which is closer to the destination. Privacy is not considered for disclosing the navigation information to the network.
(c) *Greedy Traffic-Aware Routing Protocol (GyTAR)* [12]: It is based on the concept of intersection-based routing protocol which aims to decrease the end-to-end delay and control message overhead with low packet loss. It depends on road side units because it assumes that the number of cars in the road will be given from road side units.

Our proposed routing protocol is based on *Ad hoc on-Demand Distance Vector Routing(AODV)*. AODV is a traditional routing protocol of MANET.

Adhoc on-Demand Distance Vector (AODV) [13] routing is a simple, efficient, and effective routing protocol for MANETs which is a reactive type of protocol and finds a route to destination on demand. The advanced use of AODV is in VANET. Due to the reactive nature, AODV can handle the highly dynamic behavior of VANET.

2.1 AODV-VANET

In [8] the authors propose AODV-VANET, which incorporates the features of VANET into AODV during route discovery. These characteristics include position, speed, acceleration, direction of the vehicle and the link quality between two adjacent vehicles. From this routing protocol, we can evaluate *Total Weight of the Route (TWR)* for a particular route. The *Total Weight of the Route* is represented as

$$\sum_{i=1}^{n}[(f_s \times |S_i - S_{i+1}|) + (f_a \times |A_i - A_{i+1}|) + (f_d \times |D_i - D_{i+1}|) + \left(f_q \times \frac{1}{\mathrm{LQ}}\right)] \tag{2}$$

Where,

n	the number of nodes (Vehicles)
S	Speed of the Vehicles
f_s	Speed Weight Factor
A	Acceleration of Vehicles
f_a	Acceleration Weight Factor
D	Direction Metric of the Vehicles
f_d	Direction Weight Factor
f_q	Link Quality Weightage Factor
Link Quality (*LQ*)	Link Quality between two adjacent vehicles

However, it is not apparent from the work in [8], how TWR is finally evaluated, since each term in (2) represents different unit.

3 Our Approach: Ant-AODV-VANET

We propose a new reliable routing protocol for VANET which is influenced by bio-inspired concept. Using this protocol we try to evaluate the optimized path using the unique characteristic of VANET. We consider two main metrics of the VANET to be Hop Count (HC) and Link Quality (LQ). The aim is to choose a path with lesser number of hops and high link quality. Hence the parameter (Op_{value}) to choose the final path is directly proportional to Link Quality and reversely proportional to HopCount. In our proposed routing protocol we have considered a parameter called Route Weightage Factor (RWF), which denotes the visibility of pheromone [7] deposited on a route and is thus equivalent to visibility of route. It is obvious that a route will be visible if large amounts of pheromone are deposited on the route. Since it is likely that ants will take the shorter routes, RWF is considered to be reverse of the distance (d_{ij}) between any two nodes and defined as:

$$\mathrm{RWF}_{ij} = \frac{1}{d_{ij}} \tag{3}$$

Link Quality (LQ) [14] is defined to be the ability of the link to remain stable as long as possible, to have low bit errors and enable data to reach the destination with high signal strength. A path having higher link quality than others will be chosen. Hence the concept of link quality also relates to the amount of maximum pheromone deposition in a route. The greater the amount of pheromone, more likely it is the chosen path among other paths from source to destination. However, in wireless adhoc networks and that too in VANET, reception power indicates specifically the quality of a path. This is because better reception power implies lesser packet loss. Hence in VANET we consider link quality to be equivalent to the reception power of transmitted signal. Let the reception power of transmitted signal be denoted by S_R:

$$S_R = S_T \times G_R \times G_T \times \frac{\lambda}{4 \times \pi \times d^2} \tag{4}$$

Where, S_T is the transmission power of transmitted signal, G_R is the antenna gain of receiver, G_T is the antenna gain of transmitter and d_{ij} is the Euclidean Distance between two vehicles (V_i and V_j, say). d_{ij} is represented as

$$d_{ij} = \sqrt{\left(x_{1i} - x_{1j}\right)^2 + \left(x_{2i} - x_{2j}\right)^2} \tag{5}$$

So, we can evaluate the value of Op_{value} using the following relations:

$$Op_{\text{value}} \propto \text{LQ} \tag{6}$$

$$Op_{\text{value}} \propto \text{RWF} \tag{7}$$

$$Op_{\text{value}} \propto \frac{1}{\text{HC}} \tag{8}$$

From (3), (4) and (5), we can say,

$$Op_{\text{value}} = \text{CV}.\frac{\text{LQ} \times \text{RWF}}{\text{HC}} \tag{9}$$

where CV is a constant whose value lies between 0 and 1.

Table 1 shows the mapping between the relevant parameters in Euclidean TSP ACO process and our proposed algorithm.

Table 1 Mapping between the basic parameters of ACO and VANET

Terms of Ant Colony Optimization in Euclidean TSP problem [13]	Significance in our proposed routing protocol of VANET
Number of city	Number of vehicles
Number of ant	Number of transmitted packets
Quantity of pheromone	Link quality [here, it is S_R]
Visibility of quantity	Route weightage factor (RWF)

3.1 Working Principle of Proposed Protocol: Ant-AODV-VANET

In the above Fig. 1 we represent a Vehicular Adhoc Network using $V2V$ communication. At a instance, say t_1, V_2 is a source node and V_{15} is a destination node. All nodes are dynamic. All nodes have their own routing table. For t_1 instance, V_2 wants to send any message to the destination node V_{15}. There are multiple paths to transmit the packet in this network. Using our algorithm, we want to evaluate optimized path from source to destination using important parameter of VANET. Path discovery method is divided into two parts, these are Forward Path Discovery and Backward Path Discovery. There are different paths, these are $\{V_2$-V_5-V_6-V_9-V_{11}-$V_{15}\}$ or $\{V_2$-V_5-V_6-V_9-V_{11}-V_{13}-$V_{15}\}$ or $\{V_2$-V_4-V_6-V_9-V_{11}-$V_{15}\}$ or $\{V_2$-V_4-V_6-V_9-V_{12}-V_{11}-V_{13}-$V_{15}\}$ etc. We consider some important parameters those are 'distance between two node' which is reversely proportionate to receiving signal strength, that is Link Quality of two dynamic nodes; 'visibility of the road' which is inverse of the distance (d_{ij}); 'Hop Count'. We choose the path which contains maximum Link Quality, maximum Visibility, and minimum Hop Count. For above case, there are two paths between source to destination, those contain the minimum hop count. These are $\{V_2$-V_5-V_6-V_9-V_{11}-$V_{15}\}$ or $\{V_2$-V_4-V_6-V_9-V_{11}-$V_{15}\}$. Between two paths, we can choose more optimized path depends upon the Euclidean distance of connected nodes. Because the remaining two parameters (Link Quality, Visibility of the road) of our algorithm which are directly related to distance. So, using our algorithm the optimized path is $\{V_2$-V_4-V_6-V_9-V_{11}-$V_{15}\}$.

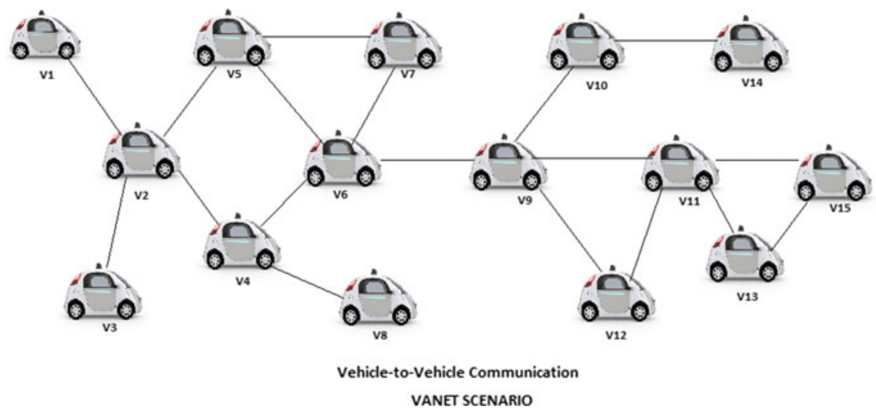


Fig. 1 VANET scenario using $V2V$ communication

3.2 Algorithm of Proposed Protocol: Ant-AODV-VANET

Algorithm 1: Forward path discovery process (at a node)

```
Input: packet P1 is received
    if (P1 is an RREQ packet)
    if (!D) then
            if (P1 is not previously received by the node)
                Send P1_Acknowledgement to the sender
                Evaluate Distance dij using Equation 5.
                Evaluate LinkQuality (SR) using Equation 4.
                HopCount = HopCount +1.
                if(D is not in NNi) then
                  Send P1 tox where xϵNNi
                        else
                            Send P1 to D
                        endif
                        Buffer P1 to particular neighbor_node.
                Evaluate Pheromone Value (Opvalue) using Eqn 9.
                Update own routing table in the current_node.
            else
              P1 is discarded.
            endif
     else
         Send RREP
         Send P1_Acknowledgement to the sender
     Endif
    elseif(P1 is Acknowledgement)
      Receive P1_Acknowledgement packet from neighbor.
    endif
```

Algorithm 2: Backward path discovery process (at a node)

```
Input: Packet P2 is received
if( P2 is a RREP packet)
      if (current_node_id != S)
            Set a neighbor_nodes (neighbor_node_id) list of current_node.
   if (neighbor_node_id != S) then search routing table
         if (more than two Opvalue are equal) then
               The lowest node_id is selected
         else
               The Highest Opvalue is selected
         endif
           Send P2 to that neighbor_node_id.
    endif
          Send P2 to S.
          Terminate the path discovery procedure.
   //Got the optimized path and terminate the process.
    else
     Terminate the path discovery procedure.
   endif.
  endif.
```

4 Results and Analysis

We have compared the routing protocols: AODV-VANET [8] and proposed Ant-AODV-VANET. We calculate the Op_{value} for both routing protocols AODV-VANET and Ant-AODV-VANET. We found out the percentage of accuracy of the measurement of Op_{value} of our proposed routing protocol. The following assumptions have been taken while considering the VANET:

(a) All Vehicles are equipped with GPS receiver, Digital Map, and sensors.
(b) No other fixed infrastructure for communication is present.
(c) Transmission range for every node is same.
(d) Average speed of the vehicles in this network is constant. So, relative position of each vehicle is approximately constant.
(e) the VANET is Highway Vehicular Adhoc Network.

We analyze the value Op_{value} of both scenarios (AODV-VANET and Ant-AODV-VANET) respect to λ of Eq. 4 and CV of Eq. 9. The National Telecommunication and Information Administration preliminary divides the radio

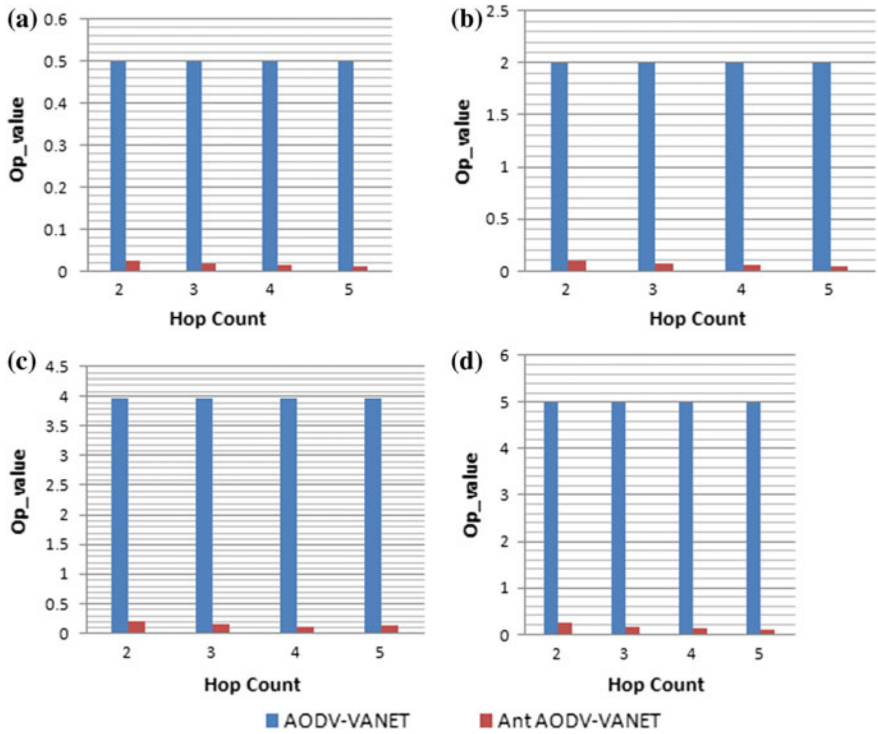


Fig. 2 **a** Op_{value} versus HopCount [λ = 10 m], **b** Op_{value} versus HopCount [λ = 40 m], **c** Op_{value} versus HopCount [λ =80 m], **d** Op_{value} versus HopCount [λ = 100 m]

spectrum in nine bands [15]. We observe the signal of average range of Wavelength for VANET environment.

In [8] the authors have mentioned about TWR. However, according to our assumptions, only the last term of TWR is relevant to the calculation of Op_{value} and TWR is then found to become inversely proportional to Op_{value}.

In Fig. 2a, b, c, and d we represent the variation in Op_{value} with change in Hop Count value. In case of AODV-VANET, changing value of Hop Count does not affect the value of Op_{value}. But, from Tables 2, 3, 4, and 5 we observe that the value of Op_{value} increases with increasing value of λ. We have considered the value of the Constant Value (CV) in both cases to be 0.5. In Ant-AODV-VANET, the Op_{value} decreases with the increasing value of Hop Count. Since Op_{value} considers all factors affecting a route selection (Eq. 9), hence the result is self-explanatory (Table 6). Thus, the trend in our proposed protocol signifies that a path with lesser hop counts will function better. It would be prudent to choose the minimum Op_{value} path. Our work is also supported by the evaluation of accuracy of path. Figure 3

Table 2 Op_{value}Versus HopCount (HC) [λ = 10m]

		HopCount			
		2	3	4	5
Op_{value}	AODV-VANET	0.5	0.5	0.5	0.5
	Ant-AODV-VANET	0.02487	0.0166	0.0124	0.00995

Table 3 Op_{value}Versus Hop Count (HC) [λ = 40m]

		HopCount			
		2	3	4	5
Op_{value}	AODV-VANET	2	2	2	2
	Ant-AODV-VANET	0.099425	0.0663	0.0497	0.03977

Table 4 Op_{value}Versus Hop Count (HC) [λ = 80 m]

		HopCount			
		2	3	4	5
Op_{value}	AODV-VANET	3.97	3.97	3.97	3.97
	Ant-AODV-VANET	0.19875	0.1325	0.099375	0.1243

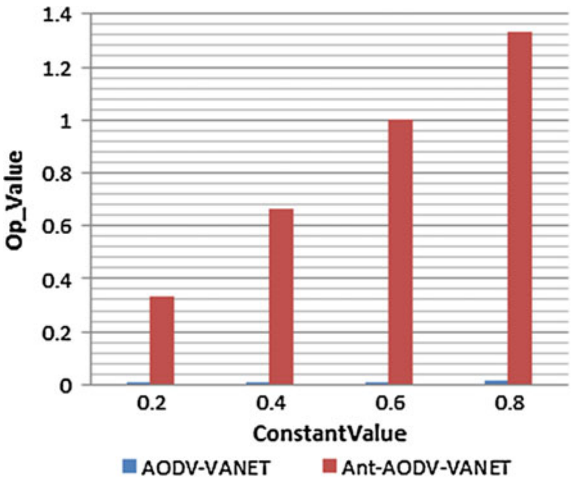
Table 5 Op_{value}Versus Hop Count (HC) [λ = 100m]

		HopCount			
		2	3	4	5
Op_{value}	AODV-VANET	5	5	5	5
	Ant-AODV-VANET	0.24875	0.166	0.1243	0.0995

Table 6 Op_{value} versus constant value (CV)

		ConstantValue			
		0.2	0.4	0.6	0.8
Op_{value}	AODV-VANET	0.004	0.008	0.012	0.016
	Ant-AODV-VANET	0.3333	0.6666	1	1.3333

Fig. 3 Op_{value} versus Constant Value



represents the variation in Op_{value} with change in CV ($0 < CV < 1$). We can evaluate 98.8% accuracy using the method [16] of measurement in the path that is evaluated by our proposed protocol with respect to only 16.8% (approx.) accuracy of the path evaluated by the work in [8]. Hence the routing protocol proposed in this paper is able to deliver more accurate and stable path.

5 Conclusion

In this paper, we propose a new routing protocol Ant-AODV-VANET that uses the concept of ACO in VANET environment. Though there exists a few bio-inspired routing protocols, the proposed approach is novel. The results validate the concept that shortest route (with least hop count) and better link quality will be the best one. The results also denote that the route found by the proposed technique is far more accurate than another. However, further simulations and analysis need to be done on the proposed protocol.

References

1. Mitra, G., Chowdhury, C., Neogy, S.,:Application of Mobile agent in VANET for Measuring Environmental Data, IEEEXplore Proceedings of AIMoC, Kolkata, India, 2014, pp. 48–53.
2. Nzouonta, J., Rajgure, N. Wang, G., Borcea, C.:VANET Routing on City Roads Using Real-Time Vehicular Traffic Information. IEEE Transactions on Vehicular Technology, Vol. 58, No. 7, September 2009.
3. Patil, K., Dhage,M.: Enhanced Optimized Routing Protocol for Vehicular Ad-hoc Network. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, October 2013, pp. 4013–4017.
4. Hinds, A., Ngulube,M., Zhu,S.,Al-Aqrabi,H.:A Review of Routing Protocols for Mobile Ad-Hoc NETworks(MANET). International Journal of Information and Education Technology, Vol. 3, No. 1, February 2013, pp. 1–5.
5. Kaur, H.,Sahni, V., Bala, M.:A Survey of Reactive, Proactive and Hybrid Routing Protocols in MANET: A Review. International Journal of Computer Science and Information Technologies, Vol. 4(3), 2013, pp. 496–500.
6. Maitipe, B.,Ibrahim,U., Hayee, M.,Kwon, E.:Vehicle-to-Infrasturcture and Vehicle-to-Vehicle Information System in Work Zones. Transportation Reasearch Record: Journal of the Transportation Research Board, No. 2324, D.C, 2012, pp. 125–132.
7. Liu, X.:Ant Colony Optimization AlgorithmBased on Dynamical Pheromones for Clustering Analysis.International Journal of Hybrid Information Technology,Vol. 7, No. 2(2014), pp. 29–38.
8. Yu, X., Guo, H., Wong, W.:A Reliable Routing Protocol for VANET Communications. Proceedings 978-1-4577-9838-2©2011 IEEE, pp. 1748–1753.
9. S, B., Sathya, S.: A Survey of Bioinspired Optimization Algorithms. International Journal of SoftComputing and Engineering, Vol. 2, May 2012,pp. 137–151.
10. Al-Rabayh,M., Malaney, R.:Performance Comparison of VANET Routing Protocols. Proceedings 978-1-4244-6252-0/11 ©2011 IEEE.
11. Paul, B., Ibrahim, M., Bikas,M.:VANET Routing Protocols:Pros and Cons. International Journal of Computer Applications, Vol. 20-No. 3, April 2011,pp. 28–34.
12. Forderer, D.:Street Topology Based Routing.Master'sThesis,University of Mannheim,May 2005.
13. Al-Rabayh,M.Malaney, R.: A New Scalable Hybrid Routing Protocol for VANETs. IEEE Transactions on Vehicular Technology, Vol. 61, No. 6, July 2012, pp. 2625–2635.
14. V, B., Duraismy, V.: Improved AODV Based Link Quality Metrics. Intenational Journal of Advances in Engineering and Technology, November, 2012, pp. 269–275.
15. www.livescience.com/50399-radio-waves.html
16. www.eworkshop.on.ca/edu/pdf/CalculatePercentAccuracy.pdf
17. Li, B., Wang, L., Song, W.: Ant Colony Optimization for the Traveling Salesman Problem Based on Ants with Memory.Fourth International Conference on Natural Computation, Proceeding 978-0-7695-3304-9/08©2008 IEEE, pp. 496–501.

Time Dispersion Parameters for Double Bounce Geometrical Channel Including Rain Fading Effect

M.C. Anand, Divya Rani and B.K. Sujatha

Abstract In this paper, a two-dimensional geometrical channel modeling is taken under consideration to simultaneously describe the double bounce time dispersion characteristics of multipath channel. For ongoing 4G wireless communication technology of cellular mobile channel for microcell dense sub urban arborous area including rain fading effect. The ToA (Time of Arrival) pdf (Probability Density Function) and RMS delay spread of the received multipath signal are provided at the mobile station (MS) according to the distribution of trees as scatterer in uniformly distributed elliptical geometry. The Marginal ToA pdf and RMS delay spread has been derived and simulated as for the proposed model. The derived formulas can be further used to evaluate Power Delay Profile(PDP).

Keywords Modeling · ToA · Multipath channel · pdf · PDP

1 Introduction

The energy came out in all directions by a transmitter may take several paths before it is obtained at the front-end side of receiver. The propagation path to be taken by radio wave depends on certain causes producing an effect, such as: type of antenna and its height, frequency, and the air conditions and terrain [1]. In a Randomly made distribution where scatterer such as trees cause attenuation, scattering, diffraction of the transmitted wave before reaching at the other end of

M.C. Anand (✉) · B.K. Sujatha
Department of Telecommunication, M.S. Ramaiah Institute of Technology, Bangalore, India
e-mail: anandschama@gmail.com

B.K. Sujatha
e-mail: bksujatha@msrit.edu

D. Rani
Department of Telecommunication, Sir MVIT, Bangalore, India
e-mail: divyarani.9934@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_24

communication system so called destination, so the radiated wave may reach at the receiver by scattering via single obstructing obstacle like tree is called single bounce attenuation so, there may be further chances of fluctuation of electromagnetic signal by scattering from double obstructing obstacle like tress [2].

Time-dispersive nature of the multipath channel determines that the maximum data rate can be handled by transmitted plane wave without even requiring channel equalization and which also facilitates the accuracy of navigational services like example vehicle location. General characterization of temporal dispersion of the multipath channel signal is essential for wide band frequency system such as orthogonal frequency division multiplexing system thus direct sequence provides a measure of flat fading or frequency selectivity of the channel which is always expressed strongly with inverse of the signal bandwidth [9].

Vegetation plays an important role on the fading events in wireless communication system. Since Tamir [3] had lifted up the idea of a (lateral wave) quality common to a group of side wave (1–100 MHz of frequency range) propagation in forested scenario, much attempt has been Putin by expertise making an desired observations to work on the foliage medium as a dependence on the propagation channel. Lately the fading effect for UHF broadcast signal power measurements were taken for both wet divisions of the year (trees in leafs) and dry divisions of the year (trees relatively out of leaf) in Akure-Ilara way of ondo state, Nigeria and the outcome got were validated against the theoretical estimation. However, in the wet time, there were more attenuation of signal made a comparison to dry divisions of the year and signal degradation is a purpose of leaf density. As lower the density, the better the signal obtained at the other side of communication system [4].

The wideband scattering of canopies had been given in detail attention regarding rain fading effect. Where sub urban environment which were able to mimic the accurate scattering behavior of vegetation canopies. The environment covers trees, forests be used to appropriate study of the multipath propagation. Above 10 GHz of frequency bands, rain has significant effect because of obstruction of signal which causes attenuation of electromagnetic wave (signals), and which increases with increasing frequency. Vegetation can be responsible for both attenuation and the fluctuation of the transmitted signal which gives a fading effect to the signal. Multipath propagation leading due to reflections from buildings, refraction from midst of tress, buildings and terrain may occur as well [5].

The Rain attenuation varies slowly while vegetation effects come out in tightly signal driving power. These trees as a scatterer block the line-of-sight between the base station (BS) antenna and the mobile station (MS) antenna, this particular Signal fading caused by tree leaves has been widely investigated [6].

2 Proposed Model Description

Figure 1 [2] elaborates a typical propagation phenomenon with some local scatterer, i.e., tress located inside the elliptical region where uniformly distributed with both transmitter (Tx) and (Rx) receiver are located on the axis of foci of the elliptical geometry. Each and every different multipath signal of incident wave arrives at the front end side of mobile station (MS) are desirably considered only after the double bounce caused by two different trees (scatterer).

Where in this particular model transmitter (Tx) assumed as (BS) base station and receiver (Rx) act as mobile station (MS), omnidirectional vertical polarized dipole antenna is considered at both Tx and Rx. The distance between BS and MS (foliage depth) assumed to be exactly 1 km and obviously height of the BS is considered to be larger than MS. The impulse response of multipath channel can be generalized mathematically in continuous time domain as [7]

$$h(t,\tau) = \sum_{i=0}^{N-1} a_i(t,\tau)\ \exp\left(j\theta(t,\tau)\, d\left(t,\tau_i(t)\right)\right) \tag{1}$$

$a_i(t,\tau)$, $\theta\,(t,\tau)$ refers to the signal strength, phase and N is number of multipath. Multipath impulse response determines the specific component for each and every time t and $\tau_i(t)$ as excess delay.

Considering effects of rainfall on the propagation of the signal due to contribution of scatterer as four different layer model as shown in Fig. 2. Following respectively air, canopy layer, trunk layer, and last but least ground layer into the 2D elliptical geometry foliage region the propagation delay scenario can be estimated [8] as

$$\tau = \frac{D\sqrt{\varepsilon_r}}{c} \tag{2}$$

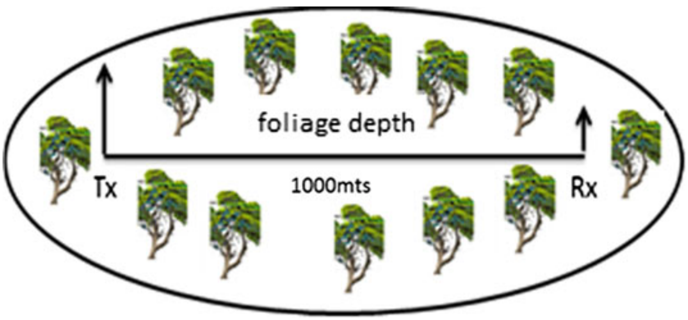


Fig. 1 Dense suburban arborous microcell scenario of trees

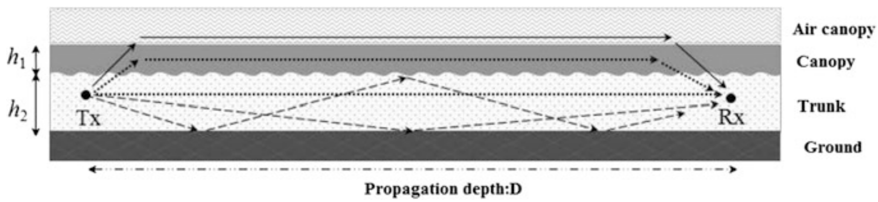


Fig. 2 Propagation phenomenon in four-layer

where C is the speed of light for free space radio wave, D is the propagation foliage depth in meters, ε_r relative permittivity of the all four homogenous layers. Where permittivity value for wet canopy layer for the dense suburban arborous area of forest type approximated between (1.1–1.3) with a assumption, most of the propagation take place at canopy layer.

3 Marginal ToA Pdf

Marginal ToA density function can be derived and obtained by two different methods for uniformly distributed elliptical geometry with trees as scatterer. Solving a mathematical equation based on area approach in order to obtain a general form of a ToA pdf for a uniformly distributed function which facilitates to give normalized values, i.e., by doing derivative of time of arrival (ToA) cumulative density function (cdf) with respect to τ where cumulative distribution function can be obtained by calculating probability of scatterer being matched inside the elliptical geometry. Second approach is integrating the Joint ToA/AoA probability density function over dθ (Angle of arrival parameter).

So according to second approach, integrating with respect to AoA will give rise to the exact value so, proceeding with it using the Joint AoA/ToA density function. Figure 3 [2] demonstrate the double bounce scenario, (r_{b_1}, θ_b) represents the pictorial view trees in two dimensional space.

The polar co-ordinates of the model are [10] obtained by transformation of rectangular co-ordinate.

$$r_{b_1} = \sqrt{x^2 + y^2} \tag{3}$$

$$\theta_b = a\ \tan\left(\frac{y}{x}\right) \tag{4}$$

$$\begin{aligned} x &= r_{b_1}\ \cos\theta_b \\ y &= r_{b_1}\ \sin\theta_b \end{aligned} \tag{5}$$

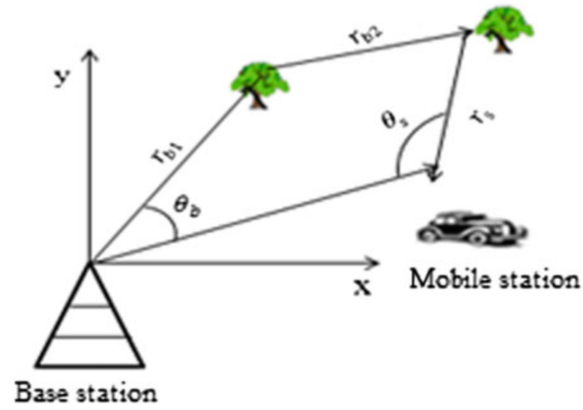


Fig. 3 Pictorial view of double bounce

Marginal ToA probability density function $f_\tau(\tau)$

$$f_\tau(\tau) = \int_{-\pi}^{\pi} f_{\tau,\theta_b}(\tau, \theta_b)\, d\theta \tag{6}$$

where Joint AoA/ToA density function can written as [2] follows

$$f_{\tau,\theta_b}(\tau,\theta_b) = \frac{(\tau^2c^2 - 2\tau c r_{b_2} - D^2 + 2Dr_{b_2})\begin{pmatrix} 2\tau c^2 r_{b_2}\cos\theta_b - 2\tau c^2 r_{b_2} - 2D\tau c^2\cos\theta_b + 2\tau^2c^2 \\ -2cr_{b_2}\cos\theta_b + +2Dcr_{b_2}\cos\theta_b + D^2c - 2Dcr_{b_2} + 2cr_{b_2}^2 - \tau^2c^3 \end{pmatrix}}{4\pi a_m b_m (r_{b_2}\cos\theta_b - D\cos\theta_b + \tau c - r_{b_2})^3} \tag{7}$$

According to definition of marginal ToA density function, substituting Eq. (3) in Eq. (2), we will arrive at

$$f_\tau(\tau) = \int_{-\pi}^{\pi} \frac{(\tau^2c^2 - 2\tau c r_{b_2} - D^2 + 2Dr_{b2})\begin{pmatrix} 2\tau c^2 r_{b2}\cos\theta_b - 2\tau c^2 r_{b_2} - 2D\tau c^2\cos\theta_b + 2\tau^2c^2 \\ -2cr_{b_2}\cos\theta_b + +2Dcr_{b_2}\cos\theta_b + D^2c - 2Dcr_{b_2} + 2cr_{b_2}^2 - \tau^2c^3 \end{pmatrix}}{4\pi a_m b_m (r_{b_2}\cos\theta_b - D\cos\theta_b + \tau c - r_{b_2})^3} d\theta \tag{8}$$

$$f_\tau(\tau) = \int_{-\pi}^{\pi} \frac{(\tau^2c^2 - 2\tau c r_c - D^2 + 2Dr_c)\left(\tau^2c^3 - 2\tau c^2 r_{b_2} + 2cr_{b_2}^2 + D^2c - 2Dcr_{b_2}\right) + \cos\theta_b(2\tau c^2 r_{b_2} - 2D\tau c^2 - 2cr_{b_2}^2 + 2Dr_{b_2})}{4\pi a_m b_m((r_{b_2} - D)\cos\theta_b + (\tau c - r_{b_2}))^3} d\theta \tag{9}$$

Above equation can be rewritten as given below

$$f_\tau(\tau) = a\int_{-\pi}^{\pi} \frac{b - e\cos\theta_b}{(g\cos\theta_b - f)^3} d\theta \tag{10}$$

Where all variable a, b, e, g, and f are

$$\begin{aligned} a &= \left(\tau^2c^2 - 2\tau c r_c - D^2 + 2Dr_c\right) \\ b &= \left(\tau^2c^3 - 2\tau c^2 r_c + 2cr_c^2 + D^2c - 2Dcr_c\right) \\ e &= -(2\tau c^2 r_c - 2D\tau c^2 - 2cr_c^2 + 2Dr_c) \\ f &= -(\tau c - r_c) \\ g &= (r_c - D) \end{aligned} \tag{11}$$

The ToA pdf seen at the base station is ditto same as the ToA pdf observed at the mobile station. Since multipath component traveled distance between base station and mobile station by mean of a tree as a scatter is independent of the point of view.

$$\begin{aligned} f_\tau(\tau) = a[\theta\{\frac{eg}{2f^4} - \frac{b}{f^3} - \frac{3bg^2}{f^5}\} + \sin\theta\,\{\frac{e^3}{f^3} + \frac{6eg^2}{f^5} - \frac{3bg}{f^4} - \frac{10bg^3}{f^6}\} \\ + \sin 2\theta\,\{\frac{3eg}{4f^4} - \frac{3bg^2}{2f^5}\} + \sin^3\theta\,\{\frac{10bg^3}{3f^6} - \frac{2eg^2}{f^5}\}]_{-\pi}^{\pi} \end{aligned} \tag{12}$$

After substituting the limits, the marginal ToA (Time of arrival) probability density function (pdf) we will get as follows:

$$\begin{aligned} f_\tau(\tau) = 2\pi c^2(\tau^2 - 2\tau\frac{r_c}{c} - \frac{D^2}{c^2} + \frac{2Dr_c}{c^2})[\frac{(\frac{D\tau}{c} + \frac{r_c^2}{c^2} - \frac{\tau r_c}{c} - \frac{Dr_c}{c^2})(\frac{r_c}{c} - \frac{D}{c})}{(\frac{r_c}{c} - \tau)^4} - \frac{(\tau^2 - 2\frac{\tau}{c}r_c + 2\frac{r_c^2}{c^2} + \frac{D^2}{c^2} - \frac{2Dr_c}{c^2})}{(\frac{r_c}{c} - \tau)^3} \\ - 3(\frac{\tau^2 - 2\tau\frac{r_c}{c} + 2\frac{r_c^2}{c^2} + \frac{D^2}{c^2} - 2\frac{Dr_c}{c^2}}{(\frac{r_c}{c} - \tau)^5})(\frac{r_c}{c} - \frac{D}{c})^2] \end{aligned} \tag{13}$$

a) **Double Bounce Mean** ($\overline{\tau}$)

It is possible to find double bounce mean just by integrating τ (time delay spread), and marginal ToA density function with respect to dτ. Where τ varies from ($\tau_{\min}$ = D/C to $\tau_{\max}$ = 5 μs) as shown below

$$mean = \int \tau f(\tau)d\tau \tag{14}$$

Now multiplying the constant term (k = r/C) from the above main equation, and it is assumed that r = 100 m.

$$\overline{\tau} = \frac{1}{(k-\tau)^5}((k-\tau)^4 + 3(k - t_{\min})^2((k-\tau)^2 + (k - t_{\min})^2)) \tag{15}$$

Integrating the main equation (marginal density function multiplied by τ) with respect to the dτ, with assumption of u = k − τ. After differentiating we will get dτ = −du which causes change in the limits

$$\int f(\tau).\tau d\tau = \int_{u=k-t_{\min}}^{u=k-\tau} (k-u)^2 h.(-du) \tag{16}$$

where

$$h = \frac{2\pi c^2}{u^5}\{u^2 - g^2[u^4 + 3g^2(u^2+g^2)]\} \text{ and } \mathrm{g} = (\mathrm{k} - \mathrm{t_{min}}) \tag{17}$$

$$\overline{\tau} = 2\pi c^2 (\frac{u^3}{3} - k\frac{u^2}{2} + (3g^6 k + 2g^2)u - 2g^2 k \log u + \frac{u^{-3}}{3} + 3g^6) \tag{18}$$

b) **Double Bounce Second central Moment ($\overline{\tau^2}$)**

Integrating the ToA equation (marginal density function) multiplied by τ^2 with respect to the $d\tau$, with assumption of $u = k - \tau$.

$$\overline{\tau^2} = \int \tau^2 f(\tau) d\tau \tag{19}$$

After simplification, double bounce second central moment expressed as

$$\cdot \int f(\tau).\tau^2 d\tau = \int_{u=k-t_{\min}}^{u=k-\tau} (k-u)^2 h.(-du) \tag{20}$$

$$\begin{aligned}\overline{\tau^2} = 2\pi c^2 (\frac{-u^4}{4} + 2k\frac{u^3}{3} - (k^2 + 2g^2)\frac{u^2}{2} - 2g^2 k^2 \log u - \frac{3}{2}g^6 u^{-2} - \frac{3}{4}g^6 k^2 u^{-4} \\ - 2g^6 k u^{-3} + 4g^2 k u)\end{aligned} \tag{21}$$

c) **RMS Delay spread (σ_τ)**

RMS Delay spread obtained by finding mean, i.e., nothing but the first central moment and second central moment and then taking square root of difference of mean from second central moment, which is denoted by (σ_τ) as follows [10]

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\overline{\tau})^2} \tag{22}$$

where

$\overline{\tau}$ = mean or first central moment
$\overline{\tau^2}$ = Second central moment
σ_τ = RMS delay spread

The calculation of RMS delay spread obtained from the temporal dispersion of the geometrical channel model constructed with the help of mathematics.

4 Simulation Result

The marginal ToA pdf and RMS delay spread are simulated in MATLAB. E in the simulation during code execution assumed as relative permittivity ($\mathcal{E}_r$) for different intensity of rain (1 to 1.3). The distance between BS-MS (foliage depth) D = 1 km, angle of arrival (AoA) θ_s for multipath signal varies from ($-\pi$, π), and delay spread τ (τ_{min} = D/C = 3.334 µs to τ_{max} = 5 µs) for the 50 different multipath. The channel sounder is a two-dimensional sounder at the centered frequency of 1.8 [GHz]. The sounder measurement and specifications were carried out in order to setup, and all those parameters are listed below in Table 1.

Table 1 Experimental investigation

S.no	Simulation parameters	Quantity
1	Carrier frequency (f_c)	1800[MHz] down link
2	Cellular Technology	4G LTE (long time evolution)
3	Tx signal band width (BW)	20[MHz]
4	Tx signal time period	0.326 µs
5	Tx and Rx antenna	Omnidirectional vertical polarized dipole antenna
6	Tx power (P_t)	46 dBm
7	Tx Gain (G_t)	18 dB
8	Rx power (P_r)	2 dB
9	Separation between (Tx-Rx)	1000 m

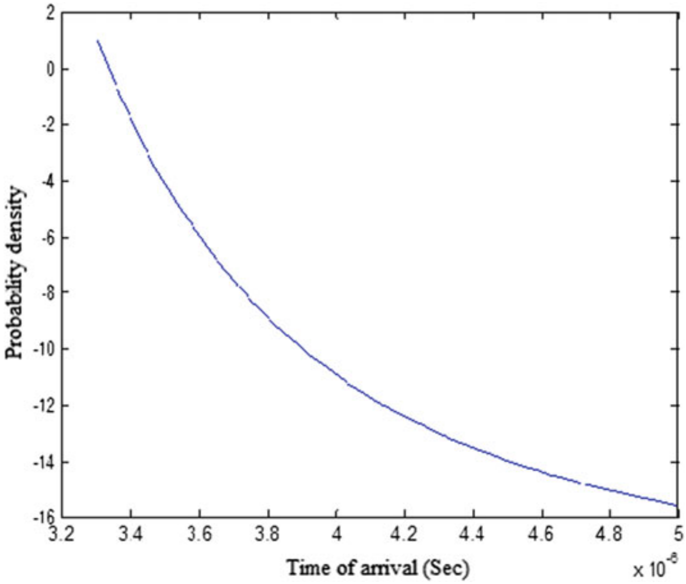


Fig. 4 Elliptical model ToA pdf

Table 2 Time dispersion parameters at different rain intensity

Relative permittivity ($\mathcal{E}_r$) based on rain intensity	Double bounce mean in micro seconds, μs (10^{-6})	Second central moment in Pico seconds (10^{-12})	RMS delay spread (σ_τ) in μs (10^{-6})
$\mathcal{E}_r$ = 1.0 (0.9 mm/h)	2.057	8.928	2.167
$\mathcal{E}_r$ = 1.1 (6.8 mm/h)	2.157	9.821	2.273
$\mathcal{E}_r$ = 1.3 (24.4 mm/h)	2.345	11.607	2.471
Ɛr = 1.1 (6.8 mm/h)	2.157	9.821	2.273
Ɛr = 1.3 (24.4 mm/h)	2.345	11.607	2.471

Based on the obtained simulation result Fig. 4 elliptical model ToA pdf is a plot of time of arrival versus probability density. As intensity of rain increases, $\mathcal{E}_r$ with wet condition canopy layer also increases this intern looking at the plot shows that there is high density of scatterer (Trees) with relatively small delays.

According to Table 2 simulation result, time dispersion parameters for the geometry of elliptical model double bounce rain fading effect are calculated based on the assumption of r = 100 m (distance between two scatterer), D = 1000 m, delay varies (τ_{min} = D/C to τ_{max} = 5 μs), and angle of arrival θ_b = ($-\pi$ to π). As rain intensity increases, $\mathcal{E}_r$ increases which leads to widening the spread of the signal respectively for higher bounce. The obtained simulation results are further used to analyze the power delay profile of multipath channel component.

5 Conclusion

In this paper, the marginal ToA probability density function (pdf) and time dispersion parameter (mean, second central moment, and RMS delay spread) for the geometry of an elliptical model based double bounce rain fading effect is derived and obtained. According to results as there is increase in rain intensity, effective relative permittivity ($\mathcal{E}_r$) increases which in turn causes the decrease in density of scatterer of multipath component, and there is also widening the spread of the signal for higher bounce. The obtained simulation results with proposed model are further useful to find power delay profile, cyclic prefix in case of different rain event and then to further analyze of the performance of 4G cellular OFDM system.

References

1. R. B. Ertel, and J. H. Reed, "Angle and time of arrival statistics for circular and elliptical scattering models," IEEE J. Select. Areas commun, vol. 17, no, 11, pp, 1829–1840 Nov, 1999.
2. M.S. Divya Rani, Anand M C, Dr.B.K Sujatha "Joint Angle and Time of Arrival Statistics of Rain Fading Effect for Second Bounce Cellular System" conducted by the IEEE International Conference on smart sensors and systems (IC –SSS 2015), December 21st–23rd 2015.

3. Tamir, T, “On radio wave propagation in forest environments” IEEE Trans. Antennas Propag., vol. 15, no. 6, 806–817, 1967.
4. Ayekomilogbon Olufemi,“UHF band Radio Wave Propagation Mechanism in Forested Environments for Wireless Communication Systems” Journal of Information Engineering and Applications ISSN 2224-5782 (print) ISSN 2225-0506 (online) Vol. 3, No. 7, 2013.
5. Frank M. Schubert, Bernard H. Fleury, Roberto Prieto-Cerdeira, and Alexander Steingass, Andreas Lehner, “A Rural Channel Model for Satellite Navigation Applications”, 6th European Conference on Antennas and Propagation (EUCAP), Aalborg University, 9220 Aalborg East, Denmark.
6. Eric R. Pelet, “Effect of Wind on Foliage Obstructed Line-of-Sight Channel at 2.5 GHz”, IEEE transactions on broadcasting, vol. 50, no. 3, september 2004.
7. Rafael F.S. Caldeirinha, Miqdad O. Al-Nuaimi “Wideband Characterisation of the Dispersive Effects on Isolated Mature Trees for 2-GHz Band Urban Microcells,” unpublished.
8. Y. S. Meng, Y. H. Lee, and B. C. Ng “ Further study of rainfall effect on VHF forested radio wave propagation with four- layered model” Progress In Electromagnetism Research, PIER 99, 149–161, 2009.
9. N. M. Khan, M. T. Simsim, and P. B. Rapajic, “A generalized spatial model for all cellular Environments”, in Proc. IEEE SympoTIC 04, Bratislava, Slovakia, 2004, pp. 33–38.
10. Papoulis, Probability, Random variables, and Stochastic process, 3rded. New York: McGaw-Hill, 1991.

Assistive Text on Hand Held Objects for Blind People

Samruddhi Deshpande and Revati Shriram

Abstract This paper presents camera-based system which will help blind person for reading. This is the framework to assist visually impaired persons to read text patterns and convert it into the audio output. To obtain the object from the background and extract the text pattern from that object, the system first proposes the method that will capture the image from the camera and object region is detected. The text which are maximally stable are detected using Maximally Stable External Regions (MSER) feature. The detected text is compared with the template and converted into the speech output. The text patterns are localized and binarized using Optical Character Recognition (OCR).The recognized text is then converted to an audio output. Experimental results shows the analysis of MSER and OCR for different text patterns. MSER shows that it is robust algorithm for the text detection. Therefore, this paper deals with analysis of detection and recognition of different text patterns on different objects.

Keywords Maximally stable external regions (MSER) · Optical character recognition (OCR)

1 Introduction

Reading is essential in day-to-day life. Printed text is present everywhere in the form of documents like reports, receipts, statements, restaurant menus, product packages, instructions, etc. There are many devices that can provide good access to common objects such as product packages and objects with printed text such as names of different brands. The ability of people who are visually poor or have significant visual impairments to read printed text and product packages will enhance independent living and foster economic and social self-sufficiency [1–4].

S. Deshpande (✉) · R. Shriram
Department of Instrumentation and Control, Cummins College of Engineering for Women Karvenagar, Pune 411052, India
e-mail: samruddhideshpande512@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_25

In order to solve the task that to extract text information from complex backgrounds with variable text patterns, this system proposes a text detection and localization algorithm that combines detection and conversion of text patterns. This actually differentiate text characters and discriminative text features from background outliers [5–7]. The problem is challenging that to automatically localize and detect the text from ROI in the captured images with different backgrounds. The text captured images are mostly contain noise. The text characters are always present with different fonts, variable scales, and colours. The text containing images are varying in orientation and alignment. Many algorithms are present for localization of text regions in image captured scene images. In assistive reading systems [8–10], it is very difficult for blind person to find the position of object of interest from the centre of the camera. In order to assist the blind person, this system will work to track the object of interest within the camera view. This proposed algorithm will effectively handle the different text patterns and different backgrounds, to extract the text information from the camera captured image. In order to solve the problems regarding this we are defining a novel feature for extraction of text from the image [11–15]. In order to detect the text from natural scene images, it is very difficult to find it [16, 17]. The natural scenes contain different shapes, sizes, orientations and complex backgrounds. Hence to detect the such text patterns, this system will help blind user to easily locate the object and get the information on the hand held objects. As OCRAs OCR is used it will recognized the text into separate letters. However, this paper is more focused on the text detection algorithms. This system will detect the text with different shapes, sizes, and orientation of text objects. In order to enhance the blind people to become independent socially and economically, assisting text is one of the helping hand for them. Today there are many system available, it is very difficult for blind user or visually impaired person to get location of that object. Such systems like s barcode scanner in Braille, but it is very difficult for blind user to locate the position of the object. So, this system is proposing to solve this problem of blind people. This system will easily detect the text patterns from the image The detected text are given as speech output to blind person [9] (Fig. 1).

2 Framework

This prototype system presents text reading from the camera captured image. Framework consists of three functional section scene capture, data processing, and audio output. The scene capture components collects the images of different objects and finds the object of interest from the image. In this system camera is used for scene capture purpose. The data processing component is used for the detection of text patterns to extract the text information from image. The data processing includes [12] (1) Object of interest detection to get the text patterns from image (2) Text localization to obtain text regions and to recognized them. The aim of the system is to assist the blind people, to ensure the object is appearing in the camera,

Fig. 1 Image samples applied for text read

it is provided with wide angle of view. To get the actual object in the view, the user can shake the object for couple of seconds. Background subtraction is one of method for detecting the object of interest from the image. The most effective tool to detect the object from the image. The algorithm must be robust to detect difference between lightening and shaking. The proposed system is able to detect the text strings properly. The laptop is used as processing device for the audio output. he audio output is used to assist the blind user. Easiest way to inform blind user is the audio output. The flowchart for the proposed system is shown in figure. The main contribution of the system is to solve the problem for blind people by using

1. Simply by detecting the text pattern from captured scene.
2. A novel method for localizing and recognising the detected text.
3. Audio output of the recognized text to the blind user. The flowchart explains the system very well.

Firstly, we are taking images using the camera. Then the proposed system will detect the object of interest from the image using MSER features. After that text patterns are detected from the image. These text patterns are binarized and localized. The localized text are recognized using OCR software tool. The recognized text patterns are given as audio output to blind people. The external regions are detected using the anchor points. Local extremas are detected at multiple scales. The characteristics function of the region is given as,

$$\mathrm{f(t)} \;= |\mathrm{I(t)} - \mathrm{I0}|/\left((1/\mathrm{t}) * \int |\mathrm{I(t)} - \mathrm{I0}| \right) \tag{1}$$

The approximation of region is up to second order moment, uses ellipse to approximate the region. The ellipse of second order is used same as that of the moment of region

$$\mathrm{mpq} = \iint \mathrm{x}^\wedge\mathrm{p} * \mathrm{y}^\wedge\mathrm{p} * f(\mathrm{x,y})\mathrm{dxdy} \tag{2}$$

$$\mathrm{q} = \mathrm{Ap} \tag{3}$$

$$\sum(1) = \mathrm{A}\sum(2) * \mathrm{A}^\wedge\mathrm{T} \tag{4}$$

$$\mathrm{p} = [\mathrm{x}, \mathrm{y}]^\wedge\mathrm{T} \tag{5}$$

where p is relative to centre of mass (Fig. 2).

3 Text Detection

MSER is used for detection of blob in the images. A number of co-variant regions are extracted from the image using MSER algorithm. There are various text detection methods available for text detection. The comparison of all the methods is listed in a following table. MSER are the connected set of components of some grey-level image, which are stable. MSER takes regions, which are nearly same through a wide range of threshold levels. The pixels which are having threshold below some threshold value are shown as white and the pixels which are above and equal to the threshold value are shown as black. This can be done by sorting all the pixels by the grey value Various text detection methods are available today [1].

The algorithm

1. Starts from the local intensity extreme point.
2. Go in every direction as a function of intensity.
3. The curve connects the points to create the region boundary.
4. Compute geometric moments.
5. Replace the region with ellipse (Fig. 3).

The external word refers to the bright or dark intensity pixels. MSER has dark or bright pixels than all the other pixels present on its boundary. All the pixels are first sorted by the grey value and then added incrementally to each component depending upon threshold. Area region is varied with respect to the threshold . When threshold is minimun the regions are maximally stable (Fig. 4).

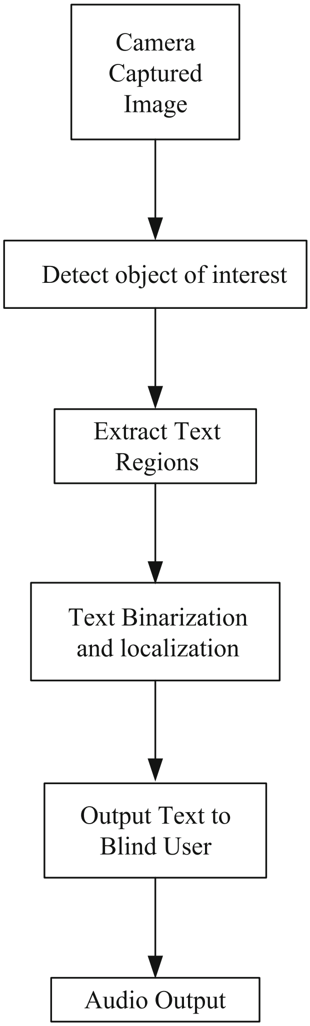


Fig. 2 Flow chart if proposed system

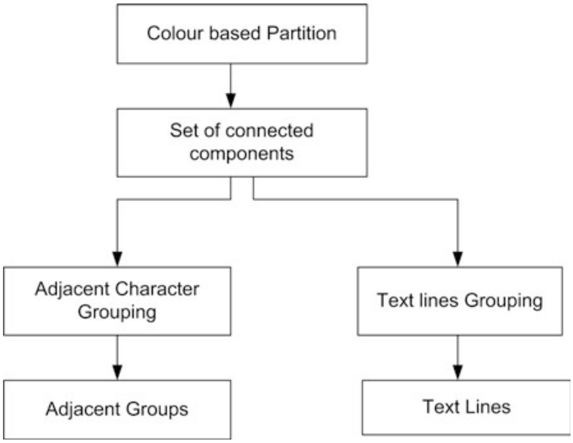
The The MSER follows following steps

1. Performing a luminance thresholding of an image, just by sweeping threshold of intensity from black to white.
2. Extract the external regions or connected components.
3. Find out the external regions are maximally stable or not, that is to calculate local minima.
4. Appropriate region is indicated by ellipse.
5. Thus, we get the region detected features.

The external regions may be rejected by the following reasons If it is too big or small. Also, when it is unstable or similar to the Parent MSER. The external regions

Fig. 3 Some results of text detection on different objects describing that detected text are highlighted in different colours and text information is extracted

Fig. 4 Flowchart for detection of text



have some properties such as they have set of continuous transformation of image coordinates. These regions are the function intensity. These regions are detected only they have the stable linearization than other pixels present locally. Detected regions do not changed even if the image is skewed.

(A) Properties of MSER

1. Invariance between the image intensities
2. Stable
3. Multiscale detection, different shapes and sizes of letters are detected.
4. Affine invariant
5. Better repeatability
6. Homogeneous regions are detected.
7. Small regions are detected
8. It does not works well with motion blur.
9. MSER is good for change in orientation and scale
10. It is best for JPEG change

4 Text Recognition

Optical character recognition is the electronic tool for conversion of typed letter form to printed text pattern. OCR is tool that converts scanned documents into text searchable form. A text region with minimum rectangular area for space are selected. OCR generates better performance to text patterns who are assigned with proper margin areas and binarized to segments of texts. OCR is tool which is used for text recognition in this proposed system. The recognized text are recorded in script files and are given to the blind user. Microsoft Speech software development kit is used to play the audio output. The user can access it accordingly [1] (Table 1).

Table 1 Comparison of different text detection methods

Detector	Feature type	Scale independent
FAST	Corner	No
Minimum Eigen value algorithm	Corner	No
Corner detector	Corner	No
SURF	Blob	Yes
BRISK	Corner	Yes
MSER	Region with uniform intensity	Yes

5 Experiments

(A) *Datastes*

Fifty images are used to evaluate this algorithm. The image resolution is in the range of 640 *480 to 1600 *1200. The analysis is based on the adjacent characteristics and stability of the text patterns. The Logitech camera is used to take the images of the objects. The camera is of 5MP with built in mic and noise reduction facility. The prototype system is developed to take images from the camera and this camera is connected to Sony vaio laptop through USB interface. The laptop performs the processing part of the system and gives the speech out[put to the blind person. The system is robust to the orientation, complex backgrounds, and lightening. Automatic detection and recognition is applied to different images separately. This system is robust and effective for different handheld objects. The system will capture the image and extracts the text patterns from it. The system will cropped the image according to the text present in the image. The cropped image is then separated to individual letters. The letters are than read as the speech output to the blind user

(B) *Evaluation of text detection*

Two techniques are used to detect the text and recognized that text in this proposed system. The results are evaluated on the dataset of captured images. This system is instructing the blind users to read the handheld objects. Using this algorithm we are correctly detecting the text in proper area. After the localization of text, these text are recognized using OCR tool This system is actually based on the reading text from camera-based images. The efficiency of the system will be improved by parallel processing of text detection and recognition. The output is giving text as the speech output to blind people.

6 Conclusion

In this work, a novel text detection and recognition algorithm is proposed. The system describes about reading of printed letters on the objects for helping the blind people. The problem for the blind persons in reading is solved by this proposed method. This system will distinguish the object from the captured image and detect the texts from the image. The proposed method is demonstrated for localizing text in natural images. MSER is used to detect the text information from the image. This will extracts the text patterns from different backgrounds. This is the robust algorithm for text detection. OCR is used for text recognition on the localized text patterns. The system will transform the localized text to the audio output for blind people. The results point out that it is be possible to achieve high performance using MSER and OCR for text detection and recognition.

References

1. Chucai. Yi, Y. Tian. and Areis. Arditi, "Portable Camera-Based- Assistive-Text-and-Product Label Reading From Hand-Held Objects for Blind Persons" IEEE/ASME transactions on mechatronics
2. X. Chen and A. L. Yuille, Detecting and reading text in natural scenes, in Proc. Comput. Vision Pattern Recognit., 2004, vol. 2, pp. II-366II–373.
3. B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, in Proc. Comput. Vision Pattern Recognit., 2010, pp. 2963–2970.
4. N. Giudice and G. Legge, Blind navigation and the role of technology, in The Engineering Handbook of Smart Technology for Aging, Disability, and Independence, A. A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. Hoboken, NJ, USA: Wiley, 2008.
5. D. Dakopoulos and N. G. Bourbakis, Wearable obstacle avoidance elec- tronic travel aids for blind: A survey, IEEE Trans. Syst., Man, Cybern., vol. 40, no. 1, pp. 2535, Jan. 2010.
6. S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, Text extraction and document image segmentation using matched wavelets and MRF model, IEEE Trans Image Process., vol. 16, no. 8, pp. 2117–2128, Aug. 2007.
7. X. Chen, J. Yang, J. Zhang, and A. Waibel, Automatic detection and recognition of signs from natural scenes, IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.
8. K. Kim, K. Jung, and J. Kim, Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.
9. C. Yi and Y. Tian, Text string detection from natural scenes by structure based partition and grouping, IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sep. 2011.
10. P. Viola and M. J. Jones, Robust real-time face detection, Int. J. Comput. Vision, vol. 57, no. 2, pp. 137–154, 2004.
11. C. Yi and Y. Tian, Assistive text reading from complex background for blind persons, in Proc. Int. Workshop Camera-Based Document Anal. Recognit., 2011, vol. LNCS-7139, pp. 1528.
12. P. Shivakumara, T. Phan, and C. L. Tan, A gradient difference based technique for video text detection, in Proc. Int. Conf. Document Anal. Recognit., 2009, pp. 6670.
13. M. Shi, Y. Fujisawab, T. Wakabayashia, and F. Kimura, Handwritten numeral recognition using gradient and curvature of gray scale image, Pattern Recognit., vol. 35, no. 10, pp. 2051–2059, 2002.

14. X. Chen and A. L. Yuille, Detecting and reading text in natural scenes, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, pp. 366373.
15. K. C. Jung, K. I. Kim, and A. K. Jain,Text information extraction in images and video: A survey, Pattern Recognit., vol. 5, pp. 977–997, May 2004.
16. Y. M. Y. Hasan and L. J. Karam, Morphological text extraction from images, IEEE Trans. Image Process., vol. 9, no. 11, pp. 1978–1983, Nov. 2000.
17. T. Kasar, J. Kumar, and A. G. Ramakrishnan, Font and background color independent text binarization, in Proc. 2nd Int. Workshop Camera-Based Document Anal. Recognit., 2007, pp. 3–9.

Intra Frame Coding in H.264 to Obtain Consistent PSNR and Reduce Bit Rate for Diagonal Down Right Mode Using Gaussian Pulse

Manjanaik N. Manjanaik and Manjunath Ramachandra

Abstract Intra prediction process of H.264 advanced video coding standard used to code Intra frame to achieve higher coding efficiency. Rate Distortion Optimization (RDO) method of H.264 use to code Intra frame in intra prediction process. This method increases more computational complexity and difficult to implement in real time applications. The previous work on Intra frame coding was achieved increased in Bit rate, degradation/loss of PSNR and reduction of computational complexity for different quantization parameters. In this paper developed a better approach for Intra frame coding using Gaussian pulse and diagonal down right mode to achieve higher coding performance. The proposed algorithm is based on Gaussian pulse which is multiplied with each 4×4 transformed coefficients of 4×4 sub macro block of Intra frame before quantization block which provides signal in smooth and avoids image get degrade and achieve reduced bit rate for coarse quantization parameters. The proposed algorithm implemented using MATLAB and compared with original JM18.6 reference algorithm of H.264. The results of proposed method are compared with previous algorithm Jin Wan et al. The simulation results of the proposed algorithm achieved consistent PSNR and reduced in bit rate of 27.38% QCIF and 30.63% CIF sequences for diagonal down right intra prediction mode using Gaussian pulse compared with previous algorithm Jin Wan et al.

Keywords H.264 · PSNR · QCIF · MPEG · QP

M.N. Manjanaik (✉)
Department of ECE, UBDT College of Engineering, Davangere 577004, India
e-mail: manjubdt2009@gmail.com

M. Ramachandra
Wipro Technolgy, Bangalore, India
e-mail: manju_r_99@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_26

1 Introduction

The Joint Video Team (JVT) developed H.264 video coding standard. This standard provides better coding performance compared to older video coding standards. It has many advanced features due to these features it achieves higher coding efficiency while it increases the computational complexity which is difficult for real-time implementation applications. To provide solutions of good quality, good compression, and low bit rate of intra frames for storage, transmission etc., video compression algorithms are essential. For Intra frame coding, AVC adopts Rate Distortion Optimization technique (RDO) with full search mode algorithm. But RDO with full search mode algorithm is computationally complex and it is very difficult to implement for real time applications. So in order to reduce complexity many researcher have been developed fast mode decision intra prediction algorithms for Intra frames. Several authors developed Intra prediction algorithms for Intra frame coding using enhanced of cost function, local edge information of macro block, early block type selection, feature of macro block, variance of macro block and approximating the cost function, etc. [1–8]. Using these approaches achieve increase in bit rate, degradation of PSNR and reduction of computational complexity or save encoding time. The previous work on Intra frame coding in AVC/H.264 using fast mode decision Intra prediction algorithm was achieved increase in bit rate, degradation/loss of picture quality, i.e. PSNR and reduction of computational complexity for different quantization parameters. The previous algorithms achieved saving of encoding time with loss of picture quality (PSNR) in addition to the increase of bit rate of intra frame. So far, in Intra frame coding reduction in bit rate, consistent of PSNR or controlled PSNR and high compression was not achieved and Gaussian pulse approach is not used for intra frame coding.

2 Methodology

This section presents the problem formulation for Intra frame coding to achieve the performance parameter of proposed work. The problem formulated as increase in bit rate and picture quality (PSNR) reduces for intra frame coding in fast mode decision intra prediction algorithm of AVC also loss of picture quality in previous video coding standards for a higher value of quantization parameters. The International standard yuv video sequences in QCIF and CIF resolution which is available as open-source downloaded from the website www.codervoice.com as input. Extracting Intra frame of yuv video sequence, processed in units of macro blocks. The macro block is divided into 4×4 submacro blocks. Each sub macro block of intra frame is processed using proposed functional blocks such as intra prediction, integer transformer, Gaussian pulse, quantization, context adaptive variable length coding to get compressed bit, inverse quantization and inverse transform to get reconstructed frame. Finally, performance parameters such as PSNR, and Bit rate of reconstructed frame of proposed algorithm are measured.

3 Proposed Diagram

Figure 1 shows the proposed system consists of different functional modules. Input intra frame is processed in terms of 4 × 4 sub macro block. Coding of each sub macro block is done with the help of intra prediction. Prediction P is obtained using prediction equations of prediction modes. The prediction block is subtraction with the current sub macro block to get residual coefficients. The transformed coefficients of residual data are obtained using 4 × 4 integer transform. The Gaussian pulses are multiplied with transformed coefficients to get Gaussian coefficients. The quantization process of AVC are quantized the Gaussian coefficients to get quantized coefficients. Context Adaptive Variable-Length Coding (CAVLC) used encode the quantized coefficients. These quantized coefficients are zigzag scanned and reordered and encoded to get the compressed bits. The prediction macro block P is added to residual block to reconstructed block by followed process of inverse quantization and inverse transform to quantized coefficients of intra frame. This process repeat till at the end of last sub block of intra frame to get reconstructed block. Finally measure the performance parameters of proposed algorithm.

4 Intra Prediction

The prediction unit of AVC used to predict within the intra frame called intra frame coding. There are nine prediction modes in intra prediction unit, such as horizontal (Mode-1), diagonal down left (Mode-3), vertical (Mode-0), vertical left (Mode-7), horizontal down (Mode-6), horizontal up (Mode-8) diagonal down right (Mode-4),

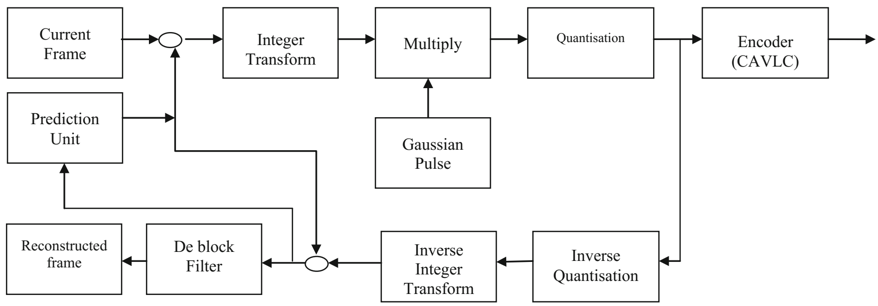


Fig. 1 Proposed block diagram

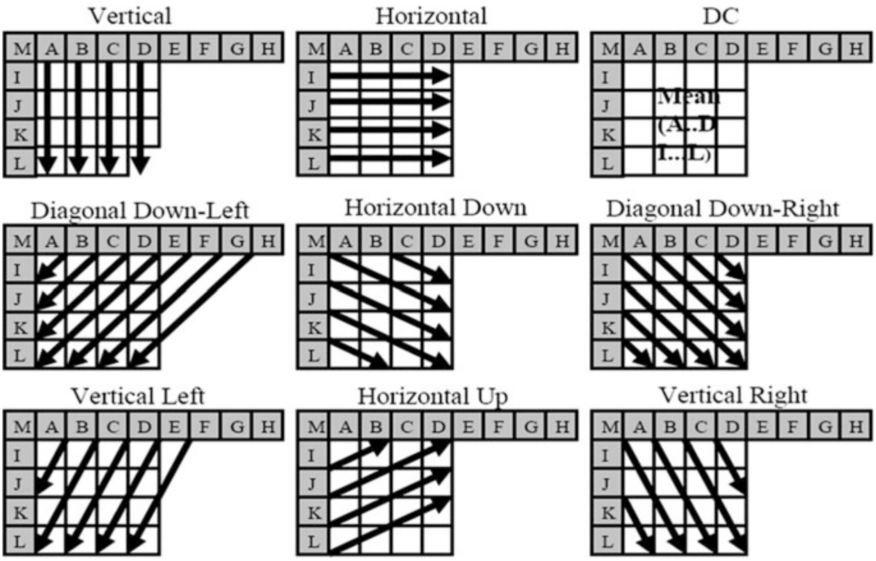


Fig. 2 4 × 4 Intra prediction modes

DC (Mode-2), and vertical right (Mode-5) which is shown in Fig. 2 in which the pixels **A–M** are reconstructed pixels. All the current pixels of **a–m** of current block are replaced based on of previously the reconstructed pixels of reconstructed block. In intra frame spatial redundancy are exploit by prediction modes. Mode 2 is intra prediction mode with no direction. Mode 4 specifies diagonal down right mode it replace all pixels/samples in the current 4 × 4 block by the average weighted samples (**A–M**) of reconstructed 4 × 4 block. In the proposed work, Mode-4 is used for Intra frame coding.

5 Gaussian Pulse

The proposed method for intra frame coding is based on the Gaussian pulse. Gaussian pulse is shaped as a Gaussian function or Gaussian distribution shown in Fig. 3 [6–7]. The Gaussian pulse equation described in Eq. (1) used for intra frame coding.

$$G(v) = \exp(-v^\wedge 2) \tag{1}$$

This Gaussian pulse gives abstract frequency domain samples without inter mixing frequency domain information among the samples. This Gaussian operation implemented before quantization process of advanced video coding standard.

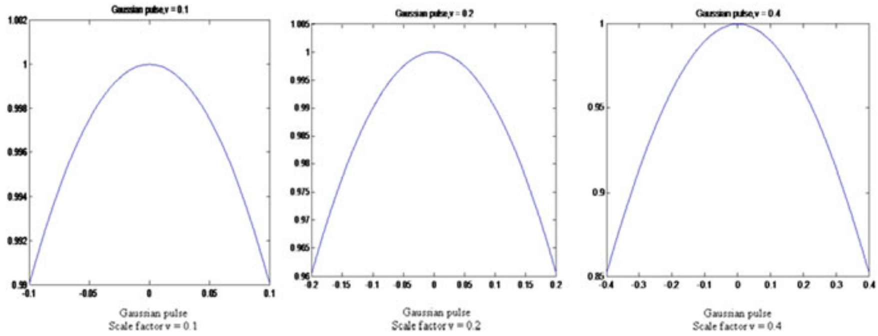


Fig.3 Generation of Gaussian pulses

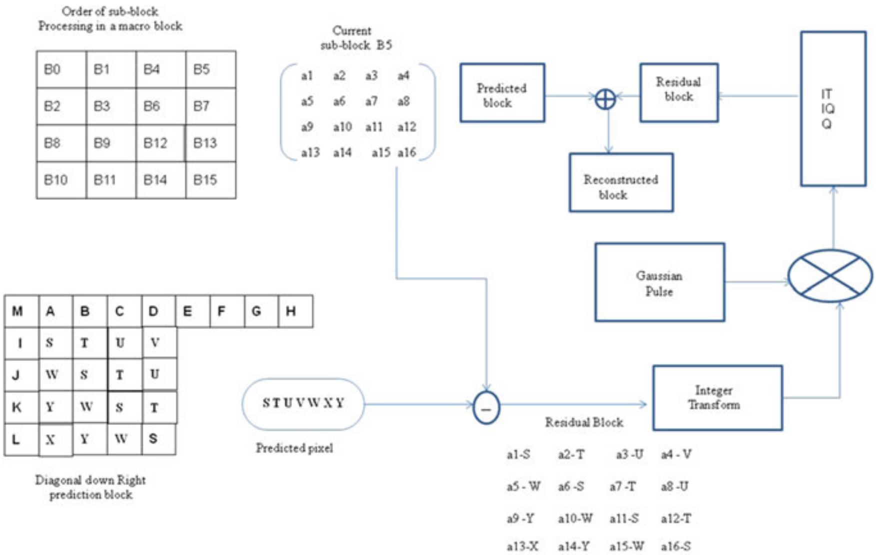


Fig. 4 Opreation of diagonal down right mode

6 Diagonal Down Right Intra Prediction

Figure 4 shows operation of diagonal down right prediction mode, there are three steps involved in diagonal down right prediction mode such as order of processing of sub block, prediction block and encoding and reconstructing intra frame. In diagonal down right prediction mode, all the pixels of current sub block is predicted by linearly weighted average of reconstructed pixels **A–M** of reconstructed block. For diagonal down left prediction mode prediction equations generated as follow

$$
\begin{aligned}
a &= f = j = p = (I+2M+A+2) \gg 2 = ((I+M)+(M+A)+2) \gg 2 = \mathbf{S} \\
b &= g = l \ = (M+2A+B+2) \gg 2 = ((M+A)+(A+B)+2) \gg 2 = \mathbf{T} \\
c &= h = (A+2B+C+2) \gg 2 = ((A+B)+(B+C)+2) \gg 2 = \mathbf{U} \\
d &= (B+2C+D+2) \gg 2 = ((C+D)+(C+B)+2) \gg 2 = \mathbf{V} \\
e &= j = o = (J+2I+M+2) \gg 2 = \mathbf{W} \\
m &= ((K+L)+(K+J)+2) \gg 2 = (L+2K+J+2) \gg 1 = \mathbf{X} \\
i &= n = ((K+J)+(I+K)+2) \gg 2 = (K+2J+I+2) \gg 2 = \mathbf{Y}
\end{aligned} \tag{2}
$$

a, b, c, d, e, f, g, h, i, j, k, l, m, n, o and p are pixels of current sub block. These pixels are predicted using previously reconstructed pixels **A–M**. Now the pixels **S T U V W X Y** are the predicted pixel. After processing of order of sub block, first block encoded and decoded using forward and reverse path of AVC. Predicted pixels (**STUVWXY**) are obtained using the diagonal down left intra prediction mode. Residual block is obtained by subtraction of predicted pixels with current block pixels. The residual pixels are transformed, Gaussian multiplied, quantized, inverse quantized, inverse transformed and add prediction to get reconstructed block. This process repeats till the end of last sub block of intra frame.

7 Algorithm

For intra frame coding, the following steps are required to achieve the objective of the proposed method

- The QCIF and CIF yuv video sequence as input.
- Read Intra frame from input test yuv video sequence (I-frame).
- Intra frame is encoded in the form of macro blocks of size 16 × 16.
- Macro block is divided into a 16 4 × 4 sub macro blocks.
- The first block of sub block is processed directly followed by forward encode path without using previously reconstructed block.
- Obtain reconstcted block using inverse quantization and transform with addition of prediction.

  Obtained prediction block of reconstruct a 4 × 4 sub block using Diagonal down right intra prediction mode.
- By subtracting current sub block with prediction block to get residual block.
- Apply integer transform, Gaussian pulse, quantization and encoder process for residual data at forward path of encoder and reverse process such as inverse quantized, inverse transform and add prediction at reconstruction path of encoder.
- Finally measure performance parameters such as PSNR and bit rate for different QP with Gaussian values.

Residual coefficients are integer transformed using the equation

$$W = A \times A^{T} \tag{3}$$

The Gaussian pulses are multiplied with transformed coefficients for different Gaussian values. The resulting coefficients are, Gaussian integer transformed coefficients.

$$G(N) = W * G(v) \tag{4}$$

These coefficients are quantized using the quantization process as

$$Z = \mathrm{round}(G(N)/Qstep) \tag{5}$$

Quantized coefficients are zig zag scaned, reordered and encoded using CAVLC.

The quantized coefficients are inverse quantized by using inverse quantization process given by

$$W' = Z.\,Qstep \tag{6}$$

W' are inverse quantized coefficients and Z are quantized coefficients. Inverse quantized coefficients are inverse transformed by inverse quantization process given by

$$X' = Ai' * \ W' * Ai \tag{7}$$

X' are inverse transformed coefficients. The inverse transformed coefficients are added with prediction.

8 Experimental Results

The performance parameters are measured using the following formula.

$$\Delta P = \mathrm{PSNR}(\mathrm{Proposed}) - \mathrm{PSNR}(\mathrm{JM\,reference}) \tag{8}$$

$$\Delta B = \frac{\mathrm{Bit\,rate}(\mathrm{proposed}) - \mathrm{Bit\,rate}(\mathrm{JM\,reference}) \times 100}{\mathrm{Bit\,rate}(\mathrm{JM\,reference})} \tag{9}$$

The simulation conditions are specified in table. The coding performance was measured by JM18.6 AVC algorithm. Simulation results of the proposed algorithm were compared with JM 18.6 algorithm. Table 1 indicate the simulation results of Intra frame in QCIF format. I Table 2 indicate the experimental results for Intra frame in CIF. The simulation results are obtained for Intra frame in QCIF and CIF sequence under different quantization parameters, Gaussian value 0.1 with diagonal

Table 1 Simulation results of intra frames: between proposed method and JinWang method

Sequence QCIF	Method	Parameter	Quantization parameter (QP)					Proposed (bit rate) %	Jin Wang [8] (bit rate) %	Remarks
			42	37	32	27	22			
News	Proposed	PSNR	37.55	37.60	37.65	37.69	37.70	**31.08**	**11.63**	**Proposed method achieves higher Bit rate saving and PSNR is Consistent**
		Bit rate	69.02	101.76	147.69	205.54	278.98			
	Reference (JM18.6)	PSNR	27.09	37.46	34.07	30.66	40.98			
		Bit rate	77.92	139.92	218.40	340.46	504.96			
Foreman	Proposed	PSNR	36.70	36.72	36.78	36.82	36.90	**28.76**	**2.98**	
		Bit rate	84.42	118.47	162.65	214.28	282.54			
	Reference (JM18.6)	PSNR	26.95	30.52	34.11	37.94	41.70			
		Bit rate	88.32	154.24	238.48	354.40	511.20			
Container	Proposed	PSNR	36.06	37.08	36.99	36.95	37.90	**23.78**	**2.61**	
		Bit rate	76.31	109.05	155.53	212.36	287.78			
	Reference (JM18.6)	PSNR	27.44	30.73	34.07	37.48	41.14			
		Bit rate	76.88	132.89	208.80	319.60	483.76			
Carphone	Proposed	PSNR	36.89	36.84	36.81	36.76	36.74	**28.15**	**5.47**	
		Bit rate	73.48	106.53	150.75	206.38	279.91			
	Reference (JM18.6)	PSNR	27.94	31.31	34.40	37.88	41.63			
		Bit rate	71.44	124.64	194.48	303.12	456.80			
Hall	Proposed	PSNR	36.85	36.88	36.92	36.93	36.93	**24.62**	**2.22**	
		Bit rate	267.83	197.14	145.87	106.88	77.00			
	Reference (JM18.6)	PSNR	27.62	31.34	34.97	38.38	41.43			
		Bit rate	75.60	129.04	200.24	300.24	438.08			
Akiyo	Proposed	PSNR	39.51	39.44	39.44	39.38	39.35	**27.89**	**3.52**	
		Bit rate	57.54	81.57	115.75	159.88	220.91			
	Reference (JM18.6)	PSNR	29.03	32.56	35.77	39.44	42.92			
		Bit rate	61.28	106.80	166.56	256.72	377.60			
Average								**27.38**	**4.74**	

Table 2 Simulation Results of Intra Frames: Between Proposed Method and JinWang Method

Sequence CIF	Method	Parameter	Quantization parameter (QP)					Proposed Bit rate %	Jin Wang [8] Bit rate %	Remarks
			42	37	32	27	22			
Mobile	Proposed	PSNR	34.85	34.82	34.82	34.80	34.78	**38.81**	**3.49**	**Proposed method achieves higher Bit rate saving And PSNR is Consistent**
		Bit rate	577.78	798.74	1054.71	1326.11	1630.99			
	Reference (JM18.6)	PSNR	23.95	27.49	31.23	35.86	40.74			
		Bit rate	683.28	1181.12	1799.12	2622.72	3660.80			
Stefan	Proposed	PSNR	35.55	35.57	35.59	35.61	35.58	**29.34**	**3.12**	
		Bit rate	402.89	567.84	769.15	996.55	1279.12			
	Reference (JM18.6)	PSNR	25.75	29.40	32.97	37.23	41.57			
		Bit rate	404.00	732.64	1135.68	1711.68	2467.84			
Football	Proposed	PSNR	35.76	35.42	35.52	35.47	35.60	**23.27**	**4.05**	
		Bit rate	388.31	578.59	844.15	1146.37	1480.07			
	Reference (JM18.6)	PSNR	26.01	28.97	32.12	36.01	40.26			
		Bit rate	356.16	642.56	1065.60	1724.56	2601.76			
Bus	Proposed	PSNR	35.79	35.71	35.64	35.62	35.58	**31.10**	**3.56**	
		Bit rate	523.07	758.92	1038.49	1336.69	1680.92			
	Reference (JM18.6)	PSNR	24.26	27.52	31.19	35.64	40.38			
		Bit rate	505.92	937.76	1573.04	2457.76	3609.96			
Average								**30.63**	**3.55**	

downright prediction mode. Table 1 shows the simulation results of proposed algorithm for 6 QCIF Intra frames under Quantization Parameters 42 37 32 27 and 22, Gaussian values 0.1 with mode-4. From the Table 1 it notices that the proposed algorithm give high performance than JinWang algorithm and JM18.6. Average of bit rate saving is about 27.38% PSNR is consistent achieved in proposed method compared with JM 18.6 for QCIF sequence under different QP with Gaussian 0.1. From the Table 1 it notices that the proposed algorithm give high performance than JinWang algorithm and JM18.6. Average of bit rate saving is about 30.63% PSNR is consistent achieved in proposed method compared with JM 18.6 for CIF sequence under different QP with Gaussian 0.1. PSNR measured in dB and bit rate measured in kbps.

9 Conclusion

The 4×4 intra diagonal down right prediction mode for intra frame coding in H.264 using Gaussian pulse, proposed work achieved consistent picture quality and averagely 30.63% bit rate saving for CIF sequence. Also averagely 27.38% bit rate saving for QCIF sequence. Thus proposed method was compared with Jin Wang method and JM 18.6 reference method. Compared to previous method the proposed method maintains PSNR consistent manner and achieved bit rate saving.

References

1. Chaminda Sampath Kannangara, *Complexity Management of H.264/AVC Video Compression*, the Robert Gordon University 2006.
2. Thomas Wiegand, Gory. Sullivan, Seniour Member, IEEE, Gisle Bjontegaard and Ajay Luthra, "*Overview of the H.264/AVC Video Coding Standard",* IEEE Transactions on circuits and systems for video Technology, Vol. No 7 2003.
3. Jijun Shi, Yunhui Shi, Baocai Yi, " *Fast intra prediction mode decision for AVC based on edge direction detection"* 2008 The Institution of Engineering and technology.
4. K.Bharanitharam, An-Chao Tsai, "*Efficient Block Size Decision Algorithmm for intra mode decision in AVC encoder" 2009* IEEE International Symposium on Multimedia
5. Pengyu Liu, Kebin Jia, "A Novel Intra-frame Prediction Algorithm Based On Macro-block's Histogram for H.264/AVC*"* IEEE Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2010
6. Rein van den Boomgaard and Rik van der Weij, "*Gaussian Convolutions Numerical Approximations Based on Interpolation"*, Intelligent Sensory Information Systems, University of Amsterdam, and The Netherlands.
7. Pascal Gwosdek, Sven Grewenig1, Andr´es Bruhn, and Joachim Weickert, "*Theoretical Foundations of Gaussian Convolution by Extended Box Filtering*".
8. Jin Wang, Yunhi Shi, Wenpeng Ding, Baocai Yin, "A Low –rank Matrix Completion Based Intra Prediction for H.264/AVC" IEEE Multimedia and Intelligent Software Technology, 2011.

iKan—A Kannada Transliteration Tool for Assisted Linguistic Learning

Prakash Anusha, Ashok Meghana, R. Swathi, C. Sahana and S. Sowmya Kamath

Abstract Kannada is the official language of the state of Karnataka and is a prominent South Indian language. Despite being one among the forty most spoken languages of the world, the rich heritage of Kannada literature along with its rich culture has remained so far limited to those who know how to read Kannada. In order to bring this to the masses, so far, there has been minimal research work in automatic translation and transliteration of Kannada works. In this paper, we present iKAN, a platform that aims to automatically transliterate and later, translate Kannada script into English language, thus providing access to Kannada Literature. A phonetic mapping scheme was developed to enable read-ability and Kannada alphabets were represented in UTF-8 format, for transliteration into English. The Unicode approach used for the representation of characters resulted in high accuracy. We demonstrate the effectiveness of the proposed approach towards building an user-friendly transliterator for Kannada, that can be used as an e-learning tool.

Keywords Assisted linguistic learning · Natural language processing Automatic transliteration · E-Learning

P. Anusha (✉) · A. Meghana · R. Swathi · C. Sahana · S. Sowmya Kamath
Department of Information Technology, National Institute of Technology Karnataka, Surathkal, Mangalore, India
e-mail: anusha.blr@gmail.com

A. Meghana
e-mail: m6gnaa@gmail.com

R. Swathi
e-mail: swathirajanna94@gmail.com

C. Sahana
e-mail: csahana95@gmail.com

S. Sowmya Kamath
e-mail: sowmyakamath@ieee.org

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_27

1 Introduction

Kannada is the thirty-third most spoken language in the world. It is spoken by the Kannada people or Kannadigas (Kannaigaru), mainly in the state of Karnataka and by Kannadiga people settled in other states in India and in the world. It is a Southern Dravidian language, derived from the Kannada-Badaga, Tamil-Kannada forms [1] and there were about 38 million native Kannada speakers as of 2007.

The Kannada language has 49 phonemic alphabets which are divided into three groups namely swaragalu—from 'a' to 'au' (vowels—13 letters); vyanjanagalu—from 'ka'-'lla' (consonants—34 letters); and yogavaahak-agalu (neither vowel nor consonant—two letters: anusvara 'am' and anusvara visarga 'aha/ah') [1]. The character set is quite similar to other Indian languages. The script is a descendant of the Brahmi script, and is convoluted like all the other Indian languages, due to the occurrence of various combinations of \half-letters" (glyphs) (for example when we split the word 'Ram', we get 'ra' + 'm', where the 'm' is called a glyph or half 'ma' which is not followed by a vowel), or symbols that attach to various alphabets and add distinctive marks.

The Kannada script is phonetically faultless, barring the existence of the sound of a\half "n" (which becomes a half m) when uttered. In comparison to a number of other written symbols, even though Kannada has only 49 characters many more unique and distinct characters that can be formed by combining the base character set to form compound characters (ottakshara—for example when we split the word 'Pusthaka', we get 'pa' + 'u' + 's' + 'tha' + 'ka', where the 'sa', as glyph, is combined with 'tha' to form an ottakshara). Each written symbol in the Kannada script is in correspondence with one syllable, while in English we have one phoneme. The Kannada script is syllabic. Some of the letters in the Kannada Language have gone obsolete. One of the letters within Kannada language which is the unclassified vyanjana having the pronunciation of 'nh' or 'inn' is now extinct. It was mostly used in Kannada Literature that originated from the Tulunadu areas of the South Kanara district. At present, none of the mainstream works employ this consonant [1].

Transliteration is the process where we take in a source script from a particular language and then transcribe each and every word to a target script of any other language. These words can be content words or proper nouns, that may be of local or foreign origin [2–4]. Machine transliteration is incredibly valuable and highly demanded in the fields of automated translation as it has wide applications in multiple-language information extraction, polyglot text and speech processing. Hence, over the past few years, there has been extensive work in transliteration and several models of various kinds have been developed. Currently, there are four main kinds—the grapheme-based model, phoneme-based model, hybrid model, and the correspondence-based model. These models are distinguished based on the types of units to be transliterated, which can include graphemes, phonemes, both graphemes and phonemes, and the correlation between graphemes and phonemes [5].

Similarly, vowels can be of two types, the Schwa(for example ‘a’,‘e’) and the Non-Schwa(for example ‘aa’,‘ee’). Schwa is the short vowel that the last consonant inherits it by default. The Non-Schwa are represented using Maatras (for example, the word ‘haadu’ has ‘aa’ which forms a maatra while ‘mara’ is by default accompanied by ‘a’), which are the modi ed shape of the vowel when combined with consonants. Each vowel gives one maatra and these Maatras also each have an UTF-8 value. An Akshara(for example, ‘ka’—(glyph)‘k’ + ‘a’) can be de ned as an orthographic manner of representing speech sounds. Aksharas are syllabic, have generalized form and always end in a vowel, with white space as boundary. Most Indian languages uses the same punctuation marks (for e.g. a period, comma etc. as used in English), but there are a few exceptions in languages such as Devanagiri.

In summary, it is a fact that all Indian languages have the same phonetic base, thus for an Indian to learn a new Indian language, it wouldn’t be hard as he would already be familiar with all the sounds. These languages differ due to various possible phontactics, that is, the permissible combination of phonemes. A pivot for every syllable is chosen amongst the available consonant symbols. Generally, the consonants have an inherent vowel in addition to the base consonant, but vowel symbols are independent vowels with no consonants. Maatra, which is a dependent vowel, non-inherent in nature, are basically vowels with a distinctive mark. The Viraam (for example, in ‘Ram’, where ‘m’ is half letter is not followed by a vowel) is added when there’s a consonant without an inherent vowel [2].

In the work described in this paper, we attempt to build a full Kannada to English transliterator. The rest of the paper is divided into 4 sections. In Sect. 2 we present a discussion on the existing techniques used for transliteration. Section 3 outlines the details of the proposed transliteration approach and the corresponding experimental results are presented in Sect. 4, followed by conclusion and possible directions for future work in Sect. 5.

2 Related Work

Transliteration is the process of conversion of a string of text from one script to another. Unlike transcription, which maps phonemes of one language into a writing system, transliteration mainly deals with representing the characters precisely by mapping graphemes from one system of writing into graphemes of another. The accuracy of transliteration has a major impact on the performance of multilingual and cross-lingual text processing applications like Information Retrieval (IR) and Machine Translation (MT) [6, 7]. In this sense, Kannada transliteration is a form of machine transliteration model that aims to uphold the phonetic characteristics of its words as much as possible, by considering both the graphemes and phonemes [8].

Very less work exists in the area of Kannada language transliteration and translation. Some relevant related work includes researchers who worked with other regional Indian languages. Prahallad et al. [2] proposed a simple approach for building transliteration editors by considering every syllable of Indian language as

Consonant*Vowel. This paper proposes numerous schemes for transliteration like ITRANS to input the scripts of Indian language. The advantages of their method include lesser number of entries in the mapping table, automatic rendering of characters and easy adaptation for new languages. However, the disadvantage is that these scripts have low user readability.

Surana and Singh [4] mainly focused on the word origin applying different techniques that make use of the properties of the script as a result of which no training data is necessary on the target side, but more sophisticated algorithms and techniques have to be used on source side. Fuzzy string matching is used in place of using a training set. Their method deals with transliteration as a Cross Lingual Information retrieval (CLIR) problem which is mainly designed for conventional applications (CLIR, MT) also for applications like input-building mechanism.

Ganapathiraju et al. [3] described a Transliteration Scheme called Om for many Indian languages based on the fact that all Indian languages are phonetic in nature. They also developed a text editor, Microsoft WinWord that integrates the Om input for many Indian languages. The representation used here is parseable by many universal language processing tools like machine translation, information retrieval, and text summarisation etc., and also exploits common alphabets of various Indian languages. Their aim was to cater the large number of people who can speak but not read the native language. The Om tool uses a transliteration mapping scheme. Figure 1 shows the transliteration mapping for Kannada.

The proposed approach aims to develop a Kannada to English transliterator framework that will eventually also include a translator intended for non-native speakers and enthusiastic language learners. Our approach is based on a phonetic mapping scheme for Kannada alphabets represented in the UTF-8 format. The focus is on designing a user-friendly transliterator for Kannada, that can be used as an e-learning tool.

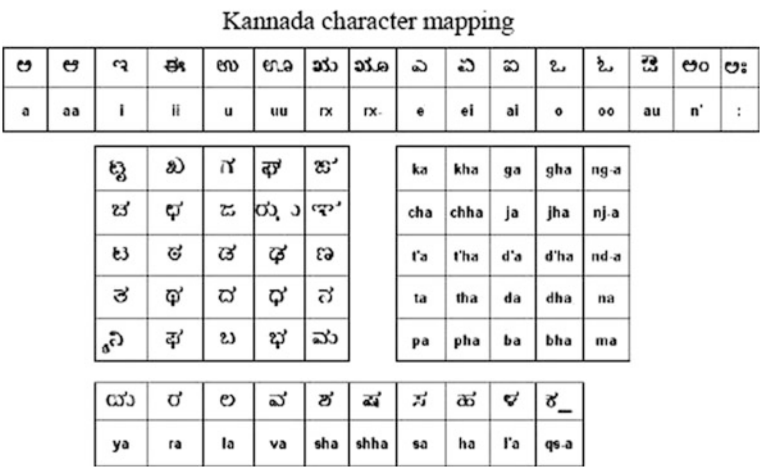


Fig. 1 Transliteration mapping for Kannada [3]

3 Proposed Transliteration Approach

Figure 2 depicts the overall system workflow of the proposed work. We have used the Unicode approach to convert Kannada words into English script and also English words written phonetically into Kannada script.

Considering, consonants as ‘C’ and vowels as ‘V’, the concept of C*V, where each syllable can have V, CV, CCV or CCCV can be considered [2]. Thus, to identify each syllable, the vowels are the anchor point, and every syllable has to end with a vowel. To represent Kannada script for these syllables, suppose we consider CV, the UTF-8 sequence of the syllable is obtained and we render its shape in

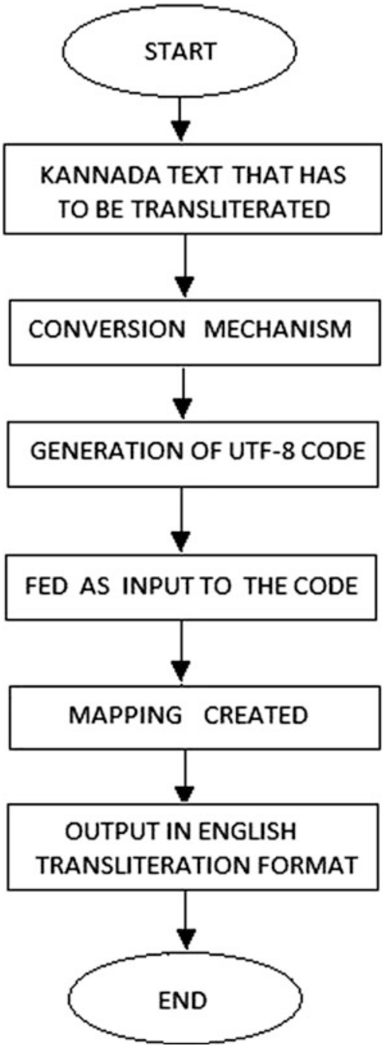


Fig. 2 Workflow of the proposed system

Unicode. The shape of an Akshara is of the form C*V and depends upon the composition of consonants, of which one is considered a pivotal symbol. In case of CCV, a Viraam/Halant($) is added between the two consonants, thus eventually rendering the shape for C$CV. Similarly, for CCCV, it is rendered for CCCV. The main processes that are carried out are as follows:

```
PUNCTUATION( c)
 do if c = "!"||","||"."
     return 1
 else
     return 0
MAP(i)
  display phonetic letters in English according to
  the mapping table

TRANSLITERATE(A , n)
for i<- 1 to n
     do if A[i] = Punctuation(A[i])
            print A[i]
     else if A[i] = virama
            continue
     else if A[i] = {ಅ to ಔ } or {ಕಾto ಕೌ}
            map(A[i])
     else if A[i] = ಅಂ,
            print "n"
      else if A[i] = ಅ:
            print "h"
      else
           map(A[i])
         if (A[i+1] = {ಕto ಱ} or {ಅ to ಔ}) or
Punctuation(A[i+1])
     print "a"
```

Fig. 3 Transliteration algorithm for mapping

1. Kannada text that has to be transliterated is identified.
2. Using the Unicode conversions, a UTF-8 form of the Kannada script is created.
3. The UTF-8 code is fed as an input to the system.
4. A phonetic mapping is done from the UTF-8 codes to the Transliterated output.
5. Finally, the transliterated output in English is produced.

In the beginning, the input Kannada words are identified and are split into sub-parts, that is, into the respective matras and the halants. Then, the sub-part elements are converted into the UTF-8 form. The newly represented UTF-8 equivalent is then mapped using the Kannada-English Transliteration dataset [9–13]. The dataset has three parts consonants, aksharas, gunithas and the two ayo-gayahas—'am', 'aha'.

A mapping algorithm shown in Fig. 3 is used to handle cases for the same and append default 'a' in case of no match, indicating no gunitha has been devised. The mapping algorithm ignores space, commas & punctuation and retains them as before. It also uses a map which consists of different forms of the Kannada alphabet —Swaras, Gunitha and consonants. Given a file containing the UTF-8 form of the Kannada script, the transliteration of the same will be made available. Finally, the output is obtained in English transliterated form.

4 Empirical Results

The proposed Kannada Transliteration System takes the UTF-8 Character Encoding of Unicode Format of the Kannada text as the input and produces the transliterated English output. Figure 4 shows the Kannada Input Text which is to be transliterated into English. It may consist of words, letters, sentences or paragraphs. Figure 5 shows the UTF-8 Input which is taken as input for transliteration. It is a variable-length encoding which uses 8-bit code units. Figure 6 shows the output of Transliteration, which is the nothing but the English form of the Kannada input.

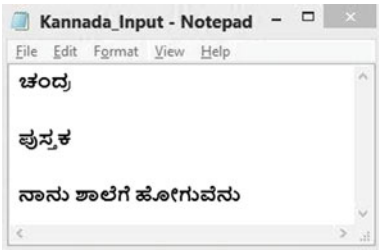


Fig. 4 Kannada input text

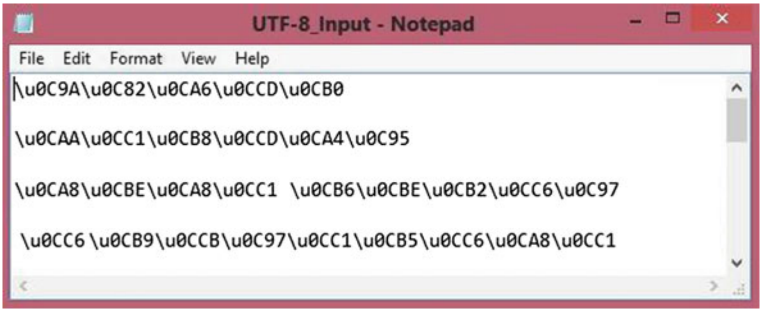


Fig. 5 UTF-8 input

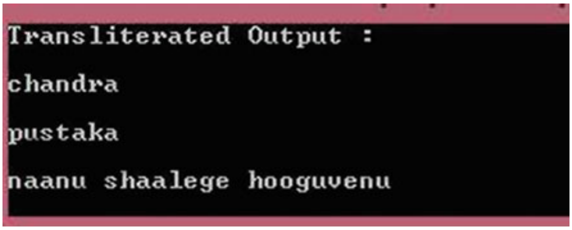


Fig. 6 Transliterated output in English

4.1 Case Study

To evaluate our Transliteration System on real language input, we tested by using the Karnataka state anthem, “Jai Bharatha Jananiya Tanujathe” as input. Currently a simple GUI has been provided which will later be developed further into a more user-friendly interface. Figure 7 shows current GUI which allows the user to either select the Kannada text input or enter Kannada text input, which is converted into UTF-8 input and transliterated output in English for the same is displayed back to the user. As can be seen in Figure 8, it is evident that the transliteration system is quite accurate. The Kannada content considered here as the input contains a wide variety of simple and complex words which are of the form C*V. The transliterated output generated is a proof to the fact that the phonetic nature of the Kannada language is preserved during the Transliteration Process.

The accuracy of the transliteration using the proposed approach was as high as 99%. One particular case where the system gave inaccurate results was when the input text included the usage of ‘anusvara’ (represented as ‘0’ or ‘um’/‘un’ phonetically) and ‘visarga’ (represented as ‘:’ or ‘aha’ phonetically) in the Kannada words. Since the English language does not contain these alphabets, expressing them in English is not possible. The phoneme of ‘anusvara’ is dependent on its contextual usage and may either be ‘am’ like in ‘pampa’ which has (0), or ‘an’ like in ‘chandra’ which also has (0). Likewise, ‘visarga’ also has conflicting phonemes, like ‘ah’ in ‘dhukha’ which has (:) and ‘aha’ in ‘namaha’ which also has (:).

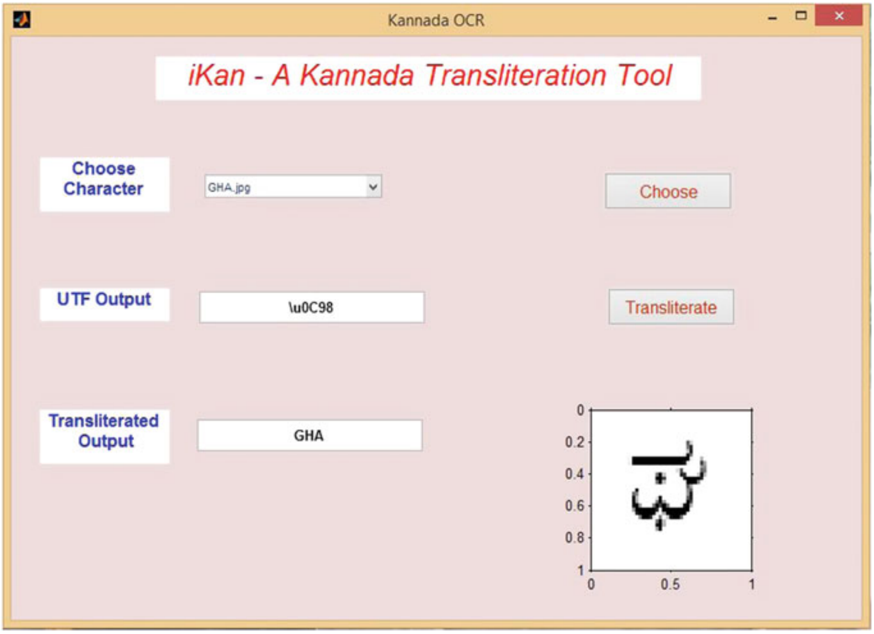


Fig. 7 GUI of the transliteration system

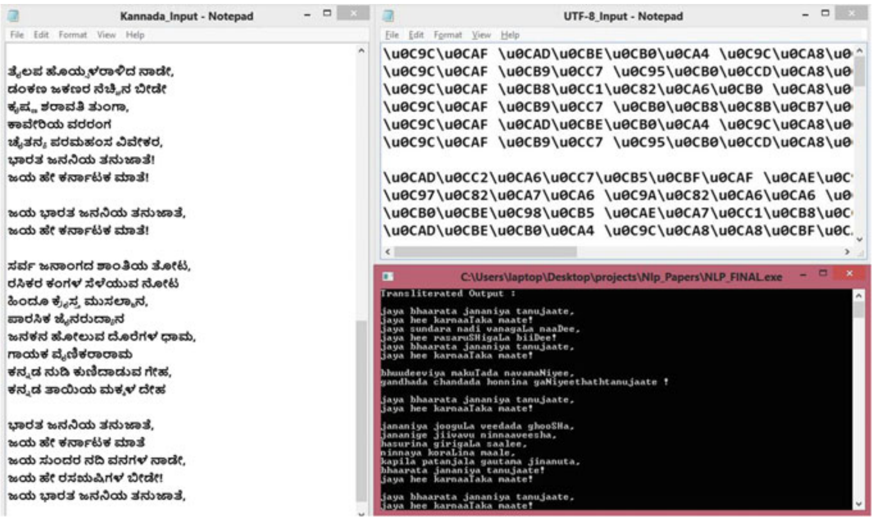


Fig. 8 The Karnataka state anthem transliterated

Since the occurrences of these different phonemes for the same letters are context-dependent, the system may sometimes fail to recognize the context and produce a different output.

5 Conclusion and Future Work

In this paper, a transliteration scheme from Kannada to English has been described. The advantage of the algorithm explained in this work, is that it has a very low running time and an almost 99% accuracy. This scheme can be extended to other Indian languages also thereby allowing transliteration from languages phonetically similar to Kannada to English. Using this framework, we intend to develop a user-friendly tool which is simple to use and also inexpensive. It would be of utmost use to people who can understand Kannada, but cannot read the script, and it can also be extended to a translator (Fig. 8).

References

1. Wikipedia: Kannada | Wikipedia, The Free Encyclopedia 2015
2. Lavanya, Prahallad, Prahallad Kishore, and Ganapa Thiraju Madhavi: “A simple approach for building transliteration editors for indian languages.” Journal of Zhe-jiang University Science A 6, no. 11 (2005): 1354–1361.
3. Ganapathiraju, Madhavi, Mini Balakrishnan, N. Balakrishnan, and Raj Reddy: “Om: One tool for many (Indian) languages.” Journal of Zhe-jiang University Science 6, no. 11 (2005): 1348.
4. Surana, Harshit, and Anil Kumar Singh: “A More Discerning and Adaptable Mul-tilingual Transliteration Mechanism for Indian Languages.” In IJCNLP, pp. 64–71. 2008.
5. Vijaya, M. S., V. P. Ajith, G. Shivapratap, and K. P. Soman: “English to tamil transliteration using weka.” International Journal of Recent Trends in Engineering 1, no. 1 (2009).
6. Shastri, G: “Kannada morphological analyser and generator using trie.” IJCSNS 11, no. 1 (2011): 112.
7. Antony, P. J., V. P. Ajith, and K. P. Soman: “Kernel method for English to Kannada transliteration.” In Recent Trends in Information, Telecommunication and Comput-ing (ITC), 2010 International Conference on, pp. 336–338. IEEE, 2010.
8. Sinha, R. M. K., and A. Jain: “AnglaHindi: an English to Hindi machine-aided translation system.” MT Summit IX, New Orleans, USA (2003): 494–497.
9. KanScan: http://www.kannadaocr.com/KanScan.html
10. Wikipedia: “Optical character recognition | Wikipedia, The Free Encyclopedia”
11. Vasavi, S and Varma, Ch and Santosh, DM and Sai Ram, S: Book Search by Capturing Text from Digital Images Using Optical Character Recognition. International Journal of Computer Science & Information Technologies,2,2014.
12. Ananthanarayana, V. S., D. K. Subramanian, and M. Narasimha Murty: “Scalable, distributed and dynamic mining of association rules.” In High Performance ComputingHiPC 2000, pp. 559–566. Springer Berlin Heidelberg, 2000.
13. Unicode Entity Codes for the Kannada Script Penn State University”, 2015

An Exploitation of Postulating Privacy in Utilization of Cloud Services with Crowdsourcing by Implementing Item Response Theory

C.J. Kavithapriya and S. Ancy

Abstract The proposed method gives a clear description about not only the crowdsourcing but also the implementation in cloud computing for effective usage and implementation of privacy. Item Response Theory (IRT) methodology is used in an implementation of Crowdsourcing in cloud services to resolve the sensitivity of items in the cloud and the users are made aware towards privacy. This methodology can be accomplished for both online or offline and also to be used in broadcasting to solvers of an unauthorized group as public accessing of solutions. In this proposed method, the essential services or the data are acquired by entreating contributions from an enormous public especially from the online community instead of traditional methods. Also, data analysis have been done for the cloud services by using the concept of crowdsourcing with the implementation of IRT (Item Response Theory) for the purpose of spreading the awareness of privacy among the people who wants to access data from the public cloud. As a future enhancement by using the IRT methodology, the data from the cloud can be analyzed accurately and provided with effective cryptosystem for providing privacy.

Keywords Cloud services · Crowdsourcing · IRT (Item response Theory) Data analysis · Data privacy

1 Introduction

Users of Mobile configures the mobile cloud in order to vend resources of the cloud and services such as data computing, collection of data and for the purpose of Consumers Overhauling—People Consuming services is entitled as Crowd

C.J. Kavithapriya (✉) · S. Ancy
Department of Information Technology, Jeppiaar Institue of Technology, Chennai 631604, India
e-mail: kavi11tech@gmail.com

S. Ancy
e-mail: sancysancy18@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_28

Sourcing via Mobile. The term Mobile Crowdsourcing is not the equivalent to long-established cloud computing which is being influenced by association with Internet. Rather, it belongs to a category among the service provided by the electronic commerce, pervasive cloud services that delivers the service through both online and offline workstations. The foremost transformation amongst the two classes of models are that all the Users of Mobile coupled to internet could be potentially remain as a service contributor in the Internet- centered mobile crowdsourcing, despite only the Clients of Mobile in the locality of mobile crowdsourcing can afford services of in non-internet centered that is local- based mobile crowdsourcing. To bring this scenario into world a specific architecture is followed.

The key components of mobile crowdsourcing are

1. Service Consumers-People Consuming Services
2. Users of Mobile
3. Servers in Centralized Mode
4. Local Servers

Service Consumers-People Consuming Services Service Consumers-People Consuming Services refer to users at both electronic and native that necessitates the services of cloud across the system of mobile crowdsourcing by outsourcing responsibilities to Mobile Clients.

Users of Mobile Users of Mobile with mobile devices can unconventionally forms a mobile cloud. They can provide cloud services, for both online Service Consumers-People Consuming Services as well as local Service Consumers-People Consuming Services. They provide service WiFi/cellular networks and by using techniques called NFC/ Bluetooth the local servers are communicated by Consumers of local service and People who Consume Services [1]. To perform this task the mobile user requires the mobile mobile cloud computing or should adopt local computing and take part in outsourcing the task.

Servers in Centralized Mode For the people consuming the services from the Internet-based servers, the perfect platform for mobile crowdsourcing is laid by Servers in Centralized Mode. In order to perform outsourcing of task and evaluation of service, the information of the cloud sourcing like records of services in historical order and profile of the various users is stored by centralized mode severs. Also, on accountability of the Servers in Centralized mode, the additional activities such as publishing the task, allocation of task [9], collection of report and processing the feedback.

Local Servers To the mobile clients in same location and for people consuming services, the aggregated task result, task which has been outsourced, Services of local crowdsourcing and the broadcasting of task are afforded by Local Servers. Each Local Server is provided with dedicated gateways of local mobile. The collection of user report and distribution of information regarding the task to the ne It is used to disseminate the task information to adjacent users of mobile is done by gateways of local mobile. To support crowdsourcing of mobile, these gateways perform querying and updating of necessary information from centralized mode servers. Nevertheless, the mobile clients do not trust the local servers and hence it is local servers are usually implemented for commercial purpose.

1.1 Applications of Mobile Crowdsourcing

Since five years, the mobile devices have transformed as sensors and "hub of information" in our day-to-day Life. The mobile computing and crowdsourcing together paves a way for extremely proficient and computing data effectively by cutting down the cost. Mobile crowdcomputing and mobile crowdsensing are the main mobile crowdsourcing applications [2].

1.1.1 Mobile Crowdcomputing

Outsourcing of data computation of mobile users is done by mobile crowdcomputing. The Mobile User's participation in the tasks which are outsourced can locally execute these tasks or unlade them to the servers of cloud depending on the resources used for computation and its own data. Due to human intrusion, mobile crowdcomputing can influence human-intelligence to have an impact with the tasks which are additionally suitable for evaluation by human than the computations of machine like entity resolution, image annotation, and sentiment analysis. The local-based mobile crowdcomputing application is Honeybee [4], which performs the outsourcing of tasks-detection of face and photography. The American Mechanical Trunk Platform, is the platform in which the crowdsourcing the computing tasks of Crowd DB [3] is performed by querying and answering. The mobile devices are used to take specific photos and execute face detection algorithm along with the user's personal evaluation.

1.1.2 Mobile Crowdsensing

Substantial economical resources along with massive efforts on technical side are required for collecting data, processing of data like sensing the environment and monitoring the data. Mobile crowdsensing usually outsource collection, processing data to the mobile users. These users of mobile are responsible for sensing data with mobile devices that are furnished with sensors. Additionally, the users of mobile process and execute the data by the methods such as mobile cloud computing or the local computing. Cloud services for collection and processing of data can be made economically efficient by provoking huge number of mobile users to participate in mobile crowdsensing. Another application of local-based mobile crowdsensing that resourcefully detect current signals of traffic and collectively transfer the detected signal information through ad hoc network is entitled "SignalGuru" [5]. These exchanged information of traffic signals can be used by the Smartphones in order to forecast and plan the traffic signals in future, guiding the drivers for making decisions while driving the vehicle. Besides the above-described applications, Medusa [6] is one in which the specific sensing data is collected by making the tasks of sensing being outsourced which includes documentation of video, auditing the video and monitoring the road. This information collected by Medusa is communicated through Secure HTTP based Wireless Communication to all the users of mobile who has been connected to the internet.

1.2 Key Challenges in Mobile

1.2.1 Egocentrism and Incentive

Forming a mobile cloud which is powerful is precarious as making the mobile users to participate in the process of mobile crowdsourcing. For the purpose of motivating the mobile users to make them involve in the crowdsourcing process includes the enticements such as Contribution by Personal, Financial Incentives, and Societal Gain. In order to afford the former mentioned incentives, the new-fangled mechanisms had been proposed by Yang et al. [7] in which it comprises reward-sharing, auction-based mechanisms. Based on the working time of each tasks performed by people consuming services are rewarded with fixed rewards and these rewards are shared by each individual participants in reward-sharing incentives. They also design a truthful auction-based incentive mechanism, where users of mobile make offers for different outsourced tasks and Service Consumers-People Consuming Services choose appropriate participants to maximize their own utilities. Both the reward-sharing and auction-based incentive mechanisms can economically stimulate the formation of a mobile cloud and also can achieve mutual benefits.

1.2.2 Task Allocation

The explicit set of tasks is allotted to the users of mobile those who can perform the processing of crowdsourcing with the outcome of proficient and precise results. A conscription algorithm in mobile crowdsourcing application by taking the agility alleyways of the mobile users has been proposed by Reddy et al. [8]. Also the algorithm proposed by Reddy et al. takes the privileges of temporal and geographic availabilities of the users of mobile which influences the delay in task in a elevated way which in turn to be deliberated for the purpose of selection of participation of the mobile users. For an illustration, if the task which has been outsourced, "Find an unoccupied basketball court at the University of Waterloo," the clients/users of mobile that are in concern about "Sport" and people those who pursue study at this University are given the highest priority for being allocated to the tasks which has to be outsourced. The basic and decisive criteria for partaking in outsourcing of task in mobile crowdsourcing entail the social influences to be examined and hence finalize the degree for each matching pair of the task. This is in order to maximize the prospective efficacy of the services provided for both the mobile users and also the services consumed by the people.

1.2.3 Security Threats

For the services provided by the cloud and for the people consuming the services provided, Security is contemplated as an important issue. However, the concept of mobile crowdsourcing has an origin from postulation that the mobile users must provide the precise results. Even though, mobile crowdsourcing excel as emerging practice in the technology field, it also involves the insistent drawback such as intrusion by the malicious users of mobile to distress the process of outsourcing of tasks. The following transgressions are performed by the malicious users such as i) fabricating the computation, ii) intuiting the results, iii) interrupt the tasks that are currently in process, iv) launching attacks in either direct or indirect mode which in turn leads to undesirable effect in the tasks that are outsourced. For providing solutions for the previously mentioned drawbacks, Zhang et al. [10] proposed an approach of estimating robust trajectory for mitigating the undesirable influences of aberrant trajectories of users and also for categorizing the normal and unusual users. Besides the above-mentioned solutions provided, there existed another solution for solving the problems of crowdsourcing, in which the computation of reputation score performed by the devices and also the credibility of the subsidized data evaluated. In the process of trust evaluation [12], not only the advantages exist but also the disadvantages such as disclosure of privacy by conjoining the multiple task reports correlated with trust values also exists. For developing a consistent crowdsourcing system, the attacks categorized as malicious has to be prevented and high-level security as A major challenge is to design sophisticated security measures to be designed.

1.2.4 PRIVACY LEAKAGE

The concept of privacy leakage deals with the revealing of the personal details about the services consumed by the people. Information of the mobile users such as the residing place of the user, individual activities and social affiliations are divulged through the spatio- temporal information which is tagged to the tasks that are being outsourced. To avoid the disclosure of the vital information about the mobile users and their outsourced tasks, a solution to be provided by using certain privacy deterrence solutions. As an illustration for the privacy leakage, in a small space of an environment in which when the users participate in outsourcing of tasks, their details about the location can be disclosed. In order to defense against the privacy issues in crowdsourcing, cryptography algorithms can be used and can prevent the snooping of data by the intruders during transmission and manipulation of data. For classifying the tasks and to ensure with the accuracy in classification without revealing the privacy, a collaborative learning method had been proposed by Liu et al. [13]. The method proposed by Liu et al. classifies the tasks as activity or the context recognition and also make use of the techniques such as regression, feature perturbation to ensure the accuracy in the classification of the tasks. Besides the above-explained methods, there also exist other methods such as anonymity for providing the effective solution for privacy leakage issues. Since the local servers are installed usually for only commercial uses and not considered as the trustworthy server mode, hence it is recommended to contribute additional attentiveness to the mobile crowdsourcing which is deployed in the local-based server mode.

1.3 Solutions for Mobile Crowdsourcing Applications

In this section, some details about the social-aware task allocation scheme [9] is to be focused to deal with the challenges such as the incentives and task allocation as mentioned in the above sections with detailed description. Also, to tone down the privacy and security breaches for the local-based mobile crowdsensing related applications, a reputation system has to be envisioned.

1.3.1 Social-Aware Task Allocation for Internet-Based Mobile Crowdcomputing

For the computation of the outsourced tasks, the users of the mobile have the capability to afford the services from cloud in the mode of the Internet- based crowdcomputing. In order to achieve the computing services as targeted and planned, the important criteria to be chosen are the task allocation and the incentives as discussed in the above sections. Generally the crowdcomputing hinge on the technology of human-intelligence and the social attributes—specialties, social activities of the mobile users. For a conventional application based on internet of the mobile a scheme called social-aware task allocation (SATA) [14] has been discussed in this section with the set of procedures.

(i) A set of tasks has been published by a consumer to perform outsource with the requirements of delay, task budget and specialties.
(ii) The task execution information- processing cost, delay, then being executed by the mobile users whoever is interested and of capable of outsourcing the task given by the customers and then request for the participation.
(iii) Based on the information of the application, the tasks have been allocated to the applicants of subset.

(iv) Execution of tasks that are outsourced with their individual data and with resources of computation are executed by the participants those who have been selected.
(v) Reports are submitted to the Services Consuming People after the task execution are completed.
(vi) Evaluation of task report done and feedback id provided to the Services Consuming People.
(vii) Auction-based mechanism for providing incentives has been adopted by the SATA scheme to inspire the mobile users to achieve benefits and perform the outsourcing task in efficient way.

2 Challenges and Elucidations of Crowdsourcing

The basic methodology used in crowdsourcing is as an initial step, the crowd has to submit the solutions which is later owned by broadcaster and at times remunerated with prizes. On using the crowdsourcing services, a huge amount of solutions and information for an amount of smaller cost from different variety of people all over the world.

2.1 Issues of Crowdsourcing

(a) The product of services being outsourced.
(b) The level of collaboration.
(c) Managerial Control Systems.

2.1.1 Service or the Product Being Outsource

Crowdsourcing not only covers the low end complexity such as simple form of data coding, data de-duplication, also the vital role played as the ratings or the evaluation for the users in the community on many sites. Nowadays, the cloud sourcing also covers more complicated human work of intelligence which also .includes the business, areas of research and development. Hence the crowdsourcing is called as “diversifying” mechanisms and provides ways to form new markets, then to override the market status of the traditional business by handling the communities in online.

2.1.2 Collaborations in Crowdsourcing

Related to the introduction of the data present in the community in order to depict the solutions or answers.

2.1.3 Managerial Level of Control Systems

In managerial level, crowdsourcing does not consider about development of skill on the potential workers. But also its approach is to uncover and leverage the skills and the potential strength of the workers.

3 Implementation of IRT in Data Analysis

The concept of Item response theory [15] was first proposed for the purpose of accessing the data and increasing the ability to access the data in secured manner. Initially it was implemented in the psychometric field and used most widely in the field of education in order to calibrate, evaluate the set of items for providing scores based on its abilities.

3.1 Key Concepts of IRT

3.1.1 Item Response Function

For the description of item response theory, the key concept of IRT, which is called as Item Response Function (1) or Item Characteristic Curve which plots the responses of participants to the test items with dichotomously score? Each test respondent influences a certain level of the ability measured by the test. The score on test is computed based on the ability of respondents for initial accessing. The probability $P(\theta)$ of that the respondents endorse the item is estimated using the trait level given as θ . The curve plots are plotted with the x-axis consisting of the trait levels and the y-axis with the probability of the respondent. As shown in the Fig. 1 it represents the logistic response model with the parameters a and b, where a defines the of discriminating power of an item and b is the parameter for the location of item.

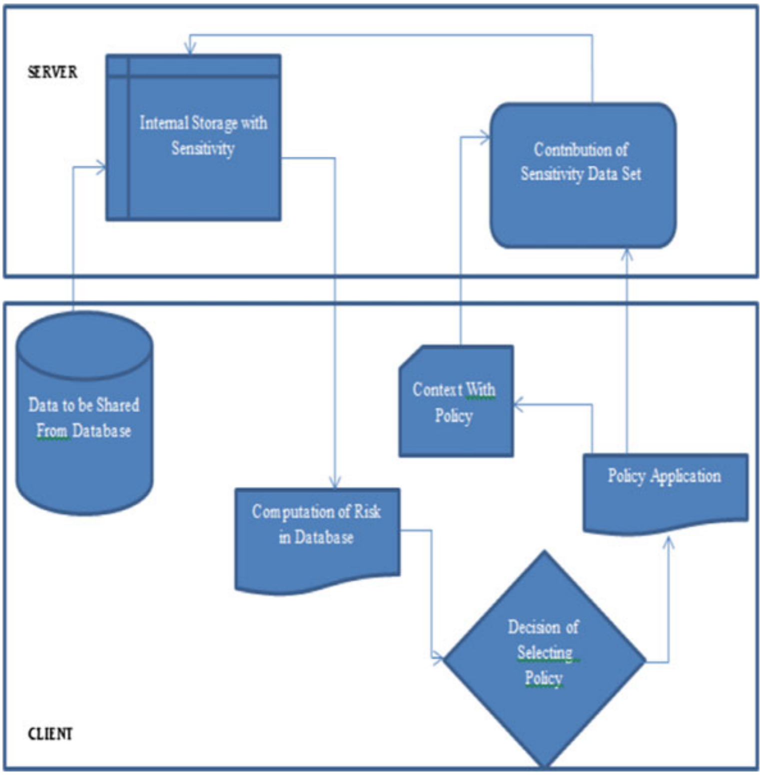


Fig. 1 Representation of client server framework of implementation of privacy in IRT

3.1.2 Computation of Characteristic Curves

The probability parameter for identification of the characteristic difference of the item is computed as

$$P(\theta) = 1/(1 + e^{-1}) \tag{1}$$

Where the description for the above variables are given below.

l − negative of the exponent a $(\theta - b)$
θ – level of ability
e − constant 2.718

This item response function accepts that the probability of the responses that are accurate is depend on the function of b and θ which in turn evaluates the items that are discriminatively equal for testing the fundamental trait levels.

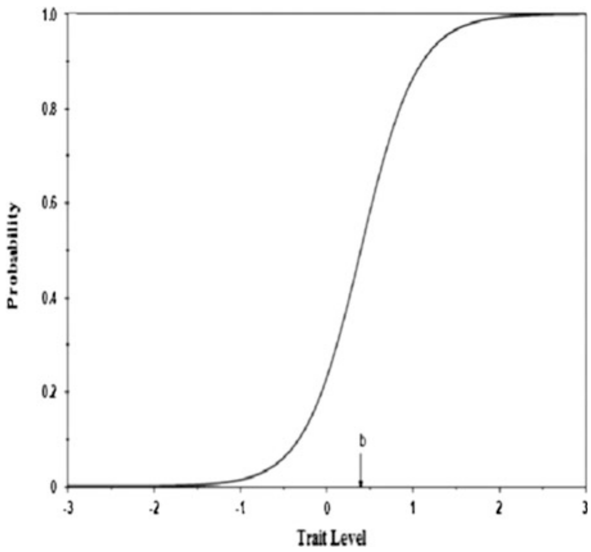


Fig. 2 Representation of Characteristic curve in IRT with the x-axis values interpreting values of the probability of the precise response from different participants

4 Implementation of Privacy Framework

The interactions between the entities end users and cloud service providers develop the framework for implementing the privacy in IRT. The roles of end user consist of data observer [11] or the data sharer in which the cloud provider can be only the data observer. The user sends the data to the single cloud service provider which acts like the repository for maintaining this user's data. As depicted in the Fig. 2, the modules of the framework are explained below.

4.1 Context Vocabulary

The context vocabulary in the framework describes the item's content feature. Considered as metadata and for sharing operation, it is taken as an environment for achieving this operation. This context vocabulary has the details about the framework for the privacy implementation and to prevent the leakage of the information about the mobile users through the intruders.

4.2 Sharing the Policy

The policy in the range of [0, 1] with different types of policies for each different items that are shared by the people in the large community. The value 0 in policy range implies that it is full transparency while 1 represents the full obscurity.

4.3 Crowdsourcing

Foe each of the process involved in the sharing operation the type of the policy applied and the context of an item are grouped together in cloud through the mechanism for preserving privacy.

4.4 Evaluation of Risk

Based on the analysis and processing of the information which is crowdsourced will eventually lead the people to know about the privacy implementation using the IRT.

4.5 Policy Recommendation

the final module in the framework recommends the different types of policies which guides the end users to share and access the information via privacy.

5 Determination of Risk Evaluation in Crowdsourcing

In the Fig. 2, the process involved in the determination of risk evaluation in crowdsourcing is depicted. Initially, the client query the server regarding the sensitivity of the operation involved in sharing of the data. The server in turn provides the different types of contexts and policies for each sharing operation [14].

5.1 Querying for Privacy Aware

The context sent by the server is a part of the Query Set which are termed as homonyms. Considering the contexts Cx and Cy in the subsets which if suppose requires the context before sending to the user to be at least two anonymity subsets. Later, the contexts CI(Cx) and CI(Cy) both will be sent to the end user and also all the other ku-2 number of contexts are appeared in the other subsets where no more possible items are being the real target context (Fig. 3; [16]).

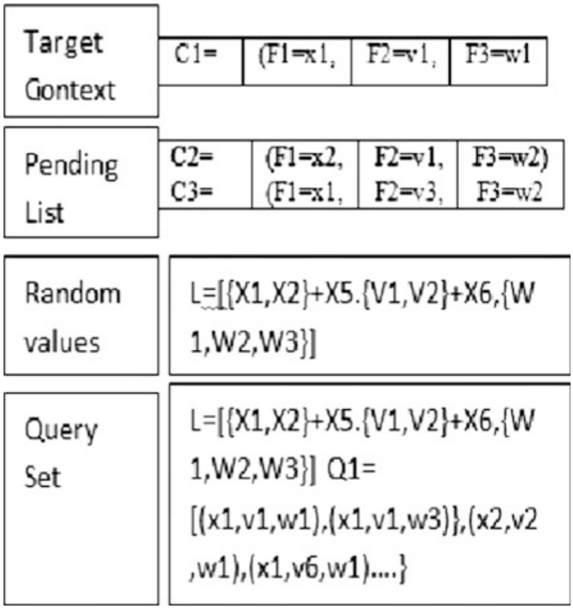


Fig. 3 This figure explains the query set formation which gives the detailed description about the target context, details about the pending list, random values of the tasks that are outsourced and the set of query

6 Computation of Sensitivity and Risk Evaluation

The two features of IRT compared to the classical test theory are the (a) calculation of group invariance for item's parameters, (b) calculation of invariance for trait levels. On focusing on the one-dimensional IRT models, the three assumptions are made as (a) single trait determining the person's response, (b) local dependence, i.e. no association between the different items, (c) reproducing the observed responses by the estimated item and parameters of person. In the crowdsourcing, by receiving the crowdsourcing information (CI) from the client, the cluster S is matched with the server and computes the sensitivity which requires the value of S(sensitivity) to be minimum of 15 contexts along with the associated policies and contexts. The sensitivities are evaluated by using the technique of Marginal Maximum Likelihood Estimation which is considered as the maximization expectation algorithm to estimate parameters of different types of models. This algorithm relies on having the adequate responses by the different people for different types of items. Also, for better understanding of the sensitivities of the different types of the items, the hit rate ratio (2) is also determined as number of queried items with available sensitivity to the total number of queried items as depicted in Fig. 4.

$$\mathrm{H(HitRate)} = \frac{\text{No. of Queried Items with Avaialable Sensitivity}}{\text{Total No. of Queried Items}} \tag{2}$$

The target context for the evaluation of sensitivity should be appeared in the k types of different subsets of anonymity subsets. All the subsets with the different

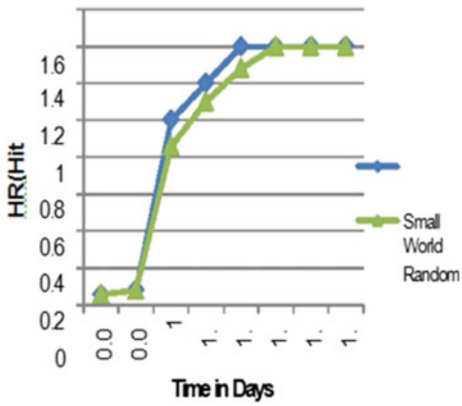


Fig. 4 Figure represents the graph plot of the effect of changing social network with sensitivity and evaluation of hit rate with respect to the time in days

values in the target subset to be available in the query set which is depicted in the Fig. 3, the representation of the query data set format.

7 Conclusions

The proposed method deals with the privacy related risk evaluation in the field of the social work. The sensitivity computation done with the techniques of using the maximization algorithm with the estimation of hit rate. Also, it concentrates on the Item Response Function (1) for the evaluation of the responses from the different people in network with dichotomously score. The proposed method deals with the basic concepts about the crowdsourcing concepts which includes the challenges in the mobile crowdsourcing, the solutions provided by different philosophers. Also describes about the security threats which is involved in the crowdcomputing and how to protect the information of the mobile users provided during the execution and processing of the outsourcing of the tasks which are allocated to the capable users by the service consuming people. For avoiding the security breaches, the innovative scheme entitled as SATA has been given as the reference in this method and also provides the framework for the implementation of the privacy with the certain criteria such as context vocabulary, sharing the policy, crowdsourcing, evaluation of risk and policy recommendation. Another approach for determining the response of the items which is given by the Item Response Theory for the risk evaluation and for plotting the graph between effect of changing social network with sensitivity and evaluation of hit rate with respect to the time in days. As a future enhancement, the sensitivity evaluation using the hit rate with respect to the time can be done for various other features in of the Item Responses from the participants in the community.

References

1. D. Huang, T. Xing, and H. Wu, "Mobile Cloud Computing Service Models: A User-Centric Approach." *IEEE Network*, vol. 27, no. 5, 2013, pp. 6–11.
2. A. Faggiani *et al.*, "Smartphone-Based Crowdsourcing for Network Monitoring: Opportunities, Challenges, and a Case Study," *IEEE Commun. Mag.*, vol. 52, no. 1, 2014, pp. 106–13.
3. J. Franklin *et al.*, "Crowddb: Answering Queries with Crowdsourcing," *Proc. ACM SIGMOD*, 2011, pp. 61-72.
4. N. Fernando, W. Loke, and W. Rahayu, "Honeybee: A Programming Framework for Mobile Crowd Computing," *Mobile and Ubiq. Syst.: Comp.,Net., and Services*, 2013, pp. 224–36.
5. E. Koukoumidis, S. Peh, and M. R. Martonosi, "Signalguru: Leveraging Mobile Phones for Collaborative Traffic Signal Schedule Advisory," *Proc. ACM MobiSys*, 2011, pp. 127–40.
6. M. Ra *et al.*, "Medusa: A Programming Framework for Crowd-Sensing Applications," *Proc. ACM MobiSys*, 2012, pp. 337–50.
7. D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing," *Proc. ACM Mobicom*, 2012, pp. 173–84.
8. S. Reddy, D. Estrin, and M. Srivastava, "Recruitment Framework for Participatory Sensing Data Collections," *Pervasive Comp.*, 2010, pp. 138–55.
9. S. He *et al.*, "Toward Optimal Allocation of Location Dependent Tasks in Crowd sensing," *Proc. IEEE INFOCOM*, 2014, to appear.
10. X. Zhang *et al.*, "Robust Trajectory Estimation for Crowdsourcing-based Mobile Applications," *IEEE Trans. Parallel Distr. Sys.*, vol. 25, no. 7, 2013, pp. 1876–55.
11. K. Huang, S. Kanhere, and W. Hu, "Are You Contributing Trustworthy Data?: The Case for a Reputation System in Participatory Sensing," *Proc. ACM MSWiM*, 2010, pp. 14–22.
12. X. Wang *et al.*, "Enabling Reputation and Trust in Privacy- Preserving Mobile Sensing," *IEEE Trans. Mobile Comp.*, 2014, to appear.
13. B. Liu *et al.*, "Cloud-Enabled Privacy-Preserving Collaborative Learning for Mobile Sensing," *Proc. ACM Sen- Sys*, 2012, pp. 57– 70.
14. J. Ren *et al.*, "SACRM: Social Aware Crowdsourcing with Reputation Management in Mobile Sensing," *Computer Commun.*, 2015, http://dx.doi.org/10.1016/j.comcom.2015.01. 022
15. Thorpe, Geoffrey L., and Andrej Favia. "Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications." (2012).
16. Harkous, Hamza, Rameez Rahman, and Karl Aberer. "C3P: Context-Aware Crowdsourced Cloud Privacy." *Privacy Enhancing Technologies*. Springer International Publishing, 2014.

Music Information Retrieval: A Window into the Needs and Challenges

Dave Jayati, Pandya Dhara, Fruitwala Pranav and Tarjni Vyas

Abstract The field of Music Information Retrieval is still very immature although it has been addressed by a substantial number of research projects over past three decades. Many problems are faced in music information retrieval due to the nature of music itself. In this paper, the basic representation of audio and video are discussed along with the challenges of music retrieval. The need for music information is discussed using the statistical survey. The paper challenges the common assumption that searching on the pitch alone is satisfactory and discusses the other features which can aid in searching. This assumption may indeed be right for monophonic music but it is insufficient for polyphonic music. Even in monophonic music the possibility of faulty results can't be ignored.

Keywords Information retrieval · MIR · Music search strategies
Monophonic features

1 Introduction

The single related judgments known by us is very limited which shows that evaluation of music information retrieval systems according to standard Cranfield model in text information retrieval world is not possible and an alternative to Cranfield approach for music has not even been proposed yet. Same as text, some kind of indexing is required for music also but the techniques required are very

D. Jayati · P. Dhara · F. Pranav · T. Vyas (✉)
Institue of Technology, Nirma University, S G Highway, Ahmedabad 382481, India
e-mail: tarjni.vyas@nirmauni.ac.in

D. Jayati
e-mail: 15mcen06@nirmauni.ac.in

P. Dhara
e-mail: 15mcen14@nirmauni.ac.in

F. Pranav
e-mail: 14mcen06@nirmauni.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_29

different. It will not be too much if we say that music IR is decades behind the text IR because that is the reality.

Nearly all music-based research which is known to us is concerned with mainstream Western music. It is the music which is not tonal and derived from a particular tradition but which is primarily based on pitch. In this paper, that bias is maintained. Thus, music for ensembles of percussion instruments (not pitch), electronic music and microtonal music have been excluded. Music IR involves the very substantial amount of music and of information science. Thus, it is cross-disciplinary. A significant amount of music perception and cognition is also involved in Music IR. Although MIR is a rather young field, and the problems of MIR are challenging [1], there are already commercial applications of MIR systems [2].

2 The Reasons for "Music IR Is Hard"

In a few words, music recognition is the mathematical analysis of an audio signal (usually in WAV format) and its conversion into musical notation (usually in MIDI format) [3]. Segmentation and Units of Meaning: There are many ways to formulate the queries and it is possible that the users are not aware of each and every way to formulate the queries. So to conflate different but synonym words are very important in the context of the user's information need. If the system does not do this the recall will suffer. Thus, conflating units of meaning normally words is a basic requirement of text IR. The conflation should be done judiciously so that the precision does not suffer. Current efforts at studying MIR system usability issues focus on user behavior exhibited in specific MIR systems [4].

3 Needs for Music Information

The emergence of audio and video data types in databases will require new information retrieval methods adapted to the specific characteristics and needs of these data types [5]. To the users, the importance of descriptive metadata and the extra-musical information is huge which can be considered as the first finding. Extra-musical information and descriptive metadata have very important commercial aspects for users which can be considered as the second finding.

In given Fig. 1, three categories are shown which are "Artist information" (74.6%), "Lyrics" (81.0%) and "Title of work" (90%) which is either extra-musical information or metadata. By studying the various data such as the t 60.7% positive responses to "Price of item", positive response rate of 67.2% to "Learn about item before purchase" and the 47.1% positive expression for "Review/rating" information, The "Artist information" numbers with the "Background information" (39.1%), the importance of the commercial aspect can be understood.

Music information \ Response	Positive %	Negative %	Don't know %
Title of work(s)	90.1	7.4	2.5
Lyrics	81.0	15.4	3.6
Artist information	74.6	23.7	1.7
Sample tracks for listening	67.4	27.3	5.3
Track listing	60.7	33.8	5.5
Price of item	51.7	41.5	6.8
Information on genre	49.1	46.3	4.6
Review/Rating by other people	47.1	47.3	5.6
Influences among artists	42.6	52.6	4.8
Background information (history, theory, etc.)	39.1	55.4	5.6
Information on different version(s) of work(s)	37.3	55.7	7.0
Artwork/Album cover	30.8	62.8	6.5
Links to related websites	29.7	62.2	8.0
Released date	21.5	71.2	7.3
Record label	15.0	77.9	7.0

Fig. 1 The responses for the possibilities of the users are for seeking the following music information? [6]

Reason \ Response	Positive: Frequency (times per month) ≤ 1 %	2–4 %	≥ 5 %	Positive Total %	Never Total %
Listen for entertainment	18.0	33.4	43.1	94.5	5.5
Build collection	28.5	39.7	20.9	89.1	10.9
Verify or identify work, artist, lyrics, etc.	30.9	31.1	11.9	73.9	26.1
Learn about artists	34.4	27.8	8.3	70.5	29.4
Learn about item before purchase	32.9	26.4	7.9	67.2	32.7
Listen for work or study purposes	15.7	21.7	22.1	59.5	40.5
Learn about music	31.8	16.0	6.7	54.5	45.5
Use for special occasions (wedding, presentation, etc.)	27.3	11.9	1.4	40.6	59.4
Learn about instrument(s)	23.0	10.5	4.0	37.5	62.4
Perform with a musical instrument	18.2	9.1	5.5	32.8	67.2
Karaoke/Sing for entertainment	16.2	8.5	7.2	31.9	68.2
Use for gadgets (ringtone, computer sound effect, etc.)	19.5	9.1	1.9	30.5	69.6
Play at certain places (café, etc.)	15.5	7.9	2.6	26.0	74.0
Use in teaching/ instruction	12.6	3.8	1.1	17.5	82.5
Academic research	8.6	3.8	1.6	14.0	86.0
Sing professionally	4.5	2.4	1.7	8.6	91.4

Fig. 2 The responses for the reasons of seeking music information [6]

It also suggests that the information is deliberately explored by the users for enhancing the experience of the music.

Music Information is sought by users as an ordinary experience which can be considered as the third finding. Music Information is sought by users for assisting the building of a collection of music which is the 4th finding and 5th Finding is Music Information is sought by users to verify or identify the artists, lyrics or work.

A huge percentage (73.9%) of the people who search for the music information to gain enough information for verifying or identifying lyrics, work, artists etc. for which obtaining actual material and item is not so appropriate strategy as to obtain "name the tune" which is shown in data. More musical information plays an important role in the music practical knowledge of users by enriching it which is suggested by the data (Fig. 2).

4 Audio and Music Representations

Music notation is a rich source of representational problems [7]. Three representations of audio and music are: what are the extremes of maximum and minimum structures at which the music notation lies and in the middle, the less-well-known time-stamped events from. The basic representation of music is as shown in Figs. 3 and 4. On each representation, numerous variations exist.

On stereo vs. mono, type of material and for audio like MP3 in which audio is compressed so that perceptually unimportant features are removed, a huge variation is possible. Enhancing the structure is tougher than reducing the structure with

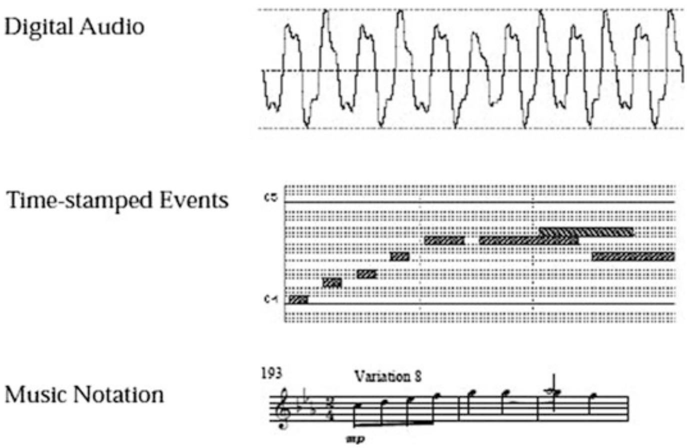


Fig. 3 Representations of music (schematic) [8]

Representation	*Audio*	*Time-stamped Events*	*Music Notation*
Common examples	CD, MP3 file	Standard MIDI File	sheet music
Unit	sample	event	note, clef, lyric, etc.
Explicit structure	none	little (partial voicing information)	much (complete voicing information)
Avg. rel. storage	2000	1	10
Convert to left	-	easy	OK job: easy
Convert to right	1 note/time: pretty easy; 2 notes/time: hard; other: very hard	OK job: fairly hard	-
Ideal for	music bird/animal sounds sound effects speech	music	music

Fig. 4 Basic representations of music [8]

Explicit structure	minimum	medium	maximum
Music representation (and examples)	Audio (CD, MP3)	Events (Standard MIDI File)	Music Notation (sheet music)
Text representation (and examples)	Audio (speech)	ordinary text	text with markup (HTML)

Fig. 5 Music vs. text [8]

reasonable quality. Comparing music and text is often very helpful as varying amount of explicit structure is there in the text as shown in Fig. 5.

5 The Four Parameters for Notes and Music Retrieval

"Recent studies in musical perception suggest that durational values may outweigh pitch values in facilitating melodic recognition." This statement has been stated by Selfridge-Field, on the other hand, some researchers states that pitch c is the most important feature. The pitch of a melody is not the only important feature which is evident.

What should be the weight of each four parameters information relatively in a given style of music is one obvious question. In mainstream music, the figures might be timbre and dynamics 15%, pitch 60%, rhythm 25%. In both the dimensions vertical and horizontal, pitch occurs. Patterns resulting from the meter are also involved in the rhythm so rhythm is not just a string of duration.

5.1 Pitch Matching and Realistic Databases

Some of the other information needed for large database and complex music other than melodic pitch is as explained here.

5.2 Salience

In music, salience depends on parameters like loudness and the texture thickness. A fraction of melodies is indistinguishable in a large group like an orchestra in total effect which can lead to something that seems very correct "matches" for the given queries.

Fig. 6 T = trill, RN = repeated notes, RE = rest, G = grace notes [8]

5.3 Confounds

Matching of contours or matching or actual profiles of successive pitch intervals is done by music IR systems which are based on melodic pitch (Fig. 6).

5.4 Cross-Voice Matching

To assume that in polyphonic music search for matching one voice at a given point of time is tempting to assume but in many cases, it will not work. Matching over voices will still be necessary despite the complete availability of voice information. It is possible to find multiple matches in cross-voice matching having different rhythms if the only pitch is considered and it will be terrible for precision as well.

5.5 Polyphonic Queries

Queries [9] will generally be polyphonic in searches through examples which almost all users does. For musically trained users, a more specialized case is applied. The users desire to find instances of chords. These type of queries are polyphonic inherent in nature which requires more other parameters than pitch.

So, melodic pitch alone only is not sufficient enough for complex music having a large database. Even melodic and Harmonic pitch together fails sometimes when searching in a large database. So, when the duration patterns and pitch both are matched, the results can be improved. Moreover, the thickness of texture and loudness both affect the salience and they should be used for adjusting the search results.

6 Monophonic Feature Selection

Monophonic sources [10] also known as melodies are used for most of the work in music information retrieval systems. In these systems, new notes begin only after the previous note gets finished sounding.

6.1 Absolute Measures Vs. Relative

Reduction of the notes into single dimension is the fundamental approach of the monophonic feature. Usually, pitch and duration are used in music retrieval systems. The dissimilarity between duration and pitch is that the invariance of the pitch is obtained through interval rather than through proportion. There are three standard expressions related to relative pitch: (1) Accurate interval between two adjacent pitches. (2) The rough shape which shows the sign and groups the value into some equivalence classes. (3) Simple contour which keeps the signs and discards the magnitudes.

6.2 Unigram Vs. N-Grams

For n-gram extraction, the approach of sliding window is used. The sequence of the node which has window size is n will be converted to a sequence of relative unigrams. Detecting the patterns which are repeating is another approach to extract n-gram. Here we assume that the frequency has a bigger role in the similarity of music for the second approach. Weights are expressed in terms of the relationship between the ratios of duration, pitch intervals. According to the weights, computation of evaluation takes place and automatic decisions are taken about where to put the boundary markers. Local maxima are used in it. N-gram window will be the nodes in a sequence which resides between the markers.

6.3 Shallow Structure

In the text information retrieval system, we have identified words as adjectives, verbs, nouns. Unlike texts, music does not have speech parts. In shallow structural feature, extractor is a technique which examines the pitches and probabilistic best fit into a key which is known, is done by the technique. Similar shallow structural techniques can be defined for duration and pitch both for complex music patterns.

7 Polyphonic Feature Selection

7.1 Monophonic Reduction

A monophonic data [11] can be deduced from a polyphonic data through selecting one note at every time stamp. The most commonly used monophonic sequence which is used to extract is the melody or theme.

7.1.1 Short Sequences

By constructing the small monophonic sequences from the polyphonic sources the monophonic reduction takes place. It is not useful when the collection or the database of the music increases. Searching for some known monophonic strings from the polyphonic database can be done by many retrieval algorithms. The selection of monophonic patterns can be done by using some clues such as repetition by some feature extraction algorithms.

7.1.2 Long Sequences

The short monophonic sequence uses the repetition and evolution for feature extraction but it may not be trustworthy. One approach is extracting the highest pitch note. In another approach, the lowest pitch is extracted and entropy measures, average pitch, music information are used by some other approaches. In some techniques, the polyphonic data is split into many numbers of sequences which are searched independently and the result is combined but this technique is not useful for overlapping notes.

7.2 Homophonic Reduction

In this technique, at a given time step, every note is selected instead of selecting at most one note at a time. Thus, reduction of the polyphonic source into homophonic source can be done by assuming the independence between the overlapping notes.

7.3 Monorhythmic and Homorhythmic Reduction

In a monophonic sequence of notes, there is an inherent symmetry. A note can be spat into its pitch and duration. Extraction of a sequence of pitch intervals is easier but polyphonic music, it is different. Naturally, the duration of that note and pitch of an individual note do not have equal weight. Whenever we are trying to extract the monophonic melody from a polyphonic source, a note with the highest pitch should be used. Whenever we are getting the homophonic source from a polyphonic source, the pitch will be comprised to the homophonic slice. There should be a problem at the reduction of monophonic and homophonic music and the reason behind this problem is that notes or sets of notes are extracted at various time steps.

8 Conclusion

In this paper, we have discussed the need and the problems with music information retrieval. It has been discussed that why music information retrieval is harder than text information retrieval. The four parameters for music notes which can aid in searching apart from pitch are discussed for the accuracy and high precision for music information retrieval. Different features of monophonic music and polyphonic music are explained which vary in nature along with different approaches and techniques used for retrieving polyphonic and monophonic music queries.

References

1. D. Byrd and T. Crawford. Problems of music information retrieval in the real world. Information Processing and Management, 38:249–272, 2002.
2. Type, Rainer, Frans Wiering, and Remco C. Veltkamp. "A Survey of Music Information Retrieval Systems." ISMIR. 2005.
3. AKoff Sound Labs (2000). What is Music Recognition? Retrieved January 31, 2001, from the World Wide Web: http://www.akoff.com/about.html
4. Cunningham, S.J., Reeves, N., and Britland, M. "An Ethnographic Study of Music Information Seeking: Implications for the Design of a Music Digital Library", Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital Libraries, Houston, USA, 2003.
5. A. Ghias, J. Logan, D. Chamberlin, and B. Smith. Query by humming—musical information retrieval in an audio database. In Proceedings of ACM International Multimedia Conference (ACMMM), pages 231–236, San Francisco, CA, 1995.
6. Lee, Jin Ha, and J. Stephen Downie. "Survey Of Music Information Needs, Uses, and Seeking Behaviours: Preliminary Findings." ISMIR. Vol. 2004. 2004.
7. R. B. Dannenberg. A brief survey of music representation issues, techniques, and systems. Computer Music Journal, 17(3):20–30, 1993.
8. Byrd, Donald, and Tim Crawford. "Problems of music information retrieval in the real world." Information processing & management 38.2 (2002): 249–272.
9. Bainbridge, D., Cunningham, S.J., and Downie, J.S. "How people describe their music information needs: A grounded theory analysis of music queries", Proceedings of the International Symposium on Music Information Retrieval, Baltimore, USA, 2003.
10. Pickens, Jeremy. "A survey of feature selection techniques for music information retrieval." Proceedings of the 2nd International Symposium on Music Information Retrieval (ISMIR). Vol. 124. 2001.
11. J. Pickens. A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. In Proceedings of the 1st International Symposium for Music Information Retrieval (ISMIR), October 2000.

Cuckoo Search-Based View Selection

T.V. Vijay Kumar, Amit Kumar and Biri Arun

Abstract A data warehouse is a central store of all entities, concepts, metadata and historical information created for doing data validation, complex mining, analysis and prediction in many organizations. A data warehouse is designed primarily as a tool for answering analytical queries, which are intricate and exploratory and have higher response times when answered using a data warehouse. Many real enterprise information integration systems compute and maintain materialized views or cache results. Unlike virtual views, materialized views store data. Since all views cannot be materialized due to storage space constraints, an appropriate subset of views needs to be selected for materialization. The selection of such a subset is an NP-complete problem. Swarm intelligence algorithms have been extensively used to resolve such problems. In this paper, the cuckoo search (*CS*) algorithm has been adapted and discretized to solve the view selection problem. Based on this, a *CS*-based view selection algorithm (*CSVSA*) has been proposed. Also, experiments were performed to ascertain appropriate parameter values for which *CSVSA* is able to select reasonably good quality *Top-K*.

Keywords Data warehouse · Decision-making · Analytical queries Materialized view selection · Swarm intelligence · Cuckoo search algorithm

T.V. Vijay Kumar (✉) · A. Kumar · B. Arun
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: tvvijaykumar@hotmail.com

A. Kumar
e-mail: er.kumaramit2009@gmail.com

B. Arun
e-mail: biriarun@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_30

1 Introduction

In the present digital age, information has become the key asset for conducting business operations, especially as an aid in business decision-making [23, 27]. With the availability of huge volumes of digital enterprise data, decision-making has become an intricate and increasingly difficult task. Decision support systems (*DSS*) were designed to help managers in making better and faster decisions. The purpose of *DSS* was to analyse accumulated business data in order to obtain information concerning the current and future business trends, and to find the correlations existing between the various business elements, etc. [27]. The effective value of the accumulated business data is most crucial for the synthesis of any reliable information. Different business processes/elements collect and use business data in a loosely coupled or independent manner; because of which data becomes inconsistent. Inconsistent data is very difficult to analyse and the information obtained using it could be misleading [15]. With a view to provide historical, subject-specific, time-variant, non-volatile, integrated and credible enterprise data for analytical purposes, the data warehouse was developed. Data from many disparate operational databases of organizations are extracted, transformed and loaded into the organization's data warehouse; which is the central repository of enterprise data [15, 17]. A data warehouse is accessed by *DSS* to extract information using online analytical processing (*OLAP*) tools [20]. On account of the rapidly growing volumes of data warehouses and the complexity of *OLAP* queries, the query response times are substantially large; with the processing even taking hours, days and months to provide the needed information [5, 11]. Such information becomes stale and a major impediment in effective decision-making. In order to decrease the *OLAP* query response time, materialized views have been popularly used [13] and the same is the focus of this paper. To facilitate data analysis, a data warehouse stores data using star schema. Star schema consists of a central fact table with many dimension tables; they are linked through a primary–foreign key relationship. A fact table consists of measures of business events like sales, profit, etc. Dimension tables describe measures of the fact table like the date of sale, the region of sale, etc. A view is formed by joining a set of dimension tables as specified by an *OLAP* query. Join operations on massive dimension tables are computationally very expensive and result in large query response times [4, 7]. Views, being virtual, need to be recomputed based on the queries posed; such an approach in the *DSS* environment is not efficient [4, 7]. Gains in terms of computations can be achieved if the views, based on the data accessed by queries that are frequently posed, can be stored in the data warehouse; such an approach would eliminate the computational cost related to recomputing a previously computed view. These precomputed views, referred to as materialized views, are smaller in size compared to the base tables and thus capable of answering queries in a shorter response time [13]. For a fact table with n dimension tables, there are 2^n possible views; therefore, due to the limited disk space and high maintenance cost, it would be impractical to think of materializing all possible views. A better approach would be to selectively determine a set of views whose materialization satisfies the disk space constraint, the maintenance cost constraint and the response time constraint [6, 21]. However, identifying/selecting such sets of views, referred to as view selection, is an *NP*-complete problem [13].

View selection refers to selecting appropriate set of views from the database schema that conform to the resource constraints [1, 6]. Views can be selected empirically [30, 31, 33] or based on heuristics like greedy [13, 28, 32, 34–38], randomized [39, 41, 42], evolutionary [40, 42, 43, 49], etc. Several techniques exist in literature that addresses the view selection problem. [13] considered the selection of views in a lattice framework and proposed a greedy algorithm using a benefit measure to select sets of views for materialization. It has an exponential run time complexity with regard to the number of dimension tables and fails to scale up for higher dimensions. [10] proposed an *AND-OR* view graph framework to represent views and their relationships; also a polynomial time algorithm was proposed to select views from it. In order to achieve computational gain, [47] proposed a *MVPP* framework to identify and materialize overlap information for a query work load. To overcome the runtime problem of *HRUA*, [25] proposed the *PBS* algorithm. [14] proposed the Genetic Local Search method (*GLS*) to select views while under space constraint. [50] used *GA* in conjunction with the *MVPP* framework to select materialized views. [1] automated the view and the index selection. [24] proposed the SimpleLocal, the SimpleGlobal and the ComplexGlobal algorithms, to address the view selection problem for multi-cube schemas. [19] applied *GA*, using the *OR*-view graph, to the maintenance cost view selection problem. [49] addressed this problem using the Evolutionary Algorithm (*EA*). [21] proposed *PGA*, which first nominates potential views and then selects beneficial views from amongst them. [18] proposed the Niched Pareto *GA* (*NPGA*) and the multi-objective *GA* (*MOGA*) to select views using the lattice framework. [9] proposed simulated annealing to select views using the *MVPP* framework. [26] proposed Particle Swarm Optimization (*PSO*) to select views using the *AND-OR* view graph. [45, 46] proposed the greedy view selection algorithms that were capable of selecting views for higher dimensional data sets. Algorithms to select frequently used views along with their sizes were proposed by [12, 34–38]. Iterative improvement, simulated annealing, two-phase optimization, genetic algorithm, memetic algorithm and differential evolution with lattice framework were used to select views in [39–44]. The mating and foraging behaviour of honeybees have also inspired view selection algorithms [2, 3, 29].This paper focuses on the use of cuckoo search (*CS*), which is a swarm intelligence technique [16], to answer the view selection problem. Accordingly, a *CS* based view selection algorithm (*CSVSA*) has been proposed. Further, experiments have been performed to ascertain the appropriate parameter values resulting in the selection of reasonably good quality *Top-K* views using *CSVSA*.

The paper is organized as follows: The Cuckoo Search Algorithm (*CSA*) is discussed in Sect. 2. Section 3 discusses the cuckoo search based view selection algorithm (*CSVSA*). Experimental results are discussed in Sect. 4. Section 5 is the conclusion.

2 CSA

To common people cuckoos are birds that are popular for their melodious voice. Scientists have studied them and have observed that cuckoos have a very unique breeding behaviour. It has been observed that some species of cuckoos lay their eggs in other (host) birds' nests. Different species of cuckoos have different strategies to lay their eggs in such host bird's nest. Cuckoos have quick egg laying ability. Some species of cuckoos wait for the host birds to lay their eggs in nests and thereafter, upon finding some opportune instance, when the host birds are away from their nests, the cuckoos surreptitiously and quickly lay their eggs in the host birds' nest amidst host bird's eggs [8, 22]. Some species of cuckoo forcefully intrude and chase away the incubating host bird to quickly lay their eggs [8, 22]. In other species of cuckoo, the male cuckoo lures away the host bird so that the female cuckoo can lay her eggs in the host bird's nest [8, 22, 48]. The shape and colour of the eggs of the cuckoos resemble the host bird's eggs so much that even the host bird fails to distinguish her eggs from the cuckoo's eggs [8, 22, 48]. At times when the host bird finds the cuckoo's egg, it removes them from its nest; many other host birds leave their nests to make new nests for themselves at other locations. The host birds incubate their eggs along with cuckoo's eggs. Since the cuckoo's eggs hatch faster than the host bird's eggs, their chicks instinctively remove the eggs of the host bird [8, 22, 48]; such a strategy increases the survivability of cuckoo chicks [8, 22, 48]. Many studies have shown that cuckoo chicks can mimic the colour and call of the host bird's chicks. As a consequence the host bird feeds cuckoo chicks along with her own chicks. Cuckoo chicks grow quicker than the host bird's chicks and eventually fly away to join other cuckoos [8, 22, 48].

The cuckoo search algorithm (*CSA*) [48] is based on the breeding behaviour of cuckoos. In order to algorithmically adapt the breeding behaviour of cuckoos, the following assumptions have been made. A cuckoo can lay only one egg at a time and can dump it in any randomly chosen host bird's nest. Nests with good quality eggs will be carried forward for future generations. The number of host birds' nests is constant. $P_\alpha \in [0,1]$ is the probability of a cuckoo egg being found by the host bird [48]. When the host birds find the cuckoo's eggs, they may either eject them or leave them in order to build a new nest. An egg in a host nest represents a candidate

Fig. 1 *CSA* [48]

Begin
Step 1: Generate initial population of n host nests x_i (i = 1, 2, ..., n)
Step 2: Initialize iteration I = 1;
Step 3: Generate a cuckoo egg randomly by Lévy flights
Step 4: Evaluate the fitness of the cuckoo egg
Step 5: Compare the above generated cuckoo egg with a randomly generated cuckoo egg from the population.
If the former is better, then replace the latter with the former in the population
Step 6: Abandon a fraction (p_α) of worse nests and generate new nests
Step 7: Identify and keep the best nests.
Step 8: Increment I by 1
Step 9: If I is less than pre-defined number of iterations, Go To Step 3
Step 10: Return the best nest
End

solution of the problem under consideration. A cuckoo egg represents a new candidate solution. *CSA* is given in Fig. 1.

CSA begins by first generating a random population of *n* host nests. For nest $x_i^{(t)}$ selected from the population, produce a new cuckoo egg, $x_i^{(t+1)}$ by applying Lévy flight to $x_i^{(t)}$ as shown below:

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus Levy(\lambda)$$

where λ is the step size and is greater than zero; in many cases $\lambda = 1$. The product $\oplus$ is entry wise multiplication. Lévy flight is a kind of random walk whose step or flight lengths are taken from a levy distribution which has characteristics of power law.

$$\text{Lévy} \sim u = t^{-\lambda},\ (1 < \lambda \leq 3)$$

The new cuckoo egg is compared with a randomly chosen nest from the population of nests. If the new cuckoo egg is fitter than the randomly chosen nest, then the latter gets replaced with the former and whereupon, P_α fraction of worse nests from the population are abandoned and an equal number of new nests are randomly built in their place. On the other hand, the good nests of the population are retained. Then ranking of the nests is done to identify and memorize the current best nest. The above steps are performed for a predefined number of iterations whereafter the best nest is produced as the output.

The above-mentioned *CSA* is adapted suitably and discretized to solve the view selection problem discussed in Sect. 1. The population of the nests considered in *CSA* would correspond to the population of the *Top-K* views, i.e. each cuckoo egg would represent the *Top-K* views in proposed algorithm *CSVSA*. The fitness of each cuckoo egg was determined based on its proximity with the host bird eggs. Here, the fitness of each set of *Top-K* views would be determined based on its *TVEC*. Lesser the *TVEC*, fitter would be these *Top-K* views. Further, the Lévy flight is accordingly discretized. *CSVSA* is discussed next.

3 CSVSA

CSVSA, which selects the *Top-K* views from a multidimensional lattice, is given in Fig. 2. *CSVSA* takes a multidimensional lattice of views *L*, the size of each view in the lattice and the maximum number of iterations I_M as input and produces the *Top-K* views *TKV* as output.

CSVSA first generates a population P_{TKV} of *Top-K* views. Next, the following steps are performed for a maximum number of iterations I_M:

A set of *Top-K* views, say TKV_i, is chosen randomly from P_{TKV}. Application of Lévy flight on TKV_i leads to generation of TKV_i'. The original *CSA* [48] proposed a Lévy flight operator, as given below:

Input: Multidimensional Lattice of views L with size of each view, maximum number of iterations I_M
Output: *Top-K* views *TKV*
Method:
BEGIN
Generate initial population P_{TKV} of n *Top-K* views TKV_i (i = 1, 2, ..., n)
WHILE (Number of Iterations $t < I_M$)
Apply Lévy flight to a randomly picked *Top-K* views TKV_i from P_{TKV} to generate TKV_i'
Compute *TVEC* of TKV_i' as given below:

$$TVEC(TKV_i') = \sum_{i=1 \wedge SM_{V_i}=1}^{N} Size(V_i) + \sum_{i=1 \wedge SM_{V_i}=0}^{N} SizeSMA(V_i)$$

where N is the number of views in a lattice, SMV_i is the Status Materialized of V_i (SMV_i=1, if V_i is materialized and SMV_i=0, if V_i is not materialized), Size(V_i) is the size of view V_i, SizeSMA(V_i) is the size of the smallest materialized ancestor of view V_i
Choose a *Top-K* views, say TKV_j, randomly from P_{TKV} (say, j)
IF ($TVEC(TKVi') < TVEC(TKV_j)$)
replace TKV_j by TKV_i' in P_{TKV};
END IF
A fraction (p_α) of worse *Top-K* views are removed and new *Top-K* views are generated by applying L´evy flights on them;
Update *TKV* with best *Top-K* views amongst *TKV* and *Top-K* views in P_{TKV}
END WHILE
RETURN *TKV*
END

Fig. 2 Algorithm *CSVSA*

$$x_i^{t+1} = x_i^t \oplus \alpha.t^{-\gamma}$$

where x_i^{t+1} represents the new solution generated around x_i^t, which is the original solution. $\oplus$ denotes the entrywise multiplication. α denotes the step size, which is taken as 1 for most problems. In *CSVSA*, the value of variable *b,* which follows the Lévy distribution, is computed. *b* represents the total number of views that are required to be changed in the *Top-K* views in order to create a new set of *Top-K* views. The value of variable *b* can be computed, as given below:

$$b = \alpha.t^{-\gamma}$$

Unlike in the original *CSA*, where α is taken as 1, in *CSVSA*, α is taken as *K*. The variable *t* corresponds to the iteration number and γ ($1 < \gamma \leq 3$) is a constant. In the beginning, the Lévy flight operator would result in large numbers. Subsequently, it would drop exponentially with increase in the number of iterations. Next, the *TVEC* of TKV_i' is computed using the following formula [39–44]:

$$TVEC(TKV_i') = \sum_{i=1 \wedge SM_{V_i}=1}^{N} Size(V_i) + \sum_{i=1 \wedge SM_{V_i}=0}^{N} SizeSMA(V_i)$$

where N is the number of views in a lattice, SM_{Vi} is the Status Materialized of V_i (SM_{Vi} = 1, if V_i is materialized and SM_{Vi} = 0, if V_i is not materialized), $Size(V_i)$ is the size of view V_i, $SizeSMA(V_i)$ is the size of the smallest materialized ancestor of view V_i. This is followed by comparing the *TVEC* of TKV_i' and the randomly picked

Top-K views, say TKV_j, from P_{TKV}. In case, *TVEC* of TKV_i' is less than the *TVEC* of TKV_j, TKV_i' replaces TKV_j in P_{TKV}. Next, p_α proportion of worst *Top-K* views in P_{TKV} are removed and replaced by the new *Top-K* views generated by applying the Lévy flight on these worst views. *TKV* is then updated with the best *Top-K* views from amongst *TKV* and *Top-K* views in P_{TKV}. After I_M number of iterations, the *Top-K* views *TKV* is produced as the output.

Next, the appropriate value of (p_α, γ) for which *CSVSA* selects the *Top-K* views, with minimum *TVEC*, is ascertained based on experimentation. The experimental results are discussed next.

4 Experimental Results

The proposed algorithm *CSVSA* was implemented using MATLAB-R2010a. To determine the suitable value of (p_α, γ) for which *CSVSA* is able to select the *Top-10* views with the minimum *TVEC*, over 400 iterations, graphs were plotted for 5, 6, 7, 8, 9 and 10 dimension data sets and these are shown in Figs. 3, 4, 5, 6, 7 and 8 respectively. The values of p_α and γ considered are (0.25, 0.5) [48], and (1.5, 3) [48] respectively. In all six graphs, *CSVSA* is able to select the *Top-10* views at a comparatively lower *TVEC* for p_α = 0.5 and γ = 1.5. This shows that p_α = 0.5 and

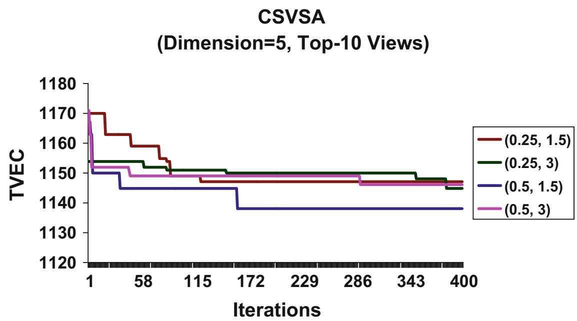


Fig. 3 *CSVSA-TVEC* Vs. iterations, dimension = 5, *Top-10* views

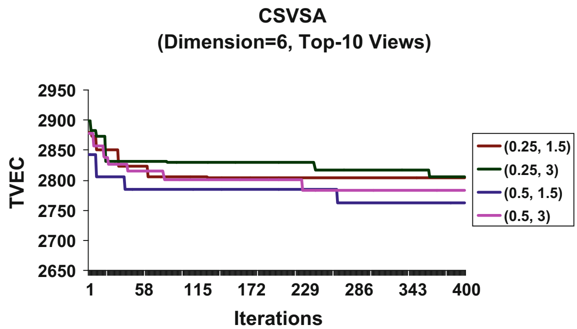


Fig. 4 *CSVSA-TVEC* Vs. iterations, dimension = 6, *Top-10* views

Fig. 5 *CSVSA-TVEC* Vs. iterations, dimension = 7, *Top-10* views

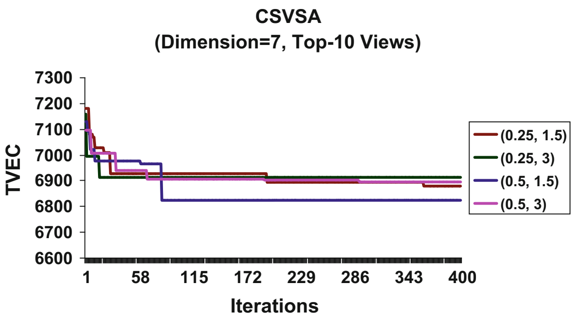


Fig. 6 *CSVSA-TVEC* Vs. iterations, dimension = 8, *Top-10* views

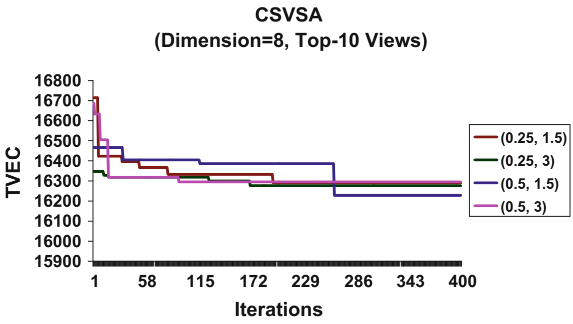


Fig. 7 *CSVSA-TVEC* Vs. iterations, dimension = 9, *Top-10* views

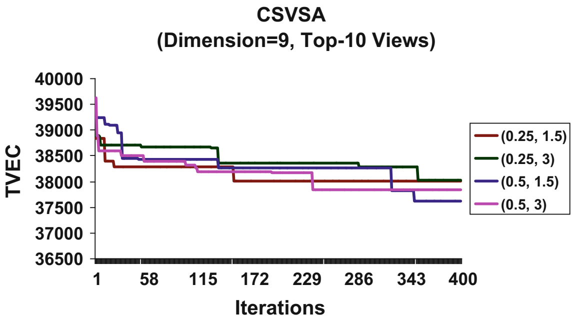
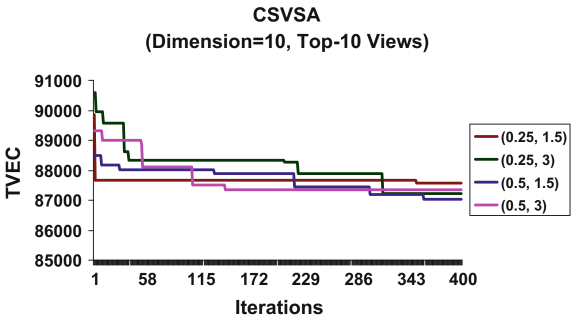


Fig. 8 *CSVSA-TVEC* Vs. iterations, dimension = 10, *Top-10* views



γ = 1.5 are the most suitable values for selecting the *Top-K* views from a multidimensional lattice.

5 Conclusion

One of the key issues concerning materialized views, i.e. view selection, is the major focus of this paper. The purpose of view selection is to select an appropriate set of views that reduces the query response time. For higher dimensional data sets, selection of the *Top-K* views is shown to be an *NP*-complete problem. Since swarm intelligence techniques have been extensively used to answer such problems, a view selection algorithm, i.e. *CSVSA*, inspired by cuckoo search, has been proposed in this paper. *CSVSA* is capable of selecting *Top-K* views for higher dimensional data sets. Further, experiments were carried out to determine suitable values of (p_α, γ) for which *CSVSA* selects the *Top-10* views, at a minimum *TVEC* for 5, 6, 7, 8, 9 and 10 dimension data sets. The experiments show that *CSVSA*, for all these dimension data sets, is able to select *Top-10* views at a minimum *TVEC* for the observed value of p_α = 0.5 and γ = 1.5. These observed values are useful and would provide better insights for generating the *Top-K* views from a multidimensional lattice.

References

1. Agrawal, S., Chaudhuri, S. and Narasayya, V. R.: Automated Selection of Materialized Views and Indexes in SQL Databases. In VLDB, Vol. 2000, pp. 496–505, 2000
2. Arun, B. and Vijay Kumar, T.V.: Materialized View Selection using Marriage in Honey Bees Optimization. International Journal of Natural Computing Research, Volume 5, Number 3, pp. 1–25, 2015
3. Arun, B. and Vijay Kumar, T.V.: Materialized View Selection using Improvement Based Bee Colony Optimization, International Journal of Software Science and Computational Intelligence, Volume 7, Number 4, pp. 35–61, 2015
4. Chan, G. K., Li, Q. and Feng, L.: Design and selection of materialized views in a data warehousing environment: a case study. In Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, pp. 42–47, 1999
5. Chaudhuri, S. and Dayal, U.: An Overview of Data Warehousing and OLAP Technology. SIGMOD Record 26(1), pp. 65–74, 1997
6. Chirkova R., Halevy A. Y., and Suciu D.: A Formal Perspective on the View Selection Problem, In Proceedings of VLDB, pp 59–68, 2001
7. Chirkova, R., and Yang, J.: Materialized Views, Foundations and Trends in Databases. Vol. 4, no. 4, pp. 295–405, 2011
8. Davies, N.B.: Cuckoo adaptations: trickery and tuning. Journal of Zoology 284: pp. 1–14, 2011
9. Derakhshan, R., Dehne, F. K., Korn, O., & Stantic, B.: Simulated Annealing for Materialized View Selection in Data Warehousing Environment,. DBA'06 Proceedings of the 24th IASTED international conference on Database and applications, pp. 89–94, 2006
10. Gupta, H.: Selection of views to materialize in a data warehouse. In Database Theory—ICDT'97, pp. 98–112, 1997

11. Gupta, H., Harinarayan, V., Rajaraman, V. and Ullman, J.: Index Selection for OLAP. In proceedings of the 13th International Conference on Data Engineering, ICDE 97, Birmingham, UK, pp. 208–219, 1997
12. Haider, M. and Vijay Kumar, T.V.: Materialized Views Selection using Size and Query Frequency. International Journal of Value Chain Management, Vol.5, no.2, pp. 95–105, 2011
13. Harinarayan, V., Rajaraman, A. and Ullman, J. D.: Implementing data cubes efficiently. In ACM SIGMOD Record, Vol. 25, no. 2, pp. 205–216, 1996
14. Horng, J. T., Chang, Y. J., Liu, B. J. and Kao, C. Y.: Materialized view selection using genetic algorithms in a data warehouse system. In Evolutionary Computation, Proceedings of the IEEE Congress on, Vol. 3.pp. 2221–2227, 1999
15. Inmon, W.H.: Building the Data Warehouse, Third Edition, Wiley Dreamtech, 2003
16. Kennedy, J., & Eberhart, R. C.: Swarm intelligence. Morgan Kaufmann, San Francisco, USA, 2001
17. Kimball, R. and Ross, M.: The Data Warehouse Toolkit. Second Edition, Wiley Computer Publishing, 2002
18. Lawrence, M.: Multiobjective genetic algorithms for materialized view selection in OLAP data warehouses. In Proceedings of the 8th annual conference on Genetic and evolutionary computation, pp. 699–706, 2006
19. Lee, M. and Hammer, J.: Speeding up materialized view selection in data warehouses using a randomized algorithm. International Journal of Cooperative Information Systems, Vol. 10, no. 3, pp. 327–353, 2001
20. Mohania, M., Samtani, S., Roddick, J. and Kambayashi, Y. (1998). Advances and Research Directions in Data Warehousing Technology. Australian Journal of Information Systems, 7 (1), 41–59, 1998
21. Nadeau, T. P. and Teorey, T. J.: Achieving scalability in OLAP materialized view selection. In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, pp. 28–34, 2002
22. Payne, R.B., Sorenson, M. D. and Klitz, K.: The Cuckoos, Oxford University Press, 2005
23. Sauter, V. L.: Decision Support Systems for business intelligence. John Wiley & Sons, 2010
24. Shukla, A., Deshpande, P. M. and Naughton, J. F.: Materialized view selection for multi-cube data models. In Advances in Database Technology-EDBT 2000, LNCS, Vol. 1777. pp. 269–284, 2000
25. Shukla, A., Deshpande, P. and Naughton, J. F.: Materialized view selection for multidimensional datasets. VLDB '98 Proceedings of the 24th International Conference on Very Large Data Bases, pp. 488–499, 1998
26. Sun, X. and Wang, Z.: An efficient materialized views selection algorithm based on PSO. In International Workshop on Intelligent Systems and Applications, pp. 1–4, 2009
27. Turban, E., Aronson, J. and Liang, T. P.: Decision Support Systems and Intelligent Systems. 7th Edition. Pearson Prentice Hall, 2005
28. Vijay Kumar, T. V.: Answering Query Based Selection of Materialized Views, International Journal of Information and Decision Sciences (*IJIDS*), Volume 5, Number 1, pp. 103–116, 2013
29. Vijay Kumar, T. V. and Arun, B.: Materialized View Selection using BCO. International Journal of Business Information Systems, Volume 22, Number 3, pp. 280–301, 2016
30. Vijay Kumar, T. V. and Devi, K.: Materialized View Construction in Data Warehouse for Decision Making, International Journal of Business Information Systems, Volume 11, No. 4, pp. 379–396, 2012
31. Vijay Kumar, T. V. and Devi, K.: An Architectural Framework for Constructing Materialized Views in a Data Warehouse, International Journal of Innovation, Management and Technology, Volume 4, Number 2, pp. 92–197, 2013
32. Vijay Kumar, T.V. and Ghoshal, A.: A Reduced Lattice Greedy Algorithm for Selecting Materialized Views, Communications in Computer and Information Science, Volume 31, pp. 6–18, 2009

33. Vijay Kumar, T.V., Goel, A. and Jain, N.:, Mining Information for Constructing Materialised Views, International Journal of Information and Communication Technology, Volume 2, Number 4, pp. 386–405, 2010
34. Vijay Kumar, T.V. and Haider, M.: A Query Answering Greedy Algorithm for Selecting Materialized Views. Lecture Notes in Artificial Intelligence, Vol. 6422, pp. 153–162, 2010
35. Vijay Kumar, T.V. and Haider, M.: Greedy Views Selection using Size and Query Frequency. Communications in Computer and Information Science, Vol. 125, pp. 11–17, 2011
36. Vijay Kumar, T.V. and Haider, M.: Selection of Views for Materialization using Size and Query Frequency. Communications in Computer and Information Science, Vol. 147, pp. 150–155, 2011
37. Vijay Kumar, T.V. and Haider, M.: Materialized Views Selection for Answering Queries. Lecture Notes in Computer Science, Vol. 6411, pp. 43–51, 2012
38. Vijay Kumar, T.V. and Haider, M.: Query Answering Based View Selection. International Journal of Business Information Systems, Volume 18, Number 3, pp. 338–353, 2015
39. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Iterative Improvement. Advances in Intelligent Systems and Computing, Volume 178, pp. 205–214, 2012a
40. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Genetic Algorithm. Communications in Computer and Information Science, Volume 306, pp. 225–237, 2012b
41. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Simulated Annealing. Lecture Notes in Computer Science, Vol. 7678, pp. 168–179, 2012c
42. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Memetic Algorithm, Lecture Notes in Artificial Intelligence, Vol. 8284, pp. 316-327, 2013
43. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Differential Evolution. International Journal of Innovative Computing and Applications, pp. 102–113, 2014
44. Vijay Kumar, T.V. and Kumar, S.: Materialized View Selection using Randomized Algorithms. International Journal of Business Information Systems, Vol. 19, No. 2, pp. 224–240, 2015
45. Vijay Kumar, T.V., Haider, M. and Kumar, S.: Proposing Candidate Views for Materialization. Communications in Computer and Information Science, Volume 54, pp. 89–98, 2010
46. Vijay Kumar, T.V., Haider, M. and Kumar, S.: A View Recommendation Greedy Algorithm for Materialized Views Selection. Communications in Computer and Information Science, Volume 141, pp. 61–70, 2011
47. Yang, J., Karlapalem, K. and Li, Q.: Algorithms for materialized view design in data warehousing environment. Proceedings of 23rd International Conference on Very Large Data Bases, Morgan Kaufmann, pp. 136–145, 1997
48. Yang, X.S. & Deb, S.: Cuckoo Search via Levy Flights. World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), India. IEEE Publications, pp. 210–214, 2009
49. Yu, J. X., Yao, X., Choi, C. H. and Gou, G.: Materialized view selection as constrained evolutionary optimization. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, Vol. 33, no. 4, pp. 458–467, 2003
50. Zhang, C., Yao, X. and Yang, J.: Evolving materialized views in data warehouse. Proceedings of the 1999 IEEE Congress on Evolutionary Computation, Vol. 2, pp. 823–829, 1999

A Regression Model for Estimating the Cooling Time of Material Using Digital Microstructure Image Analysis

P.S. Hiremath, Anita Sadashivappa and Prakash Pattan

Abstract The microstructure image analysis has gained importance in material manufacturing industries. The structural and quantitative information of the constituents of material are determined using microstructure images. The dendrite arm spacing (DAS) is a significant quantitative information that needs to be estimated. In case of aluminium alloys, the cooling time is found to be directly related to DAS measurements from dendrite structures in the images of aluminium alloys. Due to the irregular shape of dendrites, the manual quantification of DAS by the metallurgical experts is quite a challenging task. In this paper, the conventional line and circle intercept methods are automated for quantification of DAS. A regression model is designed to establish the mathematical relation between the DAS and the cooling rate of the alloy. For experimentation, the aluminium-silicon alloys, namely, Al-Si 319 and 356 materials, are considered. The regression model is tested for statistical significance, which confirms the close relationship between DAS and cooling time. The experimental results are quite promising and the proposed method has practical importance in material manufacturing industry.

Keywords Dendrite arm spacing · Cooling time · Al-Si 319 and 356 alloys
Microstructure image analysis · Regression analysis

P.S. Hiremath
Department of Computer Science (MCA), KLE Technological University,
BVBCET, Hubli, Karnataka, India
e-mail: hiremathps53@yahoo.com

A. Sadashivappa (✉) · P. Pattan
Department of Computer Science and Engineering,
PDA College of Engineering, Gulbarga, Karnataka, India
e-mail: anitaharsoor@gmail.com

P. Pattan
e-mail: prakashpattan@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_31

1 Introduction

The microstructure image analysis is important in metallurgy, especially in controlling the material manufacturing parameters. The main purpose is to determine the quantitative information from microstructure images, which is useful in understanding the material properties, failure reasons, processing history and various production control parameters. The quantitative analysis includes, estimation of volume fraction of a phase, phase shape analysis, counting and measuring grain size of various constituents from microstructure images. There are various morphological shapes of constituents that are observed when the molten material is cooled, which determine the various properties of material. Dendrite structures found in Al-Si alloy are such morphological shapes that are a challenge to metallurgical experts to characterize. The dendrite arm spacing (DAS) in dendrite structures is a characteristic parameter that provides the processing history and properties of the alloy. Figure 1 shows the example microstructure images of Al-Si alloy, which is serving automotive, aerospace, metal fabrication, foundry, welding, material testing industries effectively. Figure 2 shows a simulated dendrite structure and dendrite arm spacing (d) between two arms of dendrite structure [1].

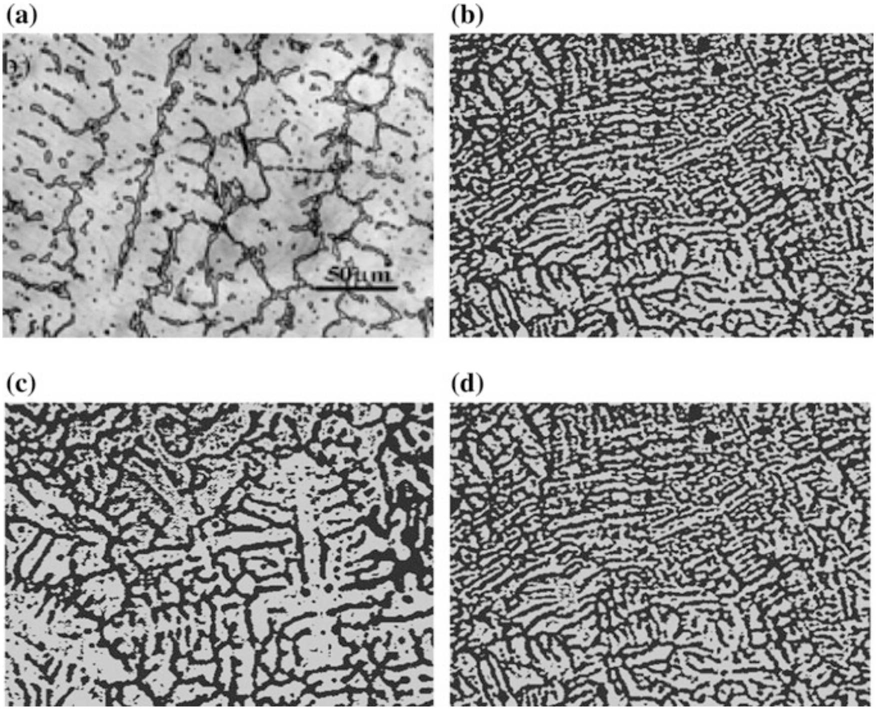


Fig. 1 Microstructure image of Al-Si alloy having dendrite structures: **a** and **b** 319 Al-Si alloy, **c** and **d** 356 Al-Si alloy

Fig. 2 Simulated dendrite structure and dendrite arm spacing (d) between two arms of dendrite structure [1]

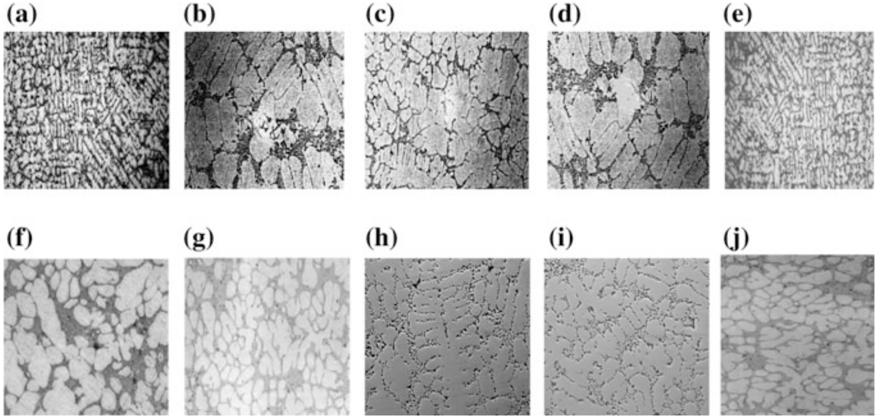


Fig. 3 Sample microstructure images: **a**–**e** Al-Si 319 alloy and **f**–**j** Al-Si 356 alloy

The experimental studies show that the spacing of arms of the dendrite in the casting depends on the solidification time, i.e., cooling time [2–4]. In [5], dendritic arm spacing in Tin-lead alloys and its significance in material properties are presented. In [6, 7], the authors have presented a theoretical analysis of the formation of a dendritic array in a directional solidification experiment. In [8–12], the growth and characteristics of dendrite structures and dendrite arm spacing are discussed.

In this paper, the objective is to propose an automated method for determining the DAS and its relation to the cooling time of material, especially Al-Si 319 and 356 alloys, based on digital microstructure image analysis. A regression model for mathematical relationship between DAS and cooling time has been established.

2 Materials Used

Unidirectional solidification experiments have been carried out on specimen material of aluminium-silicon 319 and 356 alloys [1]. Sample microstructure images of 319 and 356 aluminium-silicon alloys are shown in Fig. 3.

The alloy specimens are collected at different cooling times. Each specimen is cut, polished, and exposed under light optical microscope (LOM). The cut surface of the material is polished with nylon cloth. The microstructure images (RGB) of resolution 640 × 480 are obtained from different locations of polished surface by interfacing a computer system with LOM which is equipped with CCTV camera and image grabber card. Microstructure images are stored in .jpg format. The proposed microstructure analysis is carried out using the steps, namely, preprocessing, segmentation and finally, determining the DAS. For experimentation, the image data sets comprising 50 images of each Al-Si 319 alloy and Al-Si 356 alloys are employed. Out of 51 images of each alloy, 30 images are used for training and 21 for testing.

3 Proposed Methodology

The proposed methodology involves the microstructure image analysis to determine DAS and the regression model for cooling time estimation in terms of DAS.

3.1 Microstructure Image Analysis

Preprocessing

Generally, the microstructure images obtained using LOM have the speckle noise. The SMSF filter method [13] is applied to de-noise the image, which effectively removes speckle noise without blurring the edges of microstructure constituents. The result of preprocessing the microstructure images is a binary version of microstructure image as shown in the Fig. 4.

Segmentation

The filtered microstructure image (binary) is segmented using Otsu automatic segmentation method [14]. The discontinued edges are corrected using morphological operation, namely, hole filling.

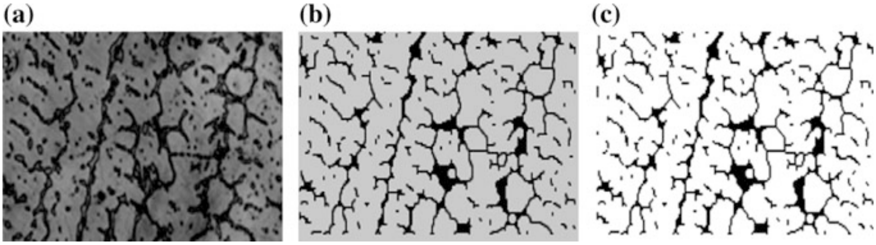


Fig. 4 Results of preprocessing: **a** Input microstructure image, **b** Filtered image, and **c** Segmented microstructure image

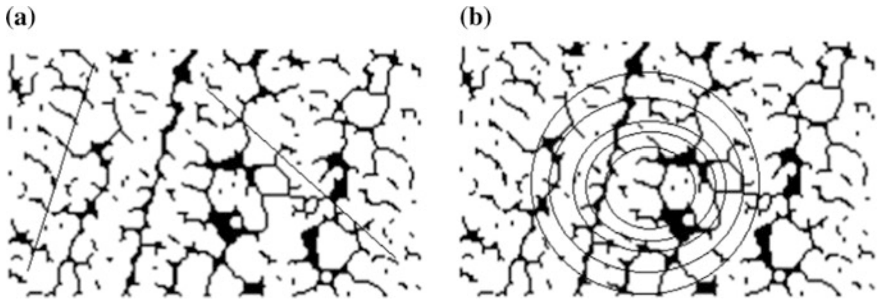


Fig. 5 Intercepts drawn automatically on microstructure image: **a** Line intercept and **b** circle intercept

Determination of DAS
The natural structure of dendrite is a tree or leaf-like or irregular structure. To determine the dendrite arm spacing (DAS), the intercept method [1] using line and circle intercepts is performed in the digital segmented image. The line and circle intercepts are drawn in digital microstructure images and DAS value is computed using Eq. (1) given as:

$$\mathrm{DAS} = \mathrm{L}/\mathrm{N}, \tag{1}$$

where L is length of line (Fig. 5a) in line intercept method, or perimeter of circle (Fig. 5b) in circle intercept method, and N is number of intersections of line (circle) with the border of dendrite structures.

The DAS is determined by drawing intercepts on different locations and the mean DAS value is computed. The average $\bar{d}$ of the DAS values obtained by using line and circle intercept methods is considered as the DAS measurement for further experimentation. This experimentation is easily repeatable using automated system, which produces authentic and accurate measurement values. For the authenticity, the experimental results are validated using manually measured DAS with the help of field experts. Table 1 shows the comparison of results obtained from the proposed method and the conventional manual method of metallurgical experts.

3.2 Regression Analysis

The regression analysis is performed to detect the correlation between DAS $(\bar{d})$ and cooling time and, also, to test if such a relation, whether assumed or calculated, is statistically significant. To learn the relation between cooling time (y) and the measured image parameter (x), viz, dendrite arm spacing, the data points are plotted (Fig. 6). The observed behavior is almost linear. The functional relationship is modeled well with linear regression using the method of least squares. The

Table 1 Comparison of DAS ($\bar{d}$) measurement obtained by using the proposed automated method and the manual method of experts

Sl. no	Material (its cooling time in s)	DAS values (micron meter)				Absolute difference between d and $\bar{d}$
		Proposed automated method			DAS by manual method (d)	
		Line intercept	Circle intercept	Average of line and circle intercept ($\bar{d}$)		
1	356 alloy (700 s)	32	32	32	30	2
2	356 alloy (800 s)	34	34	34	31	3
3	356 alloy (1100 s)	36	36	36	39	3
4	356 alloy (1200 s)	36	38	37	35	2
5	356 alloy (1250 s)	38	38	38	38	0
6	356 alloy (1500 s)	44	42	43	40	3
7	356 alloy (1900 s)	50	49	49.5	52	2.5
8	356 alloy (2100 s)	54	52	53	56	3
9	356 alloy (2200 s)	54	53	53.5	57	3.5
10	356 alloy (2300 s)	57	57	57	59	2
11	356 alloy (2350 s)	59	59	59	62	3
12	356 alloy (2500 s)	64	62	63	65	2
13	356 alloy (2600 s)	66	64	65	67	2
14	356 alloy (2800 s)	67	66	66.5	69	2.5
15	356 alloy (2950 s)	72	70	71	73	2
16	319 alloy (1200 s)	36	36	36	30	6
17	319 alloy (1250 s)	36	36	36	32	4
18	319 alloy (1300 s)	38	36	37	33	4
19	319 alloy (1500 s)	40	38	39	36	3
20	319 alloy (1750 s)	43	45	44	45	1

(continued)

Table 1 (continued)

Sl. no	Material (its cooling time in s)	DAS values (micron meter)				Absolute difference between d and $\bar{d}$
		Proposed automated method			DAS by manual method (d)	
		Line intercept	Circle intercept	Average of line and circle intercept ($\bar{d}$)		
21	319 alloy (1950 s)	49	45	47	49	2
22	319 alloy (2100 s)	52	48	50	52	2
23	319 alloy (2300 s)	56	52	54	55	1
24	319 alloy (2500 s)	59	55	57	59	2
25	319 alloy (2600 s)	63	59	61	65	4
26	319 alloy (2700 s)	68	64	66	63	3
27	319 alloy (3000 s)	74	64	69	63	3
28	319 alloy (3250 s)	77	65	71	67	4
29	319 alloy (3350 s)	79	67	73	69	4
30	319 alloy (3500 s)	82	74	78	77	1

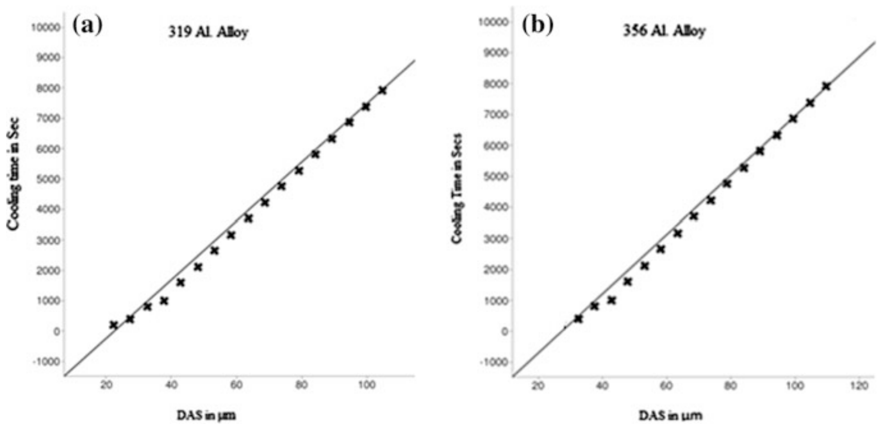


Fig. 6 Cooling time estimated using the input DAS ($\bar{d}$) by the proposed regression method for **a** Al-Si 319 alloy and **b** Al-Si 356 alloy

regression line equation obtained for 356 aluminium alloy samples is, $y = -1040.434 + 57.059 * x$ and that for 319 aluminium alloy samples is, $y = -868.487 + 57.09 * x$. The correlation coefficient for both the material is nearly 1, which indicates a strong positive correlation between DAS $(\bar{d})$ and cooling time variables. Therefore, the proposed regression model is suitable for designing an embedded system with machine intelligence applicable to material manufacturing industries in production, quality control, and material failure analysis process more effectively.

4 Experimental Results and Discussion

The results of estimated cooling time by the proposed regression model and the manual results corresponding to DAS values for two aluminium alloys, namely, Al-Si 319 and Al-Si 356, obtained from the test images are compared and shown in Tables 2 and 3, respectively.

The relative absolute differences between estimated cooling time provided by experts and determined by the proposed regression model are very small and well within the practical limits in case of both the alloys. By performing the Chi-square test at 5% level of significance [14], it is observed that there is close agreement between (i) the DAS $(\bar{d})$ values determined by the proposed automated method and the manual method, and (ii) the cooling time provided by experts and regression model, for both the Al-319 and Al-356 alloys.

Table 2 Comparison of estimated cooling time of Al-Si 319 alloy obtained by the regression model and expert values for DAS $(\bar{d})$ sample values of 319 alloy

DAS in micron meter 319 alloy	Estimated cooling time (s)		Relative absolute difference
	Expert values (O)	Regression model (E)	
31	900	913	0.1851
32	956	970	0.2020
36	1185	1198	0.1410
37	1240	1255	0.1792
44	1640	1654	0.1185
47	1814	1825	0.0663
50	2000	1996	0.0080
54	2212	2224	0.0647
57	2400	2395	0.0104
62	2667	2680	0.0630
65	2846	2851	0.0087
67	2953	2965	0.0485
69	3068	3079	0.0392
72	3240	3250	0.0307
79	3638	3649	0.0331

Tab. 3 Comparison of estimated cooling time of Al-Si 356 alloy obtained by regression model and expert values for DAS $(\bar{d})$ sample values of 356 alloy

DAS in micron meter 356 alloy	Estimated cooling time (s)		Relative absolute difference
	Expert values (O)	Regression model (E)	
31	730	728.395	0.0035
32	795	785.454	0.1160
36	1010	1013.69	0.0134
37	1068	1070.749	0.0070
44	1441	1470.162	0.5784
47	1658	1641.339	0.1691
50	1843	1812.516	0.5126
54	2031	2040.752	0.0466
57	2235	2211.929	0.2406
62	2477	2497.224	0.1637
65	2648	2668.401	0.1559
67	2787	2782.519	0.0072
69	2878	2896.637	0.1199
72	3069	3067.814	0.0004
79	3464	3467.227	0.0030

5 Conclusion

In this paper, a novel automated method for determining the dendrite arm spacing (DAS) and thereby estimating the cooling time of alloys is presented. The estimation of cooling time of alloys is well supported mathematically by the proposed regression model. The results obtained are in close agreement with expert's manual results and well within practical limits. This work serves as framework for determining many other material production parameters and paves way for building intelligent systems for controlled material manufacturing and also for material failure analysis. Further, the proposed regression model is easily amenable to deployment in embedded systems for such metallurgical applications.

References

1. ASM International Handbook Committee, ASM Handbook, Metallography and Microstructures, Vol 9, ASM International, USA, 2004
2. Spear R.E., Gardner, G.R., Dendrite Cell Size, AFS Transactions, vol. 71, pp 209–215, 1963
3. Trivedi R. and Kurz W., Dendritic Growth, Int. Mat. Rev., 39(2) pp. 49–74, 1994
4. Glicksman M.E.and Koss M.B., Dendritic Growth Velocities in Microgravity, Phys. Rev. Let., 73(4) pp 573–576, 1994
5. Bell J.A.E. and Winegard W.C., Dendrite Spacing in Tin-Lead Alloys, J. Inst. Met., 93, pp 357–359, 1963.

6. J.A. Warren and J.S. Langer, Prediction of Dendritic Spacings in a Directional-Solidification Experiment, Phys. Rev. E, 47, pp 2702–2712, 1993.
7. Makkonen L., Spacing in Solidification of Dendritic Arrays, J. Cryst.Growth, pp 772–778, 2000.
8. Emin Cadirli, Karaca I., Kaya H. and Marasli N., Effect of Growth Rate and Composition on the Primary Spacing, the Dendrite Tip Radius and Mushy Zone Depth in the Directionally Solidified Succinonitrile-Salol Alloys, J. Cryst. Growth, 255, pp 190–203, 2003.
9. Lu S.Z. and Hunt J.D., A Numerical Analysis of Dendritic and Cellular Array Growth: The Spacing Adjustment Mechanisms, J. Cryst.Growth, 123, pp 17–34, 1992.
10. Hunt J.D., A Numerical Analysis of Dendritic and Cellular Growth of a Pure Material Investigating the Transition from Array to Isolated Growth, Acta Metall., 1991, 39, pp 2117–2133.
11. Lu S.Z. and Hunt J.D., Numerical Modelling of Cellular and Dendritic Array Growth:Spacing and Structure Predictions, Mat. Sci. Eng. A, 173, pp 79–83, 1993.
12. Hiremath P.S., Anita Sadashivappa and Prakash Pattan, Analysis and Characterization of Dendrite Structures from Microstructure Images of Material, Intl‘ J. of Research in Engg. and Tech., Vol.4:5, pp. 136–140, 2015.
13. Hiremath P.S. and Anita Sadashivappa, Selective Median Switching Filter for Noise Suppression in Microstructure Images of Materials, Intl‘ J. of Image Processing, Vol. 7:1, pp. 101–108, 2013.
14. Kothari C.R. and Gaurav Garg, Research Methodology: Methods and Techniques, 3rd Edn., New Age International Publications, 2013.

Part II
Information

SDaaS: Framework of Sensor Data as a Service for Leveraging Services in Internet of Things

B.C. Manujakshi and K.B. Ramesh

Abstract From more than a decade, wireless sensor network has been adopted for various applications that call for extracting data from the geographical position which are inaccessible by humans. Although there are massive archives of literatures dealing with complexities of storage and analysis of sensory data, but the missing gap in such research work are (i) lack of considering cloud entities, (ii) does not emphasize the criticality of data complexity, (iii) non-applicability of conventional data management, and (iv) non-applicability of traditional mining techniques. Therefore, in order to address these issues, we present a novel computational framework called as SDaaS or Sensor Data as a Service which is meant for data generation, data storage, and data analysis from a single system. Supported by emulation-based technique, SDaaS is also emphasized on scalable storage system along with incorporation of user-centric knowledge mining operation. The proposed system of SDaaS is also testified with similar existing technique to find that SDaaS can perform all the three task of generation, storage, and analysis of sensory data in considerably less time and with much less usage of memory. This paper elaborates the methodology adopted along with algorithm discussion to highlight the effectiveness of proposed system.

Keywords Data analysis · Data storage · Internet of Things · Knowledge extraction · Wireless sensor network

B.C. Manujakshi (✉)
Department of Computer Science & Engineering, Acharya Institute of Technology, Bengaluru, India
e-mail: manujakshibc2014@gmail.com

K.B. Ramesh
Department of Electronics & Instrumentation Engineering, RVCE, Bengaluru, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_32

1 Introduction

The wireless sensor network has played a contributory role in gathering the surrounding data and disseminating the data from the sensors to the sink [1]. Used in various applications, e.g., habitat monitoring, environmental monitoring, industrial applications, etc., sensor nodes gather different types of data, which are not only massive but also valuable [2, 3]. A sensor network gathers a huge range of data and forward the same to the base station. However, owing to different problems of routing [4], clustering [5], security [6], it is a difficult task to ensure that the rich quality of the data being gathered from the sensor network. At present, there is much a hype of new technology called as IoT (Internet of Things), which is al about a collaborative network between sensors and cloud services. The biggest question here is IoT can only be successful if the data aggregated from the wireless sensor network are properly analyzed in lesser time, which is one of the challenging factors in IoT [7]. IoT is also about rendering the customized services to its customer in future depending upon the data being gathered from sensors. However, in order to provide such services in IoT, the complexity of the sensor data (especially in heterogeneous network that generated a higher stream of unstructured sensor data) should be analyzed [8]. Another biggest impediment is non-applicability of conventional data management software as well as data mining algorithm for such massive bits of data stream. Therefore, there is a need of a system that can address all these issues for evolving up new services by IoT in future.

Therefore, this paper presents a framework called as SDaaS, i.e., Sensor data as a service, which is meant to overcome existing problems in IoT. Section 2 discusses about the existing literatures followed by brief discussion of problem identification in Sect. 3. Section 4 elaborates about the proposed system and its contribution followed by Sect. 5 discussing about the methodology being adopted for the study. Section 6 exhibits the algorithms of SDaaS followed by comparative result discussion in Sect. 7. Finally, the paper is summarized in Sect. 8 as conclusion.

2 Related Work

This section discusses about the prior work being carried out by the prior researchers pertaining to the data management of wireless sensor network. Chen and Kang [9] have presented a framework that can perform sensory data management using earliest deadline first algorithm. The outcome of the study was evaluated using response time. Lin et al. [10] have developed similar fashion model for managing sensory data using Hadoop and MapReduce programs. The idea behind this paper was to use the extracted data to act as a service for street navigation over Android interface. Rios and Diguez [11] have presented a system that can collect, store, and can perform analysis of the sensory data using Hadoop. Guo et al. [12] have developed a mechanism that used MapReduce and indexing mechanism for an efficient processing of user query of data over cloud. Pereira et al. [13] have also used Hadoop over large datasets that are considered to be climatic data required in agricultural production. Zhong et al. [14] have presented a unique indexing mechanism that can perform an effective retrieval of data from Hadoop-based software framework. Hence, it can be seen that there are multiple studies carried out in the past as well as in recent times that is focused on using either Hadoop or MapReduce to do nearly similar kind of task of gathering, storage, and data analysis over cloud. Also all the existing system makes use of response time as the core performance factor for assessing the mechanism of presented data management techniques. However, there are some other cadre of studies where the focus is little different. Luo et al. [15] have developed a unique framework that can perform an efficient access for the data from the cloud using a unique data services. The authors have used load balancing mechanism as well as convex optimization for enhancing the optimality of the presented framework. Jin et al. [16] have developed a unique protocol that studies the multimedia data generated by the sensor network in the presence of opportunistic network. There are few more studies, e.g., [17–19], etc., that have focused on data processing over Internet of Things that connects sensors with cloud. All the above-mentioned studies have provided good guidelines for future research work and are also associated with significant issues.

3 Problem Identification

The problems that have been identified in the mechanism involved in sensory data management of the existing times are as follows:

3.1 Ignorance to Type of Data Complexity

A sensor node gathers a massive data for a definite duration of time using TDMA scheduling. Given a wireless sensor network adopts common clustering mechanism [20], there is a higher possibility of data redundancies as the same event could be captured by two or more clusters in one time. Hence, quality of data being forwarded to base station is still an open issue, which rises exponentially, when massive data is being stored thereby using the unnecessary storage unwisely as it is not possible to identify the issues of data redundancies from massive sensor data. The problem becomes much more complex if the heterogeneous wireless sensor network is considered for the study. As processing the multiple formats of the data and storing it in conventional SQL database is not possible.

3.2 Usage of Unnecessary Cloud Storage

At present, the sensory data are just stored in a definite storage location of cloud service. The biggest problem will be to segregate and perform classification of data for both storage and retrieval process instantly is never possible for such massive structure of data with definite bandwidth. Owing to non-visuality of the multiple sensory data from the heterogeneous wireless sensor network, it is never possible for a user to understand and discretize the types of the data.

3.3 Problem of Storing Unstructured Data

Normally, the data being generated by the heterogeneous wireless sensor network are highly unstructured. Owing to the nature of collaborative platform, the original metadata of the source is quite impossible to be explored in a matter of second and such problem creates an impediment toward applying conventional data mining. Moreover, as seen in prior implementations, majority of the existing approaches uses Hadoop and MapReduce software framework, whereas in reality, there are many issues in using such open source framework directly.

Hence, the problem statement of the proposed study can be defined as follows *"It is a computationally extensive task to develop a robust framework that can perform wholesome data management for wireless sensor network over cloud environment."* The next section discusses over the proposed framework that is meant to overcome the issues discussed in this section.

4 Proposed System

The prime aim of the proposed study is to develop a simple and cost-effective framework that can built a robust base of massive and complex analysis of sensor data. The proposed study is more incline toward developing a computational framework that can assist the user to perform further storage as well as data analysis and also to redefine the technique of extraction of knowledge discovery. In order to fulfill the above-mentioned goal, following research contributions were done:

4.1 A Robust Emulation

The proposed system introduces a robust emulation mechanism that can generate a massive stream of data from the sensor network.

4.2 A Scalable Storage

The system also presents a storage system that can hold the ever-growing data of the wireless sensor network over the cloud.

4.3 A Simple Knowledge Mining

The proposed system coins a term called as Sensor data as a Service or SDaaS that integrates the smart generation, storage, and analysis of the sensory data to be utilized as a smart services over cloud (Fig. 1).

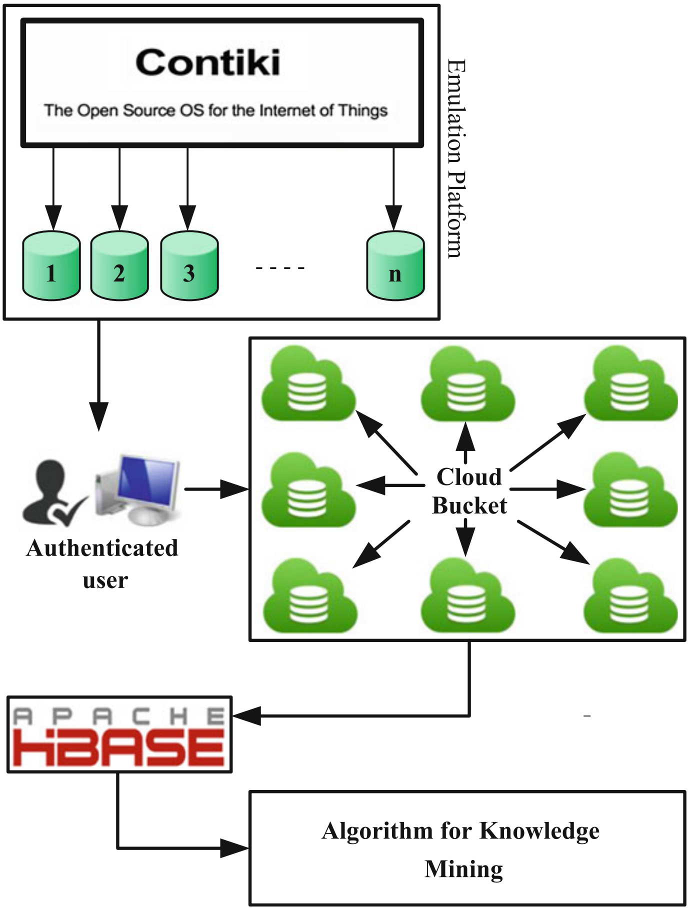


Fig. 1 System architecture of proposed system

5 Research Methodology

The development of the proposed system is carried out using analytical research approach using emulation. The research methodology is carried out in following phases:

5.1 Generation of Sensory Data

Raw sensor data could be captured from the sensor area directly or using offline standard database. Usage of both has limitation as live sensory data are hard to be accessed and reliability of study using offline data are hard to be proved. Hence, the proposed study uses Contiki which is a type of open source operation system that finds its suitability in memory constraint system, e.g., sensor networks (Fig. 2). A case study is designed where the random generations of sensory data are created. The data can be mapped equivalent to real-time sensor data as Contiki provides multiple stack of communication along with mutli-threading feature. It has also the supportability of the micro Internet protocol to the sensor network supported by event-driven kernel.

5.2 Developing a Scalable Storage System

The novelty of the proposed system is that user will be given an access as well as priviledge to store the data on respective cloud buckets (bucket is a self-created storage directory over the storage clusters). For better data management, the proposed system uses distributed database model Hbase as it supports automatic failures as well as have better supportability for query optimization. The sensory data are initially unstructured and cannot be stored in SQL-based table, hence, it is subjected to HBase, which associated its various row to each chunks of distributed data flagged with keys of the rows. Cloud bucket can be directly used to access the data using the specific keys in HBase architecture shown in Fig. 3. The distributed data with its keys are maintained on various regional server of HBase, which on request for access from bucket converts the unstructured to structured data, which is now ready to apply for any analytical operation.

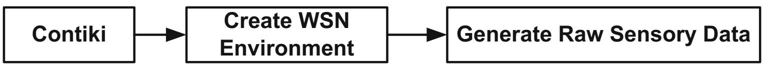


Fig. 2 Generation of sensory data

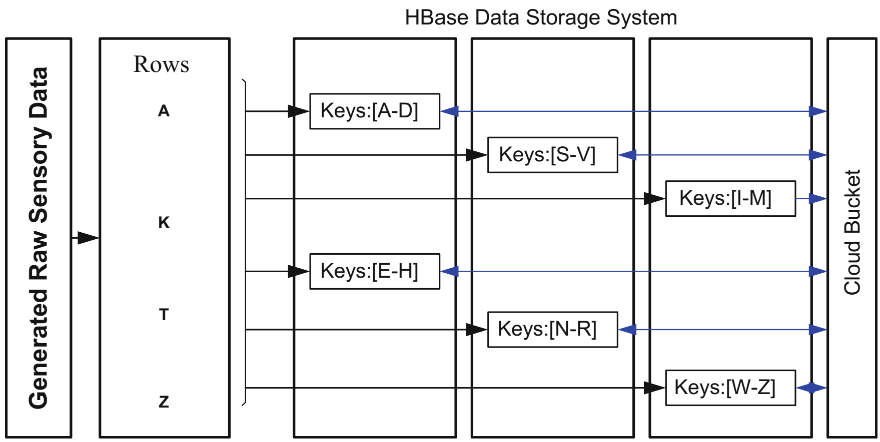


Fig. 3 Sensory data storage in HBase

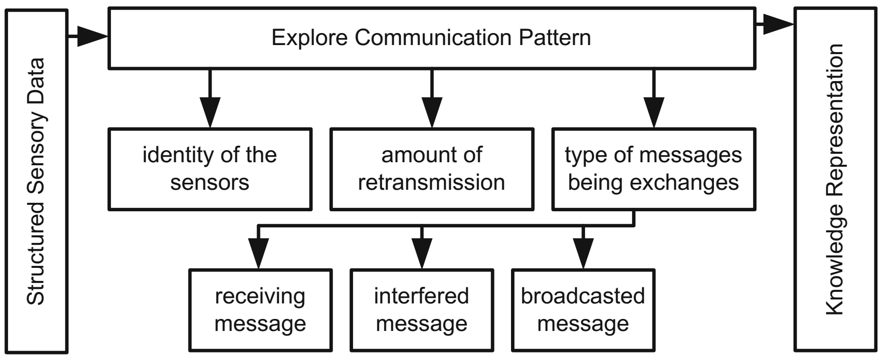


Fig. 4 Knowledge discovery from sensory data

5.3 A Simple Knowledge Mining

The data accessed from the cloud bucket performs knowledge extraction from the data and the outcome of the study will be used as Sensor Data as a Service (Fig. 4).

The proposed system will extract the transmission pattern of the sensor node, which will explicitly bring out an information about the data retained by nodes, latency, number of retransmission attempts. Such communication information will be highly valuable for the network engineers as it will give latent information about the network trends, which is normally not accessible or either studied in past. In proposed system, we define the knowledge as a specific pattern of transmission carried out by the sensor nodes in terms of communication. This knowledge can be used for visualizing the mode of workability as well as health information of the sensor during the peak traffic condition. One of the interesting fact in this part of the study is the data is being generated by runtime as a stream, where the operation of knowledge mining takes place almost instantly. As the HBase supports storage of unstructured data and give accessibility of unique structured data extraction for analysis purpose; hence, overhead on extraction on data is quite less.

6 Algorithm Implementation

The development of the proposed system SDaaS, i.e., Sensor Data as a Service is carried out using multiple softwares. We use Contiki operating system for the purpose of generating the sensory data that bear resembles with Mica Z mote. We also use centOS for running Linux to as well as MyEclipse. Linux is used for writing program for storing the data in Hbase while MyEclipse is used to design the entire framework in Java environment. The algorithm for SDaaS is discussed as below:

Algorithm for Sensor Data-as-a-Service
Input: n (number of sensor), m (message), c (clock time), s_t (session tim$_e$), S_D (Sensor data), ID (Identity information of sensor), s_m (state of message), p (Pattern)
Output: knowledge inference
Start
1. init n, m
2. set s_t=4c+rand % (4c)
3. $S_D \rightarrow [t, ID, s_m]$
4. $S_D \rightarrow$HBase
5. Extract p(HBase_data)
6. p = [p_1, p_2, ……p_n]
7. Perform inference of knowledge.
8. Store knowledge inference
End

The proposed system makes use of the Contiki operating system and Cooja simulator for generating the sensor data. Another reason for using Contiki is its ability of multitasking, dynamic loading, and handling multiple threads as compared to frequently used TinyOS. We consider the heterogeneous forms of Mica Z nodes in Contiki where the generated data is fed to Hbase which has the capability to process massive amount of unstructured sensory data with no degradation in its performance level even with increasing size of streams. The proposed system also develops a simple cloud buckets that enables the user to have an instant access to the structured data. A simple mining approach is applied that initially looks for specific patterns in the sensory data, where the term *pattern* is mainly related to communication pattern. We mainly look for some essential information of the sensors participating in the communication process, e.g., (i) identity of the sensors,

(ii) amount of retransmission, and (iii) type of messages being exchanges (e.g., receiving message, interfered message, broadcasted message, etc.) on a specific time stamp. This information will assist the network engineer to identify the communication pattern as knowledge that will lead to take necessary measures at the time of adverse environment. This is highly assistive to track nodes depleting higher energy, success rate of new routing table, compromised nodes during security breaches, latency, and respective information of the neighbor sensors involved in the process. The efficiency of SDaaS is assessed with respect to response time.

7 Result Discussion

The outcome of the proposed system is compared with the recent work carried out by Tracey and Sreenan [21]. The authors have presented a unique architecture of collaborating cloud with sensor network. The prime reason for selecting this paper for comparative analysis because of following reasons, e.g., (i) the study aims toward generating sensor data to cloud and (ii) the study also uses Contiki, Linux, and HBase for mainly data storage. Hence, owing to similar aim and tool usage, we choose to consider performing comparative analysis. However, proposed study performs additional task that enhances the performance of the sensor data management. This section discusses about the outcome accomplished from the study to prove the effectiveness of proposed system.

7.1 Analysis of Response Time

As sensor data is large enough, it is not only difficult to store but also extremely challenge to extract knowledge in less time. Thereby response time could be quite higher and may not solve the purposes for time critical applications in future of IoT considering 95% of confidence interval. A closer look into the outcome of Fig. 2 will show that proposed system has better performance of response time with increasing size of sensor data as compared to the work done by Tracey [21]. The prime reason behind this is proposed system is not only about data storage but also about conversion of complex data to simple and structured data, which can be then processed from HBase. Moreover, the proposed system uses a simple method of knowledge extraction that can understand the pattern of sensory data propagation. The simulation performed by Cooja also assists in reducing the response time to a larger extent. However, an approach designed by Tracey [21] does not have any of such features that lead to more processing time involved in storing the data to the

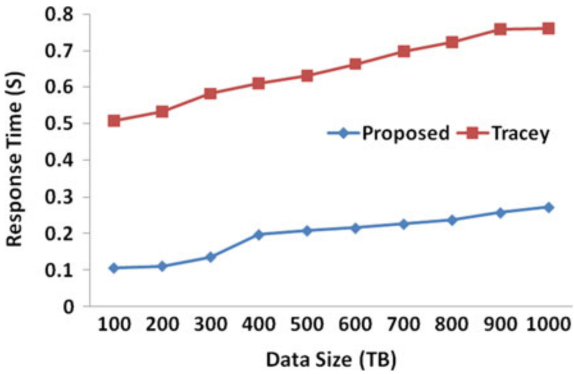


Fig. 5 Analysis of response time

Hadoop. We strongly emphasize that it is better option to use the structured data rather than unstructured data and then consider it as a service. Therefore, SDaaS ensures highly reduced response time and does not get significantly affected with increasing load of incoming streams of data (Fig. 5).

7.2 *Analysis of Memory Usage*

The proposed system considers usage of a cloud node that acts as both slave node of Hadoopand sinks point too together. After receiving sensory data, it is then forwarded to the system with Hadoop being acted as software framework. The study uses the accomplished sensor data from Contiki and store it in HDFS manner thereby rendering the file system to be only applicable for read only operation. This ensures less memory usage. Another interesting part of the proposed system is the knowledge extractionprocess. Hbase just assists to perform data conversion from unstructured manner to structured manner so that it can be easily subjected to data analysis. We apply our user-centric extraction of knowledge by extracting few significant feature vectors from the structured data. Hence, this operation does not call for using extra memory as extraction of the knowledge takes place from structured data with faster response time along with exhibition of knowledge in runtime without any storage of an intermediate memory. Therefore, SDaaS confirms both faster response time with runtime compilation to extract knowledge with lower memory consumption. Figure 6 highlights the outcome of memory usage analysis of proposed system which is comparatively lower in contrast to existing work being carried out by Tracey [21].

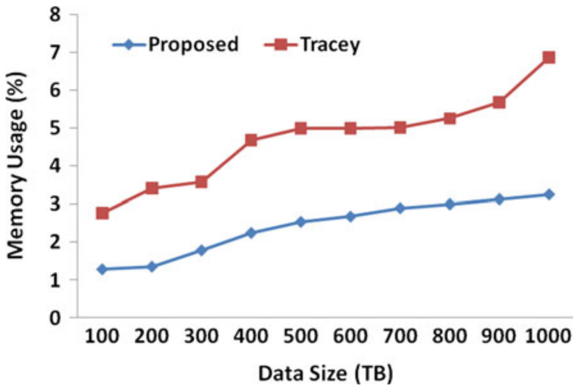


Fig. 6 Analysis of memory usage

8 Conclusion

This paper has presented a simple framework where sensory data is taken as an input and after processing by proposed SDaaS, it gives output of knowledge being extracted from the sensory data. The primary novelty of the proposed is to use the sensor data as service, which was never attempted before. The secondary novelty factor is a simple mining operation based on data whose availability is 100% guaranteed from any routing table (e.g., node identity, retransmission, messages types, timestamp). We apply Cooja Simulator and Contiki Operating system in order to generate sensory data more reliably considering the standard example of Mica Z mote. The sensory data are stored in Hadoop where the unstructured data is realized and is converted to structured data using Hbase. The structured data is then subjected to simple knowledge extraction process. We design a simple mechanism of extraction of time and event information of specific behavior of sensor node transmission that potentially assists the user to declare if the nodes are in correct states of transmission or do they require a special attention. This framework if molded to application can greatly assists various upcoming applications of IoT, where the massiveness of sensor data can be easily handled and a novel application can be developed that can extract specific event related information (e.g., when the fire has been detected, is the node is correct energy state, is the node following the correct routes, etc.). Our future work will be to further continue the work of analysis of sensor data to extract more sophisticated knowledge.

References

1. Balamurugan P. Data Diffusion and Gathering in Wireless Sensor Networks. Journal of Global Research in Computer Science. 2011 Aug 8;2(7):42–6.
2. Hull B, Bychkovsky V, Zhang Y, Chen K, Goraczko M, Miu A, Shih E, Balakrishnan H, Madden S. CarTel: a distributed mobile sensor computing system. InProceedings of the

4th international conference on Embedded networked sensor systems 2006 Oct 31 (pp. 125–138). ACM.
3. Zhang Y, Meratnia N, Havinga P. Outlier detection techniques for wireless sensor networks: A survey. Communications Surveys & Tutorials, IEEE. 2010 Apr 1;12(2):159–70.
4. Pelusi L, Passarella A, Conti M. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. Communications Magazine, IEEE. 2006 Nov;44(11):134–41.
5. Muruganathan SD, Ma DC, Bhasin RI, Fapojuwo AO. A centralized energy-efficient routing protocol for wireless sensor networks. Communications Magazine, IEEE. 2005 Mar;43(3): S8–13.
6. Muruganathan SD, Ma DC, Bhasin RI, Fapojuwo AO. A centralized energy-efficient routing protocol for wireless sensor networks. Communications Magazine, IEEE. 2005 Mar;43(3): S8–13.
7. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems. 2013 Sep 30;29(7):1645–60.
8. Barnaghi P, Sheth A, Henson C. From Data to Actionable Knowledge: Big Data Challenges in the Web of Things [Guest Editors' Introduction]. Intelligent Systems, IEEE. 2013 Nov; 28(6):6–11.
9. Chen, Liehuo, and Kyoung-Don Kang. "A Framework for Real-Time Information Derivation from Big Sensor Data." High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on. IEEE, 2015.
10. Lin CY, Li CC, Huang WH, Liao WC, Chen WM. A Sensor Data Processing and Access Platform Based on Hadoop for Smart Environments. InNetwork-Based Information Systems (NBiS), 2014 17th International Conference on 2014 Sep 10 (pp. 455–460). IEEE.
11. Rios LG, Diguez JE. Big data infrastructure for analyzing data generated by wireless sensor networks. InBig Data (BigData Congress), 2014 IEEE International Congress on 2014 Jun 27 (pp. 816–823). IEEE.
12. Guo T, Papaioannou TG, Aberer K. Model-view sensor data management in the cloud. InBig Data, 2013 IEEE International Conference on 2013 Oct 6 (pp. 282–290). IEEE.
13. de Fatima Pereira, Rosangela, et al. "Distributed processing from large scale sensor network using Hadoop." Big Data (BigData Congress), 2013 IEEE International Congress on. IEEE, 2013.
14. Zhong, Yunqin, Jinyun Fang, and Xiaofang Zhao. "VegaIndexer: a distributed composite index scheme for big spatio-temporal sensor data on cloud." Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, 2013.
15. Luo Y, Yang K, Tang Q, Zhang J, Li P, Qiu S. An optimal data service providing framework in cloud radio access network. EURASIP Journal on Wireless Communications and Networking. 2016 Dec 1;2016(1):1–1.
16. Jin, Yong, et al. "QoS guarantee protocol based on combination of opportunistic dynamic cloud service and cooperative multimedia stream for wireless sensor networks." EURASIP Journal on Wireless Communications and Networking 2015.1 (2015): 1–13.
17. Amadeo, Marica, Claudia Campolo, and Antonella Molinaro. "Multi-source data retrieval in IoT via named data networking." Proceedings of the 1st international conference on Information-centric networking. ACM, 2014.
18. Satoh, Ichiro. "A Data Processing Framework for Distributed Embedded Systems." Intelligent Distributed Computing IX. Springer International Publishing, 2016. 199–209.
19. Chen, Feng, et al. "Data mining for the internet of things: literature review and challenges." International Journal of Distributed Sensor Networks 201515): 12.

20. Cygan, D., et al. "RACE-II advanced TDMA mobile access project An approach for UMTS." Mobile Communications Advanced Systems and Components. Springer Berlin Heidelberg, 1994. 428–439.
21. Tracey, David, and Cormac Sreenan. "A holistic architecture for the internet of things, sensing services and big data." Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium, IEEE, 2013.

An Approach for Efficient Capacity Management in a Cloud

T. Roseline Philomine, Clarence J.M. Tauro and Melisa Miranda

Abstract Cloud computing is an emerging technology where computing resources such as software and hardware are accessed over the Internet as a service to customers. In the past, due to less demand, cloud capacity management was not critical. However, with the increase in demand, capacity management has become critical. Cloud customers can frequently use web-based portals to provision and de-provision virtual machines on demand. Due to dynamic changes as per the demand, managing capacity becomes a challenging task. In this paper, we discuss the emergence of cloud computing, traditional versus cloud computing, and how capacity management can be efficiently handled in a cloud. A detail on high availability of virtual machines in a cloud using the N + 1 model is discussed in this paper. With templates, many repetitive installation and configuration tasks can be avoided. We discuss the sizing of templates and the overheads of using virtual machines. We suggest ideal combinations of sizing templates to create virtual machines with optimum utilization of blades. Finally we discuss a few benefits of efficient capacity management in cloud computing.

Keywords Cloud computing · Traditional capacity management
Efficient cloud capacity · Virtual machines · ESXi host · Hypervisor
High availability

T.R. Philomine (✉)
Department of Computer Applications, Presidency College, Bangalore, India
e-mail: tproseline@gmail.com

C.J.M. Tauro
Center for Research, Christ University, Bangalore, India
e-mail: clarence.tauro@res.christuniversity.in

M. Miranda
Department of Electronics and Communications, PES University, Bangalore, India
e-mail: melisamiranda@pes.edu

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_33

1 Introduction

Cloud computing is gaining popularity due to the minimal cost of IT ownership and ready-to-use, dynamically scalable computing infrastructure. It is an innovative class of computing, based on communication that takes place over the Internet [1]. For both commercial and noncommercial computational customers, the cloud provides a promising platform for low-cost and on-demand services.

a. Public Cloud
b. Private Cloud
c. Hybrid Cloud

A public cloud or external cloud is a service where customers are provided with virtual machines on demand. There are many types of services available in public clouds, namely Infrastructure (IaaS), Platform (PaaS), Software (SaaS), Storage (STaaS) and Database (DBaaS) as a service [2]. We focus on Infrastructure as a Service in this paper. Private clouds or internal clouds are used by huge organizations. They build their own cloud environments and use them. Hybrid cloud is a combination of public and private cloud environments; this is the preferred choice for most cloud users.

Because of elasticity and low cost, there are many new challenges for cloud service providers. One of the recent challenges in cloud computing is capacity management, which has a critical impact on quality of service and profitability. Since the cloud providers are offering pay as you go services, resource demands are becoming more unpredictable than in the traditional IT environment [3]. This type of service brings fresh challenges in managing the capacity of the cloud; such services include instant virtual machine provisioning and effective utilization of resources. Capacity management plays a vital role to ensure that the cloud providers are getting good performance and gaining enough computational performance as per their requirements. Capacity management guarantees that the required computational resources are available for proper business functionality and ensures profitability for cloud service providers. Capacity management will be one of the most significant tasks for any project in a virtualized environment. But often, cloud providers struggle with inefficient capacity management because of frequent provisioning and de-provisioning of virtual machines. There must be an ongoing balance between quality of service, cost and resource utilization to get more value out of the cloud. How are we going to maintain the ongoing balance by reducing the cost and utilizing the resources for greater profitability without performance issues? Capacity management must be handled in an efficient way because of on-demand and rapid changes in a cloud. Hence the technology-oriented approach must be changed to a business-aware and holistic approach.

2 Traditional Approaches Vs. Cloud

To understand why troubleshooting performance issues and analysis of capacity in cloud environments are more difficult, we need to understand how the traditional approach differs from the cloud approach. Traditionally there was one physical machine hosting one application on one operating system. If the machine runs out of primary memory then the application running on that machine will be impacted. This will not impact any other applications and machines in the same network. But in cloud environments, all the virtual machines could be hosted on a single physical machine; there are possibilities that virtual machines are seeking more resources in order to process the user requests. Virtual machines are impacted when the memory of the host machines is utilized to its full capacity and in addition, a new application tries to access the same memory.

User expectations have changed due to the emergence of cloud computing. Users want to be connected on the go from any devices, anywhere. These user expectations or requirements lead to more innovations, which in turn increase the complexity of the system, which in turn leads to more complexity in resolving issues based on capacity and performance. Cloud computing, when compared to the traditional approach might seem to have a lot of drawbacks, but can be overcome by taking proper steps like using appropriate tools that will help in understanding the environment in detail, which helps in troubleshooting capacity-related issues. Issues can be prevented prior to their occurrence and the complexity in the cloud can be minimized with the help of these tools.

There are a few more important questions to be answered to achieve efficient capacity management in a virtualized cloud environment. The questions are as follows:

1. What am I doing with my current capacity?
2. How much more resources can I accommodate with my current capacity?
3. When will I run out of capacity?
4. What happens if I increase or decrease capacity?
5. What happens if I change my resource utilization plan?
6. Where is the best place to run my workloads?

Capacity management is broadly classified into four modules as follows:

1. Resource utilization analysis.
2. Trend analysis.
3. Capacity forecasting.
4. Modifications based on trend and forecast analysis.

We concentrate on the first module, resource utilization analysis. For efficient capacity management, it is important to analyze the current resource utilization of the cloud environment. Questions on capacity management that will help us to understand and analyze resource utilization are:

1. How many virtual machines can my virtual environment accommodate?
2. How many virtual machines does the environment currently accommodate?
3. What is the maximum capacity of the virtual machine that can be added to the existing environment?
4. If I want to add 'N' virtual machines, will the environment be able to accommodate?

Here we need to consider the performance of the environment, which is dependent on the total capacity available and utilized by the virtual machines.

3 Capacity Planning for Physical Servers

During capacity planning for a physical server in a cloud, the first question that comes to one's mind is "How many virtual machines can a physical server accommodate?" and the answer is, that it depends on the virtual CPU's and virtual memories available in the physical server. The total count also depends upon the different sizes of virtual machine combinations created on a single host machine. We will discuss the best combinations of virtual machines with different sizes, later. Before the introduction of clusters in virtualization with distributed resource schedulers and high availability, accommodating more virtual machines was sufficient [4]. But at present, considering only physical server capacity is not adequate; even the clusters have to be implemented. Considering clusters in capacity management helps to ensure that the mission-critical virtual machines running on a physical host are available 24/7 even if there is a failure.

One of the failure scenarios, also known as unplanned downtime is the outage which occurs when one or more physical server's crash unexpectedly.

Let's consider a maintenance window in which after patching we reboot the physical server. This scenario is treated as planned downtime. Customers want their virtual machines to be online in both planned and unplanned scenarios. The Virtual machines can be preserved online during a maintenance window using the vMotion [5] technique offered by VMware. vMotion is a technique in which virtual machines are migrated from one physical machine to another physical machine without causing issues to the end users. This is also known as hot migration because there is no downtime and users are not affected (SWITCH-OVER).

The virtual machines can be preserved online during an unexpected crash with minimal impact to the machines by using the high-availability feature. Fault-tolerance is a technique in which unexpected failures are estimated and prevented before the failure actually takes place (FAIL-OVER) [6].

Figure 1 above shows the use of three physical host machines to provision virtual machines. An extra physical host will be used for failover or switchover techniques as discussed earlier. High-availability features identify when any physical machine shuts down unexpectedly then it coordinates with other available physical host machines in the cluster. Once an available host is identified the virtual

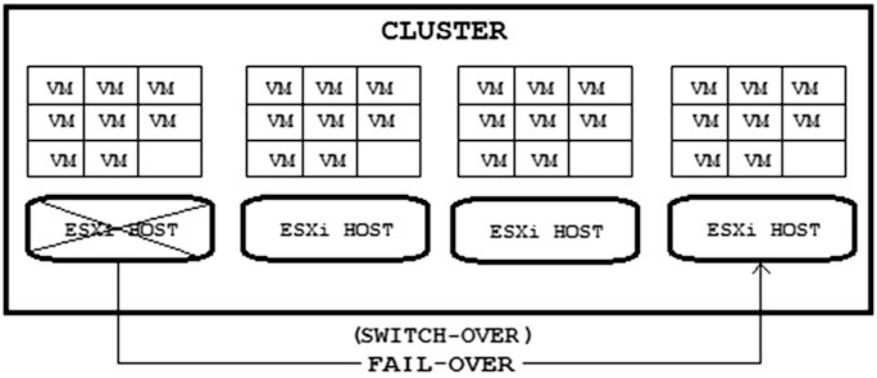


Fig. 1 N + 1 Model

machines are restarted on the identified host without human intervention. To make this happen, additional physical machines are employed to handle such scenarios. So cloud providers need to ensure that enough space will be kept in reserve to handle these scenarios during failover in a cloud.

Some cloud providers adopt the N + 1 model based on a guesstimate of physical machines. Let's consider a requirement that says we require four physical servers; and then we need to add one more physical machine according to the N + 1 model for high availability and extra space. This appears to be a good approach, but adding an extra integrated elastic sky x (ESXi) host might not be sufficient [7].

Keep in mind that capacity analysis is all about identifying how many virtual machines can be accommodated in a physical host. Thus the physical server is frequently utilized at 80% or more, which means until the physical server's memory is 80% utilized, there is a capability of adding more virtual machines on that server. According to the N + 1 model, if there is an unexpected failure then the failed host machine's load will be managed by the fifth ESXi host. But if two host machines fail at the same time then there will be a problem.

Let's consider a case where there will be a failure of two physical machines, which utilize 80% in the 4 + 1 model. Then the other three physical machines will be overutilized due to unavailability of two machines, which impacts there performance and operation. Let's consider physical machines that are only 60–65% utilized, then it is suitable for the 4 + 1 model, however we need more physical hosts to meet the requirements and avoid overutilization of machines. In reality most cloud providers may opt for more than one extra physical host machine (N + 2 or N + 3) to handle fail-over scenarios and to provide good performance and quality service. Hence good planning is required to efficiently manage the capacity in a cloud environment. We can classify the virtual machines into two major pools, like production and non-production VMs; then create clusters respectively.

To implement a distributed resource scheduler and high availability (HA) on a production cluster where the mission-critical virtual machines are hosted, the N + 1

Model provides sufficient space to handle maintenance activities when bringing down one machine at a time. There are certain limitations imposed on the number of virtual machines to be restarted by HA configuration to handle multiple unexpected failures. Another workaround is to implement the same plan on only one cluster, by turning off the restart option in cluster properties for non-mission-critical VMs. Monitoring the physical host environment plays a vital role in ensuring that we have met the requirements of a distributed host scheduler and HA. As the virtual machine count increases the physical host machines must also be increased.

4 Sizing Template

For optimum usage of blades in virtual machines, selecting the amount of virtual memory and virtual CPU is crucial. Network cards and secondary memory may be added or removed as per the need. Virtual CPU's and memories are limited resources with high demand. Allocating more virtual CPU's will reduce the return on investment and wastage of limited resources. Additionally allotting more virtual CPU's would lead to performance issues on the entire physical host machine. Moreover allotting fewer real CPU's for a mission-critical production system would lead to performance bottlenecks. Hence sizing the templates for virtual machines is crucial and should be done efficiently by balancing the return on investment and providing optimal performance for each VM. Just like virtual CPU's, oversizing virtual memories will affect the performance of all the machines on the physical host and also reduces the return on investment (ROI). Similarly undersizing a virtual memory will lead to operational issues and lack of performance. We have created a few templates based on different kinds of requirements as shown in Table 1. These templates may be useful to accomplish the requirements and also to efficiently manage the resources for better ROI.

Table 1 Blade Properties

vCPU (GHz)	VM memory
Manufacturer	IBM
Model	Blade Center HS22 V – [7871H4G x3650 M3
CPU cores	12 CPUs x 2.933 GHz
Processor type	Intel(R) Xenon(R)CPU X5670 @ 2.93 GHz
Processor sockets	2
Logical processors	6
Virtual CPU	24
Number of NICs	48 vCPU
Memory	144 GB
Operating systems	VMware ESXi hypervisor

5 Memory Overheads in Virtual Machines

Virtualization of physical memory in a cloud environment has a few associated overheads. In general, VM's can acquire two types of overhead:

1. Supplementary time required to access the memory inside a VM.
2. Other than the allocated memory to each VM, additional space maybe required for the physical host machine to store its own metadata.

A physical host machine's memory virtualization includes a slight time overhead to access memory. Processors use shadow page tables for software methods and nested page tables for hardware-based methods due to which memory access in a VM can be processed without overhead from address translation [8–10].

1. VM kernel fixed overhead
2. Each VM extra overhead

Physical memory overhead includes reservation of additional memory for each VM frame and different virtualization metadata like shadow and nested page tables. The amount of overhead memory depends upon the total number of vCPU's and the amount of memory allocated for the virtual machine OS.

5.1 *Overhead in Virtual Machines*

Each VM hosted on ESXi requires some amount of overhead memory during startup. We should know the amount of overhead memory required for startup. The table below explains the amount of overhead memory required by each VM based on the vCPU's and memory allocated.

After the startup of each VM, the amount of memory listed in Table 2 may change. The sample values were collected with VMX swap enabled and hardware MMU enabled for the virtual machine [11–14]

Table 2 Overhead memory

Size	Memory	vCPU	Overhead
Ex-small	1024	1	28.47
Small	2048	1	34.78
Small+	4096	1	47.95
Medium	4096	2	51.60
Medium+	8192	2	73.68
Large	12288	4	110.97
Ex-large	16384	4	132.89

6 Research Environment

In a cloud, virtual machines are provisioned and de-provisioned by physical machines called ESXi host. A conceptual layer is formed by the Hypervisor between the physical and virtual machines [15–17]. The resource scheduler is responsible for allowing access to the hardware for the virtual server. In cloud computing, a virtual machine CPU is called a vCPU [18] and a physical machines CPU is called a pCPU. The Memory of the physical host machine is shared among all the virtual machines deployed on that ESXi host as virtual memory [19]. This type of resistance allows creating different ranges of virtual machines as mentioned in Table 3 with different virtual machines. Different template are offered by cloud provider like extra-small (XS), small (S), small + (S+), medium (M), medium + (M +), large (L) and extra-large (XL). All these templates are mentioned below with Virtual CPU's and Virtual Memories allocated for them.

There are a number of physical host machines available in the market that can be used to build a cloud. To build a cloud many physical machines will be used.

Assuming only two physical host machines are used and the hypervisor is built on IBM Blade Center HS22 V [7871H4G] x3650 M3 server blade with VMware ESXi hypervisor operating system installed, the specification of this blade is mentioned below in Table 1. One core is equal to four virtual CPU's. Specification of the blade is mentioned below in Table 1.

As mentioned above in Table 1, the IBM Blade server 7871 x3650 M3 [20, 21, 23] supports a maximum of 144 GB Memory, 3.33 GHz CPU and 8 terabyte of disk. The available CPU and memory of the blade depend upon the license purchased. The hypervisor license cost is calculated based on the total memory and CPUs.

The amount of available cores and virtual CPU's will be counted based on processor sockets available and cores per socket on a host. If 2 processor sockets and 6 cores per socket then $2 \times 6 = 12$ total cores will be needed and if 24 logical processors and 2 processor sockets then $24 \times 2 = 48$ virtual CPU's are needed.

$$Total\,number\,of\,cores = Processor\,Sockets \times cores\,per\,socket \tag{1}$$

$$Total\,number\,of\,virtual\,CPU's = Processor\,Sockets \times Logical\,Processors \tag{2}$$

Table 3 VM template dashboard

Size	vCPU (GHz)	VM Memory
Ex-small	1	1 GB
Small	1	2 GB
Small+	1	4 GB
Medium	2	4 GB
Medium+	2	8 GB
Large	4	12 GB
Ex-large	4	16 GB

Table 4 Resource utilization table

Size	Number of VM's	vCPU Waste	Memory waste (GB)
Ex-small	40	0	104
Small	40	0	64
Small+	36	4	0
Medium	24	0	48
Medium+	18	12	0
Large	12	0	0
Ex-large	9	12	0

Table 5 Proposed plan one

Size	Number of VM's	vCPU	Memory (GB)
Ex-small	9	9	9
Small	10	10	20
Medium+	7	14	56
Ex-large	2	8	32
Total	28	41	117

7 Template Combinations

7.1 Wastage Scenario

A customer's requirements vary according to his demands. As per the customer's requirements, cloud providers have different templates available as shown in Table 1. Resource utilization is displayed in Table 4 based on template size.

Suppose the extra-small template is used to create a virtual machine on a single host then there will be 100% vCPU utilization and 27.8% memory utilization. This results to 104 GB of memory wastage. Similarly, wastage is identified for each of these templates. Resource utilization for large template-size VMs is 100%, which leads to performance issues. To overcome these issues seven physical servers should be used to host single size VMs. As the size increases, more physical machines are hosted in cloud leading to a cost-inefficient plan and memory wastage [24, 8].

7.2 Proposed Plan

Considering the high-availability VMs with N + 1 model, two plans are proposed as shown in Table 5 and 6, using these proposed plans two physical servers will efficiently utilize 80% of vCPU and memory. This proposed plan will efficiently handle the cloud environment with good performance.

Table 6 Proposed plan two

Size	Number of VM's	vCPU	Memory (GB)
Small+	10	10	40
Medium	5	10	20
Large	5	20	60
Total	20	40	120

8 Conclusions

Hence, efficient capacity management will be crucial for successful implementation of cloud environments. The comprehensive capacity of a cloud must be adequate to hold the dynamic assignments of its workload, while maintaining the performance level as agreed. The quality of service provided, should ensure that the number of VMs employed are adequate.

Efficient capacity management allows:

1. Existing infrastructure usage optimization.
2. Using the cloud resources efficiently through which cost is reduced.
3. New infrastructure purchases can be planned properly and efficiently.
4. Improvement in quality of service.
5. Unnecessary work elimination.
6. Capacity management of cloud will be consistent through N + 1 model.
7. Provides efficient capacity in a timely manner, based on the latest information.
8. Addresses the bottlenecks before the business services are severely impacted.

References

1. Shaikh, F.B.; Haider, S.; Security threats in cloud computing,Internet Technology and Secured Transactions (ICITST), 2011 International Conference for, vol., no., pp. 214–219, 12–13 Dec 2012
2. Mikkilineni, R.; Sarathy, V.; Cloud Computing and the Lessons from the Past,Enabling Technologies: Infrastructures for Collaborative Enterprises, 2009. WETICE '09. 18th IEEE International Workshops on, vol., no., pp. 57–62, 23 Dec 2012.
3. Yexi Jiang; Chang-shing Perng; Tao Li; Rong Chang; Intelligent cloud capacity management, Network Operations and Management Symposium (NOMS), 2012 IEEE, vol., no., pp. 502–505, 12–13 Dec 2012.
4. VMwarevSphereHighAvailability,[Online]http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-availability-guide.pdf, 09–10 Dec, 2012.
5. VMware vSphere vMotion, performance and Best practice, [Online].http://www.vmware.com/files/pdf/techpaper/VMware-vSphere51-vMotion-Perf.pdf, 21–22 Nov, 2012.
6. VMware vSphere Fault Tolerance, [Online]. http://www.vmware.com/products/datacentervirtualization/vsphere/fault-tolerance.html, 10–11 Nov, 2012.
7. Ahmad, I.; Anderson, J.M.; Holler, A.M.; Kambo, R.; Makhija, V.; An analysis of disk performance in VMware ESX server virtual machines,Workload Characterization, 2003. WWC-6. 2003 IEEE International Workshop on, vol., no., pp. 65– 76, 27 Dec. 2012.

8. Patidar, Shyam, Dheeraj Rane, and Pritesh Jain. “A survey paper on cloud computing.” Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference on. IEEE, 2012.
9. Leimeister, Stefanie, et al. “The Business Perspective of Cloud Computing: Actors, Roles and Value Networks.” ECIS. 2010.
10. Vaquero, Luis M. “EduCloud: PaaS versus IaaS cloud usage for an advanced computer science course.” Education, IEEE Transactions on 54.4 (2011): 590–598.
11. Böhm, Markus, et al. “Cloud computing–outsourcing 2.0 or a new business model for IT provisioning?.” Application management. Gabler, 2011. 31–56.
12. Zhang, Jie, et al. “A Framework of User-Driven Data Analytics in the Cloud for Course Management.” Proceedings of the 18th International Conference on Computers in Education.–Wong SL et al., Eds., Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education. 2010.
13. Buyya, Rajkumar, James Broberg, and Andrzej M. Goscinski, eds. Cloud computing: principles and paradigms. Vol. 87. John Wiley & Sons, 2010.
14. Di Spaltro, Dan, Alex Polvi, and Logan Welliver. “Methods and systems for cloud computing management.” U.S. Patent Application No. 12/777,218.
15. ESXi Hypervisor Architecture,[Online]. http://www.vmware.com/products/vsphere/esxi-and-esx/why-esxi.html, 03-04 Jan. 2013
16. VMware vSphere Hypervisor 5.1, [Online]. https://my.vmware.com/web/vmware/evalcenter?p=free-esxi5&lp=default, 03-04 Jan. 2013
17. Sotomayor, Borja, et al. “Capacity leasing in cloud systems using the opennebula engine.” Workshop on Cloud Computing and its Applications. Vol. 3. 2008.
18. Raghavendra, K.T.; Vaddagiri, S.; Dadhania, N.; Fitzhardinge, J.; Paravirtualization for Scalable Kernel-Based Virtual Machine (KVM), Cloud Computing in Emerging Markets (CCEM), 2012 IEEE International Conference on, vol., no., pp. 1–5, 11-12 Oct. 2012.
19. Wentao Liu; Research on cloud computing security problem and strategy,Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on, vol., no., pp. 1216–1219, 12–13 Dec 2012
20. Scoreforx3650[online]ftp://public.dhe.ibm.com/eserver/benchmarks/news/newsblurb_x3650M3_specpower_031610.pdf 28 Dec. 2011
21. Vaquero, Luis M., et al. “A break in the clouds: towards a cloud definition.”ACM SIGCOMM Computer Communication Review 39.1 (2008): 50–55.
22. Iyer, Bala, and John C. Henderson. “Preparing For The Future: Understanding the Seven Capabilities Cloud Computing.” MIS Quarterly Executive 9.2 (2010).
23. Abadi, Daniel J. “Data management in the cloud: limitations and opportunities.” IEEE Data Eng. Bull. 32.1 (2009): 3–12.
24. Karthick, A. V., E. Ramaraj, and R. Ganapathy Subramanian. “An efficient multi queue job scheduling for cloud computing.” Computing and Communication Technologies (WCCCT), 2014 World Congress on. IEEE, 2014.

Automated Negotiation of QOS Parameters Using Fuzzy Logic Based Concession Strategy for Cloud Computing Environments

B. Sunil Kamath and Rio D'Souza

Abstract Cloud service negotiation can happen in two ways, i.e. using trade-off and concession but there are other ways of accomplishing service negotiation. The one we propose below is using the fuzzy model we generate a negotiation strategy which can give better social benefit and success rate for the problem under consideration. In our model we have used fuzzy logic to give acceptance on a proposal being exchanged between cloud consumer and cloud provider. Although the strategy is similar to concession it gives better utilities when compared to the previous concession strategy. It is also observed that Fuzzy Logic-based concession (FLC) achieves better utilities for both consumer and provider 55% of the time over a random dataset as compared to only 10% for trade-off approach. The success rate and social benefit of all strategies are consolidated in a table at the end.

Keywords Cloud computing · Service negotiation · QOS · Fuzzy logic

1 Introduction

There are three service models in cloud. They are Infrastructure as a service (IaaS), Platform as a service (PaaS) and Software as a service (SaaS). The cloud computing is all about procuring Network, Storage, Compute facility from Infrastructure providers and pay them on a pay-per-use basis. Now there are some functional requirements (the various functionalities of the service) and non-functional requirements (reliability or response time of the service) attached to the price

B. Sunil Kamath (✉) · R. D'Souza
Department of Computer Science & Engineering,
St. Joseph Engineering College, Mangalore, India
e-mail: sunilk@sjec.ac.in

R. D'Souza
e-mail: riod@sjec.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_34

paid for the service got from the cloud. For example, Storage access got from Amazon S3 is a functional requirement, whereas the price and reliability of the service are its non-functional requirement.

Cloud services are provided by cloud providers and are delivered over the Internet to users. Cloud service providers/consumers will have to monitor the QOS parameters of cloud service continuously so that a guaranteed service level can be provided/consumed. Also the QOS level needs to be negotiated well in advance parallel to fixing the price of the service. This process may be automated. Any irregularities (like reliability falling below 90%) can then be reported by the opposite party and claim service credits for the lack of quality of the service.

In this paper we focus on automated QOS negotiation. There are many Negotiation strategies available for use in QOS negotiation like trade-off and concession where we assume incomplete information of opponent reserve price [1]. Also we suggest that there are few others like fuzzy logic-based concession strategy (FLC) and Bayesian updation-based negotiation strategy for learning the opponent reserve price in case of incomplete information [2]. The concession strategy starts from initial preferred value and goes on making concession in each negotiation round effectively reducing the total utility of the proposal. Whereas the trade-off strategy demands more on its more important issue and yields on its less important attribute, keeping the total utility same and generating a proposal more attractive to the opponent. Clearly, trade-off should excel all other strategy. In this paper we elaborate on fuzzy logic-based concession strategy and show that it is better than simple concession strategy [1] and sometimes excels pure trade-off.

The remainder of the paper is divided into 6 sections. Section 2 details on the related work. Section 3 is the actual problem scenario. Section 4 describes already existing negotiation solution—tradeoff and simple concession. Section 5 proposes our Fuzzy logic-based concession strategy and Sect. 6 compares all the strategies and Section 7 gives the conclusion and future work.

2 Related Work

Pan et al. [3] suggest bilateral multi-step monotonic concession negotiation protocol in which software agents on behalf of service consumer and provider act to reach a mutual agreement on some QOS (non-functional attributes like reliability, response time, reputation) values. Author [3] also suggests beginning negotiation with best offer (max profit) and making concession in subsequent rounds. Ferretti et al. [4] refutes that SLA monitoring will lead to more optimal use of resources by dynamic load balancing. The work speaks about maintaining a monitoring service in load balancer which consults SLA policy engine before a add resource request is generated. In Freitas et al. [5] the provider prepares a set of SLA templates (based on performance and fault tolerance) and then there is an agreement reached between customer and service provider based on this. Then once requests are honored, the QOS assurance mechanisms ensure that the QOS mentioned in SLA templates are adhered to (by replacing failed and delayed jobs and by booking resources, configuring, deploying the service instances). Dastjerdi et al. [6] summarizes that operations such as service discovery, scaling, and monitoring are accomplished automatically but negotiation is bottleneck if performed by humans. [6] Proposed negotiation is capable of assessing reliability of offers received from cloud providers. Also offers are generated keeping in mind the resource utilization and it concedes more on price of less utilized resources. Zheng et al. [1] refutes that negotiation is only way to resolve conflicts between service provider and consumer. Here trade-off approach outperforms concession ones in terms of individual utility and social benefit but underperforms in terms of success rate. Ortiz et al. [7] argues the need for personalized SLAs for cloud customers in the presence of multiple cloud providers each with their own pricing model and SLAs.

3 The Actual Problem Scenario

Table 1 below depicts the conflicts between service consumer and service provider in terms of two issues reliability and price. The problem scenario is assumed for Amazon S3 deployment.

The reliability can range from 0 to 100% (percentage per month) and price from 0 to 0.1 (dollars/GB of data per month). Now consider reliability of service provider which is 70% initially, even though provider can provide 85%. Service consumer has a reliability requirement of minimum 75%, although it starts negotiation initially at 90%. Clearly, there is a preference gap over reliability (70% vs 90%). Similarly consider price issue. Service provider seeks to maximize his profit so initially expects 0.08 (can come down to 0.05). Whereas service consumer is willing to pay only 0.04 initially (can come up to 0.07). Here too there is a conflict over price (0.08 vs 0.04).

Assumption (incomplete information): There are a few assumptions necessary before we negotiate, i.e. service consumer and service provider keep their values of preferred and reserved value of reliability and price secret. Service consumer knows that service provider cares for storage price more than reliability. Service provider knows that service consumer cares more about reliability than price.

Table 1 A sample starting scenario

	Min	Max	Service provider (seller)			Service consumer (buyer)		
			Res. value	Pref. value	Weight	Res. value	Pref. value	Weight
Rel.	0%	100%	85%	70%	0.1	75%	90%	0.9
Price	0	0.1	0.05	0.08	0.9	0.07	0.04	0.1

4 Existing Negotiation Solution: Tradeoff and Simple Concession

We assume linear utility function for reliability and price. The utility value for a proposal gives the level of satisfaction of the party involved in negotiation. $u_1(x)$ gives the utility for a single issue x whereas $u_2(P)$ gives the overall utility of the proposal P with n issues. The formulas are as follows:

$$u_1(x) = \frac{x - x_{\text{worst}}}{x_{\text{best}} - x_{\text{worst}}} \tag{1}$$

where $0 \leq u1(x) \leq 1$ and $u1(x_{\text{best}}) = 1$ and $u1(x_{\text{worst}}) = 0$.

$$u_2(p) = \sum\nolimits_{i=1}^{n} w_i u_1(x_i) \tag{2}$$

where w_i is the weight of issue x_i and $\sum\nolimits_{i=1}^{n} w_i = 1$.

In the above problem scenario, for a service provider reliability is lower-is-better attribute and price is higher-is-better attribute. So if the preferred proposal of service provider is (70%, 0.08) from Table 1 then u_1(price) = (0.08–0)/(0.1–0) = 0.8 and u_1(reliability) = (70–100%)/(0–100%) = 0.3. So the overall utility of the preferred proposal = 0.9 × 0.8 + 0.1 × 0.3 = 0.75.

A. A sample trade-off scenario

In the algorithm below, we assume incomplete information assumption and we go for a negotiation with two attributes, namely, reliability and price. Agent *i* sends the proposal V to agent *j* at step1. If agent *j* rejects this proposal and *j*'s counterproposal is not acceptable to agent *i* then *i* makes a trade-off from steps 4–23 where agent *i* yields on its less important attribute to create a proposal attractive to opponent (agent *j*) without lowering the total utility of the previous preferred proposal. Thus pareto optimality is achieved in this case. Concession algorithm is also similar to this one but lines 7–13 are replaced by $u_0 = u_0 - (k \times \lambda)\, u_0$ and $u_1 = u_1 - (k \times \lambda)\, u_1$. In this case the total utility gradually reduces with every round of negotiation.

Algorithm 1 Tradeoff (V, W, F, λ)

Input: Array V with Proposal values of attributes

Array W with weights of attribute

Array F with Flags of attributes. A flag indicates whether attribute is a higher-is-better / lower-is-better attribute.

λ indicates the degree of tradeoff at a time.

Output: true if succeed and false otherwise.

Algorithm:

```
agent i sends V to agent j and waits for a response
while (agent j does not accept V and its counter
proposal is not acceptable to agent i)
        k=k+1
        u_o = calculateutility (V[0], W[0], F[0])
        u_1= calculateutility (V[1], W[1], F[1])
        usum= aggregateutility (V, W, F)
        if (W[0] < W[1])
           u_0 = u_0 – (k×λ) u_0
           u_1= u_sum – u_0
        else
           u_1= u_1 – (k×λ) u_1
           u_0 = u_sum- u_1
        if (u_0< 0 or u_1< 0)
            return (false)
        else
             V[0] = restore (u_0, W[0], F[0])
             V[1] = restore (u_1, W[1], F[1] )
        if (proposal V out of bounds)
```

```
            return (false)
        else
            agent i sends proposal V to agent j and
            waits for a response
    end of while
return (true)
```

5 Fuzzy Logic Based Concession Strategy

In the negotiation here, we assume it is a bilateral one-to-one agent negotiation involving two issues, namely, reliability and price. Also we assume alternating-offers protocol between buyer and seller agents as shown in Fig. 1.

As seen in Fig. 2, the reasoning model mainly concentrates on negotiation. Offer evaluation block is fed with inputs (incoming offer and counter offer). It evaluates the offers using Fuzzy Inference System from agent’s point of view and then informs the decision-making block about the acceptance accordingly. Then the

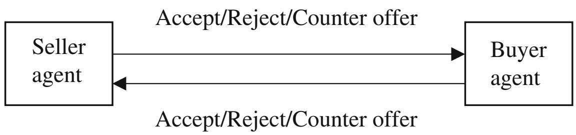


Fig. 1 Alternating-offers protocol

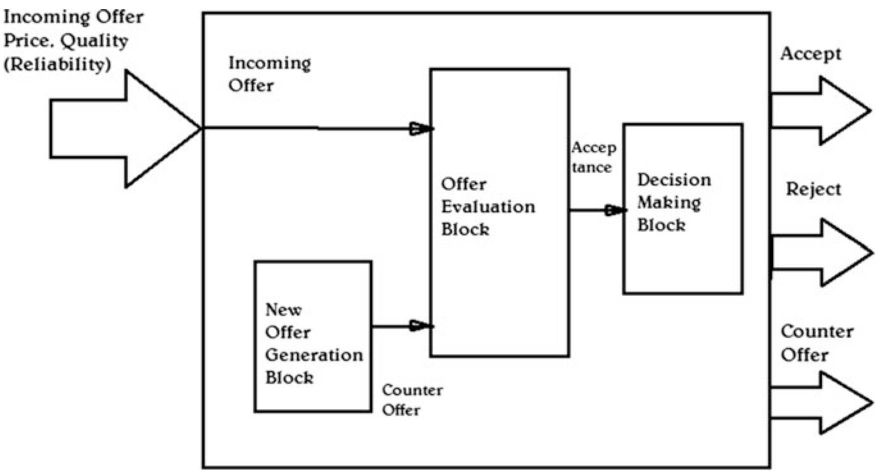


Fig. 2 Block diagram of offer evaluation block of an agent based on fuzzy logic

decision-making block makes the decision of accepting the incoming offer, rejecting the incoming offer or sending counteroffer [8].

A. New offer generation block

Here the new offer generation block is modelled as Distributed Fuzzy Constraint Satisfaction problem. The major work of new offer generation block is to find the concession value Δx_i for the next round of negotiation for each issue i. The first step is to find the concession space (CS) for agent A which is given below

$$\mathrm{CS}_{i,n}^{A} = \frac{(10 - w_{i,n}^{A})}{\left(10 - w_{i,n}^{A}\right) + (10 - w_{i,n}^{B})} * \left| x_{i,n}^{A} - x_{i,n}^{B} \right| \tag{3}$$

where A and B are the buyer and seller agents under consideration $\left| x_{i,n}^{A} - x_{i,n}^{B} \right|$ denotes the distance between the value of issue i between agents A & B. n represents the negotiation round. $w_{i,n}^{A}$ and $w_{i,n}^{B}$ denotes the level of preference on the issues by agents A and B. $\mathrm{w}_{\mathrm{i,n}}^{\mathrm{B}}$ for agent A is unknown. Therefore a guess is required to calculate $w_{i,n}^{B}$

$$W_{i,n}^{B} = \left\{ \begin{array}{ll} \min\left(w_{i,n-1}^{B} + 0.5, 9\right) & \text{for } r < 1 \\ \max\left(w_{i,n-1}^{B} - 0.5, 0\right) & \text{for } r > 1 \\ w_{i,n-1}^{B} & \text{for } r = 1 \end{array} \right\} \tag{4}$$

$$\text{Where} \quad r = \frac{\left| x_{i,n}^{B} - x_{i,n-1}^{B} \right|}{\left| x_{i,n}^{A} - x_{i,n-1}^{A} \right|} \tag{5}$$

Second step is to choose an appropriate value in the concession space $\mathrm{CS}_{i,n.}^{A}$ We have chosen it to be 20% of $\mathrm{CS}_{i,n}^{A}$ from Eq. (3).The new offer on the particular issue i is got by the following formula

$$x_{i,n+1}^{A} = x_{i,n}^{A} + \Delta x_{i,n}^{A} \tag{6}$$

In the above equation, the $x_{i,n}^{A}$ is got from preferred values supplied by the user to the agent A. Apply Eqs. (3)–(6) separately for reliability (i = 1) and price (i = 2) and get the counter offer value <$x_{1,n+1}^{A}$, $x_{2,n+1}^{A}$>.

B. Offer Evaluation Block

The counteroffer <$x_{1,n+1}^{A}$, $x_{2,n+1}^{A}$> being generated using the method mentioned above will be evaluated by comparing with incoming offer. To ease the situation, Fuzzy Inference System has been adopted. The main purpose of Fuzzy logic in our strategy is to indicate the acceptance of an opponent proposal. The acceptance using fuzzy inference process for one offer (either incoming offer/counter offer) consists of

- Fuzzification of input variables (Reliability and Price).
- Application of Fuzzy operator (AND and OR) on the fuzzified inputs.
- Apply implication method (min or prod) to get fuzzy set (before implication is applied, rules must be written for acceptance for both buyer and seller).
- Aggregate all outputs (Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set say using max method).
- Defuzzify the output (i.e. generate a single crisp value for acceptance say 70.5%)

The acceptance of incoming offer and counteroffer generated using the method above is transferred to Decision block for further processing.

C. Decision block

The following formula summarizes the decision-making process.

$$\left\{\begin{array}{c} ACCEPTANCE_{INCOMING} < ACCEPTANCE_{MINIMUM}\ REJECT \\ ACCEPTANCE_{INCOMING} \geq ACCEPTANCE_{COUNTEROFFER}\ ACCEPT \\ ACCEPTANCE_{INCOMING} < ACCEPTANCE_{COUNTEROFFER}\ COUNTEROFFER \end{array}\right\} \tag{7}$$

Here the $ACCEPTANCE_{MINIMUM}$ for an agent is got from reserve value specified for an agent. If acceptance of incoming offer is less than the minimum acceptance for negotiation then reject the offer. If acceptance of incoming offer is greater the acceptance of counteroffer then accept the incoming offer otherwise send the counteroffer to the opponent.

Finally to report the utility of the offer we need utility function. Since we are considering multi-issue agent negotiation, we can calculate the individual issue utility using our basic utility function and then the weighted (to be given by the user) sum gives the total utility of the offer for agent A. Similar calculation is done on agent B side and total overall utility is the sum of these two utilities.

Basic utility function:

$$U_i(x_i) = \frac{x_i - X^i_{\min}}{X^i_{\max} - X^i_{\min}} \tag{8}$$

where $[X^i_{\min}, X^i_{\max}]$ forms the range over which x_i operates.

Algorithm 2 FLC ( ) // Fuzzy Logic based Concession

```
create and initialize the fuzzy Inference system to deal
with acceptance of offer for buyer & seller separately.
calculate minimum buyer acceptance w.r.t. buyer FIS.
calculate minimum seller acceptance w.r.t seller FIS.
While (true)
 if (its buyer's turn)
       Map the Preferred proposal of buyer to buyer FIS
       to get the buyeracc
      Map the incoming offer (from seller) to buyer FIS
      to get inbuyeracc
      if (buyeracc < minbuyeracc)
            Report negotiation failure
      else if (inbuyeracc > buyeracc)
            Report negotiation success & print seller
            Offer to user
      else
            Generate the counter offer based on eq.(3)-(6)
            Send it to seller
if (its seller's turn)
     Map the Preferred proposal of seller to seller FIS
     to get the selleracc
     Map the incoming offer (from buyer) to seller FIS
     to get inselleracc
     if (selleracc < minselleracc)
            Report negotiation failure
     else if (inselleracc > selleracc)
            Report negotiation success & print buyer offer to user
     else
          Generate the counter offer based on eq.(3)-(6)
```

```
            Send it to buyer
    end of while
    end of procedure
```

6 Comparison of Negotiation Strategies

The Trade-off, Simple Concession and FLC simulations are carried out using Dell Inspiron laptop with Intel Core i3 processor with a speed of 2.27 GHZ 2 GB RAM on Microsoft Windows 7 Ultimate operating system. The time complexity of the simple concession is O $(1/\lambda)$ and that of FLC is dependent on Δx_i chosen in each round of negotiation, whereas the trade-off does not converge for some inputs.

For comparison we have same dataset as in [1] (as shown in Table 2) so that there is uniformity in comparison done or conclusions drawn. In the starting scenario S1–S10, a difference of 20% in reliability and 0.04 in price is maintained between preferred proposal of buyer and seller. Similarly in starting scenario S11–S19, a difference of 15% and 0.03 in reliability and price of preferred proposal is maintained between buyer and seller. S20 a difference of 35% in reliability and 0.07 in price of preferred proposal is maintained just to test the trade-off strategy for success rate.

In Table 2 all the values are normalized, i.e. reliability varies between 0 and 1 and price varies between 0 and 0.1. The weights for reliability and price are in the range 0–1 and also, i.e. sum of weights for reliability and price = 1.

A. Negotiation Protocol
   An alternating-offers protocol is engaged between the buyer and the seller in the sense first seller is going to send its preferred proposal to buyer. If buyer accepts the offer negotiation succeeds else buyer is going to generate a counterproposal based on Algorithm 1 or Algorithm 2 (trade-off or FLC) and repeat. Negotiation progresses in this way till either of the parties accept opponent proposal or both seller and buyer are dissatisfied (negotiation failure). λ in algorithm 1 is assumed = 0.1 and concession rate in algorithm 2 is assumed = 0.2.
B. Negotiation Results
   20 starting scenarios are listed in Table 2 and the corresponding results of Negotiation strategies are listed in Table 3. It can be seen that simple concession and FLC succeeds in all 20 scenarios, whereas trade-off reports failure for scenario S20.

Let us analyze the Tables 2 and 3 by considering scenario S5.

From seller's reserved offer <0.70, 0.02>, we calculate its reserved utility as 0.21. From buyer's reserved offer <0.60, 0.04>, we calculate its reserved utility as 0.600. Similarly by looking at seller and buyer preferred offers, we calculated its

Table 2 Negotiation scenarios S1–S20

Scenario		Seller reserve value	Seller prefer value	Seller weights	Buyer reserve value	Buyer prefer value	Buyer weights
S1	Rel.	0.90	0.75	0.1	0.80	0.95	0.9
	Price	0.06	0.09	0.9	0.08	0.05	0.1
S2	Rel.	0.85	0.70	0.1	0.75	0.90	0.9
	Price	0.05	0.08	0.9	0.07	0.04	0.1
S3	Rel.	0.80	0.65	0.1	0.70	0.85	0.9
	Price	0.04	0.07	0.9	0.06	0.03	0.1
S4	Rel.	0.75	0.60	0.1	0.65	0.80	0.9
	Price	0.03	0.06	0.9	0.05	0.02	0.1
S5	Rel.	0.70	0.55	0.1	0.60	0.75	0.9
	Price	0.02	0.05	0.9	0.04	0.01	0.1
S6	Rel.	0.90	0.75	0.1	0.80	0.95	0.9
	Price	0.02	0.05	0.9	0.04	0.01	0.1
S7	Rel.	0.85	0.70	0.1	0.75	0.90	0.9
	Price	0.03	0.06	0.9	0.05	0.02	0.1
S8	Rel.	0.80	0.65	0.1	0.70	0.85	0.9
	Price	0.04	0.07	0.9	0.06	0.03	0.1
S9	Rel.	0.75	0.60	0.1	0.65	0.80	0.9
	Price	0.05	0.08	0.9	0.07	0.04	0.1
S10	Rel.	0.70	0.55	0.1	0.60	0.75	0.9
	Price	0.06	0.09	0.9	0.08	0.05	0.1
S11	Rel.	0.90	0.80	0.1	0.85	0.95	0.9
	Price	0.07	0.09	0.9	0.08	0.06	0.1
S12	Rel.	0.85	0.75	0.1	0.80	0.90	0.9
	Price	0.06	0.08	0.9	0.07	0.05	0.1
S13	Rel.	0.80	0.70	0.1	0.75	0.85	0.9
	Price	0.05	0.07	0.9	0.06	0.04	0.1
S14	Rel.	0.75	0.65	0.1	0.70	0.80	0.9
	Price	0.04	0.06	0.9	0.05	0.03	0.1
S15	Rel.	0.70	0.60	0.1	0.65	0.75	0.9
	Price	0.03	0.05	0.9	0.04	0.02	0.1
S16	Rel.	0.90	0.80	0.1	0.85	0.95	0.9
	Price	0.03	0.05	0.9	0.04	0.02	0.1
S17	Rel.	0.85	0.75	0.1	0.80	0.90	0.9
	Price	0.04	0.06	0.9	0.05	0.03	0.1
S18	Rel.	0.80	0.70	0.1	0.75	0.85	0.9
	Price	0.05	0.07	0.9	0.06	0.04	0.1
S19	Rel.	0.75	0.65	0.1	0.70	0.80	0.9
	Price	0.06	0.08	0.9	0.07	0.05	0.1
S20	Rel.	0.70	0.60	0.1	0.85	0.95	0.9
	Price	0.07	0.09	0.9	0.04	0.02	0.1

Table 3 Experimental results for S1–S20

Scenario	Simple concession				Tradeoff				FLC			
	Offer		Total utility		Offer		Total utility		Offer		Total utility	
	Rel.	Price	U_{seller}	U_{buyer}	Rel.	Price	U_{seller}	U_{buyer}	Rel.	Price	U_{seller}	U_{buyer}
S1	0.80	0.07	1.416		0.82	0.09	1.587		0.75	0.09	1.52	
	Sel → Buy[1]		0.668	0.748	Sel → Buy		0.835	0.752	Sel → Buy		0.835	0.685
S2	0.76	0.06	1.32		0.91	0.05	1.347		0.88	0.05	1.348	
	Sel → Buy		0.600	0.720	Buy → Sel		0.477	0.870	Buy → Sel		0.504	0.844
S3	0.72	0.06	1.224		0.72	0.07	1.342		0.71	0.07	1.313	
	Sel → Buy		0.532	0.692	Sel → Buy		0.665	0.677	Sel → Buy		0.647	0.665
S4	0.68	0.05	1.128		0.68	0.06	1.231		0.70	0.06	1.22	
	Sel → Buy		0.464	0.664	Sel → Buy		0.580	0.651	Sel → Buy		0.548	0.674
S5	0.64	0.04	1.032		0.64	0.05	1.120		0.70	0.05	1.138	
	Sel → Buy		0.396	0.636	Sel → Buy		0.495	0.625	Buy → Sel		0.454	0.684
S6	0.76	0.03	1.032		0.97	0.03	1.20		0.90	0.05	1.298	
	Buy → Sel		0.276	0.756	Buy → Sel		0.255	0.945	Buy → Sel		0.434	0.864
S7	0.76	0.05	1.192		0.92	0.04	1.222		0.87	0.05	1.306	
	Sel → Buy		0.456	0.736	Buy → Sel		0.332	0.890	Buy → Sel		0.474	0.831
S8	0.72	0.06	1.224		0.72	0.07	1.342		0.70	0.07	1.313	
	Sel → Buy		0.532	0.692	Sel → Buy		0.665	0.677	Sel → Buy		0.647	0.665
S9	0.68	0.06	1.256		0.68	0.08	1.391		0.70	0.08	1.382	
	Sel → Buy		0.608	0.648	Sel → Buy		0.760	0.631	Sel → Buy		0.728	0.654
S10	0.64	0.07	1.288		0.64	0.09	1.440		0.69	0.09	1.461	
	Sel → Buy		0.684	0.604	Sel → Buy		0.855	0.585	Buy → Sel		0.829	0.632
S11	0.86	0.06	1.392		0.96	0.07	1.547		0.80	0.09	1.56	
	Sel → Buy		0.581	0.811	Buy → Sel		0.652	0.895	Sel → Buy		0.830	0.730
S12	0.80	0.06	1.352		0.83	0.08	1.507		0.75	0.08	1.44	
	Sel → Buy		0.596	0.756	Sel → Buy		0.745	0.762	Sel → Buy		0.754	0.75

(continued)

Table 3 (continued)

Scenario	Simple concession				Tradeoff				FLC			
	Offer		Total utility		Offer		Total utility		Offer		Total utility	
	Rel.	Price	U_{seller}	U_{buyer}	Rel.	Price	U_{seller}	U_{buyer}	Rel.	Price	U_{seller}	U_{buyer}
S13	0.76	0.06	1.256		0.86	0.05	1.307		0.84	0.05	1.281	
	Sel → Buy		0.528	0.728	Buy → Sel		0.482	0.825	Buy → Sel		0.476	0.804
S14	0.72	0.05	1.160		0.72	0.06	1.262		0.78	0.05	1.236	
	Sel → Buy		0.460	0.700	Sel → Buy		0.575	0.687	Buy → Sel		0.487	0.748
S15	0.68	0.04	1.064		0.68	0.05	1.151		0.71	0.05	1.150	
	Sel → Buy		0.392	0.672	Sel → Buy		0.490	0.661	Buy → Sel		0.458	0.691
S16	0.76	0.04	1.096		0.97	0.04	1.262		0.91	0.05	1.310	
	Buy → Sel		0.348	0.748	Buy → Sel		0.327	0.935	Buy → Sel		0.438	0.871
S17	0.80	0.05	1.224		0.92	0.04	1.284		0.88	0.05	1.316	
	Buy → Sel		0.452	0.772	Buy → Sel		0.404	0.880	Buy → el		0.477	0.838
S18	0.76	0.06	1.256		0.86	0.05	1.307		0.84	0.05	1.281	
	Sel → Buy		0.528	0.728	Buy → Sel		0.482	0.825	Buy → Sel		0.476	0.804
S19	0.72	0.06	1.288		0.72	0.08	1.422		0.70	0.08	1.391	
	Sel → Buy		0.604	0.684	Sel → Buy		0.755	0.667	Sel → Buy		0.733	0.653
S20	0.38	0.07	1.048		—-	—-	0		0.72	0.09	1.473	
	Buy → Sel		0.674	0.374	————		0	0	Sel → Buy		0.814	0.659

[1]Sel stands for seller and Buy for buyer and Sel → Buy stands for last negotiated offer was generated by Seller and accepted by buyer and vice versa

preferred utility as 0.495 and 0.765 respectively (not shown in Table 2 or 3). Now observing the S5 row of Table 3, we can say for Simple Concession Negotiated utilities (i.e. U_{seller} & U_{buyer}) 0.396 > 0.21 and 0.636 > 0.600 are better than reserved utilities calculated above. Similarly for tradeoff Negotiated utilities 0.495 > 0.21 and 0.625 > 0.600 are better than reserved utilities. The same way for FLC negotiated utilities 0.454 > 0.21 and 0.684 > 0.600 are better than reserved utilities. On the whole we can make the following observation

Observation 1: All the three strategies, namely Simple concession, Trade-off and FLC are capable of generating offers whose received utilities are better than reserved utilities.

Let us consider starting scenario S2 (see Table 3). When both parties switch from Simple concession to tradeoff approach, the buyer utility increased from 0.720 to 0.870, whereas the seller utility decreased from 0.600 to 0.477. Here its clear final offer was generated by buyer (Buy → Sel) so buyer benefits.

Same is the case with FLC scenario S2. If both parties switch from Simple concession to FLC approach, the buyer utility increased from 0.720 to 0.844, whereas the seller utility decreased from 0.600 to 0.504. Here too the final offer was generated by buyer (Buy → Sel) so buyer benefited. But consider S3, for trade-off, seller utility increased from 0.532 to 0.665 whereas buyer utility decreased from 0.692 to 0.677. Here Seller (Sel → Buy) acts as a successful offerer. Similarly consider S3 for FLC, seller utility increased from 0.532 to 0.647 whereas buyer utility decreased from 0.692 to 0.665. Here Seller (Sel → Buy) acts as a successful offerer.

Observation 2: Trade-off and FLC give better utilities for successful offerer be it buyer/seller.

Now consider S1 (see Table 3), where for trade-off, both seller/buyer utilities increased from 0.668/0.748 to 0.835/0.752. Similarly S4 for FLC, both seller/buyer utilities increased from 0.464/0.664 to 0.548/0.674. So both seller/buyer utilities increased when switching to trade-off/FLC. By analyzing S1–S20, we have found that switching to trade-off, 35% of the times only buyer benefit, 55% of the time only Seller benefit, 5% of the time both buyer as well as seller benefit and 5% of the time negotiation failure reported. Similarly by analyzing S1–S20, we have found that switching to FLC, 15% of the times only buyer benefit, 30% of the time only seller benefit and 55% of the time both buyer as well as seller benefit.

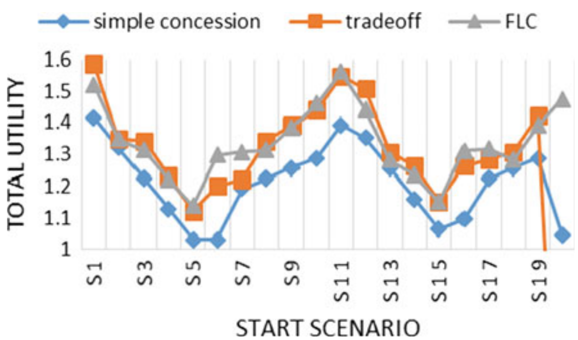


Fig. 3 Comparison of the three negotiation strategies

Observation 3: In 55% of cases, FLC gave better utilities for both buyer and seller when compared to only 10% in case of trade-off.

Now consider the total utility column of Simple concession, trade-off and FLC in Table 3. Clearly total utility column value for trade-off and FLC is better than the one for Simple concession for all scenarios S1–S19. Say, for example S1, total utility of Simple concession is 1.416, trade-off is 1.587 > 1.416 and FLC is 1.52 > 1.416. The graph in Fig. 3. Clearly, shows the curves for Trade-off (■) and FLC (▲) are higher than Simple concession (♦).

Observation 4: FLC and Trade-off give better overall social benefit than Simple concession.

Let us examine the scenario S20. Tradeoff does not seem to converge even after 20 rounds or so of negotiation. Hence negotiation fails in this case. Simple concession seem to give a total utility of 1.048 with an offer of 38% (reliability) and 0.07 (price).FLC seems to give even better total utility of 1.493 with an offer of 72% (reliability) and 0.09 (price). So the success rate of trade-off is 95% (1 out of 20 is a failure). The incomplete information assumption makes it difficult for trade-off to move towards an agreement whereas concession (both simple concession & FLC) have 100% success rate. Now consider S10, total utility of Simple concession is 1.288, whereas that of trade-off and FLC is 1.44 and 1.461. Clearly, total utility of FLC is highest of all. This is the case in scenarios S2, S5–S7, S10–S11, S16–S17 and S20. It means 9/20 cases (45%) FLC is superior even to trade-off.

Observation 5: 45% of cases FLC gives better social benefit than trade-off while maintaining 100% success rate.

7 Conclusion and Future Work

Given the scope of Cloud Computing, there is an increasing demand for hardware and software to promote this technology. This in turn raises the demand for negotiation in pricing these datacenter hardware or software to benefit both cloud provider and consumer. We have suggested throughout the paper that an automated component can do this work. We have taken note of the existing Simple concession and Trade-off strategies and we have proposed a new Fuzzy logic-based concession strategy (FLC).

From the various observations across all three strategies we conclude that FLC, though a concession strategy, has outperformed Trade-off and Simple concession in terms of social benefit and success rate. Also in 55% of the cases (S1–S20), FLC has benefitted both buyer and seller, whereas trade-off could achieve this in only 10% of the cases. Hence our strategy FLC is the most effective and efficient approach for solving the automated negotiation of price and reliability in cloud computing scenario.

Our future work involves devizing a few more negotiation strategies for cloud environment and also to dynamically switch between different strategies during negotiation based on past performance of these strategies.

References

1. Zheng, Xianrong, Patrick Martin, and Kathryn Brohman. "Cloud service negotiation: Concession vs. tradeoff approaches." In Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), pp. 515–522. IEEE Computer Society, 2012.
2. Gwak, Jeonghwan, and Kwang Mong Sim. "Bayesian learning based negotiation agents for supporting negotiation with incomplete information." In Proceedings of the international multiconference of engineers and computer scientists, vol. 1, pp. 163–168. 2011.
3. Pan, Li. "Towards a ramework for automated service negotiation in cloud computing." In Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on, pp. 364–367. IEEE, 2011.
4. Ferretti, Stefano, Vittorio Ghini, Fabio Panzieri, Michele Pellegrini, and Elisa Turrini. "QOS–aware clouds." In Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, pp. 321–328. IEEE, 2010.
5. Freitas, André Lage, Nikos Parlavantzas, and J-L. Pazat. "An integrated approach for specifying and enforcing slas for cloud services." In Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, pp. 376–383. IEEE, 2012.
6. Dastjerdi, Amir Vahid, and Rajkumar Buyya. "An autonomous reliability-aware negotiation strategy for cloud computing environments." In Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on, pp. 284–291. IEEE, 2012.
7. J. Ortiz, V. T. De Almeida, and M. Balazinska, "A vision for personalized service level agreements in the cloud," in Proceedings of the Second Workshop on Data Analytics in the Cloud. ACM, 2013, pp. 21–25.
8. Wang, Xin. "A fuzzy logic based intelligent negotiation agent." (2006)

Intrusion Detection System Based on BCS-GA in Cloud Environment

Partha Ghosh, Supriya Jha, Rahul Dutta and Santanu Phadikar

Abstract Cloud computing provides services to the user through the Internet and is cost-efficient. It comprises of a wide range of users that makes the security of data a matter of concern. Intrusion Detection System (IDS) is a multipurpose security system either in hardware or software form applied to a network that detects abnormality in packets. Detecting intrusions involve observing any changes from normal behaviour. Cloud contains huge amount of data for which high storage space is required. To reduce the required space only relevant data is stored. This is done applying feature selection procedure. Hereby, we have proposed an IDS functioning on the concept of feature selection from NSL-KDD dataset. The efficient IDS is based on our proposed algorithm BCS-GA and gives better accuracy.

Keywords Cloud Computing (CC) · Intrusion Detection System (IDS) Feature selection · Cuckoo Search (CS) · Genetic Algorithm (GA) Neural Network (NN) · NSL-KDD dataset

1 Introduction

Development of Cloud computing has aroused as a multifaceted technology with the capability to support a broad spectrum of applications [1]. It emerged as a breakaway in usage of Internet. Cloud computing is now a topic of great impact and

P. Ghosh (✉) · S. Jha · R. Dutta
Department of Information Technology, Netaji Subhash Engineering College, Kolkata, India
e-mail: partha1812@gmail.com

S. Jha
e-mail: supriya.sneha286@gmail.com

S. Phadikar
Depatment of Computer Science and Engineering, Maulana Abul Kalam Azad University of Technology, Kolkata, India
e-mail: sphadikar@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_35

has proved itself as a driver for small companies in rapidly developing world. It is an anatomy for providing various beneficial services using the Internet. It is accessible on demand and furnishes pay per use access to a puddle of shared resources [2]. Service providers in a Cloud can be basically categorized in three types as per the basic models deployed—Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) [3]. IaaS comprises of highly automated and scalable computer resources accompanied with storage in a Cloud and network empathy which can be available on demand. PaaS provides a rostrum on which specific software can be deployed or evolved. SaaS is much similar to the technique of software furnishing used erstwhile [4]. Since Cloud computing has a distributed nature with a large number of users and domains attached to it, it is highly prone to malicious activities. Information is of immense importance in any organization and thus security measures such as firewalls are deployed in each system to detect security threats [5]. Intrusions are actions that threaten the integrity, authenticity of information in a network resource by circumventing the measures provided for security [6]. The basic characteristics of an Intrusion Detection System (IDS) comprise of monitoring, analysing user and system proceedings, scrutinizing system appearance, abnormal activity analysis, etc. IDS provides trace to user entry from point of entry to point of impact, detect error in system configuration, recognize and reports alterations on any piece of data. Two prominent methodologies are used by IDS to detect any threat or attack and generate alert-Misuse detection & Anomaly detection. Misuse detection also referred to as rule-based detection is easy to configure. However, despite of its accuracy it may not always detect an attack with even slight variations. Anomaly detection observes packets for abnormal behaviour and generates alert [7]. On the basis of working environment, IDS can be categorized in two categories—Host-based Intrusion Detection System (HIDS) and Network-based Intrusion Detection System (NIDS). Exercising separate IDS for every host within a network refers to the concept of HIDS. It takes a quick view of prevailing system files and matches it with the latter one at each occurrence. If perilous system files were amended or deleted, alert is generated. NIDS is applied to a subnet. It analyzes the passing traffic on the entire subnet and has a wider scope of detecting unknown attacks. For detecting an attack a rule has to be specified or a specific characteristic has to be noted. A specific trait or characteristic of data is generalized as a feature. Feature selection forms a subset of features which accurately describe a problem [8]. Various techniques have been applied for feature selection. Genetic algorithm uses the principles of Darwin's theory for optimization and/or search problems and is has its origin in natural genetics [9]. Further, Cuckoo search algorithm is a heuristic algorithm originating from the reproductive behaviour of some cuckoo species [10]. In this paper we have proposed an algorithm for feature selection which combines the advantages of both Cuckoo Search (CS) and Genetic Algorithm (GA).

2 Related Work

A. Wasim and M. Salama [11] discussed the concept of Cloud computing. Cloud computing provides services to users enriched with the properties of elasticity and smooth scalability. B.Santosh Kumar et al. [12] explained the gist of intrusion detection. Intrusion detection can be briefly depicted as continuous observing the events of a particular system and analysing any signs of malicious activities. M. Tuba et al. [13] proposed a redesigned version of Cuckoo Search algorithm for unconstrained problem optimization. They calculated the step size of random walks by different function. That function utilized sorted array and produced better results. Li Xiangtao and Yin Minghao [14] developed a modified form of Cuckoo Search algorithm, having a self adaptive parameter. Their work was inspired by the concept of differential evolution and proposed a concept of statistical learning strategy. M. Naik et al. [15] proposed a new adaptive form of Cuckoo Search algorithm. Their effort was to fuse the step size as proportional to the fitness of a particular nest in the total search space and current generation. Also, step size calculation function is altered. That algorithm gave productive results when tested with 28 benchmark functions. In 2015, Self-adaptive Cuckoo Search algorithm for hybrid flowshop makespan problem was proposed by H. Zhonghua and Lv Xisheng [16]. Their algorithm worked on the fact that performance of Cuckoo Search depends on its parameters. They developed an algorithm which adapts the parameters on the basis of evolution processes automatically. Their algorithm used elite learning mechanism which overcame the drawback of premature convergence. X.Yan et al. [17] proposed an improved version of Genetic Algorithm. Their algorithm worked on the drawback of traditional GA, i.e. getting trapped in local minima easily. S. Ricardo et al. [18] proposed a binary form of Cuckoo Search used in solving the problem of set covering. Mujahid Tabassum and Kuruvilla Mathew [19] performed a Genetic Algorithm analysis for optimization. They applied GA on several real-world problems. Using the inferences of the above-mentioned papers we have proposed an efficient Intrusion Detection System.

3 Preliminary Study

High-dimensional datasets require abundant memory space for storage. The training time required for large datasets is also found to be quite high. Feature selection is a technique intended to form a subset of features such that redundant parts are removed [20]. Feature selection process works through two methodologies—filter-based approach and wrapper-based approach [21]. Feature selection involves the formation of subset of features, constituting the most useful features which produce higher accuracy. Different methods can be effectively applied for feature selection. Cuckoo Search is a technique applied and has its roots from the brood parasitism of cuckoo birds. Genetic Algorithm is another prominent method used for creation of feature subset. The basic idea behind GA is natural evolution and works upon solution set referred to as chromosomes to find the results.

a. Cuckoo Search (CS)

Cuckoo Search is a heuristic algorithm that works over the characteristics of cuckoo birds. The reproductive strategy of cuckoo bird is spectacular. Some species of cuckoo birds have their reproductive strategy engaged with obligate brood parasitism. They lay their eggs in variant host nests which may be of discrete types. The host bird may identify the eggs of cuckoo bird resulting in abandoning of the nest or removal of the alien eggs. Cuckoo eggs are found to be hatching early as per studies conducted. As soon as an egg is hatched they try to increase their probability of survival by removing the host eggs. Cuckoo offsprings may also impersonate the sound of host bird to gather maximum food. Cuckoo Search is characterized by three rules [22]: each cuckoo lays one egg at a time and unload it in a randomly chosen host nest, second the best nest with superior standard of eggs will proceed to the future, and aftermost the number of host nest is invarying and the possibility of discovering the alien eggs by the master of the nest is p [0,1]. On discovering alien eggs the host bird has an alternative of either thump the egg or abandons the complete nest. CS also applies the knowledge of levy flight. Levy flight can be described as a random walk where the step size is disseminated as per a specific criterion [23]. Through levy flight concept the solution is searched. CS algorithm performs random walk via levy flight. This is highly productive in traversing the entire search area. CS algorithm comprises of quite few tuning parameters and therefore can be applied to a broad range of optimization problems. CS has commendable ability of contraction and reaches the global optimum. In the overall procedure each egg laid by cuckoo represents a solution. New solution is generated using the equations depicted below cited from [24]:

$$V_{pq}{}^{t+1} = V_{pq}{}^{t} + s_{pq} * \mathrm{Levy}(\lambda) * \alpha \tag{1}$$

$$\mathrm{Levy}(\lambda) = \left| \frac{\Gamma(1+\lambda) * \sin\left(\frac{\pi * \lambda}{2}\right)}{\Gamma\left(\frac{1+\lambda}{2}\right) * \lambda * s^{\frac{(\lambda-1)}{2}}} \right|^{1/\lambda} \tag{2}$$

where λ is a constant ($1 <= \lambda <= 3$) and α is a random number generated between -1 and 1.

The step size is calculated using the equation mentioned below:

$$s_{pq} = V_{pq}{}^{t} - V_{fq}{}^{t} \tag{3}$$

where p,f $\epsilon\{1,2,\ldots m\}$ and q$\epsilon\{1,2,\ldots.D\}$ are randomly chosen indices.

The host bird learns about the alien egg with the probability value linked with equation number (4)

$$\mathrm{Pro_p} = \left(0.9 * \mathrm{Fit_p}/\max(\mathrm{Fit})\right) + 0.1 \tag{4}$$

In case the egg is not identified it grows and remains alive based on the fitness function mentioned below:

$$\mathrm{x_p} = \mathrm{x_{pmin}} + \mathrm{rand}(0, 1) * \left(\mathrm{x_{pmax}} - \mathrm{x_{pmin}}\right) \tag{5}$$

b. Genetic Algorithm (GA)

Genetic Algorithm was first proposed in its binary form by John Holland in 1970. GA is a search and optimization skill including natural selection as its functioning base [25]. Genetic algorithm works on population (chromosomes) which are sequence of binary digits i.e. 0 and 1. GA is an instructional, dynamic algorithm used in various optimization problems that gives efficient search results [26]. Genetic algorithms perform a parallel search as opposed to point by point search. GA has high ability to perform local search. Also it uses fitness values for satisfying the survival of fittest criteria which produces better output. The chromosomes are randomly generated and then the process follows to produce the desired results using genetic operators-

Selection: The stipulation of two parent chromosomes as per their respective fitness values. Higher fitness value leads to greater chance of being selected.

Crossover: Production of new offspring (children) through crossover of two parents.

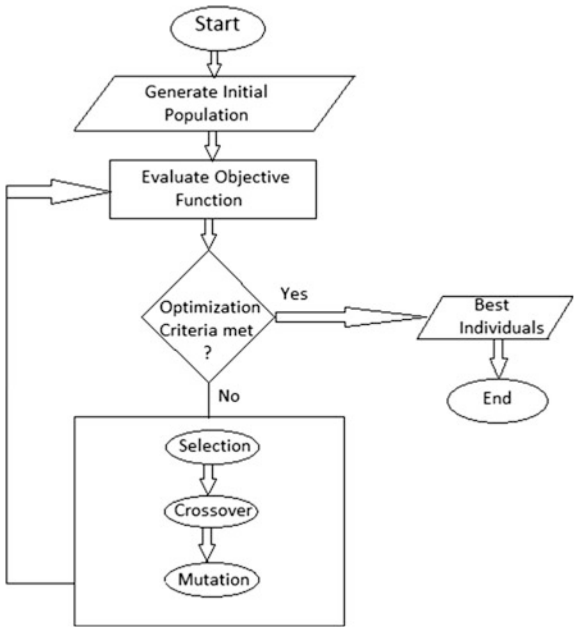


Fig. 1 Flowchart of Genetic Algorithm

Mutation:- Mutation is a genetic operator which alters each gene independently to form a fresh set of chromosomes and maintain the diversity.

Figure 1 shows the different steps of Genetic Algorithm.

4 Proposed Model

Cloud has evolved as a technology that supports secure data storage and other vital facilities through the Internet. Cloud computing concept deals with a large number of virtualized systems. It is available to multiple users on a pay and use basis. As it deals with a vast range of users it is considered to be highly vulnerable to security issues. The common safety measures of having a firewall, encryption etc. prove to be insufficient in dealing with all kinds of malicious activities. Thus for security system of data on a Cloud, IDS is applied. IDS have been widely used to detect any malicious behaviour. We have tried to form an IDS for inculcating efficient security of data stored in Cloud. High-dimensional data requires large storage and acquires more time in training. Thus we have proposed an algorithm for reducing the data using feature selection.

Our proposed algorithm combines the advantages of both Cuckoo Search (CS) and Genetic Algorithm (GA). The algorithm BCS-GA (Binary Cuckoo Search with Genetic Algorithm) as the name suggests, works on the concepts of Cuckoo Search in its binary form and also applies Genetic Algorithm procedures. We have used NSL-KDD dataset in our experiment. NSL-KDD dataset has two parts—train set and test set. NSL-KDD dataset is high dimensional and requires huge amount of memory space for storage as well as it is time consuming. Thus pre-processing and normalization is required prior to working on the dataset. Dataset pre-processing makes the NSL-KDD dataset usable in our experiment. NSL-KDD dataset comprises of 41 features along with one decision factor for detecting normal or specific attack. In 41 features of dataset, three conditional features and one decision-making feature have non-numeric values which need to be transformed in numeric values. NSL-KDD dataset constitutes of continuous and discrete values. The range of feature value is different. Normalization of dataset results in same range of features. We have applied min-max normalization paradigm in our experiment to consider all the values in the range [0,1]. On the normalized data we have applied BCS-GA for further classification. In this experiment we have reduced the dataset through feature selection procedure so that the obstacles posed by the dataset are overcome. Our experiment, functions on combination of features. The presence of specific features conveys information regarding the particular selected record. Thus binary form of Cuckoo Search algorithm effectively serves our purpose of taking into consideration a combination of features. BCS-GA algorithm forms a feature subset through which classification can be effectively done (Fig. 2).

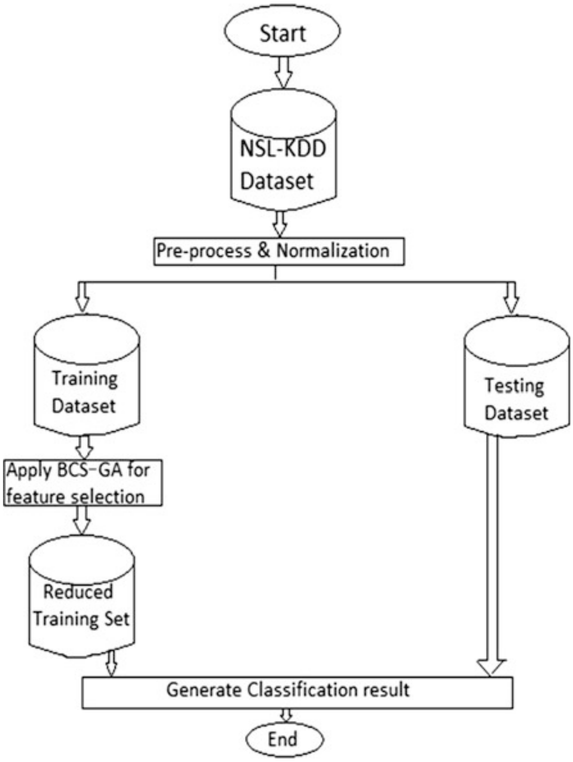


Fig. 2 Flowchart of BCS-GA

BCS-GA has been utilized for formation of appropriate feature subset. Initially with reference to CS, a random population is generated to represent a nest. Fitness value of all the nests is calculated. In search of the best possible outcomes each host nest is modified using levy flight random walk. Further the fitness and probability of survival of each nest is evaluated. The next population generation is governed by the methods of GA. Though Cuckoo Search takes into consideration nests as per their respective fitness values arranged in descending order. But through our proposed algorithm, we have modified this step using Genetic operator, i.e. selection, crossover and mutation. The values are selected randomly despite of their fitness value and modified by the process of crossover and mutation towards a better solution. This transformation has resulted in better results with high accuracy. This algorithm has the beneficiaries of Cuckoo Search method and also the evolutionary mechanism of Genetic Algorithm.

Our proposed Algorithm (BCS-GA) is depicted as follows:

Input: NSL-KDD Training Dataset.

Result: Best Feature Subset.

Begin

Randomly generate initial population and calculate the fitness of all the nests

While (iteration < max_iteration)

do

Modify every host nest using levy flights random walk

Calculate the fitness and probability to survive of each nest

Generate next population using GA

End while

Find the best feature subset using the fitness function.

End.

Our proposed algorithm has been designed in two stages. The first stage described earlier involves the creation of a feature subset. In the next phase several classifiers are applied on the algorithm and results with better accuracy are produced. NSL-KDD dataset gets reduced and a subset of relevant features is formed. Applying both the stages the algorithm accurately works on observing the network packets and detecting attacks.

5 Performance

An efficient Intrusion Detection System (IDS) is needed to classify all the attacks for securing all the Cloud data. A proper training is essential for classification of all the attacks from common conduct. An improper training of IDS will decrease the classification accuracy. Abundance of irrelevant features in the training dataset increases training time and deteriorates classification accuracy. Training of NSL-KDD dataset has been done in this paper. Using our suggested Binary Cuckoo Search with Genetic Algorithm (BCS-GA) several irrelevant features are removed from the training dataset leading to production of efficient training set.

Designing of BCS-GA includes two phases. Proposed BCS-GA is applied on the first phase of this algorithm for feature selection and classifications using different classifiers are included in the second phase. Initially NSL-KDD dataset contains 41 features. If we apply only Binary Cuckoo Search algorithm then it reduced to 24 features. But after applying BCS-GA algorithm, it successfully reduced to 16 features.

In our model, we have used Neural Network (NN) as a classifier and NSL-KDD-Test + dataset for testing the classification accuracy. If the whole train dataset is used for classification then it produced 77.027% accuracy. If only Binary Cuckoo Search is used then it produces 77.324% accuracy. But our proposed BCS-GA algorithm produces 78.229% accuracy. Confusion matrices are depicted as follows (Tables 1 and 2):

Different classifiers have been trained using KDDTrain + and Reduced dataset separately. Further testing of classification accuracy has been performed utilizing KDDTest + dataset. Classification results are shown in Fig. 3.

Table 1 Confusion matrix for testing data using NN classifier (trained by KDDTrain + using 41 features)

	Attack	Normal
Attack	7888	4945
Normal	234	9477

Table 2 Confusion matrix for testing data using NN classifier (trained by Reduced train set using 16 features)

	Attack	Normal
Attack	8223	4610
Normal	298	9413

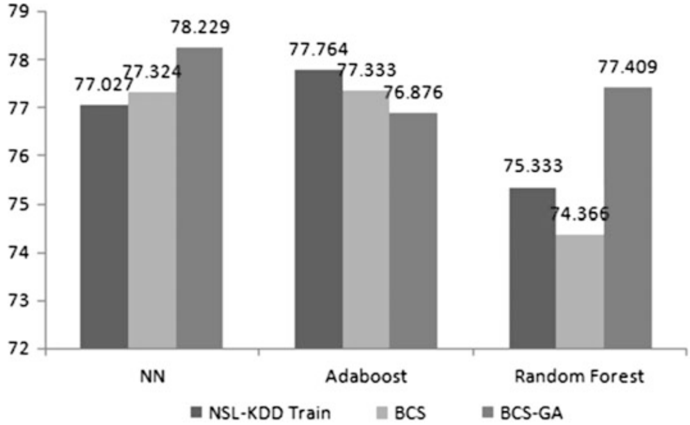


Fig. 3 Classification accuracy using different classifiers on separate datasets

6 Conclusion

In this paper we have designed an Intrusion Detection System which operates on our proposed algorithm. Cuckoo Search algorithm and Genetic algorithm contribute in forming the complete procedure of our algorithm.. With the development of technology it is evident that Cloud is very useful and inseparable part of an efficient network. Securing data in Cloud requires Intrusion Detection System. We have done feature selection through BCS-GA algorithm as the NSL-KDD dataset is very large. It reduces the training time and memory storage space required for the high-dimensional dataset. Classification after forming subset of features is comparatively easier and increases accuracy of results.

References

1. Srinivas, J., Reddy, K.V.S., Qyser, A.M.: Cloud Computing Basics. Int. J. Adv. Research Comput. Commun. Engg. (IJARCCE) 1(5), 343–347 (2013)
2. Padhy, R.P., Patra, M.R., Satapathy, S.C.: Cloud Computing: Security Issues and Research challenges. Int. J. Comput. Sci. Info. Tech. Security (IJCSITS) 1(2), 136–146 (2011)
3. Ghosh, P., Mandal, A.K., Kumar, R.: An Efficient Cloud Network Intrusion Detection System. In: Information systems design and Intelligent Applications, Advances in Intelligent Systems and computing 339, Springer India, 91–99 (2015)
4. Zhang, Q., Cheng, L., Boutaba, R.: Cloud Computing: state-of-the-art and research challenges. J. Internet Serv. Appl. 1, 7–18 (2010)
5. Hashizume, K., Rosado, D.G., Medina, E.F., Fernandez, E.B.: An analysis of security issues for Cloud Computing. In: Journal of Internet Services and Applications, Springer, 1–13 (2013)
6. Boncheva, V.M.: A Short Survey Of Intrusion detection Systems. In: problems of Engineering Cybernetics and Robotics 58, 23–30 (2007)
7. Raghav, I., Chikara, S., Hasteer, N.: Intrusion Detection and prevention in Cloud Environment: A systematic Review. Int. J. Comput. Appl. (IJCA) 68(24), 7–11 (2013)
8. Ladha, L., Deepa, T.: Feature Selection Methods and Algorithms. Int. J. Comp. Sci. Engg. (IJCSE) 3(5), 1787–1797 (2011)
9. Fang, Z., Na, L., Jinhui, L.: Application Research of the Genetic Algorithm on the Intelligent Test Paper Composition of Examination Database. Springer (ICAIC) 4, 443–448 (2011)
10. Soneji, H., Sanghvi, R.C.: Towards the Improvement of Cuckoo Search Algorithm. In: Conference on World Congress on Information and Communication Technologies, IEEE, 878–883 (2012)
11. Shawish, A., Salama, M.: Cloud Computing: Paradigms and Technologies. Inter-cooperative Collective Intelligence: Techniques and Applications, Studies in Computational Intelligence 495, Springer, 39–67 (2014)
12. Kumar, B.S., Raju, T.C.S.P., Ratnakar, M., Baba, Sk.D., Sudhakar, N.: Intruson Detection System-Types and Prevention. Int. J. Comput. Sci. Info. Tech. (IJCSIT) 4(1), 77–82 (2013)
13. Tuba, M., Subotic, M., Stanarevic, N.: Performance of a Modified Cuckoo Search Algorithm for Unconstrained Optimization Problems. In: WSEAS Transactions on Systems 11(2), 62–74 (2012)
14. Li, X., Yin, M.: Modified Cuckoo search algorithm with self adaptive parameter method. In: Information Sciences, Elsevier 298, 80–97 (2015)

15. Naik, M., Nath, M.R., Wunnava, A., Sahany, S., Panda, R.: A New Adaptive Cuckoo Search Algorithm. In: 2nd International Conference on Recent Trends in Information Systems, IEEE, 1–5 (2015)
16. Zhonghua, H., Xiaoting, D., Xisheng, L.: Self-adaptive cuckoo search algorithm for hybrid flowshop makespan problem. In: The 5th Annual IEEE International Conference on cyber Technology in Automation, control and Intelligent Systems, 1539–1545 (2015)
17. Yan, X., Luo, W., Li, W., Chen, W., Zhang, C., Liu, H.: An Improved Genetic Algorithm and Its Applications in Classification. Int. J. Comput. Sci. (IJCSI) 10(1), 337–346 (2013)
18. Soto, R., Crawford, B., Olivares, R., Barraza, J., Johnson, F., Paredes, F.: A Binary Cuckoo Search Algorithm for Solving the Set Covering Problem. Springer (IWINSE) 2, 88–97 (2015)
19. Tabassum, M., Mathew, K.: A Genetic Algorithm Analysis Towards Optimazation solutions. Int. J. Digital Info. Wireless Comm. (IJDIWC) 4(1), 124–142 (2014)
20. Ghosh, P., Mitra, R.: Proposed GA-BFSS and Logistic Regression based Intrusion Detection System. IEEE, 1–6 (2015)
21. Yu, L., Liu, H. Feature Selection for High Dimentional Data: A Fast Correletion-Based Filter Solution. In: 20th International Conference on Machine Learning (ICML), 1–8 (2013)
22. Roy, S., Chaudhuri, S.S.: Cuckoo Search Algorithm using Levy Flight: A Review. I. J. Modern Education and Computer Science 12, 10–15 (2013)
23. Jamil, M., Zepernick, H., Yang, X.: Levy Flight Based Cuckoo Search Algorithm for Synthesizing Cross-Ambiguity Functions. In: IEEE Military Communications Conference, 823–828 (2013)
24. Padhy N.P., Simon S.P. Soft computing: With Matlab Programming. Oxford Publication Book, 1st edition (2015).
25. Yan, G., Xie, G., Chen, Z., Xie, K.: Knowledge-Based Genetic Algorithms. Springer (RSKT), 148–155 (2008)
26. Sheng, L., Gao-yun, L., Jia, S., Tian-ying, S.: Research on Optimization Efficiency of Genetic Algorithms. IEEE, 1–4 (2008)

Unsupervised Document-Level Sentiment Analysis of Reviews Using Macaronic Parser

Sukhnandan Kaur and Rajni Mohana

Abstract Exponential rise in the multilingual web content affects the present-day decision support system to a great extent. To normalize such web content is the need of an hour. Reliability of decision support system broadly depends on the flawless processing of the language data present over the web. Macaronic text is one of the text usually found over the web. It is basically the text that contains number of languages in a single document instead of uniform language for whole document. To cope up with such a text, in this paper we propose a macaronic parser. This parser is language-independent and task-independent. The output of the proposed system is the normalized uniform base language text. This output can further be used in many other language processing tasks.

Keywords Sentiment analysis · Opinion mining · Macaronic language
Multilinguality · Sentiscore · Sentistrength · Punjabi text · Hindi text

1 Introduction

Sentiment analysers are highly useful in enterprise business. There are huge number of social media sites such as Twitter, Facebook, BlogSpot, Amazon, etc., which are used for collecting the reviews of people about any entity. The web users act as an advisory body for various enterprises. Business people use this data for figuring out the major and minor flaws in their products or services. This also helps them to improve their product quality. There is no language barrier to write anything over the Internet. This makes the task of sentiment analysers a bit complex. Nowadays, there are large number of sentiment analysers available for different languages. To handle

S. Kaur (✉) · R. Mohana
Computer Science and Engineering, Jaypee University of Information Technology, Solan, India
e-mail: sukhnandan.kaur@mail.juit.ac.in

R. Mohana
e-mail: rajni.mohana@juit.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_36

multilingual text is a big challenge in sentiment analysis. Along with multilingual text, people also use macaronic language over the Internet. Basically, macaronic language consists multilingual text which comprises of different languages/scripts together. With growing diversity, it has become of utmost importance that we acknowledge the existence of this kind of text. Especially, in the world where expressing opinions from anywhere in the world has become a fairly easy task. There have been a lot of studies on the information set that can be extracted from tweets and Facebook messages or posts. Twitter and Facebook information is the best way to keep a tab on the ongoing events, opinion of general public, trending topics, etc. However, one big challenge of this kind of information mining is the redundant and incongruous elements, we find along the way. Handling macaronic language not only useful in sentiment analysis but also in many natural language processing tasks such as named entity reorganization, pronoun resolution, feature extraction, etc.

However, there stands an obstacle in our way, while mining the text in one language; we seldom are able to handle a different language in the same context. We generally treat the other language/script words as foreign words, and lose major information in not treating these words. Processing this information is very useful for various automatic language processing tasks i.e. named entity recognization, pronoun resolution, automatic summarization along with sentiment analysis.

The given sentence is an example of macaronic text, it consists words other than base language.

Example1:

This is a ਚੰਗੀ *movie.*

Here the Punjabi language word (ਚੰਗੀ) of the context is taken as garbage and tossed aside and the English portion of it will be taken into consideration. With this we lost meaningful information. Here, the opining is about the movie is missed. We don't particularly know the opinion because it has been tossed aside. Similarly, if we apply the filter over the discarded words, i.e. foreign words and convert everything from Punjabi to English. We would be able to figure out the opinion about film.

Motivation

The motivation of the proposed technique is to handle the macaronic text by automatic identification of the fragments of the text belongs to different languages. The existing systems often discard the words other than the base language. The processing of the raw data often takes the text in multiple languages as an input. Sometimes, It discards text containing meaningful information. The proposed technique is designed to handle this type of discarded fragments. From example 1, it can be clearly seen that how important is the need to normalize the macaronic text for sentiment analysis. The state-of-the-art sentiment analysers give the neutral opinion about the movie although it is positive.

This arose the need to normalize the macaronic text. This paper proposes a method to fragment the text and autodetect the language used in the text based on Unicode information at a script level which is different for every language/script. The deduction of the corresponding language of the specific fragment other than the target language is also presented. Hence translate the particular foreign text into a base language. For our convenience, we have taken the English as a base language.

The remainder of the paper is organized as under: Section 2 describes the state-of-the-art sentiment analysis. Section 3 contains the detailed description of the proposed system. In Sect. 4, the proposed algorithm is presented. Section 5 investigates the results. Section 6, concludes the results.

2 Related Study

Various researchers are working in the field of sentiment analysis. They get huge success in their work. The task of sentiment analysis started decades ago. With time, it becomes the prime task for various enterprises to enhance their reputation in the market. Earliest works in the area of sentiment analysis is done by Hatzivassiloglou et al. [1]. They have used adjectives for deducing the polarity of the document. His work is then elaborated by Pang et al. [2]. They mainly focused over the supervised learning algorithms. They have used various machine learning algorithms for their work. The continuation of their work in the field of sentiment analysis is given in [3, 4], where they have used minicut algorithm for opinion summarization and also presented variousopinion mining techniques. Latest research paying more attention to the sentiment analysis over the data collected from various social sites. Researchers [5, 6] have used social media data for the sentiment analysis. Connor et al. [7] have used twitter data for sentiment analysis based on time series. Yang and Liang [8] (2010) proposed a new approach for identification of natural language, i.e. joint approach based on N-Gram frequency statistics and Wikipedia. Carter et al. [9] have used N-gram approach to identify five languages from the microblog. Later Lui et al. [10] followed same approach over the long and short textual documents. A detailed study of sentiment analysis at various levels of granularity [11] is explained by Kaur et al. Out of this literature study, we have found that the inclusion of normalization of macaronic language still needs attention. In this manuscript, we proposed a model to deal with the macaronic language for sentiment analysis.

3 Proposed System

We have taken the following approach to isolate and identify the language before embarking up on the journey to neutralize it through translation. The system design for the macaronic paerser is given in Fig. 1. Processing of the text passes through different phases of the system.

Phase1: In this phase, the input is given to the system as a web content, i.e. macaronic text/simple.

Phase 2: Based on tokens, filtration of the text is done at this level. The division of the text into base language and foreign words, i.e. other than the base language is done. The tokens which are from the base language are separated first along with their index values. The foreign words are then actually processed. These foreign

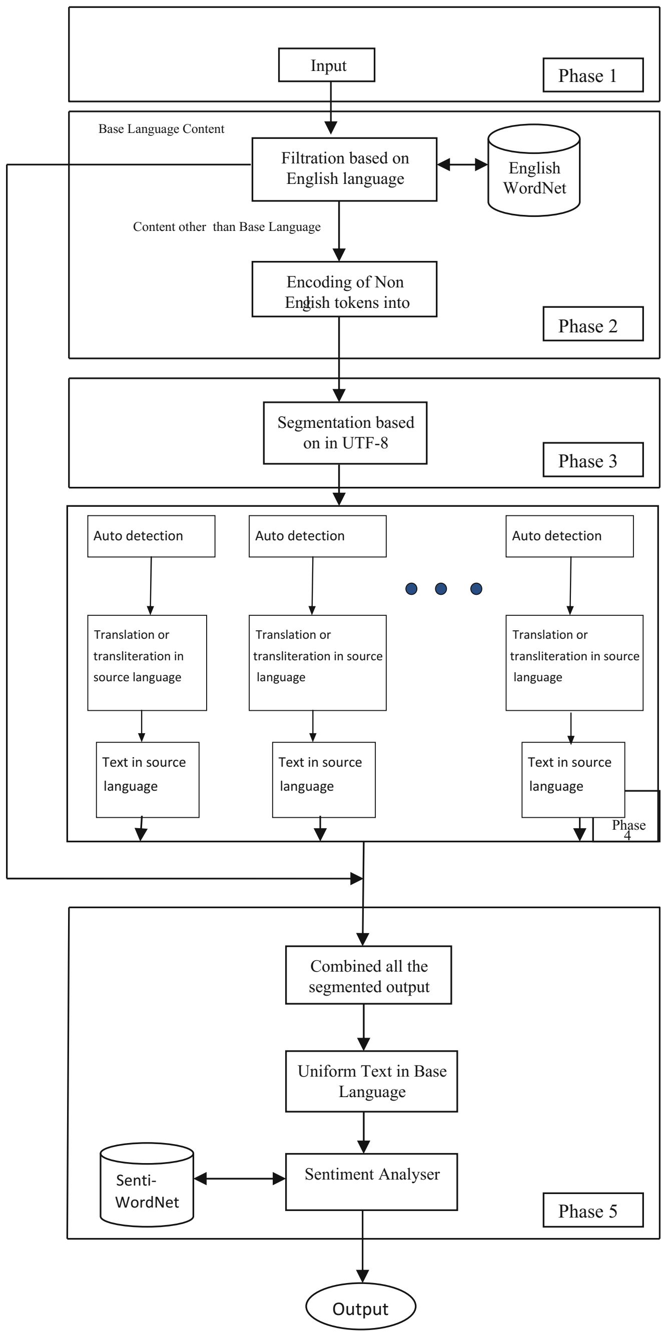


Fig. 1 Proposed system design of macaronic parser

words are tagged as non English tokens. We have used the English as base language for our work. Therefore, we have used English WordNet for the processing.

Phase3: In this phase, encoding of the tokens is done based on UTF-8. Each token is then has its encoded value.

Phase4: Under this phase, translation or transliteration is being done depending upon the number of languages we want to handle to cope up with the macaronic language. In this manuscript, the model represents the structural format of the processing of text. Under this, based on encoding automatic language detection algorithm is applied over it. After finding the language of the token, the working of translation or transliteration is started. The output of each sub-phase is the base language text.

Phase5: In the final phase of the model, whole text (English/Hindi/Punjabi)) is now converted into one language i.e. English. This converted text is then passed through the sentiment analyser to generate sentiscore for each document.

4 Proposed Algorithm

Input : Documents D = {d1, ..., dn}

'n' is the total no. of documents

Output : Document in a uniformed text ready for further language processing task.

Algorithm:

1. Set i to 1
2. for each document (d_i) in D
3. Encoding based on UTF-8 / UTF-16
4. Segmentation based on encoded document. //similar category segments are combined
5. for each segmented text
6. Detection of the language. // Hindi/English/Punjabi based on UTF-8 / UTF-16
7. Apply translation techniques
8. Content of each segment in source language.
9. End for
10. Combine all the fragmented text
11. End for
12 Apply POS tagging.
13. Processing of the formatted text for NER/ Opinion Mining/ Summarization/etc.

The proposed algorithm is very useful in dealing with the informal multilingual content present over the web. Following text describe the broad processing steps of the system. Step 1: Extracting the stop words from the mainstream sentence. We do this by selecting a base language and running the words through a dictionary of that particular language. The index value for each token is retained. These tokens become handy while plugging back at their original place.

Input: User defined text

Output: Two list containing base language words and foreign words.

Step 2: The foreign language words are entirely different from that of English (base language in our case), and thus we understand that it is not completely useless but may or may not be informative to us. Hence, we decode the same and run it through a UTF16 to UTF 8 encoder which further gives us results for the information of the script type.

Our studies have shown that Devanagiri script when converted to UTF 8 bytes, gives us a combination of three bytes. the script range starts from [224] [164] [191] to [224] [165] [129].

Input: List containing foreign words

Output: Unicode information of the script in which the foreign words are written.

Step 3: We use this information to extract the script in which the particular text is written. We don't need any sort of prior knowledge for any language for the same. All we need to know is a Unicode set for a script and run the text by it. Using that, we will autodetect the language on its own, without the help of any language experts.

Input: Unicode information of the script in which the foreign words are written

Output: The script, in which the words are written.

Step 4: After we have extracted the information, we utilize some back ground information to finalize the language for a script. However, in this paper we limit our approach to only two languages Hindi and Punjabi language script.

Hence, we run out text to the translator API, which translates our language to the particular base language, i.e. English

Input: Foreign text along with the script information

Output: Translated or neutralized text in English (Base language)

Step 5: Plug the neutralized tokens back to the original text, to make the entire picture clear.

Input: or neutralized text in English (Base language)

Output: Full input text in one base language i.e. English in this paper.

Step 6: Part of speech tagging is done to perform the various language processing subtasks.

Input: Full input text in one base language i.e. English in this paper.

Output: POS tagged sentence.

Step 7: Depends on subtasks to be performed, the tagged text is passed to the system. As in this manuscript we have used sentiment analysis. Therefore, we have passed the text to sentiment analyser.

Input: POS tagged sentence

Output: Sentiscore associated to each document.

5 Evaluation of the System

We have evaluated the system over the semantically similar dataset comprises of 200 documents i.e. English dataset and the dataset consists of the Macaronic statements which have various foreign language words studded into it. In this manuscript, we have used Punjabi, Hindi words in a particular sentence. We have applied NLTK pos tagger to find the opinionated words. From Table 1, we can see that how the presence of foreign words actually affects most of the language processing tasks. In this manuscript we have analysed the system for sentiment analysis.

Table 1 Results of sentiment analyser

Test sentence (Macaronic)	Test sentence (English statements)	POS. Macaronic	POS. English	SentiStrength (Macaronic)	SentiStrength (English)
This phone has very ਵਧੀਆ battery backup	This phone has very good battery backup	This|DT phone|NN has| VBZ very|RB ਵਧੀਆ|: battery|NN backup|NN|	This|DT phone|NN has| VBZ very|RB good|JJ battery|NN	0.108	0.293
This phone has very अच्छा battery backup	This phone has very good battery backup	This|DT phone|NN has| VBZ very|RB अच्छा|: battery|NN backup|NN|	This|DT phone|NN has| VBZ very|RB good|JJ battery|NN backup|NN|	0.108	0.293
This ਫੋਨ has better battery backup	This phone has better battery backup	This|DT ਫੋਨ| NN has|VBZ better|JJR battery|NN backup|NN	This|DT phone|NN has| VBZ better| JJR battery|NN backup|NN	0.415	0.415
This दलिचस्प movie i will surely watch	This interesting movie i will surely watch	This|DT दलिचस्प|NN movie|NN i| PRP will|MD surely|RB watch|VB	This|DT interesting|JJ movie|NN i| PRP will|MD surely|RB watch|VB	0.053	0.117
ਇਸ restaurant ਦਾ ਖਾਣਾ ਬਹੁਤ good ਹੈ	The food of this restaurant is very good	ਇਸ|NN restaurant| VBD ਦਾ|CD ਖਾਣਾ|CD ਬਹੁਤ| CD good|JJ ਹੈ| NN	The|DT food| NN of|IN this| DT restaurant| NN is|VBZ very|RB good| JJ|	0.478	0.395
इस restaurant का खाना बहुत अच्छा है	The food of this restaurant is very good	इस|NN restaurant| VBD का|CD खाना|CD बहुत| CD अच्छा|CD है|CD	The|DT food| NN of|IN this| DT restaurant| NN is|VBZ very|RB good| JJ|	0	0.395

In Table 1, for the convenience top six sentences taken to show the results, SentiScore associated with the compound sentence i.e. contain foreign words from the Hindi or Punjabi language i.e. "Test Sentence (Macaronic Statements)" under the column name "SentiStrength(Macaronic)". Other two columns show the corresponding English text and its sentiscore under the column heading "Test Sentence (English Statements)" and "SentiStrength (English)" respectively. From column "POS.Macaronic", it can be seen that how the foreign language words in a single sentence are ignored by the present day pos taggers. We have tried other taggers also, they also gave unsatisfactory results. Most of them took the foreign language words as a noun irrespective to language. This harms the language processing results to a great extent.

We have used gold standard dataset to verify the results. In Figs. 2 and 3, it can be seen that for the reviews having same semantic structure irrespective to the language have different graph slopes. The results did not follow the actual trend set through the gold standard as shown in Fig. 3. After processing the reviews having

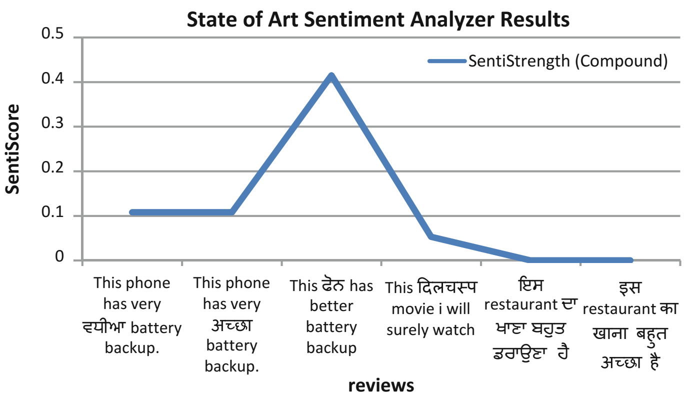
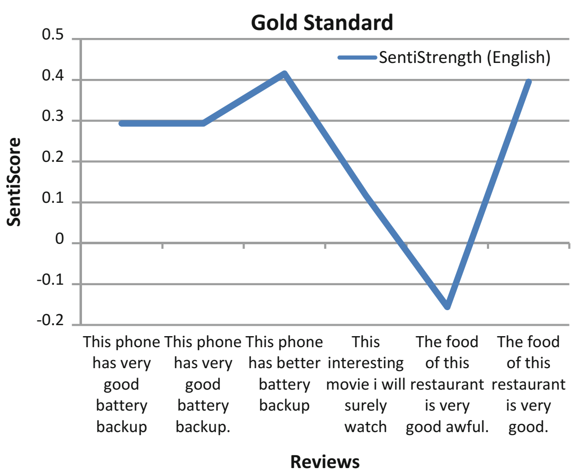


Fig. 2 Sentiscore calculated by state-of-the-art sentiment analysers

Fig. 3 Gold standard of Sentiscore



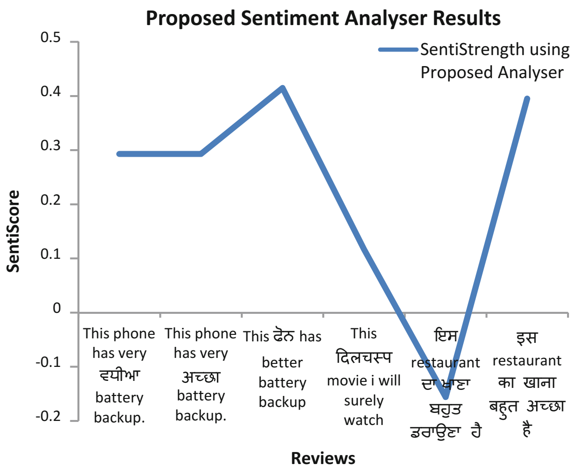


Fig. 4 Sentiscore calculated by proposed sentiment analysers

Macaronic language through the proposed sentiment analyser the results are fairly significant as shown in Fig. 4. It follows the same trend as shown in Fig. 3. This shows the significant results of the proposed system.

6 Conclusion

To recapitulate, this paper proposed macaronic parser to deal with the unnormalized web content. The results have shown the reliability of the proposed system is more as compared to the existing system. The sentiscore generated through proposed system is much closer to the gold standard of that particular dataset. Results suggest that this type of text processing is highly beneficial for the Named Entity Extraction, Feature Extraction, E-mail summarization, etc. To summarize the results, we can say that there is still a lot to do for resource scarce Indian languages to increase the feasibility of the decision support system.

References

1. V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, 1997, pp. 174–181.
2. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79–86.

3. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, 2004, p. 271.
4. B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval,* vol. 2, pp. 1–135, 2008.
5. A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in *LREc*, pp. 1320–1326.
6. A. Go, L. Huang, and R. Bhayani, "Twitter sentiment analysis," *Entropy,* vol. 17, 2009.
7. B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," *ICWSM,* vol. 11, p. 1.2.
8. X. Yang and W. Liang, "An n-gram-and-wikipedia joint approach to natural language identification," in *Universal Communication Symposium (IUCS), 2010 4th International*, pp. 332–339.
9. S. Carter, W. Weerkamp, and M. Tsagkias, "Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text," *Language Resources and Evaluation,* vol. 47, pp. 195–215.
10. M. Lui and T. Baldwin, "Cross-domain feature selection for language identification," in *In Proceedings of 5th International Joint Conference on Natural Language Processing*.
11. S. Kaur and R. Mohana, "A roadmap of sentiment analysis and its research directions," *International Journal of Knowledge and Learning,* vol. 10, pp. 296–323.

Efficient Speech Compression Using Waveform Coding in Time Domain

Subham Paul and Debashis Chakraborty

Abstract Compression of speech has been an active area of research in the field of digital signal processing for the last decade with many techniques getting standardized and utilized commercially. However, with the advent of transform coding and sub-band coding, waveform coding of speech in the time domain, being a potential remover of redundancies present in speech signals, is in need of a revamp. This paper describes a novel method for speech compression based on waveform coding in time domain that provides a good compression ratio (above 90%) and acceptable range of PSNR (35–40 db) for a particular range of sampling rates.

Keywords Speech compression · Waveform coding · RLE · DPCM
PSNR

1 Introduction

The field of data compression revolves round reducing redundancies in data so as to minimize the cost of storage as well as that of transmitting data over network channels. Removing redundancies has been performed historically since the birth of codes such as Braille and Morse, as pointed out by Salomon [1]. In the modern age of technology, we are faced with applications on a daily basis that require various forms of media such as text, images, audio and video, which have formidable size in uncompressed form. But such data cannot be exchanged properly when there are constraints on storage space and data transmission bandwidth. Technology has solved its own problem by facilitating the evolution of historical data compression techniques into more powerful ones, both lossless and lossy, that can properly

S. Paul · D. Chakraborty (✉)
Department of CSE, St. Thomas' College of Engineering
and Technology, Kolkata 700023, India
e-mail: sunnydeba@gmail.com

S. Paul
e-mail: psubham244@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_37

balance the trade-off between the amount of compression and the quality of signals required for communication [2]. With respect to applications involving voice, such as voice calls over the internet and over mobile networks, the compression applied must provide desirable results, including acceptable speech quality and real time computation to cater to the needs of the fast moving world. As a result, speech compression (or speech coding) has been a major field in digital signal processing. Speech coding involves representing the original speech signal in a compact form (so that the signal consumes less space in storage and/or transmission) and then decoding the compressed signal to produce a signal as similar to the original signal as possible. Speech is the easiest medium for humans to communicate with each other and hence the quality of speech must be preserved in the encoding/decoding process. The compression of speech is possible owing to the redundancies present in speech signals after they are digitized. Smaller size of signals leads to faster transmission, resulting in systems that enable convenient and satisfactory communication.

2 Related Work

The techniques for compressing and decompressing data are targeted towards removing redundancies, reducing storage and transmission space keeping the quality of decompressed data as close to the uncompressed data as possible [1, 3]. If the decompressed data is exactly the same as the uncompressed data, the compression is said to be lossless, otherwise the compression is said to be lossy [1]. The objective of this change in data is to reduce storage and transmission costs [4] and increasing the speed of data transfer. Speech compression methods and standardized codecs are used in numerous communication applications involving voice, such as in VoIP applications [4]. Speech compression has to be preceded by digitisation of speech, which involves sampling, quantisation and coding [4, 5]. Uncompressed digitised speech contains numerous redundancies, both repetition of data and existence of digitised audio signals inaudible to the human ear. While repetitive data in general can be contained by Run-Length Encoding [1, 3], the aspects in digital speech relating to the perception of audio can be addressed by methods such as Silence Compression [1] and Companding (notably μ-Law and A-Law) [6]. Another area that can be exploited in removing redundancies is the high amount of correlation in neighbouring speech samples [1]. Speech coding techniques can be categorized as waveform coding, sub-band coding and source coding [7]. In waveform coding, the process attempts to code the speech data so as to be able to construct a signal having a waveform similar that of the uncompressed signal [1, 7] without considering the source of speech. Differential Pulse Code Modulation (DPCM) and Adaptive DPCM (ADPCM) are excellent examples of waveform coding [1, 7, 8]. These techniques encode the differences between a speech sample and its preceding sample, thereby reducing the number of bits required to store samples of speech. Source Coding of speech involves making a model of the source of speech [1] instead of storing information about the waveform. Linear Predictive Coding (LPC) is a low bit-rate source coding technique which consists of analysis (coding) and synthesis (decoding) [7]. The AMR algorithm makes use of LPC in compressing speech to very low bit-rates [9]. It uses multiple modes of compression which are utilized as per instantaneous requirements. Speech can also be compressed using transforms [10] to produce smaller values for samples, or utilizing the output from transform techniques to remove sounds having inaudible frequencies. The latter is used in the MP3 algorithm [11, 12]. The performance of speech compression methods is measured by calculating the compression percentage (to show the reduction in storage size of the data) and peak signal-to-noise ratio or PSNR (to show the noise induced in the signal due to compression) [1].

3 Proposed Method

The proposed method is a waveform coding scheme in the time domain that involves a variety of compression techniques on the input speech signal. Like AMR [9, 13], the proposed method uses the idea of dividing the speech signal into frames (or blocks) and applying a suitable mode of compression (from 8 different modes) for the frame depending on the redundancies present in it. For each block, a proper 8-bit header is designed that contains information about the technique used for compression in the block (3-bit code), the number of bytes used for encoding the length of the block (2-bit code), and additional information such as number of bits used for encoding differences or errors, or the size of clusters, etc. (3-bit code). The header is followed by encoding the number of samples in the block. The header is followed by the stream obtained by compressing the block samples. During decompression, the compressed stream is read one block at a time and decoded according to the mode of compression used in that block. The process is based on a number of parameters which are set automatically to optimize the compression performance.

3.1 Algorithm Strategy

At first, the lower order byte of each sample of the input stream is removed and the samples having very small values are forcibly made 0 to facilitate silence compression. Then, this compressed stream is analysed, broken into blocks and further compressed using one of the 8 modes, resulting in 8 possible types of blocks produced, which are described below:

A. **Run-Length Encoding (Code: 000)**
The entire block has samples of the same value. As a result, only the value of the samples is encoded.

B. **Optimized Differential Pulse Code Modulation (Code: 001)**
A new block is created having the differences between every pair of adjacent samples in the original block and the first sample value is encoded. Then, the minimum number N_{min} in the new block is encoded followed by encoding the differences between every number of the new block and N_{min}.

C. **Linear Prediction (Code: 010)**
The first 5 sample values of the block are encoded. For the remaining samples, the current sample value is predicted using the weighted average of its previous 5 samples and the difference between the predicted value and actual value of the sample is encoded.

D. **Square Root Companding (Code: 011)**
For all sample values in the block, the square root of each value (in the range [0, 15]) is encoded.

E. **Linear Means Clustering (Code: 100)**
The block is linearly divided into clusters of pre-determined size and the arithmetic mean of every cluster is encoded.

F. **Ordinary Differential Pulse Code Modulation (Code: 101)**
The value of the first sample in the block is encoded followed by encoding the differences between every sample value (except the first sample) of the block and its preceding sample.

G. **Differential Minimization (Code: 110)**
The value of the smallest sample in the block S_{min} is encoded followed by encoding the differences between every sample value of the block and S_{min}.

H. **Unchanged (Code: 111)**
The block of samples is encoded as it is without further change.

3.2 Algorithm for Compression

Compression is performed using the following steps (Table 1):

Step 1. Obtain the sampling rate of the speech signal. Set the required parameters (Silence Boundary, Differential Encoding Boundary, Cluster Block Limit and Cluster Size) as per the obtained sampling rate:

Step 2. Pass the original stream of 16-bit speech samples into Byte Reduction Filter. Output of filter is a stream of 8-bit speech samples, modified as per silence compression requirements.

Step 3. Pass the samples into RLE Filter. Perform the method described below for every discovered block:

Table 1 Description of the parameters used in the proposed method

Parameter no.	Name of parameter	Purpose
P1	Silence boundary	Samples having absolute value less than P1 are assumed to be 0
P2	Differential encoding boundary	Coding of differences or errors is permitted only when the no. of bits to encode such values is not greater than P2
P3	Cluster block limit	Linear mean clustering is applied if size of the block is not greater than P3 samples
P4	Cluster size	Size of each cluster in linear mean clustering is not greater than P4 samples

3.1. If an non-RLE block of length less than 4 samples exists at the beginning prior to a block of samples satisfying RLE requirements, the non-RLE block elements are discarded and those many samples as added to the discovered RLE block. The RLE block is encoded using the value of samples in the block.

3.2. If a non-RLE block of length greater than 3 samples is discovered, then that block is passed into Primary Differential Encoding Filter. Here, 2 values are calculated: N1(number of bits required for differential minimization method) and N2 (number of bits required for ordinary DPCM).

3.3. If either of N1 or N2 does not exceed P2, then pass the block into Secondary Differential Filter.

3.3.1. If N1 is less than N2, then perform differential minimization, else calculate N3 (bits for optimized DPCM).

3.3.2. If N3 is less than N2, then perform optimized DPCM, else perform ordinary DPCM.

3.4. If both N1 and N2 exceed P2, then pass the block into Linear Prediction Filter.

3.4.1. If the length of the block does not exceed P3, then perform linear means clustering method.

3.4.2. If the length of the block exceeds P3, the prediction errors of the block samples are calculated and the minimum number of bits required to encode these errors is calculated as N4.

3.4.3. If N4 does not exceed P2, then perform linear prediction method.

3.4.4. If N4 exceeds P2 but is less than 8, then perform square root companding.

3.4.5. If N4 exceeds 7 then encode the block samples without further compression (unchanged block).

3.3 Algorithm for Decompression

The compressed stream is read and the blocks are identified using the header of each block. Decompression is performed on every block using the following steps:

Step 1. Obtain a fresh header from the compressed signal. Partition the header into three segments to identify the method used for compression in the block (first 3 bits—Segment 1), number of bytes to store the length of the block (next 2 bits—Segment 2) and some method specific information (remaining 3 bits—Segment 3).

Step 2. The mode of compression is identified using Segment 1. The mode indices and codes are discussed in Sect. 3.1.

Step 3. Segment 2 of the header is converted into decimal and stored as B_L. This gives the number of bytes for storing the length of the block. The next B_L bytes are read, and converted to decimal in big endian order to get the length of the current block L.

Step 4. If the mode index is either B, C, E, F or G, then Segment 3 of the header is converted into decimal and stored as B_D. This is the number of bits in which the mode-specific differences or errors are encoded in the block. For other modes this header segment is ignored.

Step 5. After analysis of the header and obtaining the length of the block, the samples of the block are decompressed according to the mode of compression used in that particular block. Whenever a sample is decoded, it is written into the decompressed file by writing a byte of value 0 followed by the decoded byte. This is because the decompressed signal has samples of length 2 bytes, accessed in little endian format.

Mode-wise Decompression:

i. **RLE**:
   I. Read the next byte and store it as X.
   II. Repeat L times:
      1. Write X.

ii. **Optimized DPCM:**
   I. Read the next byte S_0. This is the first sample of the block.
   II. Read the next byte D_{min}.
   III. Read the next L bit segments of length B_D. For each segment b_i for $i = 1$ to L:
      1. Store $D_i = D_{min} +$ decimal equivalent of b_i
      2. Store $S_i = S_0 + D_i$
      3. Write S_i
      4. Update $S_0 = S_i$

iii. **Linear Prediction:**
   I. Initialize a buffer of capacity 5 by inserting the next 5 bytes read from the compressed stream.
   II. Repeat L times:
      1. Find weighted average A of differences between two consecutive elements of the buffer.

2. Store $P = E_1 + (A * 5)$; where E_1 is the first element of the buffer and P is the predicted sample value.
3. Read the next B_D bits as segment S. Convert S into decimal and store as prediction error E.
4. Store $E_N = P + E$
5. Write E_N
6. Update buffer by removing first element, shifting the remaining elements one place towards the front, and inserting E_N at the rear.

iv. **Square Root Companding:**

I. Repeat L times:

1. Read the next four bits, convert into decimal and store as R.
2. Store $S_R = R^2$
3. Write S_R.

v. **Linear Means Clustering:**

I. Store size of cluster $L_C = B_D * 4$

II. Store number of clusters $N_C = L/L_C$

III. Repeat N_C times:

1. Read next byte and store it as M.
2. Repeat L times:
   1) Write M.

vi. **Ordinary DPCM:**

I. Read the next byte S_0. This is the first sample of the block.

II. Read the next L bit segments of length B_D. For each segment b_i for $i = 1$ to L:

1. Store D_i = decimal equivalent of b_i
2. Store $S_i = S_0 + D_i$
3. Write S_i
4. Update $S_0 = S_i$

vii. **Differential Minimization:**

I. Read the next byte D_{min}. This is the minimum sample in the block.

II. Read the next L bit segments of length B_D. For each segment b_i for $i = 1$ to L:

1. Store $S_i = D_{min}$ + decimal equivalent of b_i
2. Write S_i
3. Update $S_0 = S_i$

viii. **Unchanged:**
   I. Repeat L times:
      1. Read the next byte and store it as X
      2. Write X.

3.4 Demo Implementation

For demonstration purposes, we are taking a sequence of 16 bit speech samples S_0 as input. These are represented in little endian format, i.e. for the first two bytes B1 and B2, the sample shown is B2B1. The original stream of bytes is:

F000 FA00 FF00 FB00 0101 3C 02 F000 FF00 4401 E205
2006 EF06 1007 F105 0006 0506 0208 FF07 F205 F005 8805 …

Grouping the bytes into 16 bit samples, we get the input sequence S_0. Thus,

S_0 : 00F0 00FA 00FF 00FB 0101 023C 00F0 00FF 0144 05E2
0620 06EF 0710 05F1 0600 0605 0802 07FF 05F2 05F0 0588 …

COMPRESSION:

Removing the lower order bytes from each sample of S_0, we get a sequence S_1 of 8 bit samples:

S_1 : 00 00 00 00 01 02 00 00 01 05 06 06 07 05 06 06 08 07 05 05 05 …

Converting the samples from hexadecimal to decimal, we get sequence S_2:

S_2 : 0 0 0 0 1 2 0 0 1 5 6 6 7 5 6 6 8 7 5 5 5 …

We treat samples having values less than 4 as 0. So after modifying the required samples, we get a new sequence S_3:

S_3 : 0 0 0 0 0 0 0 0 0 5 6 6 7 5 6 6 8 7 5 5 5 …

We initialize a buffer B as empty. From the starting number of S_3 we add a number to B as long as the last 3 numbers in the buffer do not form a sequence of repetitive numbers. In this case, the insertions to the buffer B are as follows:

B = ()(0) (0 0) (0 0 0)(0 0 0 0)……(0 0 0 0 0 0 0 0)(0 0 0 0 0 0 0 0 0)
(0 0 0 0 0 0 0 0 0 5)(0 0 0 0 0 0 0 0 0 5 6)(0 0 0 0 0 0 0 0 0 5 6 6)

When the insertions stop, we extract the run length available from the beginning of B as a new sequence S_{RLE}, and keep the remaining in B.

Thus, B = (5 6 6) and S_{RLE}: 0 0 0 0 0 0 0 0 0 (Run Length = 9, Value = 0)

The mode of compression required is plain RLE, so the mode in binary is $(000)_2$. Since the blockhas 9 bytes, we need 1 byte to store the value '9'. The code for length bytes in the header will be binary of 1, i.e. $(01)_2$. The header composed is the mode code (3 bits), followed by length code (2 bits) and 3 don't care bits. The composed header in binary is thus $(00001000)_2$. The next byte is used to store the length of sample frame, which is the binary of 9, i.e. $(00001001)_2$. The next byte stores the value of the run, which is 0 in decimal, $(00000000)_2$ in binary. Thus, the bytes encoded for this blockare: 08 09 00.

Again, insertions are made to the buffer B till a run length of length 3 is obtained (the previous block was RLE, so now a non-RLE block may be present). Thus, B = (5 6 6 7 5 6 6 8 7 5 5 5). The sequence of non-RLE numbers are extracted from B as a new sequence S_{nRLE}.

Thus, B = (5 5 5) and S_{nRLE}: 5 6 6 7 5 6 6 8 7 (Length = 9)

Next, we calculate N1 (bits needed for differential minimization) and N2 (bits required for DPCM).

Minimum value in S_{nRLE} is 5. Subtracting 5 from each number, we get X_1 = 0 1 1 2 0 1 1 3 2. Thus, N1 = 2 (2 bits needed to store 3).

First value in S_{nRLE} is 5. Calculating forward differences, we get X_2 = + 1 0 +1 −2 +1 0 +2 −1. Thus, N2 = 3 (3 bits to store ± 2).

As N1 is less than N2, we opt for differential minimization. Mode code is $(110)_2$, length code is $(01)_2$, difference bits code is $(010)_2$. The composed header is thus $(11001010)_2$. The byte to store the length of the block (9) is $(00001001)_2$. The next byte is to store the first number in the block, which is 5 in decimal, $(00000101)_2$ in binary.

Next, each value of X_1 is converted into binary (2 bits each): 00 01 01 10 00 01 01 11 10. Appending these to an empty binary buffer B_2 and appending additional 0s at the rear to make the length a multiple of 8, we get B_2 = (0001011 00001011110000000). The bytes obtained from it are $(00010110)_2$, $(00010111)_2$ and $(10000000)_2$. Thus, the encoded bytes for this block are: CA 09 05 16 17 80..

Thus, for the first 18 samples (36 bytes), we have a compressed stream S_C: 08 09 00CA 09 05 16 17 80 (9 bytes).

DECOMPRESSION:

We initialize a buffer B to the contents of the compressed stream S_C and set an output stream S_{DEC}. Thus, B = (08 09 00 CA 09 05 16 17 80).

The first byte has to be a block header, which is $(08)_{16}$ in hexadecimal, $(00001000)_2$ in binary. This shows the current block is a RLE block (mode 000) and the length of the block is stored in the next 1 byte (length code 01). The next byte read is $(09)_{16}$ in hexadecimal, 9 in decimal. The next byte, having decimal value 0, is the value of RLE. Thus, a block of 9 bytes each of value 0 is appended to the output stream S_{DEC}.

Thus, S_{DEC}: 00 00 00 00 00 00 00 00 00.

Discarding the used bytes from B, the current state of B becomes (CA 09 05 16 17 80).

The first byte has to be a block header, which is $(CA)_{16}$ in hexadecimal, $(11001010)_2$ in binary. This shows the current block is a Differential Minimization block (mode 110), the length of the block is stored in the next 1 byte (length code 01) and the differences are 2 bits long (code 010). The next byte read is $(09)_{16}$ in hexadecimal, 9 in decimal. The next byte, having decimal value 5, is the least sample of the block. For 9 differences, each stored in 2 bits, we need 18 bits. These will be available in the next 3 bytes, which are 16 17 and 80 in hexadecimal. Converting them to binary and making a 24 bit string, we get the sequence S_1: 0001110000111110000000. Keeping the most significant 18 bits and ignoring the rest, we get the sequence S_2: 0001110000111110. From here, we get the 9 differences in binary (2 bits each), they are:

$(00\ 01\ 01\ 10\ 00\ 01\ 01\ 11\ 10)_2 = (0\ 1\ 1\ 2\ 0\ 1\ 1\ 3\ 2)_{10}$. Adding 5 to each, we get $(5\ 6\ 6\ 7\ 5\ 6\ 6\ 8\ 7)_{10}$. Thus, we have recovered 9 bytes for the block. These bytes are appended to S_{DEC}, which becomes S_{DEC}: 00 00 00 00 00 00 00 00 00 05 06 06 07 05 06 06 08 07.

Each byte in the output stream is preceded by a 0 byte, to make each sample 16 bits long, preserving the original nature of samples.

Uncompressed sequence: F0 00 FA 00 FF 00 FB 00 01 01 3C 02 F0 00 FF 00 44 01 E2 05 20 06 EF 06 10 07 F1 05 00 06 05 06 02 08 FF 07

Compressed sequence: 08 09 00 CA 09 05 16 17 80

Decompressed sequence: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 05 00 06 00 06 00 07 00 05 00 06 00 06 00 08 00 07

With the formulae explained in Sect. 3.5, this implementation yields compression ratio of 75% and a PSNR of 42.22 db.

3.5 Performance Measures

The following formulae have been used to obtain the compression results:

1.1.1. **Compression Factor (CF)**:

$$\mathrm{CF} = \frac{\text{Size of Output Signal}}{\text{Size of Input Signal}} \tag{1}$$

1.1.2. **Compression Ratio (CR)**:

$$\mathrm{CR} = \frac{\text{Size of Input Signal} - \text{Size of Output Signal}}{\text{Size of Input Signal}} \tag{2}$$

In this paper, CR has been expressed as a percentage.

1.1.3. **Mean Square Error (MSE)**:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(P_i - Q_i)^2 \tag{3}$$

where n = the number of speech samples in the input signal
P_i = ith sample in the input signal, Q_i = ith sample in the output signal

1.1.4. **Root Mean Square Error (RMSE)**:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{4}$$

1.1.5. **Peak Signal-to-Noise Ratio (PSNR)**:

$$\mathrm{PSNR} = 20\ \log_{10}\frac{\max_i |P_i|}{\mathrm{RMSE}} \tag{5}$$

where $\max_i|P_i|$ is the magnitude of the maximum possible sample value in the input signal. PSNR is expressed in decibel (db).

4 Results

In the proposed method, the silence compression employed is based on only one parameter, i.e. the largest sample value that should be treated as silence. This value has been set according to the sampling rate used in the particular file to be compressed. Based on experimental results, the following observation was made about the effect of changing upper limit of silence SZ (shown on the Y-axis) on Compression Ratio (CR) and Peak Signal-to-Noise Ratio (PSNR) (Fig. 1a).

Here, SZ refers to the value below which all sample values will be treated as 0. It is to be noted that SZ is the magnitude of sample value. The 8 bit samples have signed values in the range [−127, +127]. If the magnitude of a sample is under SZ, then it is considered to be part of silence and is made a sample of value 0. As per expectations, increasing the value of SZ increases the amount of compression. This depends heavily on the presence of silence zones in the speech signal. However, the

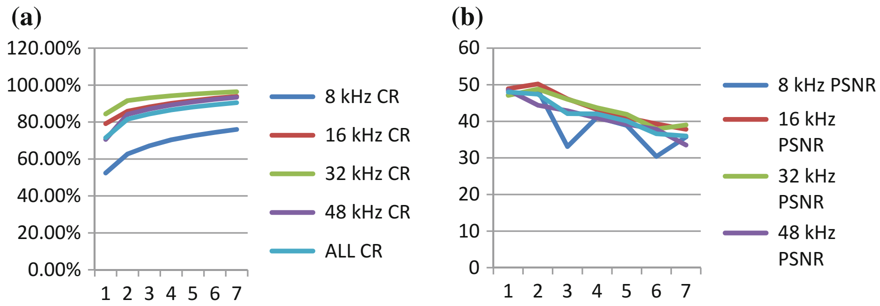


Fig. 1 Graphs showing changing **a** Compression ratio and **b** PSNR with changing silence upper limit

value of SZ must be such that it does not lead to unnecessary loss of signal quality. From Fig. 1b, it can be seen that the trend for 8 kHz signals is of a zig-zag fashion and that of other sampling rates is almost linear. We have targeted a PSNR of around 40 db after the silence compression stage in the proposed method in order to achieve a PSNR above 30 db after the final stage. Keeping this in mind, the value of SZ is selected as 4 for sampling rate of 8 kHz, and 6 for higher sampling rates. For compression, the required parameters (see Sect. 3.2) were set according to the signal sampling rate as follows (Table 2):

The proposed method was implemented on 19 standard test files (mono, 16 bit uncompressed PCM), having sampling rates in the range of 8–48 kHz and size in the range 60–800 KB. These files were downloaded from an ITU website [14]. The results for two standardized techniques, MP3 [11, 12] and AMR, for the same test files were also recorded. With respect to the compression performance (CR, PSNR and implementation time), it was observed that the proposed method worked best on the test files having sampling rate 16–32 kHz as the results of the method for such files were competent with those of MP3 and AMR. The readings obtained for these files are shown below (Table 3 and Fig. 2):

On observing the waveforms of the original and reconstructed signals for the three methods, we find that while MP3 produces a waveform which is almost indistinguishable from that of the original signal, the waveforms produced by AMR and the proposed method contain small distortions, but the overall shape of waveform is preserved. It is therefore observed that the proposed method yields the highest amount of compression among the three methods tested and also has satisfactory PSNR values.

Table 2 Parameters set for the compression process of the proposed method

Sampling rate (Hz)	P1	P2	P3	P4
8000	4	7	40	4
16000	6	3	40	8
32000	6	3	40	8
48000	6	3	80	8

Table 3 Experimental results of MP3, AMR and the proposed method (P.M.) on test files

File name	Sampling rate (Hz)	Original size (bytes)	Compression ratio for MP3 (%)	PSNR (db) for MP3	Compression ratio for AMR (%)	PSNR (db) for AMR	Compression ratio for P.M. (%)	PSNR (db) for P.M.
female1.wav (German)	16000	256046	74.69	42.128	95.01	29.027	95.92	36.811
female2.wav (German)	16000	256046	74.69	42.858	95.01	28.161	95.16	36.966
male1.wav (German)	16000	256046	74.69	42.581	95.01	27.121	96.24	37.001
male2.wav (German)	16000	256046	74.69	42.976	95.01	27.928	96.51	37.448
female1.wav (Polish)	32000	492846	75.10	28.844	97.50	30.467	97.82	39.058
female2.wav (Polish)	32000	492846	75.10	26.372	97.50	26.091	95.86	37.832
male2.wav (Polish)	32000	492846	75.10	25.733	97.50	25.857	97.14	36.947

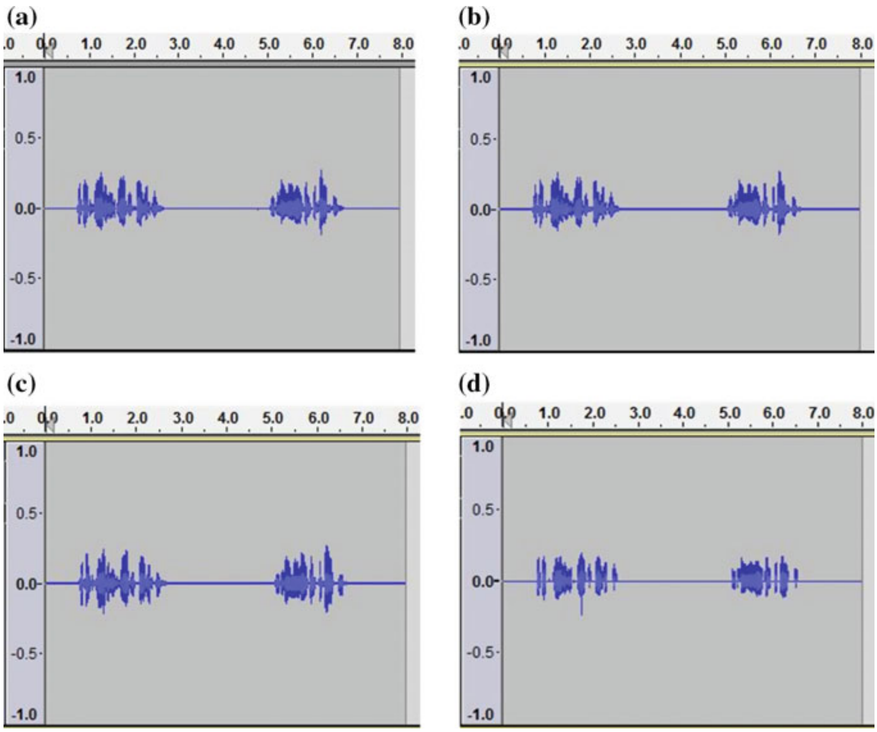


Fig. 2 The waveforms of **a** the test file "male2.wav (German)", and for **b** MP3, **c** AMR and **d** the proposed method on it

5 Conclusion

Speech compression has emerged as an essential component of modern communication systems involving voice, be it VoIP applications or voice calls in cellular networks. Popular standards such as AMR are capable of compressing speech signals considerably, but their performances can be further improved if the speech signals are re-quantized (by byte reduction and silence compression) and broken into blocks of variable lengths, as is demonstrated in the proposed method. The blocks may differ in the redundancies present in their samples, and therefore only the suitable mode of compression should be used in order to get best results. The method can be improved by optimizing the blocks, if possible, before applying the appropriate compression technique. On the whole, the proposed method is capable of identifying and removing most of the redundancies present in the tested speech files when acting on the time domain. This implies that if most of the redundancies present in the time domain can be removed, and then frequency domain waveform coding is performed on the blocks on which the proposed method gives mediocre performance, the compression can yield outstanding results, perhaps better than those of existing standards.

References

1. Salomon, D. (2007) Data Compression The Complete Reference (4th Ed). Springer, London (ISBN: 1846286026).
2. Forouzan, B.A. (2007) Data Communications And Networking (4th Ed). McGraw-Hill, New York (ISBN: 0072967757).
3. Nelson, M. – The Data Compression Book (2nd Ed). IDG Books Worldwide, Inc. (ISBN: 1558514341).
4. Sun, L. et al. (2013) Guide to Voice and Video over IP For Fixed and Mobile Networks, Computer Communications and Networks, DOI: 10.1007/978-1-4471-4905-7_2, Springer-Verlag London 2013.
5. ITU-T (1988) Pulse code modulation (PCM) of voice frequencies. ITU-T G.711.
6. Pan, D.Y. (1993) Digital Audio Compression. Digital Technical Journal Vol. 5 No. 2, Spring 1993.
7. Satyapriya, K. and Dasari, Y. (2013) Performance Analysis Of Speech Coding Techniques. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol. 2, Issue 11, November 2013.
8. Arsinte, R., Ferencz, A. and Miron, C. - Study of a Non-Linear DPCM Voice Compression Algorithm Optimised Using Neural Network Concepts (http://www.academia.edu/16775907/Study_of_a_Non-Linear_DPCM_Voice_Compression_Algorithm_Optimised_Using_Neural_Network_Concepts).
9. Choudhary, D. and Kumar, A. (2014) Study and Performance of AMR Codecs for GSM. International Journal of Advanced Research in Computer and Communication Engineering. Vol. 3, Issue 10, October 2014.
10. Vatsa, S. and Sahu, O.P. (2012) Speech Compression Using Discrete Wavelet Transform and Discrete Cosine Transform. International Journal of Engineering Research & Technology. Vol. 1 Issue 5, July 2012.
11. Raissi, R. (2002) The Theory Behind Mp3 (www.mp3-tech.org/programmer/docs/mp3_theory.pdf).
12. Brandenburg, K. (1999) Mp3 and Aac Explained. Paper presented at the AES 17th International Conference on High Quality Audio Coding, 1999.
13. ETSI (2000) Digital cellular telecommunications system (Phase 2 +); Adaptive Multi-Rate (AMR) speech transcoding. ETSI-EN-301-704 V7.2.1.
14. ftp://ftp3.itu.int/t/testsignal/GenAudio/P501/v2009_12/SpeechSequences.

Author Biographies

Mr. Subham Paul is pursuing a B. Tech degree in Computer Science and Engineering from St. Thomas' College of Engineering and Technology, Kolkata, West Bengal, India. This paper is part of the final year project on Speech Compression under the guidance of Mr. Debashis Chakraborty.

Mr. Debashis Chakraborty is an Assistant Professor in the department of Computer Science and Engineering, St. Thomas' College of Engineering and Technology, Kolkata, West Bengal, India. He has authored or co-authored about 17 conference papers in the area of Data and Image Compression.

Model Based Algorithm Validation Approach for Safety Critical Applications

R. Krishnaprasad, Manju Nanda, J. Jayanthi and Shyam Sundhar Dhage

Abstract The successful operation of safety critical systems, such as aircraft, spacecraft, underwater systems, chemical plants and nuclear plants, is largely dependent on the validation of sensor data, which provide information for control and performance. The sensor data validation is an essential step to improve data reliability. Model-Based Programming (MBP) provides appropriate algorithms to validate the sensor data. MBP provides graphical modeling, automated code generation from design models, executable specification for continuous model-based verification and validation (V&V) for early identifying design errors. In this paper we discuss the signal data validation using model-based approach to provide the analysis of validation algorithm at the design stage rather than the code stage. Analog inputs of Stall Warning System SWS/AIC computer is used as a case study to discuss the results of the model-based algorithm analysis. The analog validation algorithm of SWS/AIC system is analyzed for its completeness, time dependency, and correctness as per the requirement specification and design description. The functional and the nonfunctional requirements are analyzed. Various scenarios are analyzed to check the robustness of the algorithm. The results are plotted for further analysis. The approach can be validated complex algorithm as it provides better understanding of the algorithm.

Keywords Model-based programming · Verification & validation
Complex algorithm · SWS/AIC · Verification cases and procedure

R. Krishnaprasad · M. Nanda (✉) · J. Jayanthi · S.S. Dhage
Aerospace and Electronics Division, CSIR- National Aerospace Laboratories, Bangalore, India
e-mail: manjun@nal.res.in

R. Krishnaprasad
e-mail: krishsnaprasad.rajan90@gmail.com

J. Jayanthi
e-mail: jayanthi@nal.res.in

S.S. Dhage
e-mail: shyam_gd@nal.res.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_38

1 Introduction

Most of the industries today are highly dependent on software for their basic and advanced functionalities. Safe and reliable software operation is significant requirement for many type systems such as in aircraft and air traffic control, space applications [1]. The use of safety critical software rapidly increases for the complex safety critical systems such as military, civilian aircraft systems, nuclear plants, and medical devices, those system failures could result loss of system life, performance, signal data mismatch for system requirements, and also damage to the environment [2].

Most of the complex embedded systems are analyzed for robust scenarios under extreme failure conditions and uncertain environment. Robust scenarios are required to operate the system for longer period of lifetime with minimal attention [3]. Many of the embedded programming languages such as MATLAB/Simulink, Statemate, MatrixX, or LabVIEW [4] are capable of designing, creating, validating, simulating, and controlling the complex embedded models.

Laboratory Virtual Instrumentation Engineering Workbench (LabVIEW) [5] is one of the MBP tool from National Instruments which allows a design engineer to enjoy the benefits of model based programming. LABVIEW is a graphical programming language that uses icons instead of lines of text to create applications. In contrast to text-based programming languages, where instructions determine program execution, LABVIEW uses dataflow programming, where the flow of data determines execution.

In this paper we discuss model-based approach for signal data validation to provide the analysis of validation algorithm at the design stage rather than the code stage. The paper discusses detailed analysis, implementation and validation of analog input data used in SARAS aircraft developed as a digital computer based Stall Warning System and Aircraft Interface Computer (SWS/AIC) system [6]. The SWS/AIC computer are a dual channels system, which are powered independently from the direct current (DC). Each channel of the computer will receive ARINC 429 signals, analog signals, and discrete signals.

The work proposes the model-based workflow for validation algorithm with the case study of analog inputs used as aircraft sensors for the SARAS aircraft. LABVIEW tool suite is used to implement the proposed algorithm which satisfies all the requirements of the SWS/AIC specification document. The manually generate test cases based on the hardware–software integration (HSI) verification cases and procedures (VCP) report [7] are mapped to model–based analysis for validating the completeness and correctness of the algorithm.

The paper is organized in various sections are: Section 2 discusses about the related works done by the researchers that has been useful for this work, Sect. 3 discussed the case study and its analog input signal characteristics to be validated, Sect. 4 describes the proposed validation algorithm implemented using LABVIEW tool suite, Sect. 5 discusses the algorithms for generate the test cases from the experimental results available on the HSI_VCP report, Sect. 6 discusses the simulation results, and Sect. 7 discusses the conclusion and future work.

2 Related Work

Model-based programming has been implemented in the development of safety critical embedded software systems in industry. Usha D [8] demonstrated the design for automatic gain control (AGC) using the model-based programming approach by means of LabVIEW tool.

Bringmann [4] discusses model-based development process from requirements to code, ensuring the implemented systems are complete and behave as expected. As per Bringmann model based development allows segregation of concerns; technical aspects such as fixed-point scaling (i.e., the transformation of floating point algorithms to fixed-point computations), calibration data management, and the representation of signals in memory separated from the core algorithms thereby keeping the models as lean as possible.

Duchesne [9] describes the model-based approach to validate the data of industrial case study, which compares other data driven techniques. Jensen [10] provides a complete model-based design by coupling the system to its environment, physical processes, and embedded computations.

In this paper, we design and validate the SARAS aircraft sensors subsystems with the analog input data, functionality and validation algorithm. The simulation results provide the functionality and performance of analog input data based on the validation algorithm.

3 Case Study

The SWS/AIC computer [6] are used as case study for validating the complex algorithm using model-based workflow. SWS/AIC system consist of dual channels, which are powered independently from the direct current (DC) Essential Bus. Each channel of the computer receives ARINC 429 signals, analog signals, and discrete signals.

Each channel of the computer receives signals from the same side angle of attack (AOA) transducer, self/cross-side air data control unit (ADCU), self/cross-side attitude heading reference unit (AHRU) (normal and lateral accelerations, altitude and altitude rate information, body axis rates, airspeed and indicated dynamic pressure); the flap, and landing gear position sensors. These inputs are processed, monitored and sent as an output to caution warning panel (CWP), pitch trim servo actuator, ARINC low-speed bus and shaker actuator. Each processor has the provision for input signal selection and monitoring.

3.1 Signal Data Validation Algorithm

Validation algorithm in SWS/AIC system is from different types of signals such as analog, discrete and ARINC inputs. Each of these signals is different from one another. In the case of ARINC signals, every cycle is valid signal. Initial predefined safe values are used for the first ten fast cycles in case of analog signals and for discrete signals the initial values are used for the first twenty fast cycles. The Analog input data validation is executed once in every 25 ms. There are three variables used for each input signal to store the current, previous, last previous values. The data read every 25 ms is the current value. The control and data flow for the signal data validation shown in [Fig. 1].

The algorithm performs several functionalities such as store current value (CV), previous value (PV), and last previous value (LPV). After storing the values, comparison of absolute difference between LPV, PV & PV, CV, and LPV, PV, PV & CV within tolerance for persistence time is performed. Tolerance value for

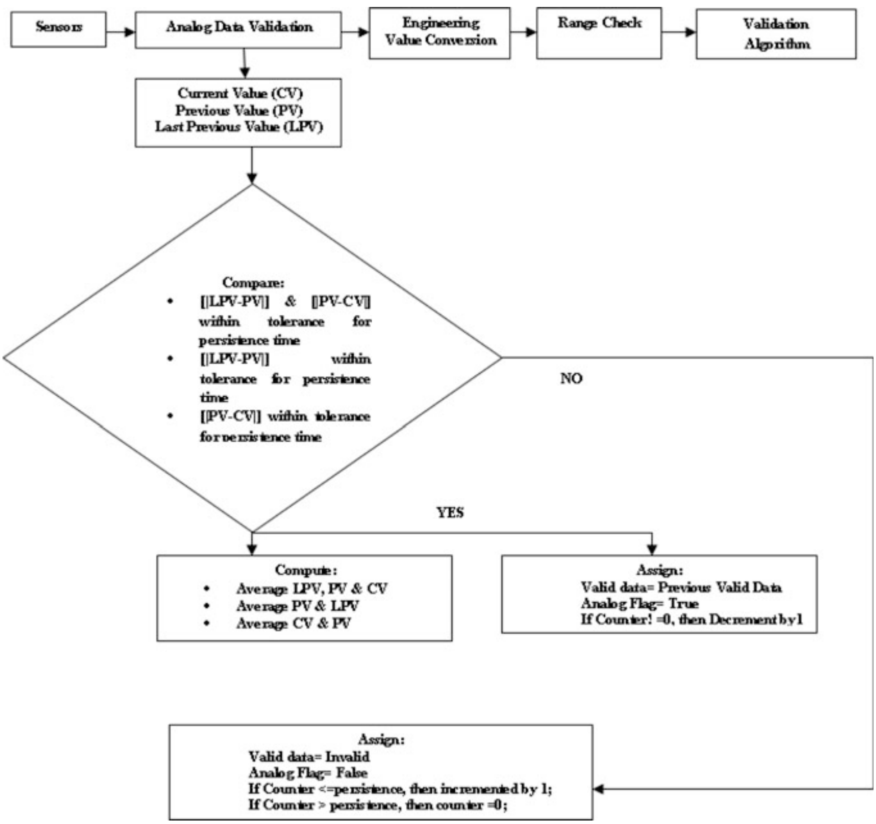


Fig. 1 Control and data flow for the signal data validation

analog signal is 2% variation of maximum analog input. Persistence time for analog input is ten fast cycles. If the comparison algorithm is satisfied, average of LPV, PV & CV, LPV & PV, and PV & CV are computed.

4 Implementation

The complete model-based validation algorithm, which is used for design and implementation of logic for model-based programming is performed using LabVIEW tool suite. This implementation model not only fully satisfies the functionalities of analog data validation followed by the SWS/AIC system requirements and also proven more robustness compare to the any other model-based programming platform. Here, we validate the analog data for pitch trim sensor functional characteristics.

4.1 Model Based Analog Signal Validation Algorithm

Figure 2 shows block diagram representation for model based validation algorithm of analog input data given by sensors. The algorithm acquires analog signals, validates the data, converts validated data to engineering values such as degrees, radians, kilograms, provides health status flag indicator by means of Boolean values (0 or 1).

Figure 3 shows the block diagram which shows the algorithm control flow. The block diagram shows the major functionalities such as acquire signal values, categorize signal as current, previous and last previous values for the acquired signal values, compute absolute differences, compare differences: |difference A| and |difference B| with respect to the given tolerance value. The comparison algorithms generate output such as validated data, counter output, engineering values of the validated data, and health status flag.

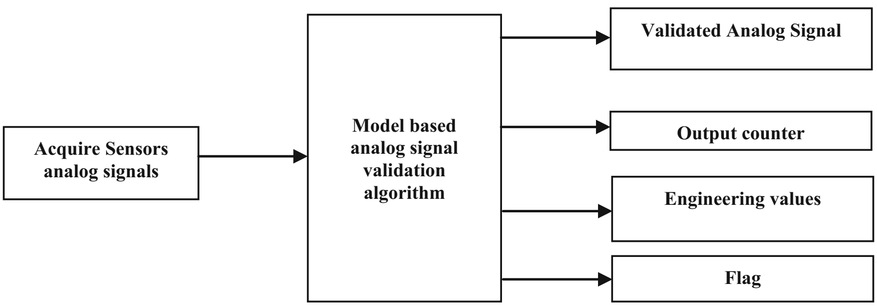


Fig. 2 Block diagram representation for model based analog signal validation algorithm

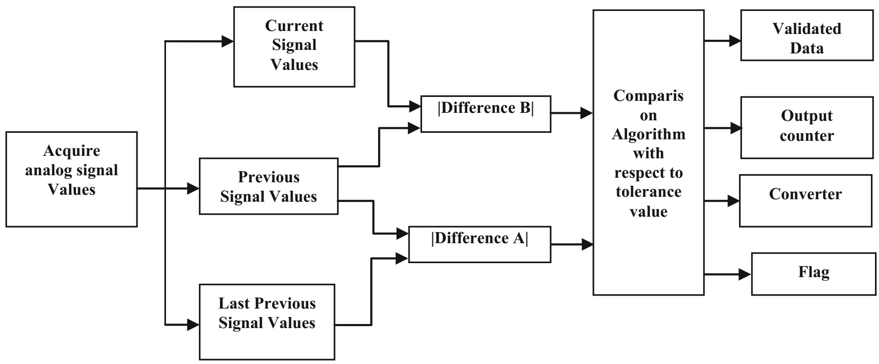


Fig. 3 Block diagram for detailed explanation of model-based analog validation algorithm

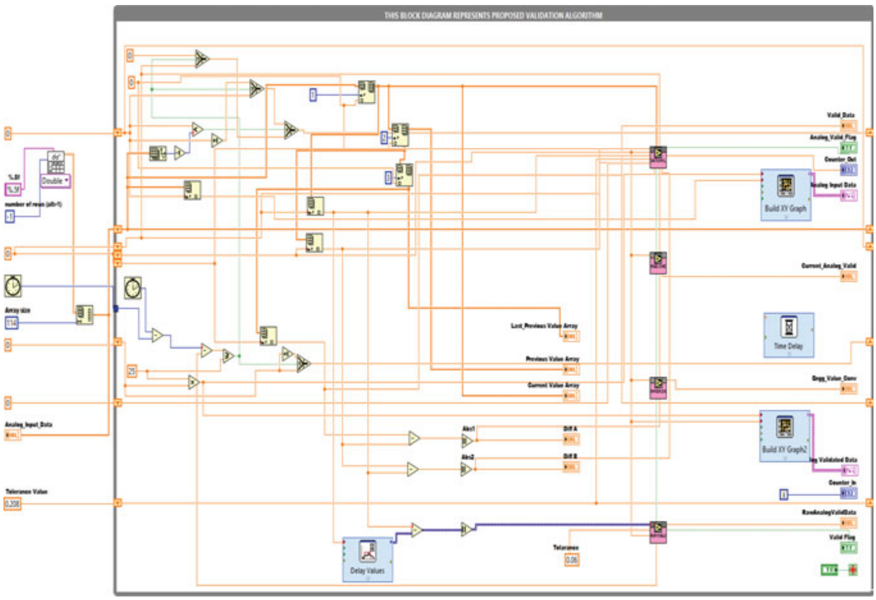


Fig. 4 LABVIEW block diagram panel representation for overall implementation model

Figure 4 shows the complete block diagram panel design for the analog validation algorithm satisfies complete functional analysis of the pitch trim sensor used in the stall warning system of SARAS aircraft and implemented using NI LabVIEW. The overall implementation model consists of four subsystems (in LabVIEW used as term subvi) such as analog validation algorithm implementation subsystem for all sensors, analog voltage range limit subsystem, other two subsystems such as analog input range test, analog input validation test for generate test cases based on the Hardware–Software Integration (HSI) verification cases and procedures (VCP) report for external input/output processing of SARAS aircraft.

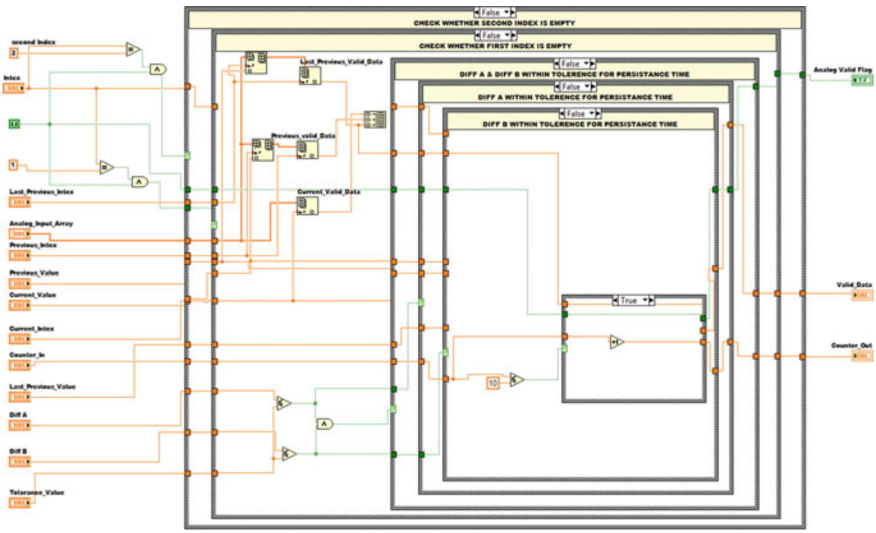


Fig. 5 LABVIEW subsystem representation for logical implementation of validation algorithm

Figure 5 shows subsystem implementation model, which not only satisfies all the logical relations depending upon the analog algorithm mentioned in the previous section and also provides expected outputs such as validated analog data, input analog waveform, validated analog waveforms, and output counter. All the functional analysis results is clearly explained in the Sect. 6 and Table 1 indicates the functional analysis result values.

4.2 *Analog Data Conversion Algorithm*

The pitch trim signal algorithm provides conversion of validated analog input data into their engineering values such as angle (degrees), mass (kilograms), torque (psig), and pressure (psi).

Figure 6 shows design for the validation algorithm of range limit the pitch trim sensor validated data. Pitch trim sensor validated data is limited in the range [min = 0 v, max = 8 v]. The range limit outputs indicate in the column of Table 1 shown in Sect. 6.

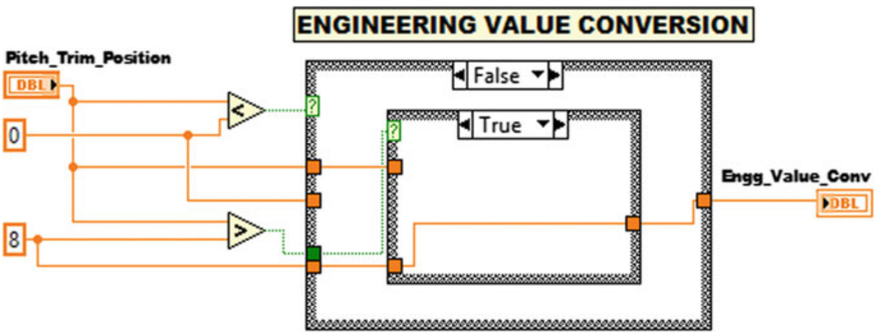


Fig. 6 LABVIEW subsystem representation for conversion of validated analog data into the engineering values

Table 1 Table indicates few of the simulation result values, which satisfies all requirements of analog data validation algorithm

Sl.No	Input	CV	PV	LPV	C_In	Diff A	Diff B	C_Out	Flag	Valid_Data	Range Limit	Angle Range Limit
1	11.3	11.3	11.3	11.3	0	0	0	0	1	0	8	8
2	11	0	0	0	1	0	0	0	1	11.3	8	8
3	10.9	11	0	0	2	0	11.3	0	1	0	0	-15
4	10.7	10.9	11	0	3	11.3	11	0	1	11	8	8
5	10.41	10.7	10.9	11	4	0	10.9	0	1	5.45	5.45	2.56689
6	9.913	10.41	10.7	10.9	5	11	0.3	1	1	11	8	8
7	9.733	9.913	10.413	10.7	6	10.9	0.487	2	1	10.9	8	8
8	9.583	9.733	9.913	10.413	7	0.3	0.787	3	1	10.7	8	8
9	9.083	9.583	9.733	9.913	8	0.487	0.68	4	1	10.413	8	8
10	8.879	9.083	9.583	9.733	9	0.787	0.33	5	1	9.913	8	8
11	8.873	8.879	9.083	9.583	10	0.68	0.65	6	1	9.733	8	8
12	8.818	8.873	8.879	9.083	11	0.33	0.704	7	1	9.583	8	8
13	8.368	8.818	8.873	8.879	12	0.65	0.21	8	1	9.083	8	8
14	8.154	8.368	8.818	8.873	13	0.704	0.061	7	1	8.8485	8	8
15	8	8.154	8.368	8.818	14	0.21	0.505	8	1	8.873	8	8
16	7.723	7.783	7.789	8	17	0.664	0.365	9	1	8.154	8	8
17	7.223	7.723	7.783	7.789	18	0.368	0.217	10	1	8	8	8
18	7.043	7.15	7.157	7.223	21	0.066	0.566	9	1	7.473	7.473	8
19	6.853	7.043	7.15	7.157	22	0.56	0.073	8	1	7.1865	7.1865	8
20	6.841	6.853	7.043	7.15	23	0.566	0.114	7	1	7.1	7.1	7.79574
21	6.701	6.841	6.853	7.043	24	0.073	0.297	6	1	7.038	7.038	7.59926
22	6.686	6.701	6.841	6.853	25	0.114	0.202	5	1	7.01367	7.01367	7.52215
23	6.485	6.686	6.701	6.841	26	0.297	0.152	4	1	6.777	6.777	6.77215
24	6.469	6.485	6.686	6.701	27	0.202	0.155	3	1	6.85667	6.85667	7.02462
25	6.319	6.469	6.485	6.686	28	0.152	0.216	2	1	6.669	6.669	6.4299
26	6.119	6.319	6.4693	6.485	29	0.155	0.217	1	1	6.65515	6.65515	6.38601
27	5.619	6.119	6.3193	6.4693	30	0.216	0.166	0	1	6.40215	6.40215	5.58425
28	5.613	5.619	6.1193	6.3193	31	0.217	0.35	1	1	6.4693	6.4693	5.79705
29	4.884	5.102	5.4023	5.6133	34	0.7	0.217	2	1	5.6193	5.6193	3.1034
30	4.675	4.884	5.1023	5.4023	35	0.506	0.511	3	1	5.6133	5.6133	3.08439
31	4.37	4.675	4.8843	5.1023	36	0.217	0.518	4	1	5.4023	5.4023	2.41573
32	4.155	4.37	4.6753	4.8843	37	0.511	0.427	5	1	5.1023	5.1023	1.46503
33	3.93	4.155	4.3703	4.6753	38	0.518	0.514	6	1	4.8843	4.8843	0.774187
34	3.685	3.93	4.1553	4.3703	39	0.427	0.52	7	1	4.6753	4.6753	0.111866
35	3.36	3.685	3.9303	4.1553	40	0.514	0.44	8	1	4.3703	4.3703	-0.854679
36	3.149	3.36	3.6853	3.9303	41	0.52	0.47	9	1	4.1553	4.1553	-1.53601
37	2.933	3.149	3.3603	3.6853	42	0.44	0.57	10	1	3.9303	3.9303	-2.24904
38	2.665	2.933	3.1493	3.3603	43	0.47	0.536	11	1	3.6853	3.6853	-3.02544
39	2.378	2.665	2.9333	3.1493	44	0.57	0.427	0	0	0	0	-15
40	2.361	2.378	2.6653	2.9333	45	0.536	0.484	1	1	3.1493	3.1493	-13.5909
41	2.191	2.361	2.3778	2.6653	46	0.427	0.555	2	1	2.9333	2.9333	-14.2754
42	2.045	2.191	2.361	2.3778	47	0.484	0.304	3	1	2.6653	2.6653	-15
43	1.824	1.849	1.855	2.045	50	0.187	0.336	2	1	2.1164	2.1164	-15
44	1.474	1.824	1.849	1.855	51	0.316	0.196	1	1	1.947	1.947	-15
45	-0.486	-0.306	-0.106	-0.1	71	0.194	0.206	0	1	-0.0104	0	-15
46	-0.694	-0.486	-0.306	-0.106	72	0.106	0.38	0	1	-0.243	0	-15

5 Analyzing Algorithm Validation by Means of Test Cases

The correct implementation of analog validation algorithm modeled using the NI LabVIEW is analyzed at the input level and at the algorithm and output level. Table 2 shows the test conditions at the input level. Table 2 shows the input test

Table 2 Table indicates input test conditions and expected outputs required for analog input range test

Test cases	Input		Expected output
	Input name	Test condition (V)	Valid data (°)
1	Pitch Trim Position (PTP)	11.3	8.0
2		8.11	8.0
3		10.85	8.0
4		7.163	8.0
5		7.156666667	8.0
6		7.15	7.964602
7		6.55	6.05279
8		3.39	−12.0
9		2.99	−14.0957
10		2.831	−14.968141
11		2.825	−15.0
12		2.819	−15.0
13		0.0	−15.0
14		−0.1	−15.0

Table 3 Table indicates test purpose conditions, input test conditions, valid flag mode behaviors and expected valid data

Test cases	Test purpose	Input			Expected output
		Input name	Test condition (V)	Valid flag	Valid data (°)
1	PTP input variation < 2% of the nominal	Pitch Trim Position (PTP)	3	1	−14.60885
2	PTP input variation = 2% of the nominal		3	1	−14.60885
3	PTP input variation > 2% of the nominal for less than 250 ms		3	1	−14.60885
4	PTP input variation > 2% of the nominal for 250 ms		3	1	−14.60885
5	PTP input variation > 2% of the nominal for greater than 250 ms		3	0	0

condition and expected output for pitch trim analog input test. Table 3 shows the input test condition and expected output for pitch trim analog input validation test.

5.1 *Algorithm for Generate Test Cases of Pitch Trim Analog Input Range Test*

Table 2 indicates input test conditions and expected outputs for generate test cases for pitch trim analog input range test available in the HSI_VCP report.

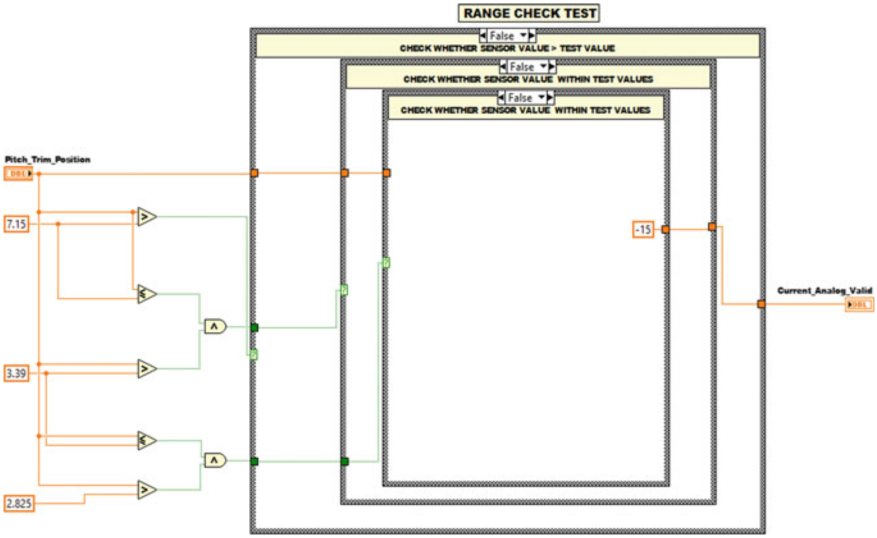


Fig. 7 LabVIEW block diagram panel design for generate test cases of pitch trim analog input range test

Figure 7 shows the implementation of the pitch trim analog input range test. The model design satisfies all the input conditions for pitch trim analog input test.

5.2 Algorithm for Analysis Test Conditions of Pitch Trim Analog Input Validation Test

Table 3 indicates input test conditions and expected outputs for generate test cases for pitch trim analog input validation test available in the HSI_VCP report.

Figure 8 shows the complete design for the validation algorithm. Table 3 provides the test scenarios for the validation algorithm. The model-based design satisfies all the test purpose conditions, input test conditions, valid flag mode behaviors and expected valid data.

6 Simulation Results

Simulation results discusses about the overall LABVIEW front panel design and all the expected simulation results given by each subsystems. The simulation results such as analog input data waveform, validated analog input waveform, validated pitch trim voltage range limit and validated pitch trim angle range limit. Table 1

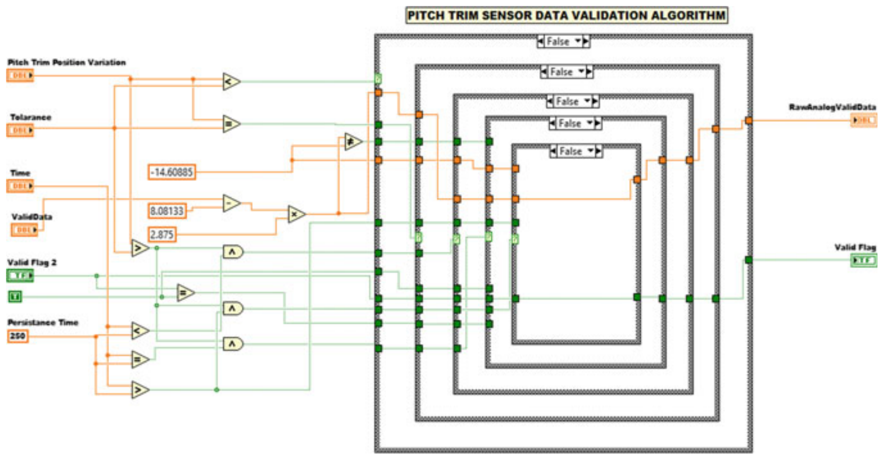


Fig. 8 LABVIEW block diagram panel design for generate test cases of pitch trim analog input validation test

Fig. 9 LABVIEW front panel representation for overall implementation model

indicates few of the simulation output data available from the various implementation models are designed on the LABVIEW tool suite.

Figure 9 shows the overall front panel design representation of analog data validation algorithm. All the available data as per the requirements for analog data validation algorithm has been tabulated in Table 3.

Figure 10a shows the analog input data waveform varies from +11.3 V to −11.3 V. Figure 10b shows the validated analog input data waveform. The corresponding values of analog input data and validated analog input data are illustrated in Table 1 (Fig. 10).

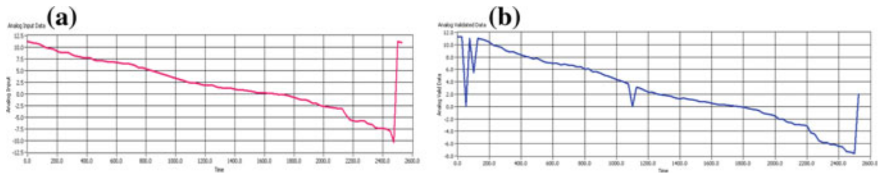


Fig. 10 **a** Analog input data waveform. **b** Analog validated input data waveform

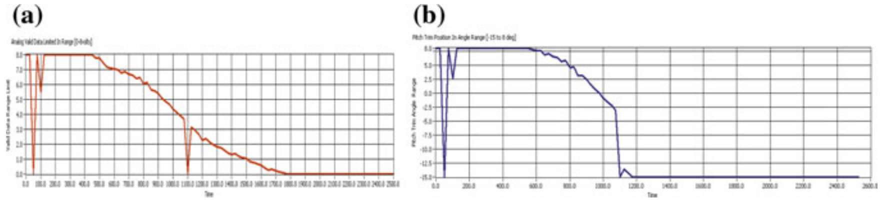


Fig. 11 **a** Waveform Representation for pitch trim voltage range limit [0–8 V]. **b** Waveform representation for pitch trim angle range limit [−15° to + 8°]

Figure 11a indicates the waveform, which limit the analog voltage input of pitch trim sensor data in between the range of 0–8 V. Figure 11b indicates the waveform, which limit the angle of pitch trim sensor data in between the range of −15° to +8°. The corresponding values of voltage range limit and angle range limit of pitch trim sensor are illustrated in Table 1.

The outcome of the model-based algorithm analysis provides an effective approach to validated complex algorithm. The approach was capable of analyzing the functional and performance properties of the validation algorithm which was not possible at the hardware testing level due to the limitation of the testing setup. The result obtained at the model level provided the confidence of the correctness and completeness of the algorithm as per the project specifications for the functionality, performance for timing and safety properties.

7 Conclusion and Future Work

The proposed mode-based approach can be used for complementing the areas where it becomes difficult to demonstrate the functionality and performance of complex functionality in a system. This model-based approach proven not only more robust to analyze the functional requirements of the system, but also generates more number of test cases based on algorithms.

This approach can be used to analyze and understand the completeness, correctness of the complex functionality in a system and also more convenient way to design and verify complex algorithm compared to other scripting languages may have more than 200 lines of codes. LABVIEW enabled in modeling and analyzing the validation algorithm of the SWS/AIC software. The analysis and the understanding of the algorithm was not only easier but also the results could be analyzed for improving the logic further.

Acknowledgements The authors thank the Director, CSIR-NAL for his support and encouragement to carry out this research work.

References

1. Pullum, L.L, "Software Fault Tolerance Techniques and Implementation", pp. 1–358(2001). http://www.artechhouse.com.
2. John C. Knight, "Safety Critical Systems: Challenges and Directions", pp. 1– 4 (2001).
3. Brian C. Williams, Michel D. Ingham, Seung H. Chung, and Paul H. Elliott, "Model-Based Programming of Intelligent Embedded Systems and Robotic Space Explorers", Proceedings of the IEEE, Vol. 91, pp. 212–237 (2003).
4. Eckard Bringmann, Andreas Kramer, "Model-based Testing of Automotive Systems", IEEE, pp. 485–493 (2008).
5. Ronald W. Larsen, "Labview for Engineers, Pearson Education", pp. 1–406 (2011).
6. J. Jayanthi, Dr. Manju Nanda, Sreekath K P and Lakshmi P, "Software Design Description for SARAS SWS/ AIC System, National Aerospace Laboratories_ ALD", Volume-10, Rep No: DR-22.
7. Nagaraj, "Hardware Software Integration (HSI) Verification Cases and Procedure for External Input/output Processing Test", National Aerospace Laboratories_ ALD, pp. 1–123 (2005).
8. Usha D, V K Agrawal, Anandita Chakraborty and Kochaleema K H, "An Approach for Model Based Programming using LabVIEW", Proceedings of the Second International. Conference on Advances in Computer and Information Technology - ACIT, pp. 191– 195 (2013).
9. Arnaud Duchesne, Georges Heyen, Philippe Mack and Boris Kalitventzeff, "Combined use of Model based Data Validation and Data driven Techniques for Process Monitoring", 17th European Symposium on Computer Aided Process Engineering – ESCAPE17, pp. 1–6 (2007).
10. Jeff C. Jensen, Danica H. Chang and Edward A. Lee, "A Model-Based Design Methodology for Cyber-Physical Systems", Proceedings of First IEEE Workshop on Design, Modeling and Evaluation of Cyber-Physical Systems, pp. 1–6 (2011).

Impact of Different Ground Plane Structures on Sierpinski Monopole Antenna

D.M.K. Chaitanya and N.V. Koteswara Rao

Abstract The objective of this paper is to analyze the impact of rectangular and circular ground plane structures on Sierpinski monopole gasket and their effect on resonant frequency, radiation patterns. Since the Sierpinski gasket is a fractal structure, the number of resonances depend on the number of iterations, the scaling factor and flare angle of the antenna. The impact of size, shape of the ground plane structure on the resonant frequencies, patterns have been observed. It is also observed the deviations in those parameters for change in the positions of the ground plane to the antenna structure with circular and square ground planes. The Sierpinski monopole gasket with rectangular and circular ground plane structures was analyzed along with changes in the positions of the ground plane using HFSS tool and the simulation analysis, measured results of the fabricated structure of the same has been reported. It was concluded that the larger the ground plane, there is a deviation in the resonant frequencies for the same structure, the more the tilt in the pattern and hence suitable for ground to airborne applications.

Keywords Ground plane · Sierpinski monopole · Fractal structure
Resonant frequency · Gain · Field patterns

1 Introduction

Monopole antennas, in general are placed on a ground plane and will have a similar performance like a dipole structure and are popular in terms of compactness. These antennas have gained lot of attention in the wireless industry because of their Omni directional characteristics, low profile, low cost and are suitable for commercial as

D.M.K. Chaitanya (✉)
Department of ECE, Vardhaman College of Engineering, JNTUH (PT), Hyderabad, India
e-mail: chaitanyadmk@gmail.com

N.V. Koteswara Rao
Department of ECE, CDAAC, CBIT, Gandipet, Hyderabad, India
e-mail: nvkr1@rediffmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_39

well as military applications. The major advantage of the monopole antenna is that the operational characteristics like gain, directivity, resonant frequencies can be improved by modifying the ground plane either in size, shape or changing the distance instead of replacing the complete antenna structure. The ground plane may be rectangular, circular in shape. The current distribution through the antenna structure determines its resonant frequency [1] and the pattern is obtained by the case parameters like size, shape and distance. The proper feeding technique must be incorporated to have impedance matching and the ground plane must be more than the quarter wavelength over the feed [2]. The emerging trends in wireless systems demands the antenna to be operated over wide band as their standards employs different operational bands. Besides this, the major desirable attributes in the antenna design are performance, low profile, with minimum cost, simple installation of the antenna. One of the blooming techniques is to construct the Fractal Structures. The Sierpinski Fractal structure is named after the polish mathematician Sierpinski in 1915 and it is developed by Puente in 1996. This paper briefs the effects of rectangular, circular ground plane structures with varying distances from the ground plane to the antenna on Sierpinski monopole and the analysis along with results. The structure was simulated in High-Frequency Structure Simulator (HFSS) tool [3] and the same has been fabricated and tested in anechoic chamber and measured results were presented in this paper.

2 Concepts and Theory

The dimension of the monopole antenna is half of the size of dipole antenna and their performance resembles as it is mounted over a ground plane. The operation of the monopole antenna structure can be verified theoretically using the image theory. The image theory states that the ground plane acts like a mirror to the antenna structure to behave like a dipole [4, 5]. The monopole as shown in Fig. 1 will have the single element on the ground plane and is fed with respect to ground. The generated electromagnetic waves are reflected or absorbed from the ground plane and they depend on the conductive, dielectric properties of the ground structure because of the change in the impedance. The impact of this impedance in turn alters the resonant frequencies, radiation characteristics of the antenna. The impedance effects can be minimized by using the twice the wavelength of the antenna structure [6]. The fields generated below the ground plane are zero, where as above the ground plane, depends on the characteristics of the antenna structure. Hence, the monopole operation is similar to the operation of the dipole structure by the theory of images with radiation above the ground plane. The only difference between dipole and monopole is that the impedance is one-half of the dipole structure with improved gain and band width performance. As the microstrip antennas finds many applications in wireless communications in recent applications because of their compactness, ease of integrations and low cost [7], a printed circuit fractal antenna can better fit into the multi band applications. The organisation of the paper is

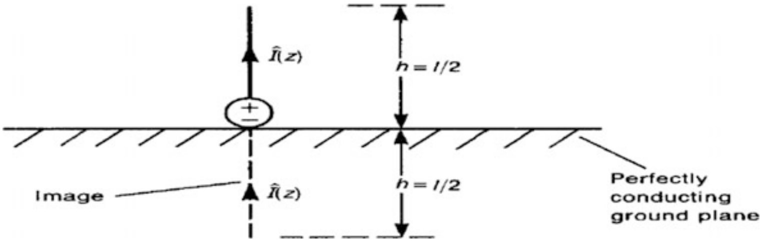


Fig. 1 The monopole antenna with perfectly conducting ground plane

divided in to different sections. The Sect. 2 explains about the theoretical concepts of monopole antenna with image theory and Sect. 3 briefs about the design of Sierpinski monopole antenna structure. The results were discussed in the Sect. 4 and conclusions are drawn in the Sect. 5 followed by references.

From the concept of the image theory, the fields can be evaluated theoretically sing the following formulae.

The electric field intensity:

$$E_\theta = \frac{j\eta I_0}{2\pi r} e^{j\beta r} \left(\frac{\cos(Bl\, cos\theta) - \cos(\beta l)}{\sin\theta} \right) \tag{1}$$

where

E_θ = Electric Field intensity
η = Intrinsic impedance
β = Phase constant
I_0 = The maximum current
L = length of the monopole antenna
r = radius

The magnetic field intensity is given by:

$$H_\varphi = \frac{E_\theta}{\eta} = \frac{jI_0}{2\pi r} e^{j\beta r} \left(\frac{\cos(Bl\,\cos\theta) - \cos(\beta l)}{\sin\theta} \right) \tag{2}$$

where

E_θ = Electric Field intensity
β = Phase constant
I_0 = The maximum current
l = length of the monopole antenna
r = radius

The average power density which is the pointing vector is given by

$$S_{avg} = \frac{1}{2}\mathrm{Re}(E_\theta X H_\varphi^*) = \vec{r}\frac{\eta I_0^2}{8\pi^2 r^2}\left(\frac{\cos(\beta l \cos\theta) - \cos(\beta l)}{\sin\theta}\right)^2 \quad (3)$$

The radiation intensity is given by

$$U = r^2 \mathrm{X} S_{avg} = \vec{r}\frac{\eta I_0^2}{8\pi^2 r^2}\left(\frac{\cos(\beta l \cos\theta) - \cos(\beta l)}{\sin\theta}\right)^2 \quad (4)$$

The directivity is given by

$$D = \frac{4\pi U(\theta, \varphi)}{\int_0^{2\pi}\int_0^{\pi} U \sin\theta d\theta d\varphi} \quad (5)$$

The radiation efficiency of the antenna is given by

$$\eta_e = \frac{p_t}{p_i} \quad (6)$$

where

p_t = Total radiated power by the antenna,
p_i = Total input power accepted by the antenna

3 Sierpinski MonopoleAntenna

The Sierpinski gasket is created by eliminating the central triangle from the main triangle [8] with flare angle of 60° [9] with number of iterations. The scaling factor of two is considered with four number of iterations to resonate the antenna at four different frequencies. The antenna structure is mounted over a rectangular, circular ground planes [10, 11] and the ground plane is positioned at different distances. The Fig. 2 shows the basic Sierpinski monopole antenna. The schematic drawn in HFSS tool for the Sierpinski monopole antenna with rectangular ground plane and with Circular ground plane are shown in Fig. 3 and Fig. 4 respectively.

The antenna is fabricated on FR4 substrate (εr = 4.4, h = 1.6 mm) and placed over a 80 × 80 cm rectangular ground plane as seen from the structure 1 where as a circular conducting ground plane having dia of 15.6 cm in the other structure. The feeding mechanism is through a coaxial SMA connector suitable for this type of configurations. The antenna is designed for 4 number of iterations with a scaling factor of two. The corresponding antenna heights are given by 4.75 cm, 2.375 cm, 1.187 cm, and 0.59 cm and the structure resonates at four different frequencies with multiples of two. The resonant frequency can be calculated by using the formula

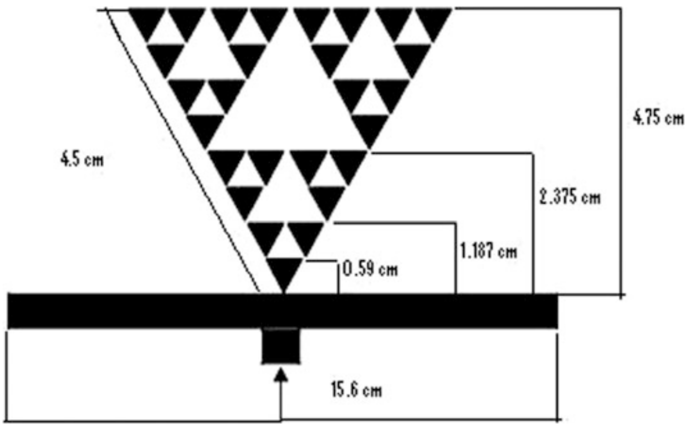


Fig. 2 Sierpinski gasket

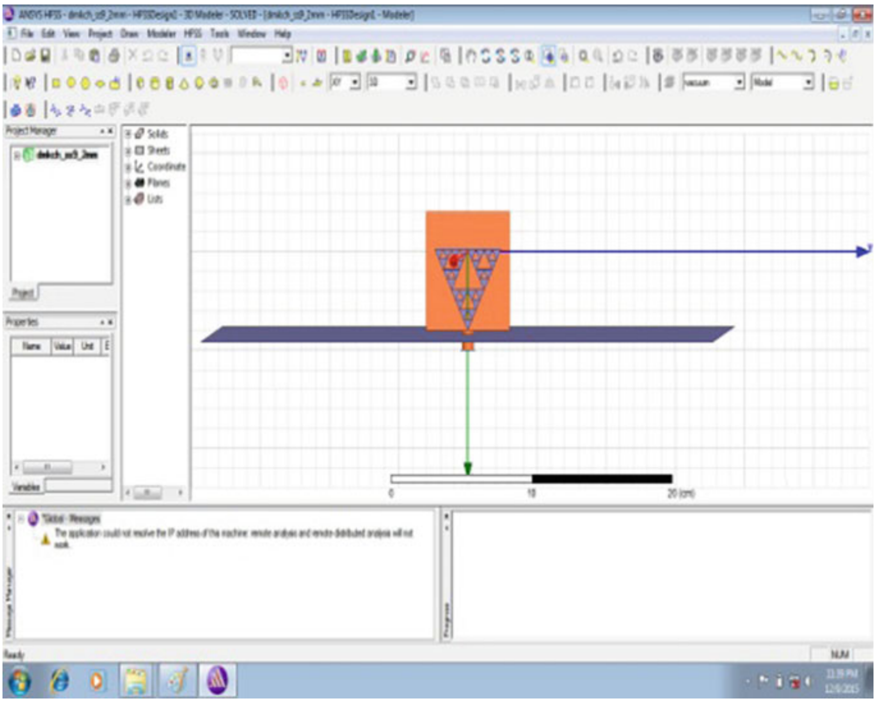


Fig. 3 Rectangular ground plane

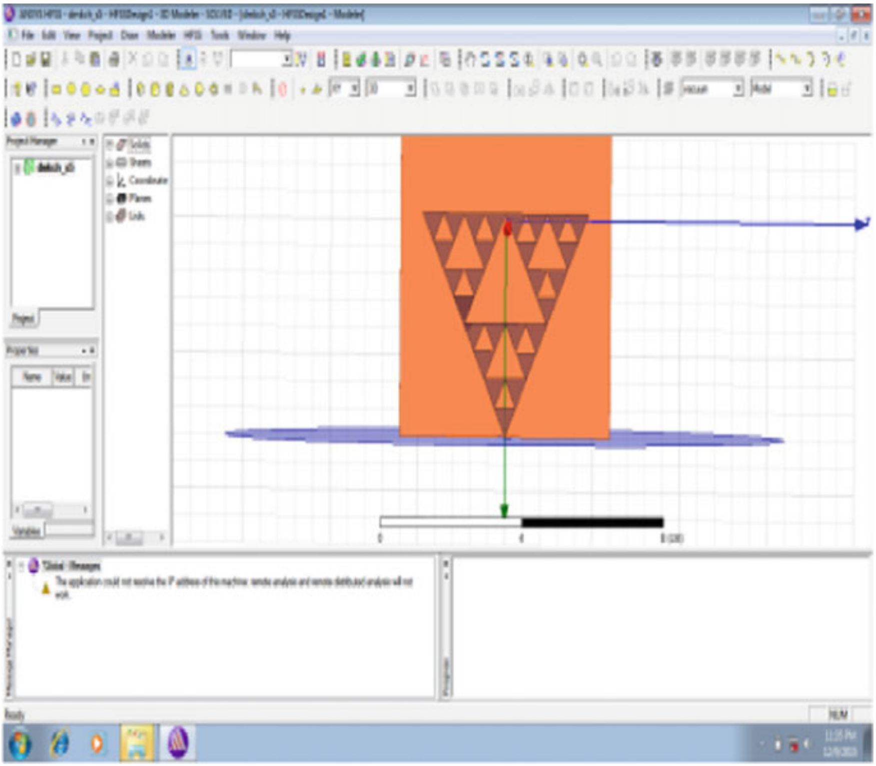


Fig. 4 Circular ground plane

$$f_r = K\frac{c}{h}\cos\frac{\propto}{2}\delta^n \tag{7}$$

where

K = Constant and is 0.152 for Sierpinski Gasket for a given substrate
α = *flare angle and is* 60^0 for equilateral triangle
δ = scaling factor and is 2 and n = iteration number

The sierpinski gaskets are simulated using HFSS software are using rectangular and circular ground planes for different distances from the ground plane to the antenna structure. The fabricated sierpinski monopole antenna along with rectangular and circular ground planes are shown in Fig. 5 and 6. The measurements were carried out for the structure using a VNA and the results are plotted in the Table 1. The corresponding simulated and measured results with 0.1 mm ground plane spacing with the antenna depicts that the antenna is resonating at four different frequencies because of its geometry.

Fig. 5 Sierpinski gasket on rectangular ground plane

Fig. 6 Sierpinski on circular ground plane

Table 1 Resonant frequencies for rectangular and circular ground planes with 0.1 mm spacing

S.No	f_r (GHz) Resonant frequencies for rectangular ground plane		f_r (GHz) Resonant frequencies for circular ground plane	
	Simulated	Measured	Simulated	Measured
1	0.79	1	1.89	1.88
2	5.07	5	4.01	4.0
3	7.56	7.62	7.89	7.9
4	11.34	11.5	15.3	15.3

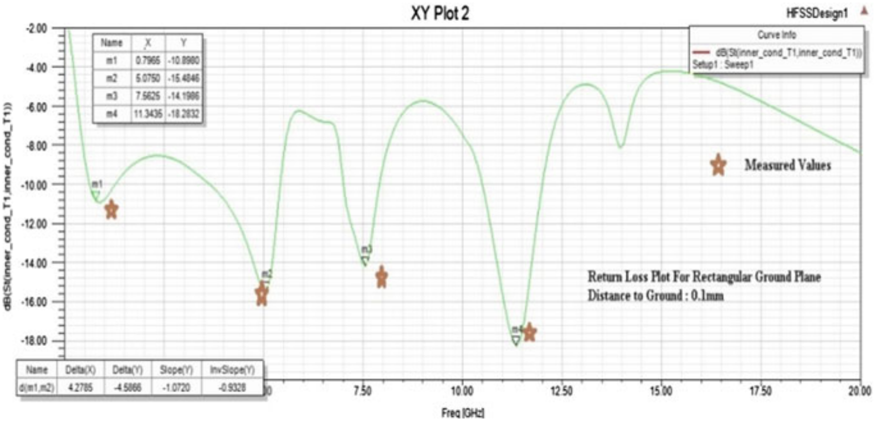


Fig. 7 Return loss plot for the rectangular ground plane with ground to patch distance 0.1 mm

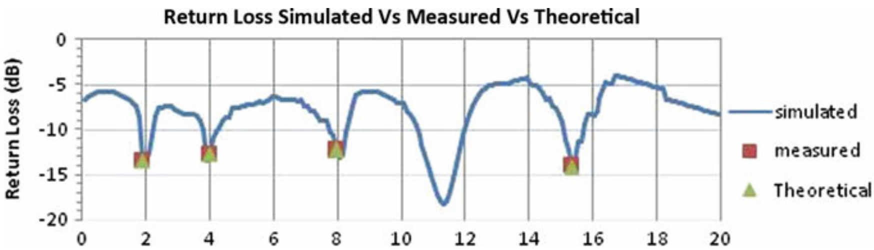


Fig. 8 Return loss plot for the circular ground plane with ground to patch distance 0.1 mm

Fig. 9 Radiation pattern for rectangular ground plane

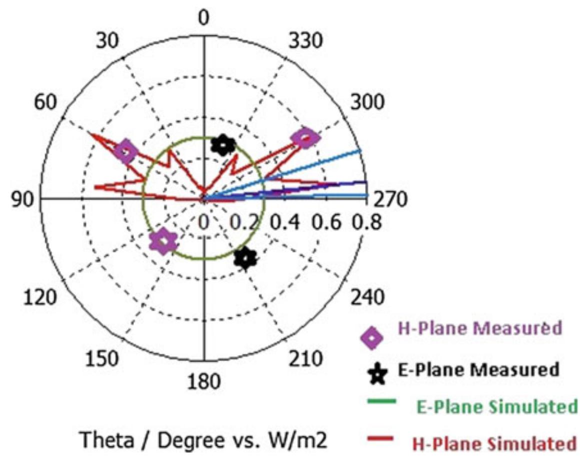
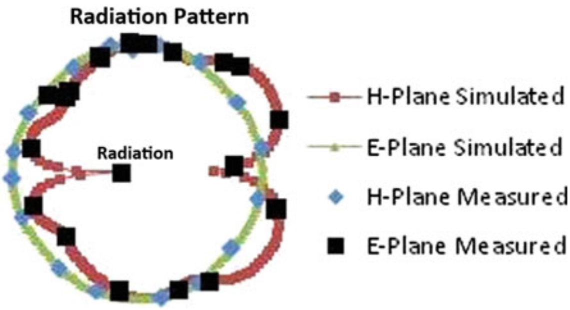


Fig. 10 Radiation pattern for circular ground plane



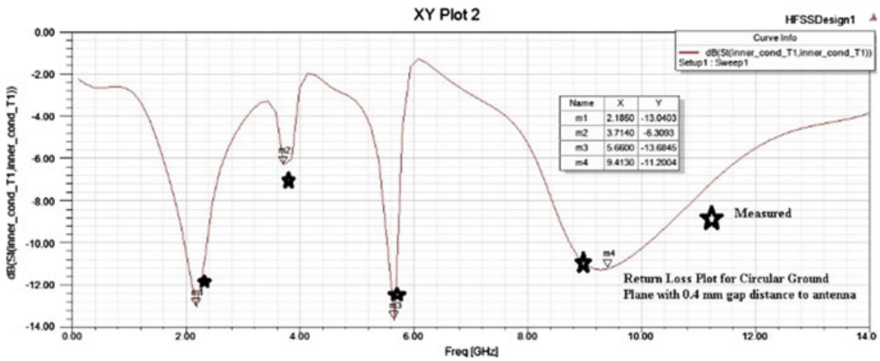


Fig. 11 Radiation pattern for rectangular ground plane with a distance of 0.4 cm from ground plane to Sierpinski monopole

The corresponding return loss plots for the same has been depicted in Fig. 7 and Fig. 8 respectively.

The radiation pattern measurements have been carried out in anechoic chamber and the corresponding pattern details are depicted in Figs. 9 and 10 for rectangular and circular ground places for 0.1 mm spacing.

The corresponding measurements for the rectangular and circular ground planes were carried out with a ground plane distance of 0.4 mm with the antenna structure and the resonant frequency plots can be observed from the Figs. 11 and 12 respectively.

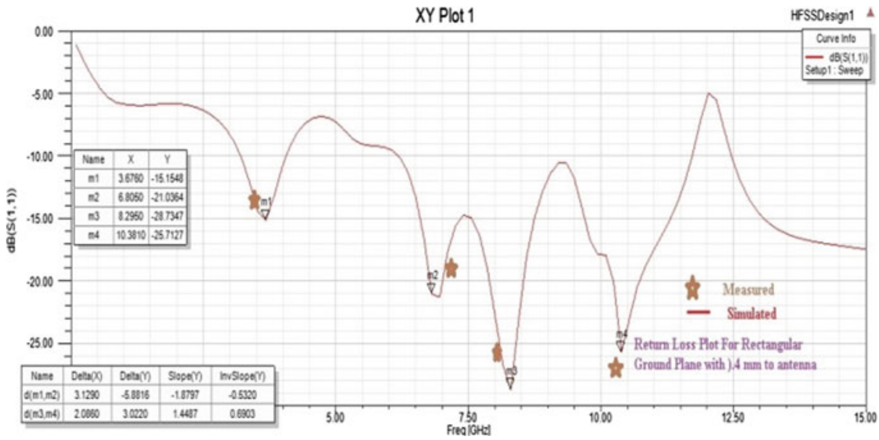


Fig. 12 Radiation patten for circular ground plane with a distance of 0.4 cm from the ground plane to Sierpinski monopole

4 Discussion of the Results

Since the fractal shapes appear at four different scaling factors, the Sierpinski monopole resonates at four different frequency bands. It can be viewed clearly from the return loss plots that the Sierpinski gasket demonstrates the log periodic behavior, and the scaling factor (δ) of two makes it to resonate at multiple of two. In particular, the antenna is matched at four frequencies for rectangular ground plane with the ground to antenna at a distance of 0.1 mm, i.e., at 11.5, 7.56, 5, 1 GHz and for Circular ground plane the resonant frequencies are given by 15.3, 7.9, 4, 1.88 GHz with the same distance.

The measured radiation patterns of the Sierpinski monopole is presented in Figs. 9 and 10 for rectangular, circular ground planes respectively. The patterns corresponding to the flare angle $\alpha = 60°$, it is observed that the tilted figure of eight above the ground plane depicts the monopole action in E-Plane and Omni directional pattern in H-Plane. As the distance from the ground plane to antenna structure is changed to 0.4 cm, there is a notable change in the resonant frequencies as shown in Figs. 13 and 14 for rectangular, circular ground planes respectively. The corresponding resonant frequencies for the rectangular ground plane are given by 9.5, 5.6, 3.7, 2.2 GHz and for Circular ground plane is given by 10.38, 8.3, 6.8, 3.6 GHz. The radiation patterns clearly show that there is a notable degree of change in the gain and directivity with the size and shape of the ground plane. It is observed that the radiation patterns are becoming more directive for the larger ground planes and more specifically with the 0.4 mm ground plane spacing as seen from the results. It is also observed that there is a drastic change in the performance of the antenna with the change of ground planes with shape, size and with the distance to the ground plane interms of return loss, radiation characteristics.

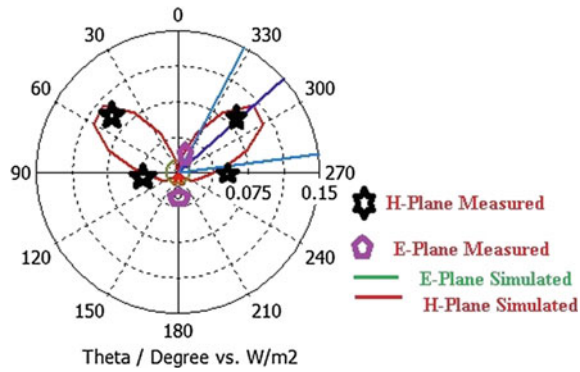


Fig. 13 Return loss plot for rectangular ground plane with a distance of 0.4 cm from gorund plane to Sierpinski monopole

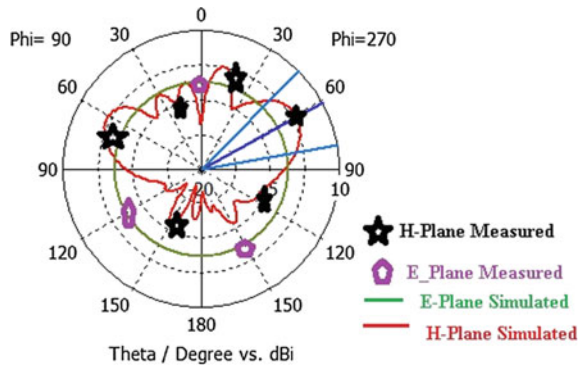


Fig. 14 Return loss plot for circular ground plane with a distance of 0.4 cm from ground plane to Sierpinski monopole antenna

5 Conclusions

The Sierpinski monopole antenna with rectangular ground plane of size 8X80 cm and circular ground plane of dia of Φ 15.6 cm along with simulated and measured results has been reported. The resonant frequencies differs with the shape and size of the ground plane however they follow the basic operation of the antenna structure. It is observed that the scaling factor of the fractal determines the structure to be resonated at a multiples of two in the present case. There are slight mismatches in the resonant frequencies due to the truncation effects and followed by the fabrication difficulties. The monopole structure has the similarity in operation of the dipole structure is observed from the results. It is noteworthy to conclude that the more the size of the ground plane, the more the tilt in the radiation characteristics of the monopole structure. The bigger the ground plane with more spacing between the ground plane to antenna makes the radiation more directional, the structures are better suitable for the airborne applications, where as the structures with smaller ground plane can be used for ground applications. The impact of the distance from the antenna to the ground plane plays a role in change in the resonant frequencies, radiation patterns as observed from the results.

These structures can be further modified to achieve reconfigurability to fit for Cognitive Radio applications simply by incorporating the switching mechanism using RF switches in the antenna structure internally or externally which further reduces the interference with improved gain and band width performance. These structure find the application in wireless communications particularly for mobile and UWB and Radar applications because of their multi band operation with increased directivity, gain and band width because of the monopole structures.

References

1. K. P Ray and G. Kumar, “Broad Band Microstrip Antennas,” Artech House, Boston and London, 2003.
2. J. Liang, C. Chiau, X. Chen and J. Yu, “Study of a circular disc monopole antennas for ultra wideband applications,” 2004 International Symposium on Antennas and Propagation, 17–21 August, 2004.
3. HFSS Tutorials on “High Frequency Structure Simulator “ Ansys, 2016.
4. D. M. Pozar, “Microwave Engineering,” John Wiley & Sons, 2005, 3rd Edition.
5. P. Singhal, “Monopole Antenna with Modify Ground Plane,” International Journal of Engineering & Technology, vol. 1, pp. 266–270, 2012
6. Constantine.A.Balanis. “Antenna Theory: Analysis and Design,” John Wiley-Interscience, 3rd edition, 2005.
7. Rakhesh Singh Kshetrimayum, “Printed Monopole Antennas for Multiband Applications” International Journal Of Microwave And Optical Technology Vol. 3, No. 4, September 2008.
8. Abd Shukur Bin Ja’afar, “Sierpinski Gasket Patch And Monopole Fractal Antenna” M.E Thesis, Faculty Of Electrical engineering, University of Teknologi, Malaysia, 2005.
9. Carles Puente-Baliarda, Jordi Romeu, Rafael Pous,Angel Cardama, “ On the Behavior of the Sierpinski Multiband Fractal Antenna” *IEEE Transactions On Antennas And Propagation,* vol. 46, no. 4, April 1998.
10. J. F. Janek and J. J. Evans, “Predicting Ground Effects of Omnidirectional Antennas in Wireless Sensor Networks,” Wireless Sensor Network, vol. 2, pp. 879–890, 2010.
11. Lusekelo Kibona “Analysis of the Effects of Rectangular Ground Plane on the Radiation pattern of the Monopole Antenna”, International Journal of Scientific Research Publications, Volume 3, Issue 11, November, 2013.

Smart Authentication Using Graphical Password for Android Devices

Aaroh Gala and Somdev Mehta

Abstract Information has become the oxygen of the modern world. To keep information protected from the so-called 'nefarious people' we need security. There are multiple authentication systems for security like text-based, pin, biometrics, graphical, etc. However, the drawbacks of these are that the user tends to keep passwords, which are fairly easy to remember, and such passwords can thus be easily cracked. This paper mainly focuses on the picture password authentication system, which is an upcoming alternative authentication system. According to research, humans tend to recollect images better than text, this system promotes the user to select a password which is difficult to guess but at the same time, can be easily remembered. This paper also focuses on overcoming some drawbacks of Persuasive Cued Click Points (PCCP), along with making the system more secure. The proposed system makes use of five images and captures one section from each of the selected images. For successful login, the users simply have to select the same sections on every image as decided earlier. Thus, users can be authenticated without having to input even a single character. To avoid brute force and any other bypass techniques, the users are allowed only three consecutive failed attempts. Upon failing those, the user account will be locked, and can be unlocked only after answering the registered security questions.

Keywords Persuasive Cued Click Point (PCCP) · Cued Click Point (CCP) Biometric Based Authentication (BBA) · Knowledge Based Authentication (KBA) Text Based Authentication (TBA) · Graphical password · Passpoints HOTSPOTS · MD5 encryption · Image compression

A. Gala (✉) · S. Mehta
Department of CSE, K. J. Somaiya College of Engineering, Vidyanagar, Vidhyavihar (East), Mumbai 400077, India
e-mail: aaroh.gala@somaiya.edu

S. Mehta
e-mail: somdev.mehta@somaiya.edu

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_40

1 Introduction

Authentication has been perceived as that process of a system which enables it to substantiate a user trying to gain access to the system. This complex age of computing has brought about a massive revolution in the authentication systems [1]. There are different types of authentication systems. The major three types of systems are Biometric-Based Authentication (BBA), Knowledge-Based Authentication (KBA) and Text-Based Authentication (TBA) [1]. BBA deals with finger print, palm print, iris recognition and voice recognition. While they are known for their accuracy, their foundation is complex and their main drawback is that, such a system requires extra hardware making these methods costly. Thus this is difficult to achieve in day to day life of common people.

In early days, text-based passwords were used for authentication. Text based passwords are nothing but a string of characters. An important usability goal for authentication systems is to support users in selecting better passwords. Users often create memorable passwords that are easy for attackers to guess, but strong system-assigned passwords are difficult for users to remember [2, 3]. Largely based on the fact of, 'something you know' is always going to be dangerous if the victim is well known to the attacker.

While the predictability problem can be solved by disallowing user choice and assigning random passwords to users, this usually leads to usability issues since users cannot easily remember such random passwords. An authentication system should encourage strong passwords while still maintaining memorability. So researchers of modern days have opted for alternative methods wherein graphical pictures are used as passwords. Graphical passwords essentially use images or representation of images as passwords. Human brain is good in remembering pictures than textual characters [4]. The major goal of this work is to reduce the guessing attacks as well as to encourage users to select more random, and difficult passwords to guess.

The main drawback of PCCP authentication is that the system selects a viewport randomly. PCCP has a shuffle option, which shuffles the viewport randomly in case the user wants a region as opposed to what was given. It might be possible that the region which a user wants does not come with this random shuffle. This may lead to creating an unintentional but strong password, which might not be easy to remember. The transition from text based password to picture based password aimed at making passwords strong and easy to remember, which is not completely fulfilled by PCCP.

The paper is categorized in five sections. Section 2 gives brief outline about various graphical authentication algorithms. Section 3 explains the proposed scheme for smart authentication. The system implementation is explained in Sect. 4. Section 5 concludes the paper and gives future direction.

2 Related Work

There are three different types of graphical passwords. Click-based password scheme, Choice-based password scheme and Draw-based password scheme.

2.1 Pass-Points

S. Wiedenbeck [5, 6] proposed pass-point graphical password scheme. Pass-Point comes under click based graphical password scheme. In pass-points password scheme, the user needs to click at five different spots on an image. To login, the user needs to click on five points in a proper sequence on an image. The main disadvantage of pass-points password scheme is HOTSPOTS (the points which are likely to be selected by the users on a particular image) [7]. The second disadvantage is that user tends to use certain patterns to remember passwords easily. These disadvantages are overcome in the next methods.

2.2 Cued Click Points

Cued Click Points [2] comes under click-choice-based graphical password scheme. Cued Click point was introduced to reduce the major problem faced in pattern and HOTSPOTS in pass-point method [8]. Instead of clicking five different point on the same image, CCP uses one-click point on five different images. The next image that appears is based on the location of the previous click point. The best feature of the CCP is that the authentication failure message appears only after all the click points are clicked. This reduces attackers from guessing the password.

2.3 Persuasive Cued Click Points

Persuasive Cued Click Points comes under click-choice-based graphical password scheme. By adding the feature of Persuasive to CPP, we increase more security as well as promote the user to select a less predictable password, which is the main goal of the authentication system [2]. In PCCP, there is a viewport which is randomly formed on the image during signup. If the user is not satisfied, then he or she can choose to shuffle [2]. This reduces the risk of HOTSPOTS. Thus making it tougher for attackers to guess the password. Main disadvantage of this is that it may create a password which is difficult to remember.

3 Proposed System

The proposed system is an implementation of the Cued click points along with changes to remove the drawback of the PCCP authentication system thus making it far more secure. The following flowchart (Fig. 1) portrays the basic flow of the proposed system.

In this system, when the user is signing up for the first time, he or she is asked to select five images. Once the user has selected the images, he or she then needs to decide the order of the images in which the images will appear. Once the order of images is selected then the user is asked to select a region on each and every image. Clicking a point on the image may not always be accurate as the user may click little bit here and there while login. To avoid this drawback, the image is divided

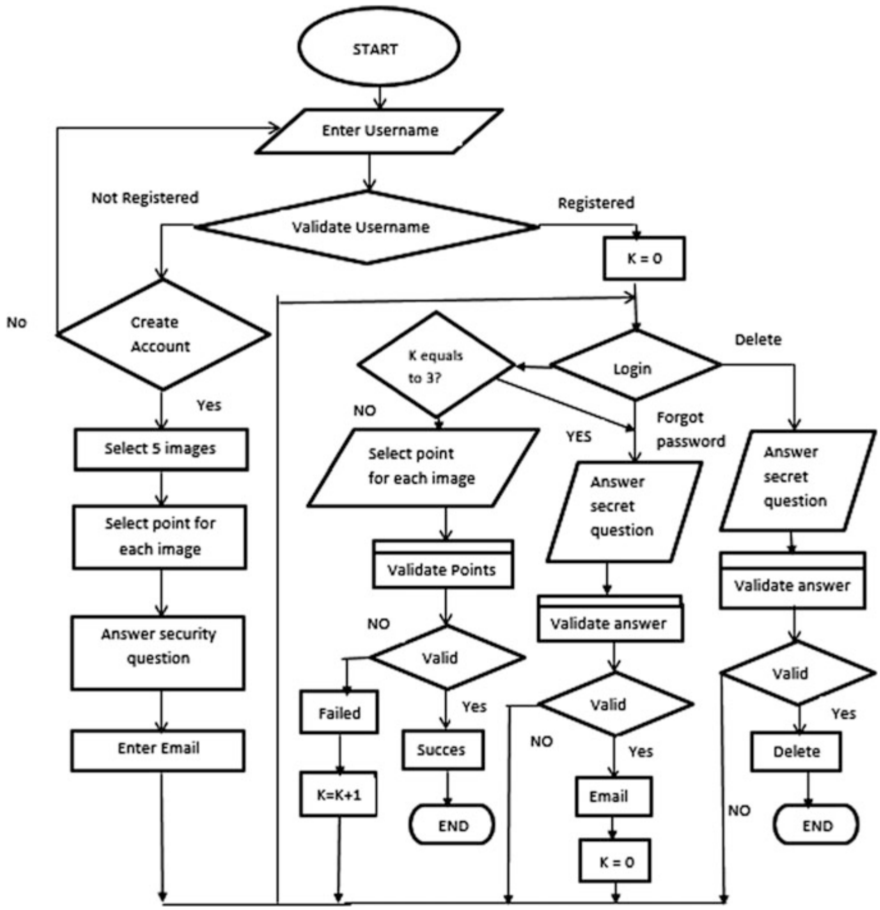


Fig. 1 Flowchart of proposed system

into 100 small regions. While signing up, the user will be able to see the boundaries of the grid. Now, the user can tap on any region to select it.

The regions along with the images will be stored in the database in an encrypted format. Then the user will be asked a few security questions which would be handy while resetting the password, when he or she cannot recollect the password. These questions along with the answers will be stored in the database.

The beauty of this authentication system is that it will only show successful login or unsuccessful login only after all the five click points are clicked. The system will not prompt or give an alert at any point in between if the user has selected a wrong region. So if a user or attacker pressed on a wrong region while login, the system will continue as if the user had pressed on the right region. It will just continue untill all five regions are clicked. The result of login will be displayed at the end. This would make it more difficult for the attacker to guess where he or she went wrong, thus increasing the security and reducing the chances of unauthorized password cracks.

4 Implementation

When the user opens the application, he/she would first see the screen shown in Fig. 2, where the user will see an empty text field, and three buttons. In the empty field the user needs to enter the username. The three available options are Login, Forgot Password and Delete Profile. To use any of the three functions entering the username is mandatory.

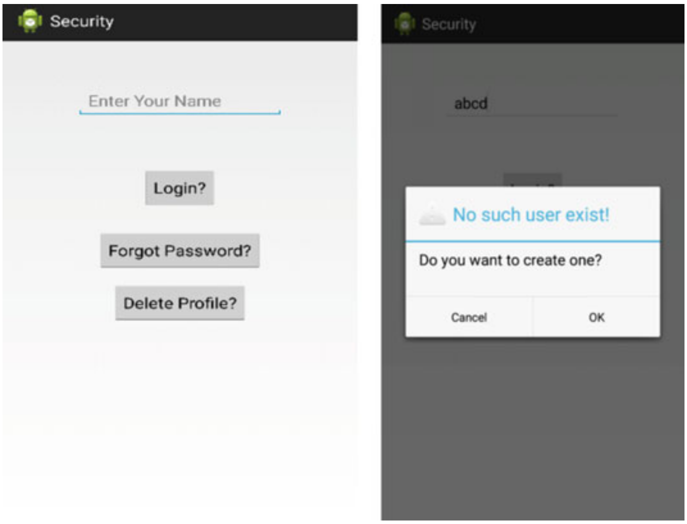


Fig. 2 Main screen

If the user does not have any account or wants to create a new account, then he/she needs to enter the new username in the text field, and then click on login. Every time the login button is clicked, the system checks if a user with provided username exists or not in the database stored in the root folder, which cannot be accessed by any user. If such user exists, then login activity is initiated, else the user will see a confirmation screen/alert box prompting the user to confirm account creation as seen in Fig. 2.

Once the user clicks on OK button, then he/she is redirected to another page to select the images as show in Fig. 3.

Now, the user needs to select five images. User can select the same, either from the gallery or can click a photo from the camera. These images are stored in a folder where all the application files are stored, having filename same as username followed by the index of the image in the selected sequence. Thus, all the user related images can quickly be accessed just by accessing image files having same name as the user name followed by image index. For example, all the images related to user xyz can be accessed by looking for image files with name xyz0, xyz1, xyz2 and so on. This folder is stored in the root folder, so the user cannot see or access it. This increases the security, as the user would not be able to manipulate any application file. Before storing the images in the folder they are compressed using the compression and scaling methods/functions of BitmapFactory class available in Android SDK. The compression is done so that the space consumed by the application, images, and application file is very low. This compression method reduces the size of the image by 90% of the original image size. Now the most

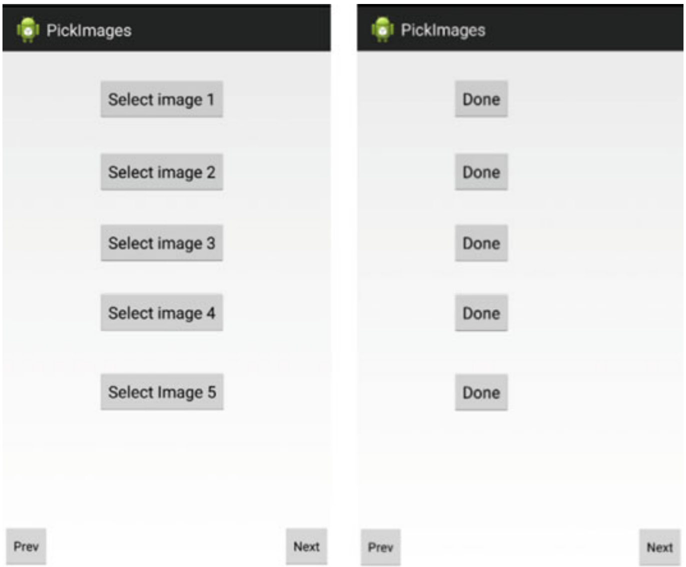


Fig. 3 Image selection

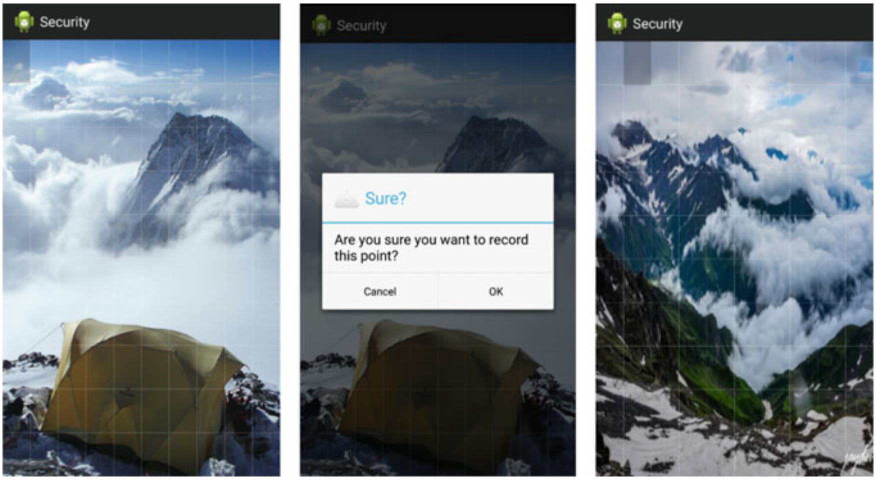


Fig. 4 Selection of sections

important task comes, which is selection of the point/grid for each and every image. This would be the password of the user. When the user clicks next, images will come up in the selected sequence, and a grid of size 10 × 10 will be visible on it as shown in Fig. 4. The grid is arranged such that every cell occupies the same area of the image irrespective of the screen size and device specifications thus maintaining integrity. This automatic adjustment is done by taking the length and height of the screen at the start, and then dividing it into 10 equal parts, resulting in dividing the screen in 100 equal grids. Each cell of the grid has a unique id. The user needs to select one of the sections/cells on each image. The image along with the grid would look similar to as seen below (Fig. 4).

When the user selects a section on the grid, it is highlighted for the user's reference. Once the user selects the point, he/she will be asked for confirmation of the point. Once the user presses OK, the section is recorded by storing its unique id. Similarly, the user will be asked to select regions on next four images. When the user selects all five points on each and every image, these points (ids associated with each section) are hashed using MD5 algorithm and stored in database in the application folder.

Then the user will be redirected to security question's page as shown in Fig. 5. Here the user needs to select three questions and answer them. These security questions will be handy when the user forgets the password, or exceeds three times continuous failed logins. The user also needs to provide an email id. An email will be sent to the user when he/she forgets the password.

Once the user filled the security question and the email id as shown in Fig. 6, this data is again hashed using MD5 encryption and stored in the database. Then the user is redirected again to the main page. Then to login, the user needs to enter the

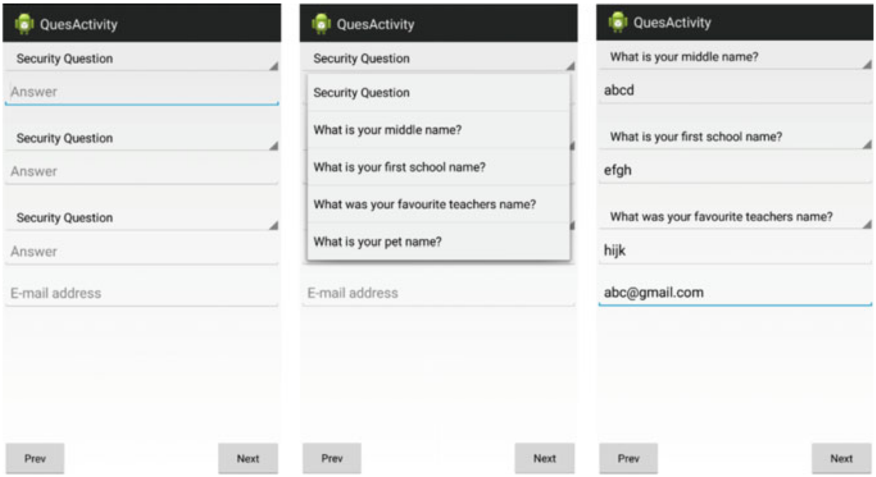


Fig. 5 Security questions

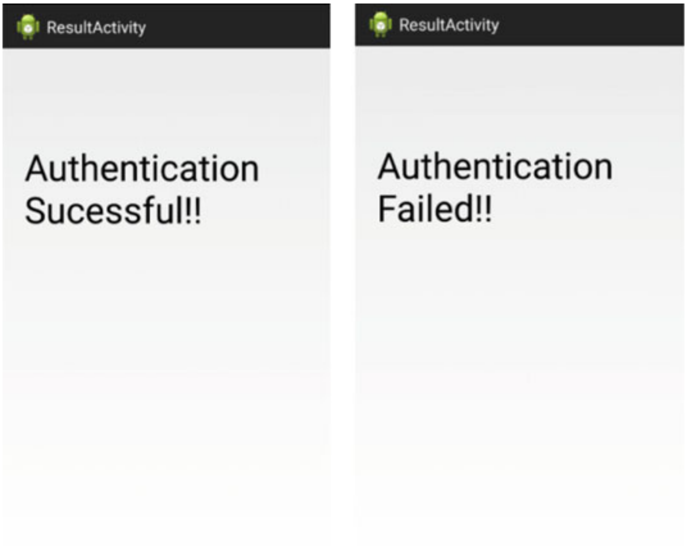


Fig. 6 Result screen

username and click on login. The same sequence of images will appear one by one, and then the user will have to click the registered points on each image.

If the user clicks on correct points then, he/she will be taken to Successful Login page. Else he/she would be taken to Failure page. Even if the user clicks on a wrong point, he/she won't be taken to the Failure page until all five points are selected by him/her.

The user will have three attempts to login, if he/she fails for three continuous attempts, then he/she needs to go through verification before any further attempts to login. After three continuous failed attempts, next time user tries to login, the user will be redirected to the verification page, where he/she needs to answer all the security questions. Only after answering all the questions correctly, he or she can attempt login.

If the user forgot the password, then user can select the forget password option on the main page as shown in Fig. 2.

When the user clicks on forgot password, user will be taken to verification page. The questions will be retrieved from the user's database and will be shown to the

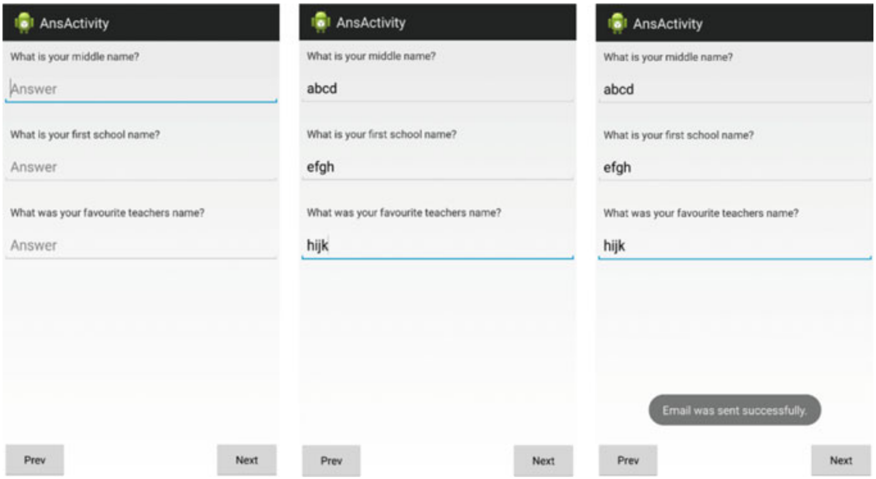


Fig. 7 Answering security questions

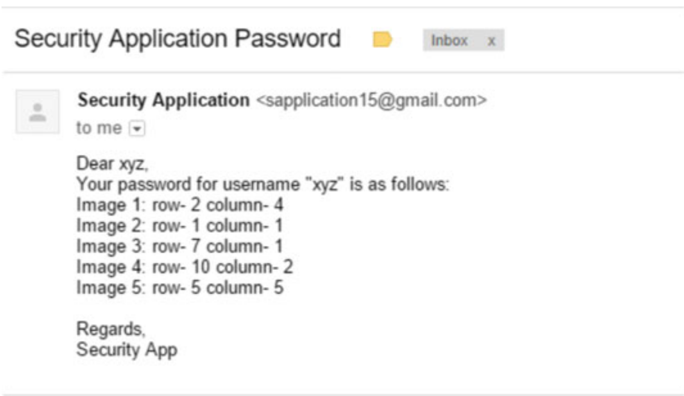


Fig. 8 Password recovery

Table 1 Comparison between text and graphical password

Basis	Text based passwords	Smart authentication
Remembrance	Simultaneous achievement of easy remembrance and difficult guess ability is inconvenient	Passwords are easy to remember yet difficult to guess
Guess ability	Usually, text based passwords contain strings that are related to user like birthdates, phone numbers, spouse name etc. These can easily be predicted and random guesses can be made	It is practically not possible for the intruder to guess the section selection of the user for all the images
Brute-force	If guesses fail, brute force can penetrate through this mechanism	Brute force is avoided by allowing user only consecutive attempts
Storage	Storage space efficient	Requires a little more storage as images are to be stored for each user profile
Time	Consumes Less time	Relatively more time is consumed
Security	Less Secure	More Secure

user. The user needs to answer all the security questions properly which can be seen in Fig. 7. If the user answered all the questions properly, then an email will be send to the registered email id given by the user at registration.

An email will be sent to the user giving the password revealing the click points of all the images in sequence as seen in Fig. 8.

5 Comparion

Comparison between texts based password and smart authentication system (Table 1).

6 Conclusions and Future Work

It is evident that even though smart authentication lacks in some aspects when compared with text based passwords, the benefits it provides clearly overpower its shortcomings. While smart authentication mechanism focuses on "easy remembrance", yet it is less vulnerable to malicious breach. In other words, it keeps authentication simple at the user's end and at the same time it poses high security barrier for potential intruders. Users do not have to go through the discomfort of strengthening their passwords by adding complexity to it, making it difficult to recollect at every login. Altogether, smart authentication sticks to basic notion of security by expecting the least from the user and challenging the intruders at the most.

The drawback of time consumption can be solved by including location based security, where the security in trusted area will be reduced, i.e. less time will be consumed and security in untrusted area will be high. For that, the user needs to add trusted and untrusted regions while registration. An additional feature can be also added, that if the users failed in three consecutive attempts then an image will be clicked using the front camera of the device without the user knowing it. It will also gather data like time, and the current location of the device using GPS, and mail it to the user's registered email id, that this person tried to access your device at this time and at this location.

References

1. Sayli Pradhan, Prachi Pandya, Rinita Nair, "Authentication using RFID and PCCP Application for Mobile Devices", International Journal of Advance Research in Engineering and Technology, Vol. 3, Issue 3, March 2015.
2. Sonia Caisson, Member, IEEE, Elizabeth Stobert, Student Member, IEEE, Alain Forget,Robert Biddle, Member, IEEE, and Paul C. van Oorschot, Member, IEEE "Persuasive Cued Click Points: Design, Implementation, and Evaluation of a Knowledge-Based Authentication Mechanism" IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 2, March/April 2012.
3. J. Yan, A. Blackwell, R. Anderson, and A. Grant, "The Memorability and Security of Passwords," Security and Usability: Designing Secure Systems That People Can Use, L. Cranor and S. Garfinkel, eds., ch. 7, pp. 129–142, O'Reilly Media, 2005.
4. Nelson, D.L., U.S. Reed, and J.R. Walling. Picture Superiority Effect. Journal of Experimental Psychology: Human Learning and Memory 3, 485–497, 1977.
5. S. Wiedenbeck, J. Waters, J.C. Birget, A. Brodskiy, and N. Memon. "PassPoints: Design and longitudinal evaluation of a graphical password system". International Journal of Human Computer Studies, 2005.
6. S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, "Authentication Using Graphical Passwords: Effects of Toler-ance and Image Choice," Proc. First Symp. Usable Privacy and Se-curity (SOUPS), July 2005.
7. K. Golofit, "Click Passwords under Investigation," Proc. 12th European Symp. Research in Computer Security (ESORICS), Sept. 2007.
8. Ms. Uma D.Yadav and Mr. P. S. Mohod,"Enhancement of Knowledge Based Authentication Mechanism using Graphical Password via Persuasion" Journal of Computer Science and Engineering, Volume 17, Issue 2, February 2013.

New Security Architecture for Big Data Hadoop

Madhvaraj M. Shetty and D.H. Manjaiah

Abstract Hadoop is a system that provides distributed job processing on large clusters using commodity servers and distributed file system that allows storing and managing petabytes of data which is collected from disparate data sources. Although Hadoop is used on private clusters behind firewalls and physical security, Hadoop is often runs in public cloud for providing shared multi-tenant service and used to store sensitive data; as a result strong authorization and authentication is required. Hadoop has two vulnerabilities that can dramatically impact its compromise resilience. That is, overloaded authentication key and the lack of fine-grained access control at the data access level. Providing security to Hadoop is more challenging due to all the interactions do not follow the classic client–server pattern: the file system is partitioned and distributed over the network, requiring authorization checks at multiple points; the system scales to thousands of servers and tens of thousands of concurrent tasks; submitted batch job is executed at a later time on nodes different from the node on which the client authenticated and submitted the job and jobs from different users are executed on the same node. To address these challenges, the base Kerberos authentication protocol is used which is supplemented by delegation, access token, and job tokens. This paper examines current security mechanisms of Hadoop clusters using Kerberos supplemented by other tokens, and then we analyzed the security problems and risks of Hadoop. Further, we proposed a new authentication protocol inside Hadoop and usage of modified tokens for improving security.

Keywords Hadoop · Security · Kerberos protocol · Delegation token
Block access token · Job token · Name node · Data node · New architecture

M.M. Shetty (✉) · D.H. Manjaiah (✉)
Department of Computer Science, Mangalore University, Mangalore, India
e-mail: madhvarajj@gmail.com

D.H. Manjaiah
e-mail: manju@mangaloreuniversity.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_41

1 Introduction

Hadoop [1, 2] is an open source, java-based framework created by Doug Cutting, which supports distributed computing on large data sets. It was originally developed on the basis of Google's MapReduce and Google File System. MapReduce is a programming model for data processing and provides the most basic mechanism for batch processing. The file System Hadoop Distributed File System (HDFS) is a scalable, distributed, and portable file system written in Java for Hadoop framework and most widely used as cloud storage [3]. It is the storage layer for Hadoop and provides the ability to store large amounts of data. The main goals of Hadoop MapReduce are to process large size data sets, cope with hardware failure and high throughput. When it was first designed, it was never intended to be used in the ways that it is used now. It was a straightforward tool for an organization to run MapReduce jobs on large amounts of data. Initially it was designed without security in mind and had no security framework. Most of the part covered in the logic on how to deal with the complexities inherent in distributed systems, such as handling of failures and coordination. It considers that the entire environment and cluster were trusted. Due to this focus, the early Hadoop project established a security stance that the entire cluster of machines and all of the users accessing it are part of a trusted network. Even though it had some authorization controls like file access permissions, a malicious user can easily impersonate a trusted user as the authentication was on the basis of password. This gives malicious user to read or modify the data first in the other's cluster and second, suppress or kill the other job to execute his job earlier than the other to complete job [4, 5]. So, there is a need for a strong authorization mechanism to protect sensitive data and also there is a need for a highly secure authentication system to restrict the access to confidential business data that are processed and stored in Hadoop framework [4].

A non-secure Hadoop configuration relies on client-side libraries to send its credentials as determined from the client-side operating system as part of the authentication protocol, this method is sufficient for many deployments that rely on physical security. Authorization checks through Access control List (ACL) and file permissions are performed against the client-supplied user ID. A user can directly communicate with a Data Node (DN) once the block location is known, this enables the unauthorized clients to impersonate authorized users and access the cluster. So, initially password was the only authentication mechanism used to verify the validity of users. As the project evolved, it became apparent that minimum there should be a mechanism for users to strongly authenticate to prove their identities at least. A well-established protocol Kerberos [6] was introduced to Hadoop. This paper examines the current security mechanisms of Hadoop clusters using Kerberos supplemented by other tokens, and then we analyzed the security problems and risks of Hadoop. Further, we proposed a new authentication protocol inside Hadoop and usage of modified tokens for improving security.

The rest of the paper is organized as follows: Section 2 presents existing scenario of security in Hadoop clusters, Sect. 3 deals with some challenges related to

Hadoop security and Sect. 4 describes methods proposed to make Hadoop cluster more secure, and Sect. 5 discusses about its security analysis and Sect. 6 concludes the paper.

2 Hadoop Security Framework

This section discusses the basic security protocols which are implemented in the current Hadoop framework. There are three primary areas which have been implemented in terms of the security framework, the Kerberos authentication mechanism for authenticating users to Hadoop, the token-based protocol for data access inside Hadoop framework and other miscellaneous protocol like RPC, data transfer protocol [6, 7]. The primary challenge in implementing security is to prevent impersonation, where a user can misbehave and emulate another user to have access on other's data in the clusters and another challenge is to deal with the trade-offs between Hadoop cluster scalability and prevention of performance degradation.

Kerberos [8] is a network authentication protocol developed at MIT as part of Project Athena. It uses private-key cryptography for providing authentication across open networks. It is a third-party authentication mechanism, in which users and services rely on a third-party Kerberos server—to authenticate each to the other. As shown in Fig. 1, there is Kerberos authentication module which has authentication server and the ticket granting server. After Kerberos is configured, Kerberos authentication is used to validate the client-side credentials. The authentication flow of Kerberos and Hadoop are shown in Fig. 1. Initially user obtains a Ticket

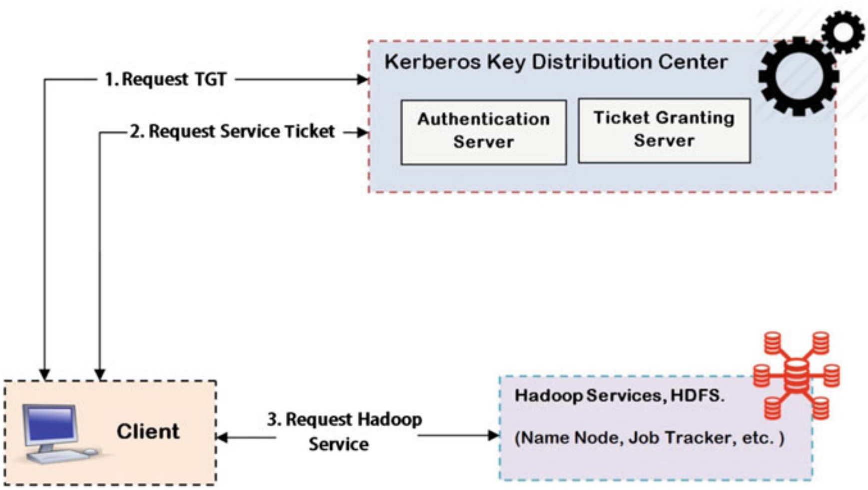


Fig. 1 Kerberos and Hadoop authentication flow

Granting Ticket (TGT) from Kerberos Key Distribution Center (KDC). After the user has TGT, the client application accesses the Hadoop Services which initiates a request for the Service Ticket (ST) that corresponds to the Hadoop Service the user is accessing. The ST is then sent as part of the connection to the Hadoop Service. The corresponding Hadoop Service must then authenticate the user by decrypting the ST using the Service Key exchanged with the KDC. If this decryption is successful the end user is authenticated to the Hadoop Service. But here performance is one important consideration as cluster grows the volume of interaction between authentication server (AS) and ticket granting server (TGT) will also grow.

Once clients and services have been authenticated to each other via Kerberos, it is important to understand that many more security controls are required beyond this initial authentication. Because developers write applications, and applications access HDFS and execute MapReduce jobs over data stored in HDFS on behalf of the users executing them. AS all services must perform tasks on behalf of the client, more security controls are needed to ensure that all of these services doing the work on behalf of the client are authorized to perform work on behalf of the client.

So token-based authentication is used for subsequent authentication. Different security tokens supplement the primary Kerberos authentication in Hadoop. Currently, there are three forms of tokens [6, 7]:

- Delegation Token—It identifies a user to a particular HDFS service.
- Block Access Token—It gives a read or write permission to a particular HDFS block.
- Job Token—It identifies a MapReduce task for its jobs.

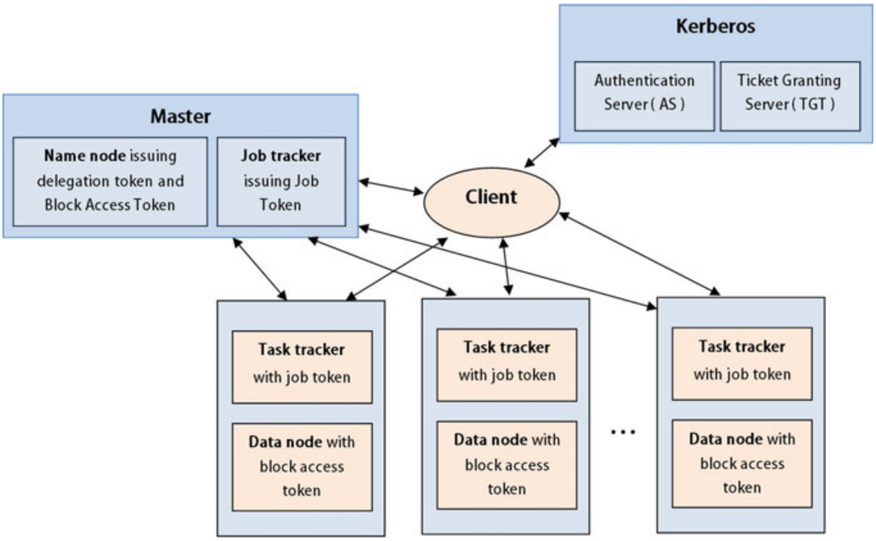


Fig. 2 Hadoop security framework

The overall architecture of security framework is shown in Fig. 2 [7]. The client will load Kerberos tickets that are in the user's ticket cache. MapReduce will also create a token cache that will be loaded by the task. When the application creates an RPC connection, it will use a token, if an appropriate one is available. Otherwise, it will use the Kerberos credentials.

2.1 Delegation Token

After initial authentication to NameNode using Kerberos credentials, a user obtains a delegation token, which can be used for subsequent authentication to NameNode as the user [7]. The token is in fact a secret key shared between the user and NameNode and should be protected when passed over insecure channels. Anyone who gets it can impersonate the user on NameNode. Note that a user can only obtain new delegation tokens by authenticating using Kerberos. When a client obtains a delegation token from NameNode, it specifies JobTracker as the token renewer. All the tasks of the same job use the same token. JobTracker is responsible for keeping the token valid till the job is finished. After that, JobTracker may optionally cancel the token.

Format of delegation token is,

$$\begin{aligned}&\text{TokenID} = \{\text{ownerID, renewerID, issueDate, maxDate, sequenceNumber}\}\\&\text{TokenAuthenticator} = \text{HMAC-SHA1}(\text{masterKey, TokenID})\\&\text{Delegation Token} = \{\text{TokenID, TokenAuthenticator}\}\end{aligned}$$

NameNode uses masterKey to generate and verify delegation tokens. If currentTime > expiryDate, the token is considered expired and any client authentication request using the token will be rejected. The sequenceNumber is a global counter in the NameNode that is incremented as each delegation token is created to ensure that each delegation token is unique. Using delegation token, client (e.g., a Task) uses a TokenID to authenticate itself with NN.

2.2 Block Access Token

Originally in Hadoop, DataNodes did not enforce any access control on accesses to its data blocks and all the file permissions are stored on NameNode. This is made possible for an unauthorized client to read a data block as long as he can supply its block ID. For this purpose, Block Access Token (BAT) is used here which authenticates client to HDFS blocks and enables its holder to access certain HDFS data blocks [7]. It is issued by NameNode and used on DataNode. BAT is generated in such a way that their authenticity can be verified by DataNode. When the client accesses the NameNode for files, authorization is checked on the NameNode and block access tokens are generated based on file permissions. These tokens are returned to the client along with locations of their respective blocks. When the client accesses DataNodes for those data blocks, block access tokens are passed to DataNodes for verification. Block access tokens are generated using a symmetric key scheme where the NameNode and all of the DataNodes share a secret key. When DataNode receives a token, it uses its own copy of the secret key to recompute the token authenticator and compares it with the one included in the token. If they match, the token is verified as authentic.

A BAT has the following format, where keyID identifies the secret key used to generate the token, and accessModes indicates mode of operation (READ, WRITE, COPY, REPLACE)

TokenID = {expirationDate, keyID, ownerID, blockID, accessModes}
TokenAuthenticator = HMAC-SHA1(key, TokenID)
Block Access Token = {TokenID, TokenAuthenticator}

A BAT is valid on all DataNodes regardless where the data block is actually stored. The secret key used to compute token authenticator is randomly chosen by NameNode and sent to DataNodes when they first register with NameNode and at regular intervals on their next heartbeats. The BAT generated by NameNode is not renewable and needs to be fetched again once the token expires.

2.3 Job Token

When the job is submitted, the JobTracker will create a secret key that is only used by the tasks of the job when identifying themselves to the framework [7]. This token will be stored as part of the job in the JobTracker's system directory on HDFS and distributed to the TaskTrackers via RPC. This token will be used for the RPC via DIGEST-MD5 when the task communicates with the TaskTracker to requests tasks or report status. Additionally, this token can be used by Pipes tasks, which run as sub- processes of the MapReduce tasks.

3 Limitations

However Hadoop was originally designed to run in a well controlled private environment. When moving Hadoop to a public cloud, there are challenges related to its security mechanisms. There is a rising concern that Hadoop in its present form may not be able to maintain the same security level in a public cloud as it does in a protected environment. Internal cloud attacks can bypass security mechanisms to compromise the safety of data and computing in Hadoop. Currently Hadoop is leading with a third-party authentication protocol called Kerberos. But one of the major limitations of Kerberos is that it is not effective against password guessing attacks. Further, Kerberos requires a trusted path to handle passwords and it does not support multipart authentication [9].

We evaluated security mechanisms in current Hadoop as follows [10–12]:

- Authentication of users is not done by Hadoop services; Hadoop itself does not issue user credentials or create accounts for users.
- There is no enforcement of access control by Data Nodes on accesses to its data blocks.
- HDFS and MapReduce communication will not travel on untrusted networks.
- Name Node and all Data Nodes share the same key to generate block access token. If the key is leaked from machines of HDFS by any attack methods, an attacker has the capability to use the key to generate arbitrary Block Tokens as he or she wants to, and accesses any data blocks in HDFS.
- Delegation Token used by map and reduce processes lacks fine-grained access control. With a Delegation Token, a process has the privilege to behave as the Hadoop user and to access all content which the Hadoop user is allowed to access. So, it can enable the attacker to impersonate the Hadoop user and access the Delegation Token's owner's data. And same delegation Token is used by all computing processes (i.e., map processes and reduce processes). So a leaked Delegation Token can be used steal a large amount of data from HDFS.

Some of the possible solutions for enhancement of security are [13]:

- Access control at the file system level.
- Access control checks at the beginning of read and write.
- Secure way of user authentication inside Hadoop.
- Security using role-based access control.

4 Proposed Work

4.1 *Modified Tokens*

We proposed new Hadoop security architecture based on providing identity to each data nodes and processes in the cluster. This model tends to enhance the security in Hadoop and provides authentication service to various components by using modified tokens. The proposed approach can prevent compromised Hadoop processes from compromising the rest part of Hadoop and provides better isolation level among Hadoop components.

4.1.1 Block Access Token

Name Node generates Block Access Tokens to authorize client/process to HDFS blocks. It uses same secret key to generate all the block tokens irrespective of number of DataNodes present. If this key is leaked, then an attacker has the capability to generate any arbitrary Block Access Token and to access any data blocks in HDFS. So here we use different keys for each data nodes in cluster to generate block tokens. These keys generated using the hardware information of that node (for example, MAC, IP, Unique ID for each data node). While cluster is booting up and ready to start, both NN and DN exchange this information. DN sends its MAC id to NN and NN computes unique id for each node based on the MAC, node number and location of the node in the cluster, and assigns to each DN. This occurs only once and this process done in secure environment. This id is regenerated only when there is an error, exception or an unauthorized request in the DN; it is done by DN by sending MAC with its previous key. After this process NN sends timestamp to all the DN with command to generate device key. DN generates the same by the below format and send acknowledgement back to NN.

$$\text{Device key} = \{\text{MAC, Unique ID, Timestamp}\}$$

When client requesting to access particular block, it uses block id in request. NN has the information about in which data node that block present and uses that data nodes details to generate Device key. This Key is used by NN to create block access token. Only that particular DN has this Device key, and gets block token, uses this token to authorize HDFS block. Further all tokens encrypted while it is transferred over network. Since device key cannot be same for long time, it should be changed at regular interval of time due to security reason. So NN sends command to regenerate Device key with timestamp, DN generates and sends acknowledgement back to NN. Also NN issues the same command for regeneration of device key in case of any exception found by the NN.

4.1.2 Delegation Token

Hadoop (NN) produces single delegation token to each client for all services, it is used by different process to authenticate themselves to NN. After initial authentication to NN, a client obtains a delegation token which is given to a job for subsequent authentication to NN. All process related to that job shares a single delegation token, but which leads to lack of fine-grained access control vulnerability, so it is good to have access control information for each process and restrict the content that process can access using this token with respect to security. This is achieved by providing a unique id to each main process when it first authenticates with NN, so after initial authentication, produced delegation token can be used only by that process & its sub processes. Also user id is mapped to process id; it is useful while applying ACL to particular process/user. So we proposed a 2-step delegation token, in first-step delegation token generated normally and in second step we add individual process id, so that NN can provide fine-grained access control based on the user/process permissions defined on ACL in NN.

Format of the process id which is generated by NN,

$$\text{Process ID} = \{\text{user id, clients IP, timestamp, sequence number by NN}\}$$

So delegation token generated using Token ID and Process ID with master key. Eventually job creates number of sub processes and generates sub process id using process id,

$$\text{Sub process id} = \{\text{Parent process id, user id, clients IP, timestamp, sequence number by Job client}\}$$

This id used with token in subsequent authentication by subprocesses and NN verifies it by extracting parent process id to get info about to which user/process it belongs and what access is permitted. By defining access control lists in NN, we can easily handle the tasks which are using other services of Hadoop.

4.2 Authentication Protocol

In this approach, we redesigned Hadoop security framework to improve security in the cluster. Our proposed work will work without the third-party protocol Kerberos, and but utilizes the fundamental concept of it. Protocol is based on dual node authentication where two nodes are used to authenticate users to Hadoop as shown in Fig. 3.

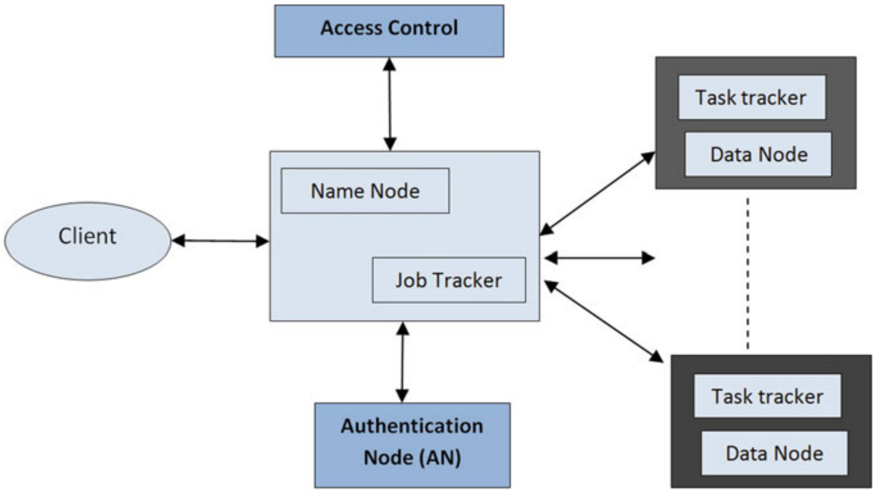


Fig. 3 New authentication protocol for Hadoop security framework

The proposed architecture uses both Name Node (NN) and Authentication Node (AN) for user authentication. Since all the sensitive authentication data is stored in AN, this node will be installed in secure zone, where there is higher security is provided to that node. This architecture is different from the multi-node authentication where all the nodes are equally exposed to client. In the proposed approach, only one node can communicate with the client and the other is hidden from the user as shown in Fig. 3 which can improve the overall security and eliminates single point of vulnerability. This dual node authentication protocol authenticates user to the Hadoop if and only if both the servers are mutually involved in the authentication mechanism. Also our proposed scheme of authentication divides users into different categories based on their roles in Hadoop clusters and offers effective security against attacks like replay attack, guessing attack, and stolen-verifier attack.

New authentication protocol scenario In this framework NN has information about the users already registered. Initially client sends username to NN, it checks whether user is registered or not, if the user is not registered, then the registration process can be taken at this time. If the user is registered then it sends this username with its unique id generated for that request/user based on timestamp, sequence number, username to AN, it will generate a token based on the information such as username, IP, timestamp, expiry date, sequence number, session keys and that token is dual encrypted by the NN unique id and user's password and sends it to client via NN. Client gets valid token to authenticate NN by decrypting token using his own password. Each time user enters password, client sends signal to NN like TRUE/FALSE which indicates success or failure of decryption. If there is three FALSE signal received by NN, means user entered wrong password three times and NN sends command to AN to cancel the token. So it will be invalid, even when client enters correct password, next user need to get new token by entering username again. By using this we can avoid password guessing attack, there by blocking user if he tries to enter wrong password more than times that we defined in configuration.

After successful decryption using his password, the obtained token is still encrypted by NN unique id is sent to NN, it decrypts with its id and verifies client by comparing token with AN, if it matched then user will be authenticated. After successful authentication, NN is responsible to check access control list defined based on their roles in Access control Node. By this model, we can define and manage the entire access control list centrally based on the user's different roles in the hadoop environment.

5 Security Analysis of Proposed Work

Currently Hadoop uses many kinds of secret keys and tokens to perform authentication and authorization among different Hadoop components. These sensitive keys and tokens may be used to attack Hadoop once leaked. In our proposed approach both the modified token provides more security avoiding Hadoop components to be compromised. In block token generation, each Data Node is forced to share a unique device key with the Name Node, so it can produce different token for different blocks in HDFS, so only that block will be compromised even if it is leaked and attackers cannot access resources beyond the compromised block. And each process/job in Hadoop can have different access privileges in our proposed model via modified delegation token. It enforces access privilege on Hadoop processes when these processes access different services. A compromised job has impact on only to the service to which it is privileged. By the using proposed authentication protocol, we can avoid attacks such as password guessing attack, replay attack, and stolen-verifier attack. Since there is no direct communication between any client to AN, there is no possibility of hacking AN. And even if NN is compromised, it will not be exposing any user credentials. Also it provides centrally manageable access control list to handle different roles of users effectively in the hadoop environment.

6 Conclusion

This paper proposed two new approaches to improve security of Hadoop using new authentication protocol inside Hadoop and by using modified tokens. The proposed authentication protocol will enhance Hadoop security as the authentication mechanism utilizes two nodes for authentication. Kerberos is effective for user authentication. However, it is extremely difficult to deploy in Hadoop cluster. Also by enabling Kerberos, it degrades Hadoop performance and makes Hadoop run much slower than without Kerberos. We believe that our proposed new security model that is integrated in Hadoop itself can make its deployment easier with less performance impact. There are many areas to improve in the various security aspects of Hadoop clusters and new technologies need to be proposed to enhance and improve the security in terms of reliability and flexibility.

References

1. Tom White, "Hadoop- Definitive guide", O'Reilly,2009.
2. Apache™ Hadoop®! Available: http://Hadoop.apache.org/.
3. H. Zhou and Q. Wen, "Data Security Accessing for HDFS Based on Attribute-Group in Cloud Computing," in International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014), 2014.
4. N. Somu, A. Gangaa, and V. S. Sriram, "Authentication Service in Hadoop Using one Time Pad," Indian Journal of Science and Technology, vol. 7, pp. 56–62, 2014.
5. E. B. Fernandez, "Security in data intensive computing systems," in Handbook of Data Intensive Computing, ed: Springer, 2011, pp. 447–466.

6. Das, Devaraj, Owen O'Malley, Sanjay Radia, and Kan Zhang. "Adding Security to Apache Hadoop." Hortonworks, IBM (2011).
7. O'Malley, Owen, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell. "Hadoop security design." Yahoo, Inc., Tech. Rep (2009).
8. Steiner, Jennifer G., B. Clifford Neuman, and Jeffrey I. Schiller. "Kerberos: An Authentication Service for Open Network Systems." In USENIX Winter, pp. 191–202. 1988.
9. Sadasivam, G. Sudha, K. Anitha Kumari, and S. Rubika. "A novel authentication service for Hadoop in cloud environment." In Cloud Computing in Emerging Markets (CCEM), 2012 IEEE International Conference on, pp. 1–6. IEEE, 2012.
10. Wang, Peipei, and Anwesha Das. "Insights to Hadoop Security Threats."
11. Xianqing, Yu, Peng Ning, and Mladen A. Vouk. "Enhancing security of Hadoop in a public cloud." In Information and Communication Systems (ICICS), 2015 6th International Conference on, pp. 38–43. IEEE, 2015.
12. Gaikwad, Rajesh Laxman, Dhananjay M. Dakhane, and Ravindra L. Pardhi. "Network Security Enhancement in Hadoop Clusters." Int. J. Appl. Innov. Eng. Manage. (IJAIEM) 2, no. 3 (2013).
13. Jam, Masoumeh Rezaei, Leili Mohammad Khanli, Morteza Sargolzaei Javan, and Mohammad Kazem Akbari. "A survey on security of Hadoop." In Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on, pp. 716–721. IEEE, 2014.

An Overview of Bio-Inspired Computing

K. Lakshmaiah, S. Murali Krishna and B. Eswara Reddy

Abstract Computing over the years has evolved from being simple a mathematical processing machine to more sophisticated problem-solving entity pushing limits around reasoning and intelligence. Along the way, lots of scientists and engineers have closely observed some of the biological processes achieving certain things in a more efficient and simple fashion than traditional computational mechanisms. This has led to development of various techniques and algorithms which try and mimic these biological processes and are categorized under, bio-inspired computing. In this paper, an overview of bio-inspired Computing is presented.

Keywords Bio-inspired computing · Neural networks · Artificial immune systems Swarm intelligence

1 Introduction

Even though there are continuous, relentless, and breathtaking advances in computer technology, we continue to be humbled and fascinated by the variety, adaptability and sophistication of the natural world around us. From the beginning computing was inspired by nature. Alan Turing asked if computers could be created which could think like humans, which has now become a reality. Bio-inspired Computing, short for biologically inspired computing, is a field of study that loosely knits together subfields related to the topics of connectionism, social

K. Lakshmaiah (✉)
JNTUH, Hyderabad, India
e-mail: Klakshmaiah78@gmail.com

S. Murali Krishna
CSE & ICT, S V College of Engineering, Karakambadi Road, Tirupati, India
e-mail: muralikrishna.s@svcolleges.edu.in

B. Eswara Reddy
CSE, JNTUA College of Engineering, Kalikiri, Chittoor 517234, India
e-mail: eswarcsejntua@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_42

behavior and emergence. It relies heavily on the fields of biology, computer science and mathematics [1].

Computers over the years has evolved from being simplex mathematical processing machine to more sophisticated problem-solving entity, pushing limits around reasoning and intelligence. Along the way, lots of scientists and engineers have closely observed some of the biological processes achieving certain things in a more efficient and simple fashion than traditional computational mechanisms. This has led to development of various techniques and algorithms which try to mimic these biological processes and are categorized under, Bio-Inspired Computing [2].

In fact the genius of bio-inspired computing is the fact that they achieve certain things in a more efficient and simple fashion than many of the already existing “traditional” computing mechanisms. This has led researchers and engineers to a rapid exploration of bio-inspired engineering. This includes development of various techniques and algorithms that mimic various biological processes.

In the process of exploring Bio-Inspired engineering, the various drawbacks of classical computing has become highlighted with even greater significance. The various areas in which the classical paradigm is failing to make a mark in (even with lots of lots of research) are Pattern Recognition, Robustness, Dealing with incomplete information, Adapting and learning based on Experience [2, 3].

This is where the need of the Bio-inspired engineering really rears its head and makes its claim. By mimicking process found in life we might be able to solve some or all of these problems, problems for which solutions cannot be found from the classical computing and thinking.

How does one arrive at a bio-inspired solution? It involves a few steps that will be explained in detail later. Before we go there let us understand the thinking behind the use of the term bio- inspired.

The term bio-inspired represents a strong relationship between a particular computing system or algorithm and a biological system which will be used to solve a specified problem. The computing system or algorithm is built in such a way that it follows a similar procedure or has similar capabilities to the biological system. The categories of bio-Inspired Computing include—Computing, systems and networking.

Bio-Inspired computing represents a class of algorithms focusing on efficient computing, for example optimization processes and pattern recognition. Bio-inspired systems rely on system architectures for massively distributed and collaborative systems, for example distributed sensing and exploration. Bio-inspired networking is a class of strategies for efficient and scalable networking under uncertain conditions, for example delay tolerant networking [4, 5].

Now we come to the design of bio-inspired solutions. What are the steps required to arrive at such a solution?

There are three steps (As shown in Fig. 1):

- Identification of analogies.
- Understanding.
- Model simplification.

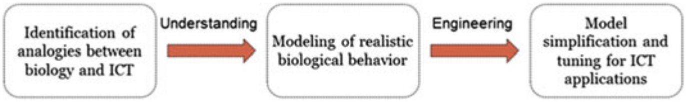


Fig. 1 Steps in designing bio-inspired solutions

Even though there is a huge overlap between Artificial Intelligence (AI) and Bio-Inspired Computing (BIC) there are certain differences between them. AI is the building of intelligent machines which act rationally and decision making is done using deductive logic. They are primarily used to improve our understanding of human intelligence. BIC is building of life-like systems which are used to improve our understanding of biology [6].

bio-Inspired Computing can be divided into two broad categories, namely Swarm Intelligence and Natural Computing.

2 Swarm Intelligence

A living organism is a complex system which works in sync with other such living organisms to produce desirable output with specific qualities like fault tolerance, distributed computing, evolution and growth.

These qualities are difficult to achieve with the classical approach. Swarm intelligence on the other hand achieves the above the goals with the desired qualities. A bio-inspired solution that mimics the collective behavior and collective processes of several living organisms to solve problems is the essence of Swarm Intelligence.

This collective behavior and intelligence exhibited by animals of similar size, similar tribe, and similar characteristics is known as swarming. For Example, milling about the same spot or migrating in some direction. This behavior is mainly seen in small insects and in the bird family. Plenty of swarming examples can be found in nature. Some examples of swarming are,

- Termites swarm to build colonies
- Birds swarm to find food
- Bees swarm to reproduce.

So swarming is done in nature. But what goal do these organisms achieve, why do they swarm? The reasons for the above organism to swarm are,

- Finding food is best achieved when done collectively—better forage
- Migration is a collective, social activity.
- Increases the strength of defense and can aid in guarding against predators.

These reasons are important because when systems are bio-inspired by swarming the goals that these systems achieve or the characteristics that they require will be inherently similar to these outcomes. Or the problems that these systems are trying to solve are usually related to the problems that swarming solves. Swarming is a powerful mechanism which can yield results which is out of reach of an individual entity if it were performing the same activity in isolation [7].

Now let's take one example in more detail and examine the swarming activity demonstrated by a flock of birds. In this case three simple rules are followed:

- Collision Avoidance
  - Rule 1: Avoid Collision with neighboring birds
- Velocity Matching
  - Rule 2: Match the velocity of neighboring birds
- Flock Centering
  - Rule 3: Stay near neighboring birds

The rules that the birds follow in a flock are very similar to the qualities desired by an algorithm or a system. The work done by every bird is similar to the computation required by a task executed by a node in a system of various nodes. Because of the simplistic design principles involved, the swarm characteristics are very appealing and can be used to simplify complex computing systems performing similar tasks.

A solution based on swarming offers the following advantages or characteristics.

Simple rules for each individual: Individual participant node performs simple compute which is necessary to meet the requirements to be in swarm to achieve the end goal. This way the compute capabilities required are minimal or just enough to follow rules of the group activity

- *No central control:* Lot of the distributed systems built in this computing era have certain drawbacks, one of them is the master slave architecture and single point of failure. Because there is a master node controlling most of the distributed activity, if the master were to fail the entire group would fail. The fault tolerance mechanism built involves use of another distributed system altogether and increases the complexity of the system. Emerging distributed architectures have the shared nothing concept which can be thought of as an inspiration from these swarm solutions. Swarm solutions do not have central authority which increases the robustness of the entire solution by reducing impact of failure.
- *Emergent:* Emergent structures are a common strategy found in many animal groups: colonies of ants, mounds built by termites, swarms of bees, shoals/schools of fish, flocks of birds, and herds/packs of mammals. An example to consider in detail is an ant colony. The queen does not give direct orders and does not tell the ants what to do. Instead, each ant reacts to stimuli in the form of chemical scent from larvae, other ants, intruders, food and build up of waste,

and leaves behind a chemical trail, which, in turn, provides a stimulus to other ants. Here each ant is an autonomous unit that reacts depending only on its local environment and the genetically encoded rules for its variety of ant. Despite the lack of centralized decision making, ant colonies exhibit complex behavior and have even been able to demonstrate the ability to solve geometric problems. For example, colonies routinely find the maximum distance from all colony entrances to dispose of dead bodies.

Finally the definition of Swarm intelligence is: "any attempt to design algorithms or distributed problem-solving devices inspired by the collective behavior of social insect colonies and other animal societies" [Bonabeau, Dorigo, Theraulaz: Swarm Intelligence].

Thus Swarm Intelligence can be used in various systems to solve different problems across different fields. Some of the places where Swarm Intelligence has already been used or is in use are,

- The Lord of the Rings film trilogy made use of similar technology, known as Massive, during battle scenes
- In telecommunication networks in the form of ant based routing
- Swarm robotics is a new approach to the coordination of multirobot systems which consist of large numbers of mostly simple physical robots. These employ a majority of swarm intelligence solutions

Particle Swarm Optimization (PSO) is a computing method that imitates the social behavior of humans or insects. It concentrates on optimizing the problem by iteratively improving the candidate solution with regard to a specific quality. Humans interact with each other continuously looking for other humans who they can bond with. As they meet more new people their experience increases and they are able to correctly choose whom to become friends with. It's an iterative process.

A somewhat similar process occurs in flocks of birds. Over a number of iterations, a group of variables have their values adjusted closer to the member whose value is closest to the target at any given moment. Imagine a flock of birds circling over an area where they can smell a hidden source of food. The one who is closest to the food chirps the loudest and the other birds swing around in his direction. If any of the other circling birds comes closer to the target than the first, it chirps louder and the others veer over toward him. This tightening pattern continues until one of the birds happens upon the food. This algorithm is simple and easy to implement (Fig. 2). The algorithm tracks three global characteristics of variables,

- Each particle's or agent's(a bird in this case)
- Global best(gBest)
- Stopping Value

Each particle is represented by two values. The velocity and personal best (pBest). Each particle (or agent) evaluates the function to maximize at each point it visits in spaces. This is called the velocity. It indicates how much the data has changed. pBest indicates the closest the particle has ever come to the target. Both

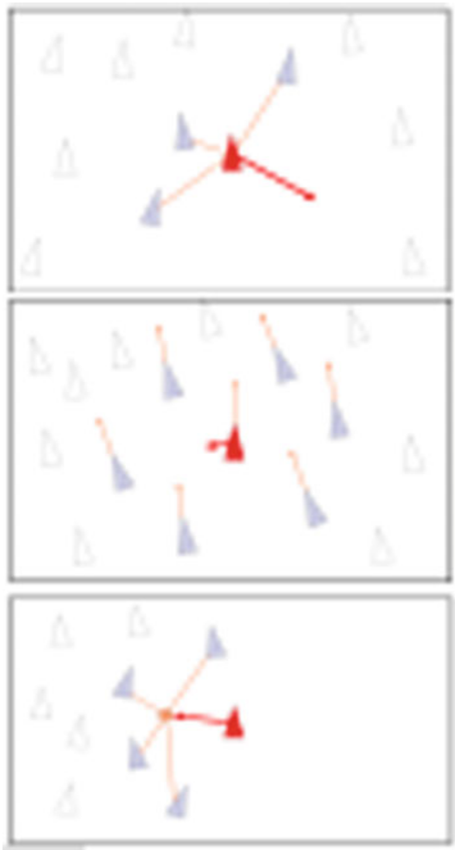


Fig. 2 Global characteristics of variables

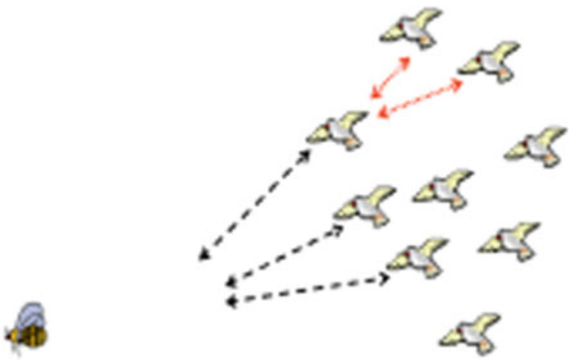


Fig. 3 Particle or agent

these values depend on the data that the particle carries. gbest represents that particle's data which is closest to the target. Stopping value indicates the point at which the algorithm should stop if no target is found (Fig. 3).

The particle's data could be anything.. In the above example (flock of birds), the data would be the X, Y, Z coordinates of each bird. The individual coordinates of each bird would try to move closer to the coordinates of the bird which is closer to the food's coordinates (gBest). If the data is a pattern or sequence, then individual pieces of the data would be manipulated until the pattern matches the target pattern.

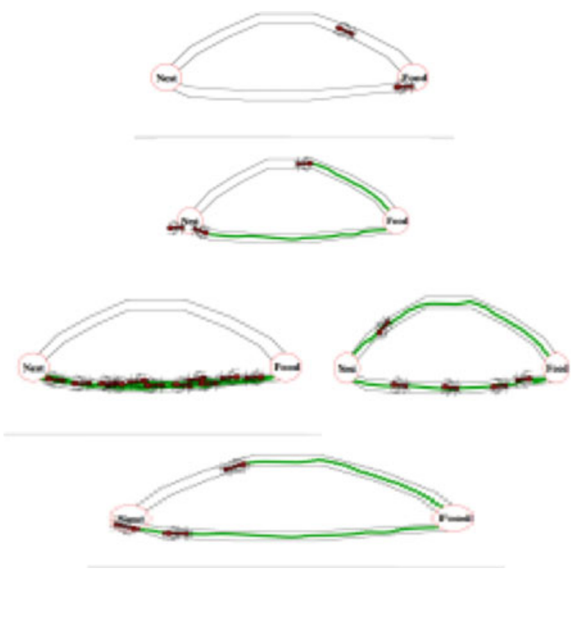
Let's take another example of a bio-inspired solution. Ant Colony optimization is a technique that has been inspired by the foraging behavior shown by ants. Foraging has been seen as one of the primary reasons for exhibiting swarm behavior in insects and birds. Ants, especially, work in a very robust and distributed fashion where they optimize for time and energy at every step.

To begin with ants wander randomly out of their colony in search of food. Once a food source is located, ants lay down pheromone trails on their way back to the colony. Also Pheromone evaporates over time. It is very likely that there would be multiple trails for ants to follow at a given point in time. The decision for choice of trail is based on the density of pheromone deposited for a particular trail. The density of pheromone depends on a couple of factors, namely the distance of the food and how frequently the trail has been used. Father the food is lesser the density of the trail. Also the more frequently traveled trails have greater density because every time an ant travels the trail it leaves behind pheromone.

Pheromone now becomes the only means of communication for the ants. This kind of indirect communication via local environment is called stigmergy. The above described mechanism provides better adaptability to local conditions, robustness, and fault tolerance by means of redundancy.

The objective of the strategy is to exploit historic and heuristic information to construct candidate solutions and fold the information learned from constructing

Fig. 4 Demonstration of Ant Colony Optimization



solutions into the history. Solutions are constructed one discrete piece at a time in a probabilistic step-wise manner. The probability of selecting a component is determined by the heuristic contribution of the component to the overall cost of the solution and the quality of solutions from which the component has historically known to have been included. History is updated proportional to the quality of the best known solution and is decreased proportional to the usage if discrete solution components. This is a typical example which demonstrates why bio-inspired computing has an upper hand over classical computing (Fig. 4).

3 Neural Networks

Artificial neural networks (ANN) are computational models inspired by an animal's central nervous systems (in particular the brain) which is capable of machine learning as well as pattern recognition. They were originally designed to mimic the human brain functions. Though the primary influence is the human brain, most of the neural modeling mechanisms have sound statistical backing and represent numerous statistical distributions and functions. Artificial neural networks are generally presented as systems of interconnected "neurons" which can compute values from inputs [8].

Neurons work by processing information. They receive and provide information in the form of spikes. An artificial neuron works in a similar way wherein it receives a set of inputs, processes them, and produces an output (Fig. 5).

McCulloch–Pitts Model of a Neuron: An artificial neural network is composed of many artificial neurons that are linked together according to a specific network architecture. The objective of the neural network is to transform the inputs into meaningful outputs [9] (Fig. 6).

Artificial neural networks are essentially simple mathematical models defining a function **f**. The function of **f** is decision-making, given the inputs and other hidden factors the function must make a decision, which leads to the output of the network. Sometimes models are also intimately associated with a particular learning algorithm or learning rule. The neural network model is designed to improve upon its performance by learning from previous experiences. This is similar to how humans

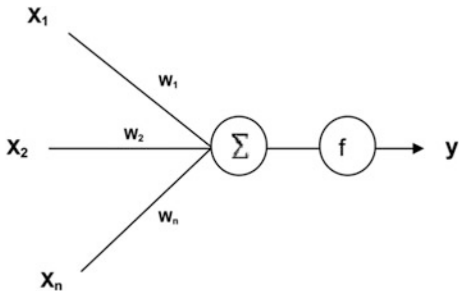


Fig. 5 Neuron work processing

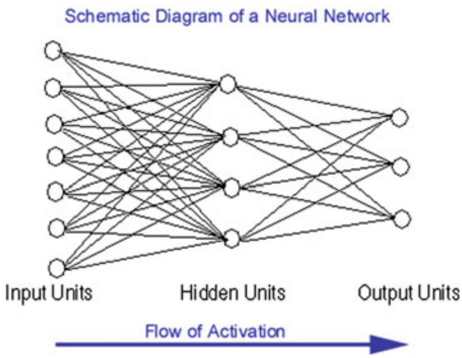


Fig. 6 Schematic and flow of activation in neural network

learn and improve their decision-making capabilities from their past experiences [10].

The following are few of the prominent areas of application for neural networks:

- Character Recognition
- Speech Recognition
- Speech Synthesis
- Signal Recognition and classification
- Trajectory control in Robotics
- Search and retrieval, especially in image search

4 Artificial Immune Systems

Artificial immune systems (AIS) are a class of computationally intelligent systems inspired by the principles and processes of the vertebrate immune system. The algorithms typically exploit the immune system's characteristics of learning and memory to solve a problem.

The field of AIS is concerned with abstracting the structure and function of the immune system to computational systems, and investigating the application of these systems towards solving computational problems from mathematics, engineering, and information technology. AIS is a subfield of biologically inspired computing and natural computation [11].

The human body's immune system is a perfect example of a learning system. It has the ability to distinguish between good cells and potentially harmful ones. AIS are learning and problem solvers based on our own immune systems. The AIS have been used to solve a wide variety of problems including:

- Computer Security
- Pattern Recognition

- Bridge Fault Detection
- Data Mining

In the human immune system, pathogens are the harmful cells which cause harm, B cells (B-lymphocytes) are the detecting antibodies. T cells (T-lymphocytes) are generated in B cells and they attack the pathogen. Memories of the previous infection are retained. This is how vaccination systems work in humans. Once encountered with a disease the human systems remembers it and learns how to fight and eradicate it. Thus, the body becomes immune to it and will not be affected by it.

In the AIS, we have something which is very similar to this. We have the immune network theory which is similar to the Network of B cells. We generalize that an initial set of cases are used as training data, the past experiences are used to make decisions and improve the performance. It is a self learning system [12].

4.1 Immune Network Theory

The immune Network theory had been proposed in the mid-seventies (Jerne 1974). The hypothesis was that the immune system maintains an idiotypic network of interconnected B cells for antigen recognition. These cells both stimulate and suppress each other in certain ways that lead to the stabilization of the network. Two B cells are connected if the affinities they share exceed a certain threshold, and the strength of the connection is directly proportional to the affinity they share.

4.2 Negative Selection

The concept of Negative selection is identification and deletion of self-reacting cells. They are typically used for classification, pattern recognition, and anomaly detection. It is done by randomly generating detector cells, which destroys anything that matches the self-cells and accepts it if it is not a self cell.

4.3 Clonal Selection

The clonal selection principle describes the basic features of an immune response to an antigenic stimulus. It establishes the idea that only those cells that recognize the antigen proliferate, thus being selected against those that do not. The main features of the clonal selection theory are that:

- The new cells are copies of their parents (clone) subjected to a mutation mechanism with high rates (somatic hyper mutation);

- Elimination of newly differentiated lymphocytes carrying self-reactive receptors;
- Proliferation and differentiation on contact of mature cells with antigens.

Here, the fit cells are allowed to grow in number, the unfit cells are slowly removed. The procedure of cloning is directly proportional to the fitness and mutation is inversely proportional. This is done by calculating the fitness, selecting the k best fit cells, clone them proportional to their fitness and mutate them inversely to fitness.

4.4 Bridge Fault Detection

Bridge is analogous to the human body. The vibrations caused in the bridge are antigens.

The self-set contains safe patterns. The detector set (B-cell) contains of unsafe and dangerous vibrations.

5 Conclusion

We have seen the different approaches bio-inspired approaches used in computing and the advantages of such methods over traditional computing methods. But the applications of BIC restrict to classification and optimization problems. It may be hard to suggest an optimal BIC technique for a given problem. The interplay between biology and computer science will hopefully help us solve many problems. It is hard to comprehend and understand the nature's way of computation. More research needs to be done to improve the BIC techniques and demonstrate real-life applications.

References

1. Biologically inspired computing by Nancy Forbes; Litton, Tasc.
2. Beyond Computation: Information Technology, Organization Transformation, and Business Performance by E. Brynjolfsson and L. Hitt, J. Economic Perspectives, Vol. 14, No. 3, Fall 2000.
3. Biologically inspired computing by Josh Bongard, University of vermant.
4. Biologically-Inspired Innovation in Engineering Design: A Cognitive Study by Swaroop Vattam, Michael Helms, and Ashok K. Goel Design Intelligence Laboratory School of Interactive Computing Georgia Institute of Technology.
5. Bio-inspired design: An overview investigating open questions from the broader field of design-by-analogy by Katherine Fu, Diana Moreno, Maria Yang, Massachusetts Institute of Technology & Singapore University of Technology and Design Republic of Singapore.

6. Biologically Inspired Design: Methods and Validation Carlos A. M. Versos and Denis A. Coelho Universidade da Beira Interior Portugal.
7. Swarm Intelligence and Bio-Inspired Computation: An Overview, Xin-She Yang, Mehmet Karamanoglu, Department of Design Engineering and Mathematics, School of Science and Technology, Middlesex University, The Burroughs, London, UK.
8. Biologically-Inspired Spiking Neural Networks with Hebbian Learning for Vision Processing Lyle N. Long1 and Ankur Gupta2 The Pennsylvania State University, University Park, PA 16802.
9. Biologically inspired neural networks for the control of embodied agents Răzvan V. Florian Center for Cognitive and Neural Studies (Coneural) Str. Saturn 24, 400504 Cluj-Napoca, Romania.
10. Bio-Inspired Computer Fovea Model Based on Hexagonal-Type Cellular Neural Network by C. H. Huang, C. T. Lin Dept. of Electr. & Control Eng., Nat. Chiao Tung Univ., Hsinchu.
11. Artificial Immune Systems by Julie Greensmith, Amanda Whitbrook and Uwe Aickelin.
12. Theoretical advances in artificial immune systems by J. Timmis,A. Hone,T. Stibor,E. Clark, Department of Computer Science, University of York, UK.

Providing Cloud Services to an Autonomous Robotic Car in Real-Time Traffic Environment

Prachi Sablani, Saurabh Kumar Prasad, Ashish Kumar Pandey and Rajesh Doriya

Abstract The era of autonomous car is yet to evolve, as the job of planning navigation path in dynamically changing environment has become a challenging task for researchers. One of the fundamental dynamically changing parameter is traffic. In this paper, we have proposed dynamic vehicle traffic routing algorithm to facilitate the robotic car to find the appropriate route from a particular source location to reach the destination. Moreover, we have demonstrated this algorithm by applying different constraints like lookahead condition and previously acquired values as heuristic. Illustrations of these variants of algorithm are analyzed, based on the real data of a city. In addition to it, we have described how providing a cloud service can enhance the computation speed of enormous and continuously changing traffic data, thereby meeting storage requirements. OpenStack is used as a software platform to provide this service to the large-scale local network.

Keywords Dynamic vehicle traffic routing · Cloud computing · OpenStack

1 Introduction

An autonomous robotic car is a device that has the capability to acquire information of its environment and helps in navigation without human efforts. Robotic cars generally operate in domains that change dynamically under unknown terrains.

P. Sablani (✉) · S.K. Prasad · A.K. Pandey · R. Doriya
Department of Information Technology, National Institute of Technology, Raipur 492010, India
e-mail: sablaniprachi@gmail.com

S.K. Prasad
e-mail: code.saurabh19@gmail.com

A.K. Pandey
e-mail: hisash0102@gmail.com

R. Doriya
e-mail: doriya1@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_43

They have control systems that extract the information obtained from sensors, which is very useful to evaluate suitable navigation paths so as to reach the desired destination. This paper outlines a goal-oriented path searching problem where a robotic car has to move from a particular source to desired destination. This is an application of sensor-based path searching [1, 2].

As the robot moves it encounters with camera sensors at every node (intersection of two paths). These sensors capture the real-time traffic information on the road and send this information to the cloud. We have used OpenStack, an open-source cloud computing software to provide this cloud service. The whole process is divided into three stages-

- **Collect**. In this stage, as soon as the robotic car reaches a particular node, the camera sensors at that node collects the real-time traffic information. Then the information is forwarded to the cloud for further processing. The development of radio standards over the past few years has enabled us to send information, captured at data-gathering devices (sensors, cameras) to the cloud, and to send commands back from the cloud to act on the information.
- **Analyze**. As soon as the cloud platform receives the real-time traffic information, it analyzes the information by deploying dynamic vehicle traffic routing algorithm (discussed in next section). This algorithm helps the robotic car in deciding the appropriate path with as less traffic as possible, by taking into account some suitable constraints.
- **Act**. The motion of the robotic car depends on the information it receives from the cloud. It chooses the most significant path according to current traffic conditions. The task of searching a path in goal-directed navigation problem is straightforward for known terrain and enough research has already been done [3]. However, for unknown terrains, it is still evolving. Earlier, the local planning of path was the center of concern and some theories have been proposed, like curvature velocity approach [4], field-based technique, which is based on the assumption that goal attracts the moving body and obstacles repel them [5], and some other like dynamic window [6] approach. Our paper focuses on path finding strategy which aims at finding the path with minimum traffic by considering an additional heuristic condition. This problem holds resemblance with the graph search problem solved by incremental search or parallel search process [7] and our algorithm is typically a modified version of A-star algorithm. Incremental search problem basically involves path finding problem where a convenient path has to found as the edge cost or the topology of graph changes [8]. Some examples are [9–13]; however, they consider different assumptions. Another one is well-known bug algorithms [14].

2 Dynamic Vehicle Traffic Routing Problem

2.1 Definition

The dynamic vehicle traffic routing algorithm is based on the fundamental that along with finding the fastest path to reach any particular location, it is crucial to select the path that is easily available. The selected path may or may not be the best path. The intensity of traffic is an important parameter to check the availability of the path. During the rush hours of the day, traffic is usually chaotic which makes it difficult for a vehicle to reach the desired location in a particular time interval. The traffic patterns, on the large-scale road network, changes continuously. Thus, this paper provides a method which evaluates the real-time traffic information at every node so that we can select the convenient path with minimum traffic.

As dynamic vehicle traffic routing algorithm is applicable for large-scale network, so to describe our algorithm, we are using a real-time map data of a city which can be represented in the form of a graph G, and V represents the set of vertices of the graph called nodes. Each vertex node represents the junction which is equipped with sensors to capture the traffic information. Every edge represents the path between two junctions. The graph is subjected to following constraints:

- The topology of graph does not change.
- The path should be available and computed in real time.
- The traveling time depends on the traffic of the road network.
- Traffic information is different at each node.

2.2 Algorithm Notation

Regarding each node, following information is given: $id(v)$ denotes the id of vertex $v \in V$, $prev(v)$ denotes the previous node of v, $neighbors(v)$ represents the set of successors of v, $goal\_dist(v)$ represents the estimated distance from v to goal node $vgoal, x\_value(v)$ represents the x-coordinate of v and $y\_value(v)$ represents the y-coordinate of v, $speed(v, v')$ represents the maximum speed of the path connecting v and $v', dist(v, v')$ represents the length of the path connecting v and v'. The algorithm determines an appropriate path from a given start node $v_{start} \in V$ to desired goal node $v_{goal} \in V$. Also, two types of values are maintained for each node: $g\_x(v)$ and $h\_x(v)$. $g\_x(v)$ is the total cost required to traverse from v_{start} to v and $h\_x(v)$ is an estimated cost between v and v_{goal}. Sum of these two distances is stored in third variable which represents the total cost and is denoted by $f\_x(v)$. Also, new_gx, new_hx and new_fx represent the recalculated values of g_x, h_x and f_x respectively. Furthermore, a priority queue is maintained namely 'openlist' which contains all the successors of already explored nodes and there is a closed list which contains all the explored nodes. $is\_openlist(v)$ stores the information whether the node v is in open list or not and same goes for *is_closedlist*(v). In order to balance the magnitude of cost, we have used two constants: alpha as 10 and beta as 1. *traffic_range* is the range of values that traffic can possess.

Following functions are used by pseudocode to describe the algorithm: *traffic*(v, v') which stores a random traffic value of the path between v and v', *openlist .insert*( ) which inserts a new element in the list before the element in the specified position, *openlist .empty*( ) which checks whether the list is empty, *openlist .erase*( ) function removes either single or multiple elements from the list, *openlist .begin*( ) function returns an iterator pointing to the first element of the list.

```
Algorithm 1:Dynamic Vehicle Traffic Routing Algorithm
Procedure initialize ()
traffic_gen()
for all the nodes v ϵ V
SET v.is_openlist = false
SET v.is_closedlist = false
SET v.prev = −1
Procedure path_find(v_start, v_goal)
initialize()
openlist.insert(v_start)
WHILE openlist is not empty, do
SET min_id = openlist.begin()
if(min_id == destn)
return
openlist.erase (openlist.begin())
SET min_id.is_openlist = false
closedlist.push_back(min_id)
SET min_id.is_closedlist = true
for all the neighbors n ϵ neighbourmin_id
SET effc_speed = speed (min_id, ni) − speed (min_id, ni)/traffic_range − traffic(min_id, n_i)
SET new_gx = min_id.g_x + dist (min_id, ni)/effc_speed
SET new_hx = alpha ∗ goal_dist (n_i)
SET new_fx = new_gx + new_hx
if (n_i.is_openlist == true)
then
if (new_gx < n_i.g_x)
openlist.erase(openlist.find(n_i)
SET n_i.g_x = new_gx; n_i.h_x = new_hx; n_i.f_x = new_fx
SET n_i.prev = min_id
openlist.insert(n_i)
end if
else if(n_i.is_closedlist == false)
then
SET n_i.g_x = new_gx; n_i.h_x = new_hx; n_i.f_x = new_fx
SET n_i.prev = min_id
openlist.insert(n_i)
SET n_i.is_openlist = true
end if
end while
store_path(v_goal)
Procedure main ()
import_data()
path_find(vstart, v_goal)
```

2.3 Algorithm Description

In the Algorithm 1, first the main method imports the data of the city from the database and then calculates the path by calling the function $path\_find(v_{start}, v_{goal})$. The $pathfind(v_{start}, v_{goal})$ method initializes the required parameters by calling $initialize(\,)$ method. This $initialize(\,)$ method, on the other hand, calls the $traffic\_gen(\,)$ method which stores the random traffic values ranging from 0 to 1, between every connected node, while the unconnected nodes have infinity as their traffic value.

In the next step, starting node is inserted in the openlist and the node with the minimum cost is extracted from the openlist. Check if this node is the goal node, if yes then path is found, otherwise continue the search. Remove this element from the open list and insert into the closed list. Now, to find the next suitable node perform *step 12* to *step 30* for all the neighbors of this extracted node.

For any neighbor, there could be three possibilities: First, if this neighbor is present in closed list, then simply avoid it. Second, if it is not in the open list, insert it into the same along with g_x, h_x and f_x values. And finally, if it is already in the open list then recalculate the values of g_x, h_x and f_x and update it if it is lower than the previous ones. Updating of the speed between the current node and its neighbor node requires the recalculation of speed as it will be constantly changing because of changing traffic conditions. The cost is calculated by taking into account the original cost (without any traffic), traffic conditions, and estimated distance from the goal. Repeat this for all the neighbors and compare their costs. The main purpose behind this comparison is to find the neighbor with the minimum cost so that it can be chosen as the next node to be traversed. Continue this process till the goal is reached.

2.4 Additional Constraint

In addition to the previous section in Algorithm 2, a lookahead condition (multi-hop) is added, according to which path can be accurately found by considering the traversal cost of upcoming nodes of immediate neighbors. For example, suppose we want to analyze the nodes up to two lookahead (two hops) then the cost up to secondary successors of immediate neighbors is analyzed. This helps us in finding the more efficient path, thereby avoiding the path having high traffic.

Thus, apart from the above-mentioned functions one more function $next\_hop(v_{start}, hop\_count, v_{goal})$ is added in the algorithm. This recursive function calculates the cost up to desired hop counts and returns the minimum cost with respect to concerned neighbor. This cost is then added to the total cost, thus producing better results.

Algorithm 2: Function for lookahead condition
Procedure next_hop (vstart, hop_count, vgoal)
1. **if**$(hop\_count == 0)$
2. **return**
3. $SET\ min\_cost = \infty$
4. **for** $all\ the\ neighbours\ n \in neighboursstart$
5. $SET\ newcost = dist(vstart, ni)/effc\_speed + alpha * goal\_dist(n_i) + next\_hop\ (n_i, hop\_count - 1, vgoal)$
6. **if** $(\,!\,n_i.is\_closedlist \&\& newcost <= min\_cost)$
7. **then**
8. $SET\ min\_cost = newcost$
9. **end if**
10. **return**min_cost

2.5 Heuristic

As a part of the heuristic, one more parameter can benefit the searching process. It is a self-learning process where we use the information from previously traversed paths to select the next path and then update the selected path for future reference. For example, suppose there is a path which has been traversed frequently then it will be having more chances of getting opted again in the near future.

Suppose, $history(v, v')$ denotes the previously recorded information regarding the possibility of opting node v' after reaching node v. This value is added to the heuristic cost, i.e., h_x and is updated everytime v' is chosen after v. Therefore, new_hx can be calculated as

$$new\_hx = alpha * estimated\_dist(n_i, v_{goal}) + beta * history(min\_id, n_i)$$

3 Use of Cloud Computing in Dynamic Vehicle Traffic Routing Algorithm

3.1 Introduction and Necessity of Cloud

Limitations due to on board computation power, storage size, and battery backup has brought the need to perform complex computations or store large data on a dedicated server's hardware, which can provide the much needed computation power and larger memory that the robot lacks. When we consider real-time traffic information of a large-scale network, like that of an entire city, then it becomes difficult and tedious for a normal machine, having general configurations, to handle such an enormous volume of data. Cloud Computing is a one such paradigm which is designed to make use of networked computers and communication system in a more cost efficient way.

The National Institute of Standards and Technology (NIST) [15] defines Cloud Computing as *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."* There are three types of service models for cloud—software as a service, platform as a service and infrastructure as a service and four deployment models—public cloud, private cloud, hybrid cloud and community cloud.

- **Software as a Service (SaaS).** The product in SaaS is an application offered to users over Internet. These applications are hosted by a cloud service provider. The customer gets the privilege of administrative access but has no control over implementation of application.
- **Platform as a Service (PaaS).** A developmental platform, designed for the programmers to create, test, and manage the applications. It provides APIs, programming languages, and other necessary tools required to develop and deploy application.
- **Infrastructure as a Service (IaaS).** The service providers outsource their infrastructure such as servers, virtual machines, storage units, networking components and other resources which the user with low-level infrastructure can use them as per their needs.

The first step in building a cloud environment involves choosing a cloud management solution. Since each solution is suited for different scenarios and has some trade-offs, this makes it a very difficult decision. Open-source frameworks offer advantage over enterprise clouds by offering the freedom to modify the source code, no licensing cost, strong user and developer community, transparency, open to extensions while reducing costs and avoiding vendor lock-in.

It is becoming more and more difficult to choose the most suitable open-source cloud platforms due to emergence of large number of cloud solutions. [16] performs a qualitative comparison of architecture and their implementation, [17–19] evaluates the features and performance of various open-source clouds on different evaluation criteria and helps in choosing the solution to suit specific needs.

3.2 OpenStack

In this paper, we have used OpenStack, a fully open-source cloud which provides IaaS solution for public and private clouds. OpenStack scales both vertically and horizontally for meeting diverse computing requirements, works with Hadoop for big data needs, and offers high-performance computing (HPC) for intensive works. OpenStack is built using python and available under Apache 2.0 license. Dashboard is a web interface for allocating and releasing of resources from end users. It also supports most of the virtualization solutions like Hyper-V, KVM, ESX, LXC, UML, Xen, QEMU, and XenServer. The main components of OpenStack include Horizon (Dashboard), Nova (compute), Neutron(networking), Cinder(storage), Glance (Image service), Swift (object storage), Heat(orchestration), and Keystone (identity).

In this paper, in order to meet storage requirements and speed up the path searching process, OpenStack launches a separate instance for each robotic car, which is a virtual machine inside the cloud. This dedicated instance runs the dynamic vehicle routing algorithm separately for each robotic car and thus harnessing the power of cloud to help the robot reach its destination.

3.3 Other Cloud Computing Platforms

- **CloudStack**. CloudStack is gaining momentum in the open-source CMS market. It has two parts—Management server and Cloud infrastructure. The former is responsible for management of cloud resources, assignment of guest VMs, assignment of IP addresses, allocating storage during VM instantiation, managing snapshots, disk images, etc. It provides a web interface for system administrators and end users, as well as API for CloudStack API and the Amazon EC2 interface. CloudStack also allows us to integrate with external load balancers like Citrix NetScaler apart from using existing load balancing algorithms.
- **Eucalyptus.** Eucalyptus (now acquired by HP) is an AWS compatible open-source CMS. It supports installation of private and hybrid clouds while Eucalyptus main target is hybrid installations and its compatibility with Amazon Web Service AWS.
- **OpenNebula**. While it is being most commonly used as a private cloud, it also supports public and hybrid clouds. OpenNebula is well suited for managing virtualized data center and most widely used among research institutions. OpenNebula include XEN, KVM, or VMware virtualization. It supports cloud interfaces like VMware vCloud, Amazon Web Services and provides features like secure multi-tenancy, user management, multiple user rolling, and quota management.

In addition to open-source cloud computing services, five leading public cloud providers are Amazon EC2, Windows Azure, IBM cloud, Google Cloud Platform, and Rackspace.

4 Experimental Results

For testing the algorithm, we have used a data from a clipped portion of open street map of New Delhi, containing 6000 nodes. We analyzed it for following conditions:

- Using no lookahead
- Using multiple lookaheads.
- Using multiple lookaheads and additional heuristic condition (history).

Table 1 shows the result for low configuration system, in which it is observed that the average travel time decreases as lookahead increases and it is further improved when additional heuristic parameter is applied. Average CPU time represents the total computation time taken by the CPU for guiding the robot from source to destination. It is also observed that CPU time increases as the lookahead increases to perform the additional calculations. We can observe the percent decrease in travel time taking lookahead as 0 as our benchmark.

Table 2 shows the result for high end cloud infrastructure, according to which algorithm's performance is drastically improved. It takes half the time to perform the same operations as compared to a low-end configuration system. Also, due to large storage capacity of cloud it becomes easy to handle the enormous data.

In the below graph (Fig. 1), we have compared travel times during different conditions like for look ahead 0, 1, 3, and 5, testing it for both history and without history condition. It is concluded that it takes less time to travel if history conditions are applied.

From the below graph (Fig. 2) it can be observed that the computation time of high-end cloud infrastructure is much less than that of low configuration system. Thus, the overall performance is improved.

Table 1 A comparison of algorithm's performance for various conditions taking alpha as 10 and beta as 1 between Source 80 and Destination 1000

System configuration	Lookahead	History	Avg. CPU time (s)	Avg. travel time (s)	% decrease in travel time
Intel Core 2 Duo CPU T6600 @2.20 GHz 3 GB RAM	0	NO	0.359	2695.456	–
	1	NO	0.751	2497.858	7.33
	3	NO	1.023	2143.342	20.48
	5	NO	2.916	1662.998	38.30
	1	YES	0.792	2123.364	21.22
	3	YES	1.211	1737.091	35.55
	5	YES	3.156	1374.513	49.00

Table 2 A comparison of algorithm's performance when performed on high-end configuration on cloud, taking alpha as 10 and beta as 1 between Source 80 and Destination 1000

System configuration (cloud)	Look ahead	History	Avg. CPU time (s)	Avg. travel time (s)	% decrease in computation time
Intel core i7 CPU 3630QM @2.4 GHz 8 GB RAM	0	NO	0.131	2695.456	63.50
	1	NO	0.299	2497.858	60.18
	3	NO	0.461	2243.342	54.93
	5	NO	1.495	1662.998	48.73
	1	YES	0.328	2123.364	58.58
	3	YES	0.618	1737.091	48.96
	5	YES	1.543	2074.513	51.10

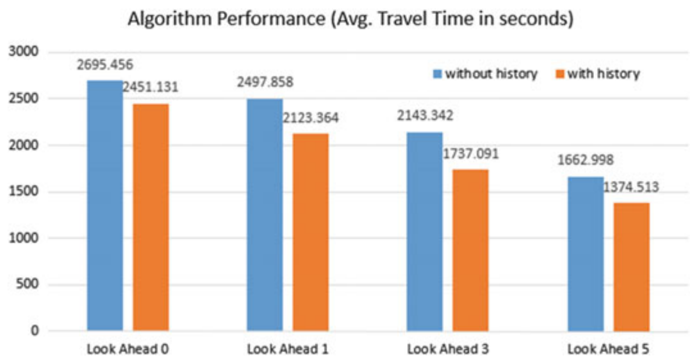


Fig. 1 Comparison of algorithm's performance for various conditions

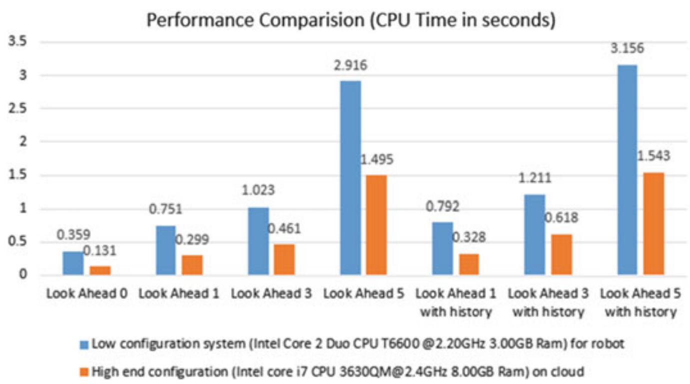


Fig. 2 Comparison of algorithm's performance when performed using high end configuration on cloud

5 Conclusion and Future Work

To sum up, we have presented an algorithm for planning navigation in unknown terrain. The paper depicts how the motion of robotic car can be aided by dynamic vehicle traffic routing algorithm, which monitors the dynamically changing traffic patterns and generates the most accurate path accordingly. Along with the traffic condition, it has taken into account the lookahead and additional heuristic condition. The experimental results show that the application of these additional conditions has significantly reduced the travel time for mobile robots. Also, the cloud service provided by OpenStack enhances the functionality of algorithm, thereby reducing the computation time and meeting the storage requirements.

References

1. H. Choset and J. Burdick, "Sensor based planning and nonsmooth analysis," in Proceedings of the International Conference on Robotics and Automation, 1994, pp. 3034–3041.
2. "Sensor-based planning, Part I: Incremental construction of the GeneralizedVoronoi graph," in Proceedings of the International conference on Robotics and Automation, 1995, pp. 1643–1649.
3. J.-C. Latombe, Robot Motion Planning. Kluwer Academic Publishers, 1991.
4. R. Simmons, "The curvature velocity method for local obstacle avoidance, "in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 1996.
5. O. Khatib, "Real-time obstacle avoidance for manipulators and Mobile robots," International Journal of Robotics Research, vol. 5, no. 1, pp. 90–98, 1986.
6. D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance." IEEE Robotics and Automation, vol. 4, no. 1, 1997.
7. M. Tao, A. Elssamadisy, N. Flann, and B. Abbott, "Optimal route Re-planning for mobile robots: A massively parallel incremental A* algorithm," in International Conference on Robotics and Automation, 1997, pp. 2727–2732.
8. G. Ramalingam and T. Reps, "On the computational complexity of dynamic graph problems," Theoretical Computer Science, vol. 158, no. 1–2, pp. 233–277, 1996.
9. G. Ausiello, G. Italiano, A. Marchetti-Spaccamela, and U. Nanni, "Incremental algorithms for minimal length paths, Journal of Algorithms,vol. 12, no. 4, pp. 615–638, 1991.
10. S. Even and Y. Shiloach, "An on-line edge deletion problem," Journal of the ACM, vol. 28, no. 1, pp. 1–4, 1981.
11. S. Even and H. Gazit, "Updating distances in dynamic graphs," Methods of Operations Research, vol. 49, pp. 371–387, 1985.
12. E. Feuerstein and A. Marchetti-Spaccamela, "Dynamic algorithms for shortest paths in planar graphs," Theoretical Computer Science, vol. 116, no. 2, pp. 359–371, 1993.
13. P. Franciosa, D. Frigioni, and R. Giaccio, "Semi-dynamic breadth-first search in digraphs," Theoretical Computer Science, vol. 250, no. 1–2, pp. 201–217, 2001.
14. V. Lumelsky and A. Stepanov, "Path planning strategies for pointautomaton moving amidst unknown obstacles of arbitrary shape," Algorithmica,vol. 2, pp. 403–430, 1987.
15. T. Grance, and P. Mell, "The NIST definition of Cloud Computing," National Institute of Standards and Technology (NIST), 2009.
16. D. Cerbelaud, S. Garg, J. Huylebroeck, "Opening the Clouds: Qualitative Overview of the State-of –the-art Open-Source VM-based Cloud Management Platforms", Proceedings of the 10th ACM International Conference on Middleware, 2009.

17. S. Wind, "Open Source Cloud Computing Management Platforms: Introduction, Comparison and Recommendations for Implementation", IEEE Conference on Open Systems, pp. 175–179, 2011.
18. I. Voras, B. Mihaljecić, M. Orlić, et al, "Evaluating OpenSource Cloud Computing Solutions", MIPRO, Proceedings of the 34th International Convention, pp. 209–214, 2011.
19. M. Mahjoub, A. Mdhaffar, et al. "A Comparative Study of the Current Cloud Computing Technologies and Offers," IEEE First Symposium on Network Cloud Computing and Applications, pp. 131–134, 2011.
20. R. Doriya, N. Wadhwa, K. Suraj, P. Chakraborty and G. C. Nandi, "Dynamic vehicle traffic routing problem: Study, implementation and analysis using ACO and GA," Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on, Kanyakumari, 2014, pp. 1164–1171.

Part III
Communication

Evaluation Energy Parameter Using Jumper Firefly Algorithm in Wireless Sensor Networks

Sandeep Bidakar, H.H. Kenchannavar and Umakant P. Kulkarni

Abstract Application such as area monitoring, forest fire detection, land slide detection, and healthcare monitoring are used for continuous monitoring. The nascent wireless technology is used to develop such applications is known as Wireless Sensor Networks (WSNs). Sensor devices have limited energy and they are battery-operated devices. Various techniques like clustering and routing are employed to minimize the consumption of energy to maximize the lifetime of the network. Clustering is a process which is used to group the sensors based on properties associated with them such as energy, position, and degree of neighbourhood nodes. On the other hand, routing is a process of establishing optimal path to the intended destination. In the sensor network, each node gathers raw data and directly diffuses it to sink. In the proposed study, the clustering algorithm selects the cluster head node which collects the data and forwards to sink. Jumper Firefly Algorithm (JFA) is bio-inspired, meta-heuristic optimization-based clustering algorithm. The main aim of the proposed study is to exploit the features of centralized clustering jumper firefly algorithm to optimize the energy parameter and improve the network lifetime.

Keywords Centralized cluster-based routing · Energy-efficient clustering
Jumper firefly algorithm · Network lifetime · Throughput · WSN

S. Bidakar (✉) · H.H. Kenchannavar
Department of Computer Science and Engineering, KLS Gogte Institute of Technology, Belagavi 590008, India
e-mail: sandeepbidakar@gmail.com

H.H. Kenchannavar
e-mail: harishhk@git.edu

U.P. Kulkarni
Department of Computer Science and Engineering, SDM CET, Davalagiri, Dharwad 580002, India
e-mail: upkulkarni@yahoo.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_44

1 Introduction

In the coming years, sensor networks have large scope of relevance and are applicable to various fields such as noticing scientific phenomenon to use it in agriculture monitoring, warehouse account management, health monitoring applications, etc. In order to have an adequate idea about the scientific phenomenon, it is essential for researchers to gather numerous measurements of scientific happenings at a particular geographic region, these measurements can be gained from distant place through remote sensing.

There is no means for observing or noticing particular region of interest from the distant place. The form of technology that helps to monitor the region of interest and scientific happenings in the environment is called as Wireless Sensor networks (WSNs) [7]. Basically WSN consists of sensor nodes and probe devices, these probe devices are deployed throughout the network. They will examine local scientific conditions, also capable of processing in-network data and include computational resources. Recent advances in IC technology led to the development of micro-electro-mechanical systems (MEMS) resulted in emergence of low-cost and low- powered wireless sensors. Fill the text from your manuscript in different sections.

Wireless sensor networks are similar to mobile Ad hoc Networks (MANET). Both possess some common properties such as distributed nature, self-organized architecture, multi-hopped, and lack of fixed infrastructure. MANET's main feature is its mobility but sensor networks have no mobility or sometimes have low mobility. Basically sensor devices possess properties like less bandwidth availability, small processing power, and high amount of redundancy and power constrained. Bluetooth technology is the popular example for MANET whose main aim is providing wireless connectivity among pair of nodes which also has same features as sensors-nets. Sensor-nets are part of WSN, whose main aim is monitoring and gathering spatial, dense, temporal and continuous data from environment and directly diffuse them to collection station.

Bio-inspired algorithms [17] are kind of meta-heuristics which imitates the nature, for solving optimization problems. Nature is a great source of motivation for solving hard and complex problems in computer science. Since it possess extremely miscellaneous, flourishing, dynamic, and fascinating phenomenon. It always discovers the optimal solution to problems and maintains absolute balance with its components. Nature-inspired algorithms coming up as new epoch (age) of computation. Still being young and emerging, results are incredible, widens the scope, and vitality of bio-inspired algorithms.

2 Related Works

Wendi B Heinzelman et al. [1] in his study "An Application-Specific Protocol Architecture for Wireless Micro-sensor Network" sensor networks require robust wireless communication protocols helps to achieve low latency and energy efficiency. Authors have presented Low Energy Adaptive Clustering Hierarchy (LEACH) [1] is type of clustering technique provides good performance in terms of energy efficiency, data aggregation, lifetime, latency and application perceiving quality. In the study authors have followed centralized LEACH technique to form cluster based on energy, location and neighbourhood nodes. LEACH architecture performs local computation to reduce amount of data transmission time. Media access control (MAC) and routing protocols enable low-energy consumption in sensor nodes. In case of LEACH-C clusters are chosen in such way that energy required for cluster members to send their information to cluster head (CH) should be minimum. Base station utilizes its global knowledge regarding network to form a better cluster which needs less amount of energy to transmit information. The numbers of cluster heads (CHs) are predetermined and fixed in number. Optimal number of predetermined cluster heads will be limiting factor for this work because fixed number of cluster heads after depleting their energy completely, they may reach to certain threshold energy level or they may die at that point network operations ends, no more transmission is possible.

I.F. Akyilidiz, W. Su et al. [7] in the study "Wireless Sensor Networks: a Survey" suggested, sensing ability and potential applications of sensor networks are exploration. In wireless sensor networks, layered communication architecture is built based on requirements of functionality and services. To optimize and exploit features of layered architecture, algorithms and protocols are developed. In this survey, the authors have dealt with factors like low cost, rapid deploying ability, flexibility of sensor networks that allow creating wide range of applications in this domain. This study motivated us to rebuild the communication architecture as per the requirements of the proposed model and rebuilding layered communication architecture will help us to achieve high-energy optimization and network lifetime.

S. Selvakennedy, S. Sinnappan, Yi Shang in their study [18] "A Biologically-Inspired Clustering Protocol for Wireless Sensor Networks", this study experimentally investigates swarm intelligence technique, analysis performed to determine number of cluster heads (CHs) to achieve goal of optimal performance in terms of energy consumption. Biologically inspired T-ANT algorithm developed. It helps to find optimal number of cluster heads (CHs) in the network and this robust algorithm also capable to handle unforeseen situations such as failures of nodes. T-ANT achieves two swarm behaviours: separation and alignment. The algorithm substantially has less state overhead in memory compared to conventional LEACH. The study focused on selecting optimal number of cluster heads but the authors have not mentioned regarding formation of clusters. Formation of cluster is important in sensor network because cluster depicts several properties of nodes such as initial energy, location, and degree of neighbourhood nodes.

Ali-Asghar salepour, Babak Mirmobin, Siamak Mohammadi, in the study [2] “An Energy Efficient Routing Protocol for Cluster based Wireless Sensor Networks using Ant Colony Optimization” presented two-level routing: first level (intra-cluster) cluster members send data to cluster heads (CH). In second level (inter-cluster) where cluster heads use Ant Colony Optimization (ACO) to find best path to base station. Only cluster heads (CH) can participate in inter-cluster routing and delay incurred with network is minimized by this bio-inspired algorithm. In the study authors presented idea of using Hybrid Energy-Efficient Distributed (HEED) clustering technique but not implemented any clustering technique in the study. ACO is a bio-inspired method for routing optimization. Idea behind ACO is ants by their behavioural nature generally discover shortest path to reach to their food source and share same route with other ants. In the study, the authors have tried optimization of routing without clustering which can be limitation to the study because in wireless sensor networks clustering plays vital role, without cluster formation it is difficult to elect cluster heads. If cluster heads are selected there is possibility that non-cluster head nodes may confuse with cluster heads they belong to, there is chance of cluster head may receive multiple duplicate packets from non-cluster head nodes. Overhead results at this stage because generally routing techniques involve path maintenance cost.

Sonam Palden Barfunga, Prativa Rai in the study [6] “Energy Efficient Cluster Based Routing Protocol for Wireless Sensor Networks,” suggested cluster-based hierarchical protocol and cluster heads (CH) are centrally elected by base station. Cluster head selection involves two stages: all nodes wishes to become cluster head are listed based on parameters like relative distance of nodes from base station, remaining energy level, probable number of neighbour sensor nodes and number of time the node has already become the cluster head. The data transmission inside cluster, between non-cluster head nodes to cluster heads and from cluster heads to base station takes place in multi-hop fashion. The current transmission round ends when the energy level of anyone cluster head reduces to half of its initial energy level. Network remains functional until all the preselected cluster heads reach their half of initial energy for this reason energy is associated with cluster heads are not completely utilized. If the numbers of clusters are chosen dynamically through clustering algorithm, that would resulted in better scalable cluster but this study lags in providing scalability.

Adil Hashmi, Nishant Goel, Shruti Goel in the study [10] “Firefly Algorithm for Unconstrained Optimization,” suggested the biological behaviours of fireflies (lightening bugs), like flashing nature of fireflies. And stated three important rules specific to Firefly algorithm: Fireflies are attracted by brighter ones regardless of their sex. Attractiveness is proportional to brightness. Brightness decreases as distance between nodes increases. Objective function determines the brightness of firefly. Advantages of firefly algorithm are: it is precise, robust, and easy to understand and supports parallel implementation. These features made algorithm popular and motivates to use it for researches and applications of sensor networks.

3 System Design

3.1 Network Model

In this section, the proposed network setup is assumed to have N number of sensor hosts, capable of gathering raw data from the environment which can be deployed over large geographic region.

For modeling such type of networks following assertions are made:

(1) All the hosts in the network have limited energy and network is stationary (no mobility).
(2) Every node periodically sense raw data and diffuse directly to base station through cluster head.
(3) Base station location is crucial factor in the network, base station may be inside or outside the sensing region of the network.
(4) Each node in the network can extend their transmission range.
(5) Each node in the network can operate as cluster head node as well as non-cluster head node.
(6) Data aggregation at the cluster heads will remove unnecessary redundant data.
(7) Data communication is performed in multi-hop fashion.

3.2 Protocol Description

The Jumper Firefly algorithm is a bio-inspired algorithm, it is developed based on performing efficient trial and error method to find an acceptable accurate solution, this encourages learning advances in the networking through experimentation. As name suggests this algorithm explains and deals about the specific kind of lightning bugs or insects which are seen typically on tropical summer nights in high temperature areas. Most important feature of these fireflies is they produce light through some biological activity and generally move toward the darker areas. We are considering each of these bugs as sensor hosts which are having some limited amount of energy.

Three important characteristics of Jumper Firefly Algorithm [11]

- All fireflies are of same kind (unisex) and these fireflies attracted by brighter ones, regardless of their type.
- Attractiveness is defined as the one of the most important property, the degree of attractiveness is directly proportional to brightness of the firefly and as distance increases the brightness gradually decreases. If there is no more bright and attractive firefly present in the region, fireflies move in random direction.
- The brightness or intensity of light is determined by value of the cost function.

The Jumper Firefly Algorithm can be described using three important factors:

3.2.1 Attractiveness and Light Intensity

Light intensity is the brightness of the firefly. Attractiveness is directly proportional to brightness. Light intensity changes according to inverse square law.

$$E(r) = E_0/r^2 \tag{1}$$

Attractiveness [10] can be given as

$$\beta = \beta_0 \exp(-\gamma \mathrm{r}^m) \tag{2}$$

According to inverse square law as distance between nodes increases energy reduces. Distance between two nodes is computed with help of Pythagoras theorem as follows:

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y)^2} \tag{3}$$

3.2.2 Cluster Setup

In the proposed cluster algorithm, cluster formation by centrally running algorithm base station forms the cluster. The process of cluster formation depends on the value of objective of the cost function [16], cost function can be given as:

$$\text{cost} = \beta \times f_1 + (1 - \beta)f_2 \tag{4}$$

3.2.3 Energy Consumption

There are two energy consumption models [8] we studied, Free Space Propagation Model and Two Ray Ground Propagation model. Energy consumption can be computed as follows:

$$\text{Energy Consumption} = \begin{cases} \varepsilon_{fs} \times d^2, \text{if } d \le d_0 \\ \varepsilon_{tr} \times d^4, \text{if } d > d_0 \end{cases} \tag{5}$$

As described above if the distance between nodes is less than or equals to zero ($d \le d_0$) free space propagation method is employed for calculation of energy consumption. If the distance between nodes is greater ($d > d_0$) then two ray ground propagation model is used.

3.3 Routing Protocol

Fill Ad hoc On-demand Distance Vector (AODV) routing protocol [4] is used in the proposed study. The routing algorithm operates in two stages first (Intra-cluster routing) where nodes in the network diffuse there data to cluster heads. Second (Inter-cluster routing) cluster heads aggregates the data and forward to sink or base station. Source host sends the Route Request (RREQ) packet to intended destination by appending destination sequence number.

If direct path is available to the destination, on receiving RREQ packet, destination replies with Route Reply (RREP) packet, if the path get established and transmission begins. In the network if path break is identified then node attached to that path broadcasts Route Error (RERR) message to all its neighbouring nodes. For the next round of transmission, nodes will follow alternate route to the destination. If direct path to destination is not available RREQ will be sent to intermediate node and next intermediate nodes forward RREQ until best optimal path to destination is found.

3.4 Jumper Firefly Algorithm

Data: Generate S sensor hosts to contain C randomly selected Cluster Heads.

Map the randomly generated position of the sensor hosts with closest (x, y) co-ordinates.

Total number of sensor hosts is t, MaxGen overall number of hosts in the network.

Result: Positions of the Cluster Heads (CHs) are determined.
Generate initial set of sensor hosts;
Formulate light intensity I;
while ($t <$ MaxGen)
if (any host is corrupt)
Put sensor host in new position stochastically;
end if
map the position of sensor host with closest (x, y)
co-ordinates;
for $i = 1$ to n (all n sensor hosts)
for $j = 1$ to n (all n sensor hosts)
if ($I_j > I_i$), move sensor host i towards j;
end if
Evaluate change in the position and update the host's position values and
light intensity;
end for j;
end for i;
Rank sensor hosts, find the current best;
end while;
post process results position and intensity;
end procedure;

3.5 Activity Diagram

See (Fig. 1)

3.6 Simulation

In this proposed study, we have simulated sensor network with help of OMNET++ [14]. It is an object oriented, modular discrete network simulator. Developed in 1998, it is still new product. It is an open-source package freely accessible. This simulation package is built based on C++ language foundation, it offers class library, GUI support (animation of network environment and graphical network editing). This kind of simulator can be used for modeling queuing networks, validating hardware architectures, traffic modeling, and to measure evaluation performance of complex systems. The performance of the Jumper Firefly Algorithm is simulated using OMNeT++3.2.

The simulation of network is done for 34 nodes in a 400 $\times$ 400 m network field and the multiple sink are located at four corners of the network field. The following Table 1 give parameters computed during the simulation run and Fig. 2 shows the OMNeT++ simulation environment, connection establishment among nodes, cluster heads are indicated with antenna symbol. The simulation runs for longer time (say for 2000 s) at the end of the simulation run maximum number of nodes in the network lose their energy and links between the nodes are withdrawn. Nodes transmit huge number of packets and lose their energy for transmission as well as reception of packet. Nodes gradually lose their energy during transmission and they will reach to minimum energy (say 200 milliJoules in the present simulation scenario) (Fig. 3).

They also lose connections with neighbour nodes. Void region get created in that area, because of loss of energy. Objective of this study is to extend the transmission range of nodes, which have sufficient energy to transmit. By extending the range toward the void region, nodes in the void region can easily transmit their packets to their nearest CH. Hence network lifetime increases and energy [15] associated with each node is completely utilized for network functioning.

Several conditions to be checked if the current node is CH and receiving node is also CH or if a current node is CH and other node is sink or if a current node is sink and receiving node is CH or if a current node is source and receiver is CH or if a sender node is CH and receiving node is source and distance within 300 m range. If condition holds true then nodes can able to communicate. Else if distance between two normal nodes is less than or equal to 100 m they can able to communicate. Lifetime value will be computed in physical layer and stored in the text file called as life.txt. We inspect the change in the energy level during simulation by inspecting energy parameter as an object.

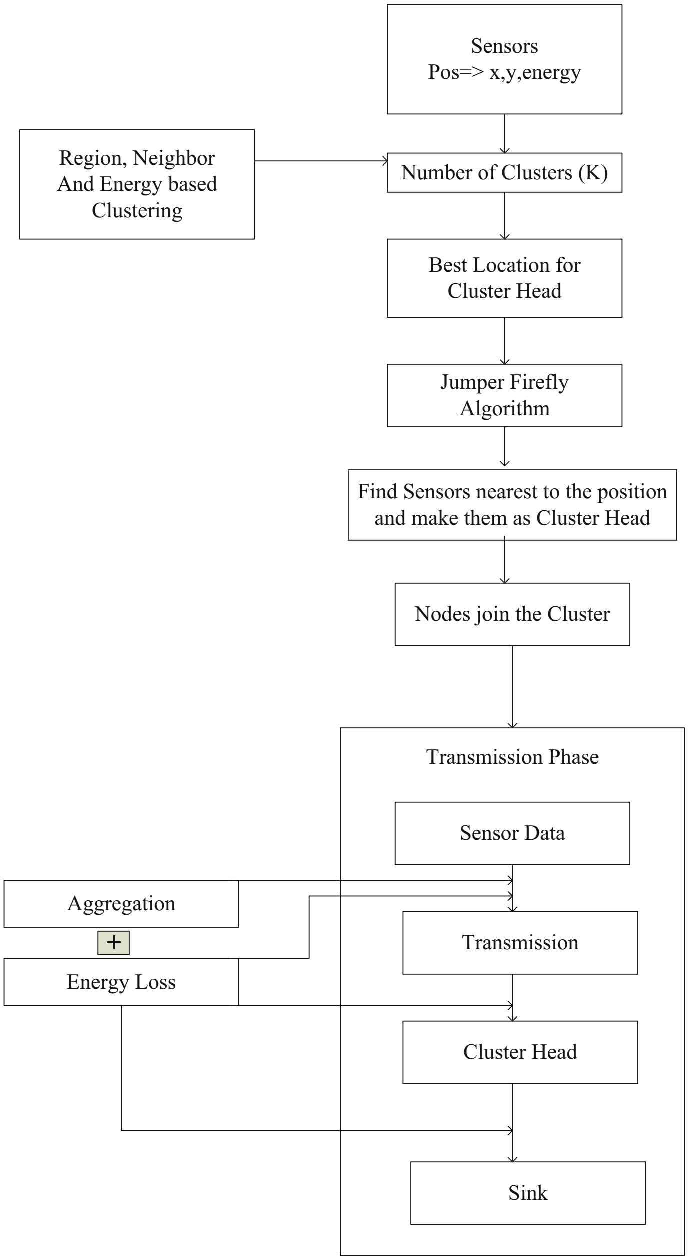


Fig. 1 Shows the activity diagram

Table 1 Parameters summary

Parameters	Value
Lifetime	338.028 s
PDR	92%
Throughput	1.403 Mbps

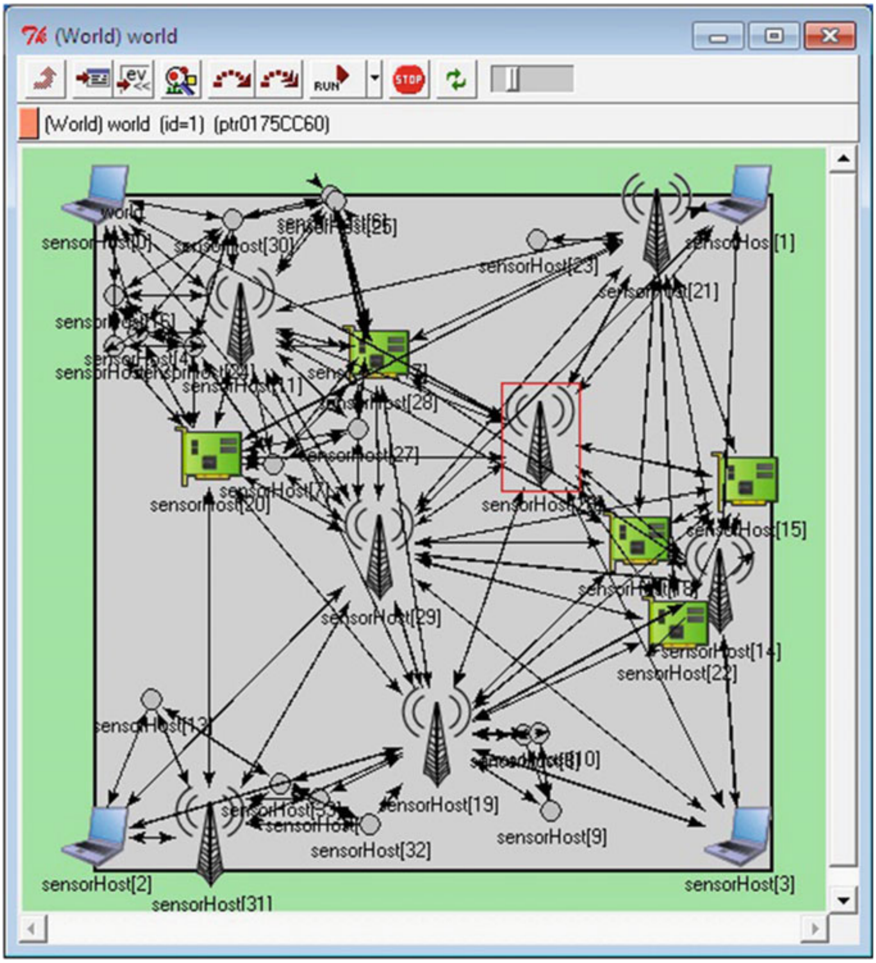
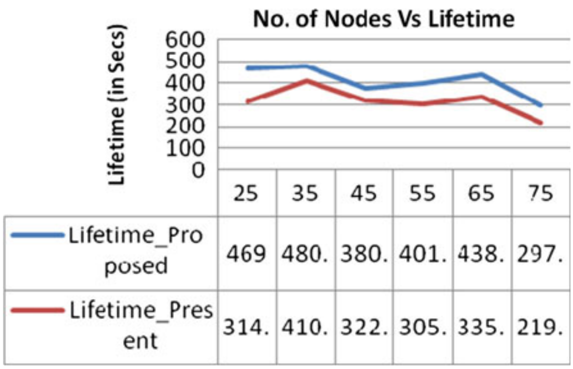


Fig. 2 Shows uniformly distributed WSN

Fig. 3 Pseudo-code of physical layer

```
Physical Layer isReachable
Physic *s = (Physic*) simulation.module(from);
Physic *d = (Physic*) simulation.module(to);
int x, y;
*parent=parentModule();
double dist, a;
bool r;
getPos(x,y);
sqrt (x_i-x_j)^2+(y_i-y_j)^2;
if dist<= range
if s->CH && d->CH || s->CH && d->sink || s->sink &&
d->CH|| s->source && d->CH|| s->CH&& d->source
&& dist<=500;
return true;
r=dist<=100;
return true;
```

Fig. 4 Shows the comparison of number of nodes versus network lifetime



	25	35	45	55	65	75
Lifetime_Proposed	469	480.	380.	401.	438.	297.
Lifetime_Present	314.	410.	322.	305.	335.	219.

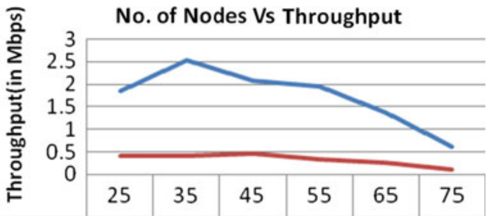
4 Result Analysis

In the analysis of results, we have compared number of nodes versus lifetime and throughput.

4.1 *Comparison of Number of Nodes Versus Network Lifetime*

Lifetime [11] is time period when first nodes dies from the start of the simulation. Graph shown in Fig. 4 clearly justifies as the number of nodes increase the lifetime also increases in proposed study. Lifetime increased by 10.04% (by taking average of lifetime values).

Fig. 5 Shows the comparison of number of nodes versus throughput



	25	35	45	55	65	75
Thrpt_pro	1.841	2.523	2.068	1.943	1.348	0.602
Thrpt_pre	0.399	0.418	0.457	0.332	0.257	0.114

4.2 *Comparison of Number of Nodes Versus Throughput*

The proposed system gives better results compared to present system. Throughput is measured in terms of bits per second. By observing the graph in Fig. 5, we can draw inference, number of nodes increases and throughput increases compared to present system. Throughput is increased by 26.29% (by calculating standard deviation of throughput values of the proposed work).

5 Conclusion

Wireless sensor networks are best suited for large geographical region in monitoring applications. Continually varying data such as temperature, pressure, humidity, health condition of patients etc., are gathered by sensor devices. There is need to have supportive methodologies to achieve successful operation of sensor networks. In the proposed work we have built the four-layered communication architecture to achieve energy efficiency. The application layer provides data needs to be transmitted to sink, routing involves establishment of routes to the destination node, MAC layer checks for the best optimal path for the transmission and most important physical layer achieves the objectives of the proposed work such as reachability of nodes in the network, energy optimization. Routing plays an important role in the sensor network, routing happens through CH. In the proposed work by considering the energy, geographical region and degree of neighbourhood CHs are elected. CH aggregates the data received by the cluster members and forwards it to sink or Base Station (BS). Cluster-based routing provides best QoS. The proposed study achieves energy optimization, by extending the transmission range of nodes toward void region in the network. Parameters like area coverage, lifetime, throughput, Packet Delivery Ratio (PDR) are computed. Simulation of proposed algorithm gives better results compared to conventional LEACH technique.

References

1. Cayirci E, I F Akyildiz, W Su and Y Sankarasubramaniam, 2002 Wireless sensor networks: a survey, *Elsevier Sci B.V.Computer networks 38*, vol. 1, no. 4, pp. 393–422.
2. S Binitha and Sathva Siva S, 2012 A Survey of Bio inspired Optimization Algorithms, *International Journal of Soft Computing and Engineering (IJSCE),* vol. 2, no. 2, pp. 137–151.
3. A Chandrakasan, B Hari and W. R. Heinzelman, 2002 an Application Specific Protocol Architecture for Wireless Microsensor Networks, *Wireless Communications IEEE Transactions*, vol. 1, no. 4, pp. 660–670.
4. S. Selvakennedy, Shang Yi and S Sinnappan, 2007 A Biologically-Inspired Clustering Protocol for Wireless Sensor Networks, *Elsevier, Computer Communications*, pp. 2786–2801.
5. Afzali-Kusha Ali, Salepour Ali-Asghar and Mirmobin Babak, 2008 An Energy Efficient Routing Protocol for Cluster based Wireless Sensor Networks using Ant Colony Optimization, *IEEE*, pp. 455–459.
6. Burfunga Palden Sonam, kumar Hiren, Rai Prativa, and Sarma deva, July 2012 Energy Efficient Cluster based Routing Protocol for Wireless Sensor Networks, *IEEE International Conference on Computer and Communication Engineering (ICCCE 2012)*, pp. 611–615.
7. Goel N, Goel S, Gupta D and A Hashmi, May-June 2013 Firefly Algorithm for Unconstrained Optimization, *IOSRJournal of Computer Engineering (IOSR-JCE)*, vol. 11, pp. 75–78.
8. Gopi Mahesh and Sarma N.V.S.N, April 2014 Energy Efficient Clustering using Jumper Firefly Algorithm in Wireless Sensor Networks, *IJETT,* vol. 10, pp. 525–533.
9. N. A. Latiff, B Sharif and C Tsimenidis, 2007 Energy Aware Clustering for Wireless Sensor Networks using Particle Swarm Optimization, *IEEE 18th International Symposium, Personal Indoor and Mobile Radio Communications(PIMRC)*, pp. 1–5.
10. Cheng W. and Shi H, 2009 AEEC: an Adaptive Energy Efficient Clustering Algorithm in Sensor Networks, *Industrial Electronics and Applications (ICIEA), 4th IEEE Conference*, pp. 3950–3954.
11. Bhagwat P and Perkins C, October 1994 Highly Dynamic Destination Sequenced Distance-Vector Routing (DSDV), Ad-hoc On-demand Distance Vector (AODV) Routing for Mobile Computers, *Proc. Of the ACM SIGCOMM*.
12. Hornig Rudolf and Varga Andras, March 2008 An Overview of the OMNeT++ Simulation Environment, *SIMUTools*, pp. 1–10.
13. Kenchannavar H.H, Kulkarni U P, January 2012 Optimized Energy-Aware Algorithm for collaboration in Visual Sensor Networks, *International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 327–335.
14. Balakrishnan Hari, Chandrakasan Anantha and Heinzelman Rabiner Wendi, 2000 January Energy-Efficient Communication Protocol for Wireless Microsensor Networks, *IEEE published in the Proceedings of the Hawaii International Conference on System Sciences*, pp. 1–10.
15. Bidar M. and Kanan H R, Oct–Nov 2013 Jumper firefly algorithm, *International Conference on Computer and Knowledge Engineering (ICCKE-2013)*, pp. 267–271.
16. Gao Teng and Jin Rencheng, 2008 A Regiional Centralized- Clustering Routing Algorithm for Wireless Sensor Networks, *IEEE*, pp. 1–4.
17. Haneveld P T Klein, K. Kopke, Lichte H S, Swigulski M, and Valentin S, 2008 Simulating Wireless and Mobile Networks in OMNeT ++ The MiXiM Vision, *Workshop on OMNeT++ (co-located with SIMUTools)*, pp. 1–8.
18. Havinga M, Meratnia Nirvana, Paul J and Taghikhaki Zahra, A Reliable and Energy Efficient Chain-clusterbased Routing Protocol for Wireless Sensor Networks, *IEEE Pervasive Systems, University of Twente, Enschede, Netherlands*, pp. 248–253.

Developing a Context-Aware Secure Model for Sensor Network Environment

Pooja Mohan and Manpreet Singh

Abstract Personalized healthcare services have improved the quality of service. Due to development of ICT and the advancement of medical care, the traditional domain of health care is replaced by monitoring the patient remotely. Applications exchange information among professional for providing care services by wearing sensors on the body of the patient as well as embedding them in the environments. Due to pervasive and mobile computing, these care services must be adapted according to changing context in dynamic environment. Developing these application require model to represent contextual information. Context ontology-based model allow application adaptation based on changing context. In this paper, we develop the healthcare context ontology for the effective handling of healthcare problems during an emergency situation.

Keywords OWL · SWRL · Privacy · Context · Security

1 Introduction

Pervasive health care provides services of health care to anyone, at anytime, and anywhere by removing restriction of time and location thereby increasing the quality of healthcare. In the underdeveloped countries, due to limited medical staff and the scarcity of finances, the quality of health care is demising. The only solution to solve this problem for providing better services is through serving the patients remotely. This requires the integration of hardware and software infrastructures for collecting patient health information in the home environment.

P. Mohan (✉)
Department of IT, GGDSD College, Panjab University, Chandigarh, India
e-mail: pooja.mohan@ggdsd.ac.in

M. Singh
University College of Engineering, Punjabi University, Patiala, India
e-mail: msgchd@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_45

In pervasive computing, for the application to work in dynamic environment, all the services must be capable to adapt themselves due to changing context. Context, we refer to any information that can be used to characterize the situation of an entity, where an entity can be a person, place, or physical or computational object [1].

Using contexts such as time, location, people, and activity, WSNs know which information is significant to transmit. It helps to retrieve the system information easily.

For dynamic environment, small, energy-efficient deployed sensors are used to recognize activities of the user using various machine learning techniques. Deploying these technologies in any applications is difficult due to heterogeneous platforms and languages [2]. Due to heterogeneity, the integration and exchange of information as well as their reusability is very difficult to achieve. As context can be considered a specific kind of knowledge, it can be modeled as an ontology. Context ontology model provides framework for representation, sharing and integration of context knowledge for various services. In this study we propose a context ontology model for smart healthcare where context reasoning is performed based on semantic web technologies. Due to this feature of semantic web to link information from heterogeneous sources allow ontology to used in various application areas such as medicine, bioinformatics, e-commerce etc. To process the content of information rather than just presenting information to humans an application make use of the Web Ontology Language (OWL). Ontology allow intelligent agents to perform reasoning on contextual information using declarative semantics as predicates written in OWL. It can be used as a knowledge base for sharing information in dynamic systems as well as it also supports the reusability [3, 4].

The paper is organized as follows. Section 2 presents the background of the work. Section 3 discusses the Approach and the methodologies, and Sect. 4 provides an overview of ontology construction and its phases. In Sect. 5, we present the overall discussion and conclusion.

2 Related Work

There are various methods for modeling context. They can be modeled by key-value, mark-up scheme-based, graphical (e.g., Unified Modeling Language), object-oriented, logic, and ontology-based approach [5]. Key-value models are being used for one specific application and also lacking the formality and sharing of knowledge across different systems. Mark-up scheme model do not represent complex constraints and contextual relationship. Others models such as graphical, object-oriented, and logic-based do not allow contextual reasoning as well as sharing of knowledge among various applications. Cooltown [6] and Context Toolkit [7], offer weak support for knowledge sharing and context reasoning.

The study in [8] focused on an ontology-based context modeling, management and reasoning process. The model proposed in the study [9] used an ontology to describe person, place and intension. The extensible context ontology is being used in pervasive computing environment by CONON [10]. There was lack of generality and no support for context classification, dependency and quality of control required for context reasoning.

Ontology-based model offers contextual reasoning based on the technology of semantic web [11]. Such models are capable to express concepts as well as their interaction in an easiest manner [12]. An ontology not only makes our model independent of programming and application environment but also allows standardization of structural context representation. It allows to infer knowledge using domain specific rules. An ontology is used to represent context into higher level (generic environmental domain) and lower level(domain specific) [13].

3 The Approach and Method

Knowledge Engineering Methodology which is an iterative approach used to identify the various classes, subclasses, and properties. There is no specific approach for developing an ontology. It uses an iterative approach of revise and refine to evolve an ontology.

Each ontology *O* contains a set of concepts (classes) *C* and a set of properties *P*. A class is a collection of individuals and a property is a collection of relationships between individuals (and data). A property that relates an individual to another individual is called object property and a property that maps an individual to a data literal is called data type property. The ontology model is consisting of set of concepts, each describing a physical or conceptual object including Person, *Environment*, *Service*, Sensors, *Location*, *Activity*. The study uses the ontological engineering process to integrate the different concepts to build the security framework for sensor network environment in the smart healthcare.

4 Ontology Development

We have used Protégé tool for development of our ontology. There are some steps for developing an ontology. First of all it is important to identify the domain and scope of an ontology. It is required to limit the scope of the model that we are going to design. The next is to check whether we can use an ontology that is being defined and used by some other application. After that it's very important to develop the Vocabulary, i.e., the list of all items as well as the properties.

After all the classes are defined in the model, the class hierarchy is defined. The development process may be top-down, bottom-up, and hybrid. Top-down process starts with the defining the most general concept and then to specific. In bottom-up it moves from specific to general. Hybrid is the combination of both approaches. The concepts, properties and restrictions, individuals of all ontologies are described as follows.

A. *Concepts*

The main class hierarchy, consisting of concepts involved in the said ontology is shown in Table 1. Thing is abstract super class for all classes. The concepts in the ontology with their brief description are as follows.

The main components of a Healthcare Ontology is shown in Fig. 1 graphically. The class hierarchy consists of person, role, platform, device, activity, location, service etc.

The class person consists of doctor, patient, nurse, administrator, manager, caretaker, insurance agent, accountant, etc. They are classified into different subclasses depending upon the role assigned to them. A doctor and an accountant can be in the role of a patient. They are assigned permission to access data as well as devices according to roles that they have assigned.

The person class is about the description of the concepts that are participants in the environment, is shown in Fig. 2. The subclasses are patient, doctor, visitor, pharmacist, nurse and insurance_agent etc. according to roles in which they are in.

The subclass ageStatus at any instance falls into any of the following category—child, adult, and old. The subclass healthStatus specifies the health condition of a patient. It may be veryCritical, critical, underObservation, average, and normal.

In the sensor network environment, the network includes the combination of hardware and the software. The class *platform* includes the collection of *hardware* and the *software* as its subclasses. The subclass *hardware* is divided into subclasses *sensor, transceiver*, and *actuators* as shown in Fig. 3.

The *sensor* subclass represents the number of sensors which are used to monitor the physiological data of a patient as well as the environmental condition. The subclasses of sensor class is *heartRate, pressure, sugar and temp*, which are used to

Table 1 Main concepts of context-aware healthcare ontology

Concept	Description
person	It includes patient, doctor, nurse, admin, manager
platform	It includes the hardware and the software used in the system
attack	Identifies various attacks in the system
alarmStatus	Forwards notifications to concerned person
activity	It lists the various activities in which a person may engaged in
location	It shows the various locations where person may be located in
device	It lists the devices that a person owns whether personal or of hospital
role	It is the set of roles taken up by the person
senParameters	It lists the parameters of sensor nodes
personalData	It is the list of attributes that a person may have
service	It categorizes the set of services
privacy	Various privacy conditions
security	Various security conditions
sensorMeasurements	It lists the parameters measured by the sensor nodes

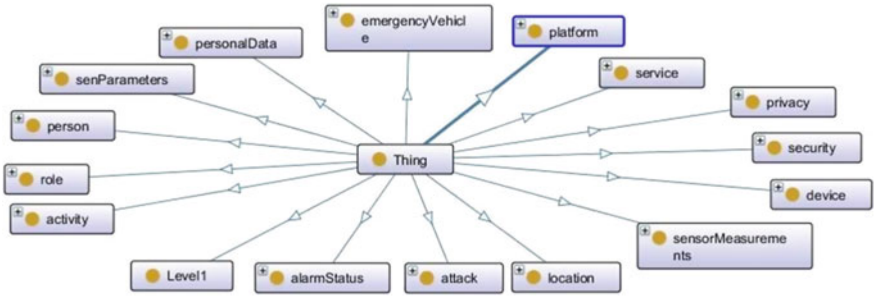


Fig. 1 Components of healthcare ontology

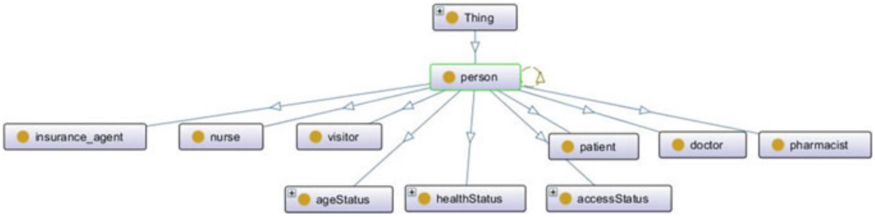
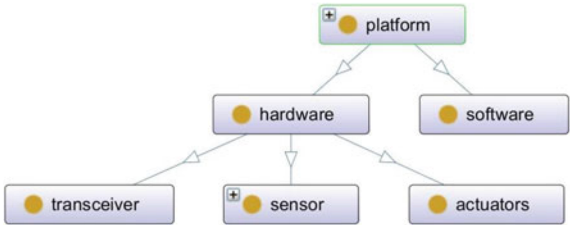


Fig. 2 Concept person ontology

Fig. 3 Concept platform and its subclasses



monitor the measurement of heart rate, blood pressure, sugar level and the temperature of a patient. First of all sensors are checked, whether they are ready to take participation in the communication. It is decided on the basis of the characteristics of the sensor network environment. If all the characteristics of nodes are in the range then sensors are ready to take part in the communication. The class *feasibleH*, feasibleP, feasibleS, and feasibleT represents all the sensor nodes which are ready for communication. All the other nodes which are not in the range are in the class *notfeasibleH, notfeasibleP, nonfeasibleS and nonfeasibleT*. The *feasibleH, feasibleP, feasibleS, and feasibleT* classes are divided into *secureH, secureP, secureS, secureT* and *insecureH, insecureP, insecureS, and insecureT* as its subclasses as shown in Fig. 4. The nodes which are free from attacks are secure nodes and is represented by *secureH, secureP, secureS and secureT* and nodes with the presence of attacks are insecure nodes represented by *insecureH, insecureP,*

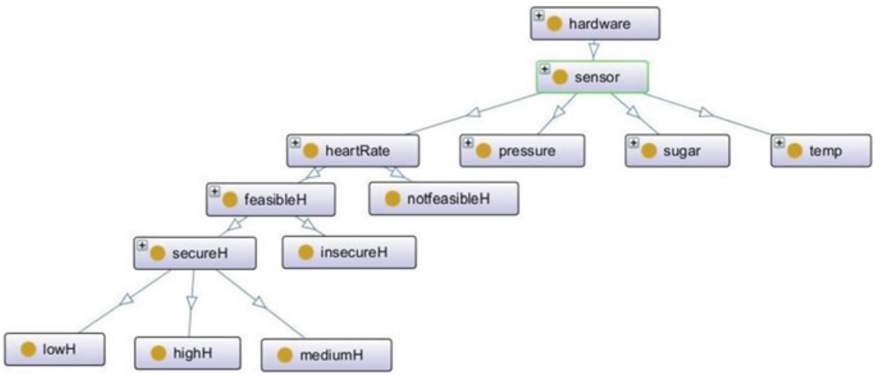


Fig. 4 Concept sensor and its subclasses

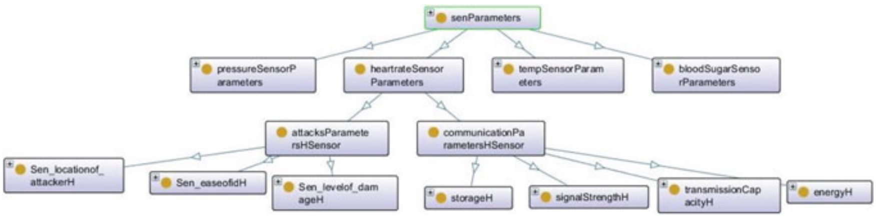


Fig. 5 Concept senParameters and subclasses

insecureS, and insecureT etc. The insecure nodes are not allowed to participate in the communication. The nodes which measures the patient heart rate is classified into *lowH, mediumH and highH* depending upon the range in which they are in. The nodes which measures the patient blood pressure is classified into *lowBP, normalBP, and highBP* depending upon the range in which they are in. The nodes which measures the patient blood pressure is classified into *lowT, highT, and veryHigh* depending upon the range in which they are in. The nodes which measures the patient blood pressure is classified into normalSugar*, preDiabetes, and diabetes* depending upon the range in which they are in.

The Class *senParameters* specifies the characteristics of various types of sensors. Each type of sensor includes features related to various attack conditions in the system as well as the communication parameters. The *senParameters* class is subdivided into subclasses *pressureSensorParameters, heartrateSensorParameters, tempSensorParameters,* and *bloodSugarSensorParameters* as shown in Fig. 5.

The *heartrateSensorParameters* subclass is consisting of *attacksParameters HSensor* and *communicationParametersHSensor* subclasses. The *attacks ParametersHSensor* subclass identifies attack conditions depending upon the value of the location of attacker, ease of identity and level of damage as shown in Fig. 6. The value assigned to subclass *Sen_locationof_attackerH* are implicit, explicit, and both.

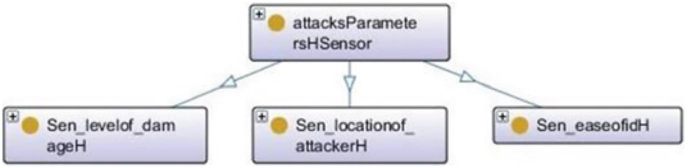


Fig. 6 Concept attacksParameterHSenor and its subclasses

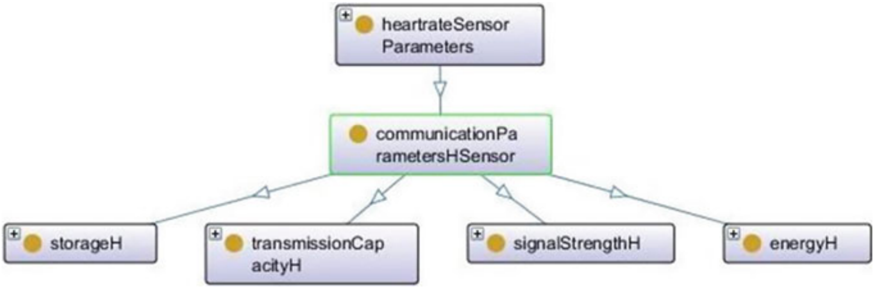


Fig. 7 Concept communicationParametersHSensor and its subclasses

The subclass *Sen_easeofidH* may contain the values easy, medium, and hard. The subclass *Sen_levelof_damageH* may be assigned the values low and high.

The *communicationParametersHSensor* subclass identifies the feasibility of sensors, whether they are ready to transfer their data. The various parameters are checked like storage level, energy level, signal strength and the transmission capacity of the sensor nodes as in Fig. 7. The value assigned to subclasses *storageH, transmissionCapacityH, signalStrengthH,* and *energyH* are low and high.

The concept *alarmStatus* identifies the various types of notifications that are being sent to different persons depending upon the health condition of the patient as well as the condition of the sensors that are being used to transmit physiological data. The various subclasses are *alarmToAdmin, alarmToNurse, alarmToCareTaker* , and *alarmToDoctor* etc. as shown in Fig. 8. The notification is sent to the administrator about all the sensor nodes which are not feasible or ready for communication. If the condition of some patient is very critical then alarm is send to the doctor, caretaker of the patient and the emergency vehicle transportation department so that patient may be transported to the hospital as early as possible. An alarm is being sent to the nurse treating the patient and the caretaker if the condition of the patient is critical. For a patient who is under observation an alarm is sent to the person who is taking care of the patient.

The concept sensorMeasurements is used to define the different range of readings measured by various types of sensors. The subclasses under *sensorMeasurements are sugarV, heartRateV, temperature, diaPressureV, and sysPressureV* etc. as shown in Fig. 9.

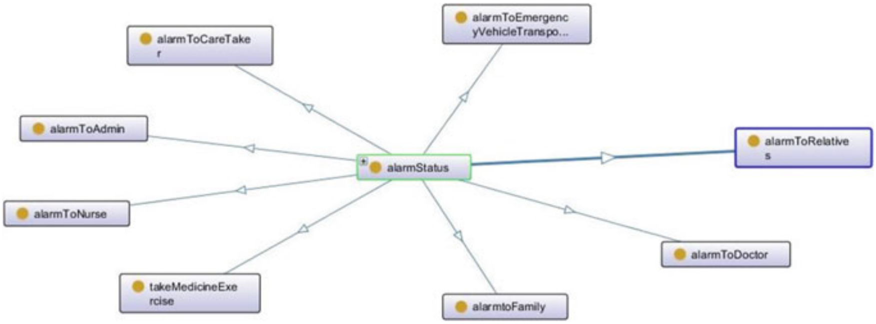


Fig. 8 Concept alarmStatus and its subclasses

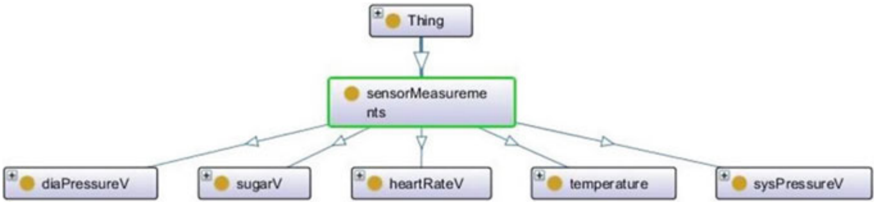


Fig. 9 Concept sensorMeasurements and subclasses

The concept *attack* represents the various types of attacks existing in the system. There are various attacks such as *internal, external, active, passive, laptop, mote-class, stealthy*, etc.

A code segment to represent the concepts implementation is shown in Fig. 10.

B. *The Properties*

OWL properties represent relationships. Properties link individuals from the domain to individuals from the range. At the detailed level object properties link an individual to an individual. Datatype properties link an individual to data values. In other words, they describe relationships between an individual and data values.

The Object Properties

The object properties, in essence, are the relationships among concepts. Table 2 shows the core object properties. Each object property may have a corresponding inverse property. If some property links individual 'x' to individual 'y' then its inverse property will link individual 'y' to individual 'x'.

There are two types of properties. One is defined property and the other is inferred property. The defined property is set by the user or is obtained directly through sensors. The inferred property is derived through the reasoning done by the inference engine on the rules written by the user.

A code segment of implementation of object properties is shown in Fig. 11.

```
....
<Declaration>
      <Class IRI="#Sen_levelof_damageH"/>
   </Declaration>
   <Declaration>
      <Class IRI="#Sen_levelof_damageP"/>
   </Declaration>
   <Declaration>
      <Class IRI="#Sen_levelof_damageT"/>
   </Declaration>
   <Declaration>
      <Class IRI="#Sen_locationof_attacker"/>
   </Declaration>
   <Declaration>
      <Class IRI="#Sen_locationof_attackerH"/>
   </Declaration>
   <Declaration>
      <Class IRI="#Sen_locationof_attackerP"/>
   </Declaration>
.....
```

Fig. 10 Implementation of concept

Table 2 Object properties of person ontology

Property	Description
Locatedin	This is about the linking of concept person with a concept named location where a person is located currently
hasRole	It is the linking of a concept person to the role that he can take
hasTrustlevel	It is how much trust a person can have towards using any resource
engagedin	It is the depiction of activities in which a concept person is involved in
hasAddress	It specifies the address from where the person belongs to

C. *Data Properties*

The data properties used in the ontology specify the range of permissible values for the given property. The values may be in the range of integer, string, boolean, etc. It links the concept with a value of some particular type (Table 3).

D. *The Individuals*

Individual instances are the most specific concepts represented in a knowledge base. An individual gives the value of the members of the class. The description is as follows. The individuals of heartRate sensor are hs1, hs2, hs3, hs4, hs5, hs6, etc. shown in Fig. 12. Similarly the individual of role are doctor, nurse, visitor, patient and insurance agent etc.

```
.....<ObjectPropertyDomain>
        <ObjectProperty IRI="#hasPresenceofAttacker"/>
        <Class abbreviatedIRI="untitled-ontology-300:feasibleS"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#hasPresenceofAttackerH"/>
        <Class IRI="#feasibleH"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#hasPresenceofAttackerP"/>
        <Class IRI="#feasibleP"/>
    </ObjectPropertyDomain>
.......
```

Fig. 11 Implementation of object properties

Table 3 Data properties of a concept person

Data property	Description
hasAccessMedicaldata	It shows whether a concept person is given access to read medical data of a patient or not
hasUserName	It specifies the concept person with a name as string
hasPassword	It specifies the concept person with password as string

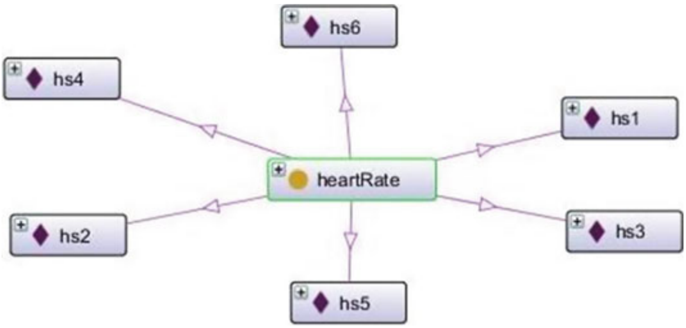


Fig. 12 Individuals of concept heartRate Sensor

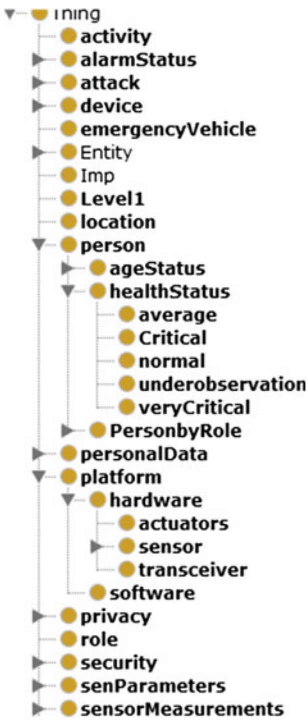
A possible implementation for individuals is given in Fig. 13.

We then identified 170 classes during the conceptualization process. Our system employs an OWL-DL ontology to represent the various types of classes and their relationships. Figure 14 shows parts of our ontology, which is engineered using the ontology editor Protégé (version 5) software package.

```
.....
<Class IRI="#energyT"/>
   </ObjectPropertyRange>
   <ObjectPropertyRange>
      <ObjectProperty IRI="#hasFullAccessDevice"/>
      <Class IRI="#device"/>
   </ObjectPropertyRange>
   <ObjectPropertyRange>
      <ObjectProperty IRI="#hasLimitedAccessDevice"/>
      <Class IRI="#device"/>
   </ObjectPropertyRange>
   <ObjectPropertyRange>
      <ObjectProperty IRI="#hasName"/>
      <Class IRI="#name"/>
   </ObjectPropertyRange>
   <ObjectPropertyRange>
      <ObjectProperty IRI="#hasPresenceofAttacker"/>
      <Class IRI="#Sen_locationof_attacker"/>
   </ObjectPropertyRange>
....
```

Fig. 13 Code segment of individual

Fig. 14 Class hierarchy



E. *Rules*

We have created various policies using Semantic Web Rule Language (SWRL). An SWRL rule consists of an antecedent and a consequent. Conditions in the consequent hold whenever the conditions in the antecedent are true. The interface for defining rules to infer new knowledge is shown in Fig. 15.

We have implemented a prototype for testing by defining various rules to infer whether a patient health condition is critical, veryCritical etc. and then setting various alarm controls. We used the HermiT reasoning engine to infer high-level situations.

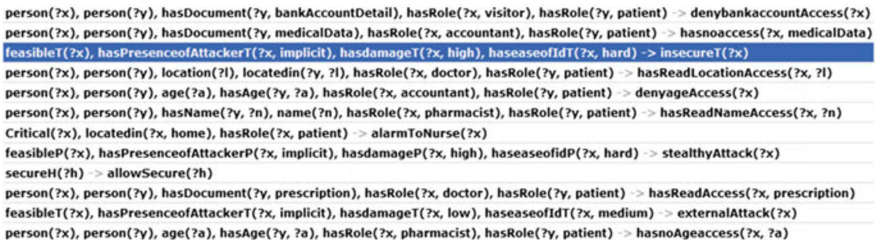


Fig. 15 SWRL interface for inference

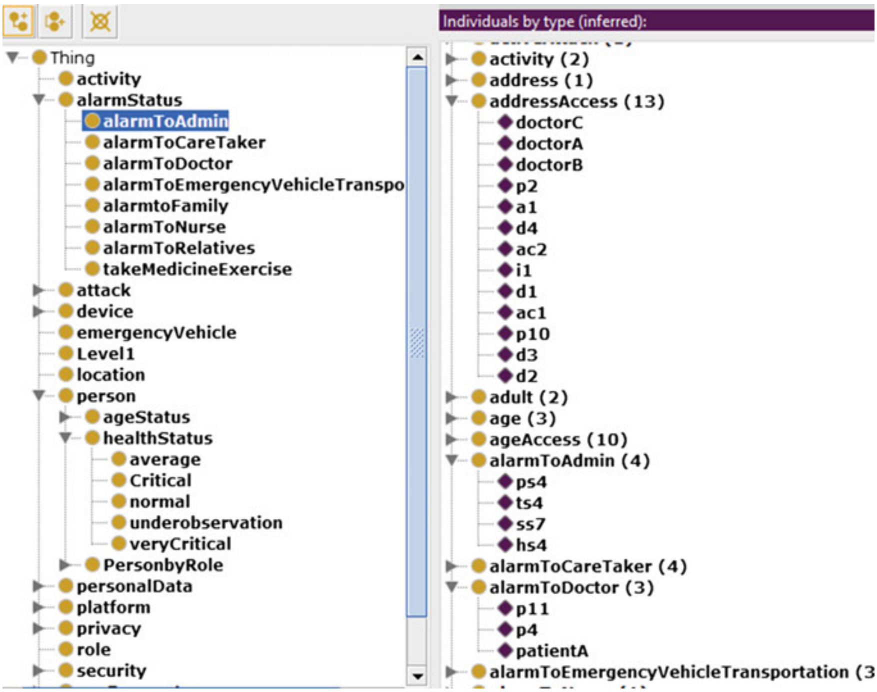


Fig. 16 Inferred concept through policies

Some higher level rules are written to infer contextual information such as what arrangements should be done in case of an emergency. The various security and privacy policies are also specified. The disclosure of patients' data is done according to roles. The permission to access a specific device is according to some predefined rules. Similarly all types of attack conditions in the sensor nodes are specified. All the nodes which are insecure are identified and the alarms are being sent to the system administrator for appropriate actions. The interface for inferring the knowledge as a result of applying various policies is shown in Fig. 16.

5 Conclusion

This study represents the context ontology for the health care. To provide better healthcare services to a patient in the secure manner without wasting time is the main focus of this paper.

We have built an ontology and a set of rules on the basis of a scenario in which a patient is critical and requires an immediate treatment. The graphical representation of the class hierarchy along with its properties, individuals as well as security and privacy policies is represented by the model. A prototype is simulated to test the environment. The future scope of the study plans to propose an architecture for a trust-based policies in the healthcare environment.

References

1. Dey, A. and Abowd, G. (2000), "Towards a Better Understanding of Context and Context-Awareness", Workshop on the what, who, where, when and how of context-awareness at CHI 2000.
2. L. Chen, C. Nugent, M. Mulvenna, D. Finlay and X. Hong (2009), "Semantic Smart Homes: Towards Knowledge Rich Assisted Living Environments, Intelligent Patient Management", pp. 279–296.
3. Huq, M. R., Tuyen, N. T. T., Lee, Y. K., Jeong, B. S., & Lee, S. (2007). "Modeling an Ontology for Managing Contexts in Smart Meeting Space". In *SWWS,* pp. 96–102.
4. Her, Yun, Su-Kyoung Kim, and YoungTaek Jin (2010), "A Context-Aware Framework using Ontology for Smart Phone Platform." *JDCTA* 4(5), pp. 159–167.
5. Mohan, P., Singh, M. (2013), "Formal Models for Context Aware Computing", International Journal of Computer Applications Technology and Research 2(1), pp. 53–58.
6. Chen, H. (2004), " An Intelligent Broker Architecture for Pervasive Context-Aware Systems". PhD Thesis.
7. Dey, A.K. (2000), "Providing Architectural Support for Building Context-Aware Applications", PhD Thesis.
8. Christopoulou, E., Goumopoulos, C., & Kameas, A. (2005). "An ontology-based context management and reasoning process for UbiComp applications". In Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies (pp. 265–270). ACM.

9. Chen, H., Finin, T., & Joshi, A. (2003). “An ontology for context-aware pervasive computing environments”. The Knowledge Engineering Review,18(03), pp. 197–207.
10. Wang, X. H., Zhang, D. Q., Gu, T., & Pung, H. K. (2004). “Ontology based context modeling and reasoning using OWL”. In Pervasive Computing and Communications Workshops. Proceedings of the Second IEEE Annual Conference, pp. 18–22.
11. Xu, Nan, et al. (2013), “CACOnt: a ontology-based model for context modeling and reasoning.” Applied Mechanics and Materials. Vol. 347, pp. 2304–2310.
12. Jang, Changbok, and Euiin Choi. (2011), “Context model based on ontology in mobile cloud computing.” Advanced Communication and Networking. Springer Berlin Heidelberg, pp. 146–151.
13. S., et al. (2010), “An ontology based context modeling approach for mobile touring and navigation system.” Proceedings of the The 2010 Canadian Geomatics Conference and Symposium of Commission I, ISPRS Convergence in Geomatics–Shaping Canada’s Competitive Landscape, Calgary, Canada, pp. 15–18.

Inter Cluster Balanced LEACH Algorithm for Energy Efficient Wireless Sensor Network

Sahul Goyal and Bindiya Jain

Abstract In today's world, wireless sensor network has been main area of research for different stakeholders. In the past, many routing techniques have been proposed to increase the active period of the Wireless Sensor Networks and cluster-based routing technique has been widely used to attain objective. In the current study, an attempt has been made to contribute in the field of WSN by implementing new Balanced Cluster LEACH (IBLEACH Protocol). The study includes the division of the clustering process into two phases, i.e., intra cluster and inter cluster, which will further help to increase the energy efficiency of the system. The results reveal that the use of IBLEACH protocol makes the network more energy efficient.

Keywords Wireless sensor network · Clusters · Smart clustering technique SCT · Intra cluster phase · Inter cluster phase

1 Introduction

Wireless sensor networks (WSN) comprise of several sensor nodes deployed randomly in a field having limited memory, processing capacity and power mounted on one board which cannot be recharged or replaced as shown in Fig. 1. Therefore, to extend the network lifetime is the major challenge in sensor networks. Recently many researchers have implemented many routing techniques with one common challenge, i.e., save the energy to enhance the life of the network.

S. Goyal (✉)
Computer Science Department, DAV Institute of Engineering & Technology, Jalandhar, India
e-mail: er.sahul.goyal@gmail.com

B. Jain
Electronics & Communication Department, DAV Institute of Engineering & Technology, Jalandhar, India
e-mail: bindiyajain29@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_46

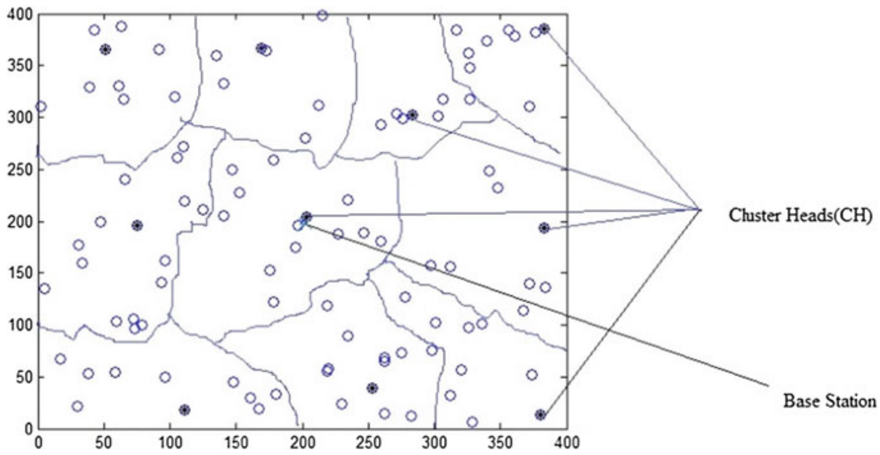


Fig. 1 Cluster based wireless sensor network

In WSN, one of the major challenges is to save the energy or to enhance the efficiency. To achieve this, paradigm is shifted from a single-level hierarchy to multi-level hierarchy and clustering is the best way to implement multi-level hierarchy in the network. A clustering technique forms 2-level hierarchy. In level 1, some member nodes close to each other form a group called clusters. In cluster, all the member nodes sense the data and forward to the one particular node called Cluster Head (CH). In level 2, cluster heads aggregates the data forward to the Base Station (BS). BS can be located within the field or at another place; [1]. The base station can be a sink node or a personal computer equipped with better power supply which is used for data storage. The major concern of clustering technique is a selection of cluster head. Clustering is an NP-hard problem to minimize the total energy dissipation in a network. For the N number of sensor nodes in WSN, there are 2^n-1 different combinations to elect the sensor node as CH or non-CH in each solution for the WSN [2].

In a wireless sensor network, there are many routing protocols based upon the principle of SI which is governed by the behavior of a flock of birds called particle swarm intelligence (PSO) that helps to select Cluster Heads (CHs). PSO developed by Kennedy and Eberhert (1995) are population-based search algorithm helps to solve real- world optimization problems by simplifying the assumptions. In WSNs, generally PSO optimizes the number of CHs based on number of nodes. In WSN, Groups of random particles initialize PSO, which explore optimal solution by updating generations. Every node is being updated with the maximum values calculated by mathematical formula called as the fitness value of node, best solution of node is known as pbest and best value in the whole swarm population referred as gbest. This gbest value of particles use to predict the number of clusters in the network and fitness value of particles are used to find CHs among clusters.

Low energy adaptive clustering hierarchy (LEACH) based on clustering technique is the most accepted algorithm in WSN. In LEACH, it performs the task in two phases called Advertisement Phase Cluster and Setup Phase. Advertisement phase of WSN uses Eq. 1 to find the cluster head from the Cluster and in setup phase every node decides his cluster head.

$$T(n) = \begin{cases} \frac{P}{1-P*\left(rmod\frac{1}{P}\right)} & \text{If n € G otherwise} \\ 0 & \end{cases} \tag{1}$$

CHs use energy very quickly and depends upon the distance from BS while associated nodes in cluster put away their energy slowly. So, LEACH assumes uniform energy consumption for CHs [3–5]. Most of research work in LEACH is how to elect Cluster Heads (CHs) to balance the energy consumption of the wireless sensor network. IBLEACH balances the energy consumption by consistently finding the optimum node for CH in cluster with the various parameters of the nodes and neighbor CHs during the advertisement phase of the round.

The rest of paper organized as follows: Sect. 2 covers the brief reviews of previous work to enhance the efficiency of LEACH protocol Sect. 3 covers the problem formulation and we introduce the approach to proposed problem. The simulation results of proposed protocol presented in Sect. 4 and in Sect. 5 we conclude our work.

2 Related Work

Routing in WSNs are different from other communication and wireless ad hoc networks [6] due to several characteristics that distinguish them like low power, limited memory, etc. In WSNs low-power signal processing architectures, energy efficient wireless routing protocols are areas of broad research in recent years. [7] LEACH is one of the first hierarchical routing approaches in wireless sensor networks that improve lifetime and energy efficiency of networks. Key features of LEACH are: (1) random rotation of the CH (2) local compression (3) localized coordination and inspiration for many clusters based routing protocols. In PEGASIS [8] protocol sensor nodes forms chain to pass information to neighbor. PEGASIS use a token system to pass the data to BS. Token passing through all the nodes collect data and designated node transmits to the BS. The disadvantage of PEGASIS is that there is necessity of knowledge about the complete path of the network. In Hybrid Energy Efficient Distributed (HEED) clustering approach cluster head selection depends upon residual energy and degree of the node. Achieves uniform distribution of cluster head [9].

Multihop-LEACH protocol CH uses other nodes called Cluster Nodes (CNs) as a communication station to transmit data to BS [10]. M-LEACH protocol change communication mode from a single hop to multi-hop between CHs and BS. The main disadvantage of M-LEACH is CHs as relay nodes consume more energy.

LEACH-DE is an improved version of LEACH protocol that is based upon Differential Evolution (DE) routing algorithm. It changes the criteria to select the CHs in WSN and divides setup phase into two sub phases, i.e. cluster-head and formation of cluster phase to reduce the communication distance or Convergence. LEACH-C [11, 12] (LEACH centralized) protocol another version of LEACH protocol uses centralized algorithm which is executed at the Base Station in place of distributed algorithm. Base Station selects CHs and their associated nodes. The biggest disadvantage of LEACH-C is the dependence of energy consumption depends on the location of Base Station (BS).

Centralized-PSO a variant of PSO algorithm, In Centralized-PSO nodes having energy above average energy are elected as the cluster heads [13]. In another variant of the PSO minimum spanning tree-PSO, Minimum Spanning Tree-PSO creates minimum spanning tree of nodes to elect CHs and associated nodes. Election of CHs are based on the energy available and Euclidean distance to its neighbor nodes. Selection of optimized routes between the nodes and its Cluster Heads on the basis of energy consumption [14]. A distributed PSO algorithm for wireless sensor networks considers hidden nodes and asymmetric links. It minimizes the number of hidden nodes and asymmetric links by increasing the transmit power of these nodes, thus increase the lifespan of the sensor network [15]. PSO-SSM (PSO with supervisor student mode) is energy aware clustering in WSNs. PSO-SSM is applicable only in those WSNs where each node has fixed omnidirectional transmission range, the sensor field should be mapped into a 2-Dimensional space. After deployment positions of the nodes are known to the base station [16].

Hybrid GA-PSO clustering algorithm using the principles of Genetic Algorithm (GA) and PSO to improve the lifetime of WSN. Hybrid GA-PSO use GA to choose the CHs and PSO to select the member nodes [17]. ELEACH [18] protocol the authors propose that residual energy is the main pattern while forming the CHs in Wireless Sensor Network. In LEACH-TLCH (LEACH Protocol with Two Levels Cluster Head) requirement of associated CH depends upon the distance and residual energy if energy of CH or Distance of CH is less than average energy and distance then cluster needs associated node. [19] MODLEACH [20] enhancement in LEACH sets up two threshold energy levels to improve the lifetime of WSN. Once, node Declared as CH will remain CH until its energy level reaches the threshold level. Difference between other LEACH protocols, and proposed Inter Cluster Balanced LEACH (IBLEACH) [21] are (1) The WSNs average energy they calculate is the average energy of all the alive nodes, but in Proposed IBLEACH average energy calculates only within the cluster, to increase the lifetime of the network. (2) Balanced clustering has done by taking threshold distance between two CHs. (3) Changing the CHs after every round result into increase the energy consumption, but in proposed IBLEACH CHs changes after a certain threshold.

3 System Model and Problem Description

A wireless sensor network consists of n sensor nodes along with one sink node with unique ids spread randomly over an area. In LEACH protocol sensor nodes grouped into clusters and sink node located at optimal location ($0.5 \times X_m$, $0.5 \times Y_m$) which collects all data from sensor nodes. Each round of LEACH protocol divides further into two phases shown in Fig. 2.

In set up phase, selection of Cluster Head depends upon desired percentage *(p)* (node has not become a cluster head from *p* rounds) nodes with the desired percentage *(p)* greater than the threshold limit selected as CH. Selected CHs broadcast massage in setup phase as an invitation to join them and associated nodes joins the cluster head having minimum distance.

In steady phase, each round divides into slots of equal time durations where in each time slot CH receives data from associated nodes and sends to sink node of Wireless sensor network.

Radio Model:

Radio model shows energy consumed by LEACH to transfer L-bits from cluster node to sink node (Fig. 3).

The total energy consumed by Cluster Node (E_{CN}) is

$$
\begin{aligned}
&E_{\mathrm{CN}} = E_{Tx} + E_{CHsel}\\
&E_{CHsel}(L,d) = (E_{\mathrm{elec}}L + \varepsilon L d^{n}) + R_{ais} * E_{\mathrm{elec}} + P_s * E_{\mathrm{elec}}\\
&E_{Tx}(L,d) = E_{\mathrm{elec}}L + \varepsilon L d^{n} \quad (\textit{if } D > d \text{ then } n = 4 \text{ otherwise } n = 2)
\end{aligned}
$$

The total energy consumed by Cluster Head (E_{CH}) is

$$
\begin{aligned}
&E_{\mathrm{CH}} = E_{CHsel} + E_{CHAD} + E_{CHcom} + E_{CHf}\\
&E_{CHsel}(d) = E_{elec} + \varepsilon d^{n}\\
&E_{CHAD} = p * n * E_{\mathrm{elec}}\\
&E_{CHcom}(L,d) = \left(N_f * E_{\mathrm{elec}}\right) + \left(E_{\mathrm{elec}}L + \varepsilon L d^{n}_{\mathrm{tosink}}\right) \quad (\textit{if } D > d \text{ then } n = 4 \text{ otherwise } n = 2).
\end{aligned}
$$

Energy gaps in different variants of LEACH protocol change cluster head after each round. This leads to wastage of energy after every subsequent round during the advertisement phase. Cluster heads must have some threshold distance between each other.

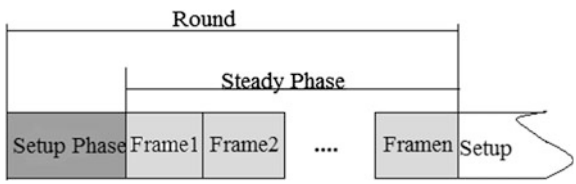


Fig. 2 Round of LEACH operation

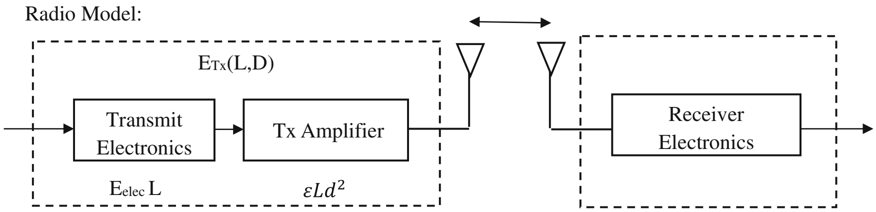


Fig. 3 Radio model

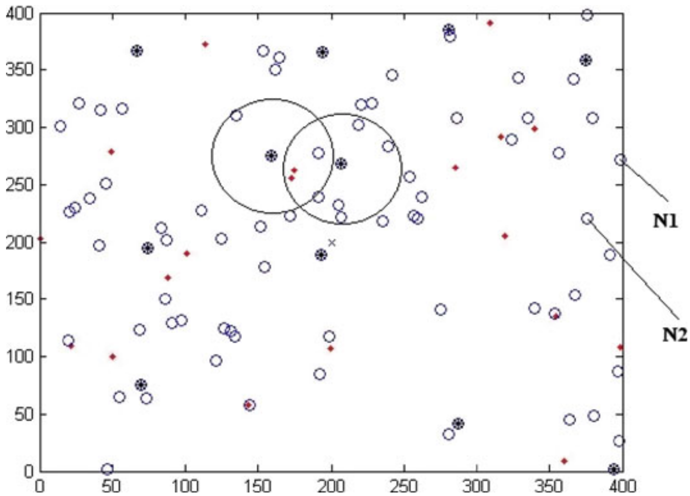


Fig. 4 LEACH protocol

4 Proposed Algorithm

The biggest disadvantage of LEACH protocol is number of cluster heads and position of cluster head is not optimal and it changes cluster head after every round. CHs should change randomly and is a basic property of LEACH, but it is unnecessary to change cluster head after every round and nonuniform distribution of cluster heads. Figure 4 shows that there is no cluster head near node N_1 and N_2, so it consumes more energy to transfer data to base station.

The Major drawback of implementing PSO in Wireless sensor network is non-localization of nodes. PSO treats all nodes of WSN as a whole, but LEACH divides all nodes into distinct groups called as clusters. CHs distant from base station tends to lose more energy than near CHs. So, in proposed solution, it also considers all the parameters to find the CH change in energy after every round. The algorithm proposed all the nodes have a face value calculates from the equation which is based upon parameters like distance, residual energy, total energy

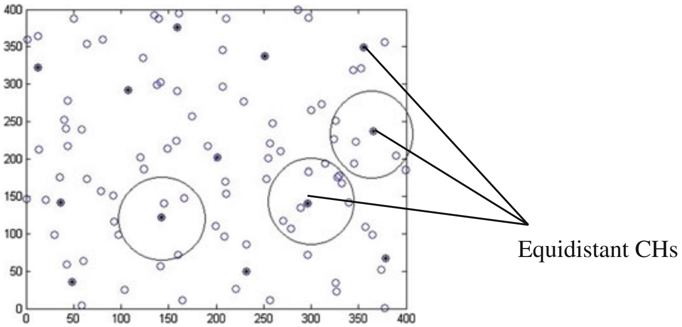


Fig. 5 Cluster pattern of IBLEACH protocol

remaining in network etc. The node with highest face value assigns as a first CH. Later all CHs are chosen based upon criteria, i.e. distance should be greater than the threshold distance and face value should be higher than remaining nodes this cycle goes on till the network is alive. Advantage of proposed algorithm is that all CHs has minimum threshold distance from each other and these are uniform in nature to remove residual nodes from the network as shown in Fig. 5.

Proposed algorithm for new routing protocol

Algorithm 1: To find the number of cluster Heads Required in WSN.
1. Calculate the number of alive nodes in Network. *2. For every alive node in WSNs Repeat steps3 & 4.* *3. If the number of clusters are equal to zero:* *I.a find the node with maximum face value and assign that as Cluster Head.* *4. If Distance of node is* ***greater than or equal to*** *threshold Distance and energy of nodes is* ***greater than*** *the threshold energy level, assign that node as a* ***Cluster heads and execute steps 5 & 6.*** *5. Compare Cluster head with all the neighboring nodes.* *6. If associated value of node is greater than associated value of Cluster head and distance is less then threshold limit then change the Cluster head with given node.*

Algorithm 2 : To find the optimalnode for Cluster Head in Cluster
1. Calculate the average energy of neighbor for alive nodes in network. *2. If energy of Cluster heads is* ***less then*** *certain limit calculate with average energy parameter. execute steps 3 & 4.* *3. Compare Cluster head with all the neighbor nodes.* *4. If associated value of node is greater than associated value of Cluster head and distance is less then threshold limit then change the Cluster head with given node.*

In PSO main criteria of selection is based upon how it calculates the face value of nodes in earlier it is assumed that all the parameter effects linearly on the face value of the nodes but in practically it gives best result when these effects exponentially on facevalue of nodes in our proposed formula of facevalue it effects the number of alive nodes in the neighborhood, distance from base station of node and number of rounds that node become doesn't act as cluster head and remaining energy of node effecting exponentially on the facevalue these parameters are effecting facevalue by following way.

$$S\,(i).fvalue = (N + (D_{\max} - S(i).d + S(i).P)^{\wedge(S(i).E - E_{\text{avg}})}$$

where:

$S(i)$.fvalue:	face value of node
N:	Number of Alive nodes in neighbor/number of alive nodes
$D_{\text{MAX:}}$	Maximum distance between base station to node
$S(i).d$:	Distance between node and base station
$S(i).P$:	number of rounds that node become non cluster head
$S(i).E$:	remaining energy of node
$E_{\text{avg:}}$	average energy of neighbor nodes

Results and Analysis:

Simulation of IBLEACH was performed in MATLAB(r2009a) with parameters:

Parameters	Value
Area	400 × 400
Number of nodes	100 Nodes
Energy	1 J
Node distribution	Random distribution
Base station location	200 × 200

Simulation results show that IBLEACH perform much better than MODLEACH in terms of throughput and network lifetime.

Figure 6 shows that IBLEACH perform better from MODLEACH and basic LEACH protocols by balancing the cluster using various parameters like distance, energy, alive neighbor nodes. As shown in Table 1, IBLEACH is ahead from MODLEACH and LEACH in terms of network lifetime. Its 1st node dies at 129th round, but in other variants it dies at 46th and 66th respectively. Similarly, in proposed algorithm 90th nodes dies at 1417th round, whereas in LEACH and MODLEACH last node dies at 1017 & 987.

Throughput is another important metric to judge the efficiency of WSN. Throughput means to know how much information will be generated by the Wireless Sensor Network. As shown in fig. 7 If a base station receives more packets from cluster heads that means more efficient protocol. Table 2 shows that in IBLEACH protocol BS receives 10,909 packets from CHs as compare to MODLEACH packets to BS are 6959. In basic LEACH packets to BS decreased to

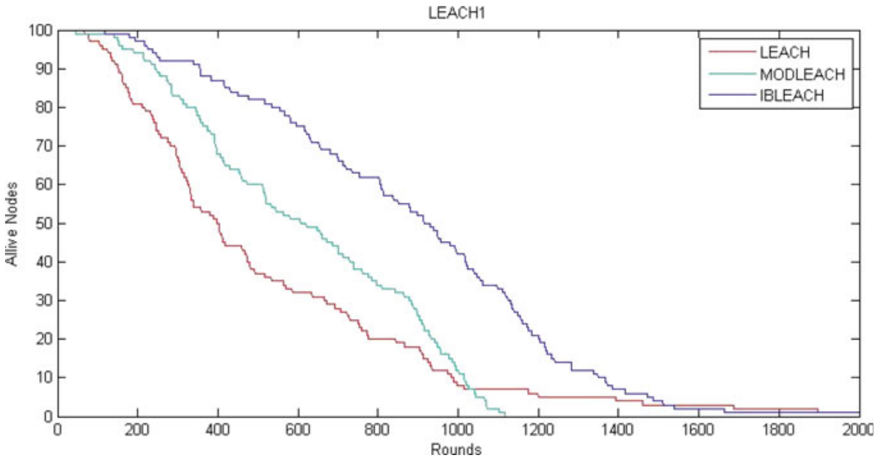


Fig. 6 Number of nodes alive

Table 1 Number of nodes alive

No. of dead nodes	1st node	50	90	100
Rounds (IBLEACH)	129	942	1417	Up to 2000
Rounds (MOD LEACH)	46	400	987	1116
Rounds LEACH	66	620	1017	1976

5234. It is observed that IBLEACH protocol performs much better than MODLEACH and basic LEACH protocol.

Number of CHs: Fig. 8 shows the number of cluster heads during each round of wireless sensor networks number cluster heads also indicates nodes alive in the network. Other LEACH protocols spent most of his lifetime with one or two CHs means less number of nodes and less information about the field, but IBLEACH spent most of his lifetime with more than 10 CHs means more information. this pattern also shown by average energy of IBLEACH protocol.

The average energy of alive nodes indicates remaining energy in alive nodes Fig. 9 indicates that the average energy of nodes in IBLEACH is far more than other protocols and it remains alive for more time than other protocols.

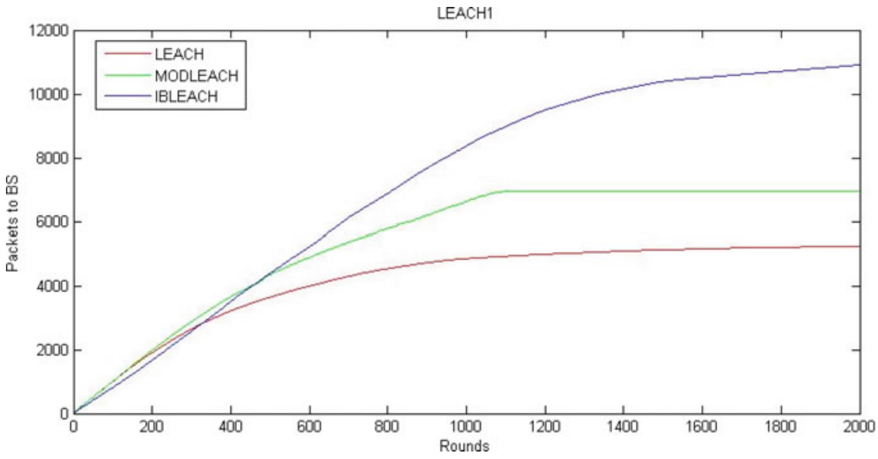
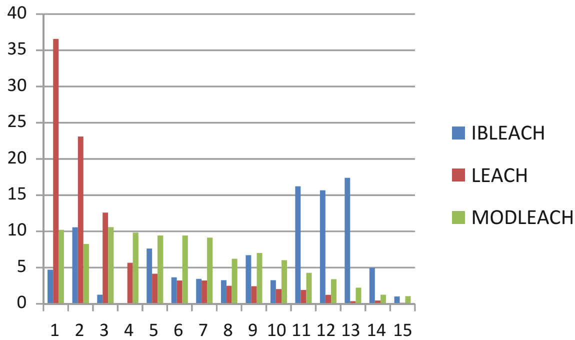


Fig. 7 Packets to BS

Table 2 Packets to base station

No. of rounds	500	1000	1500	Final
Packet to BS IBLEACH	4340	8230	10,388	10,909
Packet to BS MOD LEACH	4290	6613	6959	6959
Packet to BS LEACH	3646	4867	5152	5234

Fig. 8 Number of CHs



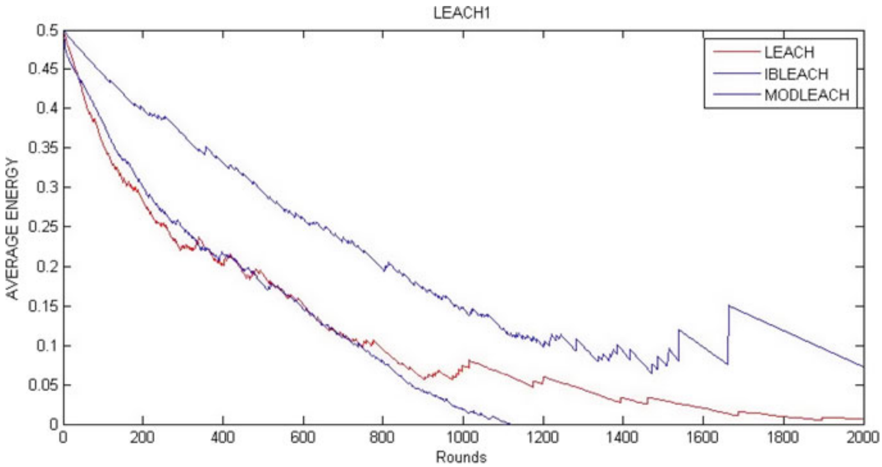


Fig. 9 Average energy of nodes

5 Conclusion

In this paper, we have proposed a routing protocol to improve LEACH by introducing IBLEACH based upon PSO. In the IBLEACH number of CHs depends upon the global environment and selection of CHs based on the local environment. IBLEACH balanced the energy consumption for WSNs by distributing the cluster evenly through the whole area, thus increasing the lifetime of the entire network. Our simulation results show that IBLEACH outperforms LEACH and MODLEACH and its improvements in terms of the network lifetime, energy consumption and throughput.

References

1. W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan(2000). Energy-Efficient CommMicrosensor Networks, Proceeding International Conference on System Sciences (pp. 1–10).
2. Bara'a A. Attea and Enan A. Khalil, "A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks," Applied Soft Computing, vol. 4, pp. 1950–1957, July, 2011.
3. Raghavendra V. Kulkarni, and Ganesh Kumar Venayagamoorthy, "Particle Swarm Optimization in Wireless-Sensor Networks: A Brief Survey‖ IEEE Trans. on Systems, Man, and Cybernetics—Part C: Applications and Reviews, vol. 41, no. 2, pp. 262–267, Mar. 2011.
4. Yang, H., & Sikdar, B. (2007). Optimal CH selection in the LEACH architecture. In IEEE international conference on performance, computing, and communications (pp. 93–100).

5. Latiff, N. M. A., Tsimenidis, C. C., & Sharif, B. S. (2007). Performance comparison of optimization algorithm for clustering in wireless sensor networks. In IEEE International conference on mobile adhoc and sensor systems (pp. 1–4).
6. C., Lee, S., Mitra, P., & Kumara, S. (2010). Distributed routing in wireless sensor networks using energy welfare metric. Information Sciences (pp. 1656–1670).
7. Fengjun Shang, Yang Lei (2010). An Energy-Balanced Clustering Routing Algorithm for Wireless Sensor Network Wireless Sensor Network (pp 777–783).
8. S Lindsey, CS. Raghavenda, (2002). PEGASIS: pow er efficient gathering in sensor information systems, in: Proceeding of the IEEE Aerospace Conference, Big Sky, Montana.
9. Ossama Younis, Sonia Fahmy(2004). HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks, IEEE Transactions On Mobile Computing (pp. 366–379).
10. X. N. Fan and Y. L. Song, (2007). Improvement on LEACH Protocol of Wireless Sensor Networks, Proceedings of 2007 International Conference on Sensor Technologies and Applications, Valencia, (pp. 260–264).
11. W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan (2002). An application-specific protocol architecture for wireless microsensor networks, in IEEE Transactions on Wireless Communications, pp. 660 – 670.
12. Padmalaya Nayak, Pallavi Shree (2014). Comparison of Routing Protocols in WSN using NetSim Simulator: LEACH Vs LEACH-C. International Journal of Computer Applications (pp 1–6)
13. N.M.A. Latiff, C.C. Tsimenidis, and B.S Sheriff(2008) "Energy-aware clustering for wireless sensor networks using particle swarm optimization". In IEEE International Conference on mobile radio communication, 1–5.
14. X. Co, H. Zhang, J. Shi, and G.Cui "Cluster heads election analysis for multi-hop wireless sensor networks based on weighted graph and particle swarm optimization". In IEEE fourth International Conference on computing, 7, 599–603.
15. Jason Tillett, Shanchieh Jay Yang, Raghuveer Rao and Ferat" Optimal topologies for wireless sensor networks"IEEE International Conference on Personal Wireless Communications.
16. S. Guru, S. Halgamuge and S. Fernando "Particle swarm optimizers for cluster formation in wireless sensor networks". In IEEE International Conference on sensor, 319–324.
17. total number of sensor nodes in WSN is N, a sensor node is either elected as CH or non CH in each solution so that there in "international conferance of Engineers and Computer Scientists, March 2011
18. Jia Xu, Ning Jin, Xizhong Lou, Ting Peng, Qian Zhou, Yanmin Chen (2012) 9th International Conference on Fuzzy Systems and Knowledge Discovery (pp 2174–2177)
19. Hongwei Chen, Chunhua Zhang, Xinlu Zong, Chunzhi Wang (2013) LEACH-G: an Optimal Cluster-heads Selection Algorithm based on LEACH, JOURNAL OF SOFTWARE (pp 2660–2667)
20. D. Mahmood, N. Javaid, S. Mahmood, S. Qureshi, A. M. Memon, T. Zaman (2013) MODLEACH: A Variant of LEACH for WSNs International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), (pp 158–163).
21. Ahmed Salim, Walid Osamy, Ahmed M. Khedr IBLEACH: intra-balanced LEACH protocol for wireless sensor networks (2014) Wireless Networks Springer (pp 1515–1525).

EPCA: Energy Preservation using Clustering Approximation in Sensor Network

A.P. Jyothi and Sakthivel Usha

Abstract At present, the existing clustering algorithm suffers from computationally expensive process resulting in less assurity of operational longevity with energy efficiency of sensor nodes in an adverse condition. The present paper introduces a novel technique called as EPCA i.e. Energy Preservation using Clustering Approximation in wireless sensor network. The technique provides two levels clustering optimization by providing multiple attributes for precise selection of clusterhead. EPCA provides less complex operation with faster response and better sustainance towards longer simulation rounds. The clustering approximation technique implemented shows superior energy preservation as compared to existing standard energy efficient protocols in wireless sensor network.

Keywords Network lifetime · Battery · Clustering · Optimization
Wireless sensor network · Optimization

1 Introduction

A wireless sensor network is globally used for collecting information from human inaccessible geographical location. Although, the concept of wireless sensor network has shown multi-dimensional application in theory but it really doesn't even exists in commercial markets. It is interesting as there are massive research papers on wireless sensor network since last two decades, but it has never met the utility of common people very largely. The prime reason behind this is a sensor node still suffers from energy problems that are the root cause of majority of other problems e.g. routing [1, 2], bandwidth [3], security [4]. From the trends of existing research, one thing is very much clear that clustering is one of the most suitable factors that

A.P. Jyothi (✉)
Visvesvaraya Technological University, Belagavi, India
e-mail: jyothiarcotprashant@gmail.com

S. Usha
Department of CSE, RRCE, Bangalore, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_47

positively or negatively affects the network lifetime of wireless sensor network. At present, there are various energy efficient routing as well as clustering techniques introduced by various researchers till date [5]. However, all teh clustering technique seriously lags inclusion of cost effective optimization (although there are various studies focusing on iterative and computationally expensive optimization theory [6, 7] in wireless sensor network). Hence, this paper discusses about a unique optimization technique called as clustering approximation that contributes to enhancing the network lifetime by providing multiple attributes for selecting clusterhead. Section 2 discusses about the prior research work carried out in the area of energy efficient clustering in wireless sensor network followed by Sect. 3 that briefs about the problems identification. The proposed contribution is discussed in Sect. 4 followed by algorithm implementation in Sect. 5. The result analysis is done in Sect. 6 followed by summary of the paper in Sect. 7 as conclusion.

2 Related Work

This section briefs about the studies being carried out towards addressing energy problems in wireless sensor network. Our prior study has reviewed existing techniques for addressing power issues [8] while this section updates the contribution of some recent research work.

Most recently, usage of Voronoi diagram for energy efficiency was witnessed in the work of Gautam et al. [9]. The network lifetime was found to significantly improve after applying ant framework in this study. Different from conventional clustering techniques in sensor network, Yu et al. [10] presented an uneven clustering for energy preservation on clusters with unequal number of nodes. Clustering technique with usage of Hausdorff distance-based grouping was introduced by Zhu et al. [11]. Unfortunately, the outcomes of such works are never found to be benchmarked. Udompongsuk et al. [12] have presented a study that performs selection of clusterhead based on moving average, a statistical-based technique. The outcome of the study was found to suffer from scalability issue and its tradeoff with energy efficiency. Pei et al. [13] have presented a study using conventional LEACH algorithm investigated over cognitive radio in wireless sensor network. The technique finds less usability over time-based applications in wireless sensor network. Various optimization technique based on swarm intelligence were also found to have better impact on minimizing energy depletion e.g. dolphin swarm optimization, elephant swarm optimization [14], bacteria foraging algorithm [15], BAT algorithm [16]. All the above mentioned studies are focused on homogenous networks dedicated to address energy problems. Studies towards clustering issues pertaining to heterogeneous network were carried by Meenakshi [17] and Patil [18]. Although, all the above mentioned techniques gives a better guidelines for future research, but all of them are potentially associated with issues. Brief discussion of such issues in made in next section.

3 Problem Identification

The past section has discussed about different energy effective procedure to enhance the system lifetime and a percentage of the standard clustering techniques, which has received a wide acknowledgment in past. In any case, the greatest problem with every system is pretty much the same. Various studies are found to consider the position of the base station at the centroid of simulation area, which prompts higher degree of routing overhead. The second issue in existing studies is selection criteria of clusterhead on the basis on residual energy. However, few studies have truly adopted energy utilizing bio-inspired procedures, yet complexities of such algorithms concerning handling time and memory utilization is missing from the latest research contribution. The third issue in that the recent studies are found with less novel features. Larger part of an existing solution is done on just single level optimization approach, whereas, there is a reasonable probability of performing optimization utilizing different levels considering different parameters included in data aggregation process of remote sensor network. The fourth problem explored in the existing studies is consideration of uncertain knowledge about the communication behaviour of the sensors. Thus, the issue articulation of the proposed study can be characterized as – "*It is a computationally difficult to build up a multi-level clustering for upgrading the system lifetime in wireless sensor applications*". The next section presents a novel method that addresses the issues identified (Fig. 1).

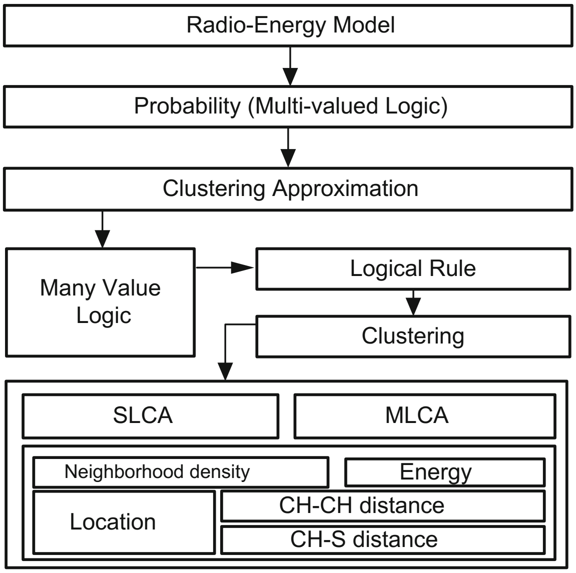


Fig. 1 Architecture of proposed schema of EPCA

4 Proposed System

The prime purpose of the proposed system is to develop a novel clustering technique which is not only simple but yet robust in its design principle in terms of scalability and energy efficiency for wireless sensor network. Figure 2 shows architecture of Energy Preservation with Clustering Approximation (EPCA). The design of proposed system starts by implementing first order radio energy model, which states definite values of energy parameters e.g. energy required for transmittance, receiving, amplification, data aggregation etc. The core intention of the proposed system is to implement a multiple valued logic that has better governance over the decision making for selection of clustering. Hence, the proposed system performs clustering using dual approximation techniques called as SLCA or Single Level Clustering Approximation and MLCA or Multiple Level Clustering Approximation. The SLCA process considers neighborhood density, energy and location of the sensors for choosing it as candidate sensors. However, confirmation of candidate sensors for being eligible to be actual clusterhead is carried out by MLCA that performs additional computation of distance between two clusterheads and distance between clusterhead and sink. Hence, it can be seen that proposed EPCA provides 5 new attributes for performing selection of the clusterhead in the wireless sensor network, which is one the novel and cost effective approach in the optimization principle.

An essential reason for implementing many-valued logic in design of EPCA is to perform proper management of uncertainties pertaining to energy optimization.

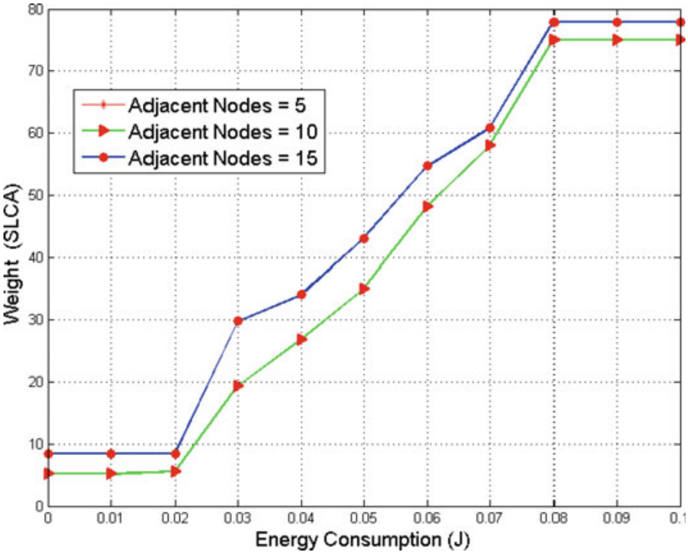


Fig. 2 Impact of adjacent sensors on weight of SLCA

A simple logical rule is developed using linguistics e.g. low, high, medium, very high, very medium etc., which has the better control over inference system towards actual selection of clusterhead. The EPCA technique takes the input of battery and neighborhood density. The technique also considers input parameter for probability that a sensor can become a cluster leader during consecutive cycle of data aggregation. As EPCA algorithm is more focused on clustering process, so it will also emphasize on the selection process of the core object, which is cluster leader. The system makes use of indicator function for designating high linguistics. The study performs the optimization of the energy using dual level of clustering. The first clustering level applies EPCA on battery power and neighborhood density, while the second clustering level subjects the first clustering level to position, clusterhead-sink node distance, and distance between two clusterheads. The development of the algorithm of the proposed system is done in analytical approach to ensure better evidential ground.

5 Algorithm Design

The proposed system is developed over Matlab considering the random distribution of sensors in the simulation area. The proposed system also allows the sink node to be location independent, which means it can be kept at any position within the range of simulation area. The system also considers a specific threshold for adjacent nodes along with 5% probability of clusterhead selection. These assumptions are intentionally done in order to perform comparative analysis with LEACH.

Algorithm Energy Preservation with Clustering Approximation (EPCA)

Input: η (Network), n (node), C (c| c is closest node to n, which is leader node), Th (threshold), i (total number of cluster), j (frequency of becoming leader), α (centroidal localization), β (fusion node-sink node distance), and γ (inter fusion node distance)

Output: Selection of cluster leader

Start

1. Initialize probability (n)$\leftarrow$FLAME (battery, neigh_dens)
2. If n $\neq$ leader
3. Than j$\leftarrow$0
4. If (j = |size(|η/i|)
5. isleader (n)$\leftarrow$false
6. Th$\leftarrow$1
7. Else Th$\leftarrow$|j/size(η)|
8. End if
9. if (arb(0,1)>Th) //arbitrary function
10. leader(n)$\leftarrow$n
11. probability (n) $\leftarrow$FLAME (α, β, γ)
12. transmit (probability (n), C) to candidate
13. For each cϵC
14. if (probability(n)<probability(C))
15. leader (n)$\leftarrow$c;
16. isleader(n)$\leftarrow$false
17. transmit (AbortCH_elect_msg, C)
18. Else isleader(n)$\leftarrow$True
19. j$\leftarrow$j+1
20. End if
21. End if
22. If (isleader(n)=true)
23. transmit(leader_msg, C)
24. Else
25. forward joining request to nearest leader.

End

The proposed algorithm is designed considering the significant charecteristics of EPCA approach in order to overwrite the clustering technique proposed in LEACH algorithm. The prime goal of the algorithm is to enhance the network lifetime using dual level of optimization. The next section discusses about the outcomes and performance analysis of the proposed study with respect to conventional LEACH algorithm.

6 Result Analysis

This section highlights the significant findings of the proposed EPCA. From application viewpoint, the result analysis is done by implementing algorithm considering Size of packet header to be 20 bytes, simulation time as 1000 s, size of transmitted packet to be 15 bytes, size of data packet to be 1000 bytes, and transmission radius of 20 m. The network model of algorithm implementation considers evaluating with 50–500 numbers of sensors with initialized energy of 0.5 J. The outcome of the study is compared with conventional LEACH algorithm and hence similar simulation parameter is retained for comparative analysis.

Figure 2 highlights the impact of adjacent nodes on weight of Single Level Clustering Approximation (SLCA). The term weight will mean a matrix consisting of elements of SLCA i.e. position, residual energy, and adjacent nodes. Hence, increasing value of weight will also mean higher probability of node to be chosen as cluster head based on these three elements. This technique is very different from existing technique where clusterhead is selected mainly from residual energy. This outcome also a state that the proposed system has higher longevity and is also scalable as increasing number of sensors positively assists in weight factor of SLCA. However, better trend is increasing weight is evident only till 90% of energy consumption, after that it shows linear properties, which means static value of weights as nodes starts dying after consuming more than 90% of energy.

Figure 3 showcases the impact of different energy factors on the weight of Multi-Level Clustering Approximation (MLCA). The outcome shows higher the residual energy more is the weight of MLCA with increasing number of adjacent sensors availability. The proposed system shows higher scalability feature by showing that increasing inclusion of sensors tends to increase the weight of MLCA that results in appropriate selection of a clusterhead. In this analysis, it is also found that residual energy is not the only factor to control the selection process as seen in many existing clustering techniques. In face a moderate (0.05 J) of residual energy is more than enough to ensure better longevity of the sensors for long terms in wireless sensor network.

Figure 4 shows the influence of two different distance factor on the weight of MLCA essentially to show the impact of coverage on weight over increasing simulation rounds. Increasing distance among the sensors will also mean lowering of coverage and connectivity over longer simulation rounds. The outcome shows that increasing distance can ensure optimal weight value for MLCA. The trend of

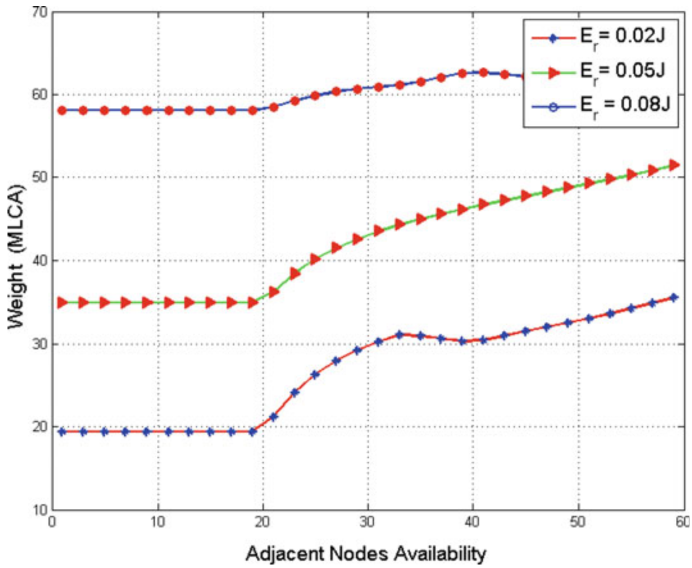


Fig. 3 Impact of residual energy on weight of MLCA

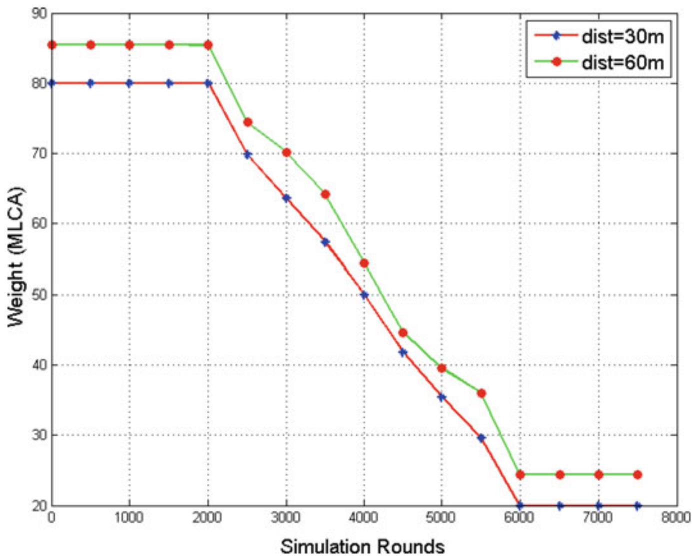


Fig. 4 Impact of distance on weight of SLCA

curve is downwards as in increasing round, node death occurs resulting in spontaneously increase in distance among the nodes, which finally lowers the weight of MLCA. Another significant outcome is that the degradation is very rigorous towards 6,000th simulation rounds onwards showing better longevity of the network lifetime in wireless sensor network.

Figure 5 exhibits outcomes of proposed system with respect to network lifetime. Interestingly, the discussion of network lifetime is done with respect to increasing clusterhead and sink distance. The outcome shows that LEACH cannot sustain long enough simulation rounds. However, EPCA has been found with better retention of alive nodes. The processing time of the proposed system for 1,000 simulation rounds is 2.44 s while that of conventional LEACH algorithm is found with 9.55 s. Therefore, there is a significant reduction of computational time for the proposed system. Owing to the dual level of optimization due to novel clustering algorithm of EPCA, the proposed system performs selection of leader using multiple attribute instead of residual energy factor (like LEACH). Another significant contribution is LEACH performs communication considering sink at the center of the simulation, whereas the proposed system doesn't have any dependency on the location of the sink. At any position of the sink, the proposed systems is capable for performing selection of the cluster leader and perform data aggregation. Moreover, the supportability of multi-level routing is quite high in proposed system, which is not present in LEACH. Hence, a simple modification of LEACH features can highly increase the network lifetime. Moreover, EPCA algorithm is incorporated with LEACH to overcome the clustering defects of LEACH. Hence, a selection criterion of cluster leader is changed to suitably enable the process of energy-efficient clustering. This process significantly optimizes energy and network lifetime.

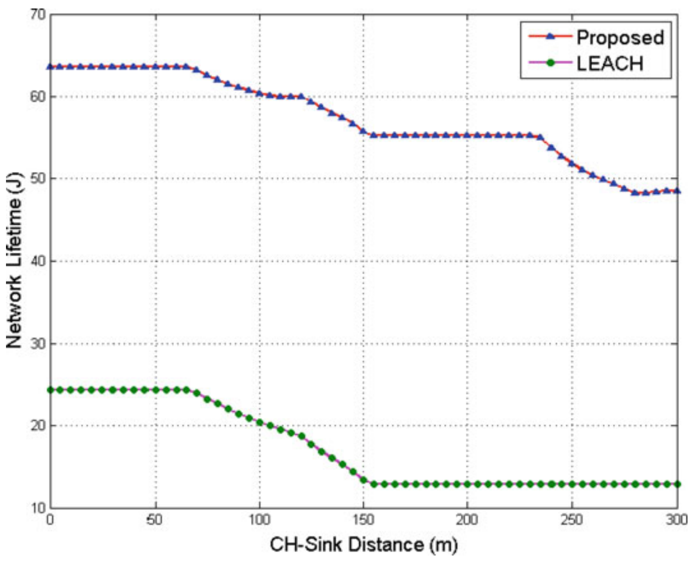


Fig. 5 Network lifetime of proposed system and LEACH

7 Conclusion

Energy depletion during communication of wireless sensor network is still a challenging problem to solve. After reviewing the existing system, it was found that clustering is one of the best alternative solutions to address such issue. However, existing techniques of clustering requires more improvement in its design principle. This paper has discussed about a single and multi level clustering approximation techniques which not only enhances the clustering scheme but also positively impacts on the maximum retention of residual energy of sensor nodes. The prime innovativeness of the study is its simple optimization technique which differs from existing clustering approaches with respect to lower iterative rounds, new conditions of selection of clusterhead, ensuring participation of all nodes during data aggregation process, etc. The study introduces a dual optimization level where selection of cluster leader is carried out considering neighborhood density, centroidal localization, fusion node-sink node distance, and inter fusion node distance. The outcome of the study was found to be highly energy efficient in contrast to conventional LEACH algorithm. The processing time performance is also quite fair enough to think about its application in resource hungry applications of sensor network. The outcome of the study shows its direct applicability on any sensors with 48 kb of physical memory.

References

1. S. P. Singh, S.C. Sharma, "A Survey on Cluster Based Routing Protocols in Wireless Sensor Networks", *Elsevier-International Conference on Advanced Computing Technologies and Applications,* vol.45, pp. 687–695, 2015.
2. M. Aslam, N. Javaid, A. Rahim, U. Nazir, A. Bibi and Z. A. Khan, "Survey of Extended LEACH-Based Clustering Routing Protocols for Wireless Sensor Networks," *IEEE International Conference on Embedded Software and Systems, High Performance Computing and Communication,* Liverpool, 2012, pp. 1232–1238, 2012.
3. M. Chitnis, P. Pagano, G. Lipari, Y. Liang, "A survey on Bandwidth resource Allocation and Scheduling in wireless sensor networks", *International Conference on Network-Based Information Systems,* 2009.
4. M. Kumar, K. Dutta, "A Survey of Security Concerns in Various Data Aggregation Techniques in Wireless Sensor Networks", *Springer Journal of Intelligent Computing, Communication and Devices,* vol. 309, 2014.
5. S. P. Singh, K. Bhanot, S. Sharma, "Critical Analysis of Clustering Algorithms for Wireless Sensor Networks", *Springer Journal of Proceedings of Fifth International Conference on Soft Computing for Problem Solving,* vol. 436, 2016.
6. M. M. Fouada, N. E. Oweisb, T. Gaberb, M. Ahmedd, V. Snaselb, "Data Mining and Fusion Techniques for WSNs as a Source of the Big Data", *Elsevier-International Conference on Communication Management and Information Technology,* vol. 65, pp. 778–786, 2015.
7. S. Babaie, T. Branch, A. Khadem, "A New Method for Improving Life Time in Wireless Sensor Network by Using Genetic Algorithm", *Springer Journal of Instrumentation, Measurement, Circuits and Systems,* vol. 127, pp. 177–184, 2012.

8. A.P. Jyothi and U. Sakthivel. “Trends and Technologies Used for Mitigating Energy Efficiency Issues in Wireless Sensor Network”, *International Journal of Computer Application*s, vol. 111, No.3, pp. 32–40, February 2015.
9. N. Gautam, S. Sofat, and R. Vig, “An Ant Voronoi Based Clustering Approach for Wireless Sensor Networks”, *Springer Journals, Social Informatics and Telecommunications,* 2014.
10. J. Yu, Y. QI, G. Wang, “An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks”, *Springer Journal of Control Theory Application,* 2011.
11. X. Zhu, L. Shen, and T-S Peter Yum, “Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks”, *IEEE Transactions On Vehicular Technology,* vol. 58, no. 2, February 2009.
12. K. Udompongsuk, C. S. In, C. Phaudphut,, “MAP: An Optimized Energy-Efficient Cluster Header Selection Technique for Wireless Sensor Networks”, *Springer Journal, Advances in Computer Science and Its Applications, Lecture Notes in Electrical Engineering,* 2014.
13. E. Pei, H. Han, Z. Sun, B. Shen and T. Zhang, “LEAUCH: low-energy adaptive uneven clustering hierarchy for cognitive radio sensor network”, Springer- EURASIP Journal onWireless Communications and Networking, 2015.
14. Bharathi M.A, B.P. Vijayakumar, D.H. Manjaiah, Cluster Based Data Aggregation in WSN Using Swarm Optimization Technique, *International Journal of Engineering and Innovative Technology,* Vol. 2, Iss.12, June 2013.
15. M.A. Bharathi, M. Mallikarjuna, B.P. Vijaya Kumar, “Bio-Inspired Approach for Energy Utilization in Wireless Sensor Networks”, *Elsevier-International Conference On Modelling Optimization And Computing,* Vol. 38, pp. 3864–3868, 2012.
16. K. Seelam, M. Sailaja and T. Madhu, “An Improved BAT-Optimized Cluster-Based Routing for Wireless Sensor Networks”, *Springer Journal, Intelligent Computing and Applications, Advances in Intelligent Systems and Computing,* 2015.
17. D. Meenakshi and S. Kumar, “Energy Efficient Hierarchical Clustering Routing Protocol for Wireless Sensor Networks”, *Springer Journal, Social Informatics and Telecommunications Engineering,* pp. 409–420, 2012.
18. P. R. Patil and U. P. Kulkarni, “Energy-Efficient Cluster-Based Aggregation Protocol for Heterogeneous Wireless Sensor Networks”, *Springer-Journal, Intelligent Computing, Networking, and Informatics, Advances in Intelligent Systems and Computing,* 2014.

Federating Disjoint Segments in Wireless Sensor Networks Using Jumping Particle Swarm Optimization

Ramya Sharma and Virender Ranga

Abstract Wireless Sensor Networks are created by placement of sensor nodes, which are small, cheap devices placed for sensing the environment and performing simple calculations. As the transmission range and battery lifetime of these devices is very low, extra relay nodes are placed in the network to maintain long distance connectivity. Due to external interferences and harsh environmental conditions, the network may become disconnected. Restoring lost connectivity is crucial for further operation of the network. This problem is shown to be an NP-hard, thus we pursue meta-heuristics to solve relay node placement problem. In this paper, we propose a meta-heuristic technique, Network Restoration through Jumping Particle Swarm Optimization (NR-JPSO), which can be used efficiently for restoring the lost connectivity. However, Particle Swarm Optimization (PSO) considers many solutions for a particular problem (called particles) and modifies their position to give an optimal solution. Our suggested approach first finds the representative node for each disconnected segment, then creates Steiner points for reconnection. These Steiner points are further used to create random spanning trees, which are used as particles in NR-JPSO to provide an optimal interconnected network.

Keywords Relay node placement · Steiner points · Spanning trees · JPSO

1 Introduction

In the previous years, the applications of wireless sensor networks (WSNs) have faced an uprising in their numbers [1]. For applications like space exploration, forest fire detection, combat field reconnaissance, and machine health monitoring,

R. Sharma (✉) · V. Ranga
Department of Computer Engineering, National Institute of Technology, Kurukshetra, India
e-mail: sharma.ramya268@gmail.com

V. Ranga
e-mail: virender.ranga@nitkkr.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_48

in order to observe a covered area and keeping certain activities in check, a number of sensor nodes are placed in the network. By positioning sensors to control unattended surroundings, the danger to human life may be avoided and the environment will continue to be monitored. These applications use sensor nodes (SNs) which have small battery life, with limited processing and communication capabilities. After deployment, the sensor nodes set up a network with the objective of data sharing and synchronizing the actions performed. To facilitate such collaboration, the said nodes must be accessible to each other. Long distance communication for sensor nodes will be costly and would exhaust them very quickly (as energy transmission is proportional to the distance between the two nodes). Thus, relay nodes (RNs) are introduced, which are used to transmit the sensed data from sensor nodes to a base station (BS). This problem of introducing minimum relay nodes in the setting so that the entire network is interconnected is a NP-hard problem and is called relay node placement (RNP) problem. This RNP problem takes two separate architectures into account, one tier WSN and two tier WSN. In one tiered WSN, both sensor nodes and relay nodes contribute in the routing. While in two tiered WSN, only relay nodes are used in routing procedure. The limitation of a vast network is that the wireless devices may fail, leaving the WSN disconnected. Due to harsh and unpredictable environmental conditions like landsides, floods or forest fires, the network faces disconnection, due to breaking of the network into smaller intraconnected segments. To restore the connectivity, a set of relay nodes can be positioned along the minimum cost path. The connectivity between every minimal cost path can be represented as a minimal cost tree.

A Steiner Minimum Tree (SMT) is a tree formed by interconnecting a given set of vertices with the help of some extra points (Steiner points) to gain low cost connectivity. In triangle geometry, the Steiner point is a specific point associated with a plane triangle. This point has the shortest distance from all the triangle vertices. Thus, Steiner points are used for connectivity of a graph. Creating a SMT is a NP-complete problem, thus we pursue meta-heuristics to tackle this problem. Jumping Particle Swarm Optimization (JPSO) is a discrete variation of Particle Swarm Optimization. We propose JPSO, Network Restoration through Jumping Particle Swarm Optimization (NR-JPSO), to solve the SMT problem to regain lost connectivity in the network. The structure of the paper is divided into five parts beginning with the introduction section. Related work is described in Sect. 2. Section 3 shows our contribution in this paper. Simulation results are shown in Sect. 4 and paper is concluded in Sect. 5.

2 Related Work

In the approach suggested by Senel et al. [2], the authors considered the problem of relay node placement which uses Spider web deployment strategy, placing relay nodes in the inward direction so that the network has better connectivity and coverage. Major shortcoming of this method is the slight increment in relay node count. The proposed approach of Lee et al. [3] for network fragmentation is an Optimized Relay node placement algorithm with help of a minimum Steiner tree on the Convex hull (ORC). ORC attempts to recognize Steiner points (SPs) which are used to populate relays such that the disconnected segments will be linked with very less number of relays. ORC positions RNs inwardly from the border of the area identified by the convex hull. The authors propose a distributed Cell-based Optimized Relay node Placement (CORP) in approach by Lee [4], whic h practise greedy heuristics and chooses to decrease the number of relays necessary for setting up a connected inter-segment topology. CORP models the area as a grid of equal-sized cells and identify the best neighbouring cell of a segment Seg_i which lay on the shortest path that joins Seg_i to the other segments, functioning in rounds. This method attempts to minimize the relay node count.

3 Proposed Solution

Our proposed solution is divided into various phases as described below:

(a) ***Phase I-Detection of the network fragmentation***
The network detects the disconnection when multiple relay nodes do not receive the transmitted data or beacon messages. These relay nodes send the beacon messages to all the other neighbouring nodes, thus discovering disconnection, if present. The relay nodes which have disconnected neighbours act as the boundary nodes in the segment. After discovering the disconnection, the reconnecting procedure is started, as shown in Phase II.

(b) ***Phase II- Reconnection***
This phase has four sub-phase as described below:

- *Selecting a Representative node for each segment*

For reconnection, every segment must be represented by some node which further processes the reconnection phase. This node is called the representative node of a segment. Thus, for an *ith* segment Seg_i, the representative node rep_rn_{ij} is selected from a set of $boundary\_rn_{ij}$, where i stands for ith segment and j stands for jth boundary node (a node which has experienced disconnection of its links with other nodes).

For selecting the representative node, first we find out the centre of mass of each segment. Then, for every segment we find that boundary relay node which is closest to all the segments' centre of mass, i.e. which has the value $\sum_{k \neq j} \mathrm{dist}\left(boundary\_rn_{ij}, \mathrm{com}_k\right)$ minimum for that segment, where *dist()* calculates distance between two nodes.

After finding out the representative nodes (stored in rep_rn_{ij} *for* ith segment), we proceed to find out the Steiner points for further calculation.

- *Finding out the Steiner points*

In triangle geometry, a Steiner point (SP) is described as an extra point introduced, which when connected to the tree vertices of the triangle incurs minimum cost, i.e. the distance between the SP and the vertices is minimum. Thus, for the reconnection of the segments, we use Steiner points.

In our algorithm, Steiner Point of a triangle is calculated to be the circumcenter of the triangle if the triangle is right angled or acute angled triangle (i.e., the circumcenter lies within the triangle or on the boundary of the triangle). Whereas, when the triangle is obtuse angled, the Steiner point is calculated as the Centre of Mass of the triangle. We randomly select three representative nodes from the network and find the Steiner point of the said triangle. We observe $(number\ of\ segments\ -1)$ Steiner points(*Spoint)* in our algorithm. As these Steiner points are potential relay nodes location, they are stored for later use. The circumcenter of a triangle is calculated in Eq. 1.

$$x = \frac{((m_{pAB} \times \mathrm{mid}_{AB}.x) - \mathrm{mid}_{AB}.y) - ((m_{pAC} \times \mathrm{mid}_{AC}.x) - \mathrm{mid}_{AC}.y))}{(m_{pAB} - m_{pAC})} \tag{1}$$

where

m_{pAB}	is the slope of AB
m_{pAC}	is the slope of AC
$\mathrm{mid}_{AB}.x$	is the x coordinate of mid-point of line AB
$\mathrm{mid}_{AB}.y$	is the y coordinate of mid-point of line AB
$\mathrm{mid}_{AC}.x$	is the x coordinate of mid-point of line AC
$\mathrm{mid}_{AC}.y$	is the y coordinate of mid-point of line AC

- *Constructing a fully Connected Graph and finding out the spanning trees*

To ensure connectivity in the network, we construct a fully connected graph, consisting both representative relay nodes as well as Steiner nodes. Out of this fully connected graph, a set of n random spanning trees is generated. We can choose n accordingly, as it represents the swarm size. Here, n is considered to be 100.

Our aim is to generate a Steiner minimum Tree. Thus, we use a modified version of Particle Swarm Optimization to find the best spanning tree.

- *Implementing the Jumping Particle Swarm Optimization algorithms*

Originally, the PSO algorithm was introduced for continuous optimization problems. To deal with combinatorial optimization problems, a variation of PSO called Discrete Particle Swarm Optimization (DPSO) was proposed in [5], in which particles explore a multi-dimensional discrete exploration space. After introduction of DPSO, many alternatives of this algorithm have been suggested. A certain DPSO approach called Jumping Particle Swarm Optimization (JPSO) algorithm has lately been presented by Moreno-Perez et al. [6] to solve combinatorial optimization problems. In JPSO algorithm, as proposed by Qu et al. [7], instead of defining the velocity of every swarm particle to change its position in the search, the particle moves (jumps) from position to position (which are solutions) in discrete search space. If a particle has good fitness in certain area, it attracts the other particles to explore the area. This particle is called an attractor.

In our proposed algorithm, we have selected three kinds of attractors:

l_best, which is the best position of the current particle till now,
g_best, which is the best position among all particles till now, and
c_best, which is the best position among all particles in this iteration.

$$v_{i,j+1} = c_0 v_{i,j} + c_1 r_1 \left(l\_best - x_{i,j}\right) + c_2 r_2 \left(g\_best - x_{i,j}\right) + c_3 r_3 \left(c\_best - x_{i,j}\right) \quad (2)$$

Equation (2) presents the original PSO equation, which can be divided into two parts, self-improvement, which means exploration of current position, and the second part which encourages the particle to explore the space with the help of attractors. Thus, a particle follows the attractors for improvement.

Every attractor has some likelihood c_x, which is the probability of selecting a particular attractor or self-improvement of the particle. To explore the space, a random number r, which has the range [0,1] is generated. This r along with likelihood c_x determines the particle move.

- If $r \in [0, c_0)$ the particle implements self-improvement.
- If $r \in [c_0, c_0 + c_1)$ the particle explores locally, i.e. *l_best* attractor is chosen for improvement.
- If $r \in [c_0 + c_1, c_0 + c_1 + c_2)$ the particle explores globally, i.e. *g_best* attractor is chosen for improvement.
- If $r \in [c_0 + c_1 + c_2, c_0 + c_1 + c_2 + c_3 = 1]$ the particle explores current particles, i.e. c*_best* attractor is chosen for improvement.

Selecting one of these four moves, the particle *jumps* across the search space for optimal solution. The space exploration of NR-JPSO is explained in the previous segment. In the self-improvement jump, the particle selects a Steiner node from the tree which has degree 1 or 2, and removes it. In case the degree is 1, the node is just consuming space as is a burden on the tree. If the degree is 2, then the respective connecting two nodes are joined together and Steiner node is removed (on the basis of the Euclidean distance). In the attractor oriented jump, the minimum cost path from attractor is selected and replaced in the current particle. If the path already exists then no changes are done. If a representative node lies in between the path then the path up to representative node is conserved. After the updation of the current particle, *l_best* and *g_best* are checked and updated. The entire procedure is shown in Fig. 1. Finally, the NR-JPSO after meeting the stopping criterion gives *g_best* as the final output, as shown in the Fig. 2.

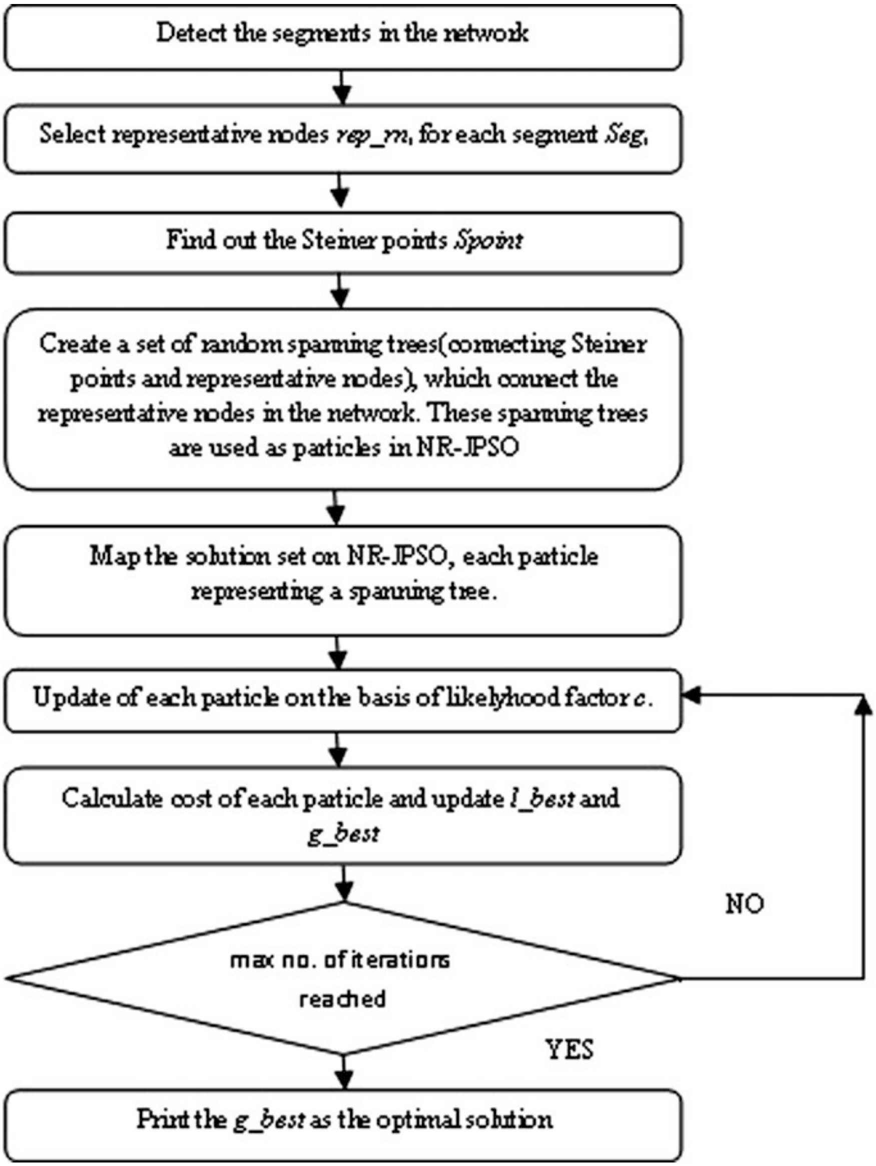


Fig. 1 Flow chart of NR-JPSO

```
Input: A particle swarm of PART number of randomly generated spanning trees span[PART], total number of nodes n
Output: An optimal spanning tree
Initialize all the attractors l_best[PART], g_bestand c_best
while(the stopping criteria is not met)
        for( each spanning tree span[i])
                generate a random number r between [0,1]      /*r is a random number which is used for defining the
                                                              likelihood of a given particle to improve itself*/

                if(r >=0 and r <= 0.24)                       //self improvement
                        self_improvement(span[i]);
                else if(r >= 0.25 and r <= 0.49)              //improvement with the help of local best attractor
                        path_replacement(span[i], lbest[i])
                else if(r >= 0.50 and r <= 0.74)              //improvement with the help of global best attractor
                        path_replacement(span[i], gbest)
                else if(r >= 0.75 and r <= 0.99)              //improvement with the help of current best attractor
                        c_best = find_current_best(span)
                        path_replacement(span[i], c_best)
                if(cost(l_best[i]) > cost(span[i]))
                        l_best[i] = span[i]
                if(cost(g_best[i]) > cost(span[i]))
                        g_best[i] = span[i]
        end for
end while
returng_best
```

self_improvement*(current_particle)* a procedure that takes the spanning tree as an input and removes a redundant Steiner node from the tree
path_replacement(*current_particle, attractor)* a procedure that take the current spanning tree and an attractor as the input, finds out the minimum cost path in the attractor tree and replaces that path in the current tree
cost*(particle)* a procedure which takes a spanning tree as an input and returns its cost

Fig. 2 Network restoration through jumping particle swarm optimization

4 Results Discussion

The comparison of the cost is done between MST created over representative nodes of each segment and the Global Best Steiner tree generated by NR-JPSO. To show this, we have assumed 6 segments and their respective representative nodes. A Minimum Spanning Tree is generated (as shown in Fig. 3) using Prim's algorithm.

After the introduction of Steiner nodes, for NR-JPSO, random spanning trees are produced for the swarm particles, one such tree shown in Fig. 4. Also, a Minimum spanning tree is created, using Prim's algorithm, to compare the random spanning tree and Minimum Steiner Tree (shown in Fig. 5). Finally, the Minimum Steiner Tree in Fig. 6 shows that the extra nodes are removed after processing.

We have taken the results with varying number of segments and having different topologies. The output of NR-JPSO is compared with the corresponding Minimum Spanning Tree, which is obtained through Prim's algorithm. The cost factor is defined by the Euclidian distance between the nodes. The size of the swarm is 100 and the number of iterations is also 100. The comparable results are shown in

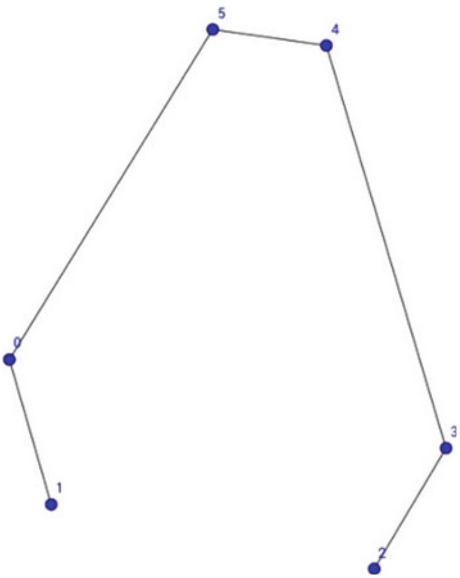


Fig. 3 MST generated by 6 segmented networks (0–5 are representative nodes of respective segments)

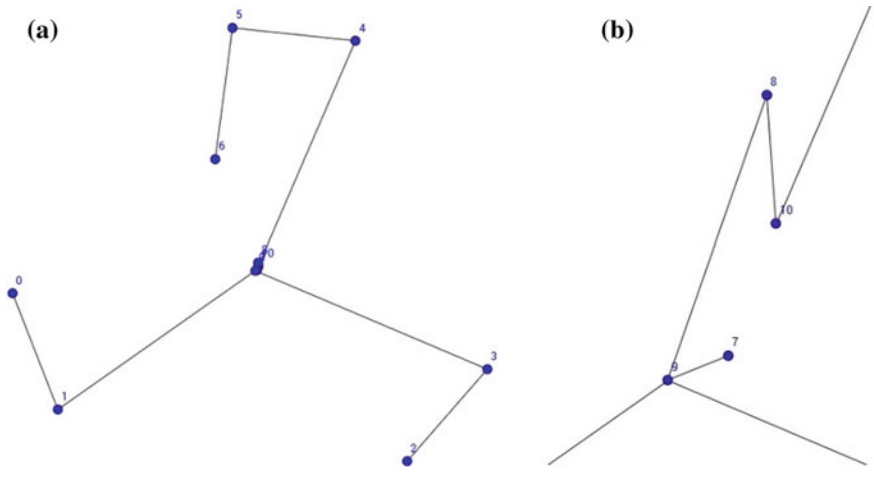


Fig. 4 **a** A random spanning tree generated after introducing 5 Steiner nodes (Zoomed Out) [6–10 are Steiner nodes], **b** A random spanning tree generated after introducing 5 Steiner nodes (Zoomed In) [6–10 are Steiner nodes]

Fig. 7. The results show that the NR-JPSO algorithm even after introducing extra nodes, produces lesser cost.

The Steiner Minimum tree generated is then populated by the relay nodes, by placing relay nodes after distance |R| is covered, where |R| is the transmission range of a relay node.

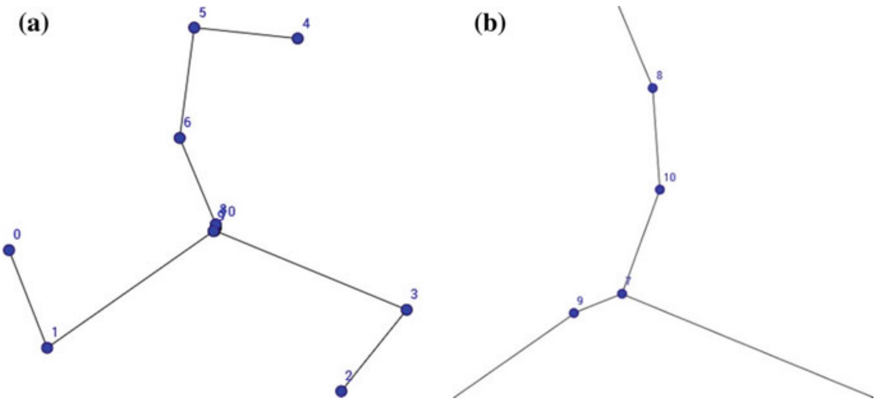


Fig. 5 **a** Minimum spanning tree generated after introducing 5 Steiner nodes (Zoomed Out) [6–10 are Steiner nodes], **b** Minimum spanning tree generated after introducing 5 Steiner nodes (Zoomed In) [6–10 are Steiner nodes]

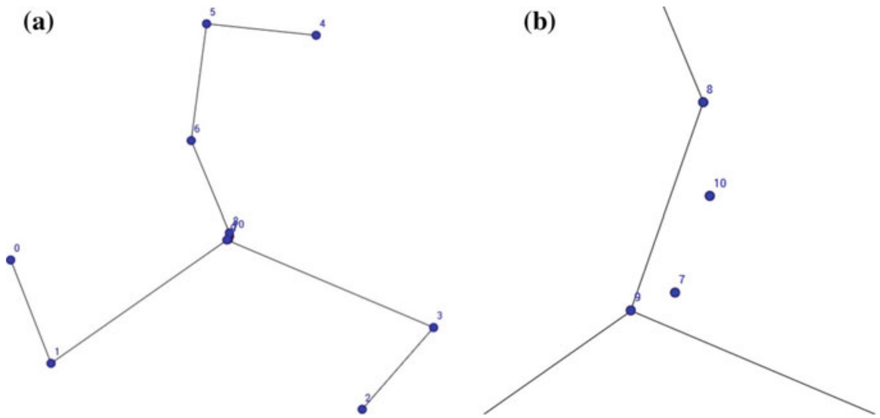
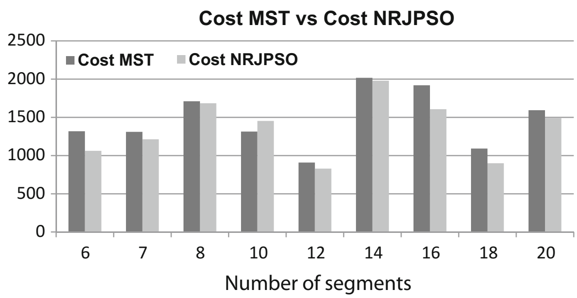


Fig. 6 **a** Global best generated after NRJPSO is implementation (Zoomed Out) [Nodes 7 and 10 are extra, hence are removed from the tree]. **b** Global best generated after NRJPSO is implementation. (Zoomed In) [Nodes 7 and 10 are extra, hence are removed from the tree]

Fig. 7 Cost versus # segments



5 Conclusion and Future Scope

In this paper, we present a technique, Network Restoration through Jumping Particle Swarm Optimization (NR-JPSO). This methodology makes use of Steiner nodes and Steiner Minimum trees and solves a NP-hard problem in polynomial time. It uses a discrete variant of Particle Swarm Optimization, Jumping Particle Swarm Optimization to generate a swarm and produce the best output, i.e. least cost Steiner Minimum Tree, by removing the surplus Steiner nodes and improving the particle's fitness function (cost).

As the results show, the cost of tree generated NR-JPSO is comparatively less than that of MST. The biggest advantage of using NR-JPSO is that it is topology independent, while the previous work in this area may fail after following heuristic techniques, NR-JPSO will provide a better result irrespective of the placement of segments.

Future Scope in this area can be improvement of the time complexity for NR-JPSO. Also, this work can be further extended for placing relay nodes or sensor nodes while establishing a new Wireless Sensor Network.

References

1. Ranga, V., Dave, M., Verma, A.K. Network Partitioning Recovery Mechanisms in WSANs: a Survey. *Wireless Personal Communications*, 72(2), 2013, pp. 857–917.
2. Senel F., Younis M., Akkaya K. Bio-inspired relay node placement heuristics for repairing damaged wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 60 (4), 2011, pp. 1835–1848.
3. Lee, S., & Younis, M. Optimized relay node placement for connecting disjoint wireless sensor networks. *Computer Networks*, 56, 2012, pp. 2278–2804.
4. Lee S., Younis M. Optimized relay placement to federate segments in wireless sensor network. *IEEE Journal on Selected Area in Communications, Special Issue on Mission Critical Networking*, 28 (5), 2010, pp. 742–752.
5. Kennedy J., Eberhart R.C. A discrete binary version of the particle swarm algorithm. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics,* 1997, Piscataway, NJ, pp. 4104–4109.
6. Consoli S., Moreno-Perez J.A., Darby-Dowman K., Mladenovic N. Discrete particle swarm optimization for the minimum labelling Steiner tree problem. *Natural Computing*, 9(1), 2010, pp. 29–46.
7. R. Qu, Y. Xu, J. P. Castro and D. Landa-Silva, Particle swarm optimization for the Steiner tree in graph and delay-constrained multicast routing problems. *J. Heuristics*, 19(2), 2013, pp. 317–342.

Part IV
Applications

Enhancing Security of Stature System by Using Sybil Detection

Neha Chauhan, Ankita Thadani and Jalpa Baria

Abstract Stature is the basically the term used for the reputation of the person on the network. People create fake identities and Sybil accounts are persistent in today's online world. We propose an algorithm based on the user trust to detect Sybil nodes in the Peer nodes, it is a distributed approach. We first introduce a basic user trust verification matrix which recompense for the weaknesses of the basic scheme. Finally, we have carried out simulations to explore the feasibility.

Keywords Reputation System · Stature System · Trust · Weighting Approach Security · Cryptography

1 Introduction

Peer-to-peer network permits the peers to collaborate users for service with no need for any centralized infrastructure. The lack of central authority over other entities on the network prone to Sybil attack [1]. A particular user can register under different random numbers, and sidestep any check which endeavors to assure the quality of the data through the plurality of contributors. The initial scheme for the Sybil is Advogato [2], intended to diminish exploitation in online services, it uses graphical approach of Distributed Hash Tables [3] to bound the probability for the routing interruption in those systems.

N. Chauhan (✉) · J. Baria
Department of CSE, Sadar Vallabhbhai patel Institute of Technology, Vasad 388306, Gujarat, India
e-mail: neha1260@gmail.com

J. Baria
e-mail: baria.jalpa@gmail.com

A. Thadani
Department of CSE, GTU, Ahmedabad, India
e-mail: ankitathadani13@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_49

Every entity can join the network and on registering a unique identity is issued by the system or by a TTP (trusted third-party). Considering the case for, Facebook [4] users can join further 9 million websites and online services by just validating that new resisted websites with their existing accounts. Each entity is anticipated to have only solitary identity and is generally anticipated to exploit this identity while intermingling with other users in the system. But considering the Sybil attack the invader counterfeit several identities, where each identity created by the entity can be coined as Sybil, which is joined by the attacker to aim for various adversarial objectives.

These annotations inspire for exploration of a way to detect Sybil accounts. The invader is creating supplementary "individual" Sybil accounts, fooling intellectual entity. Suspicious user can submit an application perception to distinguish even minute discrepancy. the majority online communities have method.

2 Definitions

> Peer trust is exacting altitude of the individual in a context likelihood by having it an entity evaluate various entities will act upon a exacting way in that context, in cooperation previous to keep an eye on the way in a perspective relates to ones own action [5].

Considering the trust in an unknown entity by the trust recommendation and propagation Guha et al. [6] identified propagations as following:

Direct Propagation As the names suggest the direct trust. Consider *I* trusts *j*, *j* trusts *k*, then we can conclude that *i* trusts *k*.

$i = j,\ j = k,\ i = k$

Co-Citation i_1 trusts j_1 and j_2, and i_2 trusts j_2. Under co-citation, it is concluded that i_2 also trusts j_1.

i1 → j1 then we can infer i2 → j1
i2 → j2

Transpose Trust Implies that *I* trusts *j* then there *j* can develop some level of trust towards *i*. We can infer *i* trusts *j*, then transpose trust implies that *j* should also trust *i*.

i.
i → j then j → i

Trust Coupling *i* and *j* both trust *k*, then trust coupling leads us to infer that *i* and *j* should trust each other since they both trust *k*.

i → k then i → j
j → k

Let *q* be an agent be a querying agent to receive the trust recommendation from an agent as lat where a is the source agent and *t* a target agent about whom the trust is being transposition. Considering after this recommendation consequence of the trust, agent *q*

$$P\ (\text{perform}\ (q,\ t,\ \psi) = \text{true}) = \text{lat}.$$

Then this act of q establishing trust in t as a consequence of a trust recommendation from **a** is said to be a simple trust propagation from **a** to q about t.

The newly established trust value P (perform $(q,\ t,\ \psi)$ = true) is said to be simple, propagated trust from **a** to q about t.

3 Background

Here is the concise general idea of present community network-based Sybil detection systems.

3.1 SybilGuard

It is decentralized organization for controlling the sybil attacks which get benefit of IP harvest and sybil attacks restricted by quantity and dimension of cluster [7].

Figure 1 gives the explanation of working of the Sybil guard algorithm.

Limitations: The main highlight of this algorithm is on the property of the entities in the network. (i) rapid incorporation, and (ii) malevolent users might produce numerous nodes and a small number of attack boundaries.

3.2 SybilLimit

SybilLimit espouses an analogous organization as SybilGuard. SybilGuard protocol permits to admit a large number of sybil nodes and it presuppose that public networks are fast mixing, that has by no means be unambiguous in actual situation [8].

There are two protocols, first is safe random itinerary procedure and a verification protocol, executes in the back end and preserves information which will be used by second protocol.

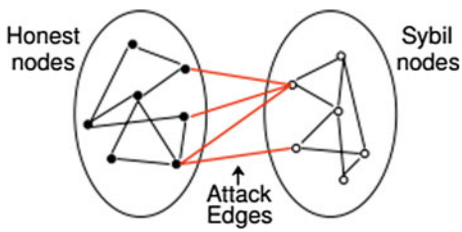


Fig. 1 Sybil guard working [1]

Grouping has:

(i) leveraging numerous instance of the random route to execute numerous small random routes,
(ii) utilize junction on edges instead of nodes.

3.3 SybilInfer

SybilInfer is an algorithm for categorization of nodes of honest entities. It is secure, Decentralized approach first such scheme is Advogato, which diminish exploitation in on-line services, followed by a proposal Distributed Hash Tables to for possible routing disruption it uses information available to the defender and is applied to peer-to-peer or distributed system [9].

3.4 GateKeeper

Gate keeper as the name suggests keep the eye on the gate in the decentralized way and makes an assumption that the network is randomly expanding day by day. So in real-life situations this algorithm suffers [10, 11].

3.5 SumUp

It is a voting service contains the covered rating values and they are connected to ith links where everyone can vote but the vote of the rating provider will only counted if and only if if there exists the trusted path between the owner and rating provider [12].

3.6 MobID

It is decentralized resistance for transportable devices hence it is called as MobID. Reduces the number of interactions with sybil attackers [13].

The collaborative devices get problem to:

1. run localization algorithms increasing the accuracy of location-based services
2. cache Web content to avoid economic costs

MobID guarantees that an honest individual accepts, and is accepted by, most other honest people with high probability (Table 1).

Table 1 Comparison of the literature survey

N	Systems		Advantages	Disadvantages	For
1	The beta reputation system [14]	C	Flexible and simple to implement	Immune to the entities which change their identity. Support binary value	e-commerce websites
2	Gossip trust [15]	D	Has speed looking scheme	Bloom filter makes it complicated	Decentralized peer networks with rank
3	DebitCredit system [16]	D	Short term misuse of reputation	Does not provides the security the system without receipt	Reward system
4	Coercion-free stature system [17]	C	Though centralized but still third party not required	No collusion occurs	Token related central system, Transaction system
5	Enhancing privacy preservation homomorphic system [18]	D	It provide more security	Works only for the trusted agent	P2p network, ranking systems
6	Sybil defenses [19]	D	Find out sybil defense mechanism	Two different kinds of algorithm	Social network, Distributed network
7	SybilDefender: [20]	C	Defense from the Sybil is done locally communities around a trusted node	For perform the algorithm here first identify the trusted node	Social network, Distributed network
8	VoteTrust: [21]	D	Communication of user is on accepting associations, globally vote sum up for Sybil prediction	Vote trust using two techniques for Sybil detection	Weighting approach, Fully distributed p2p network

4 Proof of Proposed Work

4.1 Base Work

Homomorphic system with the primitive root imposing discrete logarithmic problem [15].

pick 2 prime numbers p and q
N is equal to p x q.
p and q be secret and N be publically aviable.
Select random number x and a root g where g is the primitive root of the number p nas both are less than p. r is random number

$y = g^x \bmod p$.
Encryption

1. random integer number= r and.
   E1(rating) = (rating +r*p) mod N.
2. Select random integer number k,:
   Ei(rating) = (a,b) = (g^k mod p, y^k E1(M)mod p)

Decrypted algorithm Di (a,b) is M=b× (ax)-1 (mod p).
Additive: M1+M2=Dg (Eg (M1) ⊕Eg (M2)).

Mathematically based cryptosystem

select any two prime numbers say p and q
P=11 q=3,N=pq=33
X=3 g=7

---

g of GF(p). where g and x are smaller than p.
gf(p) here is primitive root of 11 Gf(11)=(2,6,7,8)

$y = g^x$ mod p. use this y for the encryption
y=73mod11 =2

---

encryption will be performed in following two steps:
1. E1(M) = (M+r*p) mod N.
M1 =2 r=5
E1(2)=(2+5x11) mod33
E1(2)=24
2. Select random integer number k, and the encryption algorithms are:
Eg(M) = (a,b) = (gk mod p, yk E1(M)mod p)
K=7
Eg(2)=(a,b)=(77mod11,27x24mod11)
Eg(2)=(6,3)

1.E1(M) = (M+r*p) mod N.
M2=3 r=5
E2(3)=(3+5x)mod11=25
K=11
Eg(M2)=(a,b)=(7^{11}mod11,2^{11}x25mod11) Eg(3)=(7,6)
(6,3) ? (7,6)
(110,011) ? (111,110)=(1,5)
M=b× $(ax)^{-1}$ (mod p),
M=5/1 mod11=5 M=5

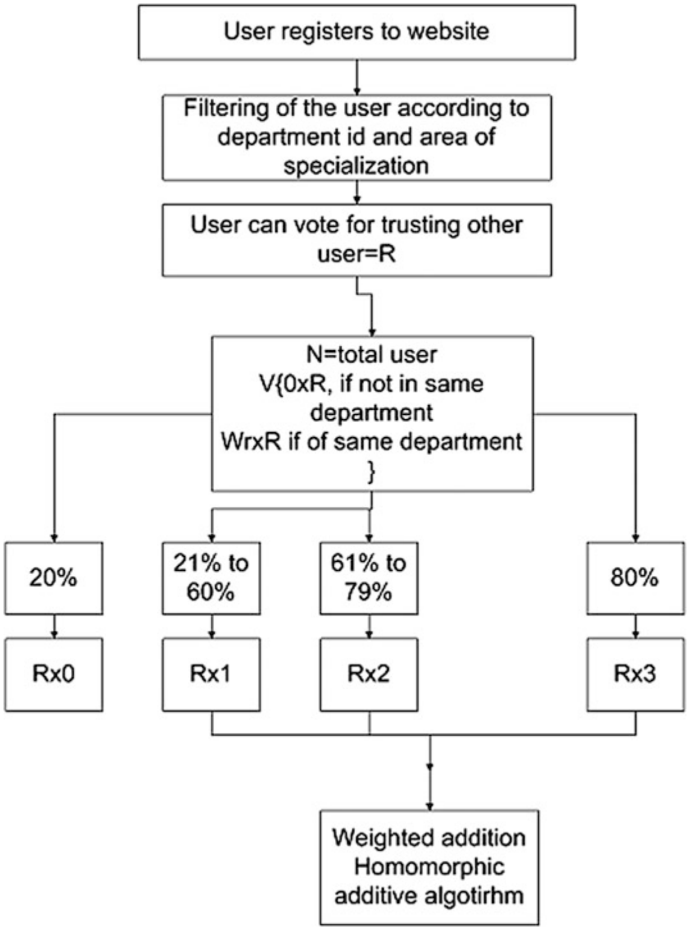


Fig. 2 Flow of working of proposed Sybil detection algorithm

Table 2 Presents the trust matrix for 10 user here number of users can be increased

U/T	1	2	3	4	5	6	7	8	9	10
1	■									■
2		■						■	■	
3			■		■			■	■	■
4				■				■	■	
5					■			■		■
6					■	■		■		■
7						■	■			■
8				■		■		■		■
9				■		■		■	■	
10						■		■		■
Overall trust level (Percentage)	**10**	**10**	**10**	30	30	50	**10**	80	40	70

4.2 Proposed Work

Here, we sketch the algorithm for Sybil detection for the security purpose. Today in real world many malicious user or fake identity available so they are given vote or rating to other people so sometimes it may be possible for honest user that does not get priority than other users. So we perform this algorithm (Fig. 2, Table. 2).

Initially we assume that,

1. M=voting capacity
   N=total node or user into network
   S=seed user it means A user who has get highest rating or above 10
   W=Weight
   R=Rating by the user for an article A
   U=user
   I(u)=initial voting capacity of user
2. PROCEDURE: Sybil Detection

   S ⟵ select or identify Seed user,

   If Uγ€S then,
   I(u) ⟵ $\frac{N}{Ns}$;
   else
   I(u) ⟵ o;
   end if

3. Weighted rating , ratting=R,
   If R < 20% than
   R*0 // Consider as Sybil User
   Else if 21%<R<60%
   R*1;
   Else if 21%<R<79%
   R*2;
   Else R>80%
   R*3;
   End if

Now, if U€S than,
Weighted addition homomorphic algorithm;
Else
User has lower rating & we can discard that user;
End if

If user 8 rates A as
$R = 2$
Trust matrix score = 3
2*3 = 6 hence rating got the weight, users in bolds are sybils

5 Experimental Evaluation

The dataset [21] holds “49,290 users with total number of rating as 139,738 487,181 trust statements. Having id i.e. anonymity. The dataset consists of 2 files.

Ratings data ratings_data.txt.bz2 (2.5 Megabytes): it contains the ratings given by users to items.

Every line has the following format:

user_id item_id rating_value
1 1000 4
“user 1 has rated item 1000 as 4”

For experiment Java language is used with php my admin at back en we set probable Sybil to be 0, as it will be worst case and we changed their ratios set of sybils as 10, 20, 30 and 40 respectively, out of 100 participants. Figure 3 shows the output as the lower is the Sybil community the higher will be the weighted rating as per the trust matrix formed above and Fig. 4 is the storage space for both base and proposed

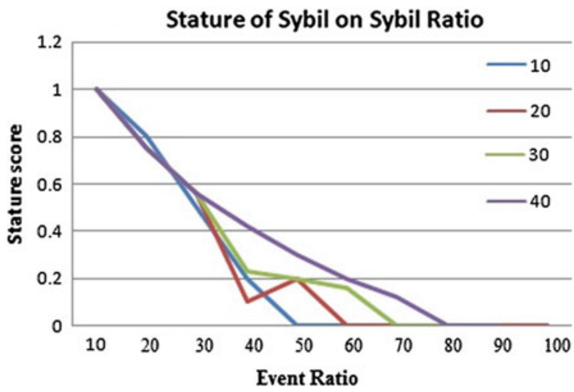
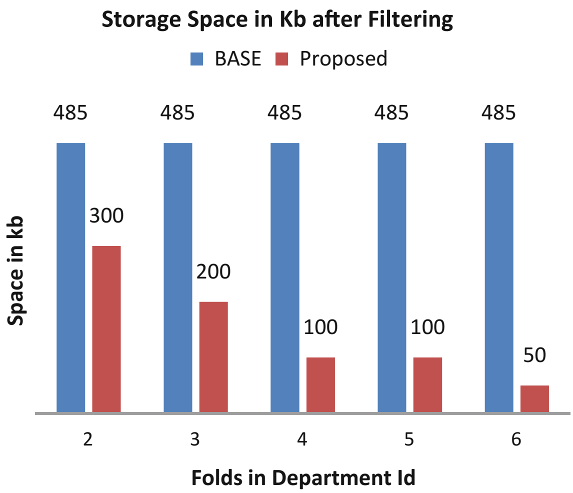


Fig. 3 Comparison of detection by Sybil ratio

Fig. 4 Comparison of storage space



6 Conclusion

We conclude that proposed modification is better as

1. New Matrix is secure and transfer trust decentralized way.
2. New Matrix is used for limiting k party cryptosystem
3. New Matrix due to filtering has lower storage space

Hence we can say that stature system can be enhancing by using the proposed algorithm. Further work can be extended by graph matrix and tags for the sybils for more security.

References

1. Newsome, James, et al. "The sybil attack in sensor networks: analysis & defenses." Proceedings of the 3rd international symposium on Information processing in sensor networks. ACM, 2004.
2. Levien, R. "from Advogato Website." *Advogato Trust Metric, URL:* http://www.advogato.org/trust-metric.html.
3. Kaashoek, M. Frans, and David R. Karger. "Koorde: A simple degree-optimal distributed hash table." *Peer-to-peer systems II*. Springer Berlin Heidelberg, 2003. 98–107.
4. Duggan, Maeve, et al. "Social media update 2014." *Pew Research Center* 19 (2015).
5. D. Gambetta. Trust: Making and Breaking Cooperative Relatioins, chapter Can We Trust Trust?, pages 213 – 237. Department of Sociology, University of Oxford, 2000.
6. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In Proceedings of the International World Wide Web Conference (WWW 2004), 2004.
7. Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. (2006). Sybilguard: defending against sybil attacks via social networks. ACM SIGCOMM Computer Communication Review, 7.

8. H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. In Proc. IEEE S&P, Oakland, CA, May 2008.
9. Danezis, George, and Prateek Mittal. "SybilInfer: Detecting Sybil Nodes using Social Networks." 2. Jøsang, Audun, and Roslan Ismail, The beta reputation system, Proceedings of the 15th bled electronic commerce conference. 2002.
10. Wei, Wei, et al. "Sybildefender: Defend against sybil attacks in large social networks." INFOCOM, Proceedings IEEE. IEEE, 2012.
11. Tran, Nguyen, et al. "Optimal sybil-resilient node admission control."INFOCOM, 2011 Proceedings IEEE. IEEE, 2011.2. Jøsang, Audun, and Roslan Ismail,The beta reputation system, Proceedings of the 15th bled electronic commerce conference. 2002.
12. Tran, Dinh Nguyen, et al. "Sybil-Resilient Online Content Voting." NSDI. Vol. 9. No. 1.
13. Quercia, Daniele, and Stephen Hailes. "Sybil attacks against mobile users: friends and foes to the rescue." INFOCOM, 2010 Proceedings IEEE. IEEE, 2010.
14. Jøsang, Audun, and Roslan Ismail, The beta reputation system, Proceedings of the 15th bled electronic commerce conference. 2002.
15. Zhou, Runfang, Kai Hwang, and Min Cai., Gossiptrust for fast reputation aggregation I peer-to-peer networks. Knowledge and Data Engineering, IEEE Transactions on 20.9 (2008):1282–1295.
16. Gupta, Minaxi, Paul Judge, and MostafaAmmar., A reputation system for peer-to-peer networks, Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video. ACM, 2003.
17. Zhou, Runfang, Kai Hwang, and Min Cai., Gossiptrust for fast reputation aggregation I peer-to-peer networks. Knowledge and Data Engineering, IEEE Transactions on 20.9 (2008):1282–1295.
18. Thadani, Ankita, and Vinit Gupta. "Enhancing Privacy Preservation of Stature System Through Homomorphic System." Emerging Research in Computing, Information, Communication and Applications. Springer India, 2015. 439–449.
19. Viswanath, Bimal, et al. "An analysis of social network-based sybil defenses." *ACM SIGCOMM Computer Communication Review* 41.4 (2011): 363–374.
20. Wei, Wei, et al. "Sybildefender: Defend against sybil attacks in large social networks." *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012.
21. Xue, Jilong, et al. "Votetrust: Leveraging friend invitation graph to defend against social network sybils." *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013.
22. http://www.trustlet.org/wiki/Downloaded_Epinions_dataset

Word Sense Disambiguation for Indian Languages

Mitul Sheth, Shivang Popat and Tarjni Vyas

Abstract Word sense disambiguation is a part of Natural Language Processing (NLP) and technique to computationally find the correct sense of the ambiguous words using the context in which they occur. It is used in many applications like machine translation, information retrieval, information extraction, etc. There are many approaches for WSD like supervised, unsupervised, etc. In this thesis all the approaches are discussed and Graph-Based Unsupervised approach is applied with global measures for Indian Language using Hindi WordNet. The motive is to find more accuracy and easy adaptability of WSD for an Indian language.

Keywords Word sense disambiguation · Natural language processing
Graph-based approach · Unsupervised learning

1 Introduction

1.1 Natural Language Processing

Natural language process is a big step forward towards field of artificial intelligence and machine learning. Because of NLP, the goal of more user-friendly machines and computers can be achieved since people can talk to their machines in a free manner [1]. Natural language processing is the ability to process human language. Humans communicate in the most natural ways through the language and since storage transformation and analysis of documents which are written in natural language are done by machines lately, NLP is necessary [2, 3].

M. Sheth · S. Popat · T. Vyas (✉)
Institute of Technology, Nirma University, S G Highway, Ahmedabad 382481, India
e-mail: tarjni.vyas@nirmauni.ac.in

M. Sheth
e-mail: 14MCEN24@nirmauni.ac.in

S. Popat
e-mail: 14MCEN14@nirmauni.ac.in

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_50

1.2 Wordnet

Wordnet is a lexical database. Princeton university has created WordNet and currently version 3.0 is available. It contains 155,000 words with approximate 117,000 synsets, e.g., automobile with its synsets are: {car, auto, automobile, machine, motorcar} [1]

WordNet has multiple synonyms of words so they can be used to remove ambiguity in words. Verbs, nouns, adjectives all are in WordNet.

WordNet is a lexical resource which contains verbs, adverbs, adjectives into sets which are called synsets. These synsets are also interlinked to form a network which can be seen by a browser. Wordnets for different languages are freely and publicly available. WordNet is very much useful in WSD.

1.3 Introduction of WSD

Word Sense Disambiguation is used to identify correct meaning of word. It is the main problem in NLP and it has may application like information retrieval, information extraction, machine learning. WSD is a method to find the proper sense of the given word. Example, word bank has many meaning in English. These kinds of words with multiple senses are called polysemous words. WSD is the process of finding out the exact sense of Polysemous word [1, 4].

1.4 Need of Research

WSD finds its many application in different areas:

- Information Extraction (IE): In particular information extraction is interested to find between specific instances of concepts, e.g., M. Smith likes fishing. But he does not like biking.
- Machine Translation (MT): Using word sense disambiguation roper translation of the word can be done which can be useful in machine translation.
- Word Processing: Spelling of word can be corrected using word sense disambiguation and also special characters can be added properly.

2 Literature Survey

In human language, ambiguity in word sense is prime concern. To remove this ambiguity many researchers have chosen many different paths and proposed many solutions for different languages.

2.1 Approaches to WSD

There are many approaches in Word Sense Disambiguation. Mainly those approaches depend on the WordNet means data set which is available or not. Based on this mainly data set available or not. They are mainly Supervised and Unsupervised Disambiguation approach. Some other Disambiguation approaches also there but mainly are those [5–7].

2.1.1 Knowledge Base Approach

In this technique mostly dictionary like WordNet or thesaurus is used in back end. The main difference between wordnet and thesaurus is: thesaurus contains only synsets which defines only one type of relation where WordNet contains more than 8 relationships among words [8, 9].

2.1.2 Supervised Approach

Supervised approach uses classification methods to disambiguate the sense classifier is focused on polysemous word and uses the different classification methods to disambiguate the particular word. For training of the classifier training data is used in which a set of examples with target word are manually tagged with senses. These senses are from the sense inventory of the WordNet or machine readable dictionary.

2.1.3 Unsupervised Approach

Unsupervised approach uses untagged corpus and does not need training data. It mainly includes two methods which are [10]:

Similarity-Based Methods

Similarity-based methods use the comparison between different senses of ambiguous word and context words. The most similar sense is the winning sense. Different methods use different definitions of senses as well as uses from sentence to whole corpus as context words. But experimental results shows that Graph-Based approaches are more accurate then similarity-based methods [11, 12].

Graph-Based Methods

There is another unsupervised approach is based on graph. In this approach, after processing the sentence find the polysemous words and then creating virtual graph on vector. These nodes are given labeled which are provide us to compute the similarity. Then based on measures the similarity value are computed and based on that sense label will be given to words [13].

2.2 *Working of Graph-Based Algorithm*

In this algorithm there are mainly three steps:

- Graph is developing with nodes which indicate senses and joint means edges as relations between those nodes.
- Between two nodes, weights are given using global measures of algorithm.
- Select senses which have the highest weights are assigned to that respective words.

Above three steps can be understood using the below example: She drank some milk.

In above sentence there are two polysemous words drinks and milk. As our algorithm finds graph using WordNet. For the word "drink" the graph shown in the fig. For word "milk" the graph will be created same manner. According to the step one then graph will be created step by step. In WordNet for words 'drink' and 'milk' there are five and four senses accordingly (Fig. 1).

After this, we need to find measures. For that there are mainly two connectivity measures which are:

2.2.1 Global Measures

Compactness

Due to high compactness vertex can easily reach to other vertex. The equation to measure Compactness is: [8]

$$\frac{CO\,(G) = \mathrm{Max} - \sum_{u \in \mathrm{V}} \sum_{v \in \mathrm{V}} d\,(u, v)}{\mathrm{Max} - \mathrm{Min}} \tag{1}$$

$$CO(G) = \frac{(5 \times 5 \times 4) - 28}{(5 \times 5 \times 4) - (5 \times 4)} = 0.90$$

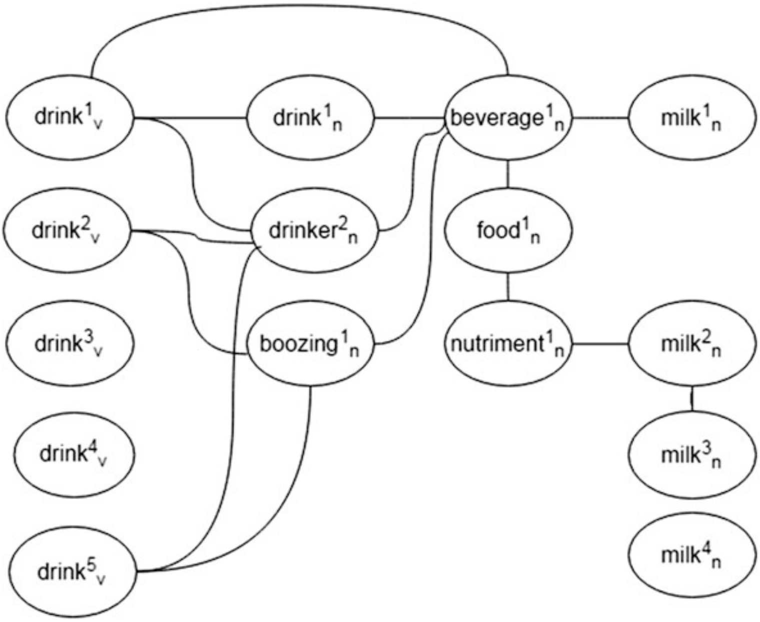


Fig. 1 Last Step Graph construction process for the sentence: "She drank some milk."

Entropy

Entropy is used to find the information stored in random variable. In graph theory, more entropy means more nodes are valuable while low entropy indicates vice a versa. Entropy equation is [8]:

$$H(G) = -\left(\sum\nolimits_{v \in V} p(v) \log(p(v))\right) \tag{2}$$

where the vertex probability $p(v)$ is determined by [8]:

$$\left\{\frac{\deg(v)}{2 \times |E|}\right\}_{v \epsilon V} \tag{3}$$

For above Graph the Entropy value is

$$P(v) = \left(\frac{3}{12}, \frac{2}{12}, \frac{4}{12}, \frac{1}{12}, \frac{2}{12}\right)$$
$$H(G) = 0.94$$

Edge Density

Edge density is a global measure. It combines both.edges in normal graph and edges in fully connected (complete) graph. The equation is [8]:

$$ED\,(G) = \frac{|E(G)|}{\binom{|V|}{2}} \tag{4}$$

For the first graph it is

$$ED\,(\,G) = \frac{6}{\binom{5}{2}} = 0.60$$

After computing the score search which sense is for the given sentence and it is the final step of the algorithm. According to sense the document will be retrieve from the number of document (Fig. 2).

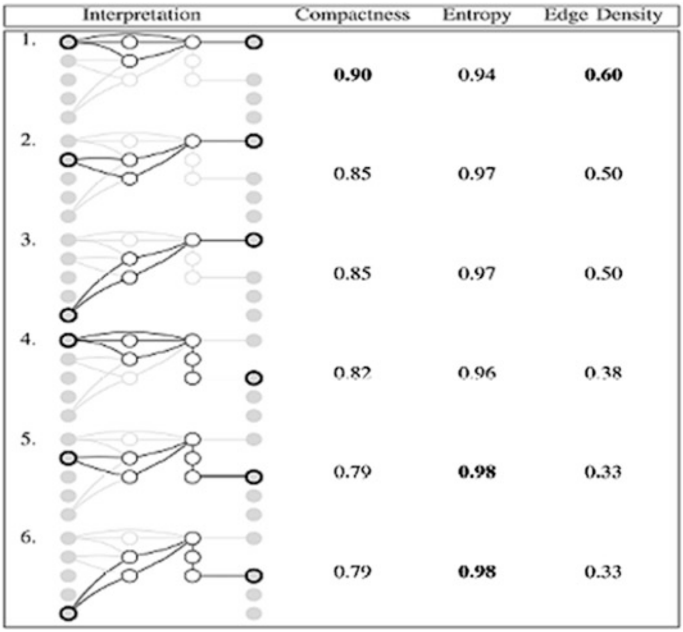


Fig. 2 Result after processing [8]

2.2.2 Local Measures

Degree Centrality

Degree centrality is based upon degree of vertex. High degree vertex is central and zero degree vertex is disconnected. Degree centrality is the relation between vertex degree and maximum degree [8].

$$C_D(V) = \frac{\deg(v)}{|v| - 1} \tag{5}$$

In the above figure For the graph 1 value is

$$C_D\left(\text{drink}_v^1\right) = \frac{3}{14}$$

Key Player Problem (KPP)

In KPP, if vertex is close to others then it is considered important. The KPP is defined as [8]:

$$\text{KPP}_{(v)} = \frac{\sum u \in V : u \neq v \frac{1}{d\,(u,v)}}{|v| - 1} \tag{6}$$

After measuring the based on the score find which sense is for the given sentence and this is the last step of the algorithm. According to sense the document will be retrieve from the number of document (Table 1).

Table 1 The best scores for the senses of drink (v) and milk (n) for each connectivity measure are shown in bold face [8]

Nodes	Degree	Kpp
drink_v^1	0.21 = ✓	0.45 = ✓
drink_v^2	0.14	0.37
drink_v^3	0.0	0.07
drink_v^4	0.0	0.07
drink_v^5	0.14	0.37
milk_n^1	0.7 = ✓	0.36 = ✓
milk_n^2	0.7 = ✓	0.27
milk_n^3	0.0	0.07
milk_n^4	0.0	0.07

3 Implementation

3.1 Implementation Unsupervised Graph-Based Approach in Hindi Language

This is the implementation of the Unsupervised Graph-based approach on Hindi WordNet. The implementation is done for the same example "she drank the milk" in Hindi. There are two polysemous words in the sentence milk and drink.

First the senses of both the words have been found. Then, we have created tree for each sense of the word to its root word and compared these senses of one polysemous word with the other polysemous word. After that, we created the LCS tree by comparing different senses of word with other and then using that LCS tree we have created the graph which can be used to apply local or global measures for finding out the accurate sense of the polysemous word. In graph, nodes are senses and edges are semantic relation between them.

Graph for Hindi words with nodes as senses and edges as semantic relation between them. Words are milk and drink with drink has 2 senses and milk has 3 senses.

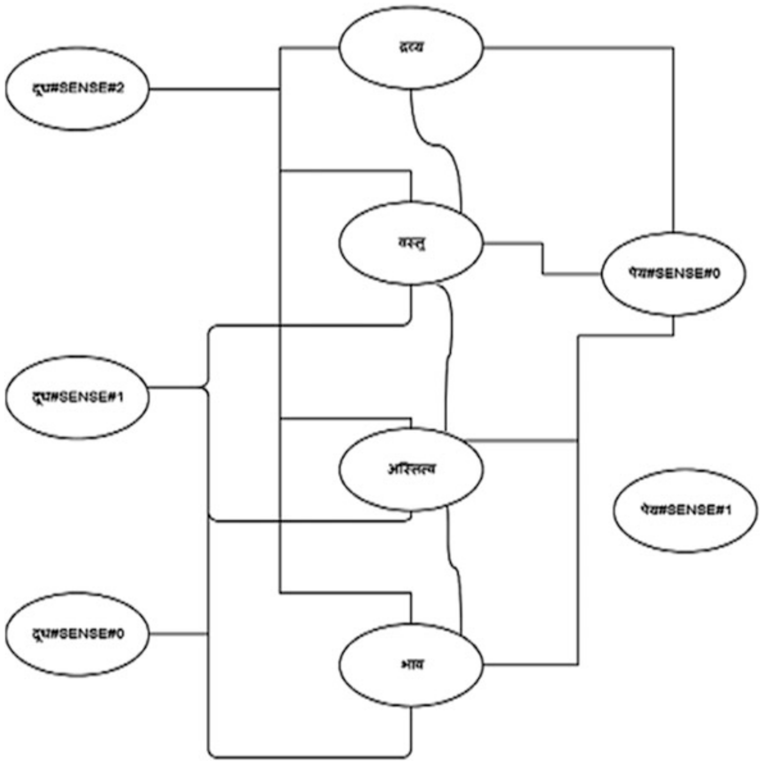


Fig. 3 Graph for Hindi words

Table 2 The best scores for the senses of drink (*v*) and milk (*n*) for each connectivity measure are shown in correct face

No. of edges	V	Senses	Results
8	5	Sense 0 = 0.8	✓
8	5	Sense 1 = 0.8	✓
11	6	Sense 2 = 0.7	×

Table 3 The best scores for the senses of drink (*v*) and milk (*n*) for each connectivity measure are shown in correct face

Nodes	Degree
milk_n^2	0.5 = ✓
milk_n^1	0.37
milk_n^0	0.37
drink_v^4	0.5 = ✓
drink_v^5	0.0

Table 4 Correct results for global measure (Example sentences: १ सूरय कई सतिारों में से एक हे | २ पाठशाला की इमारत अच्छी हे |)

सूरय	सतिारा	Score	Correct sense
Sense0	Sense0	**0.67**	✓
Sense0	Sense1	0.67	×
Sense0	Sense2	0.6	×
पाठशाला	इमारत	score	–
Sense0	Sense0	**0.6666667**	✓
Sense0	Sense1	0.52380955	×
Sense1	Sense0	0.47619048	×
Sense1	Sense1	0.5	×

Also, drink sense#2 is not connected to any other node because of empty LCS tree. Edge Density is applied as a Global measure and following are the result.

Degree Centrality is applied as a local measure and following are the result.

4 Performance Analysis

In this research, performance may depend on the Hindi words. For some words program can find different sense and using measures sense can be found using above formula. For some words all the value are same and as well as LCS tree so it is very hard to find the which sense will be applied, in this paper we have only user one method for each measure.

Here are the examples on which we have applied global measure (Table 4).

Here are the examples on which we have applied local measure (Table 5).

Table 5 Correct results for local measure (१ शरीर में दो आँख होती हे | २ कार वाहन के व्‌रग में आती हे |)

Words	Senses				Correct sense
	Sense0	Sense1	Sense2	Sense3	
शरीर	**0.25**	–	–	–	Sense0
आँख	0.16666664	**0.33333334**	0.16666667	0.125	Sense1
वाहन	**0.2857143**	–	–	–	Sense0
कार	**0.2857143**	0.21428572	–	–	Sense0

5 Conclusion

Knowledge base approaches generally go for thesaurus based algorithm which require ontological information for sense extraction. This information adds additional cost to computation as the information is provided manually.

Supervised algorithm requires training set which implies increase in efforts. It is the most cumbersome job to tag words with the sense previously and than create a model which classify the words. Manual tagging always relies on human efficiency which is obviously less than automated machines.

Unsupervised algorithms do not require any training set, however, it will be generated from context word but each new polysemous training is required. Another limitation is poor accuracy and the techniques which have higher accuracy are very complex to implement. However, it takes a huge amount of efforts to get implemented unsupervised graph-based approach in Hindi language.

As mentioned earlier that its a new approach for Indian languages so Higher accuracy and adaptability can be applied in Graph-based unsupervised approach.

References

1. R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys (CSUR),* p. 10, 2009.
2. R. a. W. J. a. B. L. a. K. M. a. K. K. a. K. P. Collobert, "Natural language processing (almost) from scratch," *The Journal of Machine Learning Research,* pp. 2493–2537, 2011.
3. A. Z. Z. Merhbene Laroussi, "Unsupervised system for Lexical Disambiguation of Arabic Language using a vote procedure," *IEEE-ICoAC,* 2011.
4. D. L. a. P. Pantel, "Discovering Word Senses From Text," in *ACM*, 2002.
5. M. S. Deepak Kumar, "FST Based Morphological Analyzer for Hindi Language," *International Journal of Computer Science Issues(IJCSI),* vol. 9, pp. 349–353, 2012.
6. S. P.Tamilselvi, "Optimal Distance Metric Function with Trigram Features for Case based Word Sense Disambiguation Using Artificial Neural Network," 2014.
7. R. a. V. P. Navigli, "Structural semantic interconnections: a knowledge-based approach to word sense disambiguation," *Pattern Analysis and Machine Intelligence, IEEE Transactions,* pp. 1075–1086, 2005.

8. R. a. L. M. Navigli, "An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation," *Pattern Analysis and Machine Intelligence, IEEE Transactions,* pp. 678–692, 2010.
9. R. Mihalcea, "Knowledge-based methods for WSD. In Word Sense Disambiguation: Algorithm And Application," in *Springer*, 2006.
10. E. A. a. A. Soroa, "Using the Multilingual Central Repository for Graph-Based Word Sense Disambiguation," in *Sixth Language Resources and Evaluation,* 2008.
11. E. a. E. P. G. Agirre, "Word sense disambiguation: Algorithms and applications," Springer Science\& Business Media, 2007.
12. R. a. M. R. Sinha, "Unsupervised graph-basedword sense disambiguation using measures of word semantic similarity," IEEE, 2007, pp. 363–369.
13. G. E. a. D. Radev, "Lexrank: Graph-Based Lexical," *Artificial Intelligence Research,* vol. 22, pp. 457–479, 2004.

Discovery of Fuzzy Hierarchical Classification Rules (FHCRs): A Genetic Algorithm Approach

Renu Bala and Saroj Ratnoo

Abstract Conventional Fuzzy Classification Rules (FCRs) are most often discovered in the form of flat '*If-Then*' rules. These flat rules increase the size of Fuzzy Rule Based Systems (FRBSs). Moreover, a large number of rules are not considered comprehensible. Most of the real world knowledge can be organized into hierarchical fashion for deciphering knowledge at multiple levels of details. However, the flat representation of rules ignores hierarchical relationships that may exist among the classes of a dataset. This paper proposes a Genetic Algorithm approach to discover Fuzzy Hierarchical Classification Rules (FHCRs). The rules discovered in this form can predict the knowledge at various level of abstraction. As a part of the GA design, a suitable encoding scheme to capture the hierarchical structure of knowledge being mined, a fitness function to measure goodness of the hierarchies and genetic operators to evolve the solutions have been suggested. The proposed approach is illustrated on a 'Land transport' dataset specifically designed for the purpose.

Keywords Fuzzy hierarchical classification rules · FHCRs · Fuzzy classification rules · FCRs · Fuzzy hierarchical production rules · FHPRs · Hierarchies Degree of subsumption · Coefficient of similarity

1 Introduction

The problem of discovering flat classification rules has been extensively studied in the area of data mining and machine learning [2]. In the past few decades, discovering hierarchical classification rules has become a priority for researchers in the field of rule mining [2–10]. A knowledge base organized in hierarchical form is not

R. Bala (✉) · S. Ratnoo
Deprtment of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar 125001, India
e-mail: renu0805@gmail.com

S. Ratnoo
e-mail: ratnoo.saroj@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_51

only comprehensible; it can make prediction at multiple levels of abstractions i.e. it can handle variable precision logic with respect to specificity. One of the efficient and easy to understand rule structure, to support the discovery of hierarchical classification rules, is Hierarchical Production Rule Systems (HPRSs) [1, 6, 4]. Hierarchical multi-label classification is another complex challenging task that requires discovery of rules in hierarchical form. In hierarchical multi-label classification problem, an instance can be assigned to more than one classes out of hundreds or thousands of the classes [13]. Popular examples of the hierarchical multi-label classification problems discussed in literature are the task of text classification [2, 3, 13] and protein function prediction [13, 5].

In real world problem domain, a machine has to reason using insufficient, vague and uncertain premises. Fuzzy Classification Rules (FCRs) are recognized as competent systems for handling noisy, imprecise and incomplete information. The FCRs of the form ‘If < X is A > Then < Y is B>’ are integral part of Fuzzy Rule Base Systems (FRBSs). The automated discovery of Fuzzy rules from databases is an active area of research. However, discovery of Fuzzy rules in flat form increases the size of the Rule base at a rate which puts impractical overheads on the inference drawing mechanism for control applications. The size of FRBSs can be effectively reduced by discovering rules in hierarchical form [5].

This paper proposes discovery of Fuzzy Hierarchical Classification Rules (FHCRs)- a fuzzy counterpart of Hierarchical Production Rules (HPRs)- using GA. We have suggested degree of Subsumption and coefficient of similarity as quantitative measures for evaluating the goodness of FHCRs. The rest of the paper is organized as follows. Section 2 introduces the essential background details. Section 3 presents the design of GA in detail. It explains the encoding scheme, fitness function, the operators used in GA design. Section 4 presents the experimental setup and results. Section 5 gives the conclusions and future directions of this research.

2 Background Details

2.1 Fuzzy Hierarchical Classification Rules

FHCRs are the extension of Fuzzy Classification Rules (FCRs) to accommodate hierarchies. A FHCR is represented as

- *If* P is X *Then* C_k
- *Generality* [C_g]
- *Specificity* [C_{s1}, C_{s2}]

Here C_k represents the rule class, C_g is the general class and C_{s1}, C_{s2} are the specific classes. The general class represents the most general concept and specific class shows the most specific concepts. Since each rule in the hierarchy inherits all the properties of its parent FHCR, it is not required to list all such properties repetitively. To make decision about the general and specific classes, degree of Subsumption and coefficient of similarity- two quantitative measures- are used. If Class C_k subsumes C_s then C_k is more general class than C_s, i.e., C_s is the specific class of C_k. The next section describes these two measures in detail.

2.2 Degree of Subsumption

To compute degree of Subsumption between two given classes C_k and C_s, we need to identify their defining properties. The defining properties for a class C_k are the distinct linguistic labels (i.e., small, medium, high) for the attributes of a given dataset (U) which have recall more than a user-defined threshold value θ_t (0.7 in this paper) in class C_k [1, 6]. The values for recall, denoted by x and y, for the ith property in the class C_k and jth property in the class C_s are computed respectively as below:

$$x_i = \frac{|(\mu(P_i)_\alpha \wedge C_k)|}{|C_k|}. \tag{1}$$

$$y_j = \frac{|(\mu(P_j)_\alpha \wedge C_s)|}{|C_s|}. \tag{2}$$

The symbol α in Eq. 1 and 2 represents alpha-cut operation on fuzzy membership degrees. We have kept the value of alpha equals to 0.55. The sets of defining properties $S(C_k)$ and $S(C_s)$ for the classes C_k and C_s are computed as below:

$$S(C_k) = \forall P_i \in U;\ y_i > \theta_t. \tag{3}$$

$$S(C_s) = \forall P_j \in U;\ y_j > \theta_t. \tag{4}$$

The Subsumption between ith and jth properties of defining sets $S(C_k)$ and $S(C_s)$ is evaluated as below:

$$\text{subsume}\,(S(C_k(i)), S(C_s(j)) = 1 \quad \text{if}\ (x \le y)\ \text{and}\ (i = j). \tag{5}$$

$$\text{subsume}\,(S(C_k(i)), S(C_s(j)) = y \quad \text{if}\ (x > y)\ \text{and}\ (i = j). \tag{6}$$

$$\text{subsume}\,(S(C_k(i)), S(C_s(j)) = 0 \quad \text{if}\ \ (i \ne j). \tag{7}$$

The overall Subsumption between the two classes is given by

$$\text{deg\_sub}(C_k, C_s) = \sum_{i=1}^{|S(C_k)|} \sum_{j=1}^{|S(C_s)|} \frac{\text{subsume}\,(S(C_k(i)), S(C_s(j)))}{|S(C_k)|}. \tag{8}$$

The degree of subsumption decides the hierarchy of classes. For example, the subsumption (Two-wheeler, Scooter) > Subsumption (Scooter, Two-wheeler), implies that the ‘Two-wheeler’ class subsumes the ‘Scooter’ class, i.e., The ‘Two-Wheeler’ class becomes the general class of the ‘Scooter’ class.

2.3 Coefficient of Similarity

A coefficient of similarity (S) of attributes between two classes C_k and C_s is defined on the basis of a 2 × 2 contingency table (Table 1)

$$\chi^2 = \frac{N(ps - qr)^2}{M}. \tag{9}$$

$$N = p + q + r + s \tag{10}$$

$$M = (p + q)\,(r + s)\,(p + r)\,(q + s). \tag{11}$$

$$S = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{12}$$

In the above Equations, N is the total number of properties present; k is the degree of freedom. The value of S shall always be between 0 and 1. Higher the value of S, more is similarity between the classes involved [10, 6]. The coefficient of similarity is helpful in deciding the hierarchy between the classes when a general class subsumes two classes with same degree of Subsumption. For example let us assume C_{car} subsume two classes $C_{\text{Big-car}}$ and C_{Sedan} with same degree of Subsumption. In such situation either $C_{\text{Big-car}}$ or C_{Sedan} can be put as specific class of C_{car}. However, computation of similarity enables the system to put $C_{\text{Big-car}}$ as the specific class of C_{car} and C_{Sedan} as the specific class of $C_{\text{Big-car}}$.

Table 1 Contingency table

$S(C_k)$ (*Two-wheeler*) ↓	$S(C_s)$ (*Scooter*) →	
	Observed	Not observed
Observed	p	q
Not observed	r	s

3 Genetic Algorithm Approach

3.1 Individual Representation

A hierarchical individual is represented as set of four different blocks. The first block specifies premise part of the rule. The second block designates the decision part of the rule. The third block specifies the generality part and the final block shows the specificity part of the individual. Here is a constraint on individuals that the rule class, the general class and the specific classes need to be all distinct. Figure 1 shows the encoding of two individual chromosomes and their mapping to the corresponding rules. The encoding scheme maps the chromosomes to rules in CNF form where there is a conjunction between different attributes. We have adopted the Michigan approach where each chromosome in the population represents a single rule. We have applied a pure binary string for encoding the premise part of a chromosome. A block of *n* bits signifies *n* consecutive linguistic fuzzy variables- '*small*', '*small-medium*', '*medium*', '*medium-large*' and '*large*'. Within a block, a '1' bit represents the presence of a linguistic term whereas a '0' bit marks the absence of any value. A block with all bits set to 1 or 0 is treated as a '*don't care*' state which indicates nonexistence of an attribute from a rule. The consequent part contains the class label of the rule and generality part contains the label for the general class of the rule. The next block consists of three bits—each for one of the specific classes—which specify that there are at most three specific classes of the rule class.

3.2 Initial Population

We got the set of FCRs by employing the algorithm proposed in [2] for the discovery of Fuzzy Censored Classification Rules [FCCRs]. The Censors which are

Genotype (Chromosomes)

Weight	Mileage	Width	Length	Boot Space	Fuel Capacity	Class	Generality	Specificity
100	111	100	100	1000	1000	1	0	3 4 0
Weight	Mileage	Width	Length	Boot Space	Fuel Capacity	Class	Generality	Specificity
010	100	010	000	1111	000	2	0	5 6 0

Phenotype (Rules)

If (Weight is low) $\wedge$ (width is low) $\wedge$ (Length is low) $\wedge$ (Boot-space is low) $\wedge$ (Fuel-capacity is low) →Two-wheeler Generality[] Specificity[Scooter, Bike]

If (Weight is medium) $\wedge$ (Mileage is low) $\wedge$ (width is medium) →Car Generality[] Specificity [Small-car, Big-car]

Fig. 1 Encoding scheme and its mapping to rule exceptions to the rules have been dropped because these are not the part of the rule structure FHCRs. Initial population of FHCRs is constructed from these Pre-discovered set of FCRs.

The initial population of the FHCRs is generated by taking rules from the FCRs and generating classes randomly in their generality and specificity parts. The number of possible combinations will increase with the increase in the number of classes present in a dataset.

3.3 Fitness Evaluation

Fitness function gives the quantitative measure to test the quality of FHCRs in the population. In the proposed approach two quantitative measures degree of Subsumption and similarity coefficient, described in Sect. 2, are considered to evaluate the fitness of an individual. Fitness function is given below.

```
If [Generality =empty and Specificity= empty) then
  Fitness=precision × recall
Else
  Fitness1= precision× recall
  Fitness2=Subsumption[Cg,Ck)×Similarity[Cg,Ck)
```

$$\texttt{Fitness3=}\sum_{s=1}^{3}\ \texttt{Subsumption[Ck,Cs)}\times\texttt{Similarity[Ck, Cs)}$$

```
  Fitness=Fitness1+Fitness2+Fitness3
```

The subsumption and similarity matrices are computed in advance for the set of FCRs in the initial population (Sect. 2, 3) to save the overhead of computation of degree of subsumption and coefficient of similarity again and again for each FHCR generated during evolution process.

3.4 Genetic Operators

A roulette wheel is used as the selection operator. The rules are selected from the same species i.e., the FHCRs of the same rule. For crossover operator, the classes either in generality part or specificity part of the FHCRs are swapped. The mutation operator generates a new individual by mutating its generality and specificity blocks.

4 Experimental Setup and Results

The proposed approach is illustrated on a dataset ‘Land transport’ specifically designed for this purpose. The dataset contains 6 continuous attributes (Weight, Mileage, Width, Length, Boot-space and Fuel-capacity) and five linguistic terms (Low, Low-medium, Medium, Medium-high, High) used for fuzzified attributes. There are 9 distinct classes in this dataset.

We have generated the initial population of FHCRs from the set of FCRs discovered by using the algorithm proposed in [2]. The population size has been kept eight times the numbers of Pre-discovered FCRs. There are eight random FHCRs generated for each of the rule species. The Subsumption and similarity matrices have been computed from these FCRs by using Eqs. 8 and 12. The fitness of individuals in the population has been calculated from the subsumption and similarity matrices. The Crossover and mutation operators have been applied only to the generality and specificity parts of the initial FHCRs. The selection and crossover operators are applied intra-species because cross breeding would result into individuals of lower fitness. The mutation probability is kept as 0.1 and the crossover probability is set to 0.6. We have run the GA for 500 generations. The additional stopping criteria adopted was no change in the fitness of individuals for last 10 generations. The FHCRs discovered for the ‘Land transport’ dataset along with their fitness is shown in Table 2.

Figure 2 shows the algorithm design of the proposed approach.

Further, the conceptual hierarchies have been generated by merging the related FHCRs in post-processing step. The description of post-processing steps is given below:

1. If smaller FHCRs are part of a bigger hierarchy then the bigger FHCR is retained and smaller ones are dropped. For example the FHCR for the ‘Two-wheeler’ class already contains the FHCRs of the ‘Scooter’ and ‘Bike’ classes.
2. If the two FHCRs have a common general class but different specific classes then these two FHCRs are merged into a single hierarchy. For example, FHCRs for the classes ‘Small car’ and ‘Big car’ both have the ‘Car’ as their general class; these two FHCRs are merged to produce a three level hierarchy. Rest of the FHCRs gets subsumed in this hierarchy.
3. Two different hierarchies will be merged only if the threshold criteria for degree of Subsumption and coefficient of similarity are satisfied. These threshold values are kept to be 0.6 for our experimentation. The two hierarchies left, one for the ‘Two-wheeler’ class and the other for the ‘Car’ class, cannot be merged because these do not satisfy the threshold criteria for degree of Subsumption and coefficient of similarity. However, a user-defined class ‘Land transport’ can be added on the top level by human intervention and introspection to merge the hierarchies.

Table 2 FHCRs discovered for Land transport

FHCRs	Fitness1	Fitness2	Fitness3	Total Fitness
If [Weight is low) ∧ [Width is low) ∧ [Length is low) ∧ [Boot_space is low) ∧ [Fuel_capacity is low) → Two-wheeler Generality[] Specificity [Scooter, Bike]	0.325	0	1.763	2.088
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) → Car Generality[] Specificity [Small-car, Big-car]	0.184	0	1.763	1.948
If [Weight is low) ∧ [Mileage is medium) ∧ [Width is low) ∧ [Length is low) ∧ [Boot_space is low) ∧ [Fuel_capacity is low) → Scooter Generality [Two-wheeler] Specificity[]	0.505	0.734	0	1.239
If [Weight is low) ∧ [Mileage is high)∧ [Width is low) ∧ [Length is low) ∧ [Boot_space is low) ∧ [Fuel_capacity is low) → Bike Generality [Two-wheeler] Specificity[]	0.396	0.734	0	1.131
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) ∧ [Length is medium) → Small-car Generality[Car] Specificity[Hatchback]	0.321	0.63	0.848	1.799
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) ∧ [Length is high) → Big-car Generality [Car] Specificity[Sedan, SUV]	0.284	0.63	1.683	2.598
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) ∧ [Length is medium) ∧ [Fuel_capacity is low-medium) → Hatchback Generality[Small-car] Specificity[]	0.115	0.692	0	0.808
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) ∧ [Length is high) ∧ [Boot_space is medium-high) → Sedan Generality[Big-car] Specificity[]	0.418	0.653	0	1.072
If [Weight is medium) ∧ [Mileage is low) ∧ [Width is medium) ∧ [Length is high) ∧ [boot_space is Low-medium) → SUV Generality[Big-car] Specificity[]	0.160	0.672	0	0.832

5 Conclusions and Scope for Future Research

We have proposed a genetic algorithm approach to discover Fuzzy Hierarchical Classification Rules [FHCRs]. FHCRs provide a mechanism to deal with taxonomical structure inherent in real world knowledge. This approach is illustrated on a dataset named as 'Land-transport' specially designed for this approach. This approach discovers rules in hierarchical format reflecting the relationship among the classes of the underlying dataset instead of simple flat rules. The rules discovered in such form can exhibit variable precision logic with respect to specificity. It can predict classes at multiple levels of Granularities.

Algorithm : Discovering Fuzzy Hierarchical Classification Rules (FHCRs)
Input : Set of FCRs Population Size=72/1 ; Mutation rate=0.1; Crossover rate =0.6
Output : FHCRs
Begin
1. Construct Subsumption and Similarity Matrices form the pre-discovered set of FCRs
2. Initialize random population P_0 of FHCRs using pre-discovered set of FCRs and generate random classes for generality and specificity parts
3. Evaluate Fitness of the population P_0 as given in section 3(C)
4. **While** stopping criteria not satisfied
   Begin
   4.1 Select two rules R_1 and R_2 of same species
   4.2 Apply crossover and mutation to produce new individuals $R_1^{'}$ and $R_2^{'}$ //keeping the rule part present in the rule fixed
   4.3 Compute fitness of $R_1^{'}$ and $R_2^{'}$
   4.4 **If** fitness of $R_1^{'}$ > fitness of R_1
      4.4.1 Replace R_1 in P_i with $R_1^{'}$.
   End if
   4.5 **If** fitness of $R_2^{'}$ > fitness of R_2
      Replace R_2 in P_i with $R_2^{'}$
   End if
   End while
End.

Fig. 2 Algorithm design to discover FHCRs

Further, [3] have also introduced the concept of Hierarchical Censored Production Rules [HCPRs] [3]. This is an interesting and concise rule structure that supports discovery of hierarchies along with exceptions at various levels. A HCPR is capable of exhibiting variable precision logic with respect to specificity as well as certainty of belief in a conclusion [10]. HCPRs have numerous applications in those rare situations where decision must be taken in real time environment and exceptional circumstances. The discovery of Fuzzy Hierarchical Classification Rules with exceptions-a fuzzy counterpart of the HCPRs- using GA is underway

References

1. Freitas, A.A. A Review of evolutionary Algorithms for Data Mining. In: Maimon O, Rokach L, [eds.): Soft Computing for Knowledge Discovery and Data Mining. Springer, US (2008) 79–111.
2. Al-maqaleh, B.M., Bharadwaj, K.K. Genetic programming approach to hierarchical production rule discovery. Trans Eng and Comput (2005) 271–274.
3. Cerri, R., Barros, R.C., de, Carvalho, A.C.P.L.F., Freitas, A.A. A grammatical evolution algorithm for generation of Hierarchical Multi-Label Classification rules. In: IEEE Congress on Evolutionary Computation (CEC) (2013) 454–461.

4. Davies, M.N., Secker, A., Freitas, A.A., Mendao, M., Timmis, J., Flower, D.R. On the hierarchical classification of G protein-coupled receptors. Vol. 23(23), Bioinformatics (2007) 3113–3118.
5. Li, Z., Zhou, X., Dai, Z., Zou, X. Classification of G-protein coupled receptors based on support vector machine with maximum relevance minimum redundancy and genetic algorithm. Vol. 11(1), BMC Bioinformatics (2010) 1–15.
6. Ricardo, Cerri, R.C.B. A Genetic Algorithm for Hierarchical Multi-Label Classification. In Proc 27th Annu ACM Symp Appl Comput, ACM, New Yark, USA (2012) 450–455.
7. Secker, A., Davies, M.N., Freitas, A.A., Clark, E.B., Timmis, J., Flower, D.R.. Hierarchical classification of G-protein-coupled receptors with data-driven selection of attributes and classifiers. Vol. 4(2), Int J Data Min Bioinforma (2010) 191–210.
8. Secker, A., Davies, M.N., Freitas, A.A, Timmis, J., Mendao, M., Flower, D.R. An Experimental Comparison of Classification Algorithms for the Hierarchical Prediction of Protein Function.Vol. 9(3), Expert Update [Magazine of the British Computer Society's Specialist Group on AI) (2007) 17–22.
9. Sun, A., Lim, E-.P., Ng, W.-K., Srivastava, J.. Blocking reduction strategies in hierarchical text classification. Vol. 16[10), IEEE Trans Knowl Data Eng. (2004) 1305–1308.
10. Sun, A., Lim, E.-P. Hierarchical text classification and evaluation. In: ICDM 2001, Proceedings IEEE International Conference on Data Mining, IEEE, San Jose, CA (2001) 521–528.
11. Tsumoto, S. Automated extraction of hierarchical decision rules from clinical databases using rough set model. Vol. 24(2),Expert Syst Appl. (2003) 189–197.
12. Bharadwaj, K.K., Saroj. Parallel Genetic Algorithm Approach to Automated Discovery of Hierarchical Production Rules. In: Mehnen, J., Köppen, M., Saad, A., Tiwari, A., (eds.) Applications of Soft Computing, Springer, Berlin Heidelberg (2009) 327–336.
13. Bharadwaj, K.K., Kandwal, R. Cumulative learning based on dynamic clustering of hierarchical production rules [HPRs). In: Proceedings of 6th international conference on machine intelligence, ICMI, Budapest, Hungary (2005) 333–338.
14. Rousu, J., Saunders, C., Szedmak, S., Shawe-Taylor, J. Kernel-Based Learning of Hierarchical Multilabel Classification Models. Vol. 7, J Mach Learn Res. (2006) 1601–1626.
15. Bharadwaj, K.K., Kandwal, R. Cumulative Learning Techniques in Production Rules with Fuzzy Hierarchy (PRFH) System. Vol. 20(2), J Exp Theor Artif Intell. (2008) 111–132.
16. Bala, R., Saroj. Discovering Fuzzy Censored Classification Rules [FCCRs): A Genetic Algorithm Approach. Vol. 3(4) Int J Artif Intell Appl. (2012) 175–188.
17. Bharadwaj, K.K., Jain, N.K. Hierarchical Censored Production Rules (HCPRs) System. Vol. 8(1), Data Knowl Eng. (1992) 19–34.
18. Jain, N.K., Bharadwaj, K.K. Some learning techniques in hierarchical censored production rules (HCPRs) system. Vol. 13(4), Int J Intell Syst. (1998) 319–344.

Classification of Fashionwear Using Deep Learning

Akshayvarun Subramanya, Prajwal Srinivas, V.M. Pramod and S.S. Shylaja

Abstract We propose a method of classifying fashionwear based on apparels and their patterns using deep learning. This helps in providing high-level features for clothing and can be used for matching user preferences. We obtained an accuracy of 73% using a Random Forest Classifier 87% and 87.5% for Pattern and Apparel Classification respectively using Convolutional Neural Networks, which is an improvement over previously computed results. We make the dataset created for this task publicly available. Our application will help in characterization of fashion wear and will improve accuracy when searching for specific clothing.

Keywords Pattern Classification · Deep Learning · Fashion Wear Apparel Classification · Convolutional Neural Network

1 Introduction

Fashion has always been a part of everyone's life over the years. Irrespective of who we are and what we do, how we present ourselves matters. It is important that we understand the concept of trends and help people access specific clothes of their choice.

In this paper, we aim to address the matter of search for fashion wear. Although image searches yield good results for such applications, we aim to provide high-level abstraction to users for helping people find the clothing of their liking

A. Subramanya (✉) · P. Srinivas · V.M. Pramod · S.S. Shylaja
PES Institute of Technology, Bangalore 560085, India
e-mail: aksvarun007@gmail.com

P. Srinivas
e-mail: srinivas.prajwal.95@gmail.com

V.M. Pramod
e-mail: vm.pramod1995@gmail.com

S.S. Shylaja
e-mail: shylaja.sharath@pes.edu

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_52

and also provide accurate results for all the above problems using Machine Learning and specifically Deep Learning methods. We first explored the possibility of using a Random Forest classifier for this application. Although the results were good, we wanted to improve upon that. The rise of Deep Learning methods and the improved performance that it gave over other methods for computer vision tasks motivated us to look in that direction. Convolutional Neural Networks have been shown to perform really well for image related applications. As proved by [1], Convolutional Neural Networks provide very good results for image classification problems.

This paper aims to provide an overview of the methods that are currently in practice and how we have tried to improve and use them for our application.

The datasets and the network architecture that we are trying are available to the public and we encourage people to try out the image query to realize the performance of our methods. Section 2 discusses related works that have taken place relevant to the proposed method. Section 3 provides a brief overview of the Random Forest Classifier and Convolutional Neural Networks. Section 4 explains the experimental results and the practical difficulties that we faced. Section 5 provides conclusion and the future work that we aim to carry out and improve upon performance of the system.

2 Related Works

The problem of classifying clothing attributes has seen growing interest in the recent years. One of the main reasons of classifying clothes is that the detection and description of images or videos can be achieved [2] use Random Forests and Transfer Forests to obtain 38.28% and 41% accuracy respectively [3] use a modified CaffeNet model built on top of the default AlexNet architecture using the Imagenet pre-trained weights. They achieve an overall accuracy of 50.2% for clothing attribute type classification. Some related works use CNNs to learn a similarity measure between query and shop items [4] by treating the street-to- shop match as a binary classification problem and compare the classification results with human evaluated results for the same. The fashion apparel detection proposed by [5] use geometric priors obtained by R-CNN and location priors obtained by pose of person and location of the fashion items, which are combined with appearance-based-posterior given by SVM to obtain the final posterior. In this paper, we give a combined approach to classify an image based on the pattern, type of the clothing. We use existing work on multiclass apparel classification [2, 3] as baselines and collect a new dataset each for clothing Apparel Classification and clothing pattern classification.

3 Apparel and Pattern Classification of Fashionwear

3.1 Random Forest

Random forests are built by combining the predictions of a multitude of decision trees, each of which is trained in isolation. Unlike in boosting where the base models are trained and combined using a sophisticated weighting scheme, typically the trees are trained independently and the predictions of the trees are combined through averaging. The building blocks of the random forests, the decision trees [6] are constructed by selecting the candidate splits based on information gain computed at each node. Since the problem is multiclass classification we use Random Forests. Random Forests work well with high dimensional input and have been known to improve prediction accuracy as well as decrease overfitting.

3.2 Convolutional Neural Networks

Convolutional Neural Networks are biologically inspired variants of Multilayer Perceptrons. Inspired by the animal visual cortex system, this has been proved to be one of the most accurate visual processing systems. A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. But the presence of the convolutional and the sampling layer differentiates between the traditional neural network and the convolutional neural networks. It has been found that these two aspects have very nice properties such as translational invariance and these networks although very deep in their network architectures are easier to train when compared to standard neural networks with same number of layers. A feedforward neural network consisting of L layers can be represented as

$$f(x) = f_L(\ldots(f_2(f_1(x; w_1)w_2)\ldots); w_L) \tag{1}$$

where w_i corresponds to weights at ith layer and f_i is the ith layer

For our application, we use the NVIDIA Digits platform to help us in training and testing our algorithms. Readers are suggested to go through the website of NVIDIA Digits to familiarize themselves and try to understand how to use the same.

AlexNet Architecture: The AlexNet architecture (shown in Fig. 1) was one of the first Deep Learning Network that was used for Image Classification [1]. The network consists of alternating Convolution Layers and Max Pooling Layers. The network is 7 Layers deep. The ordering is *conv-maxpool-conv-maxpool-conv-conv-conv-maxpool-fc-fc-softmax* where *conv* corresponds to Convolutional Layer, *maxpool* corresponds to Max Pooling Layer and *fc* corresponds to Fully Connected

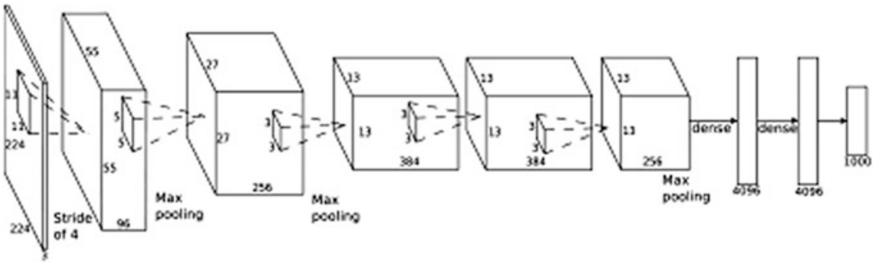


Fig. 1 AlexNet architecture

Table 1 Details of pattern v1 dataset

Class name	Data size
Animal print	211
Checked	1155
Floral	270
Polka	340
Printed	6329
Ripped	317
Solid	12380
Striped	971
Washed	200
Text	458
Sum	22631

Layers as seen in traditional Neural Networks. *Softmax* corresponds to final layer which provides a distribution of outputs across the 1000 classes. The advantage of using this model is that we do not require handcrafted features, as it will try to *learn* better features or high dimensional representations of the data which has been shown to be much richer and more informative when compared to previously used features. The architecture can be understood that the initial layers understand simple features such as edges and as the layers become deeper and deeper, more abstract features such as faces, objects etc. are understood. Here ReLU is used for non-linearity of activations and Stochastic Gradient Descent (SGD) with 30 epochs for training the network.

$$w := w - \eta \nabla J(w) \tag{2}$$

In the equation above, w is the weight matrix of the neural net, and a gradient update is performed using the learning param- eter η. The Images are of dimension 256×256 and we used a 75:25 train-test split for evaluating our algorithm (Table 1).

4 Experiments

4.1 Pattern Classification

Pattern Classification involves the detection of specific regions of clothing which have a uniform pattern such as striped shirts or checked shirts, non-uniform patterns such as printed shirts or floral dresses. We collected a dataset containing 22329 fashionwear images spread over 10 classes namely, solid, checked, floral, polka, printed, animal-print, striped, ripped, text and washed as shown in Table 1. We then create two datasets from the collected input dataset, (a) pattern rf, for the Random Forest classifier, obtained by reducing the size of the input images to a 50×50 patch containing only the area of interest(apparel part of the input image) since HOG feature descriptor gives a high dimensional feature vector and it is invariant to geometric and photometric transformations [7] and (b) pattern v1 for CNN, obtained by cropping the input images to include only the area of interest.[1] The first model RF pattern was generated from building a Random Forest Classifier on the pattern rf dataset. Table 2 shows the classification report, the classes solid, printed and checked have the highest f1-score. This is due to the large number images in these classes compared to the other classes which have relatively few images. We obtained an overall accuracy of 73%. The next model alexnet pattern was generated using Convolutional Neural Network and trained on pattern v1 dataset. We used AlexNet layer configurations. We obtained an accuracy of 87% with alexnet pattern as shown in Fig. 2b.

4.2 Apparel Classification

Apparel Classification involved classification of Dresses into different classes such as Blazers, Dresses, Jeans, Jump Suits, Kurtis, Palazzo Trousers, Sarees, Sherwani, Shirts, Suits, Tops, Trousers and T-shirts. We created a dataset which contained 12759 images distributed across different classes approximately uniformly as shown in Table 3. We use NVIDIA Digits to help us in this task and it provides a user-friendly GUI to perform classification tasks (Table 3).

Each model has its own network architecture and can be used for specific tasks. For this problem, we choose AlexNet which has been known to provide good results for Image Classification problems. Using this architecture we obtain an average accuracy of about 87.5% for the apparel classification task, over 30 epochs and use an exponentially decreasing learning rate starting from 0.01.

The graphs related to this task are found in Table 4.

[1]The dataset is available at — https://goo.gl/AAdEFM.

Table 2 Results of random forest classifier

Class	Precision	Recall	F1-score	Support
Animal print	0.00	0.00	0.00	39
Checked	0.71	0.11	0.20	305
Floral	0.00	0.00	0.00	47
Polka	0.50	0.02	0.05	42
Printed	0.68	0.67	0.67	1611
Ripped	0.00	0.00	0.00	8
Solid	0.78	0.94	0.85	3073
Striped	0.89	0.22	0.35	232
Text	0.00	0.00	0.00	12
Washed	0.00	0.00	0.00	44
Avg/total	0.73	0.75	0.71	5413

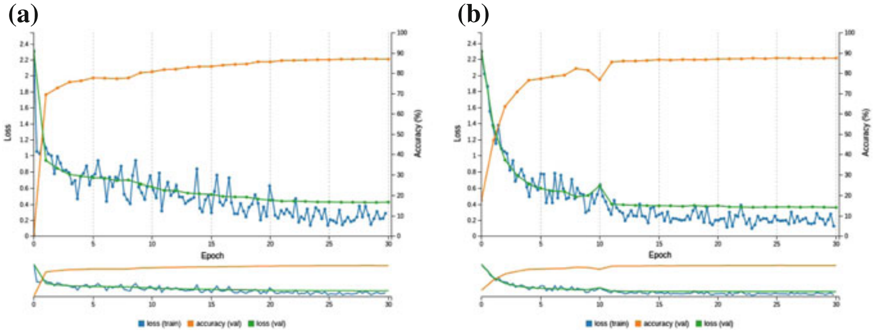


Fig. 2 Results using CNN for **a** Apparel Classification, **b** Pattern Classification

Table 3 Details of category v5 dataset

Class name	Data size
Blazers	1793
Dresses	1013
Jeans	1313
Jumpsuits	947
Kurtis	1065
Palazzo trousers	582
Sarees	1124
Sherwani	402
Shirts	1218
Suits	390
Tops	1050
Trousers	1037
T-shirts	825
Sum	12759

Table 4 Results

Task	Avg. Accuracy
Random forest for pattern classification	73%
CNNs for pattern classification	87%
CNNs for apparel classification	87.5%

5 Conclusions and Future Work

We presented a novel application of Deep Learning methods. Our results were computed over manually created datasets which have been made available to the public. The results obtained are an improvement compared to previous works, we further aim to obtain more real world data and try to increase performance.The authors gratefully acknowledge support from PES University towards funding our research.

References

1. A.Krizhevsky, G.E.Hinton, I.Sutskever. *"ImageNet Classification with Deep Convolutional Neural Networks*". Advances in neural information processing systems, 2012.
2. Lukas Bossard, Matthias Dantone-. *"Apparel Classification with Style*". ACCV, 2012.
3. Brian Lao, Karthik Jagadeesh. *"Convolutional Neural Networks for Fashion Classication and Object Detection".*
4. MH Kiapour, X Han, S Lazebnik, AC Berg, TL Berg. "*Where to Buy It: Matching Street Clothing Photos in Online Shops*". ICCV, 2015
5. K Hara, V Jagadeesh, R Piramuthu. *"Fashion apparel detection: The role of deep convolutional neural network and pose dependent priors*". arXiv preprint arXiv:1411.5319.
6. Criminisi, A., Shotton, J., Konukoglu, E. "*Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning".* Technical Report MSR-TR-2011-114, Microsoft Research (2011).
7. Dalal, N., Triggs, B. *"Histograms of Oriented Gradients for Human Detection".* CVPR, 2005.

An Efficient Algorithm for Frequent Trajectory Itemset

P. Geetha and E. Ramaraj

Abstract Large size database utilization and huge data stream generation in various applications (network traffic analysis, opinion mining, etc.) requires frequent pattern mining. The diversity existence in the database consumes more memory, time required for storage and huge rules to extract the frequent patterns. With the evolution of frequent mining algorithms, the spatial trajectory location prediction is an attractive research area in the data mining field. This paper focus on the real-time frequent item set mining based on Vague Space Partition (VSP) algorithms under the tree-based structures. This paper discusses the Tree-based Space Partition of Trajectory Pattern Mining (TSPTPM) algorithm to reduce the rule, time, and memory consumption and to extract the frequent patterns from the real-time datasets. Initially, the clustering algorithm groups the transactions based on the user choice and sequentially construct the Heap Tree with the maximum transaction ID as the root node. Then, the odd and even number of transactions is placed as the child nodes of the root node. The proposed work utilizes the VSP algorithms to convert the flexible spatial patterns into sequences that reduce the time consumption. The optimal pattern extraction and the effective partition by heap tree structure with VSP offer the considerable reduction of rules in mining. The proposed TSPTPM validates its performance on mining time for various datasets namely, chess, mushroom, connect and accident with various numbers of cluster items. The comparative analysis proves that the proposed TSPTPM algorithm reduces the runtime, memory consumption and the number of rules compared to the traditional frequent mining algorithms.

Keywords Clustering · Data reduction · Execution time · Frequent mining Heap tree · Memory consumption · Space partition

---

P. Geetha (✉) · E. Ramaraj
Department of Computer Science and Engineering, Alagappa University, Karaikudi 630003, Tamil Nadu, India
e-mail: geeth.ganesan@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_53

1 Introduction

Recent developments in several fields such as science, technology and business rise the production and storage of huge data. Thus, an intelligent analysis is the major requirement to extract the relevant data forms the large size database. Frequent Itemset Mining (FIM) is an important part of the data analysis and it extracts the patterns that are frequently occurred. The representation of database is split up into two categories namely, horizontal and vertical. The form of representation is same as the original information in the database refers horizontal. alternatively, the model utilizes the transaction identifier to convert the row-wise representation into the column-wise representation called vertical form. The fast frequent mining algorithms prefer the vertical database format instead of horizontal due to less time consumption. The increase in transaction size causes more memory occupation and hence there is need of secondary memory creation that also leads to more time-consuming.

The analysis and the evaluation of datasets depend on the generalization and the association rules. The traditional association rule mining methods incorporate rules to extract the frequent patterns from the real-time datasets. Some approaches have the less number of rules and their usefulness is less. Hence, there is a need for effective mining algorithm to improve the usefulness with less number of rules. The increase in size of applications may cause the uncertainty in data handling. The Possible World Semantics (PWS)-based discriminative instances handling with tuples efficiently handles the uncertainty in two ways such as attribute and tuples. The confidence level of tuples-based mining results is based on the probability measures that initiate the Probabilistic Frequent Item set creation (PFI). Alternatively, the rapid changes in object locations affect the performance of attribute-based mining. Hence, the inclusion of spatial and temporal attributes increases the frequent pattern prediction.

The study of diverse pattern extraction algorithms extend the mining task into frequent graph mining. But, the combinatorial representation problem is the major concern in graph mining methods, particularly in activity monitoring applications. The automatic analysis of large field with the multiple cameras, transmitters and receivers was the major problem in activity monitoring applications. The assumption of item sets as the variable makes the frequent mining as better. One of the association rule mining algorithms such as Apriori suffers from various issues such as irrelevant item generation, unusual event handling, and more number of iterations. The inclusion of utility rise up the higher utility pattern algorithms and their implementation was difficult due to the existence of differences in transactions. The unavailability of systematic problem in high-utility mining causes the major problem. The Hidden Markov Model (HMM) utilization for high-utility pattern mining leads to difficulties during the cross-validation. Due to the unavailability of feature selection in the HMM, the frequent features are not extracted. Hence, the tree-based structures are evolved to extend the applicability of frequent mining algorithms. One of the tree-based approaches called Linear-Prefix Tree (LP-tree) based on linear growth algorithm that perform the mining tasks quickly with the multi-pass information. Alternatively, the maximum utility growth Algorithm-based tree structure assures the candidates pruning with single-pass. But the execution time, memory occupation and the rule consumption are the major problems in high-utility mining tasks. Hence, there is a need of efficient algorithm to provide the trade-off between the reduction and mining performance.

The contributions of proposed TSPTPM algorithm are listed as follows:

- The construction of heap on the basis of constant time supports the frequent-pattern mining on sublinear time basis.
- Polynomial runtime reduction minimizes the overall execution time by a heap structure that improves the efficiency of the system.
- The application of Vague Space Partition (VSP) algorithm raises the density of sequences which reduces the scanning time considerably.

The paper organized as follows: The detailed description about the related works for the frequent mining, extraction of frequent trajectory patterns in Sect. 2. The implementation process of tree based space partition of trajectory pattern mining is discussed in Sect. 3. The performance analysis for frequent trajectory pattern extraction is presented in Sect. 4. Finally, the conclusions about the application of proposed tree based bit mask search algorithm presented in Sect. 5.

2 Related Work

This section illustrates the various issues in the traditional frequent mining and item set mining algorithms. Calders et al. discussed the frequent item set mining [1] by maintaining the compact summary of extracted patterns. The non-trivial compact summary maintenance on real-time synthetic datasets extended the applicability. The typical variations of database (horizontal and vertical) involved the conversion and that leads to excess memory and time consumption. Vo et al. introduced the Dynamic Bit Vector (DBV) mining [2] for memory and time consumption reduction. The utilization of fixed size of bit vectors corresponds to the number of transactions increased the memory consumption for large size transactional database. Vo et al. utilized the lookup table approach [3] with the subsumption concept that measured the support and confidence to reduce the memory. Support and confidence measures incorporate various rules for calculation. Vo et al. utilized the combination of traditional association rules, non-redundant rules and minimal non-redundant rules called Most Generalization Rules (MGAR) [4] that offered trade-off between the time consumption and mining performance. The arrival of DBV extended the mining tasks into the image classification process. Fernando et al. constructed the Frequent Local Histogram (FLH) [5] by selection of visual primitives. With the utilization of visual cues and the global information utilization built the bag-full FLH image representations.

The emerging database-related applications such as sensor monitoring, activity monitoring, etc. leads to the uncertain database creation. Wang et al. handled the uncertain database with Possible World Semantics (PWS) [6]. The approximation based on Poisson binomial distribution function discovered the frequent item sets. Lee et al. discussed the Graph-Based Mining algorithm [7] for spatio-temporal database. The mapping graph with the created trajectory information list and the DFS traversal mined the frequent trajectory patterns. The adjacency property utilization in list creation effectively reduced the search space. But, the GBM were unsuitable in some applications like social network analysis, biochemical informatics due to the computational bottleneck problem. Cheng et al. utilized the Apriori, mining growth algorithms for reduction of exponential size and the graph patterns applicability [8]. The inclusion of multiple cameras in activity monitoring made the analysis process as the difficult one. Liu et al. discussed how the number of transmitters and receivers are reduced by using the RF tag arrays [9]. The circuit design in the RF System On chip (SOC) architecture increases the test data size. Basu et al. developed the bit mask selection and the dictionary selection technique [10] for the testing time and the memory consumption reduction. The number of resources and the arbitrary length input sequences are the major problem in string search algorithms.

Venkatesan and Ramaraj introduced the searching technique [11] to reduce the size of the scanning dataset. The novel dataset introduction and the scanning offered the less time consumption. More number of iterations, irrelevant pattern generation and the unusual events handling are the major problems observed in rule mining algorithms. Abaya introduced the set and set frequency to eliminate the non-significant keys which are the major cause for problems in rule mining algorithms [12]. The unaware of differences in transactions leads to huge size rules. Boaddh et al. introduced the scanning technique to estimate the differences and thereby the time consumption were reduced [13]. The non-determination of full generality caused the problem in frequent mining. Cameron et al. discussed the bitwise data parallelism on the structuration of SIMD and GPU architectures [14]. By using these architectures, the additional resources are scaled down. The number of entities inclusion degraded the mining performance in space and time consumption. Geetha and RamaRaj proposed the Up + Growth algorithm [15] to prune the contender entities for huge utility mining entities. Traditional algorithms do not concentrate on the compression process. But, the authors concentrate on the rise of density of bit streams that provided the acceptable compression.

USpan [16] and Sequential Pattern Tree (SPT) [17] construction that extended the applicability to sign language recognition. Each transaction item have the specific significance. But, the traditional algorithms mined without considering this issue. Vo et al. proposed the fast Frequent Weighed Item set mining (FWI) with Diffset strategy and Tree-based theorems [18]. An inefficient maintenance of numerous pointers caused the frequent mining performance degradation. Pyun et al. proposed the Linear-Prefix Tree (LP-tree) [19] based on LP-growth algorithm that utilized the minimum information for mining process and provide the linear correspondence access between the nodes. The large number of candidate generation affected the high-utility mining performance adversely. Yun et al. proposed the MU-Growth tree structure [20] for capturing the information in single-pass with more execution steps. But, the execution time, memory consumption were more in the traditional approaches. Hence, frequent mining performance improved by using the proposed method of Tree based Space Partition of Trajectory Pattern Mining (TSPTPM).

3 Tree Based Space Partition of Trajectory Pattern Mining

This section illustrates the implementation of proposed Tree based Space Partition of Trajectory Pattern Mining (TSPTPM) search algorithm in sequential processes.

1. Clustering
2. Heap Tree formation
3. Tree based search
4. Vague Space Partition

The flow diagram of proposed method is shown in Fig. 1. Initially, the dataset is passed through the clustering process. The formation of clusters is based on the user choice. For example, the user choice is 1000, then the transactions in accident dataset group the 1000 transactions in a single cluster. Similarly, 30 clusters of each 1000 entries are formed. The remaining entries are included in another cluster. Then, the heap tree is formed from the clustered transactions. In that, the transaction with maximum ID considered as root node, then, the nodes correspond to even number of transactions regarded as left nodes and the nodes with odd number of transactions as right nodes to the root node. The tree based transaction data base is applied to the Vague Space Partition (VSP) algorithm which performs scanning process. The time interval between start and end point is calculated. The VSP based prefix span is used in this paper to extract the frequent mining of given dataset reduces the computational time effectively.

3.1 Datasets

An open-source data mining library termed as SPMF offers various item sets, utility item sets for frequent mining. There are six datasets are collected from the http://fimi.cs.helsinki.fi/data/ such as chess, mushroom, connect and accident. The transactions and the number of distinct items for various datasets listed in Table 1.

For example, the sample transaction data is listed in Table 2. The Table 2 indicates the number of transactions and associated items. Here, there are 5 number

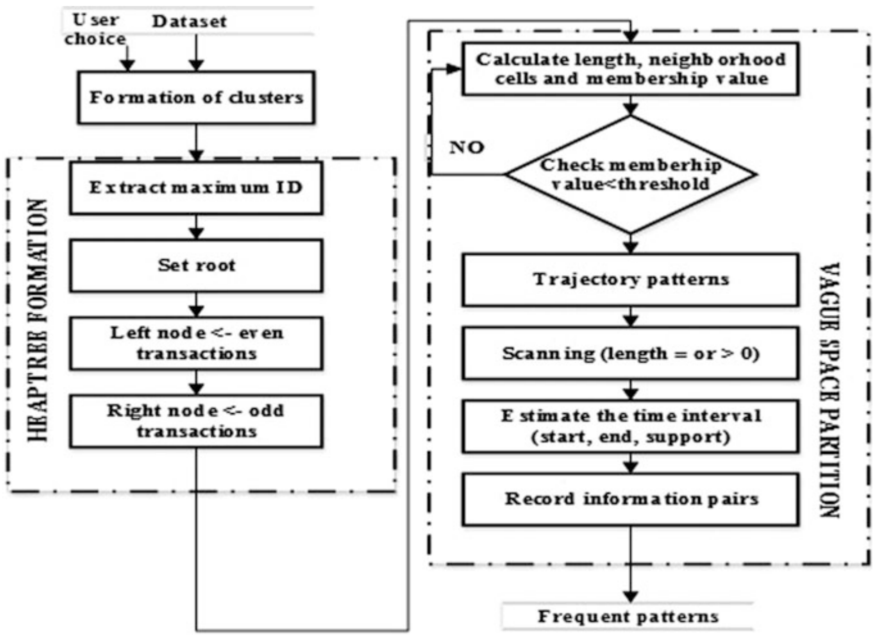


Fig. 1 Flow diagram of proposed TSPTPM algorithm

Table 1 Dataset description

Dataset name	Number of transactions	Number of distinct items
Chess	4100	86
Mushroom	8345	134
Connect	65,557	170
Accidents	345,184	478

Table 2 Transaction data

Transactions	Items
T1	1,2,4
T2	1,3,4,5
T3	2,4
T4	1
T5	1,2,3,4

of transactions are taken for analysis. A number of transactions have both 1 and 2 are 2. Then a number of transactions have 1 are 4. Therefore, the support and the confidence are calculated by

The attractiveness between the transactions is evaluated by two measures namely, support and confidence.

- *Support*
  The probability of particular transaction with respect to all required items define the support. Let us consider the two items X and Y. The support declares the probability of occurrence of X and Y in the same transaction. It is defined as

$$S = \frac{\text{Number of transactions with } X \text{ and } Y}{\text{Total number of transactions}} = \frac{2}{5} = 0.4 = 40\%. \quad (1)$$

- *Confidence*
  The degree of certainty for a given rule refers confidence. The $c\%$ transactions containing both X and Y specified by the rule $X \Rightarrow Y$. The confidence c is represented as

$$c = \frac{\text{Number of transactions having both } X \text{ and } Y}{\text{Number of transactions containing } X} = \frac{2}{4} = 0.5 = 50\% \quad (2)$$

From these calculations, it is observed that 40% of transactions contained X and Y and the 50% transactions containing X also contained in Y. The support is considered as threshold values for successive processes in this paper.

The proposed algorithm comprises following steps:

Step 1: Initialize the transactions from the datasets (Tr)

Step 2: Form the clusters $[C_n] = \text{clustering}(Tr_n)$

Step 3: Construct the tree $(T_H) = \text{Heap tree}(C_i)$

Step 4: Apply the space partition approach to the formed tree $\text{patt}_F = \text{space } \textit{partition}(T_H)$.

Step 5: Output the frequent data mining patt_F

3.2 Clustering

The first and foremost process in proposed TSPTPM algorithm is clustering. In this process, the transactions from the datasets are extracted and grouped together as a cluster of each transaction with the ID is formed. The incremental clustering algorithm to form the cluster is as follows:

Clustering

Input: *Dataset (D), number of clusters(K), threshold*
Output: Clusters(C_n)

1. *Initialize the clusters as null*
2. *Initialize the transactions as* (Tr_n)*=transactions of (D)*
3. *Set centroid of cluster as* x_i
4. *Initialize the counter as count=1.*
5. *For all* $x_i \in Tr_n$
6. *For all Cluster* $\in$ *Clusters*
7. *If* $\|x_i - centroid(Cluster)\| < threshold$
8. *Update* $centroid(Cluster)$;
9. *Assign the ID to transaction;*
10. *Count++*
11. *End if*
12. *Clusters* $(C_n) \leftarrow cluster \cup newcluster$
13. *End for*
14. *End for*

Initially, the transactions from the datasets are extracted and stored in (Tr_n). The initialization phase includes the cluster centroid, counter declaration. For each value of transactions, the new value of cluster on the basis of comparison between the centroid and the threshold value. The ID is assigned to newly added transactions to the clusters. After all transactions are grouped in the cluster with ID, heap tree formation takes place.

3.3 Heap Tree Formation

A specialized tree refers heap tree based algorithm finds the patterns with maximum ID. A comparison based sorting algorithm divided the transaction set into two regions sorting and unsorted region. The iterative procedure of shrinking makes the unsorted region into sorted region with highest ID consider as root node. The heap tree construction algorithm receives the cluster of transactions as an input. Then, for each transaction in a cluster, it extracts the ID of transaction. The transaction array, transaction with maximum ID and the no. of transactions are considered for tree construction algorithm refers to down(). The tree construction process is repeated until all the clusters are used. In to down() algorithm, the transaction with maximum ID is set it as root node. Then, the odd value of ID count is placed right to root node and the even value of ID count is at the left node to root node. The insertion of new transaction to the heap tree consists of the following processes:

- Check whether the new transaction is the last part of a cluster.
- Newly added transaction considers as the child node to the root node only if the condition is satisfied.
- Check the child node whether it is greater than or less than the root node.

- The child node will become the root node if it is greater than root node. Otherwise, it remains as it is.
- Finally, the heap tree is formed for all the transactions in the clusters.

Heap Tree

Input: Clusters(C_n), count n
Output: The tree T_H

1. *For each cluster* C_i for (i=*1….n*)
2. *for each transaction in cluster*
3. *Extract maximum ID*
4. *Start=maximum ID;*
5. *While(start* ≥ 0*)*
6. *To down (Tr, start, count-1)*
7. *Start=start-1;*
8. *End for*
9. *End for*

To down(Tr, start, end)

1. *Root=start;*
2. *While root*2+1* $\leq$ *end do*
3. *Child=root*2+1;*
4. *Swap=root*
5. *If(Tr[swap]<Tr[child])*
6. *Swap=child*
7. *If child+1* $\leq$ *end and Tr[swap]<Tr[child+1]*
8. *Swap=child+1*
9. *If Swap=root*
10. *Return*
11. *Else*
12. *Swap(Tr[root],Tr[swap])*
13. *Root=swap;*

The heap tree formation using an algorithm is explicitly implemented in MATLAB environment shown in Fig. 2. From the Fig. 2, it is observed that the node 11,689 consider as a root node. The first level of child is attached to left and

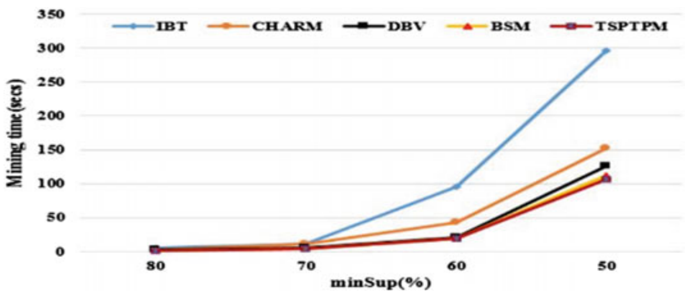


Fig. 2 Mining time analysis

right leaf nodes of root node 11,689. Then, the sequential formation, such a way that the nodes with an even number of transactions attached left of the root node and with odd values attached to the right to the root node. Finally, the constructed tree by using the heap tree algorithm is passed to the Space Partition algorithm.

3.4 Tree-Based Space Partition Search

The tree formation provided the necessary separation of item sets with maximum ID, even and odd. The tree structure contains n number of item sets. For each item set, frequent pattern extraction is provided by using the algorithm as follows:

Tree based space partition search

Input: *Item set in tree (n), transaction ID (T_{id})*
Output: *output data base (T)*

1. For k=1 to n
2. For j=1 to i
3. If (j=1)
4. t[k, j]=1;
5. Else
6. t[k,j]=0;
7. End if
8. T[1, j]= $t[T_{id}, j]$ VSP(T[1,j])
9. For each transaction in input file
10. For each item in transaction
11. Pos=(item-1)/log(n)
*12. If (item%(2*n)=0)*
*13. Item =2*n;*
14. Else
*15. Item=item%2*n;*
16. VSP $[T_{id}, pos]$;
17. End if
18. End for
19. End for

The item sets in the tree and transaction ID are given as the input to the algorithm. For each item sets vague space partition of id and position is calculated. The position value is updated by the logarithm of ‘n’ number of transactions.

3.5 Space Partition

Space Partition (SP) represents the frequent items provided the required compression and compaction of data. The item sets in each transaction referred as a frequent state if support value greater than the user-defined threshold value. The input tree which represents the item sets in transactions transformed to the numerical data in the SP process. The transaction file represented by SP algorithm converted to array for future processing.

Vague Space Partition (VSP)

Input: *Tree data base (T), membership threshold(α), neighborhood value(n), length of data base (l),*
Output: *VSD*

1. *For each value in T*
2. *For each location point*
3. *Calculate the distance between two nodes*
4. *Choose the nearest neighborhood and corresponding membership*
5. *Compare it with the membership threshold*
6. $VSD = itemswithminimummebershipvalue$
7. *End*
8. *End*
9. *Prefixspan(VSD)*

The transactions from the tree converted to the item set format. The item set format is masked by heap tree based form. The nearest neighbor from the space partition for each item set and the pre-order transaction for searching one item performed. The comparison between the obtained results with the item sets in a membership threshold value of the mining patterns. Otherwise, the comparison between the transactions continued and the support level is calculated.

Prefix span

Input: *VSD, trajectory pattern (t), length (g), Projected database (Proj(VSD))*
Output: *Frequent pattern set*

1. *If (g=0)// Scanning process*
2. *Find frequent grid cell in projected database (Pro(VSD))*
3. *For each transaction (x) in VSD*
4. *t=x;*
5. *Record information* $\in (T_{id}, time_{end}, support)$
6. *Else*
7. *Compare the time intervals for two sets (* $time_i, time_j$*)*
8. *Calculate support and append to information pair.*
9. *Accumulate total support in Proj(VSD).*
10. *End*
11. *Frequent pattern set=Information pair(* $T_{id}, time_{end}, support$*)*

The proposed VSP based on the prefix span algorithm is proposed contain scanning process. The length of *VSD* obtained from VSP is calculated. Then, check whether the length is equal or greater than zero. If it is equal to zero, the information pair that contains transaction ID, the time interval and support value are calculated. For length greater than zero, the new information pairs are calculated for each transaction. The time interval between the selected and neighborhood transaction are estimated. The support corresponding to time is also calculated. The value of the projected database updated with the new calculated support. The prediction of time interval between each transaction provides the status of frequent occurrence. The obtained information pair denoted the necessary frequent patterns.

4 Results

The Tree-based Space Partition of Trajectory Pattern Mining (TSPTPM) proposed in this paper used to obtain the frequent patterns in real-time datasets available in Frequent Mining Item (FMI) set. The mining time analysis of proposed algorithm with various number of items is presented. The implementation of TSPTPM algorithm is carried out in Intel (R) Pentium (R) CPU G 630 (2.70 GHz) with 2 GB RAM memory with Windows 7. The algorithms are coded in MATLAB R2009b. The real-time databases from http://fimi.ua.ac.be/data/ are used for implementation. The proposed algorithm is compared with the traditional algorithms such as index bit table and DBV-FI from [2], CHARM [3], Bit Stream Mask (BSM) search algorithm [13] in terms of run time and memory consumption with various minimum support values. Moreover, the analysis of rules reduction also presented as a comparison between Minimal Non-redundancy Associated Rule generation (MNAR), Most Generalization of Association Rules (MGAR) [20] with the proposed TSPTPM method. The implementation of TSPTPM algorithm is carried out in Intel (R) Pentium (R) CPU G 630 (2.70 GHz) with 2 GB RAM memory with Windows 7. The algorithms are coded in MATLAB R2009b. The real-time accident database from http://fimi.ua.ac.be/data/ are used for implementation with features described in Table 1. The evaluation of time and memory consumption are performed for various support values (minSup).

4.1 Mining Time Analysis

The analysis of mining time of proposed TSPTPM and the existing Apriori algorithms with the different cluster items are listed in Table 3. The comparison of mining time of the proposed TSPTPM algorithm with the traditional algorithms Index Bit Table (IBT) DBV [2] CHARM [3], with the different support values are graphically illustrated in Fig. 2.

From the Fig. 4, it is observed that for a support of 80% the deviation of mining time for TSPTPM withIBT, CHARM and DBV, BSM are 4.4, 2.25, 1.47 and 0.92 secs respectively. Similarly, for other support values (70, 60 and 50), the TSPTPM approach efficiently reduced the time consumption compared to traditional algorithms

Table 3 Mining Time Analysis

Cluster items	Mining Time Analysis							
	Chess		Mushroom		Connect		Accident	
	Apriori	TSPTPM	Apriori	TSPTPM	Apriori	TSPTPM	Apriori	TSPTPM
600	1.4096	0.8933	0.9513	0.6997	2.5896	0.9258	2.4506	1.3088
1200	1.4485	0.6551	0.9821	0.787	2.5576	0.985	1.3406	0.9978
1800	1.1149	0.6869	0.9426	0.7856	2.5671	1.1233	1.3598	1.1679
2400	0.9976	0.6546	0.9567	0.8567	2.5789	1.3249	1.3686	1.1341
3000	0.7976	0.6886	0.9673	0.8943	0.9456	0.8982	1.3347	1.119
3600	0.7456	0.6562	0.9736	0.8678	0.9787	0.8946	1.4198	0.986

4.2 Memory Consumption

The comparison of memory consumption of proposed TSPTPM algorithm with the traditional algorithms, Index Bit Table (IBT) DBV [2], CHARM [3], with the different support values listed in Fig. 3. It is observed that the raise of bit vector density by the bit stream mask search representation algorithm effectively reduces the memory consumption.

From the Fig. 3, it is observed that for a support of 80% the deviation of memory consumption for TSPTPM with IBT, CHARM, DBV and BSM are 12.8, 200.54, 6.04 and 5.71 MB respectively. Similarly, for other support values (70, 60 and 50), the TSPTPM approach efficiently reduced the memory consumption compared to traditional algorithms.

4.3 Rule Generation Analysis

The comparison of number of rules generation of proposed TSPTPM algorithm with the traditional algorithms MNAR, MGAR [4] with the different support values listed in Table 3. The table shows the rule generation for traditional and proposed TSPTPM algorithm. It is observed that the utilization of heap tree prior to bit stream mask search provides the reduction in number of rules. The performance analysis also presented in Fig. 4. From the Fig. 4, it is observed that for a support of 80% the deviation of rule generation for TSPTPM with MNAR, MGAR are 49.67% and 7.11% respectively. Similarly, for other support values (70, 60 and 50), the TSPTPM approach efficiently reduced the number of rules compared to traditional algorithms.

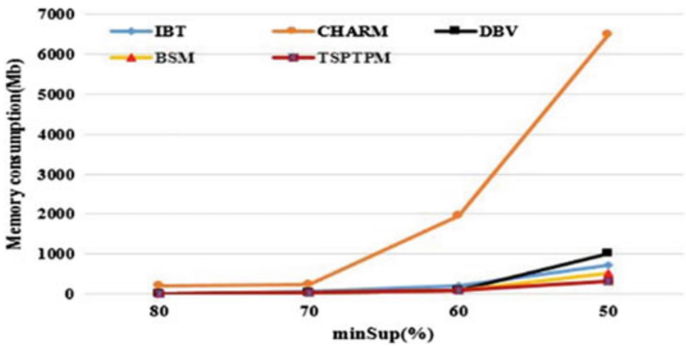


Fig. 3 Memory consumption analysis

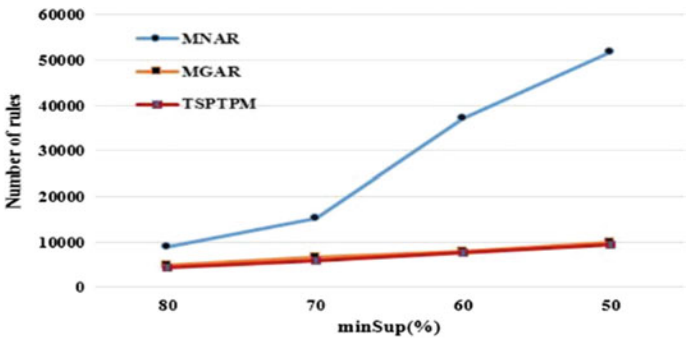


Fig. 4 Rule generation analysis

5 Conclusion

This paper focused on the real-time frequent item set mining based on Vague Space Partition (VSP) algorithms under the tree-based structures. This paper discussed the Tree-based Space Partition on Trajectory Pattern Mining (TSPTPM) algorithm for the reduction in rule, time, and memory consumption and frequent pattern extraction from the real-time datasets. The clustering grouped the transactions and the tree structure selected the root node with maximum ID. Then, the odd and even number of transactions is placed as the child nodes to the root node. The VSP algorithms utilization converted the flexible spatial patterns in sequences provides less time consumption. The optimal pattern extraction and the effective partition by heap tree structure with VSP offered rules reduction for mining. The mining time analysis of various datasets, namely chess, mushroom, connect and accident with various numbers of items in the cluster are presented. For the support value of 80%, the proposed TSPTPM algorithm reduces the mining time by 17% (support value 80%) compared to BSM. Similarly, the memory consumption of TSPTPM is reduced to 12.96% over the existing BSM. The incorporation of heap tree with the VSP algorithm in TSPTPM reduces the rule consumption by 7.11% compared to the existing MGAR method.

References

1. T. Calders, N. Dexters, J. J. Gillis, and B. Goethals, "Mining frequent itemsets in a stream," Information Systems, vol. 39, pp. 233–255, 2014.
2. B. Vo, T.-P. Hong, and B. Le, "Dynamic bit vectors: An efficient approach for mining frequent itemsets," Scientific Research and Essays, vol. 6, pp. 5358–5368, 2011.
3. B. Vo, T.-P. Hong, and B. Le, "DBV-Miner: A Dynamic Bit-Vector approach for fast mining frequent closed itemsets," Expert Systems with Applications, vol. 39, pp. 7196–7206, 2012.
4. B. Vo, T.-P. Hong, and B. Le, "A lattice-based approach for mining most generalization association rules," Knowledge-Based Systems, vol. 45, pp. 20–30, 2013.
5. B. Fernando, E. Fromont, and T. Tuytelaars, "Effective use of frequent itemset mining for image classification," in Computer Vision–ECCV 2012, ed: Springer, 2012, pp. 214–227.
6. L. Wang, D. W.-L. Cheung, R. Cheng, S. D. Lee, and X. S. Yang, "Efficient mining of frequent item sets on large uncertain databases," IEEE Transactions on Knowledge and Data Engineering, vol. 24, pp. 2170–2183, 2012.
7. J. Lee, Y.-A. Chen, and W.-C. Ip, "Mining frequent trajectory patterns in spatial–temporal databases," Information Sciences, vol. 179, pp. 2218–2231, 2009.
8. H. Cheng, X. Yan, and J. Han, "Mining graph patterns," in Frequent Pattern Mining, ed: Springer, 2014, pp. 307–338.
9. Y. Liu, Y. Zhao, L. Chen, J. Pei, and J. Han, "Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays," IEEE Transactions on Parallel and Distributed Systems, vol. 23, pp. 2138–2149, 2012.
10. K. Basu and P. Mishra, "Test data compression using efficient bitmask and dictionary selection methods," IEEE Transactions onVery Large Scale Integration (VLSI) Systems, vol. 18, pp. 1277–1286, 2010.
11. N. Venkatesan and E. Ramaraj, "High Performance Bit Search Mining Technique," International Journal of Computer Applications, vol. 14, pp. 15–21, 2011.
12. S. A. Abaya, "Association Rule Mining based on Apriori algorithm in minimizing candidate generation," International Journal of Scientific & Engineering Research Volume, vol. 3, pp. 1–4, 2012.
13. J. Boaddh, M. P. Urmila, and N. Bhargava, "A Theoretical Formulation of Bit Mask Search Mining Technique for mining Frequent Itemsets".
14. R. D. Cameron, T. C. Shermer, A. Shriraman, K. S. Herdy, D. Lin, B. R. Hull, et al., "Bitwise data parallelism in regular expression matching," in Proceedings of the 23rd international conference on Parallel architectures and compilation, 2014, pp. 139–150.
15. P. Geetha and E. R. Raj, "A Frequent Trajectory Path Mining Using Bit Mask Search and UP Growth + Algorithm," in World Congress on Computing and Communication Technologies (WCCCT), 2014 2014, pp. 17–20.
16. J. Yin, Z. Zheng, and L. Cao, "USpan: an efficient algorithm for mining high utility sequential patterns," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 660–668.
17. E.-J. Ong, H. Cooper, N. Pugeault, and R. Bowden, "Sign language recognition using sequential pattern trees," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012 2012, pp. 2200–2207.

18. B. Vo, F. Coenen, and B. Le, "A new method for mining Frequent Weighted Itemsets based on WIT-trees," Expert Systems with Applications, vol. 40, pp. 1256–1264, 2013.
19. G. Pyun, U. Yun, and K. H. Ryu, "Efficient frequent pattern mining based on linear prefix tree," Knowledge-Based Systems, vol. 55, pp. 125–139, 2014.
20. U. Yun, H. Ryang, and K. H. Ryu, "High utility itemset mining with techniques for reducing overestimated utilities and pruning candidates," Expert Systems with Applications, vol. 41, pp. 3861–3878, 2014.

Intelligent Transparent Email Security as Security-as-a-Service from Cloud

Deepak H. Sharma, C.A. Dhote and Manish M. Potey

Abstract In Security-as-a-service model the emphasis is on security being provided as a cloud service; i.e. security delivered through the cloud in place of any on-premise security implementations. In this paper we discuss Intelligent Transparent Email Security-as-a-service (IntEmailSec-aas) provided as a cloud service. The IntEmailSec-aas for email security mainly involves cleaning viruses, spam mails, phishing mails, and other types of malicious software included in email before it reaches a user's organization network. The incoming email stream is cleaned and then delivered to the organization. The outgoing email is also scanned according to defined policies and then allowed on public network. The provision for sending encrypted attachment has also been made. The encryption is done in an intelligent manner based on users' security need. The service also allows storage of emails in a central repository to be used for eDiscovery based on various parameters. The advantages of this approach are complete security management through use of multiple engines from various vendors, improved performance for client devices and no maintenance of hardware/software on the client organizations' premises. Since the IntEmailSec-aas is provided as cloud service, the management is easy and efficient through Web-based console anywhere, anytime. The access to users is provided simply through web browser. The IntEmailSec-aaS implemented as a cloud service can benefit the user with all the advantages offered by Security-as-a-service (SECaaS). A proof-of-concept (POC) of IntEmailSec-aaS is implemented and evaluated. This POC is implemented in public cloud environment. The freshness in the approach is that this POC is integrated with other Security-as-a-service options to provide a portal through which various security services can be provided.

D.H. Sharma (✉) · M.M. Potey
Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai 400 077, India
e-mail: deepaksharma@somaiya.edu

M.M. Potey
e-mail: manishpotey@somaiya.edu

C.A. Dhote
Department of Information Technology, PRMIT & R, Amravati 444601, India
e-mail: vikasdhote@rediffmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_54

Keywords Cloud computing · Email security · Security-as-a-service
Email security-as-a-service

1 Introduction

Cloud computing is a new and fast evolving model, with new improvements and capabilities being done regularly by research community around the world. Cloud computing has its roots in large-scale distributed computing technology. It is in fact an extension of grid computing, distributed computing, and parallel computing [1]. Nicholas Carr has equated the rise of cloud computing in the information age to electrification in the industrial age. Carr has argued that in the emerging future different organizations will simply plug into cloud (computing grid) for the computing resources they need [2].

Security-as-a-service [3] model emphasizes on security being delivered as a cloud service; i.e. security provided through the cloud in place of on-premise security implementations. The security-as-a-service model can also add to the functionality of existing on-premise implementations by working with them in hybrid manner. Email Security-as-a-service primarily involves cleaning viruses, spam mails, phishing mails, and other malicious software included in email before it reaches the user's organization network. The incoming email stream is transparently cleaned and then delivered to the client's organization. The provision for sending encrypted attachment has also been done in an intelligent manner based on users' need.

The advantages of this approach are complete email security management through use of multiple engines from various vendors, improved performance for client devices and no maintenance of hardware and software. The management is also easy and efficient through Web-based console anywhere, anytime. It is also often used with outgoing email for data leakage prevention. It protects confidential information in email, defends against leaks of this secret information and can ensure compliance with common international and industry data protection regulations. It can also provide Security-as-a-service for email backup and archiving. This service simplifies searching of data emails and it also eases the archive management and administration with a cost-effective solution that doesn't require any additional hardware or software.

This service allows for storage of organization's emails and attachments in a centralized repository. This centralized repository allows effective management by indexing and search on a number of different parameters, including date, recipient, sender, subject and content known as eDiscovery. Thus the organizations can protect their mission-critical email infrastructure. The threats can be from outside including spam, phishing, malicious software, unpredictable email volumes and other forms of objectionable or dangerous content before it hits the enterprise gateway. The Email SecaaS services are used to reduce email threats arising from viruses, phishing attacks, spam, denial of service and other operational disruptions. It allows users to send encrypted attachment using algorithms of users' choice. The algorithm can be chosen intelligently depending on multiple levels as per low, high or custom level.

Email Security controls [4, 5] generally focus on:

- Malware protection in incoming and outgoing mails
- Filtering of spam
- Content filtering
- Archiving of emails
- Encryption
- Reporting

In this paper an Email Security as a service (EmailSec-aaS) framework is described and POC implementation is discussed. In particular this IntEmailSec-aaS is an on-demand portable, and available pay-per-use cost model. The paper addresses various issues regarding security delivered as cloud service. This paper discusses various issues related to email security in following sections. Section 2 introduces the concept of IntEmailSec-aaS and discusses related work. Section 3 describes scope, its main components and a proof-of-concept (POC) prototype implementation of IntEmailSec-aaS in public cloud. Section 4 evaluates the prototype IntEmailSec-aaS. Finally, Sect. 5 concludes the paper and discusses future work.

2 Email Security Related Work

According to paper [4], the Email security services can be one of two service models: fully outsourced model and enterprise augmentation model. In the first service model the entire mailbox and user interface is outsourced to a cloud security provider. In the second service model the security process is added to an existing enterprise email implementation. In a fully outsourced model, it is responsibility of security service provider to handle all threats in the email as a channel (spam, phishing, malware propagation, etc.) The security service provider also provides a user interface for email and assists an organization's end users also. In the enterprise security extension model, an existing on-premise email deployment is improved by providing additional cloud-based security services and functionalities.

Email SecaaS [4] addresses many requirements for information security. These requirements include:

- sending and receiving email as per standard protocols
- to prevent malware infections via email
- Removing of unwanted spam messages

- Authentication of email users
- Security of clients and secure email access by remote devices
- Data Loss Prevention tools integration
- Storage of email records
- Encryption of emails with secure key management
- Management of emails and logging service

A related work has been done by authors of paper [6]; the paper discusses intelligent cloud based email encryption and Decryption System, the authors have proposed a framework for improving security of cloud-based email messages. The goal is to encrypt email mail messages from users' mailbox before they are actually sent. Their encryption algorithm helps in protecting users against email interception, phishing attacks, relaying of previous messages, spoofing, eavesdropping and provides high level of privacy. Their intelligent cloud based machine encryption and decryption system uses machine compressed data, combined machine-generated keys that use algorithms that are aligned with user email address and login details. It combines machine-generated keys encryption with public key encryption. So, each message is encrypted using machine encryption algorithm that requires combined machine-generated keys.

In a similar manner the authors in paper [7] have described a Secure Email Attachment as a Service (SeaS) with an aim to provide scalable and secure storage solutions for email attachments. Their model combines the traditional security of corporate email servers while offloading the email attachments storage and management to the cloud in a secure manner. A Secure Virtual Diffused File System (SVDFS), an IDA-based secure file system is used to secure the data in the cloud. Client-Server architecture is described to guarantee confidentiality, integrity and availability of data in cloud.

In paper on Implementation of Secure Email Server in Cloud Environment [8], the authors have focused on the problem of email contents disclosure, and discussed a secure mail server implemented into cloud as IaaS (Infrastructure as a Service). Their model provides Security by tuning Transport Layer Security (TLS), and SMTP-AUTH which use Simple Authentication and Security Layer (SASL) as a security mechanism and platform. They have focused on implementation and configuration of a secure email server in Linux platform (Ubuntu OS) to used Local-based Email Transfer Agent.

Another relevant work has also been done in Identification of Spam Email Based on Information from Email Header [9], they have focused on several features in the email header. These features have been used to identify and classify spam messages efficiently. The fields considered by the authors from the email header are Received field, From field, and Number of Receivers can be used to identify probable spam message in efficient manner.

3 Int-Emailsec-AAS POC Implementation

The architecture of an Email Security-as-a-service mainly involves cleaning viruses, spam, phishing emails, and other malware included in email before it reaches users' organization's network. The incoming email stream is cleaned and then delivered to the organization as shown in Fig. 1.

A proof-of-concept prototype of IntEmailSec-aaS is implemented in public cloud. The IntEmailSec-aaS framework can be applied in all types of cloud implementations. All Email IntSecaaS components are constructed and build in the form of virtual machines in the public cloud. The on-demand elasticity, portability, use of security mechanisms from various vendors are some of the characteristics of IntEmailSec-aas. All these features can be applied by starting the VM instances dynamically on the go.

The focus here is on Email Security to be delivered as a cloud service; i.e. security provided through the cloud. The POC system architecture is shown in Fig. 1, the main components of the system are IntEmailSec-aaS core—the core

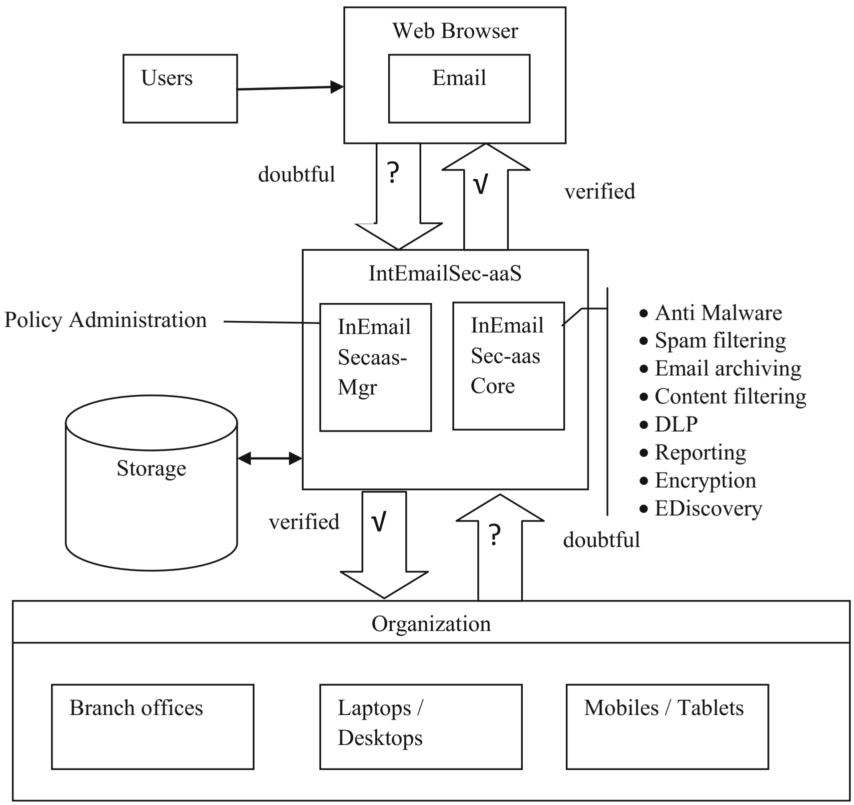


Fig. 1 Implementation of POC EmailSec-aas

IntEmailSec-aaS functions have been implemented in this module, IntEmailSec-aaS manager—managerial functions like policy administration, and Intelligence have been implemented in this module. On the one side of IntEmailSec-aas is public internet from where email comes to the organization and on the other side of IntEmailSec-aaS is the Client Organization. As demonstrated in the figure when email reaches IntEmailSec-aas it is in doubtful state. The IntEmailSec-aas core module cleanses it by applying all policies and then delivers it client organization. A client organization may have different types of users' viz. at branch offices, desktops, laptops and mobile phones etc. The emails leaving from organization network can also be scrutinized by applying appropriate policies before going on public internet.

The IntEmailSec-aas core module has other functionalities like email archiving, content filtering, Data Loss Prevention, reporting, encryption as per the policies defined. A storage unit as central repository is connected to IntEmailSec-aas for archiving and eDiscovery of emails stored.

The users can exercise the service by simply using a web browser. There is also a client side application that allows the client to encrypt the files before they are attaching them to their emails. This helps achieve protection of Data-in-Transit. The IntEmailSec-aas Core provides options to users to choose between multiple encryption levels. In the prototype implementation symmetric encryption algorithms are used. The key for encryption/decryption is generated from user's passphrase by using inbuilt API of PHP. The key management and distribution issues in symmetric key cryptography have been minimized by using the passphrase for generation of key. The encryption/decryption is done at file level.

The implementation has been done in form of a web server running under windows OS Virtual machine in a public cloud setup. The IntEmailSec-aas Core and IntEmailSec-aas Manager have been implemented in separate Virtual machines. In this way the entire Email Security functionality can be managed. It is provided as a cloud service to the client through browser and different types of devices (Desktops and Mobiles devices etc.) can also access it. The snapshots of the POC implementation are shown in Fig. 2.

4 Email-SECaaS Evaluation

Several experiments were conducted to evaluate the effectiveness of our proof-of-concept prototype of IntEmailSec-aaS in public cloud. The evaluation has been done on the basis of discussion given in our paper [10, 11] and a white paper [12]. The following criteria have been considered for the evaluation purpose:

- Reliability: the service will be provided in form of multiple web servers running in the cloud environment. The redundancy of servers will lead to high reliability and high availability to the clients. The POC was tested in form of two web servers to provide uninterrupted IntEmailSec-aaS services to the clients.

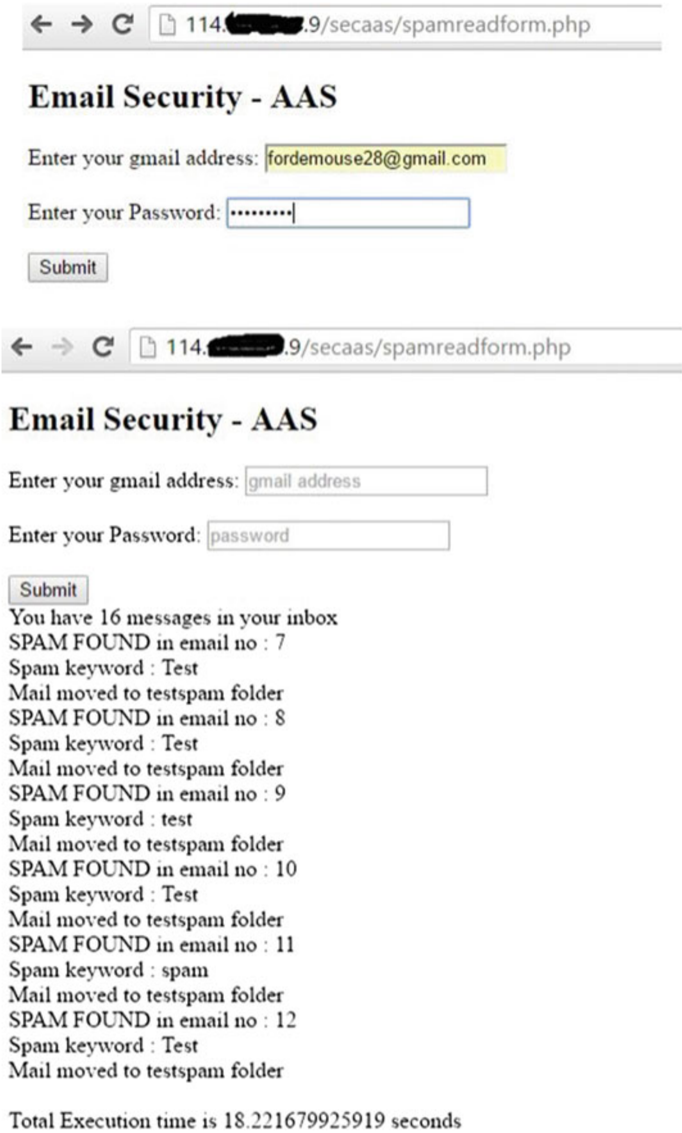


Fig. 2 Snapshot of EmailSec-aas POC

- Effectiveness: to make the service more effective, the core module handles multiple functionalities like Anti-malware, spam filtering, content filtering, Archiving, Encryption and eDiscovery. The clients can use any device and access the service through browser.
- Performance: the performance was tested by comparing the average time taken for Email security as cloud service w. r. t. standard legacy Email security

mechanism for a normal web server. The overall overhead also depends on the traffic in public cloud, but it does not increase by more than 20–25%, which is fairly good given the advantages it offers over legacy systems.

- Flexibility: the solution can work with existing legacy systems as well. Since the POC implementation is in the form of PHP programs it can easily work with legacy Encryption systems. It can provide more flexibility to customers to choose varying functionalities of security as per their need.
- Control: the client uses a web browser to access the service and it can be accessed from various devices viz. desktops and handheld mobile devices. A central portal is provided through which all the policies can be easily administered.
- Privacy and Security: the IntEmailSec-aas filters all doubtful inbound email traffic before entering clients' organization and all outbound doubtful email traffic before going on public internet. This filtering is done based on the policies defined. This ensures the privacy and security of Users' data.
- Cost of ownership: the cost of ownership is borne by the cloud security service provider. The client does not invest in anything in on-premise solution. The client will have to pay only on the basis of pay-per-use model. Since IntEmailSec-aaS is available as cloud service it is only charged to customer in form of Operational Expenses (OPEX) model.

5 Conclusion and Future Work

In this paper, we introduced IntEmail SecaaS, in the form of a framework that enables the cloud service provider to provide email security as a cloud service in public cloud. The freshness in the approach is that the POC is integrated with other Security-as-a-service options to provide a portal through which various security services can be provided. The primary functionalities implemented in POC are spam filtering, and attachment Encryption. IntEmail SecaaS is compatible with prominent cloud features including portability, elasticity, and pay-per-use service. The approach was implemented as a collection of VMs in public cloud to comply with the cloud model. The users exercise the use of service by simply using browser; it provides access for various devices from desktops to mobile clients. This solution can work with existing on-premise platform based implementations in a hybrid manner to enhance their security capabilities. The advantages of this approach are complete security management through use of multiple engines from various vendors, better performance for client devices and no maintenance of hardware and software. The management is easy and efficient through Web-based console anywhere, anytime.

The future work will involve enhancing the functionality of POC by adding more security features. The system availability, reliability and performance can be enhanced by creating replicas of core VM to distribute the traffic load and to prevent single point of failure.

References

1. Shuai Zhang, Shufen Zhang, Xuebin Chen, Xiuzhen Huo, "Cloud Computing Research and Development Trend", 2010 Second International Conference on Future Networks, IEEE 2010.
2. Nicholas Carr,"The Big Switch Rewiring the world from Edison to Google", W.W. Norton & Mar and C. Y. Law, "Secure Email Attachment as a Service (SeaS)," *2015 International Conference on* Co. January 2008.
3. Tim Mather, Subra Kumaraswamy, and Shahed Latif, 2009, Cloud Security and Privacy. An Enterprise Perspective on Risks and Compliance, O'Reilly Media, 336.
4. CSA SecaaS Implementation guidance: Email Security September 2012.
5. Cloud Security Alliance, SecaaS Defined categories of service 2011.
6. T. Ayodele and D. Adeegbe, "Cloud based emails boundaries and vulnerabilities," *Science and Information Conference (SAI), 2013*, London, 2013, pp. 912–914.
7. K. K. Mar and C. Y. Law, "Secure Email Attachment as a Service (SeaS)," *2015 International Conference on Cloud Computing Research and Innovation (ICCCRI)*, Singapore, Singapore, 2015,pp. 106–111.
8. N. K. M. Madi, S. Salehian, F. Masoumiyan and A. Abdullah, "Implementation of secure email server in cloud environment," *Computer and Communication Engineering (ICCCE), 2012 International Conference on*, Kuala Lumpur, 2012, pp. 28–32.
9. S. Bin Abd Razak and A. F. Bin Mohamad, "Identification of spam email based on information from email header," *Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on*, Bangi, 2013, pp. 347–353.
10. Deepak Sharma, Dr. C A. Dhote, Manish Potey, 'Security-as-a-Service from clouds: A survey' IIJC Vol 1 Issue 4, October 2011.
11. Deepak Sharma, Dr. C A. Dhote, Manish Potey, 'Security-as-a-Service from Clouds: A comprehensive Analysis', IJCA Volume 67-Number 3, April 2013.
12. Websense white paper, Seven Criteria for Evaluating Security-as-a- Service Solutions, 2010.

A Novel Hardware Selection Algorithm for Aerial Vehicles with Multi-rotors

K.R. Narahari, B.S. Aruna Prabha, M.S. Goutham, V. Rohit, Damodhar S. Kamath and P. Vaishak

Abstract From recent trends, Unmanned Aerial Vehicles have been pitched into the world of aerial surveillance; especially where humans could not step forward due to security and safety issues. The design of such vehicles must possess rugged and robust hardware specifications. There has been a heap of generic work done on the hardware selection criteria which is scattered over different sources and not organized in a standard chain model. The main contribution and novelty of the present work is to rescreen the truncated flow into an algorithm in a linear and conditional model; to efficiently decide on the hardware components such as Propellers, BLDC Brush Less DC Motors (BLDC) motors and electronics related to Drive assembly of any Unmanned Aerial Vehicle (UAV) with *multi-rotors*. The results at each stage of the presented model have been verified for the best fit with the practical scenarios with error approximations. The metrics used for the work comprises of Static thrust Static thrust (Initial acceleration), RPM and Power ratings of drive assembly.

Keywords Unmanned Aerial Vehicle (UAV) · Static Thrust Calculator (STRC)
Optimum Propeller Calculator (OPC) · Flight Time Calculator (FTC)
Electronic Speed Controller (ESC) · Brush Less DC Motors (BLDC)
Static thrust (Initial acceleration) · Dynamic thrust (Aerial speed)

K.R. Narahari · B.S. Aruna Prabha (✉) · M.S. Goutham · V. Rohit · D.S. Kamath
P. Vaishak
Department of EEE, Nitte Meenakshi Institute of Technology,
Bangalore 560064, India
e-mail: bs.arunaprabha@gmail.com

K.R. Narahari
e-mail: naraharikr1994@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_55

1 Introduction

Ever since the dawn of technology is in the proliferation stage, all semiconductor incorporated devices have been designed smart enough to take independent decisions based on real-time circumstances, meanwhile sacrificing their functional capabilities to reduce human efforts and risks for a given application. UAVs have become a common area of interest for diversified engineering and non-engineering platforms due to their rugged performance and due to their less dependency on human control by understanding circumstances playing around them which not only reduces human efforts at the implementation stage but also helps in reducing the amount of risk to be handled in situation such as enemy surveillance [13], disaster management [5], power distribution sectors and spying of minerals and ores [13, 2], etc. A statistical data analysis has detected their growth as an exponentially increasing characteristic with huge number of UAVs hanged up in the space for the connected world [14]. Presently wide range of path-finding, surveying, and intelligent monitoring algorithms have been implemented with these vehicles in almost every engineering discipline.

This paper has been organized into six sections. The first section deals with introduction of UAVs, second section discusses parameters used for the algorithm and third section deals with mathematical modelling of an UAV with multi-rotors. A detailed algorithm has been pitched in the fourth section and their corresponding results and characteristic comparisons have been discussed in fifth section. The last section concatenates the presented algorithm with present world circumstances and thus justifying as a Conclusion.

2 Comparison Parameters

Electronic Speed Controller (ESC) [3] is a combinational electrical circuit embedded with logical switches to convert direct supply (received from battery) to implicit 3 phase supply [16, 6] to keep the motor running continuously.

Static Thrust Calculator (STRC) [15] is a sophisticated online tool which takes inputs as propeller diameter, propeller pitch and number of revolutions per minute of propeller and generates information regarding the static thrust and power rating of BLDCBrush Less DC Motors (BLDC) [20] motors.

Optimum Propeller Calculator (OPC) [10] is an online calculator taking inputs as rated and maximum RPM [16] of a BLDC motor and generates useful information about the different combinations of propellers that can be used.

Flight Time Calculator (FTC) is an optimistic online calculator used to check the best fitness of the result with theoretically calculated flight time.

E-Calc [4] is an online platform to check the compatibility and reliability between selected individual hardware systems.

3 Mathematical Modelling

The following mathematical modelling of an UAV with multi-rotors bridges the relation between propeller specifications, RPM, magnitude of static and dynamic thrusts [20] at variable airspeed.

From Newton's second law of motion we know that "*Force equals to rate of change of momentum*" i.e.,

$$P = \frac{\mathrm{d}}{\mathrm{d}t}(mu) \tag{1}$$

where, P is the Thrust force in newton.

As we are concentrating on static thrust force, the air accelerated molecule's rate known as mass flow rate [20] ($\dot{m}$). The amount of mass offered by air molecules change with respect to time is assumed to be constant. The velocity is taken as difference between the velocity of air molecules (u_{ac}) which started out stationary and the velocity of accelerated air molecules (u_e) due to propelling action. The Eq. 1 becomes,

$$P = \dot{m} * \Delta u$$

$$P = \dot{m}(u_e - u_{ac}) \tag{2}$$

Mass flow rate of given propeller which is forming a circular disc with diameter 'y' (in inches) and swept area A in the air with an air density ρ is given by,

$$\dot{m} = \rho A u_e$$

$$\dot{m} = \frac{\rho \pi y^2 u_e}{4} \tag{3}$$

Substituting Eq. 3 in Eq. 2 we get,

$$P = \frac{\rho \pi y^2}{4}(u_e^2 - u_e u_{ac}) \tag{4}$$

From Eq. 4, u_e is assumed as nearly equal to the pitch velocity of propeller and neglecting second part of difference term for time being, the simplified equation is obtained as follows.

$$P = 1.225\frac{\pi(0.0254 * y)^2}{4}\left(N * 0.0254 * T * \frac{1\mathrm{min}}{60\,\mathrm{s}}\right)^2 \tag{5}$$

where, N is number of revolutions swept by the prop and T is the pitch of the propeller in inches. To make the Eq. 5 to work for practical conditions, two numerical coefficients [20], calculations related to dynamic thrust force and parameters related to the physical dimensioning of a propeller [20] are added. One of numerical coefficient corresponds to numerical constant and the other to power constant.

The final simplified equation thrust force is given by,

$$P = 1.67 * \left[4.392399 * 10^{-8} * \frac{N * d^{3.5}}{\sqrt{T}} * \left(4.23333 * 10^{-4} \ * NT - u_0 \right) \right] \quad (6)$$

where, d is the diameter of the propeller in inches and u_0 is the aerial speed.

From Eq. 6 it can be observed that a relation between magnitude of Static/Dynamic thrust and propeller specification has been generated for required number of revolutions. Also, a factor of 1.67 is multiplied with P, so as to map the theoretical value on to the empirical value which considers air flow profile on helical twist geometry [14] of the propeller and assumptions related to compensation between velocity difference [14] at both ends of the propeller over a period of time and hence above equation [Eq. 6] gives an increased accuracy and the result that can be confidently implemented for further calculations.

4 Noval Algorithm

In the present section, a general and unique algorithm has been proposed. For better understanding of algorithm at every stage, the algorithm has been explained in detail by taking an example of an UAV having a propeller as *6" × 4"* with 6 rotors and results pertaining to the same has been discussed at same stage and the algorithm is not strictly restricted to the assumed inputs.

4.1 Choosing Propeller Specifications

The important parameters which define any aerial vehicle's performance in accordance with hovering are the forces acting on it. As these forces are created by the spin offered by a propeller; they play a vital role in initially starting any aerial vehicle. The important parameters which have to be taken care are propeller diameter [20, 14] and propeller pitch [20, 14].

- A high pitch of propeller corresponds to high initial acceleration as it moves more in free space on a 360° revolution in the forward direction and leads to less aircraft speed.
- On the other hand, a prop with less pitch corresponds to less initial acceleration and high aircraft speed.
- By looking at the trade-off between static and dynamic thrusts, a proper pitch has to selected by considering available market resources.

- In general, a propeller with high pitch should be chosen so as to put the vehicle faster and quicker in air; which otherwise consumes more power from battery leading to lesser speeds and increased flight time.
- The diameter of propeller should be selected based on the physical dimension of the copter, availability and propeller stalling. Apart from these constraints there are not much limitations on selecting diameter.

4.2 Determining Total Effective Thrust

The spinning action of the propeller accelerates mass flow rate of air molecules and creates a difference between velocities of the same. As the accelerating action creates mass difference in the air molecules, a lifting force [6] is generated which can put the vehicle onto air which is generally measured as mass and collectively known as Thrust. The mathematical equation accounting for this particular factor has been multiplied by numerical constant so as to ensure higher endurance of vehicle and is given by the following relationship.

$$F = \frac{W}{n} \times 2 \tag{7}$$

where, W is weight of UAV including payload and n is no. of rotors

4.2.1 Hand-Calculation

Considering total payload to be carried by the UAV as 1.25 kg and assuming n as 6 (given) and substituting the inputs in Eq. 7 we get,

$$F = \frac{1.25 * 2}{6} = 0.4167\text{kg}$$

$$F = 0.4167 * 9.81$$

$$F = 4.0878N \tag{8}$$

4.3 Determining Propulsion Speed in RPM

This step revolves around calculating a minimum number of revolutions required by drive assembly to hover the vehicle in air. By validating and substituting static thrust force from Eq. 7, the number of revolutions required for the vehicle to take off the land at zero aerial speed is determined.

$$RPM = \sqrt{\frac{P(\text{in newtons}) * 1.67 * \sqrt{T}}{18.596976 * 10^{-12} * T * d^{3.5}}} \tag{9}$$

4.3.1 Hand-Calculation

Substituting propeller pitch (*T*) as 4″ and propeller diameter (*N*) as 6″ and *P* from Eq. 8 in Eq. 9 we get,

$$RPM = \sqrt{\frac{4.0878 * 1.67 * \sqrt{4}}{18.596976 * 10^{-12} * 4 * 6^{3.5}}}$$

$$RPM \cong 18,625\,\text{rev/min} \tag{10}$$

4.4 Plotting Thrust Versus Aerial Speed

A plot of static and dynamic thrust values is generated for different values of aerial speeds by using conventional tools such as excel spreadsheet [14] using the relation as mentioned in Eq. 6

From the Fig. 1, x- intercept is depicting the benchmark point as 70 mph which indicates the speed at maximum thrust [20] for hovering with good stability and the y-intercept points to a magnitude of static thrust value, i.e., 0.6958 kg. The gradual decrease represents a linear drop in static thrust with respect to increasing aerial speed as the u_e will be more at flight time when compared to u_{ac}

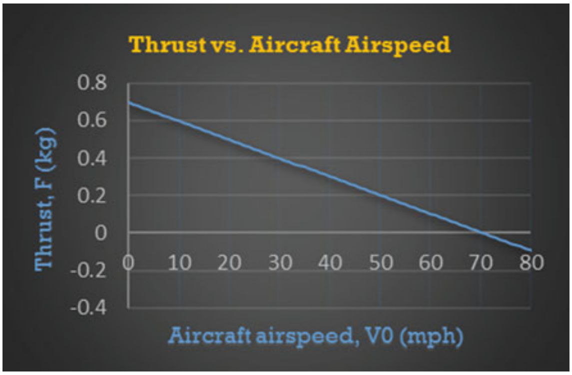


Fig. 1 Thrust versus Airspeed characteristics of a 6″ × 4″ propeller spinning at 18,623RPM

4.5 Validating Static Thrust Using STRC

This part of algorithm deals with validating the magnitude of static thrust result from the STRC and verifying the same value with magnitude of spreadsheet [14]; to check the error deviation when compared to the result of online calculator for effectively continuing further steps. The error deviation must fall within 20% for better synchronization of results from discrete platforms.

From Fig. 2, it can be noticed that the static thrust value is 0.56 kg and the spreadsheet calculated value is 0.6958 kg.

$$\%\text{Error} = \frac{0.6958 - 0.56}{0.6958} \times 100$$

The above numerical produces an error deviation of 19.51% which is falling within the expected range.

4.6 Determining Power Rating of Drive Assembly

This stage involves estimation of the mechanical power required per motor to lift the assumed amount of thrust by using STRC as shown below.

The highlighted context from Fig. 3, corresponds to a mechanical power that has to be multiplied by a factor of 1.55 as shown in Eq. 11 to derive desired mechanical power from the motor.

$$\text{Elec.Power} = 1.55 \times \text{Mech.power}(\text{in}\,W) \tag{11}$$

The numerical coefficient 1.55 has been estimated by performing numerical methods of analysis.

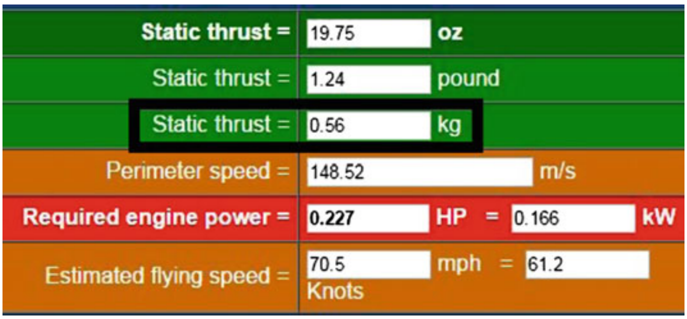


Fig. 2 Verification for magnitude of Static thrust determined from Fig. 1 using STRC

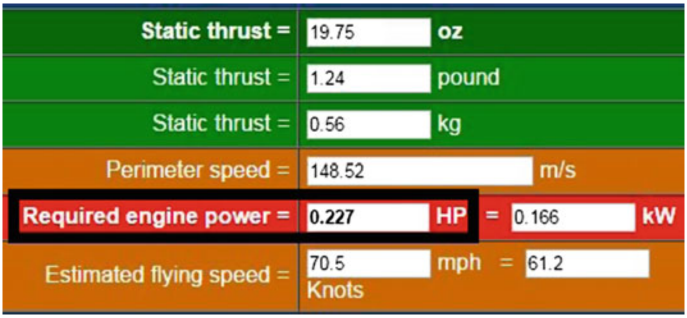


Fig. 3 Evaluation of Mechanical power required for a rotor running at 18,623RPM using STRC

4.6.1 Hand-Calculation

$$\text{Electrical Power} = 1.55 \times 0.166$$

$$\text{Electrical Power} \cong 260W$$

By looking at the electrical wattage and rated RPM a suitable motor is selected from the available market sources.

4.7 Compatibility Check for Selected Propeller Specifications

At this standing it is crucial and better to have a respectable corrrelation between the calculated data to approve that temporarily locked parameter at different stages will correspond to build an outstanding subsystem by using OPC [10].

From Fig. 4, a surmise can be confirmed that calculated systems could be systemized as a coherent subsystem dragging all calculated parameters under single platform.

4.8 Determining Li-Polymer Battery Specifications

This section deals with determining mAh rating [3] and battery voltage. For confident estimation of battery specs, a significant flight time has to be assumed and to substitute in the following mathematical relation.

$$\text{mAh} = \frac{t * \{(X * n) + Y\} * 1000}{60} \tag{12}$$

Optimal Propeller Configurations									
Diameter	Pitch	RPM	Perimeter Speed	Thrust kg	Thrust pound	Thrust oz	Fly speed km/h	Fly speed mph	Fly speed knots
7	3	19241 RPM	179 m/s	1.08 kg	2.38 pound	38.14 oz	88.0 km/h	54.7 mph	47.5 knots
6.5	3.7	19806 RPM	171 m/s	0.85 kg	1.88 pound	30.04 oz	111.7 km/h	69.4 mph	60.3 knots
6.5	2.9	21481 RPM	186 m/s	1.00 kg	2.21 pound	35.34 oz	94.9 km/h	59.0 mph	51.3 knots
6.3	4	20119 RPM	168 m/s	0.78 kg	1.71 pound	27.35 oz	122.6 km/h	76.2 mph	66.2 knots
6	5	19932 RPM	159 m/s	0.63 kg	1.38 pound	22.09 oz	151.9 km/h	94.4 mph	82.0 knots
6	4	21471 RPM	171 m/s	0.73 kg	1.60 pound	25.63 oz	130.9 km/h	81.3 mph	70.7 knots
6	3.5	22448 RPM	179 m/s	0.79 kg	1.75 pound	28.02 oz	119.7 km/h	74.4 mph	64.7 knots
6	3	23632 RPM	188 m/s	0.88 kg	1.94 pound	31.05 oz	108.0 km/h	67.1 mph	58.3 knots
5.7	3	25304 RPM	192 m/s	0.82 kg	1.81 pound	29.00 oz	115.7 km/h	71.9 mph	62.5 knots
5.5	4.5	23184 RPM	169 m/s	0.60 kg	1.32 pound	21.10 oz	159.0 km/h	98.8 mph	85.8 knots
5.5	4	24112 RPM	176 m/s	0.65 kg	1.43 pound	22.82 oz	147.0 km/h	91.3 mph	79.4 knots
5.25	4.75	24227 RPM	169 m/s	0.54 kg	1.20 pound	19.13 oz	175.4 km/h	109.0 mph	94.7 knots
5	5	25417 RPM	169 m/s	0.49 kg	1.08 pound	17.32 oz	193.7 km/h	120.3 mph	104.6 knots

Fig. 4 Cross-verifying specifications of calculated parameters of drive assembly

Fig. 5 Verifying Flight time calculated from Eq. 12 using FTC

where, t is Flight time in minutes, X is Average current drawn by a single motor in A and Y is Effective total current drawn by other subsystems except motor assembly in A

The Eq. 12 ends up generating required magnitude of battery capacity [3] and suitable battery voltage has to be chosen depending on the datasheet of motor from Sect. 4.6.

4.8.1 Hand Calculation

From selected motor datasheet, $X = 10\text{A}$ @ 10 V and consider Y as 2A. It has decided on required endurance as 8 min and hence battery capacity is decided by using Eq. 12

$$\text{mAh} = \frac{8 * \{(10 * 6) + 2\} * 1000}{60}$$

The above equation returns a magnitude approximately equal to 8000 mAh and voltage must be greater than 10 V.

Figure 5 confirms that selected power bank will deliver the required flight time as assumed at the starting of this section.

4.9 Estimating "I" Rating of ESC

Calculating the current rating of an ESC is very important as it bridges the power flow between battery and motor drive assembly. As the name indicates, the function of controller is to amplify the current [16, 3] and convert direct source to pulsating alternating current [16, 6] at regular intervals. The rated current of ESC is given by,

$$\text{Amperage of ESC} = \frac{\text{Motor wattage}}{\text{Battery volts}} \tag{13}$$

4.9.1 Hand-Calculation

Selected Brush Less DC Motors (BLDC) motor's datasheet comprehends its wattage as 260 W and battery voltage is identified as say 11.1 V i.e., more than 10 V

$$\text{Amperage of ESC} = \frac{260}{11.1} = 23.42A$$

The above result can be approximated and continuous current rating of Electronic Speed Controller (ESC) can be taken as 30A.

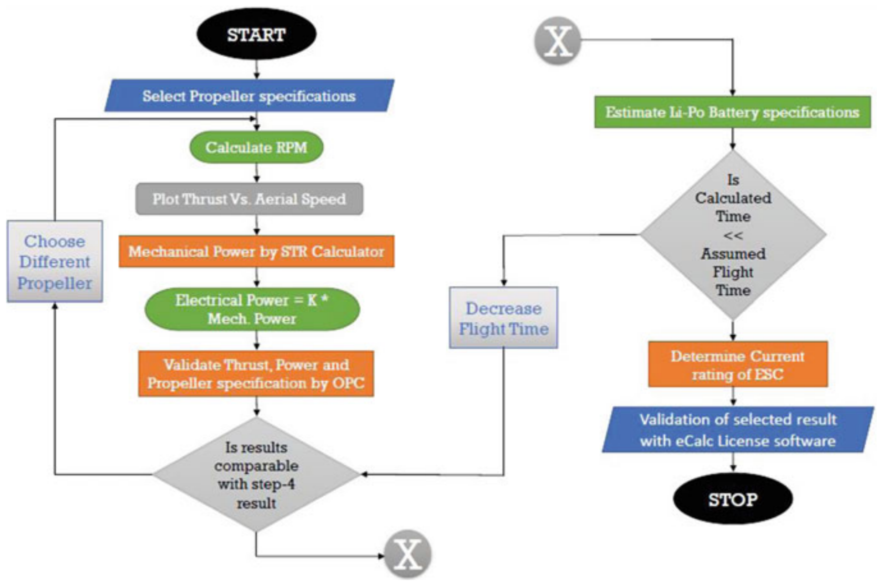


Fig. 6 Flowchart model for Hardware selection of an UAV with multi-rotor

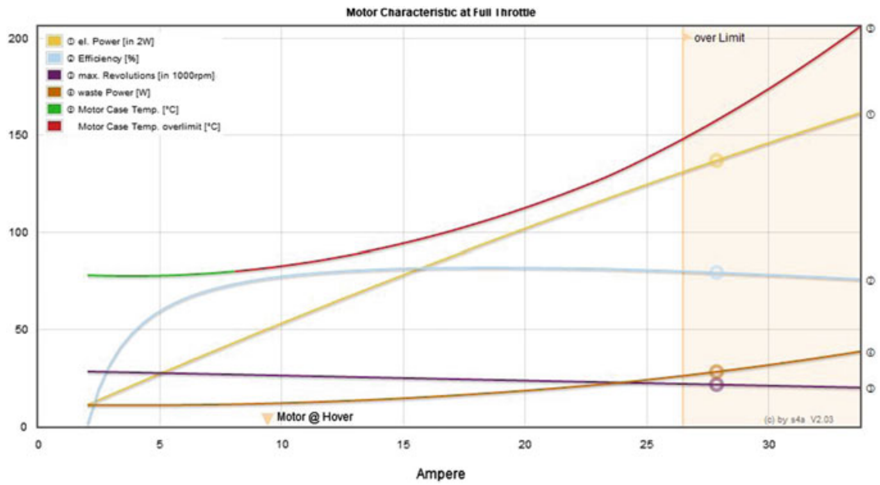


Fig. 7 Depicts the relation between electrical & leakage power, motor operating & over-limit temp., efficiency of vehicle, number of revolutions with reference to current

Figure 6 represents pictorial presentation of generic algorithm which is divided into conditional blocks to ensure linear flow without any inconvenience. All the blocks are already discussed in the algorithm sect. 4.1–4.9

5 Results

Till last section, every hardware component has been selected and verified as individuals according to algorithm discussed in Sect. 4. Since the work is concentrated on the process of building an optimistic algorithm for hardware of a multi-rotor UAV, it's time to combine every individual hardware component and to check its performance and stability characteristics as a system. A set of cumbersome mathematical equations which can correlate performance of every selected hardware component with each other will be required to accomplish the same. In the present work a common licensed software tool known as *e-Calc Multirotor* [4] has been used to verify the characteristics of the entire vehicle and also to gather information about precautionary measures which has to be taken for the optimistic flight (Fig. 7).

6 Conclusion

This paper explains the need and importance of selection of hardware for UAV'sUnmanned Aerial Vehicle (UAV). In the present scenario, first priority has been given to implement the UAVs with various algorithms for diverse applications. Indeed, it is also important to give a second thought towards those first actions so as to achieve optimum results for designed algorithms. The present work has been concentrated in depth on the second thought by generating a linear chain model type algorithm to include supportive hardware to an aerial vehicle with multi-rotors for the best performance of the vehicle. The presented work has been differentiated from existing methods by working out a hand calculation method and different online tools so as to verify the calculated results for designing a multi-rotor aerial vehicle with its best hardware components. The future scope present work can be extended by using numerical iterative methods to replace hand calculations and more rugged and robust curve fitting algorithms to get best plots for the entire system.

References

1. Anderson, John David, and J. Wendt. *Computational fluid dynamics*, Vol. 206. New York: McGraw-Hill, 1995.
2. Chen, ongBo, JianQiao Yu, XiaoLong Su, and Guanhen Luo. "*Path planning for multi-UAV formation*", Journal of Intelligent & Robotic Systems77, no. 1 (2015): 229–246.
3. Coetzee, Louis, and Johan Eksteen, *The Internet of Things-promise for the future? An introduction, IST-Africa Conference Proceedings, 2011*, pp. 1–9. IEEE, 2011.
4. E-Calculator, http://www.ecalc.ch/xcoptercalc.php?ecalc&lang=en.
5. Fogelbeg, Johan. "*Navigation and Autonomous Control of a Hexacopter in Indoor Environments*." (2013).
6. Hawkins, N, *Hawkins Electrical Guide, Number One*." Theo. Audel& Co., New York (1914).
7. Jung, Joey, Lei Zhang, and Jiujun Zhang, eds. *Lead-Acid Battery Technologies: Fundamentals, Materials, and Applications*, (2015).
8. Larrauri, Juan I., Gorka Sorrosal, and Manuela González. "*Automatic system for overhead power line inspection using an Unmanned Aerial Vehicle—RELIFO project*", Unmanned Aircraft Systems (ICUAS), 2013 International Conference on, pp. 244–252. IEEE, 2013.
9. Matič, Gašper, Marko Jankovec, David Jurman, and Marko Topič, *Feasibility study of attitude determination for all-rotating unmanned aerial vehicles in steady flight*, Journal of Intelligent & Robotic Systems 80, no. 2 (2015): 341–360.
10. Optimal Propeller Calculator, http://personal.osi.hu/fuzesisz/opc_eng/index.html.
11. Qi, Xin, Juntong Qi, Didier Theilliol, Youmin Zhang, Jianda Han, Dalei Song, and ChunSheng Hua, *A review on fault diagnosis and fault tolerant control methods for single-rotor aerial vehicles*, Journal of Intelligent & Robotic Systems 73, no. 1–4 (2014): 535–555.
12. Rashid, Muhammad Harunur, *Power electronics: circuits, devices, and applications*, Pearson Education India, 2003.
13. Shilov, Kirill E., Vladimir V. Afanasyev, and Pavel A. Samsonov. "*Vision-based navigation solution for autonomous indoor obstacle avoidance flight.*", International Micro Air Vehicle Conference and Flight Competition (IMAV2013), Toulouse, France. 2013.
14. Spreadsheet for Thrust and aerial speed plot, https://docs.google.com/spreadsheets/d/10GAMX5bRebUeykCyxRatTkdPYTtbORuJfY-Wn3myEeU/edit?usp=sharing.
15. Static Thrust Calculator, http://personal.osi.hu/fuzesisz/strc_eng/.
16. Theraja B. L, *Textbook of Electrical Technology*. Granite Hill Publishers, 2008.
17. Valavanis, Kimon P., ed. *Advances in unmanned aerial vehicles: state of the art and the road to autonomy*. Vol. 33., Springer Science & Business Media, 2008.
18. Wang, Yin, and Daobo Wang, *Variable thrust directional control technique for plateau unmanned aerial vehicles,* Science China Information Sciences (2016).
19. Xia, Chang-liang. *Permanent magnet brushless DC motor drives and controls*. John Wiley & Sons, 2012.
20. Yechout, Thomas R. *Introduction to aircraft flight mechanics*. Aiaa, 2003.

Efficient Binarization Technique for Handwritten Archive of South Dravidian Tulu Script

P.J. Antony, C.K. Savitha and U.J. Ujwal

Abstract The segmentation of text from degraded document images is a challenging task comes under an image processing area with application of handwritten character recognition. Tulu is one of five major Dravidian language with many historical documents available in handwritten form. This paper presents an overview of Tulu script and adaptive image contrast based binarization technique for extracting text from a degraded background of Tulu paper document images. In this paper, we improve the quality of the text in the degraded document image using OTSU with several edge detection algorithms (i.e. canny, Sobel, and total variation) techniques applied to the degraded document image and Adaptive threshold with several edge detection algorithms (i.e. canny, Sobel, and total variation) techniques applied to the degraded document image. Finally, the qualities of these output images evaluated by PSNR and MSE. The best combination of threshold and edge detection techniques is selected by testing several degraded documents.

Keywords Degraded document · Tulu script · Binarization

1 Introduction

A general framework for offline handwritten character recognition system involves Preprocessing, Feature extraction, Classification, Recognition stages. The preprocessing stage involves image processing steps of RGB to Gray scale conversion,

P.J. Antony · C.K. Savitha (✉) · U.J. Ujwal
Department of Computer Science and Engineering KVGCE Sullia DK,
Visvesvaraya Technical University, Belgaum, India
e-mail: cksavithaharish@yahoo.com

P.J. Antony
e-mail: antonypjohn@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_56

Binarization, Segmentation of lines, words, characters. Binarization is the process of converting grayscale image to a binary image, which is mainly used to extract text from degraded backgrounds of Ancient manuscript images.

Tulu is India's well-developed Dravidian language with script of its own. There were numerous documents written in Tulu script, dates back to the fifteenth century. These manuscripts may be used to record Mantras, culture, astrology, astronomy and medical information. Protection and preservation of such valuable materials achieved with digitization of documents by scanning or by taking photos. Digitized documents may be of poor contrast due to noise, damages, spreading ink from backside of paper [1]. Figure 1 shows deteriorations such as uneven background appeared in the Tulu handwritten document image.

Efficient binarization technique is an essential step to extract knowledge from handwritten historical documents and to improve overall performance of the system. Image binarization is the process of converting image into black and white pixels by choosing an appropriate threshold value. Selecting a single threshold, which is compatible with entire image is a difficult task. So there is a need of adaptive binarization to choose an optimal threshold for each area of the image. The proposed method uses an adaptive estimation of parameters for different stages of binarization, hence robust for different category of degraded Tulu documents.

The rest of the paper organized as follows. Section 2 presents a literature survey on different approaches used for preprocessing of ancient manuscripts. Section 3 presents proposed methodology for binarizing Tulu Historical documents. Section 4 gives Experimental results. Last section gives the conclusion.

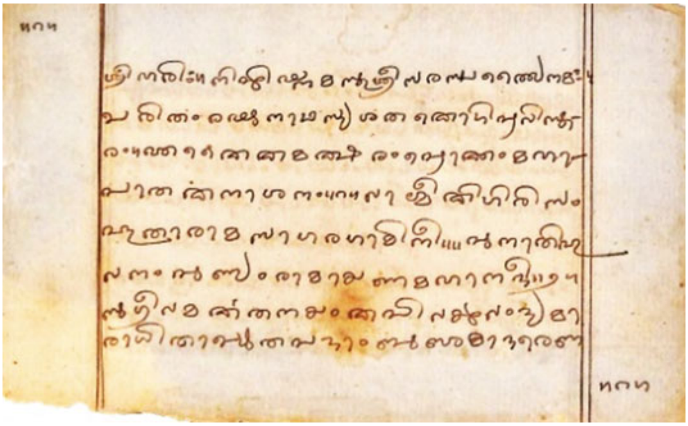


Fig. 1 Image of Tulu handwritten Paper document (Collected from National Trust for Computation and Archival of Oriental Media, Udupi)

2 Literature Survey

Many binarization techniques primarily designed for modern documents which include Otsu's thresholding techniques [2]. Article [3] explains Otsu's thresholding techniques for palm leaf manuscripts. Such global thresholding is not sufficient for segraded documents without bimodal patterns. An entropy-based binarization technique for noise elimination of document image written on both sides are explained in paper [4]. Hybrid method for foreground/background separation is explained in paper [5].

Review article [6] explains three methods of binarization as thresholding methods, hybrid methods and non threshold method with hybridization as the best choice. To eliminate the problem of single global or local thresholding, adaptive thresholding is proposed in the article [7] where the local threshold is performed for individual pixel. In the same way decompose algorithm is proposed in the paper [8]. In this technique entire region is decomposed into subregions and the for each sub region thresholding is performed automatically to extract text from degraded document. Paper [9] explains optimal threshold from gray level histogram based on discriminant analysis as solution to multithresholding problems. From the literature, it has been observed that it is difficult to select a single algorithm suitable for all types of degraded document [10] and it is mainly compared based on the type and nature of document images. So there is a need of algorithm to automatically select combination of algorithm to achieve binarization results.

3 Proposed Framework

Digitizing the documents can be done by scanning the input corpus of Tulu documents into any image file format. Such scanned document images may contain noise, degradations, variations. Proposed framework uses Adaptive contrast map and automatically choose a combination of threshold technique with an edge detectionapproach to extract text from a degraded background as shown in Fig. 2.

3.1 Adaptive Contrast Map

First, colored Tulu document is converted into Grayscale using RGB to Gray conversion [3] method using Eq. (1)

$$Y = 0.3\mathrm{R} + 0.59\mathrm{G} + 0.11\mathrm{B} \tag{1}$$

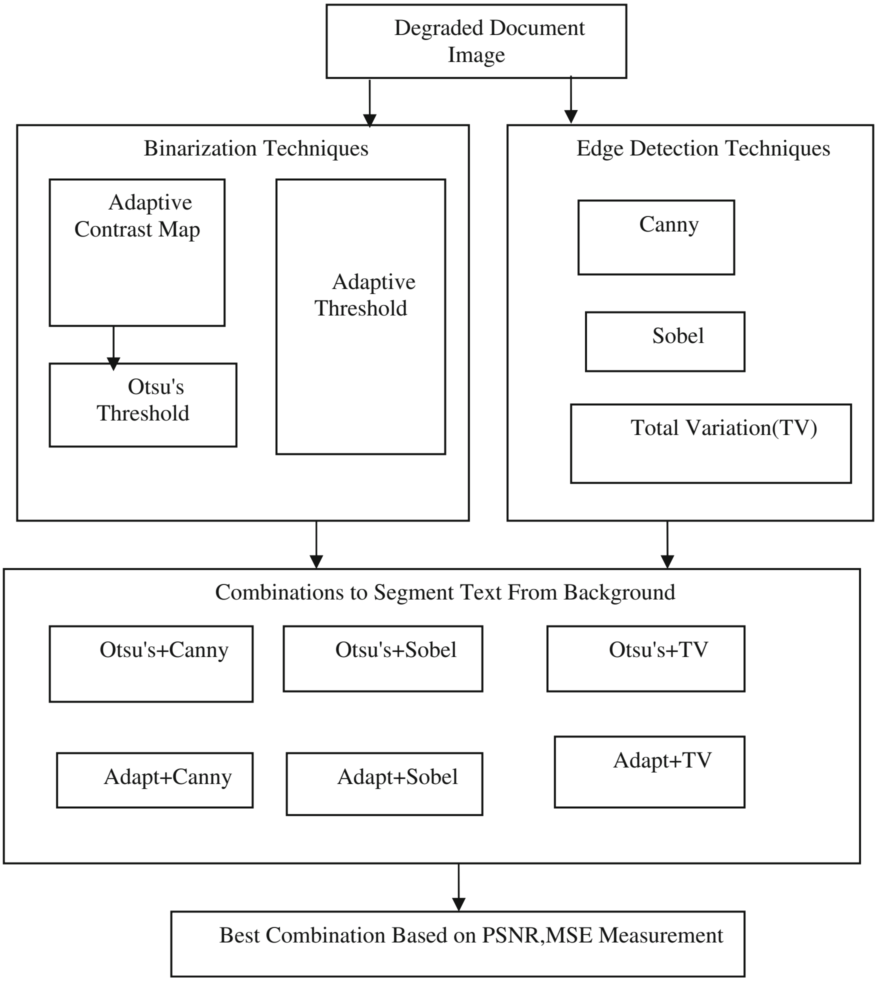


Fig. 2 Fremework for segmentation of text from degraded Background

where R indicates Red, G indicates Green and B indicates Blue. To extract text stroke edges from less variation background, Gradient of image [1] is obtained using Eq. (2)

$$\mathrm{C}(i,j) = (\mathrm{Imax}(i,j) - \mathrm{Imin}(i,j)) \tag{2}$$

where C (i, j) denotes the contrast of an image pixel (i, j), Imax (i, j) and Imin (i, j) denote the maximum and minimum intensities within a local neighborhood windows of (i, j), respectively. If the local contrast $\mathrm{C}(i, j)$ is smaller than a threshold, the pixel is set as background directly. Otherwise, it will be classified into text or background by comparing with the mean of $\mathrm{Imax}(i, j)$ and Imin (i, j).

As the degradation of background increases, the contrast of image is needed to detect the only stroke edge pixels [1] of the Tulu document text properly which is obtained using Eq. (3).

$$\mathrm{C}(i,j) = \frac{\mathrm{Imax}(i,j) - \mathrm{Imin}(i,j)}{\mathrm{Imax}(i,j) + \mathrm{Imin}(i,j) + €} \tag{3}$$

where ∈ is a positive, but infinitely small number that is added in case the local maximum is equal to 0.

To handle bright text properly, adaptive image contrast is used where local image contrast is combined with the local image gradient of the input image as given in the Eq. (4). It also detects many non stroke edges of background of degraded document, where degradation may be due to bleed through, lighting, ageing etc.

$$\mathrm{Ca}(i,j) = \alpha C(i,j) + (1-\alpha)(\mathrm{Imax}(i,j) - \mathrm{Imin}(i,j)) \tag{4}$$

where C (i, j) denotes the local contrast in Eq. (2) and (Imax (i, j) − Imin (i, j)) refers to the local image gradient that is normalized to [0, 1]. α is the weight between local contrast and local gradient that is controlled based on the document image statistical information.

3.2 Binary Map

Next, detection of the text stroke edge pixel candidate is achieved by using binary mapping techniques such as Otsu's global thresholding method and Adaptive threshold method. Otsu's thresholding technique separates the pixels into two clusters and then it finds the means of each cluster and square of the difference between means. It computes the histogram and each intensity level's probability, by setting up initial $q_i(0)$ and $\mu_i(0)$.Then computes $\sigma_b^2(\mathrm{t})$ by updating q_i and μ_i, which the point at which desired threshold corresponds to the maximum.

The adaptive threshold technique is another binarization technique which obtained by using formula

$$\mathrm{O}(i,j) = 0, \text{ if } \mathrm{g}(i,j)\ \mathrm{XS}^2 < t(i,j)$$
$$255, \text{ Otherwise}$$

This technique is useful because comparing a pixel to the average of neighboring pixels will keep hard contrast lines and ignore soft gradient changes. The advantage of this technique is that only a single pass through the image is required.

3.3 Edge Detection

To detect the edges of the image, edge detection techniques such as Canny, Sobel, Total variation methods are used. Canny's edge detector uses two adaptive thresholds and it marks edges close to real edge locations. It is tolerant to shade artifacts.

In the same way edge detection is also achieved with Sobel operator which uses 3×3 convolution kernel. The kernels respond to vertical and horizontal edges.

The discrete Total Variation (TV) model is a very successful method for image restoration because of its ability to preserve edges in the image due to the piecewise constant regularization property of the TV norm.

The discrete gradient operator is defined by

$$\nabla : \boldsymbol{R}^{n^2} \to \boldsymbol{R}^{n^2}$$

with

$$((\nabla \mathbf{f})_{\mathbf{j,k}} = \left((\nabla f)^{x}_{j,k}, (\nabla f)^{y}_{j,k}\right)$$

$$(\nabla f)^{x}_{j,k} = \{ \begin{matrix} f_{j+1,k} - f_{j,k} & if\, j < n \\ 0 & if\, j = n \end{matrix}$$

$$(\nabla f)^{x}_{j,k} = \{ \begin{matrix} f_{j+1,k} - f_{j,k} & if\, j < n \\ 0 & if\, j = n \end{matrix}$$

for j, k = 1,…n, Here $f_{j,k}$ refers to the $(jn + k)$th entry of the vector f (it is the (j, k)th pixel location of an $n \times n$ image). The discrete total variation of f is defined by

$$\boldsymbol{TV}(f) := \sum_{1 \le j,\, k \le n} \sqrt{(\nabla f)^{x\,2}_{j.k} + (\nabla f)^{y\,2}_{j.k}}$$

These edge detection techniques are used to extract a large amount of non-stroke edges.

3.4 Segmentation

The improvement in the binary map can be achieved through the combination of thresholding with the edges which is detected by edge detection methods. The combination of Otsu's thresholding with Canny, Sobel, Total variation as well as a combination of Adaptive threshold with Canny, Sobel, Total variation (TV) is used to extract text from degraded background images. Accurate text stroke edge pixels extraction is possible with these combinations. The preprocessing steps of text extraction from degraded background using combination of Otsu's threshold and canny edge detection is shown in Fig. 3.

The segmentation process involves the local threshold estimation and postprocessing stages to improve the binarization results. In this step, after detection of high contrast stroke edge pixels, threshold estimation performed locally. We calculate the most frequent distance between two adjacent edge pixels in horizontal direction and use it as the estimated stroke width. The edge pixel candidates are selected by row by row horizontal scanning of the edge image. The histogram is constructed that the frequency of the distance between two adjacent candidate pixels as shown the Fig. 4. The stroke edge width can then be approximately estimated by using the most frequently occurring distances of the adjacent edge Pixels.

Algorithm1: For detection of width of edge
INPUT: Tulu degraded document image Iimg and
Two fold (Binarised) Text stroke edge detected image Eimg
Steps:

1. Obtain width w and height h of Iimg
2. **for** each row j = 1 to h in Eimg
3. While scaning,If pixel is belongs to background, label it as 0
   Adjacent pixel is edge, labeled as 1
4. Select pixels with high intensity by removing lower intensity pixels in the same row
5. Select pair of adjacent pixels and calculate distance between them.
6. end **for**

OUTPUT: Select most frequently occuring distance as edge width by computing histogram of calculated distance

Postprocessing
The binarization result is further improved by postprocessing. Here precise selection of the edge pixel set is obtained by the filtering of isolated foreground pixels

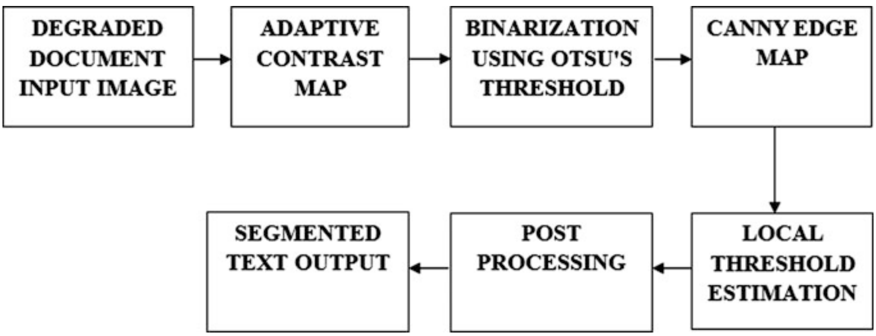


Fig. 3 Text extraction from degraded background using Otsu thresholding with Canny edge detection

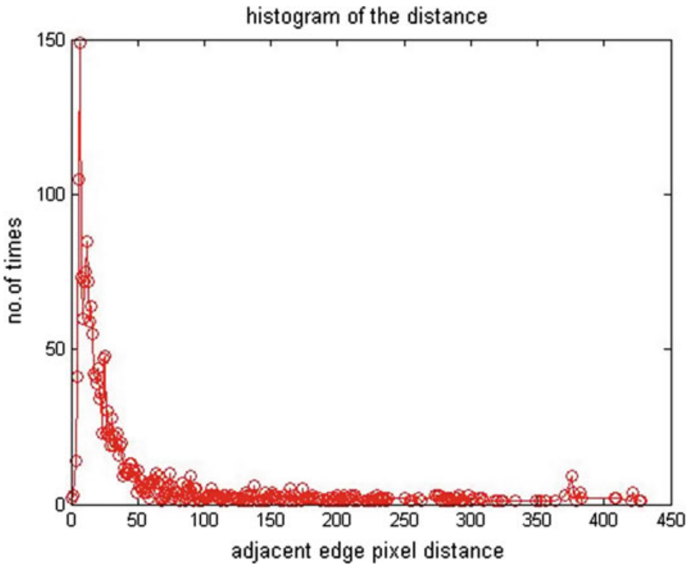


Fig. 4 Histogram of distance between adjacent edge pixel

from foreground pixels that do not connect with other foreground pixels. The neighborhood pixel pair belongs to the different classes if that lie on symmetric sides of a text stroke edge pixel (i.e., either the document background or the foreground text). One pixel of the pixel pair is therefore labeled with the other category if both of the two pixels belong to the same class. Finally, several logical operators are used to filter out some single-pixel artifacts along the text stroke boundaries.

4 Experiments and Results

For the experiments we have used a dataset of 48 images of Tulu manuscript named Sundara Khanda episode of Ramayana, which is collected from the National Trust for Computation and Archival of Oriental Media, Udupi. The algorithm is implemented using MATLAB Version R2014. For the evaluation we used measures of PSNR and MSE. How close an image to another is measure using PSNR. It is computed using the formula

$$\mathrm{PSNR} = 10\log\left(\frac{C^2}{\mathrm{MSE}}\right)$$

where $\mathrm{MSE} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\mathrm{I}(x,y) - \mathrm{I}'(x,y)\right)^2}{MN}$

And C is the difference between foreground and background. The higher PSNR value indicates most similar images. MSE is mean square error between two images. The lesser value of MSE indicates more similar pictures. Therefore, proposed method selects best combination based on higher PSNR, Low MSE. Along with considering time to complete the process. Figures 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 shows binarization results by using thresholding, edge detection as well as a combination of both. PSNR, MSE and elapsed time to complete the process is shown in the Table 1.

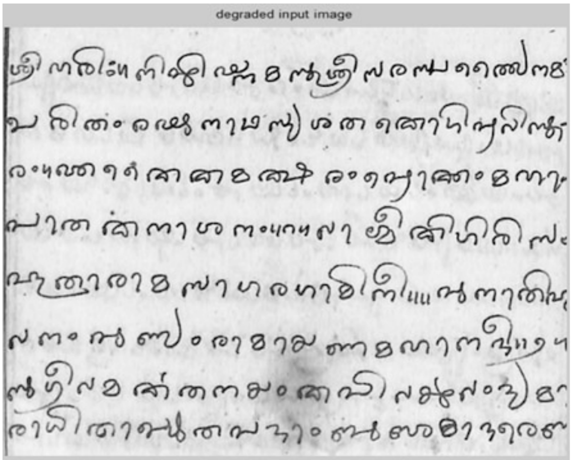


Fig. 5 Degraded input image

Fig. 6 Adaptive contrast of an image

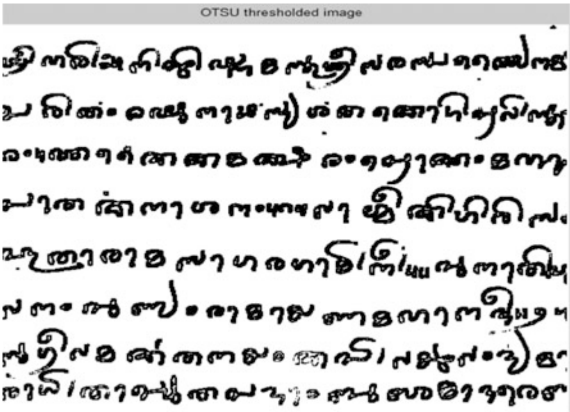


Fig. 7 Otsu's Threshold image

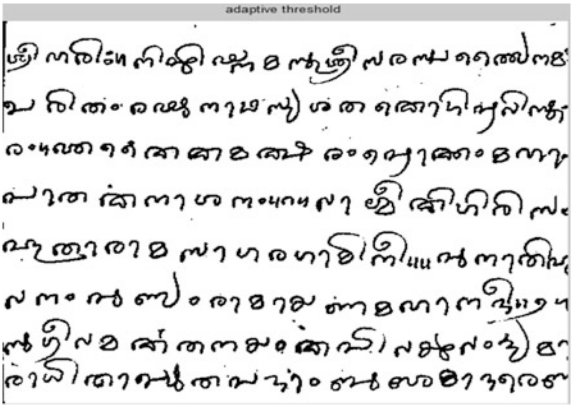


Fig. 8 Adaptive Threshold image

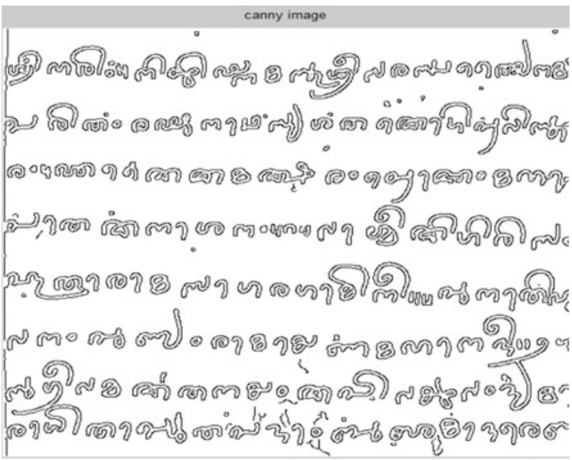


Fig. 9 Canny Edge detection

Fig. 10 Sobel edge detection

Fig. 11 Otsu combined with Canny

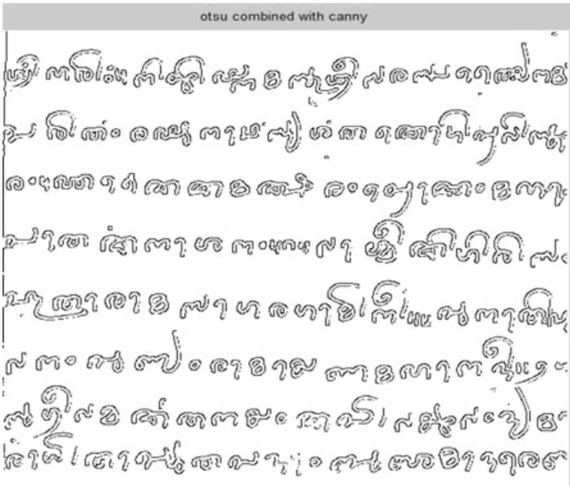
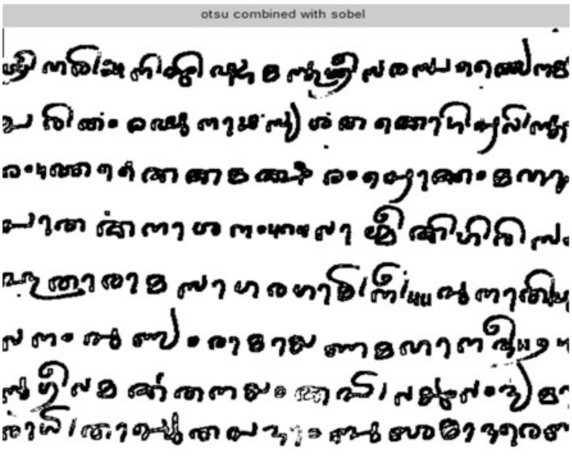


Fig. 12 Otsu combined with Sobel



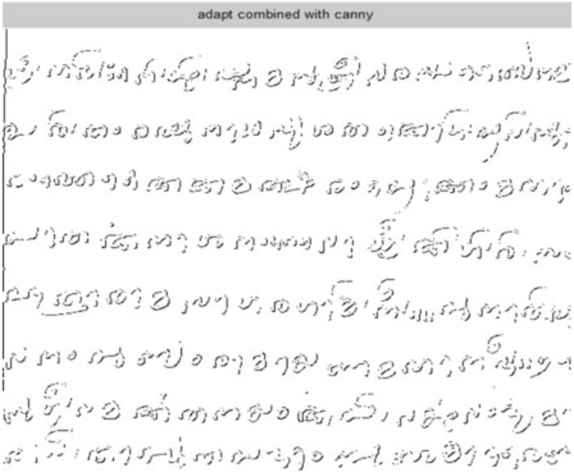


Fig. 13 Adapt combined with Canny

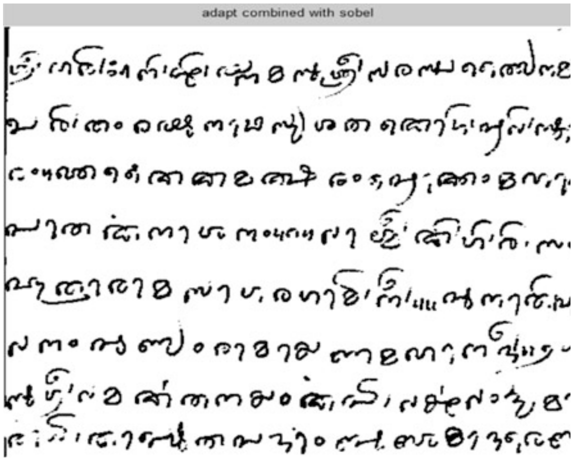


Fig. 14 Adapt combined with Sobel

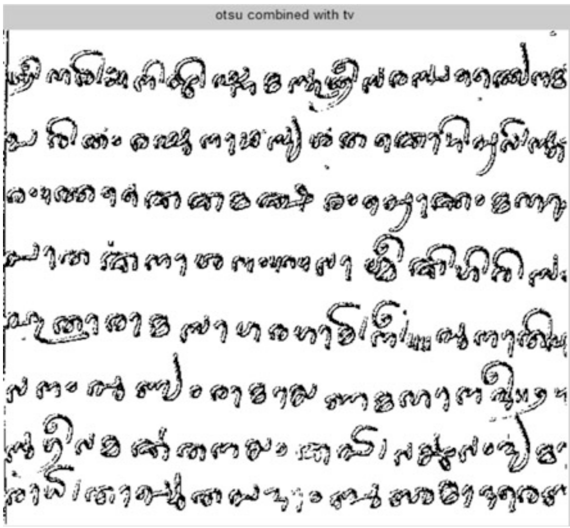


Fig. 15 Otsu combined with TV

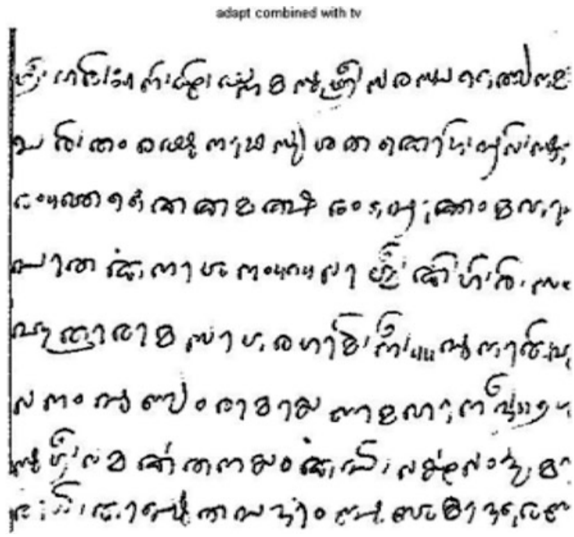


Fig. 16 Adapt combined with TV

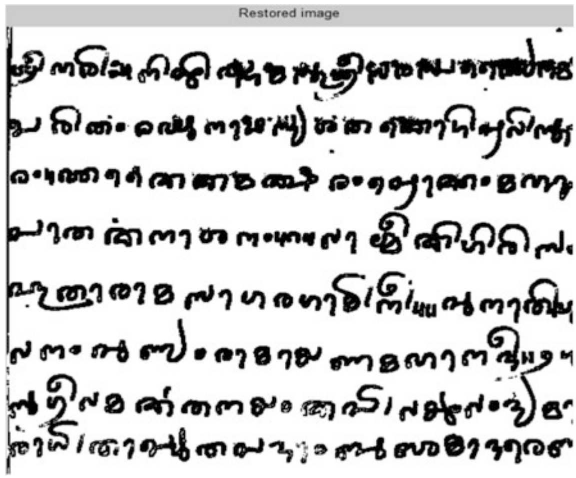


Fig. 17 Otsu with Canny

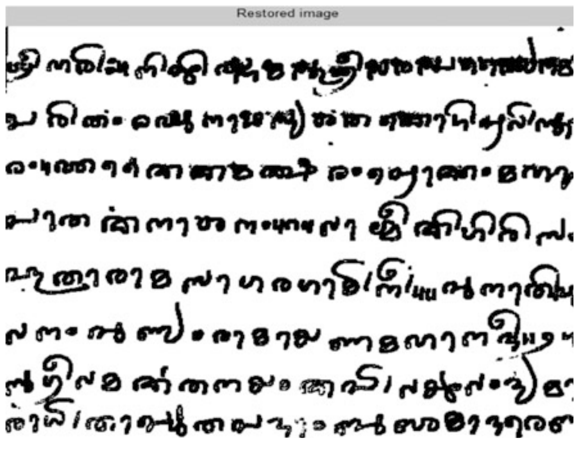


Fig. 18 Otsu with Sobel

Fig. 19 Otsu with TV

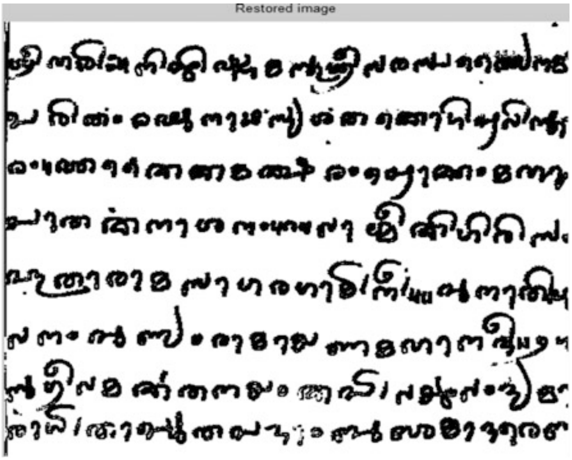


Fig. 20 Adapt with Canny

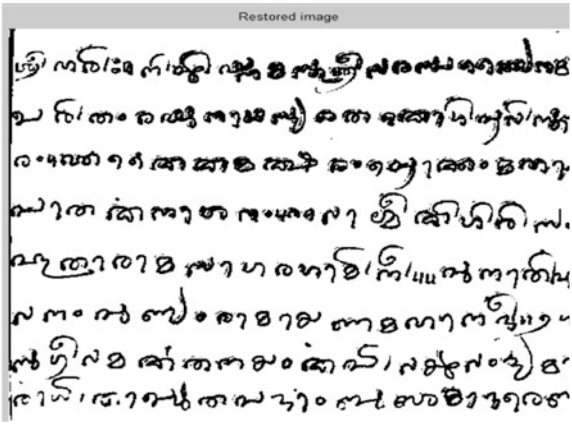
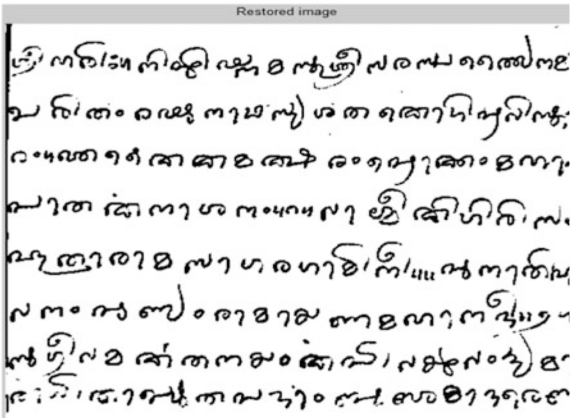


Fig. 21 Adapt with Sobel



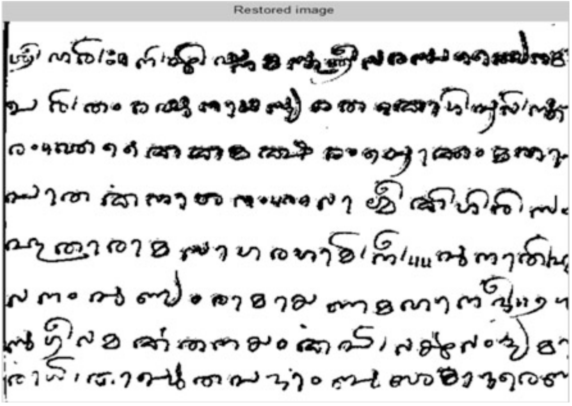


Fig. 22 Adapt with TV

Table 1 Binarization results of six combinations of thresholding techniques on TULU manuscript

	Otsu_Canny	Otsu_Sobel	Otsu_TV	Adapt_Canny	Adapt_Sobel	Adapt_TV
PSNR	58.3974	58.3601	58.3124	61.2893	63.3601	61.2890
MSE	0.0940	0.0949	0.0969	0.0483	0.0300	0.0483
TIME	22.3081	22.8697	537.7510	23.6030	22.1646	539.0459

Restored Images

From the Table 1 it has been observed that for the document page in Fig. 1, the Adapt_ Sobel combination is best with PSNR is 63,360,103 dB, MSE is 0.0300 and Elapsed time is 22.1646 s. In the same way experiment is repeated with different pages of Tulu manuscript and good binarization result with high PSNR, lowest MSE obtained by testing different combination of thresholding and edge detection techniques.

5 Conclusion

Tulu is one of five noteworthy Dravidian dialect with numerous historical documents available in handwritten form. The system proposed in this paper can be used for efficient extraction of text from such highly degraded Tulu documents. The quality of degraded document is improved by adaptive contrast map with automatic selection of a better combination of thresholding with edge detection technique. Experiments demonstrate that the proposed technique outflanks most reported archive binarization strategies in term of the PSNR, MSE with reasonable Elapsed time. The system can be extended to segment lines, characters in order to detect and recognize handwritten characters in Tulu language in future.

References

1. Su, Bolan, Shijian Lu, and Chew Lim Tan. "Robust document image binarization technique for degraded document images." *Image Processing, IEEE Transactions on* 22.4 (2013): 1408–1417.
2. Govindan, V. K., and A. P. Shivaprasad. "Character recognition—a review."*Pattern recognition* 23.7 (1990): 671–683.
3. Surinta, Olarik, and Rapeeporn Chamchong. "Image segmentation of historical handwriting from palm leaf manuscripts." *Intelligent Information Processing IV*. Springer US, 2008. 182–189.
4. Mello, Carlos AB, and Rafael D. Lins. "Image segmentation of historical documents." *Visual2000, Mexico City, Mexico* 30 (2000).
5. Boussellaa, Wafa, Abderrazak Zahour, and Adel Alimi. "A methodology for the separation of foreground/background in arabic historical manuscripts using hybrid methods." *Proceedings of the 2007 ACM symposium on Applied computing*. ACM, 2007.
6. Yahya, Sitti Rachmawati, et al. "Review on image enhancement methods of old manuscript with the damaged background." *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on*. Vol. 1. IEEE, 2009.
7. Sauvola, Jaakko, and Matti Pietikäinen. "Adaptive document image binarization." *Pattern recognition* 33.2 (2000): 225–236..
8. Chen, Y., and G. Leedham. "Decompose algorithm for thresholding degraded historical document images." *Vision, Image and Signal Processing, IEE Proceedings-*. Vol. 152. No. 6. IET, 2005.
9. Otsu, Nobuyuki. "A threshold selection method from gray-level histograms."*Automatica* 11.285–296 (1975): 23–27.
10. Chamchong, Rapeeporn, Chun Che Fung, and Kok Wai Wong. "Comparing Binarisation Techniques for the Processing of Ancient Manuscripts." *Cultural Computing*. Springer Berlin Heidelberg, 2010. 55–64.

Application of Image Processing for Detection and Recognition of Objects in Traffic Video

B. Ashwini, B.N. Yuvaraju, B. Deepashree and P.S. Venugopala

Abstract Intelligent transportation systems (ITS) are advanced application that aims to provide variety of services with respect to transport and its management issues. It enables the users to make safer and better use of transport network. These applications provide various information and help in coordinating the traffic flow. Image processing has played a major role in ITS. The edge detection technique has a very significant role in Computer Vision. Edge detection helps in filtering out unwanted information and preserves only the important properties of the image. A comparison of various Image Edge Detection techniques is done using MATLAB 2012R and the analysis is recorded. The paper focuses on edge detection techniques such as Sobel, Canny, Prewitt, and the impact of noise on it. The Experimental results find the edges of original image, notice the quality of image with the help of PSNR, RMSE, and Correlation Coefficient. This paper also focuses on optical flow method and gaussian mixture model method used to detect moving objects and an analysis is done.

Keywords Edge detection · Noise · PSNR · SNR · Correlation coefficient
Optical flow · Gaussian mixture model

B. Ashwini (✉)
Department of ISE, NMAMIT, Karkala 574110, India
e-mail: ashwinib@nitte.edu.in

B.N. Yuvaraju
Department of CSE, NIE, Mysuru, India
e-mail: yuvarajubn@gmail.com

B. Deepashree · P.S. Venugopala
Department of CSE, NMAMIT, Karkala 574110, India
e-mail: deepashreeb93@gmail.com

P.S. Venugopala
e-mail: venugopalaps@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_57

1 Introduction

Edge detection is the method used to identify and detect sharp changes or interruption in the image. These interruptions are sudden variations in the intensity of the pixel which distinguish boundaries of objects in frame. Detection of edges is tough in images having noise as the noise and the edges contain high frequency. Reduction in noise may result in out of focus and distorted edges.

Optical flow method is used to detect motion in video surveillance. It has got two methods Lucas–Karne and Horn–Schunck method.

The Gaussian probability density function has a single extension called the Gaussian mixture model. The Gaussian mixture model consists of a combination of Gaussian probability density function, which include its mean, standard deviation, and weight. The background image is the Gaussian probability density function of the mean pixel value in this paper gaussian mixture models to detect foreground in a video. In this paper three edge detection methods, optical flow, and gaussian mixture model methods are used and a comparative study has been done to know which method has a better result on various images and videos.

2 Literature Review

In the paper [1] authors focus a comparative study of edge detection techniques such as Sobel, Robert, Prewitt and impact of noise on it. Paper [2] says Edge detection has a key role in image processing and computer vision. It is front line of image processing for object detection. Comparative analyses of different edge detection methods are presented. The results say that the canny edge detection operator gives better performance value. Paper [3] presents a detailed study of mage segmentation based on edge detection methods which tells about techniques of "Gradient-based and Laplacian-based Edge Detection." The paper [4] gives details about the basic concepts of various filters and apply these filters on the image taken. [5] says combining of an image with a 2-D filter which is sensitive to large gradients is one of the traditional method of edge detection. [6] implementation of the three edge detection techniques on images with salt and pepper noise is done and they are analyzed based on the PSNR, RMSE, and CoC evaluation parameters.

The proposed approach exploits gray color information for both background subtractions to improve object segmentation. This paper [7], shows a new edge detection operator, which is Log_Sobel. It suits online detection. [8] proposed system aims to build an automatic traffic monitoring system which can replace or reduce manual traffic monitoring. In [9] classification and counting of vehicles based on different video–image processing techniques like object detection, edge detection, frame differentiation, and the Kalman filter were proposed.

In this paper, "Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance" [10] stationary cameras are used for detecting moving objects in videos. Computer vision and image processing techniques are used to develop the system based on optical flow estimation. Noise are removed by using median filters, thresholding operations remove unwanted objects. Also blob analysis sets the object type restrictions. Detection and tracking of moving objects is done successfully.

In this paper, [11] the system uses Gaussian Mixture Model (GMM) and Optical Flow methods for object tracking. The average, standard deviation, and weight are the three different approaches considered by GMM. Quick calculations can be done in optical flow. Lack of complete object tracking is a setback of this method. The GMM approach gives full results of the operation but long computing time with more noise are its drawbacks. Thus combination of these two methods and image filtering results in the successful tracking of objects.

"Traffic Image Processing System" [12] paper says, increase in waiting time during heavy traffic has a cascading effect on fuel consumption and environment. The author make use of web cameras mounted on street lights to capture still images of the traffic which then undergo a series of steps related to Image Processing and Image Analysis. Finally, calculate the traffic volume and operate the traffic light's timer accordingly. The system helps in minimization of several factors like waiting time, fuel consumption, and congestion.

"Automatic Vehicle Detection and Counting Algorithm" [13] this study proposes an algorithm to detect and count vehicles passing a certain point for video-based monitoring of traffic flow. The unique feature of this algorithm is to calculate an approximate value of speed while counting vehicles using GMM background modeling, object histogram, and pyramidal Lucas–Kanade method. In addition, another feature of this algorithm is to use the pre-processing video that is converted into distance coordinate system in consideration of radar sensor and data coordination.

In the paper "Computer Vision Based Vehicles Detection and Counting for Four Way Traffic" [14] the author tells that Moving vehicles detection evaluating the motion of vehicles from the given video. Optical flow method, frame differencing method, and background subtraction are the different methods for motion detection. From these the most commonly used method is background subtraction. Counting is done using Gaussian mixture model (GMM) and BLOB analysis is done once the object is detected.

"Moving Object Detection Using Matlab" [15, 16] implements foreground detection-based moving object detection and vehicle tracking algorithm. An AVI file is read and it's R, G, and B color components are extracted from it. Moving objects are detected using various operations. Thresholds at various phases are decides the possibility of identifying the moving object with respect to its size. Moving objects also tracked using MATLAB platform. The paper focuses on focus is given on (unmanned aerial vehicle) UAV.

3 Edge Detection Techniques

(i) *Sobel Edge Detection*—In digital images, the approximate partial derivation in gradient is computed by the Sobel operator. In terms of computations, the

edge is based on the edge convolving with the integer, small valued filter, and separable in vertical and horizontal directions. Mathematically, the approximations of the derivative can be calculated by using two 3 * 3 kernels which are combined with the original image. One kernel is simply stated on position and the other rotated by 90°.
For the input image the kernels can be applied separately one by one, for giving separate values of the gradient component in various orientations (defined as G_x and G_y). It can be joined together to get the estimation of orientation of the gradient. The absolute value of magnitude of the gradient and the magnitude of gradient is given by:

$$\mathrm{Mod}\,G = \left(G_{x2} + G_{y2}\right)^{1/2} \tag{1}$$

Typically, an estimated magnitude of image is computed using:

$$\mathrm{Mod}\,G = \mathrm{Mod}\,G_x + \mathrm{Mod}\,G_y \tag{2}$$

Equations 1 and 2 gives the angle of orientation of edge with the object relative to the pixel grid producing values to the spatial gradient.

$$q = \mathrm{arc\ tan}\left(G_y/G_x\right)$$

(ii) *Prewitt Edge Detection*—The maximum responses which are directly from the kernel are obtained by the use of Prewitt Edge Detector. The prewitt edge operator or detectors are used for the measurement of two components, i. e., horizontal edge components and vertical edge components. These two components (vertical and horizontal) use different kernels.

(iii) *Canny Edge Detection*—It is considered as standard edge detection techniques. Newer algorithms adopt canny edge detection method. The noise can be reduced and suppression can be minimum in this detection method.

A. *Optical Flow*

Optical flow is used in our system to provide statistical traffic flow information. Optical flow plays an essential and critical role which affects the final performance of the system. Thus, application-specific assumptions must be considered to choose an algorithm for optical flow estimation. The displacement between two image frames is calculated by taking two frames at times t and t + δt at every position. These methods are based on local Taylor Series approximation.
There are two methods used:

1. Lucas–Kanade: It is a widely used differential method for optical flow estimation. The flow is assumed to be in a local neighborhood of the pixel the flow is assumed to be constant. The least squares criterion is used to solve the optical flow equation. The Lucas–Kanade method is also less

sensitive to image noise. It does not provide flow information in the interior of uniform regions of the image.

2. Horn–Schunck: The aperture problem is solved by a global constraint provided by the Horn–Schunck method. The Horn–Schunck algorithm minimizes distortions in flow thereby showing more smoothness in the image.

B. *Gaussian mixture method*
Gaussian Mixture Model (GMM) is probability density function represented as a weighted sum of its function. GMM is often used in image processing as it can approximate smooth shape of the distribution. Mean, standard deviation, and weight are the components of the Gaussian probability density function. The background image is the mean pixel value of the Gaussian probability density function.

4 Parameters for Evaluation

(i) *Mean Square Error*—It is the difference between the estimated value and the actual value. The mean square error is the squared error averaged over the $M \times N$ array.
MSE = 1/MN ΣΣ (f1 (i, j) − f2 (i, j))2
i = 1 j = 1
where f1 is output image and f2 is input image. Its value must be less.

$$\text{RMSE} = \sqrt{\text{MSE}}$$

(ii) *Peak Signal to Noise Ratio*—It is the ratio of the maximum possible powers of signal to that of the power of noise. It can also be represented as a log function of maximum value of image and mean square error.
PSNR = 10 LOG (2552/MSE), here MSE is the mean square error. The PSNR value should be high.

(iii) *Correlation Coefficient*—It is the measure between the predicted values and the actual values calculated earlier. The value of correlation coefficient lies between 0 and 1. The value of the correlation coefficient is 0 when there is no match between the values. It increases in its value as the relationship strength between the predicted values and actual values increases. A perfect match of the values will have coefficient of 1.0. Maximum value of correlation better it is.
Correlation(r) = NΣXY − (ΣX)(ΣY)/sqrt([NΣX2 − (ΣX)2] [NΣY2 − (ΣY)2])
Here N = number of pixel of the image
X = input image, Y = output

5 Algorithms

A. *To Compare Edge Detection Method*

Two types of images are considered. They are car image and camera man image. The steps are as follows:

1. Read the image.
2. Convert the image to double.
3. Apply edge detection methods to the filtered image.
4. Evaluate the PSNR, RME, and CoC parameters.
5. Compare the results.

B. *Tracking Vehicles Using Optical Flow*

This model tracks vehicles in a video by detecting motion using optical flow. The vehicles are segmented from the background by thresholding the motion vector magnitudes. Then, blob analysis is used to identify the vehicles.
Steps:
Initialization:

1. Create system objects outside of the main video processing loop.
2. Optical flow object is created for calculating direction and speed of the detected objects in motion.
3. Two objects are created for analyzing optical flow vectors.
4. Filter object are created for removing speckle noise introduced during segmentation.
5. Morphological closing object is created to fill holes in blobs.
6. Initialize a system object for blob analysis to segment cars in the video.
7. Morphological erosion objects are created for eliminating unwanted objects.
8. Objects created for drawing the bounding boxes and motion vectors.
9. The object created in the previous step will display the number of tracked vehicles. System objects are created to display the input video, video displaying the motion vectors, the video with the thresholded data and the final result.

C. *Tracking Vehicles in Video*

1. The processing loop is created to track the vehicles in video.
2. The output video shows the cars which were tracked by drawing boxes.

Tracking Vehicles Using Gaussian Mixture Model

This model illustrates the use of Gaussian mixture models to detect foreground in a video. Once the foreground is detected, the binary foreground images are processed using blob analysis. Finally, bounding boxes are drawn around the detected vehicles.
Steps:
Initialization:

1. System object is created to read video from avi or mp4 file. In this project an mp4 video is used.
2. Color space converter System object is created to convert the image from RGB to intensity format.
3. System object is created to detect foreground using Gaussian mixture model.
4. A blob analysis System object is created to segment vehicles in the video.
5. System object for drawing the bounding boxes around detected vehicles is created.
6. A System object is created and configured to write the number of vehicles being tracked.
7. System objects are created to display the results.
   Stream Processing Loop

   1. A processing loop is created to track the vehicles in the input video.
   2. System objects, previously initialized are used by this loop. This loop terminates when the end of input file detected by the VideoFileReader

6 Results and Analysis

Edge detection in car

Figure 1 shows the four sub-images where the input image is the black and white or a gray image. The remaining sub-images show the edge detected images, i.e., canny, sobel, and prewitt.

Sobel edge method with high PSNR, low RMSE, and high

CoC value shows better performance results (Graph 1).

Edge detection in car

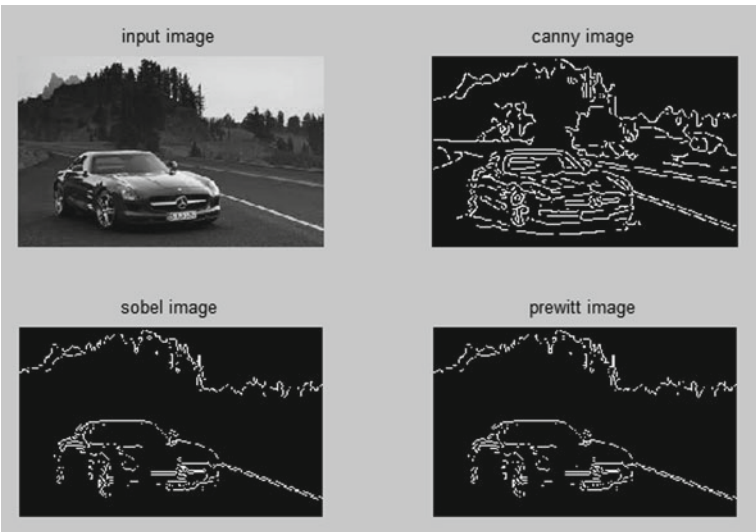


Fig. 1 Car image

Table 1 Calculated values for car image (Fig. 1)

Method	PSNR	RMSE	CoC
Canny	34.4664	23.4339	−0.0521
Sobel	38.0310	10.3120	0.0485
Prewitt	38.0135	10.3545	0.0456

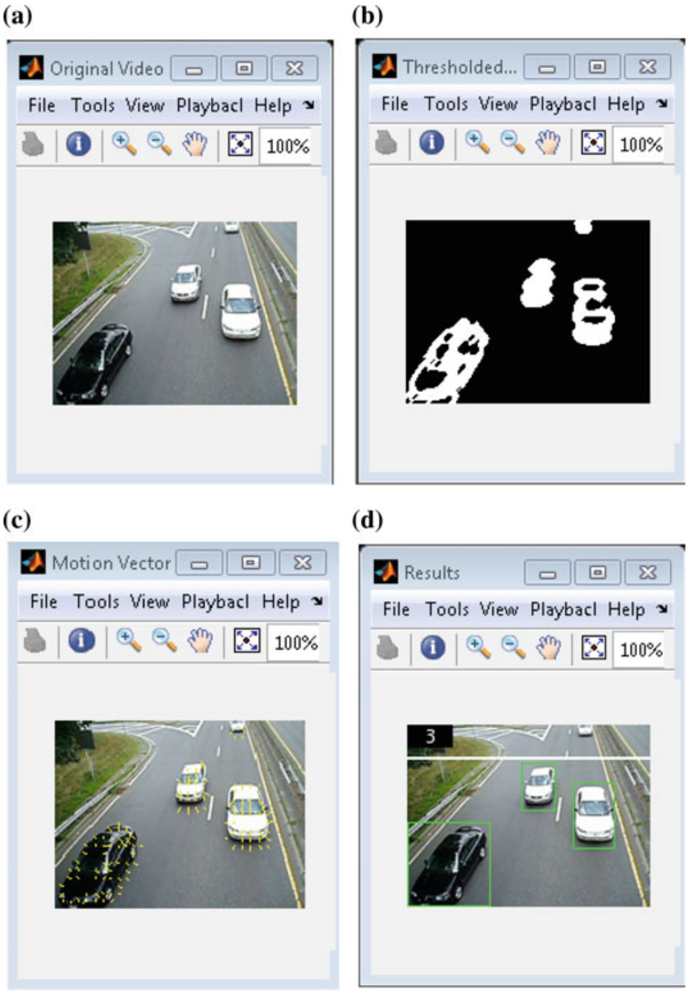


Fig. 2 Optical flow method

Similarly edge detection were carried on three more images. The result obtained was similar to the car image given in this paper

Optical flow method

Sample1

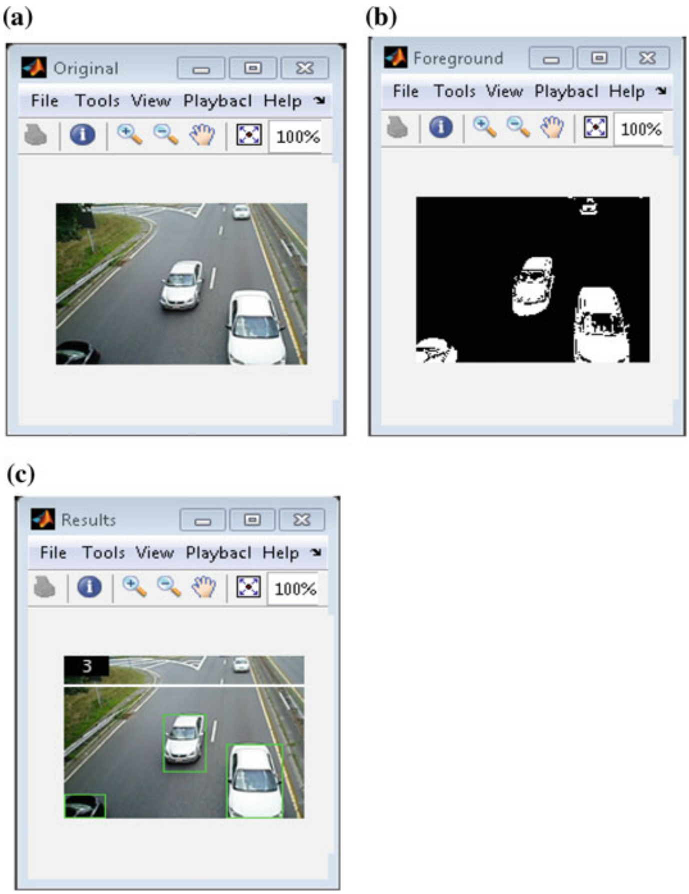


Fig. 3 Gaussian mixture method

Graph 1 Depiction of Table 1

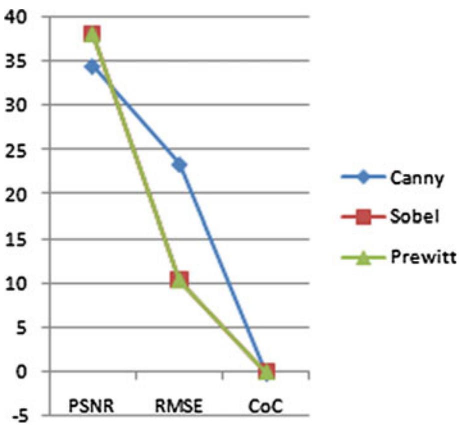


Figure 2 shows the optical flow method applied on the sample video: Fig. 2a shows original video, Fig. 2b shows thresholded frame where the frame is converted o gray image and background subtraction is done, Fig. 2c motion vector detected in the frame and the resultant video, and Fig. 2d shows the result frame where bounded box is drawn on the moving objects present in the RIO.

Gaussian mixture method

Sample1

Figure 3 shows Gaussian mixture method applied on the video: Fig. 3a shows original video, Fig. 3b shows foreground detected from the frame, and Fig. 3c shows the result frame where bounded box is drawn on the moving objects present in the RIO.

Both optical flow and Gaussian mixture method was applied on three more sample video to get the appropriate results.

Comparative analysis between optical flow method and Gaussian mixture model method

Sample video	Optical flow		Accuracy (%)	Gaussian mixture		Accuracy (%)
	Manual counting	Algorithm count		Manual counting	Algorithm count	
1	10	10	100	10	10	100
2	4	7	58	4	8	50
3	14	12	85	14	10	71.4

By the comparison we can conclude that optical flow method has better performance and accuracy than the Gaussian mixture method.

7 Conclusion and Future Work

Intelligent transport system (ITS) along with the help of different wired and wireless technology has provided innovative services and applications related to different modes transportation and traffic. ITS provides the users with more information; help them make a smarter and better choice of the transport networks. Vehicle tracking is done effectively in the present project. Two methods, optical flow and Gaussian mixture methods are used to identify the vehicles in the video. For further expansion of the project vehicle classification need to be done using Matlab.

References

1. "A Comparative Study of Edge Detection Techniques", Ranjeet Kumar Singh, Shashank Shekhar, Ram Bhawan Singh, Vinit Chauhan, International Journal of Computer Applications (0975 – 8887) Volume 100—No. 19, August 2014.

2. “Study and Comparison of Different Edge Detectors for Image Segmentation” By Pinaki Pratim Acharjya, Ritaban Das & Dibyendu Ghoshal Global Journal of Computer Science and Technology Volume XII Issue XIII Version I 30 (DFD) 2012.
3. “A Study of Edge-Detection Methods” Sonam Saluja, Aradhana Kumari Singh, Sonu Agrawal International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013.
4. “A Comparison of various Edge Detection Techniques used in Image Processing” G.T. Shrivakshan IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012 ISSN (Online): 1694-0814.
5. “Comparison of Edge Detection Technique in Image Processing Techniques” Ireyuwa. E. Igbinosa International Journal of Information Technology and Electrical Engineering Volume 2, Issue 1 ISSN:-2306-708X February 2013.
6. “A Survey on Edge Detection Techniques using Different Types of Digital Images”, R. Swarnalakshmi International Journal of Computer Science and Mobile Computing, Vol. 3 Issue. 7, July- 2014, pg. 694–699.
7. “Moving Object Detection Using Background Subtraction Algorithm Using Simulink” Mahamuni P. D, R. P. Patil, H.S. Thakar International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308 Volume: 03 Issue: 06 | Jun-2014.
8. “Research and analysis of Image edge detection algorithm Based on the MATLAB” Guowei Yang, Fengchang Xu/XPuro/c Procedia Engineering 15 (2011) 1313–1318. elsevier.
9. “Morphological Change Detection System for Real Time Traffic Analysis” Anuradha S. G., K. Karibasappa and B. Eswar Reddy International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 8 (2015), pp. 143–150.
10. “Object Counting using MATLAB”, Varun Sharma, International Journal of Scientific & Engineering Research, Volume 5, Issue 3, March-2014 614 ISSN 2229-5518.
11. Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance, Sepehr Aslani, Homayoun Mahdavi-Nasab, World Academy of Science, Engineering and Technology International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol: 7, No: 9, 2013.
12. Moving Object Tracking using Gaussian Mixture Model and Optical Flow, Abhishek Kumar Chauhan, Prashant Krishan International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 4, April 2013.
13. Traffic Image Processing System, Pranav Maheshwari, Deepanshu Suneja, Praneet Singh, Yogeshwar Mutneja, International Journal of Computer Applications (0975 – 8887) Volume 118—No. 23, May 2015.
14. Computer Vision Based Vehicles Detection and Counting for Four Way Traffic Meru A. V, Mujawar I. I. International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 2, February 2015.
15. Automatic Vehicle Detection and Counting Algorithm, Hyeok Jang, In-Su Won, Dong-Seok Jeong IJCSNS International Journal of Computer Science and Network Security, VOL. 14 No. 9, September 2014.
16. Moving Object Detection Using Matlab, Gottipati Srinivas Babu, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August—2012.

Neuro-Genetic Hybrid Approach to Predict siRNA Efficacy in Plant Viruses

Ranjan Sarmah, Shahin Ara Begum and Mahendra K. Modi

Abstract RNAi represents an extraordinarily powerful method to silence the gene expression using synthetic plasmid or viral encoded small interfering RNAs (siRNAs). So, finding an effective siRNA against a specific mRNA from the branch of predicted sequences is found to be an intriguing challenge among the researchers till date. Although various siRNA datasets with their gene silencing efficacy are available in the mammalian cell, but the same is scanty in plant genomes. The present study involves the development of a hybrid neuro-genetic approach, i.e., optimization of Back Propagation Neural Network (BPNN) model combining with Genetic Algorithm (GA) to predict siRNA silencing efficacy targeting against various plant viruses. BPNN model has been developed based on three different features, namely nucleotide position preferred, nucleotide position avoided, and thermodynamic characteristics. Identical methodology have been employed with a published dataset of 240 HIV siRNA sequences with silencing efficacy to compare the efficiency of the proposed model. BPNN performance is further optimized by employing GA for weight initialization. GA-BPNN has shown more promising results compared to BPNN in terms of accuracy and potential approaches.

Keywords Plant viruses · siRNA · Efficacy · GA-BPNN

1 Introduction

Given their emerging and re-emerging nature, viruses like *grassy stunt virus*, *Oryza sative endorna virus*, *Arobdiossis thaliana*, *Saccharomyces cerivisiae*, etc., pose a global plant health concern. In the absence of therapeutics against majority of

R. Sarmah · S.A. Begum (✉)
Department of Computer Science, Assam University, Silchar 788011, Assam, India
e-mail: shahin.begum.ara@gmail.com

M.K. Modi
Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, Assam, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_58

viruses there is a pressing need of developing antiviral agents which are more effective [1]. Targeting plant viruses, RNAi has surfaced as a prospective therapeutic tool. RNAi is a process through which a specific sequence of messenger RNA is degraded [2]. This process involves the use of ribonucleus enzyme as a dicer, which slices the long dsRNA precursors into shorter (19–23) units. Notably, in RNAi pathway siRNA plays a significant role. In order to determine the chromatin structure of a genome in RNAi-related pathways siRNA also acts as an antiviral mechanism.

The siRNA is usually characterized by a short RNA sequence with nineteen nucleotides and a $3' \, 2 - \text{n}$ T overhang. Based on certain siRNA functional features, many siRNAs can be selected, synthesized, and used to knock down the targeted gene expression. Therefore a methodology which may increase the probability of choosing an efficient siRNA can, to a larger extent slash down the cost of experimentation and time required for validation.

In order to predict effective siRNAs, innumerable siRNA designing rules as well as prediction algorithm have been proposed [3, 4]. Based on different parameters like presence and/or absence of specific nucleotides at different positions, stable thermodynamic parameters, etc., by different workers, namely Reynolds et al. 2004, Ui-Tei et al. 2004, Amarzguioui et al. 2004, Klingelheafer et al. 2009, Liu et al. 2012, Wang et al. 2010, Shabalina et al. 2006 [4–10]; have reported their siRNA design guidelines. So, it is vital to select better functional features for designing of effective siRNA since every siRNA design guideline against a given mRNA are not effectual in the same way.

To select effective siRNA with a higher knock down efficacy of target genes from a larger set of siRNA database, different machine learning algorithms has been attempted. To predict effective siRNA and their silencing efficacy various machine learning techniques like Back Propagation Neural Network (BPNN), General Regression Neural Network (GRNN), and Support Vector Machine (SVM) have been used and the performance shown by these methods have proved to be better than general siRNA design rules [11–15].

The predictive capabilities of neural network models have been effectively proven by different authors [16–22]. Heusken has reported about the screening method of functional siRNAs with the use of Artificial Neural Network [23]. A combination of features employing different machine learning techniques like ANN, SVM, KNN, and REP tree have been used by many researchers in order to predict siRNA efficacy [15, 24–26]. A work on Artificial Neural Network application based on thermodynamics and composition features has been carried out by Shabilina where the results found were compared with biopredsi [10].

The present study seeks to assess the performance of the GA-based BPNN, i.e., GA-BPNN compared to BPNN. In this study, different position specific and thermodynamic features have been used. The parameters selected from these features are the input parameter to the BPNN model. Further, to optimize the Neural Network structure, the number of nodes in the hidden layer is determined by using Moody criteria [27]. Genetic Algorithm has been used for weight initialization and to optimize the performance of BPNN model. To assess the performance of

GA-BPNN in comparison to BPNN, a published dataset of 240 HIV siRNA data are considered.

2 Materials and Methods

2.1 Preparation of Dataset

The training dataset is a collection of 305 siRNA sequences that includes 44 nos. from *Rice grassy stunt* virus [28], 30 nos. from *Oryza sativa endorna* virus [28], 191 nos. from *Arabdiosis thaliana* [29], 40 nos. *Saccharomyces cerivisiae* [28]. The numerical (quantitative) efficacy of the considered 305 siRNA sequences has been calculated by considering inhibitory-Score algorithm of i-Score [30] and s-Biopredsi.

Furthermore, a known dataset of HIVsirDB consisting of 240 siRNA sequences with efficacy was extracted [31] for cross validation of the model performance.

2.2 siRNA Features Selection

Three different siRNA sequence features such as nucleotide position specific features (preferred and avoided) and thermodynamic properties have been considered in the selected dataset. A total of 14 numbers of parameters for nucleotide position preferred and 15 numbers of parameters for nucleotide position avoided have been considered in the studied dataset as shown in Table 1. Previous workers have recognized that the thermodynamic stability plays a significant role in the effective designing of siRNA [10, 15]. A total of 11 numbers of parameters are considered for thermodynamic features (Table 1). All the selected parameters have been considered as input to the BPNN network and the efficacies are considered as the target to the network.

2.3 Normalization of Dataset

Before network training, the siRNA data have been normalized between zero to one by using Eq. (1):

$$Y = (X - \min p)/(\max p - \min p) \tag{1}$$

Table 1 Position dependent nucleotide preferences, avoidances, and thermodynamic characteristics

Nucleotide and their respective positions		Thermodynamic features
Preferred	Avoidance	
U at position 1	C at position 1	ΔG of sense-antisense of siRNA duplexes (T_1)
C at position 19	A at position 19	Intra molecular structure stability of siRNA antisense strand (T_2)
U at position 13	G at position 1	Dinucleotide ΔG values at the 5′ ends (T_3)
U at position 2	G at position 2	Dinucleotide ΔG values at the 3′ ends (T_4)
A at position 7	G at position 7	Internal Stability for antisense positions 1 (T_5) Internal Stability for antisense positions 2 (T_6) Internal Stability for antisense positions 6 (T_7) Internal Stability for antisense positions 13 (T_8) Internal Stability for antisense positions 14 (T_9) Internal Stability for antisense positions 18 (T_{10})
U at position 7	G at position 13	
A at position 1	G at position 6	
A at position 10	C at position 7	
U at position 14	A at position 17	
C at position 17	A at position 5	
A at position 14	G at position 14	
U at position 3	C at position 14	
C at position 18	A at position 11	
U at position 5	G at position 16	
	U at position 19	Difference of ΔG between position 1 and position 18 (T_{11})

where, X is the value of the respective data, Y is the normalized value, max$p$ is the maximum value of the sample data, and min$p$ represents the minimum value of sample data. Then the dataset is randomly divided into 80% and 20% of total data for training and a testing, respectively.

2.4 Network Architectures and Learning Methods

In the present study, BPNN model has been implemented with a single hidden layer as shown in Fig. 1.

2.5 Weights and Other Parameters

For optimization of the networks, different parameters like number of neurons in the hidden layer and epochs, momentum constant (MC), learning rate (LR) are

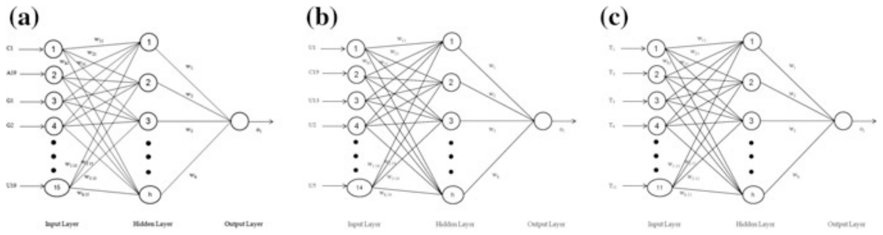


Fig. 1 **a** Nucleotide position avoided. **b** Nucleotide position preferred. **c** Thermodynamic features

initialized and modified with learning algorithm. The number of parameters in the input layer for nucleotide position preferred, nucleotide positions avoided and thermodynamic features are 14, 15, and 11, respectively. There is a single output layer relating to the silencing efficacy of siRNA sequences. The training process has been carried out for several number of times to obtain an optimized network by varying the number of nodes in the hidden layer, momentum values, learning rate, and number of iterations.

2.6 Identification of Model Performance

The optimum network model have been recognized on the basis of performance obtained by root mean square error (RSME) shown in Eq. (2) and correlation coefficient (CC) shown in Eq. (3) between target and predicted values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(o_i - t_i)^2} \tag{2}$$

$$\text{CC} = \frac{\sum_{i=1}^{n}(o_i - \bar{o})(t_i - \bar{t})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2 \sum_{i=1}^{n}(t_i - \bar{t})^2}} \tag{3}$$

where, o_i represents the network output and t_i represents the target output for the ith input pattern. On the other hand, $\bar{o}$ represents the average of network outputs, $\bar{t}$ is the average of target outputs, and n represents the total number of events.

3 Back Propagation Neural Network (BPNN)

Back propagation neural network offers most excessively used neural network learning algorithm called the Back propagation (BP) algorithm. Back propagation prevails over the constraints of single-layer networks and made an awesome breakthrough from the *single-layer perceptron* network with a sophisticated learning rule. BPNN is also considered as the most suitable learning method for multilayer networks.

The BP algorithm trains a multi-layered feed-forward neural network for a set of input samples with familiar classifications. For each input given to the network, the network verifies the output response corresponds to its input sample and finally calculates the error value by comparing the known and desired output. On the basis of error, the weights are then adjusted. Here, in this study, *Widrow-Hoff delta learning rule* is used for back propagation algorithm where the weight adjustment is done through *mean square error* (*mse*) of the output obtained against the sample input. The set of these samples is repeatedly provided to the network till minimization of error.

For each unit in hidden layer, the calculation of weighted net output corresponds to the input unit is given by Eq. (4):

$$\mathrm{net}_j = \sum_{i=1}^{n} x_i w_{ij} + w_0 = \sum_{i=0}^{n} x_i W_{ij}, \tag{4}$$

where, net_j represents the jth hidden layer value and x_i represents the ith unit value in input layer. W_{ij} represents the connection weight between these two units. W_0 represents the bias.

The diagrammatic representation BPNN model is shown in Fig. 1.

The values of the independent parameters C1, A19, G1, G2, … for nucleotide position avoided, U1, C19, U13… for nucleotide position preferred, and T1, T2, T3… for thermodynamic features are fed into the network through input layer nodes. *Levenberg-Marquardt-Optimization* algorithm is used to train the network and produces an expected result (o_1) in the output layer. The number of nodes in the hidden layer is determined by using *Moody criteria* as shown in the Eq. (5):

$$n = \sqrt{n_i + n_0} + a \tag{5}$$

where, n_i represents the input layer nodes; n_0 represents the output layer nodes; and 'a' is a constant between 1 to 10.

In Fig. 1, w_{11}, w_{12}…, are the weights between hidden and input layer; w_1, w_2... are the weights between hidden and output layer. *Nguyen-Widrow* [32] method is used to generate the initial weights.

The network includes different activation functions *viz.*, log-sigmoid, tan-sigmoid and linear (given by Eqs. 6, 7 and 8 respectively) in the hidden and output layer with different values of numbers of neurons, epochs, LR and MC as shown in Table 2.

Table 2 Network parameters

Activation function		No. of hidden neuron	Learning rate	Momentum	Epochs
Hidden layer	Output layer				
Log-sigmoid	Log-sigmoid	5–14	0.1–0.9	0.1–0.9	1000–5000
Tan-sigmoid	Linear				

$$f(x) = \frac{1}{(1 + \exp(-ax))}, (\log - \text{sigmoid}) \tag{6}$$

$$f(x) = \frac{(1 - \exp(-ax))}{(1 + \exp(-ax))}, (\text{hyperbolic tangent sigmoid} (\tan - \text{sigmoid})) \tag{7}$$

$$f(x) = x, \quad (\text{linear}) \tag{8}$$

where, a represents the slope parameter of the sigmoid function.

The output (y_j) of the jth hidden node is given by Eq. (9):

$$y_j = f\left(\sum_{i=1}^{n} w_{ji} x_i\right) \tag{9}$$

where, x_i represnts the input values; n represents the number of input nodes; w_{ji} represents the weight between ith input and jth hidden node and f is the activation function associated with jth hidden node.

The network output is obtained by the Eq. (10)

$$o_1 = \left(\sum_{j=1}^{h} w_j y_j\right) \tag{10}$$

where, w_j represents the weight between the jth hidden node and output unit and h represents the number of hidden nodes.

Some of the BPNN training and testing cases for siRNA efficacy prediction of plant dataset and HIV dataset with three different features are shortlisted in Tables 3 and 4. The short list is done from 540 tested cases of BPNN in both the cases. The best case among the three features of the BPNN model performance is graphically represented in Fig. 2a, b for plant siRNA and HIV dataset, respectively.

Table 3 Training and Testing results of BPNN for Plant dataset (Neuron = 11, MC = 0.6, Epoch = 2500)

Features	Training/ Testing	RMSE/ CC	LR 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Nucleotide position avoided	Training	RMSE	0.0896	**0.0731**	0.0775	0.0785	0.0859	0.1112	0.0999	0.0975	0.0836
		CC	0.8367	**0.9092**	0.8966	0.8938	0.8749	0.7774	0.8170	0.8340	0.8775
	Testing	RMSE	0.0972	**0.0828**	0.1011	0.0833	0.0815	0.1211	0.1012	0.1008	0.0937
		CC	0.8136	**0.8749**	0.8063	0.8764	0.8521	0.7556	0.8056	0.8263	0.8656
Nucleotide position preferred	Training	RMSE	0.1221	0.1191	0.1180	0.1215	0.1205	0.1217	**0.1145**	0.1191	0.1300
		CC	0.7191	0.7486	0.7294	0.7400	0.7293	0.7298	**0.7493**	0.7486	0.6890
	Testing	RMSE	0.1245	0.1226	0.1219	0.1236	0.1231	0.1233	**0.1189**	0.1226	0.1331
		CC	0.7045	0.7368	0.7216	0.7351	0.7189	0.7197	**0.7398**	0.7369	0.6803
Thermodynamic features	Training	RMSE	0.1042	0.1034	0.1039	**0.1015**	0.1036	0.1052	0.1019	0.1064	0.1083
		CC	0.8094	0.8073	0.7941	**0.8146**	0.8131	0.7860	0.8057	0.7932	0.7936
	Testing	RMSE	0.1102	0.1087	0.1093	**0.1076**	0.1090	0.1115	0.1081	0.1123	0.1201
		CC	0.7968	0.7897	0.7836	**0.8098**	0.8019	0.7727	0.7981	0.7829	0.7865

Table 4 Training and testing results of BPNN for HIV dataset (Neuron = 11, MC = 0.6, Epoch = 2500)

Features	Training/ Testing	RMSE/ CC	LR 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Nucleotide position avoided	Training	RMSE	0.1129	0.1037	0.0929	0.0985	**0.0874**	0.1053	0.1062	0.0995	0.0965
		CC	0.4040	0.5153	0.6453	0.6210	**0.6555**	0.5228	0.5329	0.6021	0.6458
	Testing	RMSE	0.1219	0.1128	0.1037	0.1012	**0.0963**	0.1165	0.1157	0.1015	0.1009
		CC	0.3936	0.5049	0.6343	0.6131	**0.6449**	0.5112	0.5211	0.5896	0.6397
Nucleotide position preferred	Training	RMSE	0.1062	0.1015	0.1137	0.1030	0.1045	0.1084	0.1130	0.1070	**0.1004**
		CC	0.5073	0.5043	0.4076	0.4433	0.5027	0.3997	0.4274	0.4603	**0.5775**
	Testing	RMSE	0.1102	0.1130	0.1254	0.1124	0.1169	0.1174	0.1211	0.1151	**0.1086**
		CC	0.4958	0.4971	0.3961	0.4319	0.4938	0.3869	0.4128	0.4511	**0.5668**
Thermodynamic features	Training	RMSE	0.0999	0.0991	**0.0990**	0.1001	0.1071	0.0997	0.1042	0.1054	0.1061
		CC	0.5972	0.6009	**0.6012**	0.5963	0.5128	0.6004	0.5817	0.5791	0.5699
	Testing	RMSE	0.1109	0.1098	**0.1085**	0.1106	0.1166	0.1105	0.1129	0.1163	0.1169
		CC	0.5866	0.5918	**0.5919**	0.5851	0.5037	0.5910	0.5724	0.5701	0.5603

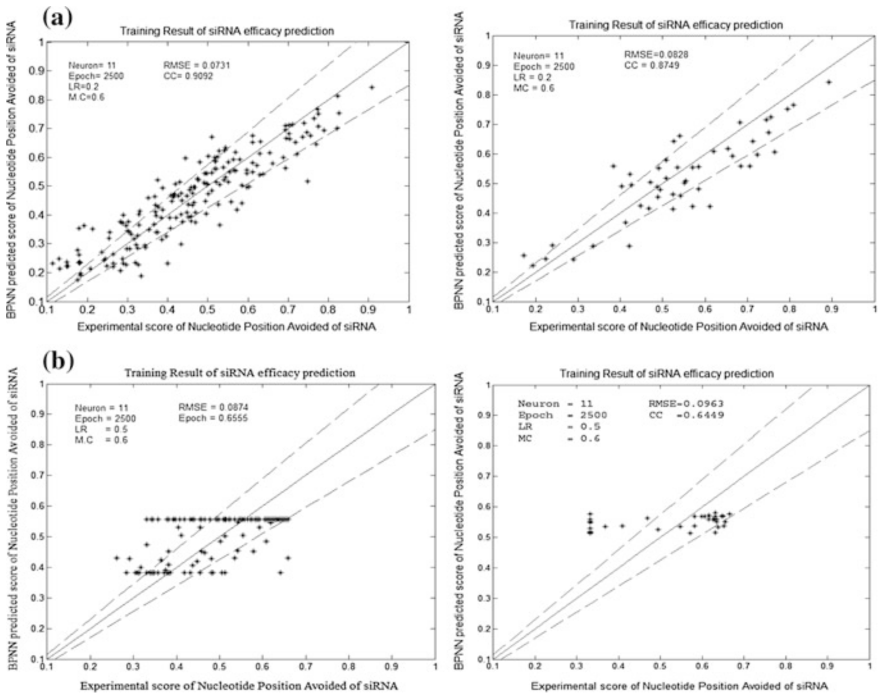


Fig. 2 **a** BPNN model performance for training and testing of siRNA efficacy prediction for plant dataset. **b** BPNN model performance for training and testing of siRNA efficacy prediction for HIV dataset

4 Weight Optimization Using Genetic Algorithm (GA)

The model performance of BPNN can further be optimized by initialization of weight using Genetic Algorithm. GA is a computerized search technique and is used extensively among optimization algorithms based on natural genetics and selection mechanics [33, 34]. GAs begins with a random population of promising results to a problem referred to as chromosomes. A fitness function can be used to evaluate chromosome representation. Here, in this study, the fitness function is the average deviation calculated between target and predicted values of siRNA efficacy. After calculating the fitness values, new chromosomes are obtained by choosing two chromosomes and then apply crossover and mutation operations. This process would continue unless a predefined termination criterion is satisfied.

GA-BPNN weight optimization process mainly has three stages. In the initial stage, connection weights of each neuron have been represented as gene segments. Fitness evaluation of these connection weight has been obtained in the second stage by constructing the corresponding neural network. The inverse of the error function is considered as the fitness function shown in Eq. (11)

Table 5 GA-BPNN training-testing results for plant siRNA dataset

Epoch	Nucleotide position avoided				Nucleotide position preferred				Thermodynamic			
	Training		Testing		Training		Testing		Training		Testing	
	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
180	0.0838	0.8733	0.0852	0.8706	0.1177	0.7561	0.1129	0.7502	0.1041	0.8072	0.1059	0.7988
200	0.0801	0.8947	0.0822	0.8898	**0.1010**	**0.8116**	**0.1075**	**0.8097**	**0.1002**	**0.8191**	**0.1031**	**0.8163**
220	0.0760	0.9008	0.0801	0.9085	0.1117	0.7836	0.1186	0.7769	0.1024	0.8088	0.1053	0.8012
240	**0.0732**	**0.9102**	**0.0784**	**0.9093**	0.1188	0.7577	0.1229	0.7516	0.1037	0.8075	0.1044	0.8009
260	0.0768	0.8976	0.0810	0.8902	0.1054	0.8096	0.1110	0.8008	0.1045	0.8067	0.1061	0.7978

Table 6 GA-BPNN training-testing results for HIV siRNA dataset

Epoch	Nucleotide position avoided				Nucleotide position preferred				Thermodynamic			
	Training		Testing		Training		Testing		Training		Testing	
	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
180	0.0876	0.7453	0.1065	0.6259	0.0953	0.6696	0.1015	0.6552	0.0931	0.7059	0.0996	0.6967
200	0.0865	0.7486	0.1059	0.6351	**0.0917**	**0.6715**	**0.1001**	**0.6521**	**0.0908**	**0.7109**	**0.0973**	**0.7018**
220	0.0851	0.7558	0.1102	0.6218	0.1011	0.6588	0.1054	0.6476	0.0915	0.7189	0.0987	0.6984
240	**0.0820**	**0.7587**	**0.1029**	**0.6546**	0.0987	0.6659	0.1025	0.6507	0.0928	0.7174	0.0991	0.6975
260	0.0844	0.7525	0.1075	0.6270	0.1112	0.6617	0.1201	0.6419	0.1011	0.6908	0.1059	0.6854

$$E = \frac{1}{2}\sum_{p=1}^{N} (t^p - y^p)^2 \tag{11}$$

where, t^p represents the target, y^p represents the network output for pth training pattern, and N represents the total number of training patterns.

The last stage involves the application of selection, crossover and mutation process and this process would be repeated until the error is lesser than a predefined value.

On the other hand GA-BPNN learning process has two phases. In the first phase, GA is employed in the BPNN network to search for sub-optimal connection weights. And finally the adjustments of weights are performed by using back propagation algorithm (BP).

Some of the best results of training and testing of the network are reported in Tables 5 and 6 for Plant and HIV dataset, respectively.

From Table 5 and 6, it is found that the hybrid GA-BPNN model shows better results than BPNN model. The best case among the three features of the GA-BPNN model performance is graphically represented Fig. 3 a, b for Plant siRNA dataset

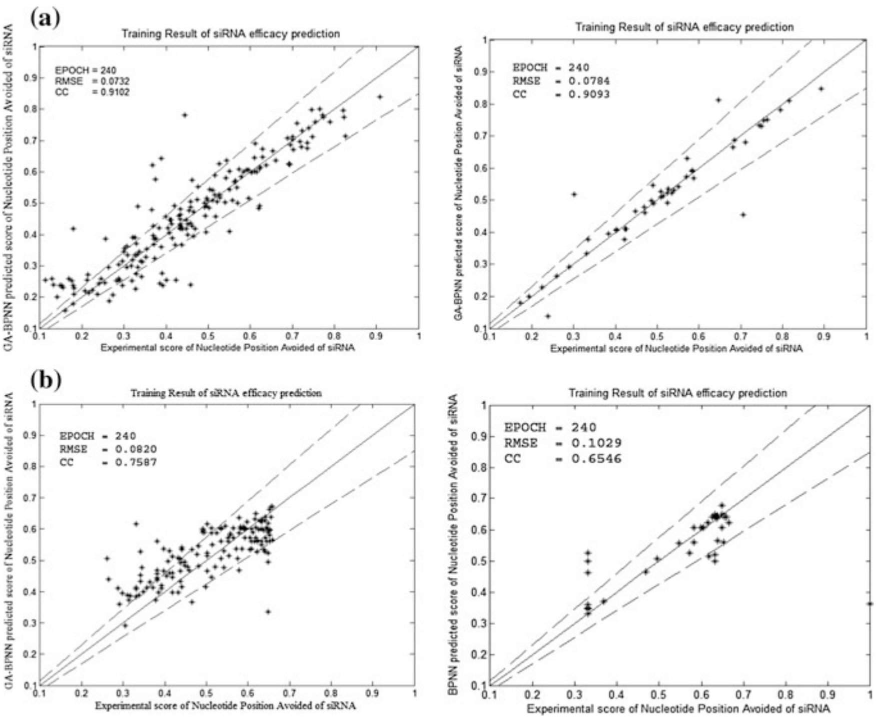


Fig. 3 **a** BPNN training-testing results of siRNA efficacy prediction for plant dataset. **b** GA-BPNN training-testing results of siRNA efficacy prediction for plant dataset

and HIV dataset, respectively. The initial population size was 25 and BPNN was run by single iteration for every individual chromosome (i.e., weights). In the reported table, the number of iterations is represented by epoch.

5 Conclusion

The present investigation has been initially involved in the implementation of BPNN model to predict the efficacy of 305 plant siRNA sequences. To cross validate the proposed model performance, a HIV dataset of 240 siRNA sequences with silencing efficacy has been taken into consideration. Three different features *viz*. nucleotide position avoided, nucleotide position preferred, and thermodynamics are considered for the proposed study and 15, 14, and 11 numbers of parameters, respectively are selected as the input parameter to the model. BPNN model performance has further been optimized in combination with the GA that performed the weight optimization of network connection. In both the dataset under consideration for siRNA efficacy prediction, the hybrid neuro-genetic model, i.e., GA-BPNN have provided better results as compared to BPNN model.

Further experimentation may be carried out with features selection which may help to reduce computational time while improving classification accuracy.

References

1. S.K. Mann, P.L. Kashyap, G.S. Sanghera, G. singh and S. Singh, "RNA Interference: An eco-friendly tool for plant disease management," Transgenic Plant Journal, Global Science book, pp. 110–126, 2008.
2. A. Fire, S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver and C.C. Mello, "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans," Nature, vol. 391, pp. 806–811, 1998.
3. S. Takasaki, Y. Kawamura and A. Konagaya, "Selecting effective siRNA sequences by using radial basis function network and decision tree learning," BMC Bioinformatics, vol. 7(Suppl 5), pp. S22, 2006.
4. A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall and A. Khvorova, "Rational siRNA design for RNA interference," Nature Biotechnology, vol. 22, pp. 326–330, 2004.
5. K. Ui-Tei, Y. Naito, F. Takahashi, et al., "Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference," Nucleic Acids Res., vol.32, pp. 936–948, 2004.
6. M. Amarzguioui and H. Prydz, "An algorithm for selection of functional siRNA sequences," Biochemical and biophysical research communications, vol. 316(4), pp. 1050–8, 2004.
7. J. W. Klingelhoefer, L. Moustsianas and C. Holmes, "Approximate Bayesian features selection on a large meta-dataset offers novel insights on factors that effect siRNA potency," Bioinformatics, vol. 25, pp. 1594–601, 2009.
8. Q. Liu, H. Zhou, J. Cui et al., "Reconsideration of in-silico siRNA design based on feature selection: a cross platform data integration perspective," PLoS One, vol. 7, e37879, 2012.

9. L. Wang, C. Huang and J.Y. Yang, "Predicting siRNA potency with random forests and support vector machines," BMC Genomics, vol. 11(Suppl.3), S2, 2010.
10. S.A. Shabalina, A.N. Spiridonov and A.Y. Oguresov, "Computational models with thermodynamic and composition features improve siRNA design," BMC Bioinformatics, vol. 7(65), pp. 1–60, 2006.
11. R. Teramoto, M. Aoki, T. Kimura and M. Kanaoka, "Prediction of siRNA functionality using generalized string kernel and support vector machine," FEBS Lett., vol. 579, pp. 2878–2882, 2005.
12. M. Mysara, J.M. Garibaldi and M. Elhefnawi, "MysiRNA-designer: a workflow for efficient siRNA design," PLoS One, vol. 6, pp. e25642, 2011.
13. A. S. Peek, "Improving model predictions for RNA interference activities that use support vector machine regression by combining and filtering features," BMC Bioinformatics, vol. 8, pp. 182, 2007.
14. J. P. Vert, N.Foveau, C. Lajaunie and Y. Vandenbrouck, "An accurate and interpretable model for siRNA efficacy prediction," BMC Bioinformatics, vol.7, pp. 520, 2006.
15. A. Qureshi, N. Thakur and M. Kumar, "VIRsiRNApred: a web server for predicting inhibition efficacy of siRNAs targeting human viruses," Journal of Translational Medicine, vol. 11, pp. 305, 2013.
16. G. Ge, G.W. Wong and B. Luo, "Prediction of siRNA knockdown efficiency using artificial neural network models," Biochemical and Biophysical Research Communications, vol. 336, pp. 723–728, 2005.
17. A.M. Chalk, C. Wahlestedt and L.L. Sonnhammer, "Improved and automated prediction of effective siRNA," Biochemical and Biophysical Research Communications, vol. 319, pp. 264–274, 2004.
18. M.C. Giddings, A.A. Shah and O.V. Matveera, "Artificial Neural Network Prediction of antisense ologodeoxy nucleotide activity," Nucleic Acids Res., vol. 30, pp. 4295–4304. 2002.
19. P.D. Mosier and P.C. Jurs, "QSAR/QSPR Studies using probabilistic neural network and general regression neural networks," J. Chem. Inf. Comput. Sci., vol. 42, pp. 1460–1470, 2002.
20. T. Niwa, "Using general regression and probabilistic neural network to predict human intestinal absorption with topological descriptors derived from two dimensional chemical structures," J. Chem. Inf. Comput. Sci., vol. 43, pp. 113–119, 2003.
21. M. Mohamed, M. Elhefnawi and J.M. Garibaldi, "MysiRNA: Improving siRNA efficacy prediction using a machine-learning model combining multi-tools and whole stacking energy (ΔG)," J. of Biomedical Informatics, vol. 45(3), pp. 528–534, 2012.
22. R. Sarmah, S.A. Begum and M.K. Modi, "In SILICO Approach to Design Potential Small Interfering RNA (siRNA) against Plant Pathogens and Diseases," Int. J. Inv. Res. & Dev., vol. 4(12), pp. 93–107, 2015.
23. D. Huesken, J. Lange, C. Mickanin, J. Weiler, F. Asselbergs, J. Warner et al., "Design of a genome-wide siRNA library using artificial neural network," Nat Biotechnol., vol. 23, pp. 995–1001, 2005.
24. W.J. Pan, C.W. Chen and Y.W. Chu, "siPRED: predicting siRNA efficacy using various characteristic methods," PLoS One, vol. 6, e27602, 2011.
25. O. Filhol, D. Ciais, C. Lajaunie, P. Charbonnier, N. Foveau, J.P. Vert and Y. Vandenbrouck, "DSIR: assessing the design of highly potent siRNA by testing a set of cancer-relevant target genes," PLoS One, vol. 7, e48057, 2012.
26. S. Sciabola, Q. Cao, M. Orozco, I. Faustino and R.V. Stanton, "Improved nucleic acid descriptors for siRNA efficacy prediction," Nucl. Acids Res., vol. 41, pp. 1383–94, 2013.
27. J. moody, in: From Statistics to neural networks, edited by V. Vherkassky, J.H. Friedman and H. Wechsler, Part II, Springer Publications (1994).
28. http://www.ncbi.nlm.nih.gov/probe/?term=siRNA.
29. http://asrp.danforthcenter.org/.

30. M. Ichihara, Y. Murakumo, A. Masuda et al., “Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities,” Nucl. Acids Res., vol. 35(18), pp. e123, 2007.
31. http://crdd.osdd.net/servers/virsirnapred/dataset.php?dataset=240_hivsir.
32. D. Nguyen and B. Widrow, “Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights,” Proceedings of the International Joint Conference on Neural Networks, vol. 3, pp. 21–26. 1990.
33. S. Rajasckaran and G.A. Vijayalakshmi Pai, “Neural networks, Fuzzy logic and Genetic algoriths, Eastern Economy Edition, 2003.
34. A. Fujail, S.A. Begum and A.K. Barbhuiya, “Neuro-Genetic approach to predict scour depth around vertical bridge abutment”, Proceedings of Fourth International Conference on Soft Computing for Problem Solving, vol.335 of the series Advances in Intelligent Systems and Computing, pp. 147–157, 2014.

Discovering Latent Social Networks Through Detection of Explicit and Implicit Taste from Online Reviews

Bhaskarjyoti Das and V.R. Prathima

Abstract Usage of community-related information can vastly improve effectiveness of customer-focussed applications such as recommendation systems. A key ingredient of community is homophily, i.e., the tendency of individuals to connect with similar individuals. A customer's taste defines such homophily. So any community built based on taste makes lot of sense. It is a challenge that only some part of taste may get explicitly declared by the customer while he leaves behind a taste footprint implicitly in online reviews. Typically community building mechanisms in online review portals are not based on taste and are only dependent on customers' declaration of fan followership or friendship. Communities built this way tend to cover only a small part of the customer population. In this paper, we present an alternative approach by putting the onus on the portal, i.e., an unsupervised framework to detect latent customer network and communities by discovering explicit and implicit tastes.

Keywords Key phrase · Annotation · Online reviews · Social network Community · Taste

1 Introduction

We can build a social network of the customers participating in an online review portal by representing them as nodes and by representing the observed connections or similarities a sedges between nodes. In these portals, the customers are encouraged to declare other customers as friends or declare themselves as "fan" of other reviewers. These declared relationships can define edges in the social network.

B. Das (✉) · V.R. Prathima
Department of Computer Science and Engineering, Rajiv Gandhi Institute of Technology, Bengaluru, India
e-mail: Bhaskarjyoti01@gmail.com

V.R. Prathima
e-mail: prathima.vr.2010@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_59

The current mechanisms for edge formation are driven by the customers but this approach so far has met with limited success. Traditionally, online review portals are rich in review content but not very famous for vibrant customer communities as in Facebook. Usually, communities built by such mechanism cover a small part of the total customer population and so applications like recommendation systems, marketing, etc., can only make a very limited utilization of social network structures. Our proposal is not to keep community building a customer driven process but to take a proactive approach to detect latent communities through detection of taste by using online review data.

2 Our Approach for Detecting Social Network

In this paper, we are proposing the concept of taste as the building block for homophily that is instrumental in community building. We propose to utilize the user generated content in the online review portals to uniquely identify the customers. The possible upside of this approach is enormous as customer identification is a formidable challenge for any business. But capturing customer's taste from unstructured content such as text review is a challenging problem. However, if this can be done, meaningful communities based on similarities of taste can be detected and put to good use by applications such as recommendation systems. For example, accurate understanding of a customer's taste can bring in the key element of serenditipity in a recommendation system.

In an online review forum, a customer willingly expresses his opinion mostly for things that he truly cares about. This is how a customer leaves a footprint of his taste in online text reviews. This part of his taste is implicit while the explicit part of it consists of preferences made evident by his choice of a category of business offering certain features. We need both to understand the customer's taste and can represent both as vectors. In our proposed framework, the taste vector for a customer is combination or concatenation of these two vectors. While the explicit taste vector can be easily formed from the customer preference data available in online review portals, creation of implicit taste vector is more involved. Our framework uses a two steps process to form the implicit taste vector. The first step attempts to summarize text reviews into key topics. We have evaluated and prototyped existing approaches before formulating ours which is discussed in details in a subsequent section. The second step consists of representing semantic content of identified key topics by suitable annotations or tags. Medelyan et al. [1], proved that automatically extracted keywords perform equally well compared to human annotated tags. In our adopted approach, forming the implicit taste vector is both a key topic extraction and text annotation problem.

After deriving the explicit and implicit taste vectors using our proposed framework, we examine, in a step by step manner, the advantages that we can gain in building a social network model for the customers. The adjacency matrix in a social network can be derived out of the similarity matrix if we assume that the calculated similarity between customers will lead to edge formation between customer nodes. We have looked at four options, i.e., the customer's self-declared relationships (such as friends), similarity based on multiple customers' shared relationship with the same businesses (for example, many customers ate in the same restaurant), a more detailed similarity based on attribute analysis of all the businesses visited by a customer (for example, known categories or attributes of all the restaurants a customer went to) and derived similarity based on implicit taste vector derived from online text reviews. The first option is the current mechanism of community detection in online review portals. The second and third options are the implementation of explicit taste while the last option deals with implicit taste vector derived from text reviews. The new edges and additional modularity that these incremental steps bring in can prove that taking this taste-based approach will uncover many new edges and consequently new communities that were unseen so far.

For the reasons discussed later in this paper, both automatic key topic extraction and automatic tag recommendation do not have any ready to implement and remarkably accurate solution. The existing researches work well with the research dataset for a particular domain but may not work well for every other domain. It is also hard to find a dataset from a similar domain to benchmark against. So, in our research, we have taken a light-weight approach based on prototyping with a rich real life dataset. We have selected restaurant business from Yelp Challenge Dataset as it fits the criteria well. We have taken the full review dataset for a particular restaurant to develop our taste framework and taken the full review dataset for restaurant business of a particular city (Edinburgh) in our final step of evaluating the four incremental strategies of building richer social network.

The rest of the paper is organized as follows. Firstly, we summarized the various researches and our strategy for key topic extraction from unstructured text. Secondly, we described our strategy for getting a suitable annotation to represent these key phrases so that an implicit taste vector can be formed. Thirdly, we presented our definition of explicit and implicit taste vectors, provided an example, and summarized our framework. Finally, we summarized our implementation and shared the analysis, i.e., the impact on community detection by examining the adjacency matrices formed by the four incremental approaches as indicated above. In the conclusion, we state our findings and indicate the next stage of our research.

3 Extracting Key Topics from Online Text Reviews

3.1 Evaluating Existing Strategies for Key Topic Extraction

The common approach irrespective of the domains involves finding the words or phrases representing different topics and sorting them using some logic to find the most important ones. The importance of topical phrases is determined by the discriminating power of the phrase instead of frequency alone. Additional challenges are the semantic aspect of the text and the inherently unstructured nature of online reviews.

Irrespective of the domain, extraction of key topics typically has two steps, i.e., finding the candidate key words or phrases and making a selection based on a criterion. The first step is usually a heuristics-based either on linguistic such as a combination of part of speech tags or on statistical analysis such as n-grams analysis. Necessary pre-processing consists of removal of stop words, unnecessary punctuation, control character, web URLs and HTML fragments. Syntactic metadata such as part of speech (POS) tag, can be an effective heuristic. We have various strategies (statistical, supervised, and unsupervised) for selection of key topics from the candidate set. In a statistical approach, appropriate term weighting system is needed while using frequency measure to choose top n candidates. TF IDF (term frequency, inverse document frequency)-based top n keywords has been popular but the generated list always is not extremely meaningful. Another approach has been to use an external resource such as Wikipedia to ascertain the importance of the candidate phrase. In non-negative matrix factorization (NMF), we can use dimensionality reduction techniques after TF-IDF-based processing. The statistical approaches produce somewhat acceptable results but they all suffer from the same shortcoming that it is hard to proxy the semantic aspect of the text by statistical means.

The key challenge in supervised approach is lack of availability of training data. The goal in this is to train a classifier on documents annotated with key phrases to determine whether a candidate phrase is a key topic or not [2]. However, we have decided not to focus on supervised approach in our investigation as in web-based customer review, getting an annotated domain specific training set is quite difficult and so this approach, though attractive, may not be practical.

For online review, unsupervised approach is more suitable as this does not require training data. This approach can be mainly classified into various dimensionality reduction techniques and graph theoretical approaches. In a graph theoretical approach, first, a cluster of documents is viewed as a network of information units and subsequently either clustering is attempted or a graph theoretical analysis is done. The success of clustering approach depends on the quality of clusters and unfortunately that is usually hard to achieve. The basic graph theoretical approach consists of representing text as a bag of words and then making a graph out of this by building a similarity matrix. The approaches vary in terms of how the candidate words or phrases are chosen in the first place, what similarity metric is used and on what basis the nodes in this graph will be ranked to find the top n key phrases. The candidate words can be n-grams, sentences as bag of words or specific parts of speech that are likely to form key phrases. The similarity metric can be cosine similarity when the nodes are sentences or co-occurrence when the nodes are words. We can model the importance of an information unit in terms of centrality in this graph and an optimum threshold can be chosen to extract the top n nodes. Alternatively, the concept of "prestige" is a more apt way to extract the influential key words. In Textrank, a text summarization algorithm, Mihalcea and Tarau [3] applied PageRank [4] on the graph of sentences after the sentences were represented as bag of words. In a dimensionality reduction approach, Blei et al. [5] proposed Latent Dirichlet allocation (LDA) which visualizes each document as a probabilistic distribution of topic and each topic as a probabilistic distribution of words. Finding key topics is the reverse of this, i.e., from words to topics. However, in LDA, there is a need of a fair amount of trial and error to come with topics containing words conveying consistent theme.

3.2 Our Approach for Key Topic Extraction

In our work, we made some extensions to the previously mentioned graph theoretical approaches (Fig. 1). Firstly, we have refined generation of the candidate key phrases by using sentiment/opinion analysis techniques. Based on the insight that subjective sentences contain more key phrases, we shrank the target review corpus retaining only subjective sentences containing polarized opinions about an object or its features. We did this by doing a sentence specific subjectivity analysis using a classifier (Naïve Bayes) trained on a standard dataset (IMDB movie review). In our sample restaurant dataset, we found 1777 subjective sentences and pruned 530 objective sentences. Secondly, we have extended the basic concept of Textrank for key topic extraction. We started with extracting key words defined by POS patterns

Algorithm for enhancing Textrank

- Take out raw sentences from raw review text
- Create index of tokens based on raw reviews
- Convert each sentence to a list of words
- Tokens are assigned POS(part of speech) tags
- Filter based on POS tags to extract adjective and noun tokens
- If Word2Vec :
  - Train Word2Vec model to build model vocabulary
  - Set of words from the above model is the candidate key word list
  - Use trained Word2Vec model to derive similarity between each pair of words
- If WordNet
  - Use WordNet to derive the semantic similarity between each pair of words
- Draw the graph of keywords with candidate word list by creating an edge when the similarity value is above a cut off.
- Apply Pagerank on the graph and sort on Pagerank. Take top n percent.
- Combine the pair of key words (into key phrases) which are adjacent as per index in preprocessed text
- Return the key phrases

Fig. 1 Algorithm for enhanced Textrank (adjective-noun pairs). We then applied Page Rank on this, sorted the keywords on Page Rank and chose to retain only certain top portion (say 70%) of this list. We then combined key words into key phrases by checking adjacency in the original text. Thirdly, we have further refined the approach by incorporating semantic distance in Textrank. We replaced the co-occurrence-based similarity metric in Textrank by the concept of semantic distance using WordNet and as a second alternative, we derived the similarity metric from a Word2Vec [6] model built based on the distributional properties of words in our corpus. In both cases, after replacing the similarity metric, we have applied PageRank on the graph as before. The enhanced Textrank brought in remarkable improvement to the result compared to pure TF-IDF-based strategy. The semantic similarity based on WordNet is quite meaningful but sadly its efficacy is limited by its vocabulary.

4 Finding Annotations to Represent Key Topics

An annotation is applied to a web document for various purposes, i.e., classifying a document to a category or sub-category, identifying sentiment polarity, identifying the skill level required from the reader, identifying the overall concept, or just a representative summary. In online review domain, coming up with annotation that represents a customer's taste is more of a summarization than categorization problem. Much of the existing research on automatic tag recommendation has heavy reliance on domain specific ontology. In the ontology hierarchy, parent child hierarchy is used to recommend a tag that represents a parent and to eliminate duplication, the common parent of two competing child tags is chosen from the ontology tree. While this ontology-based generalization approach works well for annotating a document, it does not work well when annotation is used to represent a taste. Hence we framed the problem of coming up with a suitable annotation as a second level of text mining problem where we need to reduce the dimensionality of the many key phrases we come up with in the first stage of the framework. Having already done a fair amount of mining to find meaningful keywords, our next task is to do a second level of mining to reduce duplications while retaining fair amount of details. Essentially, each selected key phrase is a small document in a document set that we have to further analyze. We have tried out three approaches. Firstly, we have done TF-IDF-based top 20 keyword extraction. Secondly, we have used non-negative matrix factorization (dimensionality reduction technique) to extract top 20 dimensions and top three keywords in each; we then extracted the top 20 words from the basket of words representing these dimensions. Finally, we have done K means clustering with the basket of words representing the extracted key phrases and picked up the central word for each of these 20 clusters, eliminating any duplicate. For us, the clustering strategy yielded good results.

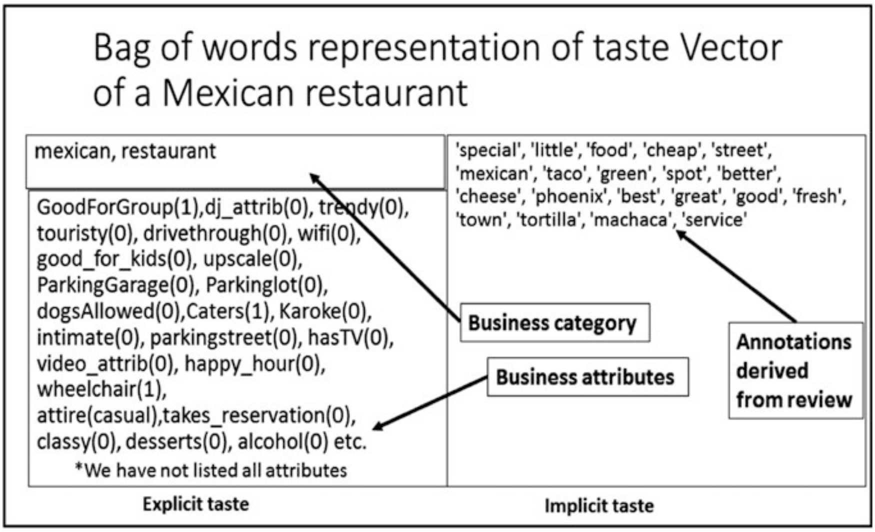


Fig. 2 Taste vector representation

5 Forming Taste Vectors and Summarizing the Framework

We have modeled explicit taste as a bag of words (Fig. 2) where each word is an explicitly known attribute of the business. In the restaurant dataset we worked with, these tags are available as category and attribute tags of the business. The category signifies the type of restaurant whereas the attribute tags are various facilities offered. When a customer visits many restaurants with these attribute/category tags, it shows an explicit pattern of his preferences. The explicit taste of a customer can be derived by aggregating his numerous review records.

For describing the implicit taste, we follow a similar bag of words strategy. Here the words are annotations derived by mining the review texts as described in the previous sections. For finding the implicit taste of a customer, we examine all the reviews by the specific customer. At the end, we come up with an implicit taste vector like a bag of words model for a document. This will vary from customer to customer. For example, for a customer in Texas who loves to frequent restaurants for grilled steak, will likely have “steak” and “grill” in his taste vector compared to another customer who is a Japanese food enthusiast having “sushi” in his implicit taste vector. The taste vector in our framework can be composed by appending implicit taste to the explicit taste.

The diagram above (Fig. 3) summarizes the proposed framework. To begin with, we have used classification-based sentiment analysis technique to filter out the text that is clearly subjective. After that, we have used unsupervised approach (graph theoretical method) to find the key topics. We have extended the Textrank

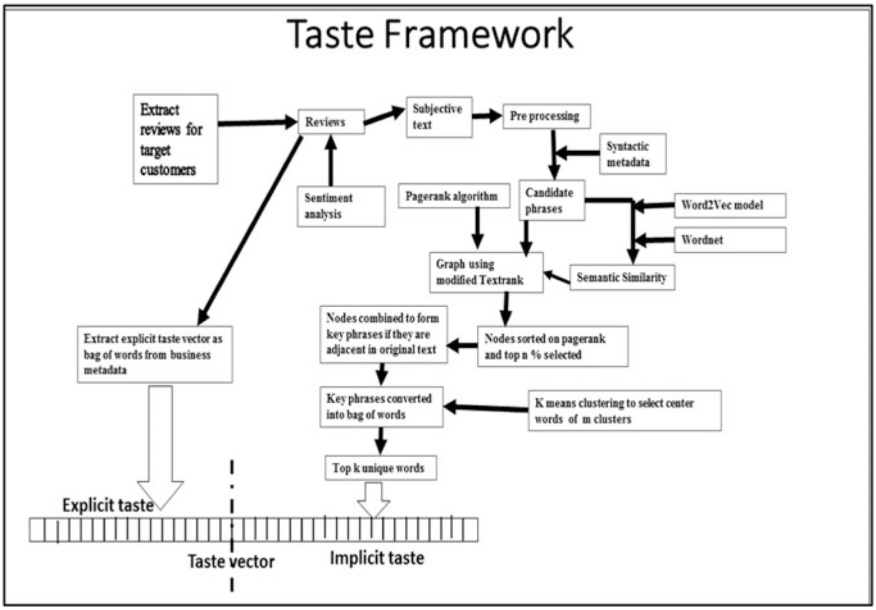


Fig. 3 Proposed framework to detect explicit and implicit taste

algorithm for key phrases (instead of sentences) and replaced the co-occurrence-based similarity metric with metric-based on semantic similarity. For automatic annotation, we treated this as a summarization problem instead of ontology- based categorization. We have addressed the use of techniques such as K means clustering on bag of words formed from selected key phrases.

6 Implementation and Analysis of the Result

In our implementation, first we analyzed the full Yelp Data Set so that we can make an appropriate small subset for us to work on. We concluded that Restaurant business has the maximum amount of businesses, reviews, and participating users and derived statistics for different cities along with the corresponding number of business, users, as well as text reviews. For the final stage of analysis, given the hardware constraint, we extracted the merged dataset for restaurant business in Edinburgh considering only businesses having more than 50 reviews and users having more than five reviews. At this stage, the Edinburgh dataset had 6470 users, 47 businesses, and 2353 reviews. Out of these users, only 1077 users participated in Edinburgh reviews. The user number is high as the friend relations expressed by many of these 1077 users span all over USA. The dataset is not balanced in the sense that the default friend network looks quite rich considering a small city like

Edinburgh. However, this perception gets corrected if we look at the overall friend relation participation across all of Yelp and that is very small.

In parallel, we made prototypes of various key phrase extraction schemes to experience the results first hand and formulate our strategy. We then followed up with an implementation of the taste framework that addresses both key phrase extraction and determination of tags to determine all customers' explicit and implicit taste vectors for Edinburgh.

In the final stage, we have used various open source libraries (Gephi, Networkx, SNAP) to do the social network analysis after forming the adjacency matrix in each case for Edinburgh dataset (Table 1). The option 1 depicts that declared friend network. Option 2 corresponds to similarity based on common businesses visited. Option 3 represents another dimension of this similarity by considering the category tags of these businesses. We have not yet considered the business attributes in our prototype but doing that will add a predictable column. Option 1 & 2 represent the explicit taste vector. Option 3 is an implementation of implicit taste vector. We see that adding these options drastically improved modularity of the social network and consequently number of detected social communities. We have chosen Louvain method for this detection and any other community detection method is expected to prove the same. Average clustering coefficients have continuously increased as options got added. We see that there is an increase of the total number of edges and average degree. We have to keep in mind that the figures for these values in the default option are looking healthy primarily because of the dataset we have chosen.

Table 1 Analysis of impact of taste vectors on a small dataset

		Default: Friend network	Similarity add on Option 1	Similarity add on Option 2	Similarity add on Option 3
Average degree	Both in and out degree are considered	17.684	8.065	189.956	8
Total no of edges	Total number of edges in the network	14,416	52,191	5,80,503	136
Maximum degree	Maximum degree in the degree distribution	2910	915	1078	17
Average clustering coefficient	Mean value of individual coefficients	0.063	0.840	1.0	0.944
Modularity	Louvain method [7]	0.543	0.554	0.046	0.116
No of communities		44	5410	2	3

7 Conclusion

In this paper, we have put the onus of detecting richer social network and communities on the online review portal instead of the existing practice of relying on the customers. Our hypothesis is: by analyzing the customer's taste footprint in the review portal's data, we can detect a far more extensive social network and communities which remain latent and are never accessed. By doing so the existing landscape of customer focused applications will vastly improve. In our implementation, we have built a framework to form the customer's taste vector and we proved our hypothesis by examining the additional richness of the social network that can be contributed by these detected taste vectors. Our research effort has been exploratory and the results are empirical due to lack of baseline dataset in such domains. As continuation of our research, we want to formalize our effort by measuring the effectiveness of our framework through building a recommendation system leveraging taste-based community and comparing with some traditional recommendation baseline. We also plan to investigate further on the concept of implicit taste and build more applications using machine learning techniques.

References

1. Olena Medelyan, Eibe Frank, and Ian H. Witten. Human-competitive tagging using automatic keyphrase extraction. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. 2009. p 1318–1327.
2. Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig G. Nevill-Manning. KEA: Practical automatic keyphrase extraction. In: *Proceedings of the 4th ACM Conference on Digital Libraries,* 1999. p 254–255.
3. Rada Mihalcea and Paul Tarau. Textrank: Bringing order into texts. In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. 2004. p 404–411.
4. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks*. 1998. 30 (1–7):107–117.
5. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*. 2003. 3:993–1022.
6. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean of Google Inc. Distributed representation of words and phrases and their compositionality. In: *Advances of Neural Information Processing Systems 26 (NIPS).* 2013.
7. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), p 1000.

Digitally Integrated Store Commerce System

Aritra Roy, Sauvik Bal and Sagarika Ghosh

Abstract This paper aims to identify certain advantages and pitfalls of the conventional store shopping and online shopping (e-commerce), and form a new hybrid system of commerce, incorporating the advantages of both parent systems and fixing their short comes. Customers can enjoy the advantages of both type of shopping with additional features like Pick-And-Leave (PAL). All the stores around the world can be provided with a single platform to integrate their business with the internet and take it to the global scale ensuring authenticity of products and cost efficient marketing. This system will provide a high level of data integration from the market for better analysis.

Keywords Physical store · Online store · E-billing · Data analysis
Individual preference · RFID tags · RFID reader · Wireless ZigBee module
Secure intelligent cart

1 Introduction

The birth of a need has always ensured the birth of new technology. And the modern human civilization has seen the advent of many cutting edge technologies being developed conquering the needs of the modern human being and modernizing the society. One routine activity which a human being spends considerable amount

A. Roy (✉) · S. Bal · S. Ghosh
Department of Computer Science & Engineering,
University of Engineering & Management, Jaipur 303807, India
e-mail: ari.roy09@gmail.com

S. Bal
e-mail: sauvikbal@gmail.com

S. Ghosh
e-mail: sagarikaghosh.wbut2010@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_60

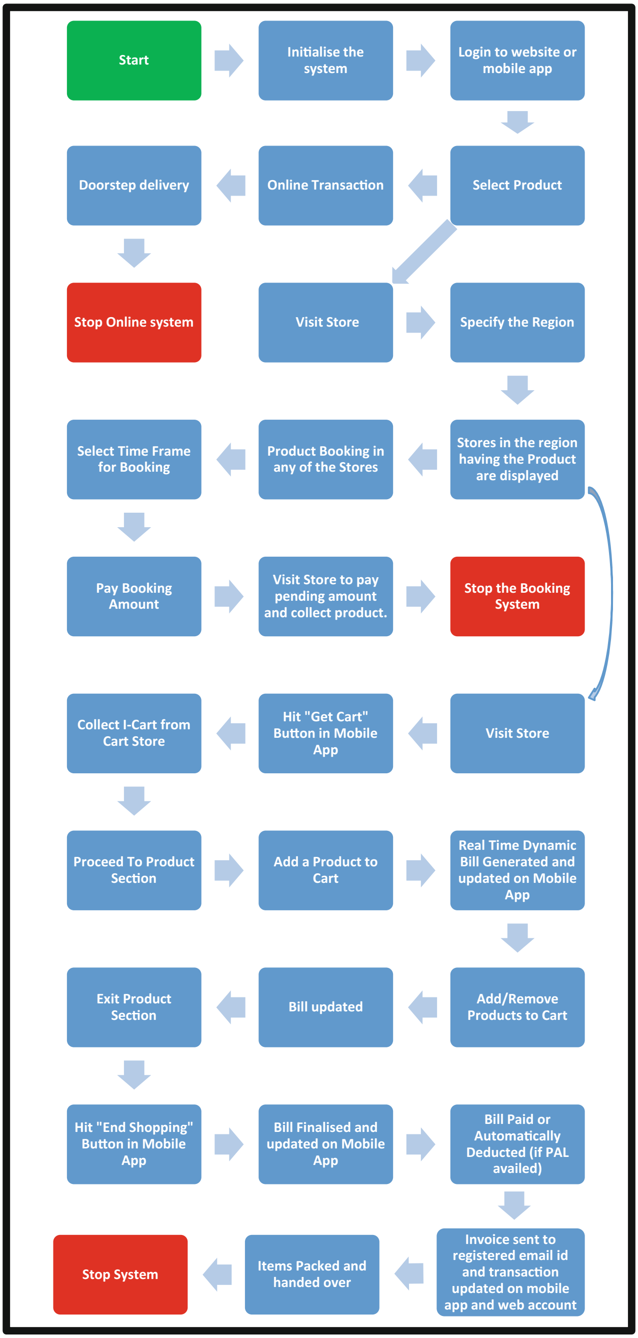
of the time and energy is Shopping. The introduction of e-commerce and the modernization of stores with product sensing and e-Billing using RFID and ZigBee have brought drastic development in shopping experience by minimizing the time and effort of customers and maximizing the options for choice of products [1]. The proposed system will fuse the features of shopping in absentia (e-commerce) and shopping in person to provide with an integrated shopping environment for the customers and the business owners. The proposed system comprises of physical stores, an online store that will synchronize the data from all physical stores, an automated store management system (ASMS) for the physical stores, a mobile application for the customers, a data analysis system integrated with the online store, an intelligent cart (I-Cart) and a cart store for the carts which shall be integrated with the mobile application. The automated physical store will bear the advantages of the conventional store shopping and the online store will bear the advantages of online shopping. The mobile app will be the connecting link between the physical and digital features of the system. The intelligent cart will have an RFID module integration to sense the products and facilitate automatic billing. All data in the processes will be encrypted. The website and the mobile app will provide a global platform for the stores enabling them to reach a larger number of customers and enhance their business. All the transaction data from the stores will be integrated in the online system and statistical analysis of this data will give a more accurate picture of the market and also behavior of the market based on regional, economical, seasonal and other constraints. Individual preference of a customer can also be tended to using the statistical data from the data analysis system.

2 Literature Survey

With the advent of online shopping [2], the store commerce around the world has faced many challenges. People prefer to shop online rather than facing the overcrowded and tiresome shopping from stores. Overcrowded billing counters in the malls and the tiresome work of going store to store to select a product discourage customers. There has been a considerable research on a secure smart shopping cart for easy billing and product identification. The use of RFID tags for product identification and billing makes shopping at stores much easy and swift [3, 4]. The security of the data is also ensured by the intelligent cart through high level encryption. But few of the aspects remain unattended in the present scenario of shopping in person like there is still no integration of the stores with the internet, individual preference of customers cannot be maintained and all the vital data about behavior of the market from the stores go unnoticed due to lack of integration and data management system.

3 Proposed System

A. Block Diagram of the System



B. Algorithm of the System

1. Start
2. Customer logs into his/her account through website or mobile app.
3. Customer selects product(s).
4. Customer chooses any one from online transaction or visit stores.
5. If online transaction is opted, the transaction is completed online according to existing procedures and product(s) will be delivered at doorstep.
6. Stop the online system.
7. If customer opts to visit stores, he/she specifies the region to look for stores.
8. All the stores in the specified region having the product(s) are displayed along with the price they offer, address of the store, special deals on the product(s), user rating of the store and ranking of the store according to the user rating of the store corresponding to the concerned product(s).
9. Customer chooses whether to book the product(s) in a store by advance payment or not.
10. If product booking is chosen, customer chooses a time frame varying from 12–48 h and accordingly a certain percentage of the price (percentage depends on the store) is paid by online transaction and the product(s) remains unavailable for sale for the next 12–48 h depending on the time frame selected.
11. Customer visits store.
12. If product(s) is (are) booked, customer pays the amount pending (if any) and takes product(s).
13. Stop booking process.
14. If product(s) is (are) not booked, customer goes to the Cart Store.
15. Customer enters registered customer id or username in the Cart Store interface or clicks the "Get Cart" button on the mobile interface.
16. An I-Cart linked with the customer id pops out of the cart store.
17. Customer takes the cart and goes for shopping the product(s) in the products section of the store.
18. Customer adds first product to the cart.
19. Cart sends product id to the reception application in the store.
20. A real time dynamic bill (RTDB) is generated for the customer.
21. Bill is updated and displayed in the mobile application and web interface, or is set as text in the customers mobile.
22. Customer adds or removes products from the cart.
23. Cart sends corresponding id of product leaving or entering the cart to the reception application.
24. Reception application updates bill and notifies the customer through mobile application or text messages and also syncs the data with the web application
25. Customer ends shopping by coming out of the product section and hitting the "End Shopping" button.

26. Reception application finalizes the RTDB and updates the same in the mobile application or texts the customer on the number registered.
27. Customer pays the bill in cash, cheque (depending on the store), card, online through mobile application or bill is auto deducted if pick ad leave service is availed.
28. Invoice is sent to the registered mail id and is updated on the mobile application and web server.
29. Items are packed and handed over to the customer.
30. Stop the system.

C. Elements of the System

Physical Stores: Physical Stores are independent business units integrated with the system. They can have sub-units functioning under them as independent business owners. They store products and carry out business according to their business policies. Though they are independent business units, yet the physical stores are a part of the system and thus have to conform to the common standard of the system. Physical Store refers to malls, marts, stores, shopping complex, etc. The Automated Store Management System (ASMS) in the physical stores handles all the transactions local to the store only.

Online Store: Online Store is the vital element in the system. It synchronizes all other elements in the system and maintains all data. It interacts with both the Physical Store and the Mobile App. The Online Store Server stores data from all Physical Stores and Mobile Apps. It hosts the Web Interface for the users. It maintains user accounts of customers where all the details of the customer and his/her transactions are maintained, and Personalized Shopping Experience is provided. The Online Store System handles all the Global Transactions. It provides the foundational structure for the complete system. It comes integrated with many features like Navigation to locate nearby stores, rating of stores, filters, etc.

Mobile App: The Mobile App provides, mobile and user friendly user interface to the users. It serves as a connecting link between physical and digital features while shopping. It interacts with both Physical Store and the Online Store, depending on the type of interaction.

Automated Store Management System: It comprises of an Intelligent Cart (I-Cart), a Reception Application and Server. It is responsible for the complete functioning and management of the Physical Store.

I-Cart: It is a simple shopping cart equipped with a sensor which senses different products when added to cart [5]. Whenever a product is added to an activated I-Cart, it senses the product and transmits the product code to Reception App.

Reception App: It is responsible for the complete auto-management of the Physical Stores. Its scope is limited to the Store in which it is installed. It maintains constant relation with the I-Cart and Customers' Mobile Apps in the store to facilitate auto-billing and store data. It also handles all the security features of the store. E. g. security camera, check-in and check-out details, etc. It is linked to and also maintained by the Online Store. All data from the reception server is shared with the Online Store by the Reception App automatically and both remain in sync.

Reception Server: It stores all data generated by the Reception App. It maintains the Inventory data. It maintains data from security feeds. It stores customer info relating to the store.

Data Analysis System: This system is integrated with the Online Store. It implements various data analysis methods on the data stored in the Online Store Server to produce various statistical data. It gives instrumental information about market trends, product feedback, regional level market developments, store market sale developments, demand of products according to age, economical categories, etc. of the customers. It gives data on individual preferences of a customer to provide personalized shopping experience to customers and to keep them updated about the market according to their personal preferences.
The PAL Feature: The feature allows a customer to shop in a store without ever having to face the billing desk. On activating the PAL Feature, customer can visit a store, shop and leave without any physical billing. To avail this feature, customer must save his/her debit/credit card with the Online Store System. PAL transactions are handled directly by the Online Store System. When the customer ends shopping and submits cart, the bill is generated and deducted from the customer's card on confirming an OTP from the app and invoice is sent to the customer's app and e-mail.

D. Advantages of the Proposed System

- Stores get a global platform for doing business with a larger range of customers.
- Shopping from stores become as easy as online shopping.
- Product selection is easy and can also be verified in person.
- Easy billing and swift purchase of products.
- Secured transactions provide hassle-free experience.
- Use of mobile app ensures portability of system.
- Integration of data from various stores ensures more accurate statistical data.
- Market analysis becomes easier.

E. Future Scope

- The system can be made compatible with the people not using smartphones by providing an alternative to mobile app.
- A packaging unit can be developed for even swifter checkout.
- System can be made compatible for very small scale stores.

4 Conclusion

By means of this paper, we intend to simplify the shopping experience in stores as well as integrate it with the internet for a wider range of business for local stores. Easy billing and tending to the individual preference of the customers through proper data analysis is also an important aim fulfilled by the proposed system. Few of the aspects like compatibility for people without smartphones can be resolved to make the system more robust in the future.

Acknowledgements I express my deep sense of gratitude and indebtedness to my parents, family members and friends and the people who have always guided and supported me. I sincerely acknowledge the encouragement and support given to us by my guides Mr. Sauvik Bal and Mrs. Sagarika Ghosh in completion of my paper. I would also thank our university "University of Engineering and Management", which provided me the opportunity to fulfill our cherished goals. I would most heartily like to thank the Almighty for everything He blessed me with.

References

1. Komal Ambekar, Vinayak Dhole, Supriya Sharma, Tushar Wadekar, "SMART SHOPPING TROLLEY USING RFID," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 10, October 2015.
2. Udita Gangwal, Sanchita Roy, Jyotsna Bapat, "Smart Shopping Cart for Automated Billing Purpose using Wireless Sensor Networks", SENSORCOMM 2013: The Seventh International Conference on Sensor Technologies and Applications.
3. Zeeshan Ali, Reena Sonkusare, "RFID Based Smart Shopping and Billing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013.
4. Galande Jayshree, Rutuja Gholap, Preeti Yadav, "RFID Based Automatic Billing Trolley," International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014).
5. Chihhsiong Shih, Bwo-cheng Liang, Cheng-zu Lin, Nien-Lin Hsueh, Pao-Ann Hsiung "An Automatic Smart Shopping Cart Deployment Framework based on Pattern Design," 2011 IEEE 15th International Symposium on Consumer Electronics.

Load Balancing Routing Mechanisms for Wireless Mesh Networks: A Survey

D.S. Keerthi and T.G. Basavaraju

Abstract Wireless Mesh Networks (WMNs) are a special case of mobile ad hoc networks. With the proliferation in wireless technologies and explosive advancements in internet, WMNs have emerged as an outstanding wireless technology in providing efficient broadband services over a larger geographical area. WMNs are self-healing, self-configured, self-optimized and self-managed, and highly fault tolerant which makes them distinct from other wireless technologies. Routing is an act of discovering a path through which a packet can traverse. Thus an effective routing provides the best and a lossless data path for a successful communication. Routing protocols with metrics are used to evaluate the best path. WMNs have abundant number of routing protocols and metrics to route the traffic from source to destination. However, most of them take into account at most one metric such as interference, packet loss, load, etc. Load balancing in routing is an important network solution that lessens the burden on a node by distributing the traffic among other nodes of the same network. This improves node utilization and reduces packet drop ratio which in turn maximizes the throughput and increases the network capacity. This paper presents a summary of the load balancing routing protocols and metrics, their functionality, features, comparison, and their theoretical details.

Keywords WMNs · Traffic-load · Interference · Metric · Protocol

D.S. Keerthi (✉)
Department of E & C Engineering, MCE, Hassan 573202, India
e-mail: dsk@mcehassan.ac.in; keerthikumards@gmail.com

T.G. Basavaraju
Department of Computer Science & Engineering, Government SKSJTI, K R Circle, Bengaluru 560001, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_61

1 Introduction

Wireless networks have undergone a vibrant revolution in recent days. These wireless networks use radio signal frequency at the physical layer to establish wireless connection among the network nodes. They connect the devices like laptops, computers and cell phones to the internet or any other business network applications. This wireless network technology has replaced the traditional wired networking that mainly relies on the cables to connect network devices. Unlike wired technology, wireless networks provide mobility, more bandwidth and do not use cable wires for connection. They are mainly intended to be used as access networks for an internet or PSTN. However, they are expensive and suffer from security and QoS issues.

With the advancements in wireless networks, Wireless Mesh Networks [1] have dominated the wireless networking in the present decade which have realized the concept of network connectivity anytime and anywhere with much lesser complexity and low cost. They have completely simplified the network deployment and are highly tolerant against network failures and possess the potential for wide broadband connectivity. Even though WMNs face some technical challenges like load balancing, optimal routing, network configuration and mobility management, they provide a cost effective, flexible and reliable solution for extending services in broadband communication.

The rest of the paper is organized as follows. Section 2 gives an overview of wireless mesh network which includes architecture and the concept of routing in wireless mesh networks. Section 3 emphasizes on some of the load balancing routing protocols and routing metrics. Tables 1 and 2 give the comparison of various routing protocols and routing metrics.

2 Wireless Mesh Networks

Wireless Mesh Network architecture [2] consists of mesh routers, mesh nodes, and gateways (Fig. 1). In WMNs each node can operate both as a router and gateway. Mesh routers have additional routing functions to facilitate mesh networking and they are the backbone of WMN. Mesh routers have very rare mobility and operate just like stationary routers. A mesh router is usually equipped with multiple wireless interfaces built on either the same or different wireless access technology to improve the flexibility of mesh networking. It has the capability to achieve good coverage even for a much lower transmission power through multi-hop communication. They route the network traffic to and from the gateways but do not terminate or initiate the traffic. Mesh clients are either mobile or stationary and can form self-organized ad hoc networks. They access services by sending requests to wireless backbone network. In contrary to mesh routers, mesh clients usually have a single wireless interface. Mesh clients are the end user devices like desktop

Table 1 Comparison of routing protocols

Sl. no.	Routing protocol	Classification	Routing metrics	Mobility	Loop free	Load balancing	Scalability/ Reliability	Throughput	End to end delay	Multicasting	Interference
1	DSR	Reactive	Shortest path	YES	YES	NO	NO/YES	Decrease with mobility	NO	NO	NO
2	DSDV	Proactive	Shortest path	YES	YES	NO	NO/YES	Decrease with mobility	NO	NO	NO
3	OLSR	Proactive	Shortest path	YES	YES	NO	NO/YES	Better	NO	NO	NO
4	LQSR	Reactive	HOP count RTT, ETX	YES	YES	NO	NO/YES	Good	NO	YES	NO
5	AODV	Reactive	FAST and Shortest path	YES	YES	NO	NO/YES	Decrease with mobility	NO	YES	NO
6	MR-LQSR	Reactive	WCETT	YES	YES	YES	NO/YES	Good	YES	YES	NO
7	MR-OLSR	Proactive	IWCETT	YES	YES	YES	NO/YES	Good	YES	YES	YES
8	IALBR	Reactive	Shortest path	YES	YES	YES	NO/YES	Good	YES	YES	YES
9	LAR-MR	Reactive	LAR	YES	YES	YES	NO/YES	Good	YES	YES	YES
10	LBGD-AODV	Reactive	LBGD	YES	YES	YES	NO/YES	Good	YES	YES	YES
11	NAV-based AWARE	Reactive	NAV	YES	YES	YES	NO/YES	Good	YES	YES	YES
12	DLB for gateway LB	Reactive	NGW with shortest path	YES	YES	YES	NO/YES	Good	YES	YES	YES
13	MODVWLS	Reactive	Link State + Discriminator (HOP count)	YES	YES	YES	NO/YES	Good	YES	YES	YES
14	AODV-ST	Reactive	ETX or ETT	YES	YES	YES	NO/YES	Good	YES	YES	YES

Table 2 Comparision of routing metrics

Sl. no.	Routing metrics	Inter-flow interference	Intra-flow interference	Load balancing	Multi-channel	Isotonicity	Stability	Multi-radio/Single radio	Layer
1	Hop Count	NO	NO	NO	NO	YES	YES	Single	Network
2	ETX	NO	NO	NO	NO	YES	NO	Single	Network
3	ETT	NO	NO	NO	NO	YES	NO	Single	Network
4	WCETT	NO	YES	NO	YES	NO	NO	Multi	Network
5	MIC	YES	YES	YES	YES	YES	NO	Multi	Network
6	LAETT	NO	NO	YES	YES	YES	NO	Multi	Network
7	WCETT-LB	YES	YES	YES	YES	YES	NO	Multi	Network
8	ILA	YES	YES	YES	YES	YES	NO	Multi	Cross layer
9	LBIARM	YES	YES	YES	YES	YES	NO	Multi	Cross layer
10	LARM	YES	YES	YES	YES	YES	NO	Multi	Network
11	ACA	NO	NO	YES	YES	YES	NO	Multi	Link
12	ETT-LB	NO	NO	YES	YES	YES	NO	Multi	Cross layer
13	IDAR	YES	YES	YES	YES	YES	NO	Multi	Cross layer

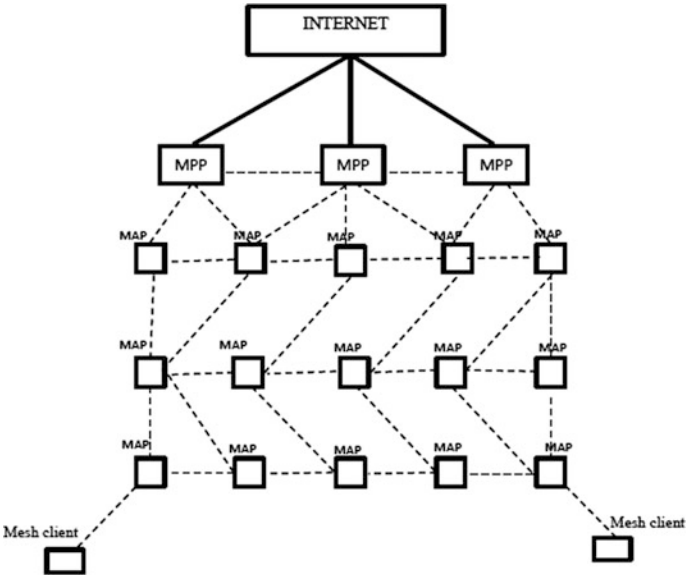


Fig. 1 Wireless mesh architecture

PC/Laptop, IP phone, PDA, Pocket PC, RFID reader. Gateways are mesh routers with wired or wireless interfaces. The gateways are equipped with multiple interfaces to connect both wired and wireless networks, hence they are expensive. The network performance is influenced by the placement of the gateways in that network.

2.1 Routing in WMNs

2.1.1 Routing Protocols

In order to improve the network performance of the WMNs [1], they must be accompanied with a routing protocol, which plays an important role in choosing the best routes, avoiding congested links with high interferences and supporting application parameters (i.e., bandwidth, delay, and packet losses). Since wireless mesh networks are a type of packet switching networks, the routing protocol directs the packets from their source to their final destination through intermediate nodes. Over 70 routing protocols exist for these types of networks [2, 3]. Each protocol adopts a routing strategy. Based on these routing strategies, the mesh routing protocols are characterized into Proactive routing, Reactive routing and Hybrid routing protocols.

2.1.2 Routing Metrics

A routing metric is a cost computed by routing algorithm to select an appropriate path for forwarding the traffic. The metric assigned to each path is available in the routing table and is computed using several techniques based on the selected routing algorithm in use. Several parameters used for computing a routing metrics are Hop count, End to End Delay, Jitter, Path reliability, Path speed, load, Bandwidth, Latency, Maximum transmission unit, Channel Diversity, Interference, and so on. Some of the basic routing metrics in wireless mesh networks are: Hop Count, Expected Transmission Count (ETX), Expected Transmission Time (ETT), Weighted Cumulative Expected Transmission Time (WCETT), Metric of interference, and channel switching (MIC).

3 Load Balancing in Wireless Mesh Networks

The distribution of workload across the network for adequate usage of network resources is called load balancing. Load balancing eliminates bandwidth bottlenecks and minimizes inefficiencies by directing traffic over the least expensive link or across multiple connections. Load balancing aims to maximize the throughput, optimize resource use, minimize delay constraints, and avoid overload on a single node.

Load balancing can be performed in two ways:

(i) Path-based load balancing.
(ii) Gateway-based load balancing.

In path-based load balancing, network performance is enhanced by sharing traffic among different paths toward the gateway. In gateway-based load balancing, traffic is shared among available gateway that is traffic is rerouted from congested gateway toward under-utilized gateways to eliminate the bottleneck problem. Sections 3.1 and 3.2 discusses in detail about load balancing schemes for WMNs.

3.1 Load Balancing Routing Protocols

(i) *Multi Radio-Link Quality Source Routing (MR-LQSR)*
MR-LQSR [4] is a source routed link state protocol that operates over multiple channel and multiple interfaces using WCETT routing metric. Basically, LQSR is derived from dynamic source routing (DSR) with route discovery, route maintenance and a mechanism for metric maintenance. The link quality is calculated by assigning equal weights to all possible links that is channel assignment; bandwidth and loss are broadcast to all nodes in the network to find the best path to the destination. The advantage of this

protocol is that, it achieves tradeoff between delay and throughput because it considers channels with good quality and effectively balances the load as it uses multiple channels for routing the packets. Another advantage of this protocol is that it provides loop free service because it uses source routing.

(ii) *Multi-Radio-OLSR Protocol (MR-OLSR)*

MR-OLSR [5] is an extension of LQSR protocol provides enhanced throughput and supports load balancing in multi-radio WMN environment. It improves channel throughput by forwarding the data traffic through different paths to prevent congestion. The path quality is calculated based on the Improved Weighted Cumulative Estimated Transfer Time (IWCETT) metric. The IWCETT is an improvement of WCETT metric that eliminates interflow interference. It uses IFQ (Interference queue) indicator, i.e., length of a data queue on the selected channel for a node.

IWCETT is defined as

$$\mathrm{IWCETT} = (1-\alpha)\sum\nolimits_{x=1}^{n} \mathrm{LL}_x + \alpha \max_{1 \le x \le k} Z_j \tag{1}$$

where,

LL_x	= Link Load
A	= Vulnerable value
$Z_j = \sum\nolimits_{\text{Hop is on channel } j} \mathrm{LL}_x \quad 1 \le j \le k$	= Cumulative transmission time for multi-hop channel j

$$\mathrm{LL}_x = \mathrm{LL}^j_{P-Q} = (1-\mu)\mathrm{ETT}^J_{P-Q} + \mu\,\mathrm{CI}^j_{x-y} \tag{2}$$

μ	= Tradeoff between intra-flow and inter-flow interference
CI^j_{P-Q}	= Congestion Indicator between nodes P and Q

$$\mathrm{CI}^j_{P-Q} = \frac{\mathrm{IFQ}^j_p}{\mathrm{Q}}$$

where, If Q^j_p = data queue at the node P on channel j.

The load balancing in MR-OLSR can be achieved by channel allocation/assignment and path schedule/selection mechanism.

(iii) *Interference-Aware Load Balanced Routing (IALBR) Protocol*

IALBR Protocol [6] is an improvement of AODV protocol with an objective to provide Load balancing by capturing interference among the neighborhood nodes with consideration that the probability of channel busy is the load of a node. The route selection criterion of this protocol is based on the traffic load of the intermediate nodes. This protocol is implemented by a cross-layer design, i.e., between Route layer and MAC layer. The route table in IALBR has same difference in route entry with that of AODV routing protocol. IALBR stores load values in both route table and in the RREQ packets to facilitate traffic load calculation by the destination node. The destination node will select the least traffic load for a congestion free transmission of packets.

(iv) *Load-Aware Routing (LAR) Protocol for multi-radio WMNs*
LAR protocol [7] uses load balancing mechanism and load aware channel selection strategy for route discovery process, which helps to find the best path with minimum traffic load from source to destination and reduce Inter/Intra-flow interference in multi-radio environment.
Load calculation: let $G = (V, E)$ for multi-radio network. The regional traffic load at node 'a' on the channel 'c' is given by

$$T^{\mathrm{C}}(\mathrm{a}) = \sum_{\mathrm{K} \in \mathrm{N}^{\mathrm{C}}} Q_K^C \tag{3}$$

$\mathrm{N^c}$ = Group of neighbor nodes of 'a'
$\mathrm{Q_K^C}$ = Average queue length at channel 'c' of node K

The load of the link is given by

$$\mathrm{LL}(l^{\mathrm{c}}) = \mathrm{T^c(a)} + \mathrm{T^c(b)} \tag{4}$$

If $P = \{l_1^{c_1} \ldots\ldots\ldots\ldots\ldots l_2^{c_2}\}$ be a path between source and destination. Then the total load of the path is the cumulative load of the links on the path

$$\mathrm{TL}(p) = \sum_{n=1}^{2} \mathrm{LL}\left(l_n^{c_n}\right) \tag{5}$$

In channel assignment mechanism, the receiver identifies the recommended channel for the reverse path between source and destination.

(v) *Load-Balancing Gateway Discovery Routing Protocol (LBGD-AODV)*
LBGD-AODV [8] is an improvement of AODV routing protocol with the objective to provide load balancing among mesh routers and gateways. It supports multiple internet gateways where normal AODV supports only one gateway. This protocol performs load balancing and route discovery in a single process. The existence of new gateway in the network is advertised by the protocol by broadcasting hello control message "gateway_hello_interval" to the network for every interval. This makes mesh nodes and mesh routers to gather information regarding complete network topology such that, effective load balancing can be achieved between mesh router and mesh gateways. This improves the network throughout.

(vi) *Network Allocation Vector (NAV)-based Aware Routing Protocol*
NAV-based Aware Routing [9] protocol is an improvement of Directional-AODV (D-AODV) protocol. This protocol uses Network Allocation Vector (NAV) routing metric which is a virtual carrier sensing mechanism to find whether the medium is busy or idle, that is it finds the minimum congested route to perform efficient load balancing. The route discovery is same as that of D-AODV but it uses NAV indicator in RREQ packet to find the congested path. In this mechanism, each node calculates its own NAV_SUM for a give interval 'τ' during which the node is idle or free. Busy portion of the node is calculated by $\overline{\mathrm{NAV\_SUM}_i}$ i.e.,

$$\overline{\mathrm{NAV\_SUM}_j} = \beta\,\overline{\mathrm{NAV\_SUM}_{i-1}} + (1-\alpha)\mathrm{NAV\_SUM}_i \tag{6}$$

where,

α = weight factor varies between 0 and 1
i = ith update period

The path session between node/router to the gateway is as follows. A source node broadcasts RREQ packet with $\overline{\mathrm{NAV\_SUM}_j}$ to the network, the intermediate node compares the $\overline{\mathrm{NAV\_SUM}_j}$ of the received RREQ packet with its own $\overline{\mathrm{NAV\_SUM}_j}$ value. If it is smaller than its own value $\overline{\mathrm{NAV\_SUM}_j}$ then it updates its $\overline{\mathrm{NAV\_SUM}_j}$ and rebroadcast to its neighbor's. Once the gateway receives RREQ packet, it turns on the timer for an interval to receive some more RREQ packets with the least value of $\overline{\mathrm{NAV\_SUM}_j}$. If any RREQ packet having value of $\overline{\mathrm{NAV\_SUM}_j}$ that is less than previous RREQ packet, then gateway will select the new route otherwise uses the route specified in previous RREQ packet.

(vii) *Distributed load balancing protocol (DLB) for gateway load balancing*
The objective of DLB protocol [10] is to forward or distribute the data traffic between gateways, i.e., reroute data traffic from highly or heavily loaded gateway to the underutilized gateway. Another major difference from the existing protocol is that it also considers contention and interference which in turn increases the network capacity and throughput. For the change in network topology this protocol converges quickly since gateways are connected through wired medium and its execution time is low. The drawback of the protocol is that it follows the shortest path routing or Nearer Gateway (NGW) approaches when the number of sinks increases. If all the domains are highly congested or least congested, it follows the shortest path (hop count) or NGW approach for gateway selection.

(viii) *Mesh on Demand Distance Vector based on Weighted Link State Protocol (MODVWLS)*
MODVWLS protocol [11, 12] is an improvement of AODV protocol. This protocol uses a combination of link state routing and distance vector routing

for route discovery. It considers the combination of three performance metrics namely throughput, queue size of the buffer and bandwidth as weighted link state information. In addition to link state information, it also considers the hop count between source and destination when multi-paths exist in the network. It selects the optimum path having minimum weight and hops. Thus it avoids congested path or node and improves the load balancing effectively. For route discovery it uses RREQ message.

3.2 Load Balancing Routing Metrics

(i) *Load Aware Expected Transmission Time (LAETT)*
LAETT [13] is an extension of ETT metric with the objective of facilitating the bandwidth requirement for the flow and to facilitate the buffer for the upcoming request in order to distribute the load across the network

$$\mathrm{ETT}_{xy} = \mathrm{ETX}_{xy} \times \frac{s}{B_{xy}} \tag{7}$$

ETT_{xy} = Expected transmission count on link (x,y)
S = Length of the packet
B_{xy} = Bit rate/transmission rate

$$B_{xy} = \frac{B_x}{\lambda_{xy}}$$

B_x = Transmission rate of node x
λ_{xy} = Link quality factor
λ_{xy} = 1 for a link of good quality

For Load balancing consideration, the remaining capacity (RC_x) of each node is

$$\mathrm{RC}_x = B_x - \sum_{K=1}^{N} f_{xk} \times \lambda_{xy} \tag{8}$$

f_{xk} = Transmission rate of the present/current flow N_x that traverse node 'x'.
The cost of a flow on the rest of the BW is calculated by factor 'xk'
An optimum transmission consumes minimum resources.
LAETT_{xy} is defined by

$$\mathrm{LAETT}_{xy} = \mathrm{ETT}_{xy} \times \frac{S}{\frac{RC_x + RC_k}{2\lambda_{yk}}} \tag{9}$$

The above equation consists of two parts: the second part captures the available bandwidth at either node. If two paths have equal ETX weight then LAETT metric considers the path having available bandwidth.

Advantage of LAETT metric is that it uses load aware isotonic routing considering link quality/available bandwidth on each node. The disadvantage is, it does not take an account of interference.

(ii) *Weighted Cumulative Expected Transmission Time Load Balancing (WCETT-LB)*

WCETT-LB [14] is an extension of WCETT with the mesh routers by using congestion aware routing and traffic splitting mechanism. The congestion aware routing works as follows:

The traffic concentration at each node is examined by the average queue length in a particular path with the threshold value considered at each node. If it is higher than the threshold value then the path is considered to be a congested one. The traffic load on all the paths can be evaluated by calculating actual time required for transmission which is the ratio of average queue length over the transmission rate (b_i).

Let N_i be the traffic concentration at ith node chosen as next hop to transmit/forward packets by several nodes, then congestion at that node increases. This can be represented as follows

$$\sum_{\text{node } i \in p} \min(\mathrm{ETT}) N_i \tag{10}$$

where, min (ETT) = Smaller ETT in the network. If it is higher, then there is more congestion at the node.

The WCETT-LB can be formulated as follows

$$\mathrm{WCETT} - \mathrm{LB}(p) = \mathrm{WCETT}(p) + L(p) \tag{11}$$

where, $\mathrm{L}\,(\mathrm{p}) = \sum_{\text{node} i \in p} \frac{QL_i}{b_i} + \min(\mathrm{ETT}) N_{\mathrm{i}}$

The traffic splitting works as follows:

Let δ be the threshold value for path switching. When an updated WCETT-LB is received by a router from its neighbor and it is found congested, then router distributes/balances the load by finding the best path. This is based on following criteria/notation.

i.e., if $(\text{WCETT-LB}_{\text{current}} - \text{WCETT-LB}_{\text{best}}) \geq \delta$

Then the router will move to the best path otherwise remains on the current path.

(iii) *Interference Load Aware Routing Metric (ILA)*
ILA [15] metric aims at finding paths with reduced level of interference, higher data rate, minimum packet loss ratio and less congestion for routing the traffic and load balancing, thereby delivering high throughput.
The ILA routing metric is a combination of two components: Metric of traffic Interference and channel switching cost (CSC).

(a) Metric of Traffic interference (MTI)
This metric considers interflow interference, i.e., the interference that occurs when the two neighboring nodes on the same channel compete with each other for the channel bandwidth.
The level of interference is dependent on the load generated by the interfacing nodes.
The MTI metric can be defined as

$$\mathrm{MTI}_x(\mathrm{C}) = \begin{cases} \mathrm{ETT}_{xy}(\mathrm{C}) \times \mathrm{AIL}_{xy}(\mathrm{C}) & ; Nz(\mathrm{C}) \neq 0 \\ \mathrm{ETT}_{xy}(\mathrm{C}) & ; N_l(\mathrm{C}) = 0 \end{cases} \tag{12}$$

where, AIL_{xy} is the average interfering load by the neighboring nodes when the transmission is between nodes x and y over channel C and is given by

$$\mathrm{AIL}_{xy}(\mathrm{C}) = \frac{\sum_{Nl} \mathrm{IL}_{xy}(\mathrm{C})}{N_z(\mathrm{C})} \tag{13}$$

where, $N_z(\mathrm{C}) = N_x(\mathrm{C})\mathrm{U}N_y(\mathrm{C})$.
$\mathrm{IL}_{xy}(\mathrm{C})$ is the interfering load by the neighbor and $N_z(\mathrm{C})$ is the group of interfering neighbors of node x and node y.
ETT_{xy} denotes the packet drop ratio and the difference in transmission rate of links.

(b) Channel Switching Cost (CSC)
The CSC metric captures intra-flow interference which occurs between the nodes in the same flow. The CSC metric is defined as

$$\mathrm{CSC}_x = \begin{cases} W_1\, \mathrm{CH}\,(\mathrm{prev}(x)) \neq \mathrm{CH}\,(x) & 0 \leq W_1 \leq W_2 \\ W_2\, \mathrm{CH}\,(\mathrm{prev}(x)) = \mathrm{CH}\,(x) & \end{cases} \tag{14}$$

where,

x	= node equipped with multiple radios
$\mathrm{CH}(x)$	= channel that node x uses to transmit to its next hop
CH (prev (x))	= channel used by the previous hop of node x

W_1 and W_2 are the weights such that $0 \leq W_1 \leq W_2$
Therefore ILA can be given as

$$\mathrm{ILA}(p) = \alpha \times \sum_{\text{link}\, x \in p} \mathrm{MTI}_x + \sum_{\text{node}\, x \in p} \mathrm{CSC}_x \tag{15}$$

where, ‘p’ is the path and α is the scaling factor.
Thus CSC increases the throughput and MTI decreases the packet delay due to the neighboring nodes.

(iv) *Load Aware Routing Metric (LARM)*
The objective of LARM metric is to perform load balancing and is incorporated in LBM protocol. LARM [16] considers the difference in transmission rate, packet loss ratio, traffic load, and both inter-flow/intra-flow interference in multi-radio mesh environment to find an optimal path.
Considering a multi-radio network, let G = (V, E) be the multi graph where V represents graphs of mesh routers and E represents multi groups of edges.
The packet delay is caused by the interference and traffic load at the current node.
Consider a channel ‘C’ and two nodes *x* and *y* competing for the channel C to route the traffic.

$$\mathrm{Q}\ (\mathrm{l}^{\mathrm{c}}) = \sum_{k \in Nc(x)UNc(y)} \mathrm{Q}_k^i \tag{16}$$

where, l^c is the link between the neighboring nodes x and y interfering on the channel C, $N^c(x)$ and $N^c(y)$ are the set of nodes x and y.
The link load l^c is given by

$$\mathrm{LL}(\mathrm{l}^{\mathrm{c}}) = \mathrm{ETT}(\mathrm{l}^{\mathrm{c}}) \times \mathrm{Q}(\mathrm{l}^{\mathrm{c}}) \tag{17}$$

where, ETT (l^c) expected transmission time of the time on channel ‘C’.
LARM is given by

$$\mathrm{LARM} = (1 - \beta) \times \sum\nolimits_{r=1}^{m} CL(r) + \beta \times \max\{\mathrm{CL(r)}\} \quad 1 \leq \mathrm{r} \leq \mathrm{m} \tag{18}$$

where, β is a tunable parameter that has range between $0 \leq \beta \leq 1$ and ‘m’ is the total number of channels available at the routing path.
CL(r) is the channel load of channel ‘r’ on the path.
P = $\{l_1^{c1}, l_2^{c2} \ldots l_n^{cn}\}$ between source and destination.

(v) *Airtime Congestion aware (ACA) Metric*
ACA Metric [17] is a combination of Airtime link cost metric and Round Trip Time (RTT) that operates in multi-radio WMNs with the objective to provide congestion free and quality path. This metric balances the load

uniformly at routers. The channel utilization factor between nodes is calculated by airtime link cost metric which helps to identify link quality.
The airtime link cost at particular link l is given by

$$\mathrm{AL_1} = \left[\mathrm{x_{alc}} + \mathrm{y_p} + \frac{\mathrm{B_t}}{r}\right] \times \frac{1}{1 - e_f} \tag{19}$$

where $\mathrm{x_{alc}}$, $\mathrm{y_p}$ and $\mathrm{B_t}$ are constants

R = packet/data rate
e_f = frame error rate
$\mathrm{B_t}$ = frame size

The objective of RTT is to bypass highly congested link.
The ACA metric for path ‘P’ is defined as

$$\mathrm{ACA(P)} = (1 - \beta) \sum_{\text{link } l \in p} \mathrm{AL}_c + \beta \sum_{\text{link } l \in p} \mathrm{AL}_c\, \mathrm{RTT}_l \tag{20}$$

where, β = tunable parameter = 0.3
It also considers probability of frame loss along link. Higher ACA value degrades the network performance.

(vi) *Expected Transmission Time—Load Balancing (ETT-LB)*
ETT-LB [18] is the extension of ETT metric with the objective of providing Load balancing. In addition to ETT, this metric also considers the utilization of the link of each node on the path. The average link traffic at a node defines the expected delay time (EDT).

$$\text{Time required for continuous load traffic transmission} = \frac{\text{Fraction of link traffic}}{\text{Transmission rate}}$$

Figure 2 shows a continuous data being transmitted from S to D through an intermediate node F, a node ‘N’ sends an additional data to node D through

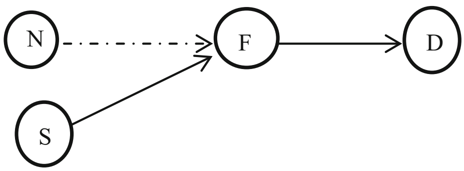


Fig. 2 Load centralization at node F

F. Thus the load is centralized at intermediate Node F. This can be decentralized by the ETT-LB metric as follows

$$\text{ETT-LB}(\alpha) = (1-\alpha)\sum_{\text{link}\, l\in P}\text{ETT}_l + \alpha\sum_{\text{link}\, l\in P}\text{EDT}_l \tag{21}$$

$$\text{EDT}_l = \frac{\text{Data } l}{\text{Bandwidth}}$$

where, α is a tunable factor varies between $0 \leq \alpha \leq 1$.

(vii) *Estimated Transmission Delay (ETD)*
ETD [19] considers the packet loss ratio and transmission delay along a path, which is given by

$$\text{ETD(p)} = \frac{\text{TD}(p)}{1-\text{PL}(p)} \tag{22}$$

where, TD (p) = Packet transmission delay for path ‘p’
PL (p) = Packet loss ratio a path P and is given by

$$\text{PL(p)} = 1 - \pi_{\text{link}\, l\in p}(1-\text{PF}_l)\times(1-\text{PR}_l) \tag{23}$$

where, PF_l and PR_l = Probability of the packet loss of link l in both forward and reverse directions, respectively.
Thus IDA is defined by

$$\text{IDA(p)} = \alpha\times\text{ETD(p)} + \beta\times\text{CLI(p)} + \text{CSLC(p)} \tag{24}$$

where, α and β = tunable factors.

4 Conclusion

This paper presents a detailed study on Wireless Mesh Networks (WMNs), classification of architecture, features and its applications. Depending on the network capacity and throughput, different architecture is adopted. Wireless mesh networks find extensive applications as they are less expensive, simple to deploy, minimum installation time, and they can be easily integrated with other wireless technologies. Routing is the key feature in Wireless Mesh Networks. An efficient routing can be achieved with the help of Routing Protocols and Routing Metrics. This paper also discussed some of the existing routing protocols and routing metrics and has presented their theoretical and mathematical details. Load balancing gives solution to distribute the incoming traffic load among different nodes. Load balancing is accomplished either through gateways or routing paths. This paper gives a review on some of the load balancing routing protocols and metrics. The paper also shows a comparison table of routing protocols and routing metrics. Thus an efficient load balancing and minimized interference can be achieved by selecting an appropriate routing protocol and routing metric depending upon the network conditions and requirements.

Acknowledgements This work is supported by the Technical Education Quality Improvement Program Phase II with the World Bank assistance. The authors would like to thank State Project Facilitation Unit, Directorate of Technical Education, Government of Karnataka, National Project Implementation Unit, MHRD, New Delhi for their support and encouragement.

References

1. F. Akyildiz, X. Wang, and W. Wang “Wireless mesh networks: a survey”, in Computer Networks Journal, vol. 47, 2005, pp. 445–487.
2. Eslami, M, Karimi, O, Khodadadi T “A survey on wireless mesh networks: Architecture, specifications and challenges” Control and System Graduate Research Colloquium (ICSGRC), 2014, pp. 219–222.
3. M. L. Sichitiu “Wireless mesh networks: Opportunities and challenge,” in Proc. of the Wireless World Congress, (Palo Alto, Ca), May 2005.
4. Richard Draves, Jitendra Padhye and Brian Zill “Routing in Multi-radio, Multi-hop Wireless Mesh Networks” proceedings of the 10th annual international conference on mobile computing and networking, pp. 114–128, New York, USA, 2004.
5. Guangwu Hu, Chaoqin Zhang “MR-OLSR:A link state routing algorithm in multi-radio/multi-channel Wireless Mesh Networks”, 18th Asia-Pacific Conference (APCC), 2012, pp. 883–888.
6. Jing Feng, Ran Xia, Huaibei zhou “Interference-Load Balanced Rouring in Wireless Mesh Networks” Wireless Communications, Networking and Mobile Computing (WiCom), 2007, pp. 1730–1734.
7. Anh-Ngoc Le, Dong-Won Kum, You-Ze Cho “Load-aware routing protocol for multi-radio Wireless Mesh Networks” International Conference on Communications and Electronics (ICCE), 2008, pp. 138–143.
8. Iqbal M, Xinheng Wang, Wertheim, D Xu Zhou “Load balanced multiple gateway support in wireless mesh networks for broadband services” Wireless and Optical Communication Conference, 2009, pp. 1–5.
9. Dong-Won Kum, Won-Kyeong Seo, Jae-In Choi, You-Ze Cho “NAV-based Load Aware Routing protocol in Wireless Mesh Networks”, Information and Communication Technology Convergence(ICTC), 2010, pp. 218–219.
10. Gálvez, J.J, Ruiz, P.M., Skarmeta A.F.G. “A distributed algorithm for gateway load-balancing in Wireless Mesh Networks” Wireless Days(WD), 2008, pp. 1–5.
11. Xian Xiao-dong, Fu Yun-qing, Wang Song-jian, Liang Shan “A routing protocol based on combinative metric for wireless mesh network” International Conference on Networking, Sensing and Control(ICNSC), 2009, pp. 803–806.
12. Bari S.M.S, Anwar F and Masud M.H. “Performance study of hybrid Wireless Mesh Protocol (HWMP) for IEEE 802.11 s WLAN mesh networks” International Conference on Computer and Communication Engineering (ICCCE), 2012, pp. 712–716.
13. H. Aiache; V. Conan; L. Lebrun; S. Rousseau “A load dependent metric for balancing Internet traffic in Wireless Mesh Networks” 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems(MASS), 2008, pp. 629–634.

14. Ma L, Denko M.K. “A Routing Metric for Load-Balancing in Wireless Mesh Networks” 21st International Conference on Advanced Information Networking and Applications Workshops, (AINAW), 2007, Volume: 2 pp. 409–414.
15. Shila D M, Anjali T “Load aware Traffic Engineering for Mesh Networks” Proceedings of 16th International Conference on Computer Communications and Networks (ICCCN), 2007, pp. 1040–1045.
16. Anh-Ngoc Le, Dong-Won Kum, You-Ze Cho and In-Soo Lee “LARM:A Load-Aware Routing Metric for multi-radio Wireless Mesh Networks”, International Conference on Advanced Technologies for Communication (ATC), 2008, pp. 166–169.
17. Kapadia K.N, Ambawade D.D. “Congestion aware load balancing for multi-radio Wireless Mesh Network” International Conference on Communication, Information & Computing Technology (ICCICT), 2015, pp. 1–6.
18. Sooyeol Yang, Jeongbae Yun, Jeongbae Yun, Honggil Lee and Kijun Han “Analysis and Proposal of Wireless Mesh Network’s Routing Metric for WBAN”, First International Conference on Networks and Communications (NETCOM), 2009, pp. 400–403.
19. Qiming Tian “A New Interference-Delay Aware Routing Metric for Multi-Interface Wireless Mesh Networks” 6th International Conference on Wireless Communication Networking and Mobile Computing (WiCOM), 2010, pp. 1–5.
20. Sooyeol Yang, Youngmi Baek, Junhyung Kim, Keuchul Cho and Kijun Han “A routing metric for load balance in wireless mesh networks”, 11th International Conference on Advanced Communication Technology (ICACT) Volume: 03, 2009, pp. 1560–1565.
21. Ekram Hossain, Kin K. Leung “Wireless Mesh Networks: Architectures and Protocols”, Springer, 2008.
22. Peipei Zhao,Jiansheng Qian “The Study of Wireless Mesh Network Routing Algorithm Based on Interference Model”, Fourth International Conference on Genetic and Evolutionary Computing(ICGEC), 2010, pp. 98–101.
23. Benyamina, D, Hafid, A. and Gendreau, M. “Wireless Mesh Networks Design-A Survey” Communications Surveys &Tutorials, IEEE, Volume: 14, Issue: 2, 2012, pp. 299–310.

A Novel Encryption Scheme for Small Enterprises

Mohit Wadhwani, Ankita Thadani and S.P. Medhane

Abstract A cryptographic algorithm is very important in securing the confidential data while transmitting over the network. Modern cryptography is suggesting a variety of encryption schemes for protecting & securing the data. This paper focuses on developing a new method to generate the key & algorithm by using cryptographic techniques which will all together make an encryption scheme which is secure for generating the key. The proposed scheme does not require any specialized hardware or software, so basically it is low-cost & flexible for peer to peer networks & entity communication, which is based on the matrix. Finally, the paper concludes that simulation results are important for the feasibility of the algorithm.

Keywords Secure communication · Sheltered message passing communication
Low budget algorithm

1 Introduction

Security of the data to be transferred is the crucial between the sender and receiver while communication. Information security is all times have set a fundamental part in each area associated with the communication of any confidential data. As the time is changing there is continuous requirement for a robust and powerful complex encryption technique. Presently, the superfluity for enciphering any plain text information subsists as well as encompasses accomplishment of the information protection to an immense level.

M. Wadhwani (✉) · S.P. Medhane
Department of Information Technology, Bharati Vidyapeeth University
College of Engineering, Pune, India
e-mail: mohit7024@yahoo.in

A. Thadani
Computer Science Department, GTU, Ahmedabad, India
e-mail: ankitathadani13@gmail.com

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_62

To maintain the equilibrium among the flexibility and robustness has always been a significant constraint of sustaining the altitude of security & encryption. Securing the data [1] is the need for encryption. Generally the algorithm used to encrypt the data deals and aims at preserving the privacy of the data concealment & discretion of the data. Still while choosing any enciphering technique to be applied to the data the factors needs to be considered are how it performs, robustness, rapidity, dimension protection, and complexity are the important aspects which need to be considered. For enciphering any data there is a need for the key which can be generally classified as the "private key (symmetric)" [2] and "public key (asymmetric)" [3]. symmetric as the name suggest is the algorithm where both the parties exchanging the information shares the same key where as asymmetric key there are different keys by the two parties. If any intruder any how can capture the information being send, this could lead to catastrophic bang for the organization's chaos.

Consequently, proficient scheme [4] which is emplaning the characteristic of marmalade the "Confidentiality, Integrity and Authenticity" [5] are of more importance.

2 Background

Below, Table 1 consists of the various protocols used for key generation with the advantages and the disadvantage of each and also for the network it is suitable for.

3 Proof of Work

The new idea incorporates of the three steps which is form plain text key 1 is generated key 2 is generated from key 1 below is the mathematical proof of the system depicting the working of algorithm.

3.1 Encryption

1. *Key 2 creation From Key 1:*
   (a) select K1 = key 1
   convert K1 to B(K1) which is binary
   (b) matrix A = binary counterpart by writing alongside the rows.
   (c) Transpose A to A^T
   (d) $B(A^T)$ = binary of A^T
   The resultant binary number to hexadecimal.

Table 1 Comparison of various key generating schemes

	System/protocol	Pros	Cons	Suitable for
3.1	Chaotic cryptography using external key [1]	Precise discussion of trust	Development of the formalization for trustis not yet complete changing identities	Situational trust
3.2	Random key material distribution [2]	Nonce for session id	Deficient in of authentication	Peer to peer network
3.3	Chebyshev polynomials [4]	Secure and robust	Complex and inefficient explanation	Multimedia network
3.4	Jacobian elliptic Chebyshev rational maps [5]	Re-keying messages is irrelevant to current group size	For large groups only	Large group network
3.5	Spanning treeKey management (STKM) [7]	Low complexity communication	Complex No collusion resistance	Wireless network
3.6	Exclusion basis system (EBS) [8]	"binary tree" stores the keys	Storage overhead Some of the users know the k number of keys	Wireless network
3.7	An efficient certificate less encryption for secure data sharing [9]	Encrypted data uploaded on cloud sensitive data	Communication overhead	Cloud
3.8	Non-invertible matrices on Hill cipher [10]	No inclusion of the third party Robust	Mathematical complexity linear nature	Peer to peer network
3.9	Invertible, involuntary and permutation matrix generation Hill cipher [11]	Overcome the weakness of the Hill cipher	Rank of the matrix difficult to derive	Peer to peer network
3.10	Advanced substitution technique and symmetric key generating algorithm [12]	Adapted to modern processors	Symmetric and parallel structure	Suited to Smart cards

$$\mathrm{HEX}\left(\mathrm{B}\left(\mathrm{A}^{\mathrm{T}}\right)\right) = \mathrm{key}\,2$$

Key 2 is in the hexadecimal form

(e) K1 to decimal number HEX(add decimal number)
(f) Result = HEX(PT)−HEX(add decimal number)
(g) S(txt) = Swap result digit among themselves
(h) binary(S(txt)).
(i) output = Inverse(2's complement(binaryS(txt)))

2. *Generation of key 3:*
   (a) hex(output)
   (b) key 3 = first and the last part hex(output) $\times$ 2
3. *Cipher Text Formulation*
   (a) split (hex code) as follows.
   (b) output 1 = The first block is Private Key shared with the receiver.
   (c) output 2 = The second block is key with cipher text

3.2 Decryption

We get Cipher text as input.

Step 1 Partial cipher text extraction

(a) Output 2 received = The block is Transport Cipher Key.
(b) Output 1 received = The block is Private key shared.
(c) Join (output 1, output 2) convert to hex which is Key 3.

Step 2 Partial extraction of key 3

(d) We remove transport cipher key and append former 2 and the final 2 hex codes of cipher text.
(e) We divide this new value from Key 3 to get Key 2.
(f) Convert Key 2 into its binary equivalent.
(g) We do the Inversion of this binary equivalent.
(h) We do 2's complement of above Inversion.

Step 3 Plaintext Extraction

(i) Convert the binary equivalent code into Hex code.
(j) We swap the digits of above Hex code.
(k) We add Hexadecimal equivalent to above Hex code.
(l) Convert above Hex code into original Plaintext (Fig. 1).

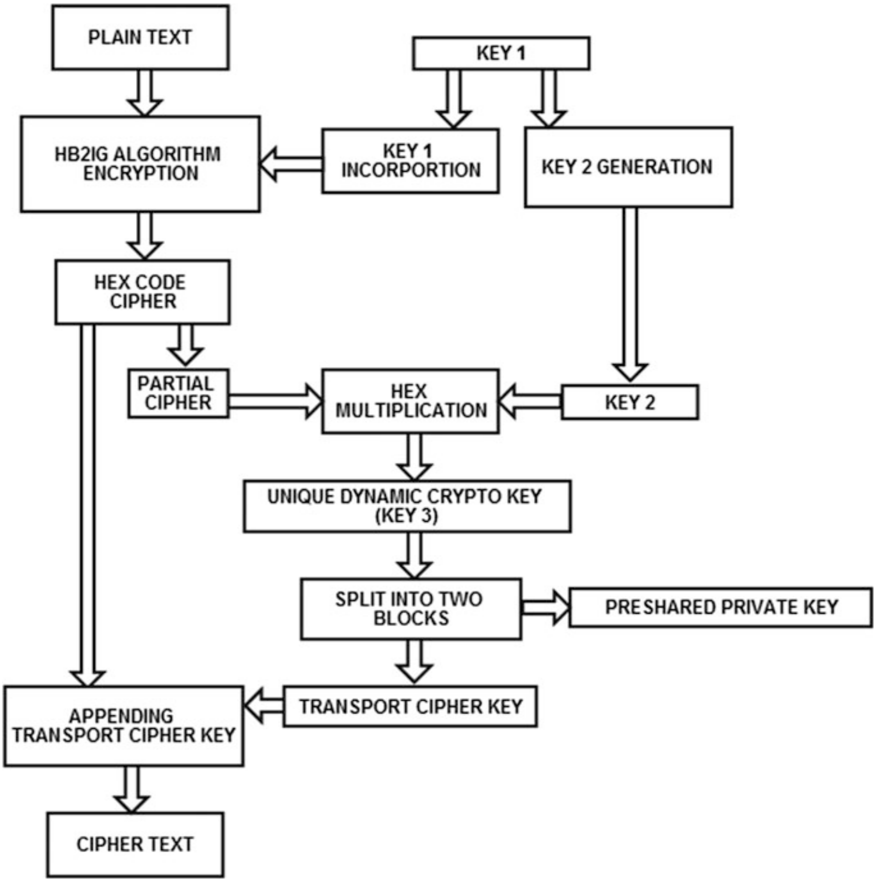


Fig. 1 Working of the encryption system

3.3 *Mathematical Proof of the Proposed Work*

Plaintext: H E L L O

Key 1 = "SK"

Hexadecimal equivalent: 53 4B

Binary (SK) = "0101 0011 010 01 011"

$$\text{Matrix } A = \begin{matrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{matrix}$$

$$\text{Transpose } A^T = \begin{matrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{matrix}$$

Row form

0001 1010 0101 1101

Hexadecimal form

Key 2 1 A 5 D

Step 2: After Key generation

Plaintext:	H	E	L	L	O
Hexadecimal Form:	48	45	4C	4C	4F

Key 1 S K

Decimal Equivalent: 31223

$$3 + 1 + 2 + 2 + 3 = 11$$
Hexadecimal of 11 = 0B

Hexadecimal code after substracting 11

48	45	4C	4C	4F
- 11	- 11	- 11	-11	- 11
37	34	41	41	44

Hex code after swapping digits of each code:

73 43 14 14 44

Binary equivalent of each hexadecimal value

01110011 01000011 00010100 00010100 01000100

2’s complement

1's complement:	10001100	10111100	11101011	11101011	10111011
2's complement:	10011101	11001101	11111100	11111100	11001100

Inversion

01100011 00110010 00000011 00000011 00110011

Hex Code: 62 32 03 03 33

Step 3: Key3 generation

$$\text{Key}\,2 = 1\text{A}5\text{D}$$

The first and last part of Hex code sequence are

62	32	03	33
62			33

Key 2	1	A	5	D
Decimal	1	10	5	13
Ascii	62	32	03	33

Key 3 63 42 08 46

Pre-shared private key: 63 42
Transport cipher key: 08 46

Cipher Text: 08 46 62 32 03 03 33

Decryption
To convert the cipher text into original plain text,

3.4 Decryption

We get cipher text as input.

private key shared: 63 42

The received cipher text:

Cipher Text: 08 46 62 32 03 03 33

1. *Extracting key 3 from cipher text:*

$$\text{Key:}\quad 08 \quad 46$$

63 and 42 are previously shared the first two cipher text join them

$$\text{key}\,3 = \quad 63 \quad 42 \quad 08 \quad 46$$

2. *Extracting Key2:*
Cipher text after removing first 2 values:

$$62 \quad 32 \quad 03 \quad 33$$

key 3 ÷ Cipher text after removing first 2 values = key 2
(63 42 08 46) ÷ (62 32 03 33) = 1 10 05 13
Therefore, Key 2 = 1 A 5 D

3. *Key 1 Extraction:*

$$\text{Converting key 2 into binary value} = \quad 0001 \quad 1010 \quad 0101 \quad 1101$$

Writing it in the matrix form

$$\text{Matrix}\quad B = \begin{matrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{matrix}$$

$$\text{Row form:}\quad 0001 \quad 1010 \quad 0101 \quad 1101$$

$$\text{Transpose Matrix}\quad T = \begin{matrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{matrix}$$

$$\text{Binary equivalent} = \quad 0101 \quad 0011 \quad 0100 \quad 1011$$

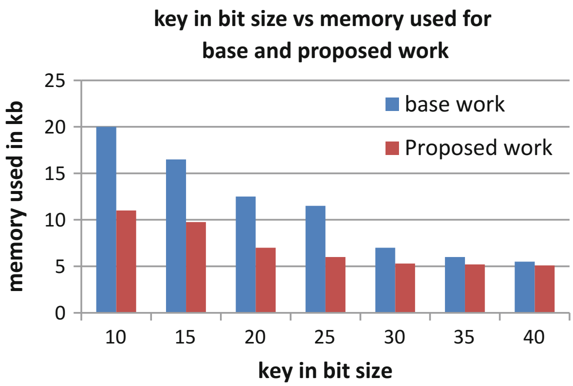
$$\text{Hexadecimal equivalent:}\quad 53 \quad 4B$$

$$\text{ASCII of above binary number} = \text{SK}$$

$$\text{Key 1} = \text{SK}$$

Decimal (SK) = 21 3 23(3 is for space)
Sum of digits of above equivalent = 11
Hex equivalent of above sum = 0b
Adding 0b with each of the hex code in the sequence:
54 48 49 53 20 49 53 20 43 4f 4e 46 49 44 45 4e 54 49 41 4c

Fig. 2 Comparison of the base and proposed work in terms of storage space



Hexadecimal code:	73	43	14	14	44
Swapping digits of each code:	37	34	41	41	44

Adding hexadecimal equivalent adding 0B

37	34	41	41	44
+11	+11	+11	+11	+11
48	45	4C	4C	4F

Plaintext: H E L L O

4 Experimental Evaluations

For experiment Java language Fig. 2 shows the output of the proposed approach which has the key size and the size of the memory in kb.

As proposed work has key generation thrice so the security and complexity will increase.

5 Conclusion

In this paper, a new advanced secure cryptographic approach has been projected which is resultant form the fundamental cryptographic algorithms and procedures are introduced and a new key generating process is presented,

1. The key generation imposes randomized encryption
2. Key 3 are generated from the plain text which leads to the most complex form of the base algorithm.

Acknowledgements I acknowledge here my debt to those who have contributed significantly to this proposed paper. I indebted to my internal guide Mr. Sampat Medhane, Department of Information Technology, Bharati Vidyapeeth University, College of Engineering, Pune for helping me and his experience is very helpful to me.

References

1. Robling Denning, Dorothy Elizabeth. *Cryptography and data security*. Addison-Wesley Longman Publishing Co., Inc., 1982.
2. Davies, Donald Watts, and Wyn L. Price. Security for computer networks: and introduction to data security in teleprocessing and electronic funds transfer. John Wiley & Sons, Inc., 1989.
3. Rivest, Ronald L. "The RC5 encryption algorithm." Fast Software Encryption. Springer Berlin Heidelberg, 1994.
4. Wheeler, David J., and Roger M. Needham. "TEA, a tiny encryption algorithm." Fast Software Encryption. Springer Berlin Heidelberg, 1994.
5. Matthews, Robert. "On the derivation of a "chaotic" encryption algorithm." Cryptologia 13.1 (1989): 29–42.
6. Guan, Zhi-Hong, Fangjun Huang, and Wenjie Guan. "Chaos-based image encryption algorithm." Physics Letters A 346.1 (2005): 153–157.
7. Gao, Tiegang, and Zengqiang Chen. "A new image encryption algorithm based on hyper-chaos." Physics Letters A 372.4 (2008): 394–400.
8. Pareek, N. K., Vinod Patidar, and K. K. Sud. "Discrete chaotic cryptography using an external key." Physics Letters A 309.1 (2003): 75–82.
9. Camtepe, Seyit A., and Bülent Yener. "Key distribution mechanisms for wireless sensor networks: a survey." Rensselaer Polytechnic Institute, Troy, New York, Technical Report (2005): 05–07.
10. Bergamo, Pina, et al. "Security of public-key cryptosystems based on Chebyshev polynomials." Circuits and Systems I: Regular Papers, IEEE Transactions on 52.7 (2005): 1382–1393.
11. Kohda, Tohru, and Hirohi Fujisaki. "Jacobian elliptic Chebyshev rational maps." Physica D: Nonlinear Phenomena 148.3 (2001): 242–254.
12. Blundo, Carlo, et al. "Efficient key management for enforcing access control in outsourced scenarios." Emerging Challenges for Security, Privacy, and Trust. Springer Berlin Heidelberg, 2009. 364–375.

Performance Analysis of Efficient Position-Based Opportunistic Routing for MANET

N.S. Kavitha, P. Malathi, Jim Mathew Philip and S. Pravinth Raja

Abstract Mobile Ad hoc Network (MANET) is the kind of ad hoc networks that dynamically form a temporary network by connecting the mobile nodes without any support of central administration and wires. This work coins the difficulties of delivering the data packets to the MANET circumstance in a reliable fashion and in good timely manner, which reside to be a highly dynamic in nature. In MANET, there are several routing algorithms, which consume topology information to make routing decisions at each node. Practical routing protocols must handle intermittent connectivity and the absence of end-to-end connections. An Efficient Position-Based Opportunistic Routing (EPOR) protocol is proposed and it will be able to take the advantage of stateless property, which uses two terms (1) geographic routing (2) broadcast nature of wireless medium. When the packet transmission starts, neighbor nodes have chance to overhear the transmission of messages, and it will act as a forwarding candidate. Thus, the node which was overheard the transmission information will forward the packets by enhancing the informations. The apt forwarder will identify the falsely relay node within a limited amount of time. The proposed algorithm removes the drawbacks of the existing GPSR and AOMDV. The additional latency occurred by the local route recovery has maximum being reduced and the duplicate relaying due to packet reroute can also be decreased. The proposed scheme of Virtual Destination-based Void Handling (VDVH) combined with EPOR protocol in a communication hole. The EPOR result demonstrates that can able to succeed the greater performance even under high mobility of nodes with considerable overhead and the new VDVH scheme also works well. This paper focuses on various performance evaluation metrics such as latency, throughput for the EPOR routing protocols and experimentally compared the performance of three routing protocols using NS-2 simulator.

N.S. Kavitha (✉) · J.M. Philip · S. Pravinth Raja
Department of CSE, Sri Ramakrishna Institute of Technology, Coimbatore, India
e-mail: nsksrit@gmail.com

P. Malathi
Bharathiyar Institute of Engineering for Women, Salem, India

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1_63

Keywords Medium access layer · Mobile ad hoc network · Efficient position-based opportunistic routing · Position-based routing

1 Introduction

Mobile Ad hoc Network (MANET) deals with number of mobile nodes that are free to move randomly. This is not in the case of the existing Internet, where the router topology is fundamentally static in nature (barring router failures or network configuration). In a MANET, the nodes are mobiles and inter-node connectivity may changes frequently due to dynamic configuration of network topology. In this track, we focus our attention on current protocols that affords the connectivity of route in MANET, such as Routing protocols [1]. The challenging problem of MANET especially the routing approaches where designed efficiently and reliably by utilizing the limited resources. An intellectual routing approach is recommended that proficiently uses the limited resources. Meanwhile the configuration of routing needed to adapt to the dynamic changes of network conditions such as network size, network partitioning, and traffic density [2].

The former interest of wireless networking is infrastructure network (wired networks). In wired networks, two main algorithms are used and they referred as link state and distance vector algorithms. In link-state routing algorithm, each of the Mobile Node (MN) or terminal periodically updates the network status by using a flooding strategy [3]. The update does by the process of broadcasting the costs- of-link-state routing of its neighboring nodes to other MNs. When each of the MNs receives an update packet, thus they update the entire network view and their link-state information. The Shortest path algorithm of link state provides the next hop node with the minimal distance from the source node to reach the desired destination. In distance vector routing, each node updates the network status by a periodical propagation of signals among the nodes for attaining the distance to every node [1, 4]. The conventional link-state routing and distance vector routing algorithm not supported in large MANETs. Due to periodic or frequent changes of network route updates in large network that may consume adequate resources such as bandwidth, power consumption that leads to frequent recharge of mobile terminals battery and increases the channel contention.

Conventional topology-based Mobile Ad hoc Network routing protocols is reasonably prone to node mobility. The crucial reason for occurrences of vulnerability is due to the pre determination of routes before data transmission [5, 6]. Owing to the frequent and rapid change of topology leads to hard maintainability of paths during transmission. Eventually, the discovery and recovery of routes consumes more time that reduces the mobile nodes energy. In this case, if the path breaks then the data packets will lose or delayed for long time until the reconfiguration or reconstruction of the route has handled that causes interruption during the transmission of data packets.

Geographic routing (GR) utilizes the location information to the purpose of forwarding data packets, which will utilize the hop-by-hop routing fashion [6]. To select the next hop forwarder, the Greedy forwarding takes the largest positive progress toward the destination. In order to trigger the route around communication voids, void handling mechanism is used. The key aspect of GR ensures the scalability and high efficiency thus the maintenance regarding the end-to-end routes is not required. However, GR is sensitive to the factual error of location information [7]. In greedy forwarding, an operation to select the next hop is quite distinct. Here the neighbor, who is located away from the sender or source that takes up the position as the next hop. If the node moves away from the coverage area of sender, it will leads to the transmission failure. A famous Geographic Routing (GR) protocol is GPSR. In this feedback of MAC-layer failure, offer the chance to reroute the packet. When node mobility increases, the performance evaluation is incapable under the simulation. Due to the nature of wireless medium broadcasting, the transmission of a single packet leads to numerous reception. If this type of transmission used as backup, then the strength of the routing protocol considered as enhanced.

In opportunistic routing, the concept of multi-cast routing has already demonstrated. However, most probably link-state topology DB is used to select and ensures higher preferences for the forwarding candidates or relay nodes. The periodic network-wide measurement is required to acquire the loss rates of inter node and it is impossible because of the dynamic nature of MNs [7]. So far, in these protocols, the batching tends to delay the data packets and not preferred for delay sensitive applications [8, 9]. Presently, the location aided opportunistic routing guided to forward the packets directly by using location information but it's designed as of other static mesh networks and aimed to provide the efficient throughput. However, opportunistic forwarding fails to investigate the robustness of the network for transmission of data packets [6, 10]. The proposed novel Efficient Position-based Opportunistic Routing (EPOR) protocol deals the forwarding candidates by using the MAC interception for caching the packet that has been received [7]. If the best forwarder not forwards the packet in an allocated time slot, a then suboptimal candidate will successfully collects/receives and forward the packets without any interruption to the data transmission. Thus, EPOR also ensures the robustness by exploiting per packet basics through multi-path accessing.

2 Routing Algorithm

1. Greedy Perimeter Source Routing (GPSR)
   MAC-layer failure feedback: As used in DSR [3], we the data packet exceeds the limitation of retransmission tries then DSR receives the notification from the 802.11 MAC layer. It excludes the congestive collapse that indicates the

recipient out of the frequency range [6, 10]. If the retry goes beyond the limit, then it attains the failure of transmission. By using, the MAC- layer failure feedback may convey the failure state as earlier as possible to the GPSR, otherwise it carries the task through the neighbor's expiration timeout interval (4:5B).

Interface queue traversal: Regarding the MAC- layer feedback, the implementation expresses the strong result. While an IEEE 802.11 interface continuously retransmits the data packets at the packets head of its queue. The head-of-line block normally waits to receive the link-level acknowledgement from the recipient. This head-of-line block used to reduce the offered transmits cycle of an interface [8]. In this situation, notification regards the retransmit retry makes queue of the packet to traverse for the interface and clears out all the failed transmission recipients addresses [11]. Then it passes all these packets to the routing protocol for retransmission or re-forwarding to another next hop.

While receiving greedy mode data packet for forwarding, a node looks its neighbor table for capturing the geographically located closest neighbor to the desired packet's destination. If the neighbor is closer to the destination, then the greedy mode packet forwards the packet to the closest node. Suppose, no neighbor is closer, then the node marks the packet under the perimeter mode [9]. However, most probably link-state topology DB is used to select and ensures higher preferences for the forwarding candidates or relay nodes. The periodic network-wide measurement is required to acquire the loss rates of inter node [9]. It is impossible because of the dynamic nature of MNs. So far, in these protocols, the batching tends to delay the data packets and not preferred for delay sensitive applications.

2. Ad Hoc on Demand Multicast Distance Vector Routing (AOMDV)

   AOMDV's many characteristics were shares with AODV. The technique based on the distance vector routing algorithm, uses an approach as hop-by-hop routing. Furthermore, AOMDV finds the retransmission routes using a route discovery procedure [12]. The difference between AOMDV and AODV is estimates by the discovering the number of routes that establishes in each of the nodes route discovery to attain the desired destination.

   In terms of AOMDV, it broadcast the Route Request (RREQ) from source to destination. It establishes the multiple reverse paths both at destination node as well as at intermediate nodes. The reverse path configuration from source to destination and intermediate node to destination makes multiple forward paths. The intermediate nodes reverse the path with an alternative path and it is useful in reducing the frequency of the route discovery. The depth of the AOMDV protocol ensures the discovery of multiple paths that are disjoint and loop free [13] and uses a flood-based route discovery, to finds the paths efficiently. However, it does not involve the special control to the packets [10]. In fact, AOMDV and AODV cause the overhead due to the extra RREPs and RERRs for discovery of multipath that harms to maintain in routing control packets (i.e., RREQs, RREPs, and RERRs).

3. Efficient Position-Based Opportunistic Routing

   An Efficient Position-based Opportunistic Routing (EPOR) mechanism which can be organized without any of the complex [14] modification to MAC protocol and meanwhile achieves the multiple candidate reception without dropping the additional benefits of collision avoidance [6, 10] that was provided by IEEE 802.11. The scheme of Virtual Destination-based Void Handling (VDVH) exhibits the advantages of greedy forwarding and opportunistic routing can achieve the transmission while handling communication voids.

   In this paper, EPOR focuses both overhead as well as usage of bandwidth due to the duplicate relay candidates for transmission also look the utilization of buffer usage. Through performance analysis, EPOR achieves gain ratio by little overhead cost due to the proper selection of forwarding candidates among the area and limitation schemes [9, 14].

   A node needs to satisfy the following conditions to locate in the forwarding region/area:

   1. Should achieve the positive progress to reach the destination [7]
   2. The distance between the source and next hop relay node should not go beyond the half of the transmission range (R) (i.e., R/2).

   Therefore, all forwarding candidates can hear each other. The key point of EPOR states that, if an intermediate node receives the same ID (i.e., same source address and sequence number) [15], then it will drop that packet from the best forwarder packet list.
4. Opportunistic Forwarding

   The design of EPOR depends upon opportunistic forwarding and geographic routing. The MNs aware their own location and their neighboring positions directly. The MNs neighborhood location informations are register in the routing table of each of the MN. Exchange of information uses one–hop beacon in the packets header [16]. To achieve the desired destination position the location registration and look up services maps the nodes addresses to locate the desired destination nodes position as in [6, 10]. In this scenario, location services schemes are more reliable and efficient. For example, when the destination replies to the requested source, destination transmits the reply by acquiring the long-range radio with low bit rate. It can implement by periodic beacon. When a source needs to transmit a packet, first, it should collect the destination location and its address attached with the packet header. Due to the movement of destination node, the true destination diverge the multi hop path and packet may drop even if the packet delivers to the neighborhood destination. To handle such issues, optimal check introduced for the destination node. At each stage of hopping, the nodes that forward the data packet will check its neighbor list to check whether the destination node presents within its own transmission range. If so, the packet will transmit directly to the destination node, likewise location prediction scheme for destination as mentioned in [4].

   In opportunistic forwarding, in order to require packet from various candidates, either it must adapt to any one of the following, an integration of routing or an

IP broadcast and MAC protocol. Foremost, it prone to MAC collision due to the absence of collision avoidance that supports to broadcast the packets in IEEE 802.11. Later it need a complex coordinates and it is quite difficult to implement. In EPOR, the similar scheme which was uses in MAC multicast mode explained in [15]. Here the packet transfer in unicast mode in IP layer (the best forwarder/relay node which makes the leading positive progress for attaining the destination and it is set as next hop) and receives the multiple candidates reception using MAC interception. The significance of RTS/ACK/DATA/CTS reduces the collision. Due to the medium reservation, all other nodes within the ranges of sender's node may overhear (eavesdrop) on the successful transmission of the packet as a higher probability in ratio.

3 Simulation Results

The performance of EPOR has evaluated and compared with AOMDV and GPSR in NS-2 simulation. It simulates under a variety of network topologies. The GPSR is a representative geographic routing protocol and AOMDV is a famous multi-path routing protocol. The parameters utilized in the NS-2 simulation listed in Table 1. The model uses the random waypoint for node's mobility without pause. The speed of the network mobility degree may vary from minimum to maximum. The minimum speed of the node is set to 1 m/s. The following metrics used for comparison of performance:

- *Packet delivery ratio*: The ratio of the number of data packets received at the destination(s) to the number of data packets sent by the source(s).
- *End-to-end delay*: The average and median end-to-end delay evaluated together with the cumulative distribution function of the delay.
- *Path length*: The average end-to-end path length meant (number of hops) for successful packet delivery.
- *Packet forwarding times per hop* (*FTH*): The average number of time, a packet has forward from the perspective routing layer to deliver a packet over each hop.

Table 1 Simulation parameters

Parameters	Values
MAC protocol	IEEE 802.11
Propagation model	Two-ray ground
Transmission range	250 m
Mobility model	Random Way Point (RWP)
Traffic type	Constant Bit Rate (CBR)
Packet size	512 MB
Number of nodes	100
Simulation time	200 μs

Fig. 1 Packet delivery ratio for communication hole

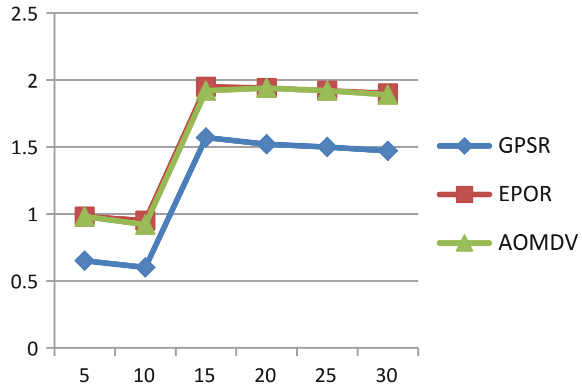
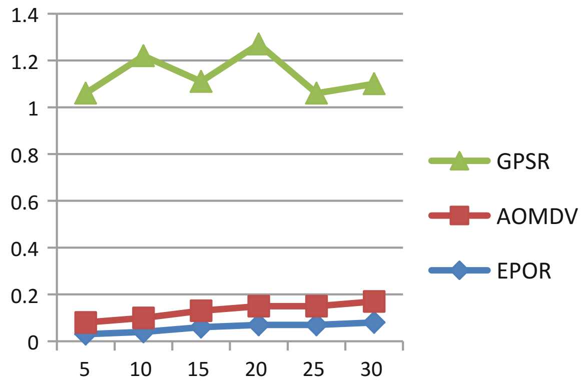


Fig. 2 Average end-to-end delay for communication hole



- *Packet forwarding times per packet* (*FTP*): The average number of times a packet has forward from the perspective routing layer to deliver a packet from the source to the destination.

A. Simulation Parameters

As by Fig. 1, we can observe that in the face of communication hole, GPSR's void handling mechanism fails to work well [17] (Figs. 2, 3 and 4).

Even when the maximum node speed is 5 m/s, only 90% of the data packets get delivered which is relatively poor compared to the other protocols [18]. As for EPOR, the improvement is not so significant since in the current implementation, VDVH is unable to deal with all cases of communication voids.

However, when the node mobility is high (e.g., when the maximum node speed is larger than 25 m/s), EPOR still performs better.

With respect to the path length, the end-to-end hops of GPSR are the largest due to the usage of perimeter mode, while EPOR still achieves the shortest path length [19].

Fig. 3 Median end-to-end delay for communication hole

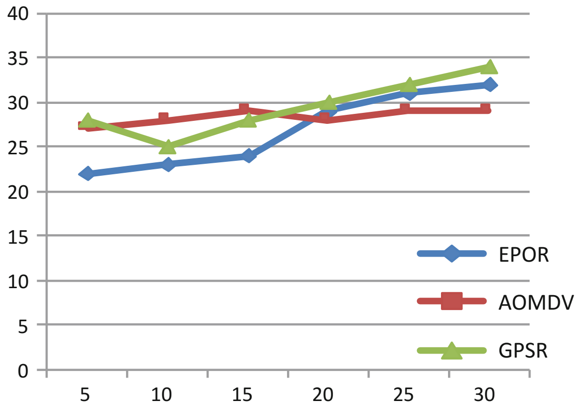


Fig. 4 Path length for communication hole

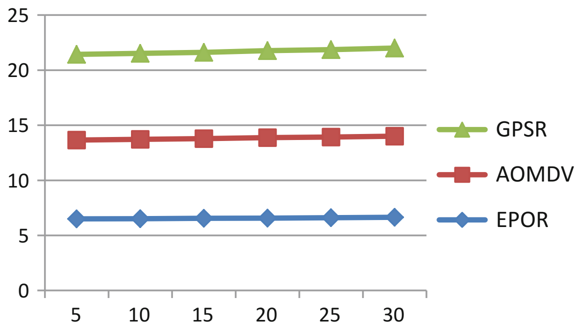
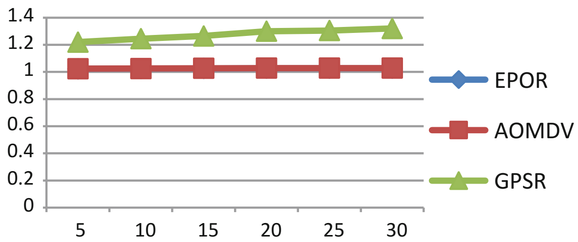


Fig. 5 Packet forwarding times per hop for communication hole



As for the result of FTH and FTP (Figs. 5 and 6), EPOR outperforms the other two as usual while GPSR performs worst indicating the perimeter mode of GPSR is incapable of working well in mobile environment (Figs. 7 and 8)

Fig. 6 Packet forwarding times per packet (FTP) for communication hole

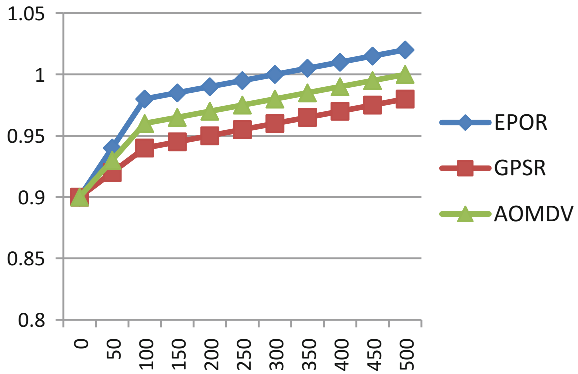


Fig. 7 Packet delivery ratio for multi fold

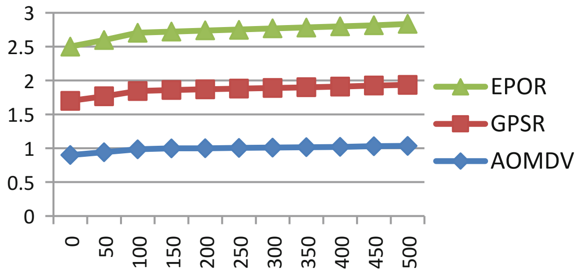
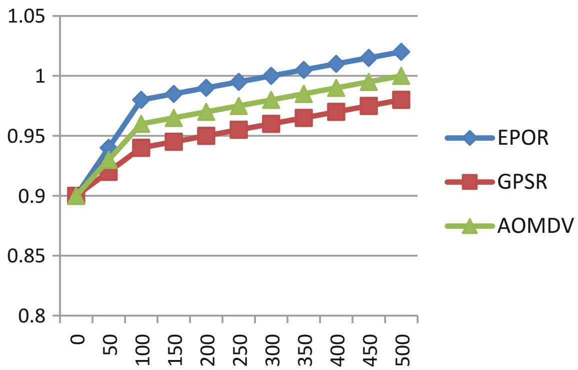


Fig. 8 CDF for single flow



4 Conclusion and Future Work

In MANET, we face the problems of frequent changes in network topology that leads to break the link would either lose the packets. Thus a novel routing protocol EPOR which takes the advantage of geographic routing and broadcast nature of wireless medium and ensures the better performance in case of link break. Through simulation, EPOR confirm the effectiveness and efficiency under high PDR.

By involving the forwarding candidates in EPOR, the property of backup is useful in the case of broken route and it will recover in timely manner. An overhead also addressed by opportunistic routing. This method can combined with a new proposal of multiple sources with a single destination in a parallel processing method [10], which will make the PDR to be increase in an energy efficient way by combined with load balancing.

References

1. Anil Kulharni, Arati Gowri, “A Reliable Data Delivery in MANET Simulation Study”, International Journal of Computer Science and Network Security, Jul 2014.
2. Ashema Harti, “Study of Impact of Mobile Ad-Hoc Networking and It’s Future Application”, International Journal of Information Technology, Jun 30, 2008.
3. Ederral Pable Ferreri da Cruz, Lusiss Edwardo Gottardo, Franciele Pereira Rossini, Vinicius Souza Oliveira and Lucas Cellin Perreira, “Evaluating Feasibility of Using Wireless Sensor Networks in Coffee Crop Through Simulation of AODV, AOMDV, DSDV and Their Vawanis With 802.15.4 MAC Protocol”, International Journal of Complete Network and Communication, Nov 2011.
4. Jim Mathew Philip, Dr. Malathi.P and Kavitha. N.S “Promoting fairness in mobile ad hoc networks”, Bonfring International Journal of Research in Communication Engineering, vol-1, December 2011.
5. Jubiin Sebastia E, Sreeraj V.R., Tauheed Islam, “Reliable Data Delivery Mechanisms for Dynamic MANETs”, International Journal of Science and Research, Mar 27, 2011.
6. Kalaivani.P, Sathya.G, Senthilnathan.N, “Dynamic Data Routing In Manet Using Position Based Opportunistic Routing Protocol”, International Journal Of Research In Computer Applications, Dec 2014.
7. Kavitha.N.S, Dr. Malathi.P and Jim Mathew Philip, “Improving throughput and controlling congestion collapse in mobile ad hoc networks” Bonfring International journal of research in communication engineering, vol-1, December 2011.
8. Kavitha.N.S, Dr. Malathi.P, lavanya.R and Jim Mathew Philip “Efficient load balancing algorithm for MANET”, International Journal of Applied engineering Research Vol-9, 2014.
9. Katroth Blakrishna Maruthiram, Karo Suresh Babu, “Performance Comparison of DSDN, OLSR, DSR and AODV MANET Routing Protocol in Traffic Condition”, International Journal of Science and Research, Aug 29, 2014.
10. Shengbo yang, chai kiat yeo, bu sang lee, ”Towards Reliable Data Delivery For Highly Dynamic Mobile Ad Hoc Networks”, IEEE Transactions On Mobile Computing, 2011.
11. Umamaheswaran. A, Gopikrishnan. S, Saranya. R “Reliable Data Transfer on Dynamic Nodes Using Packet Hiding Methods in Ad Hoc Networks”, International Journal of Innovative Research in Science, Engineering and Technology, Jul 18, 2013.
12. Kavitha. N.S, Malathi. P & Jim Mathew Philip, “Congestion Control in Mobile Ad Hoc Networks”, International Journal of Network and Mobile Technologies, Vol. 2, No.1, 2001.
13. Komma Reddy Dhana Lakshmi, Raja Ramesh, “Energy Efficient Position Based oppertunistic Routing Protocol and MANETs”, International Journal of Computer Science and Network Security, Jun 2014.
14. M. Dhevraj, G. Gridhar, “Encrypted and Consistent Data Transaction in Dynamic MANETs”, International Journal of Advance Research in Computer Science and Software Engineering, Jul 2013.
15. Mitali Sinha, Mira Ran Debbarma, Sangita Rani Bhownik, Trina Si, Ihunu Dennarma, “On The Performance Analaysis of AODV Protocol in MANET”, International Journal of Innovative Research in Computer and Communication Engineering, vol 2, Issue7, jul 2014.

16. Safna Sulaiman, Tintu Alphansa Thomas, "A Neighbour Coverage-Based Rebroadcast in MANETs Based on Energy Efficient Rebroadcast Probability", International Journal of Computer Science and Network Security, Oct 2014.
17. Prerna Setia, Hunat Uppal, "Simulation And Performance of Analysis of Proactive, Reactive and Hybrid Routing Protocols Using Static and Mobile Environment in MANET", International Journal of Science and Research, Oct 2013.
18. Shengbo Yang, Chai Kait Yeo, Bu Sung Lee, "Towards Reliable Data for Highly Dynamic MANET", IEEE Transactions, Mar 7 2011.
19. Ramkumar. V. D, "Efficient Routing Protocol for Highly Dynamic Mobile Ad Hoc Networks", International Journal of Computer Science and Mobile Computing, Jul 2013

Author Index

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1

Subject Index

N.R. Shetty et al. (eds.), *Emerging Research in Computing, Information, Communication and Applications*, https://doi.org/10.1007/978-981-10-4741-1