IITM Journal of Information Technology

Volume 4 (January - December 2018)
ISSN No. 2395 – 5457
SPECIAL ISSUE

INSTITUTE OF INNOVATION IN TECHNOLOGY & MANAGEMENT
Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as 'A+' by SFRC
NAAC Accredited Institute
(ISO 9001-2008 Certified Institute)

Ministry of Electronics and Information Technology
Government of India

NCIRCSIT'2018

National Conference on InnovativeResearch in Computer Science and Information Technology: "Advanced Computing Trends & Cyber Security"

February 17, 2018

IN TECHNICAL COLLABORATION WITH
COMPUTER SOCIETY OF INDIA (CSI)

SPONSORED BY
DEFENCE RESEARCH AND DEVELOPMENT ORGANIZATION (DRDO)

SUPPORTED BY
MINISTRY OF ELECTRONICS & INFORMATION TECHNOLOGY (MeitY)

Editors: Dr. Geetali Banerji, Dr. Ankit Verma

IITM Journal of Information Technology

IITM Journal of Information Technology is a National Annual Journal of Information Technology intended for professionals and researchers in all fields of Information Technology.Its objective is to disseminate experiences, ideas, case studies of professionals in Management and Information Technology to propagate better understandings. Its focus is on empirical and applied research and reflections that are relevant to professionals of Management and Information Technology with academic standards and rigor within purview. The views expressed in the Journal are those of authors. No part of this publication may be reproduced in anyform without the written consent of the publisher.The Journal intends to bring together leading researchers and Information Technology practitioners from around the country.

Editorial Board

Patron

Sh. J.C. Sharma

Sh. Ravi Sharma

Editor in Chief

Prof. (Dr.) C.P. Chawla

Editors

Dr.Geetali Banerji

Dr.Ankit Verma

Editorial Team

Dr. Prabhat Kumar Vishwakarma

Dr. Rashmi Jha

Prof. Harsh Arora

Prof. Madhu Chauhan

Prof. Priti Sharma

Prof. Pramod Soni

Prof. Pooja Saigal

Prof. Charul Nigam

Prof. Govind Murari

Prof. Himanshi Bhambri

Prof. Akhil Kumar

Prof. Kanika Dhingra

Prof. Kanika Bhalla

Prof. Isha Shingari

Prof. Vishakha Sehdev

Prof. Suruchi Sinha

Prof. Sushma Malik

Prof. Deepti Gupta

Prof. Ankita Miglani

Prof. Nemi Rishi

Advisory Board

Dr. G Athithan, DS and DG (MED, CoS, & CS), DRDO

Dr. P. K. Saxena, OS/SC 'H", Director, SAG, DRDO

Dr. V.K. Panchal, Former Associate Director, DTRL, DRDO

Smt. Anu Khosla, Scientist 'G', Director SAG, DRDO

Dr. Prashant Goswami, Director, CSIR, NISTADS

Dr. Sujit Bhattacharya, Sr. Principal Sci, CSIR, NISTADS

Dr. Arvinder Kaur, Professor, USIT, GGSIPU

Dr. Navin Rajpal, Professor & Dean, USIT, GGSIPU

Dr. B.V.Ramana Reddy, Dean, USET, GGSIPU

Dr. A. K. Saini, Professor, USMS, GGSIPU

Mr. Sanjay Mohapatra, President, CSI

Shri. Gautam Mahapatra, Vice President, CSI

Prof. A. K. Nayak, Hon. Secretary, CSI

Mr. Manas Ranjan Pattnaik, Hon. Treasurer, CSI

Dr. Anirban Basu, Immd. Past President, CSI

Dr. Ajay Kumar, IAS, Additional Secretary, MeitY, GOI

Shri. Arvind Kumar, Scientist G, MeitY, GOI

Dr. A S Kamble, Senior Director, MeitY, GOI

Dr. Rakesh Maheshwari, Scientist G, MeitY, GOI

Sh. Pawan Duggal, Advocate, Supreme Court of India

Dr. Sushila Madan, Professor, University of Delhi

Dr. Suresh Chandra, Professor, IIT

Dr. P. C. Saxena, Professor Emeritus, JNU

Dr. M. N. Hoda, Director, Bharti Vidyapeeth

Dr. Varun Monga, Professor & Dean, MVN University

Dr. Sunil Khatri, Director, Amity University

Dr. Shucheta Bhasin, Professor, Kurukshetra University

Dr. Jay Shankar Prasad, Professor & Head, MVN University

Dr. Kanak Saxena, Professor, SATI, Vidisha

IITM Journal of Information Technology

Annual Journal of Institute of Innovation in Technology & Management

Volume 4 **January-December, 2018**

CONTENTS

Research Papers & Articles

Page No.

- An Improved Approach for Accurate Weather Forecasting 3-8
  - Shubham Aggarwal, Sunil Kumar Chowdhary
- Sentiment Analysis of WhatsApp Group Chat 9-16
  - Lakshya Munjal, Anmol Arora, Neetu Narwal
- Deep Learning: Powerful Metaphor to Understand World 17-19
  - Nishant Nirmal, Suruchi Sinha
- Transformation of Business from E-commerce to M-commerce 20-27
  - Sushma Malik, Prashant Kumar
- Security Issues and Challenges while Outsourcing Computing Resources for E-Business 28-31
  - Prabhat Kumar Vishwakarma, Geetali Banerji
- Optimizing Operating Time using Clustering Technology in Wireless Sensor Networks 32-35
  - Vishakha Verma, Ankit Verma
- Color Image Segmentation Using Twin Support Vector Machine 36-41
  - Pooja Saigal
- Past Present & Future of Data Mining 42-44
  - Priti Sharma, Shikha Sharma
- Comparative Analysis of Automation Testing Tools 45-50
  - Charul Nigam, Sushma Malik
- A Review of Various Energy Optimization Routing Protocols in Wireless Sensor Networks 51-56
  - Nemi Rishi, Akhil Kumar
- Impact of 3D Printing Technology in Medical Science 57-61
  - Stuti Kapoor, Aparna Bhardwaj, Vishakha Verma, Ankit Verma
- Big Data Analytics Approach using Mapreduce Algorithm on Hadoop 62-67
  - Harsh Arora, Mayank Gupta, Himanshi Kohli
- A Review on Exploring the Recent Trends and Future Scope in Artificial Intelligence 68-71
  - Amanpreet Kaur, Isha Shingari
- Effective Design Prevention Approach in Software Process for Achieving Better Quality Levels 72-75
  - Kanika Bhalla, Shivangi Kharbanda
- A Descriptive Framework of Data Mining Approach and its Techniques in Multiple Fields 76-80
  - Govind Murari Upadhyay, Shivangi Kharbanda, Shubhangi Khandelwal

BLANK

An Improved Approach for Accurate Weather Forecasting

Shubham Aggarwal*
Sunil Kumar Chowdhary**

Abstract

As recurrence and impact of some weather events are increasing, therefore need to tackle such events is the concern. Climate change is one of the factor causing such increase in magnitude and frequency of extreme weather events. Weather forecasting system can help in managing such events. Daily weather forecasting and monitoring promotes resource management and disaster preparedness. This research focused on the daily weather conditions of cities of different countries like Shanghai, Manila, Tokyo, Kolkata and Mexico. Weather conditions of these cities were monitored and recorded from different weather forecasting websites on daily basis. Based on this recorded data an improved weather forecasting model can be developed through statistical analysis of temperature, wind speed, quantitative precipitation of extreme weather events happened among these cities in recent decades. Using Minitab statistical software, the accurate weather providing website was analyzed. The deviation of these readings helped in analyzing how accurate are the reading provided by the weather forecasting websites with respect to calculated mean of the readings.

Key Words: Weather Forecasting, Management, Minitab, Standard Deviation

I. Introduction

Weather is referred as the atmospheric conditions of a particular location and time. Application of science and technology to predict the weather conditions is known as Weather forecasting. Temperature, wind speed, rain, atmospheric pressure, humidity etc are the major attributes that affects weather condition of a particular region. Predicting accurate weather conditions has always been the most challenging task. Factors like climate change and global warming makes the prediction to be more difficult and challenging. Rapid growth in population in India and their policies of adopting urbanization has led to increase various socio-economic impacts due extreme weather events such as cyclones, flood, hail storm, heat and cold waves, drought, thunderstorms.

Many of today's weather forecasting systems rely on observations and analysis done by meteorologists using conventional principles. As many new models are being developed for weather forecasting every year but most of them doesn't provide accurate values. The need is to use the correct algorithms for effective and accurate prediction of weather information. Also, different online weather forecasting stations provide different weather conditions; therefore the need to develop a model that can guide user to the best possible source from where accurate weather conditions can be fetched arises. Weather Forecasting proves to be a useful in various info-science fields like disaster management, early warning system in airport or naval system, helping farmers in protecting their crops against any extreme weather conditions. Today's weather forecasting system includes combination of various technologies, computer models, trends and patterns knowledge, observations etc. Utilizing these

Shubham Aggarwal*
Student of B.Tech,
Amity School of Engineering and Technology,
New Delhi, India
Sunil Kumar Chowdhary**
Department of Information Technology,
Amity University Uttar Pradesh,
New Delhi, India

technologies accuracy of weather forecasting can be calculated for five days in advance.

II. Data Preprocessing

Daily weather conditions of some extreme weather affected cities of the world were monitored and recorded. This research focused on the daily weather conditions in Shanghai, Manila, Tokyo, Kolkata, Mexico. The weather conditions of these cities were monitored and recorded on daily basis for up to 20 days. Weather conditions were recorded on the basis of current temperature, quantitative precipitation and wind speeds. The data was formulated in form of table. The columns represent five cities of different countries affected by weather events and the rows shows weather forecasting websites. [9] [10] [11] [12] [13] To compare variations in the recorded weather conditions, we selected top weather forecasting websites- Accu weather, intellicast, weatherbug, weather underground and yahoo weather.

III. Literature Review

This section of paper investigates some recent researches in weather forecasting domain that have been done. It overviews the models, techniques, management processes that have already implemented in this domain.

V. Bjerknes and LF Richardson initiated numerical weather prediction in early 1990s [4]. This was much related to the initial value problem of statistics in mathematics. Models used for accurate weather prediction were based on linear regression. These models were easy and simple to understand. Multiple regression model was formulated further to improve the accuracy. In [5], research defines solution to weather forecast using Map Reduce Framework. Researchers describes how data set and patterns of the recent years helps in prediction of storm and Support Vector Machin(SVM) helped in classifying the data.

In [1], the researchers performed statistical analysis of weather conditions using historical weather forecast data of the National Oceanic and Atmospheric Administration [NOAA]. Important findings was that there were significant variations in forecast accuracies, observed probabilities of precipitation were significantly lower than forecasted one, Forecasting organizations generally under predict wind speeds by a large margin for days when wind speed exceeds 20 mph.

In [3], the authors developed various models to predict daily weather forecast in cities like Albay, Tiwi. Multilayer Feedforward Artificial Neural Network (MLFANN) architecture model was developed and trained with Resilient Propagation (RPROP) algorithm. The research also evaluated the Artificial Neural Network (ANN) model capability in weather forecasting.

Structural and Time series based weather prediction algorithms can be used with IOT devices in this domain [6]. Support Vector Regression, Linear Regression, Multiple Linear Regression (MLR), Auto Regressive Integrated Moving Average (ARIMA) are some of these algorithms. Using them the author observes the behaviour of forecasting algorithms. In [8], Multiple linear regression (MLR) equations is used for developing a model for weather prediction. Author used the coefficients of these regression equations for generating a prediction pattern for weather.

This study focuses on developing a model that could help in providing most accurate weather information among the selected weather providing website and to create warning system by examining the weather conditions of these cities during extreme weather events in recent decades.

IV. Methodology

A statistical analysis is performed on the weather reading recorded for 20 days. Among these different weather forecasting websites best one was chosen on the basis of total deviation. Also, historical weather data of these cities during extreme weather events were examined and their attributes such as wind speed, quantitative precipitation were analyzed.

Using Minitab statistical software, a statistical analysis has been done. The mean of the current temperature from different weather providing websites have been

calculated and based on these calculation, deviations was calculated for different website. All the deviation for different cities by a particular website were summed. And based on this obtained deviation least deviated value was figure out among different obtained deviation for particular date.

23-Oct-17	**Recorded temperature in celsius**						**Deviation Table**					
	Shanghai	Manila	Tokyo	Kolkata	Mexico		Shanghai	Manila	Tokyo	Kolkata	Mexico	**Total Deviation**
Accu Weather	23	29	23	35	21		1.8	0.3	0.2	0.6	0.6	**3.5**
Intellicast	22	30	23	34	20		0.8	0.7	0.2	0.4	0.4	**2.5**
Weather Underground	22	30.5	24	35	19		0.8	1.2	0.8	0.6	1.4	**4.8**
Weather Bug	20	28	23	34	21		1.2	1.3	0.2	0.4	0.6	**3.7**
Yahoo Weather	19	29	23	34	21		2.2	0.3	0.2	0.4	0.6	**3.7**
Mean	**21.2**	**29.3**	**23.2**	**34.4**	**20.4**							

Figure 2: Sample Dataset

V. Results and Analysis

Table-1: Sum of deviations from mean value (for current temperature)

Date	**Accuweather**	**Intellicast**	**Weather Underground**	**Weather Bug**	**Yahoo weather**
21 Sep 2017	**2.2**	4	2.6	3	4.5
22 Sep 2017	4.7	2.6	**1**	1.6	6.2
23 Sep 2017	3	3.6	**2**	7	5.5
24 Sep 2017	1.7	**1**	2.6	2.2	2
25 Sep 2017	**2.4**	3.8	4.2	**2.4**	6
26 Sep 2017	4.4	2.6	6.6	**2.4**	8.4
27 Sep 2017	1.6	1.6	3	**1.2**	3.8
28 Sep 2017	5.6	**3.4**	5.8	3.6	4.8
29 Sep 2017	3.5	**2.5**	4.8	3.7	3.7
30 Sep 2017	3	3.2	**1.8**	2.8	2.4
01 Oct 2017	3.2	**2.2**	3.8	3.6	2.8
02 Oct 2017	3.4	**2.2**	3.2	3.6	6
03 Oct 2017	4	**2**	**2**	2.2	3
04 Oct 2017	4.6	**3**	5	3.8	8.4
05 Oct 2017	3.2	**1.4**	3.8	3	5
06 Oct 2017	5.2	3.6	3.8	**3**	6
07 Oct 2017	4.4	**1.8**	3.8	3.2	2.4
08 Oct 2017	**1.6**	3	3.6	3.8	4.8
09 Oct 2017	3	**1.4**	4.8	3.6	2
10 Oct 2017	3.2	**2.4**	7	6.4	5

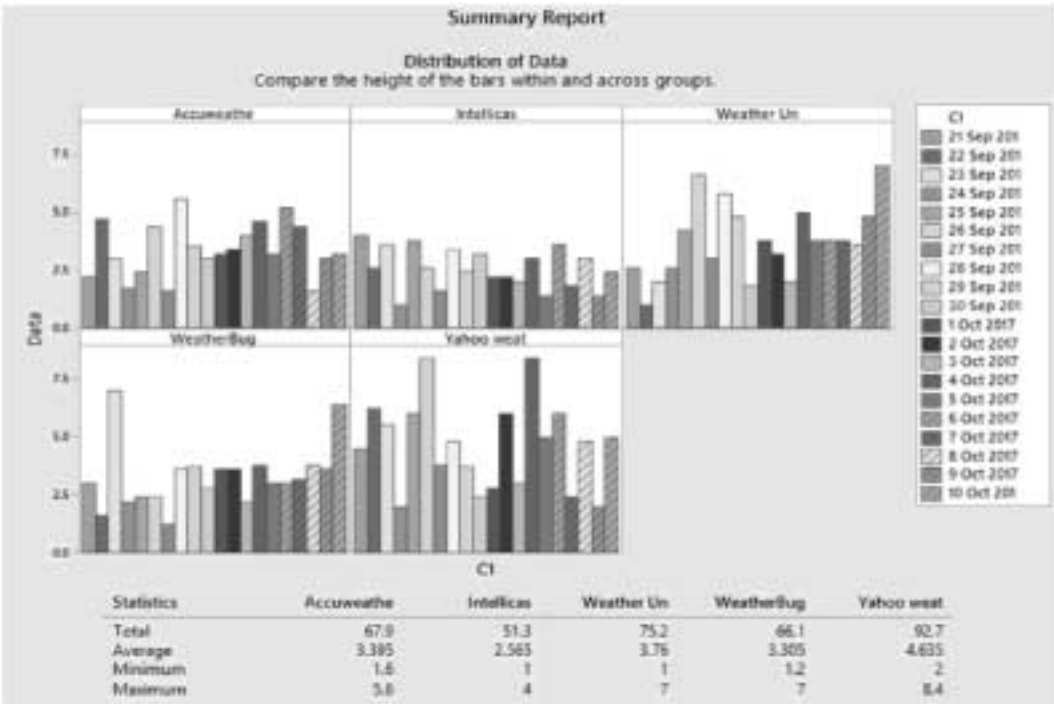


Figure 2: Minitab analysis report

The weather data for current temperature for this research has been recorded for 5 different cities since 21 Sep 2017 – 10 Oct 2017, for total 20 days from 5 different weather providing websites. The value under each website representing column represents sum of the deviations of the five selected cities.

Steps in obtaining deviation:

1. Consider the recorded data for calculating deviation.
2. Firstly, mean temperature was calculated for five different cities taken by summing the readings obtained by different websites (Fig. 2 Sample dataset)
3. Deviation table was made by calculating deviation from recorded temperature. (**Deviation** = Recorded temperature- Mean temperature)
4. Table-1 was obtained by summing the deviation obtained from five selected cities for particular website.
5. Using Minitab statistical software, an analysis has been done and the least deviation providing website was found.

This analysis has been done on the recorded current temperature. Minitab analysis report show four different statistics on basis of which the least deviated weather providing website was found. It can be seen that Intellicast weather providing website have 51.3 least total deviation. Also, the average deviation of this website was found to be minimum among these five websites that is 2.565. Also, minimum and maximum deviation of Intellicast is minimum ranging from 1-4 as compared to other websites. Therefore, from this analysis we concluded that the Intellicast provides least deviated weather reading and is most accurate than other websites. Yahoo weather website on other hand shows 92.7 maximum total deviation as well as the other statistics were very high as compared to other websites. So, Yahoo weather results to be least accurate in providing weather information.

The table-2 shows major disastrous weather events in these cities in past few decades. In Philippines,

Table-2: Major disastrous weather events in recent decades

City/Country	Event	Date	Wind speed	Precipitation	No. of Deaths
Philippines	Typhoon (Durian)	(Nov 25-Dec 6), 2006	195 kph-250kph	460mm	1399
	Typhoon (Ike)	(Aug 26-Sep 6), 1984	165kph-230kph		1492
	Typhoon (Haiyan)	(3-11 Nov), 2013	230kph-315kph		6241
Mexico	Hurricane (Wilma)	(16-27 Oct), 2005	295kph	1633.98mm	< 87
	Hurricane (Patricia)	(20-24 Oct), 2015	345 kph	193mm	< 10
Kolkata	Cyclone Storm BOB 04	(10-12 Nov), 2002	100kph	110mm	173
	Cyclone (Alia)	(26-27 May), 2009	110kph-120kph		149
Shanghai	Heavy Rainfall	(10 July), 2016		370mm	160
Japan	Tsunami	(11 March), 2011			1000 <
	Tsunami	(17 Jan), 1995			6000

Source: *www.imd.gov.in/pages/services_cyclone.php*
http://www.reuters.com/article/us-climatechange-lima-losses
https://en.wikipedia.org/wiki/Typhoons_in_the_Philippines
http://climatescorecard.org/2017/04/08/mexico-extreme-weather-event/
https://reliefweb.int/report/china/current-extreme-weather-events

three major extreme weather events have been seen in history. Typhoon Durian, Ike, Haiyan were the most disastrous events that causes more than 10000 deaths in total, millions homeless and damages to property. All of them proved to be worst typhoon of the century in Philippines having an average speed above 250 kph. Also, the frequency of these typhoons is maximum in November month. In Mexico, hurricane Wilma and Patricia were the major events causing over 100 deaths and the frequency of these disasters were maximum in October month. Both hurricanes with average wind speed above 300 kph results in heavy rainfall with an average height of 900mm. In Kolkata, cyclone storm bob 04 and Alia caused heavy rainfall with average windspeed of 110kph and the estimated fatalities were above 150. Both the cyclones existed for 2 days. In 2016, Shanghai faced 370mm heavy rainfall causing 160 deaths and flood.

The impact of such weather events doesn't seem to decrease. But the graph of deaths due to these events shows a significant fall. This is due to enhanced weather forecasting and warning systems, better disaster management strategies introduced all over the world and researches in this field. Using data of table-2, we can design a warning system that can warn in advance if the forecasted weather conditions of the day is near to or matching the weather conditions of above extreme weather events. This helps in managing the impact of disaster in advance and alarming the agencies and disaster management forces to tackle.

VI. Conclusion

This study shows how accurate weather conditions are provided by different weather forecasting websites and research on extreme and disastrous weather events in these cities in last few decades helps in knowing the weather conditions favourable for disaster and helps

in managing the disaster earlier by generating warnings. Intellicast website was found to be the most accurate weather providing website on the basis of deviation from the mean for current temperature through statistical analysis. As weather forecasting plays a vital role in all our day today activities. Therefore, it is an important area of research in human life. So, an improved approach for accurate weather forecasting was shown based on a comprehensive statistical analysis.

References

1. Halton Pi and Audrey W. Zhu, "A Method for Improving the Accuracy of Weather Forecasts Based on a Comprehensive Statistical Analysis of Historical Data for the Contiguous United States" 2014.
2. Sunil Navadia, Pintukumar Yadav, Jobin Thomas and Shakila Shaikh "Weather Prediction: A novel approach for measuring and analyzing weather data" International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) 2017.
3. Karla Louissa Marie D. Sobrevilla, Archie G. Quiñones, Kharl Vincent S. Lopez, Virna T. Azaña "Daily Weather Forecast in Tiwi, Albay, Philippines using Artificial Neural Network with Missing Values Imputation" 978-1-5090-2597-8/16/$31.00c 2016 IEEE.
4. M. N. Sulaiman, Shahi, R. Atan, "An Effective Fuzzy C-Mean and Type-2 Logic for Weather Forecasting,". Journal of Theoretical and Applied Information, vol. 5 (5),2009.
5. Mr. C.P Shabariram, Dr. K.E.Kannammal, Mr. T. Manojpraphakar, "Rainfall analysis and rainstorm prediction using MapReduce Framework," Jan. 07 – 09 2016 International Conference on Computer Communication and Informatics (ICCCI) Coimbatore, INDIA, ISSN: \978-1-4673-6680-9/16/$31.00.
6. Dr. Bashirahamad Momin and Gaurav Chavan, "An Integrated approach for Weather Forecasting over Internet of T hings: A Brief Review". I-SMAC International conference (I-SMAC 2017)
7. Krishna, G. V. , "An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis. International Journal of Computer Applications", 2015.
8. Paras, S. M., "A Simple Weather Forecasting Model Using Mathematical Regression. Indian Research Journal of Extension Education", 2016.
9. www.accuweather.com
10. www.intellicast.com
11. weather.weatherbug.com
12. https://www.wunderground.com
13. https://www.yahoo.com/news/weather

Sentiment Analysis of WhatsApp Group Chat

Lakshya Munjal*
Anmol Arora**
Neetu Narwal***

Abstract

Sentiment Analysis has emerged as a field that has attracted a significant amount of attention in the last decade because of a wide variety of applications that could benefit from its results like customer preference of products, news analytics, marketing, share preference in stock market, social media (Twitter, Facebook). It helps a company to find out how their customers feel about their product by the feedback provided by them. After the introduction of Web 2.0, social media came into existence and now it has became a very important part of everyone life. Social Media gives us the opportunity to find our long lost friends who we haven't talked to after school. It also gives us the information that what our friends and family are doing. In this research paper, we are trying to analyze the behaviour of students from WhatsApp group chat. There are certain emotions which a human have like anger, joy, positivity, stress. The messages sent by students in the group chat can be used to find out their behaviour during a particular period of time, like 10 days before exams and post exams. We have used R programming language to generate bar graph and word cloud. The process that has been followed to find out the sentiments of a student is very simple. WhatsApp group chat is extracted from the phone and can be easily opened in computer system. Then this chat is divided into different columns namely Date, Time, Sender number, and Message. Frequency of the messages in the group shows how active students are during a specific period of time and it is easy to understand that there would be less messages during the exam period. Messages contain certain misspelled words, these words are ignored and all the emoji's are removed from the chat so that it can be easily used in R Language. All the blank lines are also removed because they are certainly of no use. R has certain Libraries which are used to generate these bar graphs and word cloud. The area of Sentiment Analysis is still in its early stages of development and there are situations when we cannot understand the sentiment of a person from the message he/she has sent.

Key Words: Sentiment Analysis, Whatsapp, Behaviour Pattern, Social Media, Web 2.0.

I. Introduction

Sentiment analysis sometimes known as opinion mining or emotion AI refers to the use of natural language processing, text analysis to systematically identify, extract, quantify, and study affective states and subjective information[1]. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation, affective state (that is to say,

Lakshya Munjal*
Student of BCA,
Department of Information Technology,
Maharaja Surajmal Institute, New Delhi

Anmol Arora**
Student of BCA,
Department of Information Technology,
Maharaja Surajmal Institute, New Delhi

Neetu Narwal***
Department of Information Technology,
Maharaja Surajmal Institute, New Delhi

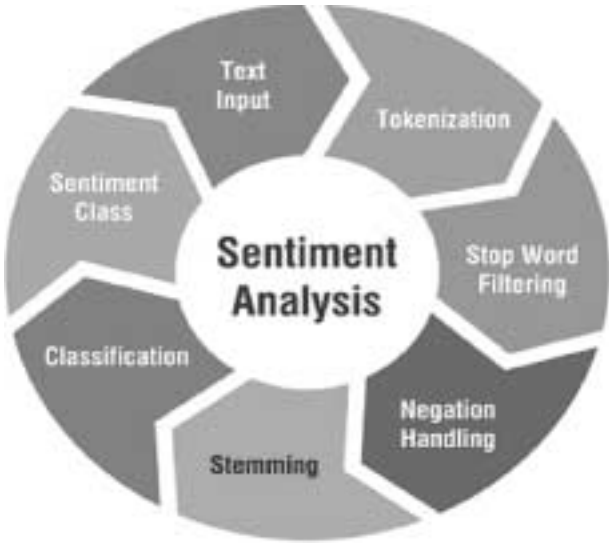


Figure-1: Sentiment Analysis process

the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry", "sad", and "happy".

The rise of social media such as blogs and social networks has fueled interest in sentimental analysis. With the rapid increase in the number of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and actioning it appropriately, many are now looking to the field of sentiment analysis. Further complicating the matter is the rise of anonymous social media platforms such as 4chan and Reddit.

The problem is that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. The fact that humans often disagree on the sentiment of text illustrates how big a task it is for computers to get this right. The shorter the string of text, the harder it becomes.

Even though short text strings might be a problem, sentiment analysis within microblogging has shown that Twitter can be seen as a valid online indicator of political sentiment. Tweets' political sentiment demonstrates close correspondence to parties' and politicians' political positions, indicating that the content of Twitter messages plausibly reflects the offline political landscape.

II. Related Research Work

EmmaHaddi [2] in his research work has performed sentiment analysis by focusing on some words and their frequency in the document. The terms that appear often in the whole document but seldom in a collection are more informative stating the meaning of the document in comparison to terms that just appeared

once. Also, the position of words in a document can completely change the meaning of a sentence.

Sakthivelet. al. [3] in their research work prepared a set of questions and half-structured interviews that were performed on the students and the teachers. Open questions helped the students and teachers to tell their stories in a flowing manner and there were some other findings that wouldn't have been discovered with closed end questions.

Bouhniket. al. [4] for their research work primary collection of data was done at places in the University where high number of students were around liking dining area, sports arena, libraries. Students were asked to complete a questionnaire. This research helped to find out the primary feature of WhatsApp that was used by students and how effective is WhatsApp to their studies and social life.

Tshwane [5] in their research work focused on how the social media has affected the daily life of students. A set of questions were asked and interviews were conducted to collect the data which was required for analyzing the effect on the social and academic performance of students. Some questions were semi-opened-questions and these questions had no restrictions how they should be answered providing the participants with opportunity to write what they couldn't have written in closed-end-questions.

Yeboah et. al. [6] their research is based on scientific citations. The information that has been collected for this research is from some other research papers that are related to the field and study. Data is also extracted from databases so that any key article is not missed.

[14]. Khatibet. al. [7] in their research paper gave the basic idea of a natural language based Android App Polarity analysis system for reviews which is capable of evaluating the reviews and provides result in appropriate form. First, the reviews are collected from google play store and then their polarity is identified as positive or negative by comparing them with the already present datasets. And, the total result is judged afterwards.

Dabhadeet. al. [8] in their research paper illustrated the research area of Sentiment Analysis on movie reviews or product reviews like android apps. Data is collected from various sources – review sites, blogs, forums or social networks. This data is analyzed using natural language processing system or machine learning approach after the data is filtered through a parser or spell checker.

Patni etl. al.[9] presented an interactive visualization system, SentiView, which analyzes public sentiments from text posted via media such as forums and predicts the short-term trend of the sentiments about events being discussed. Public opinions are studied and represented in relationship map consisting of comment points, relationship lines and interactions and then categorised into opposing, neutral or supporting sections.

III. Methodology Used

We used WhatsApp group chat for analyzing the sentiment of students during different period of time. We have tried to find out the behaviour of students prior to exams and during exam time also.

R Language is used for creating graphs and word cloud for the purpose of better understanding. The main focus while creating graphs was on adjectives in a particular sentence. Positive adjectives were given positive polarity and vice versa. Steps followed to creating graphs and word clouds are:

1. Document containing natural language is converted into input text to be given in the program; Corpus(VectorSource(filename))
2. All the punctuations in input are removed; tm_map(wordCorpus, removePunctuation)
3. Data is transformed into lower case letters; tm_map(wordCorpus, content_transformer (tolower))
4. Stop words are added into the data; tm_map(wordCorpus, removeWords, stopwords ("english"))
5. Extra white spaces are removed; tm_map (wordCorpus, stripWhitespace)
6. Unwanted words are filtered out; str_replace_all(string, pattern, replacement)

7. Presence of eight different emotions and their corresponding valence in the text file is calculated; get_nrc_sentiment(whatsapp_chat)
8. Tightly coupled collection of variables is created which is used as the fundamental data structure; data.frame(colSums(chat[,c(3:12)]))
9. Color palette is retrieved from RColorBrewer Color Ramp Palette; brewer.pal(7,"YlGnBu")
10. Word cloud is developed; wordcloud(words = wordCorpus, scale=c(5,0.1), max.words=100, random.order=FALSE, rot.per=0.35, use.r.layout=FALSE, colors=pal)
11. Graph is plotted; ggplot(data = sentimentTotals, aes(x = sentiment, y = count))

IV. Experiment and Results

The data that was extracted from WhatsApp group chat and it was filtered to remove any empty lines and emoticons. The adjectives were used for sentiment analysis and the frequency of messages was used to find out at what time of the day students are most active and the behaviour of students during Exam time. Messages were divided into 2 groups: prior to 8 pm and post 8 pm to analyze how active students are at that time.

Word clouds and bar graphs are generated using R Language. These data representation techniques are very interactive and easy to understand.

Results according to Number of Messages

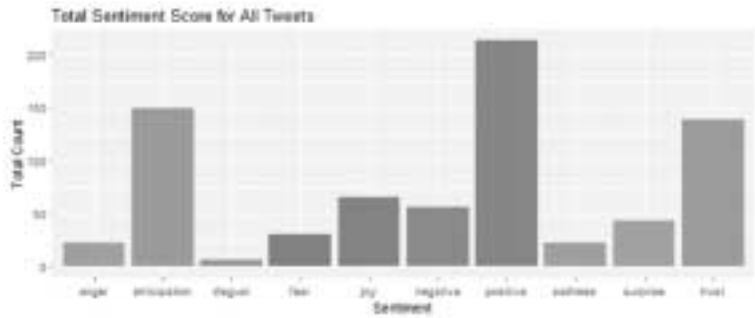


Figure-2: Sentiments Graph for Oct'16, Nov'16 and Apr'17

Figure-3: Word Cloud for Oct'16, Nov'16 and Apr'17

The following bar graph is for the period of October 2016, November 2016 and April 2017. During these months there are no exams. The above graph depicts that during the following 3 months, the behaviour of students is more positive and happy rather than being angry or sad.

Almost 64% of the messages that were exchanged during this period of time showed a positive behaviour. This can also be understood from the graph that the bar for Positive sentiment is way higher than that of Negative sentiment. It should also be noted that anger, disgust, fear, sadness are all Negative sentiments.

This word cloud depicts the words that were most frequently used in the conversation. This word cloud is for the period of October 2016, November 2016 and April 2017. The count of messages changes the size of a word cloud. Higher the number of messages, bigger is the word cloud. It can be easily understood that the main words are "maam" and "student".

Now as we move our focus to the month of December 2016 which is the month of examination we can see that the frequency of messages has drastically dropped and there is a sudden change in the behaviour of students due to exam stress.

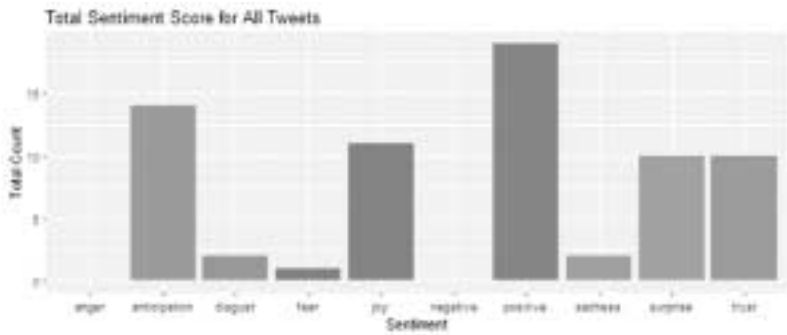


Figure-4: Sentiments Graph for Dec'16

Looking at this bar graph of December 2016, the attitude of students is positive most of the times. There were no negative messages and no feeling of anger. So the exam stress did work in the opposite manner that it works making all students positive and happy. 86% of the messages sent during the month of December 2016 were positive. The bars of disgust, fear and sad sentiment are so small that they did not had much impact on the positive sentiment.

Figure-5: Word cloud for Dec'16

A very small word cloud is generated for December 2016. What we can easily derive by looking at this word cloud is that the frequency of messages was very less and the focus of students here is mainly on the words "mam", "thank", "full" and "good". During exam time students thank their teachers when there queries are answered and as a result the word cloud has words mam and thank in big font size.

The sentiments of the group that are analyzed using R Language totally depends upon the frequency of messages. So in order to do a good comparison and good understanding of data, we generated a graph that shows the frequency of messages every month.

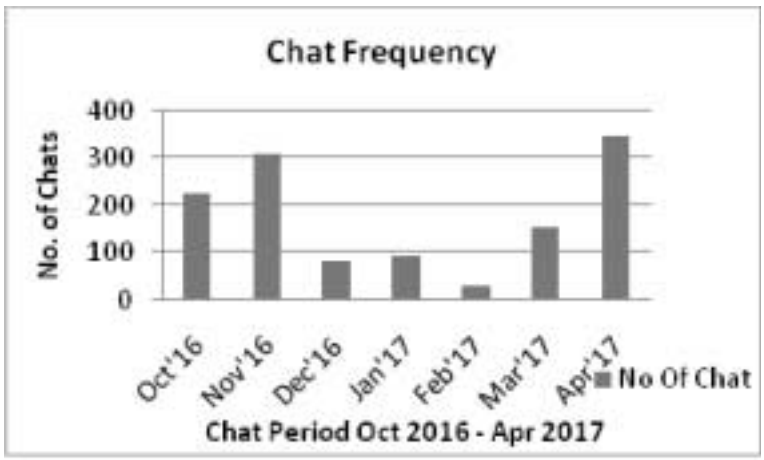


Figure-6: Message Frequency

Table 1: Message Count

Month	Message Count
October'16	228
November'16	308
December'16	82
January'17	96
Febuary'17	30
March'17	156
April'17	347

Months prior to examination, i.e., November 2016 and April 2017 have large number of messages exchanged. The reason could be there are practical's or assignment submissions during that time. January has less number of messages because the session starts from almost mid-January so there is not much interaction between the students and teachers.

Results according to Time

WhatsApp group chat is extracted from the phone and then that data is cleaned for making word clouds and graphs. Now we try to make a graph based on the time of the day when message was sent. Here the division is done based on what time of the day message was received. We have 2 categories: 10 am – 8 pm and post 8 pm.

Table-2: WhatsApp traffic

Month	10 am – 8 pm	8pm onwards	Total Messages sent
October'16	135	111	246
November'16	178	130	308
December'16	53	29	82
January'17	79	17	90
February'17	28	2	30
March'17	90	66	156
April'17	211	136	347

Table-3: Message frequency during day time v/s night time

Month	10am – 8pm (% messages)	8pm onwards (% messages)
October'16	55	45
November'16	58	42
December'16	65	35
January'17	82	18
February'17	93	7
March'17	58	42
April'17	61	39

Data for all the months was distributed and added to find out that every month how many messages are sent during day time and how many are sent during night time.

Depending on that bar graphs are generated.
Table 2 shows the number of messages that were sent during the day time every month and messages sent during night time.

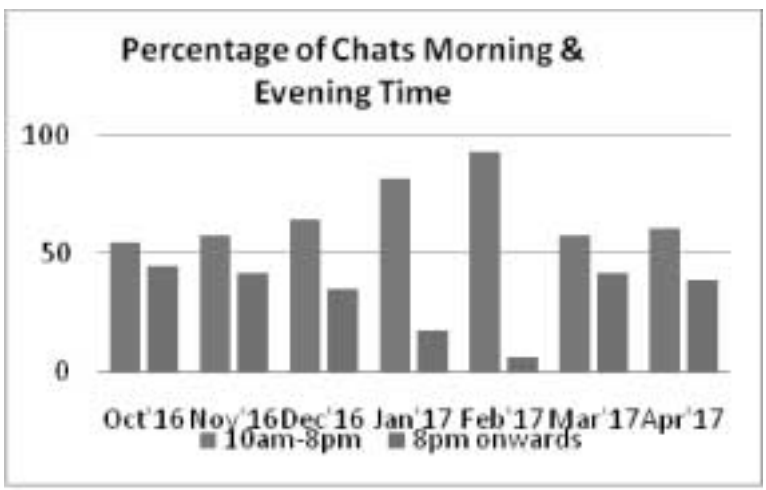


Figure-7: % messages graph

This data is now converted into percentage in order to make a meaningful bar graph. This bar graph will show that which time of the day is more dominant during every month.

Formula that will be used is:

Percentage of messages = (MC/TL) * 100

Where

MC -> message count during day time/ night time

TL -> total messages sent during that month

The following graph is generated from Table 3 data.

From the graph presented in Figure 7 it is observed that the frequency of messages is higher during the day time rather than being higher during night. The reason is that the group chat is of a class Official group that has all the teachers and students. Teachers sleep early at night so it is preferred to ask about queries during day time. Due to this reason the frequency of messages is higher at night.

V. Conclusion

It can be concluded that Social Media data can be used to provide sentiment analysis of users. This study focuses on whatsapp chat of College students and their behavior pattern during whole semester and towards examination time. In future work we will try predicting the behavior of students from their chats on the social media. It has been concluded that social media is a powerful and reliable source of information to know about person behavior.

References

1. https://lct-master.org/files/MullenSentimentCourseSlides.pdf
2. EmmaHaddi, Sentiment Analysis: Text-preprocessing, Reader Views And Cross Domains
3. M. Sakthivel, G. Hema, International Journal of Computer Science and Mobile Computing, Pg 231-239
4. Bouhnik, D., & Deshen, M. Journal of Information Technology Education: Research, Pg 217-231
5. Richard ShambareTshwane ,Journal of Economics and Behavioral Studies , Pg 542-549.
6. Johnson Yeboah (Lecturer)*,George Dominic Ewur (Lecturer), Journal of Education and Practice, Pg 157-163
7. K.C.Khatib, T.D. Kamble, B.R. Chendake, G.N. Sonavane, Social Media Mining for Sentiment Analysis, Pg 373,374,375
8. Shital S. Dabhade and Prof. Sonal S. Honale, International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 5, May 2015 pg. 123-129

Deep Learning: Powerful Metaphor to Understand World

Nishant Nirmal*
Suruchi Sinha**

Abstract

Deep Learning is an extensive and exhaustive version of neural network. It is a machine learning technique which helps in categorically defines unsupervised data and unclassified data. With the volume of data growing exponentially, deep learning is proving to be one of the finest techniques to characterize each the object and translate them into the real-time.

Key Words: Deep Learning, Neural Network, Unsupervised Datasets

I. Introduction

The inventors have always dreamed about creating a machine that could think. These desires date back to at least the time of ancient Greece. And when the programmable computing devices were first conceived, people again started wondering if they could become so intelligent that they might be able to develop their own thinking skills. Hundred years down the road, and man built one such machine. Today, Artificial Intelligence (AI) is a thriving field with many practical applications and active research topics[1].

In the early days of Artificial Intelligence, the field rapidly tackled and solved numerous problems that are intellectually difficult for human beings. But, the true challenge for AI proved to be the tasks that are easy for people to perform, problems that we solve intuitively, that feel automatic to us, like recognizing spoken words or faces in images.

What could be the solution to such problems? It is to allow the computers to learn from their own experiences. If they start getting knowledge from experience, the need for human operators to specify all the required knowledge that the computer needs will vanish. In other solutions, the computer also tries to understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts. This enables the computer to learn complicated concepts by building them out of simpler ones. If a graph is drawn showing how these two concepts of learning through experience and simplification of concepts are built on top of each other, the graph is deep, with many layers. Therefore, this approach to AI is called **Deep Learning**.

Nishant Nirmal*
Department of Information Technology,
Jamia Milia Islamia, New Delhi

Suruchi Sinha**
Department of Information Technology,
IITM, Janakpuri, New Delhi

II. Background

Broadly, AI is the computer-based exploration of methods for solving challenging tasks that have traditionally depended on people for solution. Such tasks include complex logical inference, diagnosis, and visual recognition, comprehension of natural language, game playing, explanation, and planning [2].

III. Motivation Behind Deep Learning

Deep Learning is a subset of subset of Artificial Intelligence (Machine Learning). Ironically, abstract and formal tasks that are among the most difficult mental undertakings for a human being are among the easiest for a computer. Computers have been able to defeat the best human chess players for a long time but only recently, they have begun matching some of

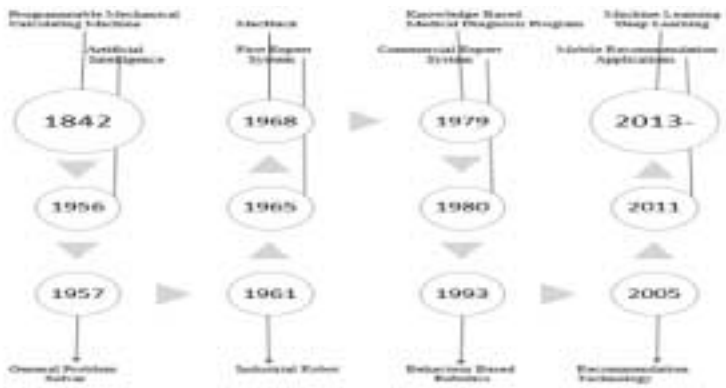


Fig. 1: Timeline of Artificial Intelligence

the abilities of average human beings to recognize objects or speech. A person's everyday life requires an immense amount of knowledge about the world. Much of this knowledge is subjective and intuitive, and therefore difficult to articulate in a formal way. Computers will also need to capture this same knowledge in order to behave in an intelligent way. One of the key challenges in Artificial Intelligence is how to get this informal knowledge into a computer.

Most of the Artificial Intelligence projects have sought to hard-code knowledge about the world in formal languages. A computer can reason automatically about statements in these formal languages using logical inference rules. This is known as the knowledge based approach to Artificial Intelligence. But there have been many difficulties in these systems relying on hard-code knowledge. We have always struggled to devise formal rules with enough complexity to accurately describe the world. Therefore, it was suggested that AI systems need to acquire their own knowledge, by extracting patterns from raw data. This capability is known as **Machine Learning**. The introduction of machine learning enabled computers to tackle problems involving knowledge of the world and make decisions that appear subjective.

The performance of the simple machine learning algorithms depends heavily on the representation of the data they are given. And this dependence is a general phenomenon that appears throughout computer science and even daily life. In computer science, operations such as searching a collection of data can proceed exponentially faster if the collection is structured and indexed intelligently. People can easily perform arithmetic on Arabic numerals but arithmetic on Roman numerals much more time consuming. It is not surprising that the choice of representation has an enormous effect on the performance of machine learning algorithms.

Deep Learning is also, a specific kind of machine learning. It is inspired by the structure of the human brain and is particularly effective in feature detection. This also involves feeding the system with large volumes of data. But since it is inspired by our brain, we first need to understand how the neural network works.

IV. Neural Networks and Working of Deep Learning:

A **Neural Network** passes data through interconnected layers of nodes, like the vast network of neurons in the brain, classifying information and characteristics of a layer before passing the results on to other nodes in subsequent layers. The difference between a neural network and a deep learning network is contingent on the number of layers: A basic neural network may have two to three layers; on the other hand a deep learning network may have dozens or hundreds of layers.

Deep learning achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones.

The most effective results of Deep Learning can be observed in feature detection, and it does so in the same way our brain intuitively detects different features. When we try to recognize a square from other shapes our brain first checks whether there are four lines associated with a figure or not. If it finds four lines, it further checks if they are connected, closed, perpendicular and that they are equal as well (Nested Hierarchy of Concept). So, we take a complex task and brake it in simple, less abstract tasks. This is exactly what Deep Learning does, but at a large scale [3].

Similarly, if we make a system that must recognize whether the given image is of a cat or a dog. If we try to solve this as a typical machine learning problem, we will define features such as if the animal has whiskers or not, if the animal has ears & if yes, then if they are pointed. In short, we will define the facial features and let the system identify which features are more important in classifying a particular animal.

Now, Deep Learning takes this one step ahead. It can automatically find out the features that are important for classification between a cat and a dog, whereas in Machine Learning we had to manually give the features.

- Deep Learning first identifies what are the edges that are most relevant to find out a Cat or a Dog.
- Further, it builds on a hierarchically structure to find out what are the combination of all the shapes and edges. For example, whether whiskers are present, or whether ears are present, etc.[5]
- After consecutive hierarchical identification of complex concepts, it then decides which of these features are responsible for finding the answer.

V. Conclusion

Over the last few years, Deep Learning has been applied to numerous problems, ranging from computer vision to natural language processing. In many cases Deep Learning outperformed previous work. Three major reasons behind the breakthrough of (deep) neural networks are:

i. The availability of huge amounts of training data.
ii. Powerful computational infrastructure.
iii. Advances in academia [4].

Since then, deep learning systems have started outperforming not only the classical methods, but also some of the human benchmarks in various tasks like image classification or face recognition. This has created the potential for many new applications leveraging deep learning to solve real-world problems, like:

- Medical Image Analysis
- Face Recognition
- Language Translation
- Prediction and Diagnosis
- Anomaly Detection and Security
- Speech Recognition
- Heuristic Classification
- Game Playing and many more.

References

1. Representation Learning: A Review and New Perspectives, Yoshua Bengio, Aaron Courville, Pascal Vincent, Arxiv, 2012.
2. The monograph or review paper Learning Deep Architectures for AI (Foundations & Trends in Machine Learning, 2009).
3. Deep Machine Learning – A New Frontier in Artificial Intelligence Research – a survey paper by Itamar Arel, Derek C. Rose, and Thomas P. Karnowski.
4. Graves, A. (2012). Supervised sequence labelling with recurrent neural networks(Vol. 385). Springer.
5. Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. 75 pages, 850+ references, http://arxiv.org/abs/1404.7828, PDF & LATEX source & complete public BIBTEX file under http://www.idsia.ch/~juergen/deep-learning-overview.html.
6. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

Transformation of Business from E-commerce to M-commerce

Sushma Malik*
Prashant Kumar**

Abstract

Now a day's mobile phone becomes common to common people, we can say that people are becoming "Mobile addict" and also it's available at very nominal and reasonable cost and that's why the number of users has drastically increased. Users also use the internet on mobile device. Although people have started using E –Commerce through mobile, so its call M-Commerce.

E-Commerce means, all forms of commercial transactions based on the electronic processing of data whereas in Mobile Commerce (M-Commerce), the buying and selling of goods and services through mobile phones means mobile equipment which is wireless handled device are used for buying and selling the goods and services of an organization. M-Commerce is based on the wireless application technology (WAP).

E-Commerce and M-Commerce are not having any boundaries, so they help organization to move from local to global market area. Number of companies and organization are doing business through E-Commerce and also move to new M-Commerce for business because major Indian population has adopted mobile phones with the advancement mobile technology. M-Commerce is the next generation of E-commerce which facilitate user to access internet without need of a place plug-in. The usage of mobile phones are not limited like for making call, receive messages, listing songs and videos but it also used for other approaches like vending, acquire and do many more other activities like get online information for traveling, online booking etc.

The main purpose of this research paper is that to make aware the readers of the current scenario and status of M-Commerce or mobile Commerce. Even many people have started using of E-Commerce but still they hesitate to use M-Commerce because of the some security issues electronic payment problems and some time the complexity of the mobile application etc.

Key Words: E-commerce, Mobile Commerce, Tools used in M-commerce

I. Introduction

Before explain the E-commerce and M-commerce first of all we understand, what is commerce? In a simple word, Commerce is Exchange of goods or services generally on a small and large scale from one place to another place or across the city, state, or even national boundaries. In traditional commerce user can purchase things easily but they face some problems like distance, availability of time and cost difference and not safe payment mode. But after changing technologies in business environments, users defeat these problems with the E-commerce.

E-Commerce or Electronics Commerce is a methodology of modern business. It can be broadly defined as the process of buying or selling of goods or services using an electronic medium such as the

Sushma Malik*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Prashant Kumar**
Department of Information Technology,
Banarasidas Chandiwala Institute of
Professional Studies,
Dwarka, New Delhi

Internet. E-commerce is using a website, which trade products or services directly from the website using a shopping cart or shopping basket system and allows user to pay through cards, e-banking, or even cash on delivery. Users can purchase anything right from a Insurance Policy to pen, by sitting comfortably in their workplace or residence and can be gift it to someone sitting miles apart just by click of a mouse on the website. It offers several benefits to businesses or users like easily reachable to a fast growing online communities, providing unlimited products to choose and services, helping to move from local area to global market and 24x7 time zone services at low operating costs of internet. Simplicity of internet access and navigation on the websites are the critical factors for the rapid adoption of E-commerce.

Mobile Commerce (M-commerce) is the subset of electronic-commerce (E-Commerce), which includes all e-commerce transactions, and these transactions are carried out by using a hand held device which is called mobile. M-commerce define exchange of goods over the internet by the use of mobile phones Or M-commerce, is about the explosion of applications and services that are becoming accessible from Internet-enabled mobile devices. M-commerce is nothing but use of mobile devices to complete the transactions.

The time and space limitation of commerce are removed and user can access it any time they need. It is a result of combining two strongly emerging trends: electronic commerce and mobile. It is online service from anywhere, anytime and on any device and is providing new business opportunities. In simple terms:

M-Commerce = E-Commerce + Wireless web (Through mobile)

Fig. 1: What is M-Commerce?

II. Features of M-Commerce

The following are the features of Mobile commerce, making it one of the effective ways of doing Commerce:

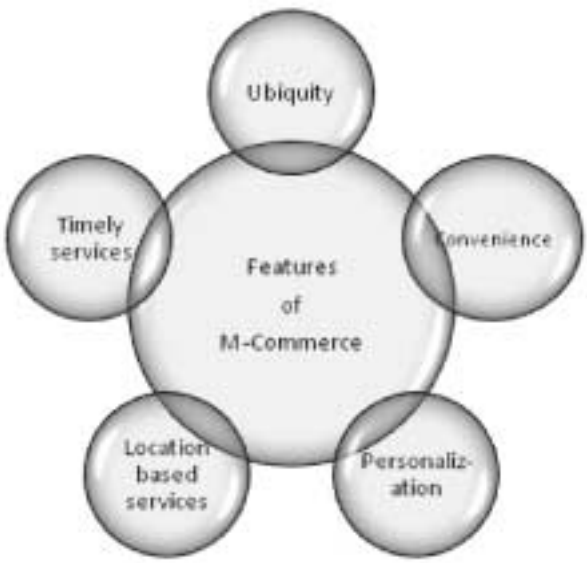


Fig. 2: Features of M-Commerce

1. **Ubiquity**: Customers can buy products or services from anywhere independent of their current geographic location.
2. **Convenience**: Companies are connected with their customers through GPRS Services.
3. **Personalization:** Customers can personalize the M-Commerce sites as per their requirements.
4. **Location based services**: GPS allows companies to propose goods and services as per the location based of the customers.
5. **Timely services**: M-Commerce provides real time availability of services because customers are directly connected.

Table I: M-Commerce v/s E-Commerce

DOMAIN	E-Commerce	M-Commerce
Definition	E-commerce refers to all the activities of buying and selling products and services with the use of electronic systems on the internet.	M-commerce refers to the process of buying and selling products and services with the use of internet /cellular data through wireless handheld devices.
History	It introduced in 1970's.	It comes into actions in 1990's.
Devices used	PC, Laptops.	Hand held Devices that is Smart phones.
Connectivity	Smaller	Larger owing to the bigger number of mobile users
Mobility	Limited	Less limited because of lighter weight and smaller size leading to easier to carry
Reach	Only at the places where the electricity and the internet are available	Broader due to its portability
Security	Less Secure as compare to M-Commerce.	It more secure because customers have authentication numbers.
Convenient	E-Commerce is less convenient because not easy to carry laptop or PC.	M-Commerce is more convenient because smart phones are easy to carry.
Platform used	Web Stores	Web stores (mobile version/mobile app)
Usage	More complicated user interface and more functions available.	Simple usage because all the functions are simplified.

III. Tools for Mobile Commerce

In present market, number of mobile tools available for the companies to engage customer in business, the following are some of the technologies that are help in reaching the customers:

1. **SMS (Short Message Service):** SMS is short messaging service which consists of 160 characters in black and white. It is the generally used and cheapest form of mobile marketing. Companies can send bulk messages to the customers.
2. **MMS (Multi-media Message Service):** MMS contains time slide show of images, text, audio and video.
3. **Mobile web applications:** customers can access the web pages of the companies through mobile phones or content of other web pages and they can do business.
4. **Wireless Network (Bluetooth/Wi-Fi):** wireless network tool is used by retailers, mall owners and small business holders because their range is small.

Fig. 3: Tools for Mobile Commerce

These tools are user to circulate the massage within a 10 meter range.

5. **Location based Marketing:** this help to locate the customer's location through global positioning system (GPS) technology and companies used this tool to deliver the offers related contents to the customers to their particular location.
6. **Voice:** many companies used voice based marketing. Interactive Voice Response (IVR) is a very popular tool for that. An IVR is a programmed telephony system that interacts with customers to gathers information. It accepts a combination of voice telephone input and touch-tone keypad selection and provides appropriate responses in the form of voice, fax, call-back, e-mail and perhaps other media. It consists of mobile device, software applications, a database and a supporting infrastructure. Interactive voice response (IVR) is very popular and companies use to offer various services to their customers. A customer has to follow the instructions recorded on the system at the other end.

IV. Applications and Uses of M-Commerce

Today's large number of people use the mobile phones to access the websites as compare to computers. That's why M-commerce has the big chance to connect big business as well as small business to customers on a large scale. In this sense, mobile phones act as a digital tool that allows organizations and customers to reach to one another more easily than the earlier modes. That's why M-Commerce has gained major popularity in various sector of the society in the last few years. Some of its applications are as follows:

1. **Money Transfer:** Now a day's mobile become the small banks through which user can transfer the money from one person to other person by using the mobile apps like various mobile wallets.
2. **Mobile Ticketing:** mobile ticketing is the process through which customers can order, pay for, get and validate tickets at any time by using the mobile phones. This process can be done with the help of simple downloaded application in mobile. The confirmation of tickets booking and electronic tickets are send to customer's mobile phones through SMS or Email. It is used in various areas like Airline ticketing, Movie ticketing, Railway and Bus ticketing, Concert/ Event ticketing.
3. **Mobile Banking:** Now a day's Banks allowed to customers to access their account and use it to

complete transaction such as online payment for purchasing products, receive notifications, and transfer money from account to other banks. It also includes:

- Mini statements and checking of account history.
- Checking the balance of the account.
- PIN provision, change of PIN and reminder over the internet.
- Cash in Cash out transactions on an ATM.

4. **Education:** Education sector is also affected by the M-Commerce. There are various educational apps are available through which user can access the number of contents, books and journals which helps the readers and students to access different types of information for their research from different places by sitting at one corner.
5. **Entertainment:** Entertainment has always played a vital role in Internet application and is probably the most popular application for the younger generation. Mobile commerce apps allow user to download games, images, music and video files at anytime and anywhere, and it also makes on-line games.
6. **M-Shopping:** Mobile shopping is another flourishing sector which came into existence after the successful introduction and growth of online shopping. Various mobile apps are available which help in shopping on mobile and mobile is easy to port/carry from one place to another so customer prefer it more than computers.
7. **Health Care:** The cost of health care is high and M-Commerce can help to reduce it. By using the technology of M-Commerce, general practitioner and nurses can remotely access and update patient records immediately which are time consuming process in the past. This improves efficiency and productivity and also reduce administrative overheads, and improve overall service quality of this sector. Various applications of this sector is also available on the mobile platform which help the customer to search the best doctor in their area and also fix the appointment with the doctor.
8. **Inventory Tracking and Dispatching:** On time delivery is critical for the success of today's businesses. M-Commerce allow industries to keep track of its mobile inventory and make on time deliveries of the products to the customers.
9. **Mobile Brokerage:** Stock market services offered through mobile devices also become more popular and are known as Mobile Brokerage. For that user can use the mobile apps and it allow the user to react on market expansion in a timely basis and introspective of their physical location.
10. **Traffic:** The users who use mobile phones are the ideal clients of the M-Commerce. Traffic control is usually a major headache for many metropolitan areas. Using the GPS technology of mobile can easily improve the flow of traffic in many ways. For example, determining the driver's exact position, giving directions on driving , and advising on the current status of traffic in the area.

V. Problems and Constraints of M-Commerce

1. **Number of Customers is not technology savvy:** Many customers will not be using mobile phones for transaction especially in the rural areas.
2. **Most of the merchant do not have updated technology:** Most of the vendors do not have advanced technology which can perform transactions with mobile phones.
3. **Many products cannot be purchased:** Many products cannot be purchased through mobile. So it plays a limited role that to in high end products.
4. **New method of commerce for many customers:** In rural areas, M-Commerce is a new concept of commerce and also in urban areas this is new phenomenon which takes time to catch by customers.
5. **Delay in Reverse transaction:** In mobile commerce if we return the purchases, consumers

are confused how to reverse the order and to get the money. It is also a time consuming process.

6. **Concerns of Security, privacy and reliability:** M-Commerce sites are less secure and less reliable for online payments through credit or debit cards.
7. **Limited speed:** Mobile internet has limited speed as compare to computer systems. So some times M-Commerce sites are not work properly on mobile.
8. **Storage capacity:** Storage capacity of mobile has limited to store important information.

VI. Benefits of M-commerce in India

Now a days, M-Commerce market grows rapidly because the number of mobile increasing day by day. M-Commerce facilitate works very easily and smoothly in the customers busy schedule. It some benefits are as follows:

1. **User friendly:** Today's websites are designed so much user friendly. M-Commerce sites use different classification of products variety which helps customers to choose needy product from the numbers of products. Website predicts option helps customers to find the product more easily and rapidly.
2. **Easy to bear mobile device:** Mobile is a hand held device. So user can easily carry the mobile with them any where customers wants.
3. **Low internet connectivity area:** M-commerce websites are efficiently used where the internet connectivity is less and website is taking more time to upload. For opening sites on mobile required less internet data as compared to opening of sites on laptop or PC.
4. **Secure transactions:** For doing the secure transactions, confirmation code are sent on the E-Mail or mobile phones. M-commerce also gives an assurance of secure transactions by after filling this code the transaction will be processed, So that chances of wrong transactions are very less and unsecure transitions can be eliminated.
5. **Convenience:** Mobile device is very convenient for opening of M-Commerce sites and very easy to handle. Mobile device is easy to carry by customers; it helps customers to avoid 0going to cyber cafes to access the internet.
6. **Saves time:** To complete the M-Commerce transaction customer not needs to plug in his/her laptop or PC and wait for the system to reload. But with the mobile device customer can complete the transaction just hit a single button on M-Commerce sites. So it's save most of the customer time.
7. **Electronic Wallet**: A mobile payment is one of

the main reasons for the success of m-commerce, because mobile device act as an electronic wallet for mobile payments through mobile payment applications.

8. **Flexibility accessibility:** Customers are flexible to access the M-Commerce sites as well as at the same time can be login on to various mobile messengers like Facebook, Twitter, Gmail and other networking platforms

VII. Drawback of M-Commerce in India

Each coin has two sides. Same M-commerce also has some drawbacks over its benefits, which are as follows:-

1. **Lack of Internet Connectivity:** Number of places in India, in which access of internet is not available. Also3G networks are not available at so many states.
2. **Language Barrier:** Mostly customers are not aware about the English language. So that for the transaction over internet through mobile devices, language becomes one of the major factors to purchases, hire and sell a particular product or services.
3. **Less Graphic Resolutions:** Mobile devices have less resolution as compare to computer/laptop, for that reason consumers are least interested to buy a particular product because product is not properly or exactly shown in mobile devices as compare to computers /laptops/notebooks.
4. **Lack of Awareness:** Customers are not aware about the term M-Commerce. And still afraid to adopt to purchase belongings online as well as they are feeling uncomfortable to buy a product through M-commerce.
5. **Less Number of Smart phone users:** Customers are not using the smart phones in the remote areas because of less internet connectivity.
6. **Less Functionality:** Mobile Phone present less functions as compared to Laptops. Functionality is limited for mobile phones.
7. **Security issues:** Customers have some security issues while making the payment. Some time speed of internet does not allow customers to make the payments efficiently. Also the threat of hacking, phishing is always there as customers don't have security software available in their phones. While customer using apps to complete the transaction, their identity and personal detail and bank credentials are used by the app store.
8. **ROPO- Research Online and Purchase Offline:** Most of the people have ROPO tendency. Numbers of customers still prefer viewing products and comparing online but they still using offline method to purchase products.

9. **Complex mobile applications:** Mobile applications of M-Commerce are available but still they are complex for the customers as compare to E-Commerce websites.

VIII. Findings

After the study, we include that there is lots of work which is to be done for M-Commerce in India. We know M-Commerce is the new industry in India and to run this type of industries, internet connectivity and mobile networking are the necessary requirements. But still in some areas of India are lacked internet connectivity and mobile networking. Some population is still not believed on the functioning of the M-Commerce. These are the main reasons for the slow rate of growth of M-Commerce in India.

But M-Commerce market grows in India because so many growth factors such as new generation's perception towards M-Commerce, growth of financial area and increasing the numbers of smart phone buyers are favorable for their growth. But still it is a small single step in the long journey of developing M-Commerce in the nation.

IX. Suggestions

M-Commerce industry run well in the India, but still it need some positive steps for the proper growth of M-Commerce in India. There should be taken essential steps to include these growth factor such as wireless network technology, building of infrastructure to internet connectivity, gives proper awareness of internet, aware more and more population about the uses and security of Ì-commerce applications by the Government of India.

Following are the some weak area's where government should take necessary action for improvement:

- Affordability of mobile devices
- Mobile internet connectivity
- Mobile payments
- Security
- Low tariffs-high revenue
- Proper Government policies

X. Conclusion

Mobile Commerce is well-known as M-Commerce. M-Commerce are used to purchase and trade goods by using the wireless devices like smart phones that have operated with Internet access. The easiness mode of communication with merchandise and service presenters, in 24 hrs services and without any location barriers are the main reason for an everyday popularity increase of M-Commerce. Both the telecommunications industry and the business world are starting to see M-Commerce as a major focus for the future.

This research paper shows that M-Commerce make its position in the Indian Market with its challenges and issues. The main reason for its usage is the affordability of smart phones. These phones are not allow to user to make audio and video call but user can now access every feature of business from viewing the product, giving the order of the product making the payments and getting their product at the home door in just few clicks. Services are not limited, user can also use M-Commerce services like online ticket booking, user can easily manage their bank accounts transfer the money from one account to other accounts in a just few minutes. As every coin has two aspects so M-Commerce also may have drawback like security issues due to wireless connections, lack of awareness, less usage in rural areas where no transportation and no internet connectivity.

M-Commerce has gain a huge popularity in its initial phase in India. Mobile commerce plays important role in the human life. We can say that M-Commerce is the next generation mode of business and its future seems to be extremely safe in India.

References

1. Rajabion Lila, "Critical Factors for Adoption of Mobile Commerce Services", International Journal of Business and Social Science Vol. 6, No. 12; December 2015.
2. K Panneerselvam Murugadoss, "Mobile Commerce – A Mode Of Modern Business", Asia Pacific Journal of Marketing & Management Review, ISSN 2319-2836, Vol.2 (7), July (2013).

Security Issues and Challenges while Outsourcing Computing Resources for E-Business

Prabhat Kumar Vishwakarma*
Geetali Banerji**

Abstract

Since last couple of years, cloud computing has been the key reason for the growth of IT-enabled businesses. Cloud Computing may now be envisioned as the next generation computing.The lucrative features behind the wide applicability of Cloud Computing are –To facilitate quick start up, low set up cost, flexibility and scalability. The purpose of this paper is to analyze its functional framework which forms basis for its suitability for E-business growth and development and further to diagnose its security breaches and their counter measures.

Key Words: APT, DoS, API, UI, SaaS, PaaS, IaaS.

I. Introduction

Remote accessibility of Hardware and software services without thinking so much about its physical infrastructure establishment is the main reason of the wide popularity of the cloud computing.. Based on the services provided by cloud computing they are classified as –

IaaS- Infrastructure as a Service.
PaaS-Plateform as a Service
SaaS-Software as a Service

Due to location flexibility, fast query processing, high availability, ability to store huge amount of data and most importantly low operational and maintenance cost are the factors which focuses the orientation of Organizations more towards business growth rather than thinking towards its technical side.

Deployment Models of Cloud Infrastructure

The deployment models of cloud infrastructure are – private cloud, community cloud, public cloud and hybrid cloud. The major functional properties of the deployment models are mentioned herewith in the following diagram.

For business promotion and awareness creation among public obviously the favourite choice will be through the deployment of public cloud. But in public cloud customer does not have control over his data, and access to this data can be granted inadvertently to entrusted parties.

The main obstacle to the wide acceptance of cloud storage, even with all advantages of public cloud infrastructure, is concern over the data confidentiality and integrity and possibility of access by untrusted users. If cloud service providers can provide data security guarantees, the customer will be assured of data safety on the cloud from internal and external threats.

Prabhat Kumar Vishwakarma*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Geetali Banerji**
Department of Information Technology,
IITM, Janakpuri, New Delhi

II. Literature Review

Right decision at right time is the key factor for the economic growth of any business. Therefore the role of IT in business in dynamic era cannot be overlooked. Whether to setup IT infrastructure for the business or to outsource the computing resources for the business is now the challenging issue. In the growing competition, financial success is important, because it is foundation of business survival.In dynamic competitve era, low investment and maximization of

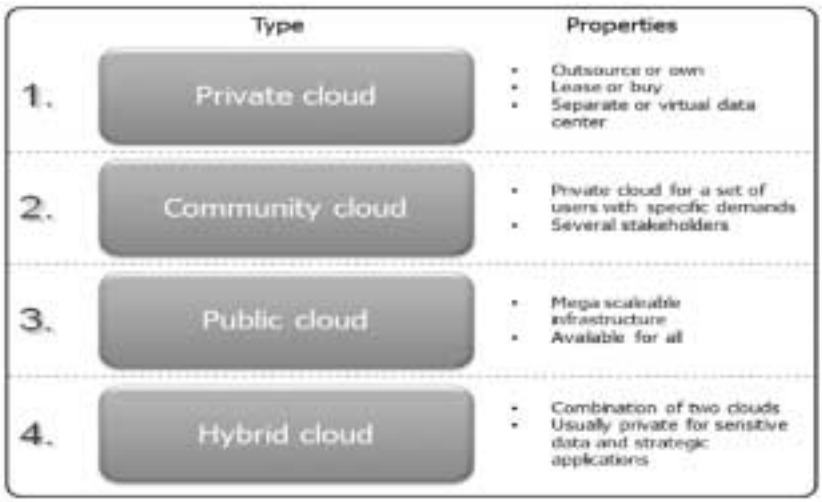


Figure-1: Deployment Model

ROI is now the core expectation in business.Proper resource utilization is one of the positive indicator for maximizing ROI and subsequently Revenue. Fully relying on technology without compromising the secracy of the business policy will lead to heavy initial investment and rapid change in technology will reinforce quick updation /modification in the technology infrastructure of the business.this is also not the smart business policy in current scenario. Cloud based resource utilization is now the necessity in business.The increasing business turnover yearwise for IaaS and PaaS is depicted in figure-2(Source: IDC)

As per NIST the most acceptable definition of cloud computing is that "Cloud computing is a model for enabling 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. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models." The Cloud computing model NIST defined has three services.

Models and four deployment models. The three service models, also called SPI model, are: Cloud Software as Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS). The four deployment models are: Private cloud, Community cloud, and Public cloud and Hybrid cloud. Compared

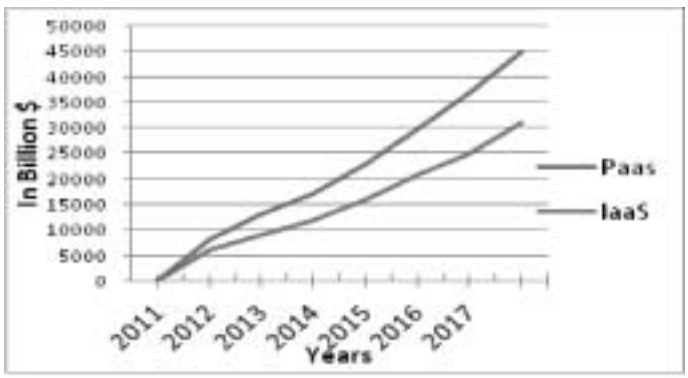


Figure-2: Comparative financial growth of PaaS and IaaS

with the traditional IT model, the cloud computing has many potential advantages. But from the consumers' perspective, cloud computing security concerns remain a major barrier for the adoption of cloud computing.

It is very common practice in any business that during peak load the resources will be optimally utilized but during off load the resource will drastically be under utilized.using cloud computing ,resource sharing can easily be well synchronized with load which will be based on category of cloud i.e. public cloud, community cloud,private cloud or hybrid cloud. At the infrastructure level, resource sharing means supporting many platforms and applications on the same physical resources, using the virtualization techniques of IaaS .At the platform and application levels, providers can achieve resource sharing through multi-tenancy. For PaaS, this means applications from many clients running on the same operating system. For SaaS, it means clients sharing the same application instance. By doing all these means and, using specially-designed application architectures, one can achieve better utilization of the underlying assets than if each client uses either a separate application instance on the same operating system or another virtualized instance sharing hardware. The economical, scalable, expedient, ubiquitous, and on-demand access to shared resources are some of the characteristics of the cloud that have resulted in shifting the business processes to the cloud.

III. Threats in Cloud Computing Environment

While working in cloud computing zone it is quite common to get leverage of the technologies related with web services, web browsers as well as virtualization. Therefore the technical issues related with these technologies will get inherited into colud computing infrastructure. It is also required here to undersatnd the concept of virtualization. The concept of virtualization is helpful in creating, managing and sharing the virtual web resouces in same way as it is being done for managing real web and computing resources.This ideology has been smartly used in distributed computing environment. A lot of applications in cloud computing are based on the virtualization concept therefore to secure the virtual resources in cloud computing is equally important as to secure the real web resources. Virtual machines have two boundaries viz. **virtual** and **physical**. Due to interconnection complexity, the created loop holes are the major reasons of high level of vulnerability and high risk of DoS attack.

Inspite of several advantages, the cloud computing persists serious data security issues. If the security issues of cloud computing are not controlled and monitored in an organized way, it will adversly affect the business. In cloud computing environment, initial interaction is through API and User Interfaces therefore they are the driving forces of cloud computing. Cloud API's IP address is vulnerable and unsafe between the cloud and user.The cloud may be the major target to the hackers as it contains high valued centralized data store and its centralised access. A lot of data damages are by the malicious insiders. The insiders may be current or former employees, contractors or parteners having the previledge of accessing the data.

Advaced Persistant Threats (APT) are also desiged and modelled for long term cyberattck by the hackers by just providing the leverage of ongoing access of the networks at initial level and futher exploited as phising attack.

Shared technology issue in cloud computing is also one of the big issue as cloud provider provides services to several tenants so there is the need of multifactor authentication.

As availability of computing resources is needed to multitenants therefore it may be on the risk of DoS.

IV. Countermeasuring Security in Cloud Based Computing

First and foremost requirement for countermeasuring security is the need of following guidelines of CSA which better explains about the issue related with the leakage of customer data while using cloud computing infrastructure based on virtual netwok.

Upto certain extent the architecture ontology approach explained in several research papers is better option to counterfeit the security issues realted with cloud based

computing.It provides a controlled mechanism for secured storage, API and access management.

Proxy based architecture are also proposed in some researrch papers to control the SQL Injection Attack.For handling cross site scripting attack active content filtering technique is a better solution.

ARP (Address Resolution Protocol) and RTT (Round Trip Time) based sniffing detection tools can be easily used to counter measure the network sniffing attack.On the basis of information exchange, cooperative and orchestrated association among various Intrusion Detection Systems, the DoS attack can be prevented.

V. Conclusion

No doubt to say that Cloud based Computing is next generation IT-Enabled Computing Technology. And It's application has a satisfactory outcome where there is a huge initial investement requirement in the Informtion Technology area.But it's applicability into the E-Commerce where sensitive information such as credit card, debit card details are needed in the cloud based infrastructure are suspicious. Authetication, accessing and revealing the confidential data is now the primary concern in Cloud computing. To hire and use of computing resources is cost effective but there is no surity that this data cannot be misused by third party consulting Cloud Computing Service providers. At the very first level it seems that while using cloud based services and even sharing confidential information if consumer is further not intersted to continue it then due to lack of clear legal cyber act the probability of being exploitation,blackmailing of cosumer based on sensitive data by service providers or by third party consumers associated with service providers , may increase.In fact there is a need of strict policy formulation and implementation on legal aspects about violation of data security or misuse of personal data resources available with cloud service providers, no matter whether consumer is further interested or not interseted to continue the cloud based services. Especially in legal context in Indian scenario the information technology act, cyber laws needs to be more strengthen in terms of misusing personal and confidential online resources available with cloud based computing environment. Otherwise Technology based, cost effective and transparent governance based on cloud computing will have no such remarkable impact in business growth as it is being expected from it.

References

1. Hasan Omar Al-Sakran,"Accessing Secured Data in Cloud Computing Environment",International Journal of Network Security & Its Applications (IJNSA) Vol.7, No.1, January 2015.
2. Mohammed J. Novel ,"Approaches to Big Data Management" International Journal of Computer Science and Information Technology Vol. 3, Issue 1January 2015 , pages 96-105
3. Subashini S, Kavitha V "A survey on Security issues in service delivery models of Cloud Computing". J Network Computing and Applications 34(1) (2011) pages 1–11.
4. D. Gollmann, "Securing Web Applications", Information Security Technical Report, vol. 13, issue. 1, 2008
5. R. Gellman, "Privacy in the Clouds:Risks to Privacy and Confidentiality from Cloud Computing," 2009.
6. Cloud computing use case discussion group, "Cloud Computing Use Cases: White Paper" available from: http://cloudusecases.org, 2012
7. Wang, W., Li, Z., Owens, R. & B. Bhargava, "Secure and Efficient Access to Outsourced Data", in Proc. of ACM Cloud Computing Security Workshop, 2009 pp. 55-65.
8. Yang, Ching-Nung & Jia-Bin Lai, "Protecting Data Privacy and Security for Cloud Computing Based on Secret Sharing", International Symposium on Biometrics and Security Technologies (ISBAST),(2013 pp 259 – 266.
9. Yang , Kan; Jia, Xiaohua; Ren, Kui & Bo Zhang,) "DAC-MACS: Effective data access control for multi-authority cloud storage systems", INFOCOM, Proceedings IEEE , 2013 pp 2895 – 2903
10. https://www.zetta.net
11. http: //www.opengroup.org/cloud/ cloud_for_business /p7 .htm

Optimizing Operating Time using Clustering Technology in Wireless Sensor Networks

Vishakha Verma*
Ankit Verma**

Abstract

Limited Energy resource of the batteries of the sensors is a challenging constraint in the design of wireless sensor. The operating time of WSN utilized in the applications is restricted by the limited resources. As data communication disperses the energy resource of the network the major role is played by the routing algorithm in the energy efficiency of WSN. The delivery of data from sensors to the base stations is done by the energy efficient cluster-based routing algorithm. These algorithms are heuristic. Heuristic algorithms do not assure about the optimal solution but still we use them because they are simple and are used for optimization of large sensor networks. This article offers a model to accomplish the optimal solution for cluster-based routing protocols in WSN.

Key Words: Sensor networks, Routing, Cluster networks, Battery, Linear programming, Optimization

I. Introduction

If we consider energy efficiency in WSN, the common problem is sending maximum data from all the sensor nodes to the base station until the first sensor node runs out of the battery. In this routing, clustering is a process that divides the network into interconnected sub-structures called clusters [1]. Each cluster has a cluster-head as a coordinator within the sub-structure which acts as a medium for data transfer between the nodes [2]. The Cluster heads communicate with each other and send data to the Base Station. Figure 1 shows the routing based on clusters where cluster heads are appointed in network.

These algorithms are heuristic. Heuristic algorithms do not assure about the optimal solution but still we use them because they are simple and are used for optimization of large sensor networks [3]. This article offers a model to accomplish the optimal solution for cluster-based routing protocols in WSN [4]. Problem is solved as integer linear programming problem and optimal solutions are computed using ILP Solvers [5]. The performance of heuristic algorithms and optimum solution for system lifetime problem is evaluated using these solutions.

The research is organized in the following manner: Previous work done in energy efficiency using cluster-based routing then a clustering-based routing model is developed and implemented in a simple network of one cluster and then multiple clusters and at last the results of models are presented [6].

Vishakha Verma*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Ankit Verma**
Department of Information Technology,
IITM, Janakpuri, New Delhi

II. Energy Efficiency Using Cluster Based Routing

High energy nodes are responsible for gathering data from the low energy nodes and then they perform data aggregation before sending data to the base station [8]. Different nodes are grouped together to form clusters and the node having high remaining energy are nominated as Cluster Head. In every cluster, the nominated Cluster Head receives the data and aggregate data from all the sensor nodes in the cluster [9]. The data size of all the sensor nodes is same. The size of aggregated data done at Cluster Head is same as the data size of every sensor node in the cluster [10].

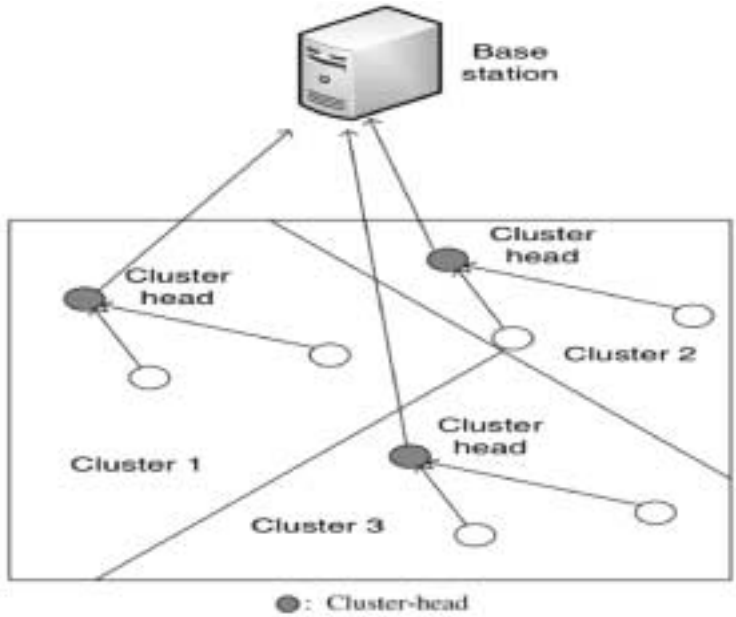


Fig. 1: Routing Based on Clusters

Data aggregation is done before sending the data to the Base station. Data Aggregation has an advantage that it reduces the amount of the information sent to the base station which in turns saves energy [11]. Figure 2 shows that node N1 require more energy to reach cluster head CH2 than CH1.

In the sensor applications, Sensor nodes send data periodically to the Base Station [12]. One round of data transmission is equivalent to the duration of time to send a unit of data to the BS. Energy which is used to send a unit of data to the base station is lost by the sensor nodes at the end of each round of data transmission. Total number of rounds which are used for sending data to the Base Stations until the first nodes is known as the lifetime of the sensor Networks.

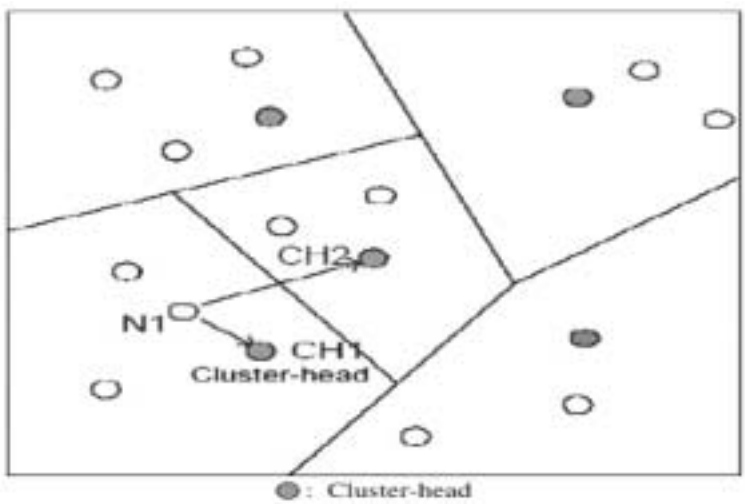


Fig. 2: Energy from Node to Cluster Head

III. Low-Energy Adaptive Clustering Hierarchy

LEACH which is Low energy adaptive clustering hierarchy is a TDMA-based MAC Protocol which is integrated with clustering. The principle of LEACH is to lower the energy consumption which is required to create and maintain clusters in order to improve the life time of a wireless sensor network. Network is made up of nodes some of which are called Cluster heads. The task of cluster head is to collect data from their surrounding nodes and pass it on the base station. LEACH is dynamic because the job of cluster-head rotates [17]. Nodes that have been cluster head cannot become cluster head again for P rounds, where P is the desired percentage of cluster heads. Each node has a 1/P Probability of becoming cluster head again [16]. The nodes which are located at a large distance from Base Station will consume more energy in comparison to those nodes which are located near to base station while sending data to the Base Station. After every round of transmission Residual Energy is calculated. A code is transmitting by all the nodes and contains information about the Residual energy and their Identification [18]. If the calculated Residual Energy of a node is greater than the Residual energy of all the other nodes then that particular node having Residual Energy greater elected as a Cluster Head. The operation of LEACH consists of several rounds with two phases in each: Setup phase and Steady phase.

In the first step, the Cluster Head sends the Advertisement Packets to all the nodes and inform Cluster nodes that they have become a cluster head [19]. For balancing the energy consumption, the node which already elected as cluster head cannot be the cluster head again until all the nodes in the cluster have become cluster head once.

In the second step, The Advertisement Packets sent by the Cluster head is received by the Non-Cluster head [13]. The Non-Cluster head send the joint request to the cluster head informing that they are members of the cluster under that particular cluster head.

In the third step, Transmission Schedule is created by the Cluster Head for the other nodes of the Cluster. In accordance to number of the nodes in the cluster, TDMA Schedule is created [14]. The member nodes of the cluster transmit the data according to the time schedule.

Second Phase of LEACH is known as Steady Phase. Cluster Head receives all the data from the Cluster Node [15]. Data aggregation is done by the Cluster Head and then it forwards the data to the Base Station either directly or through other Cluster Head along with the Static Route defined in the Source Code. After a time, limit, Network goes back to the Setup Phase.

IV. Conclusion

Routing protocol based on clustering is shown in this paper. Sensor node never works on individual data but it takes list of activities and work with all nodes. As handling of huge sensors are typical so they are divided into clusters and cluster heads are assigned. Cluster head collect information for their clusters and pass to base station. Selecting cluster is not easy. Battery life and energy of nodes are compared to select it as cluster head. LEACH is the low energy adaptive clustering hierarchy, which is used for solving problems related to wireless sensor networks.

References

1. WB Heinzelman, AP Chandrakasan, H Balakrishnan, Energy-efficientcommunication protocol for wireless microsensor networks, in 33rd HawaiiInternational Conference Systems Sciences, 2000, p. 10. Vol.2. ISBN: 0-7695-0493-0.
2. WB Heinzelman, AP Chandrakasan, H Balakrishnan, An application specificprotocol architecture for wireless microsensor networks. IEEE Trans. Wirel.Commun 1(4), 660–670 (2002).
3. I Saha Misra, S Dolui, A Das, Enhanced-efficient adaptive clustering protocolfor distributed sensor networks, in IEEE ICON, 2005, p. 6. doi:10.1109/ICON.2005.1635525. Vol.1, ISBN:1-4244-0000-7.

4. J Zhang, B Huang, L Tu, F Zhang, A cluster-based energy-efficient schemefor sensor networks, in The Sixth International Conference on Parallel andDistributed Computing, Applications and Technologies (PDCAT'05), 2005, pp.191–195. doi:10.1109/PDCAT.2005.3. ISBN: 0-7695-2405-2.
5. R Chang, C Kuo, An energy efficient routing mechanism for wireless sensornetworks, in The 20th International Conference on Advanced InformationNetworking and Applications (AINA'06), 2006. Vol.2, p. 5. doi:10.1109/AINA.2006.86. ISSN: 1550-445X; ISBN: 0-7695-2466-4.
6. JN Al-Karaki, AE Kamal, Routing techniques in wireless sensor networks: asurvey. IEEE Wirel. Commun. 11(6), 6–28 (2004).
7. J Deng, YS Han, W Heinzelman, Balanced-energy sleep scheduling schemefor high density cluster-based sensor networks. Comput. Commun. 28,1631–1642 (2005).
8. GLPK programming, 2011. http://www.gnu.org/software/glpk/
9. T Nguyen Thanh, Energy-efficient routing algorithms in wireless sensornetworks. PhD thesis (Monash University, Australia, 2008).
10. L Guo, Y Xie, C Yang, Z Jing, Improvement on LEACH by combiningadaptive cluster head election and two-hop transmission, in 9th International Conference on Machine Learning and Cybernetics, 2010, pp.1678–1683.
11. T Qiang, W Bingwen, W Zhicheng, MS-LEACH: a routing protocolcombining multihop transmission and single-hop transmission, in PacificAsiaConference on Circuits, Communications and Systems, 2009, pp. 107–110.
12. M Farooq, AB Dogar, GA Shah, MR-LEACH: multi-hop routing with lowenergy adaptive clustering hierarchy, in 4th International Conference onSensor Technologies and Applications, 2010, pp. 262–268.
13. J Chen, H Shen, MELEACH-L: more energy efficient LEACH for large scaleWSNs, in Wireless Communications, Networking and Mobile Computing, 2008,pp. 1–4.
14. SG Nash, A Sofer, Linear and Nonlinear Programming (McGraw-Hill, NewYork, 1996).
15. T Kanungo, DM Mount, NS Netanyahu, A local search approximationalgorithm for k-means clustering, in The Eighteenth Annual Symposium onComputational geometry, 2002. Volume 28, Issues 2–3, pp. 89–112.
16. S Lindsey, C Raghavendra, Power-efficient gathering in sensor informationsystems, in IEEE Aerospace Conference, 2002. Vol.3, p. 3-1125-3-1130.doi:10.1109/AERO.2002.1035242.0-7803-7231-X.
17. WB Heinzelman, Application-specific protocol architectures for wirelessnetworks. PhD dissertation (Massachusetts Institute of Technology, 2000).
18. T Nguyen Thanh, V Phan Cong, The Energy-Aware Operational Time ofWireless Ad-hoc Sensor Networks. ACM/Springer Mobile Networks andApplications (MONET) Journal, Volume 17, August, 2012. doi:10.1007/s11036-012-0403-1.
19. T Nguyen Thanh, Heuristic Energy-Efficient Routing Solutions to Extend theLifetime of Wireless Ad-Hoc Sensor Networks (Springer, LNCS 7197, 2012), pp.487–497. ISBN: 978-3-642-28489-2.

Color Image Segmentation Using Twin Support Vector Machine

Pooja Saigal*

Abstract

Image segmentation is an automated process of partitioning an image into non-overlapping regions. It is a challenging task in computer vision and is a critical step in image understanding. In recent years, various image segmentation algorithms have been developed. In this paper, we present the idea of segmentation for color images based on automatic pixel classification using Twin Support Vector Machine (TWSVM). This algorithm extracts color features from the images and uses them as training data for TWSVM. The TWSVM classifier is trained to distinguish between object and background pixels. This segmentation algorithm takes advantage of the local information of the image. It is shown experimentally that the proposed method outperforms other state-of-the-art segmentation methods.

Key Words: Image Segmentation, Homogeneity, Gabor Filter, Twin Support Vector Machine, K-Means Clustering.

I. Introduction

Image segmentation is a method of partitioning an image into various non-overlapping regions so that each region is nearly homogeneous, but the union of two or more regions is not [1]. It is a low-level vision task andis a critical task for image analysis and pattern recognition. In the recent years, a wide variety of methods are proposed for image segmentation and these methods can be categorized into five major categories [1–3]:

- Histogram thresholding-based methods [4,5]: Histogram thresholding converts a multilevel image into a binary image, where 0 value is assigned to background pixels and 1 to objects or foreground pixels of an image based on a comparison with some threshold value T. T is intensity or color value and it can be a constant or could vary with the location in the image, thus called as global and local thresholding respectively. These methods are popular due to their simplicity and efficiency.

Pooja Saigal*
Department of Information Technology,
IITM, Janakpuri, New Delhi, India

- Clustering-based methods [6,7]: Clustering is a process which replaces a set of homogeneous pixels by a cluster, where the pixels may belong to a cluster due to some similarity in color, texture etc. The popular clustering methods are K-means and fuzzy c-means (FCM) algorithms which iteratively partition an image into K clusters. These algorithms use color, texture, and spatial information, or a weighted combination of these factors to group the pixels. The pixels are assigned to the cluster that minimizes the variance between the pixel and the cluster center.
- Edge detection-based methods [8]: Edge detection methods identify the pixels in the image that correspond to the edges of the objects seen in the image. This generates a binary image highlighting edge pixels. Various edge detection algorithms used are Sobel, Prewitt and Laplacian operators.
- Region-based methods: Region-based segmentation is used to associate individual pixels of an image to sets of pixels called regions which could correspond to an object or a meaningful part of some object.
- Graph-based methods [9]: These methods make

use of a graph in which the nodes represent the image pixels and arcs link the neighboring pixels. The segmentation process aims at minimizing the weight that cuts a graph into sub-graphs. Such algorithms are computationally expensive.

Many of these algorithms have proved to be successful in various applications, but none of them is generally applicable to a large range of images. Recently, advanced approaches, such as SVM [10], have been successfully utilized in image segmentation. SVM classifiers have good generalization ability and therefore achieve better segmentation, but they are sensitive to pixel-level features and initialization. Yu and Chang [11] presented a method for solving scenery image segmentation by applying the SVMs methodology. Wang et al.[12,13] presented a color image segmentation algorithm by combining SVM and fuzzy c-means. They extracted color features and steerable filter based texture features from the image. Fuzzy c-means based algorithm is used for clustering the image pixels using color and texture features, which are used as training data for SVM algorithm for image segmentation. Wang et al. [20] propose a segmentation algorithm that uses Least Square Support Vector Machine (LS-SVM) as a classifier.

In this paper, we improve the color image segmentation algorithm given by [20]. We present a segmentation approach which is simple and effective as compared to other segmentation algorithms. The paper is organized into various sections where Section II presents an overview on TWSVM. Section III describes the pixel level color feature extraction. Section IV explains the segmentation process. The experimental setup is given in Section V and the paper concludes in Section VI.

II. Twin Support Vector Machine

Support vector machines (SVM) were introduced by Vapnik and is based on statistical learning theory and structural risk minimization principle [14]. A remarkable improvement over SVM was given by Khemchandani et al. as TWSVM [17]. TWSVM is a binary classifier that classifies data points using two non-parallel hyperplanes instead of a single hyperplane as in the case of conventional SVM. The two non-parallel hyperplanes are obtained by solving two Quadratic Programming problems (QPPs) of smaller size compared to a single large QPP solved by conventional SVM. Consider a binary classification problem of classifying m_1 datapoints belonging to class +1 and m_2 data points belonging to class -1 in the n -dimensional real space R^n. Let matrix $A \in R^{m_1 \times n}$ represent the datapoints of class +1 and matrix $B \in R^{m_2 \times n}$ represent the datapoints of class -1. Given the above stated binary classification problem, linear TWSVM seeks two non-parallel hyperplanes in R^n

$$x'w^{(1)} - b^{(1)} = 0 \quad \text{and} \quad x'w^{(2)} - b^{(2)} = 0 \tag{1}$$

such that each hyperplane is closest to datapoints of one class and farthest from the datapoints of other class. A new datapoint is assigned to class +1 or -1 depending upon its proximity to the two non-parallel hyperplanes. The linear TWSVM classifier is obtained by solving the following pair of QPPs

$$\begin{aligned} &\underset{w^{(1)},b^{(1)}}{\text{Min}} \ \frac{1}{2}\left\|Aw^{(1)} - eb^{(1)}\right\|_2^2 + c_1 e' q_1 \\ &\textit{subject to} \ -(Bw^{(1)} - eb^{(1)}) + q_1 \geq e, \ q_1 \geq 0e \end{aligned} \tag{2}$$

and

$$\begin{aligned} &\underset{w^{(2)},b^{(2)}}{\text{Min}} \ \frac{1}{2}\left\|Bw^{(2)} - eb^{(2)}\right\|_2^2 + c_2 e' q_2 \\ &\textit{subject to} \ (Aw^{(2)} - eb^{(2)}) + q_2 \geq e, \ q_2 \geq 0e \end{aligned} \tag{3}$$

Here $c_1, c_2 > 0$ are penalty parameters, and $q_1 \in R^{m_1 \times 1}, q_2 \in R^{m_2 \times 1}$ ack variables. One can then write the dual of QPPs of Eq. (2) and Eq. (3) and solve them to get the hyperplanes Eq. (1). Here we observe that TWSVM is approximately four times faster than the usual SVM. This is because, the complexity of the usual SVM is no more than m^3 (where $m = m_1 + m_2$ is the total number of data points), and TWSVM solves two QPPs, each of which

is roughly of size $\frac{m}{2}$. Thus, the ratio of runtimes is approximately

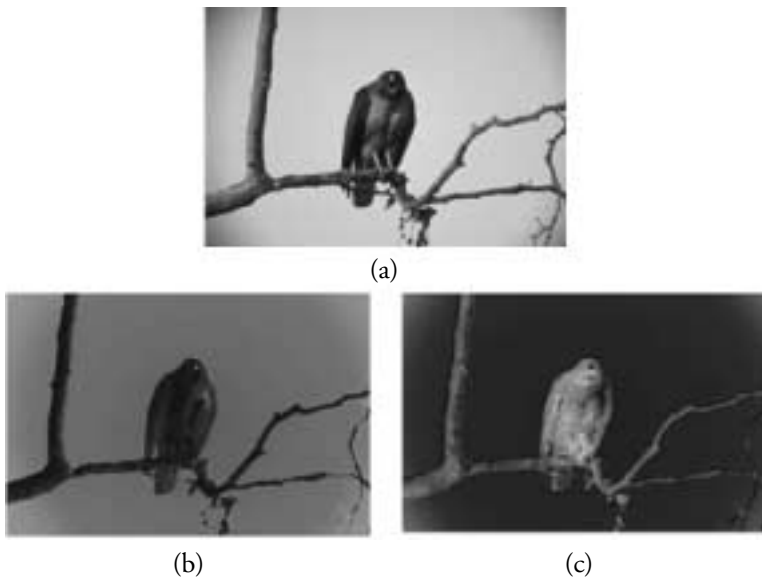
$$\frac{m^2}{2\times\left(\frac{m}{2}\right)^2}=4.$$

III. Pixel-Level Features

The problem of image segmentation is regarded as a classification task, where the task is to associate a label with every pixel in the image. so, it is important to extract effective pixel-level features. Color is a dominant and distinguishable low-level visual feature. The color feature used for this work is homogeneity. Here, each pixel of the image is classified as belonging to only one homogenous region i.e. it either belongs to the object or to the background.

Pixel-level Color Feature

The HSV color space is selected to extract pixel-level color features. The advantage of using HSV color space is [15]: (1) the brightness or lightness component is irrelevant to chromatic information of the images and (2) the chromatic component consisting of hue and saturation is intuitive. The effect of lighting on the color feature can be eliminated by considering only the HS color space to extract pixel-level color feature.



(a)

(b) (c)

Fig. 1: (a) Color image; (b), (c) H and S components respectively

The range of H and S are [0, 360] and [0, 1], respectively. To maintain uniformity between H and S components, the range of H is scaled to [0, 1]. Fig.1 shows the H and S components of a color image in HSV color space. In this paper, the pixel-level color feature homogeneity is extracted. The feature consists of two components: discontinuity and standard deviation. Let $P_{x,y}=\left(P^H_{x,y},P^S_{x,y},P^V_{x,y}\right)$ represent the three color components of a pixel at location (x,y) in an M X N image. The components of homogeneity are calculated in the following manner:

1) Calculate Discontinuity

The discontinuity $c^k_{x,y}$ measures the degree of abrupt changes in the color levels of pixel component $P^k_{x,y}\ (k=H,S)$ that is, the discontinuity is described by its edge value, and is generally obtained by applying edge detectors to the image. There are many different edge operators: Sobel, Canny, Derish, Laplace etc. Due to the simplicity and effectiveness, the Sobel operator is used for calculating the discontinuity and the

magnitude of the gradient at location (x,y).

$$c_{x,y}^{k} = \sqrt{G_{x'}^{k^2} + G_{y'}^{k^2}} \quad (k=H,S) \tag{4}$$

where $G_{x'}^{k}$ and $G_{y'}^{k}$ are the components of the gradient in the x' and y' directions respectively.

2) Calculate Standard Deviation

We assume that the signals are ergodic, so standard deviation $v_{x,y}^{k}$ gives the contrast within a local window, and is calculated for a pixel component as follows:

$$v_{x,y}^{k} = \sqrt{\frac{1}{d^2} \sum_{(m,n) \in \Omega_{x,y}} \left(P_{x,y}^{k} - \mu_{x,y}^{k}\right)^2} \tag{5}$$

where $x \geq 2$, $m \leq M-1$, $y \geq 2$ and $n \leq N-1$.

$\mu_{x,y}^{k}$ is the mean of the color levels of pixel component $P_{x,y}^{k}(k = H,S)$ within window $\Omega_{x,y}$ and is defined by

$$\mu_{x,y}^{k} = \frac{1}{d^2} \sum_{(m,n) \in \Omega_{x,y}} P_{m,n}^{k} \tag{6}$$

3) Find Pixel-level Color Feature

Homogeneity is related to the local information extracted from an image and determines the uniformity of the region. An image can be segmented into non-overlapping homogeneous regions by making an effective use of homogeneity. The local homogeneity of image pixel at the location (x,) for color component is represented by

$$h\left(P_{x,y}^{k}\right) = 1 - E\left(P_{x,y}^{k}\right) \times V\left(P_{x,y}^{k}\right) \tag{7}$$

where

$$V\left(P_{x,y}^{k}\right) = \frac{v_{x,y}^{k}}{\max\left\{v_{x,y}^{k}\right\}}, \; E\left(P_{x,y}^{k}\right) = \frac{c_{x,y}^{k}}{\max\left\{c_{x,y}^{k}\right\}} \tag{8}$$

Here, and are, respectively, the standard deviation and the discontinuity of image pixel at the location (x,y) for color component .The pixel-level color feature $CF_{x,y}$ of the image pixel $P_{x,y}$ at location (x,y) is calculated as

$$CF_{x,y} = \left(h\left(P_{x,y}^{H}\right), h\left(P_{x,y}^{S}\right)\right) \tag{9}$$

IV. The TWSVM based image segmentation using color features

In this work, we propose the use of TWSVM [17] for pixel classification. TWSVM is proved to be faster than SVM and is more effective classifier. In this paper, we present a color image segmentation technique using automatic pixel classification with TWSVM. We first extract the pixel-level color features and use them as input of TWSVM classifier. The training samples are selected by using k-means clustering, with k=2. The k-means clustering method generates a binary image of the original image and assigns a label to each pixel as belonging to object or background. We, randomly, select 1000 pixels each from the object and the background, as the training dataset for TWSVM.

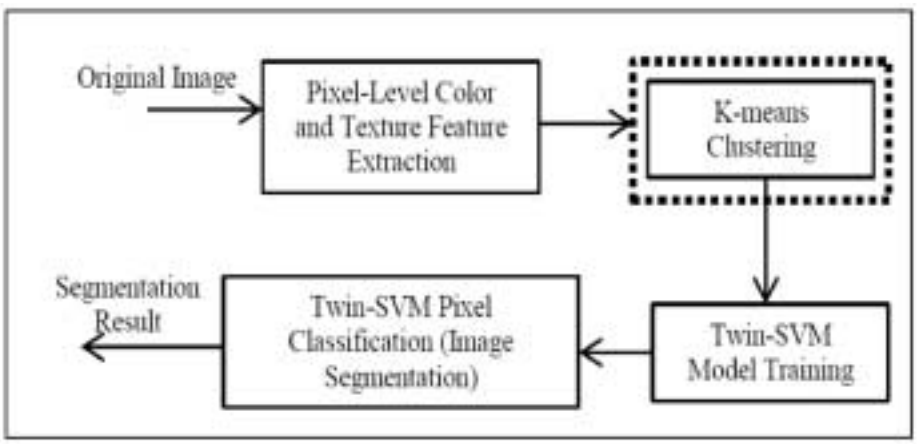


Fig. 2: Color image segmentation using pixel classification with TWSVM

V. Experimental setup and Performance evaluation

The proposed method is used to segment an image into distinct regions. The images are selected from Berkeley Segmentation Database (BSD) [18]. The BSD comprises of various images from the Corel dataset and contains ground truth of 300 images for benchmarking image segmentation and boundary detection algorithms. The content of the images are landscapes, animals, portraits and various objects. The metric used for the quantitative evaluation of the proposed algorithm is segmentation Error Rate Average Normal Precision (ER) [1]. It presents the ratio of misclassified image pixels over the total image pixels, and the error rate is defined as:

$$ER = \frac{N_f + N_m}{N_t} \times 100\% \quad (12)$$

where Nf is the number of false-detection image pixels, Nm denotes the number of miss-detection image pixels, and Nt is total number of images pixels.

For experimentation, the local window size is 5 X 5. The linear classifier is used by TWSVM. Fig. 4 shows the comparative segmentation results. Table I presents the quantitative evaluation, in terms of error rate (ER), for four different image segmentation algorithms.

Table-I: The ER achieved on Barkeley Segmentation database images

Image	Error Rate (ER)			
	Edison Method [19]	Mean-shift Method [6]	LS-SVM [20]	Proposed Method
1	16.52%	9.54%	9.16%	3.67%
2	4.83%	8.72%	1.49%	1.57%
3	7.0%	7.6%	6.2%	0.84%
4	10.81%	9.16%	8.97%	8.90%

VI. Conclusions

In this paper, an innovative approach for color image segmentation based pixel classification is proposed with twin support vector machine. The proposed method combines the pixel-wise information of color image and the generalization ability of TWSVM classifier. The color feature, homogeneity, is extracted which is in consideration of local human visual sensitivity for color pattern variation in HSV color space. Then, the pixel-level color features are used as input of TWSVM classifier. TWSVM model is trained by selecting the training samples with k-means clustering. Finally, the color image is segmented with the trained TWSVM model. The experiments, done on various images taken from Berkeley image Segmentation Database, indicate that the proposed algorithm outperforms various color image segmentation algorithms. The proposed segmentation algorithm can be extended to be applied on videos.

References

1. R. Unnikrishnan, C.E. Pantofaru and M. Hebert, "Toward objective evaluation of image segmentation algorithms", IEEE Transaction on Pattern Analysis and Machine Intelligence 29 (6) (2007), pp. 929–943.
2. K.C. Ciesielski, Kcies and J.K. Udupa, "A framework for comparing different image segmentation methods and its use in studying equivalences between level set and fuzzy connectedness frameworks", Computer Vision and Image Understanding 115 (6) (2011), pp. 721–734.
3. J.E. Francisco and D.J. Allan, "Benchmarking image segmentation algorithms", International Journal of Computer Vision 85 (2) (2009) pp. 167–181.

4. M.E. Yuksel and M. Borlu, "Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic", IEEE Transacation on Fuzzy Systems 17 (4) (2009), pp. 976–982.
5. K.S. Tan and N.A.M. Isa, "Color image segmentation using histogram thresholding–fuzzy C-means hybrid approach", Pattern Recognition 44 (1) (2011), pp.1–15.
6. D. Comaniciu and P. Meer, 'Mean shift: a robust approach toward feature space analysis", IEEE Transaction on Pattern Analysis and Machine Intelligence 24 (5) (2002), pp. 603–619.
7. B. Sowmya and B.S. Rani, "Colour image segmentation using fuzzy clustering techniques and competitive neural network, Applied Soft Computing 11 (3) (2011), pp. 3170–3178.
8. J. Wang, L.L. Ju and X.Q. Wang, An edge-weighted centroidal voronoi tessellation model for image segmentation, IEEE Transaction on Image Processing 18 (8) (2009), pp. 1844–1858.
9. Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts", IEEE Transaction on Pattern Analysis and Machine Intelligence 23 (11) (2001), pp. 1222–1239.
10. H.X. Xu, W.H. Cao and W. Chen, "Performance evaluation of SVM in image segmentation", 9th International Conference on Signal Processing (ICSP 2008), Beijing, China, 26–29 Oct, 2008 pp. 1207–1210.
11. Y.Y. Hui and C.C. Chang, "Scenery image segmentation using support vector machines", Fundamenta Informaticae 61 (3–4) (2004), pp. 379–388.
12. X.Y. Wang, T. Wang and J. Bu, "Color image segmentation using pixel wise support vector machine classification", Pattern Recognition 44 (4) (2011), pp. 777–787.
13. X.Y. Wang, Q.Y. Wang, H.Y.Yang and J. Bu, "Color image segmentation using automatic pixel classification with support vector machine classification", Neurocomputing 74(2011), pp. 3898–3911
14. V.Vapinik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York,2000.
15. C.F. Juang, S.H. Chiu and S.J. Shiu, "Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation", IEEE Transaction on Systems, Man, and Cybernetics, Part A 37 (6) (2007), pp. 1077–1087.
16. J.F. Khan, R.R. Adhami and S.M.A. Bhuiyan, "A customized Gabor filter for unsupervised color image segmentation", Image and Vision Computing 27 (4) (2009), pp. 489–501.
17. Jayadeva, R. Khemchandani and S. Chandra, "Twin Support Vector Machines for Pattern Classification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, NO. 5, May 2007, pp. 905-910
18. C. Fowlkes and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues", IEEE Transaction on Pattern Analysis and Machine Intelligence 26 (5) (2004), pp. 530–549.
19. C.Christoudias, B.Georgescu and P.Meer, "Synergisminlow-levelvision", Proceedings of the 16th International Conferenceon Pattern Recognition, Quebec City, Canada, vol.IV, August 2002, pp.150–155.
20. H.Y. Yang, X.Y. Wang, Q.Y. Wang and X.J. Zhang, "LS-SVM based image segmentation using color and texture information", Journal of Visual Communication and Image Representation 23 (2012), pp. 1095–111.

Past Present & Future of Data Mining

Priti Sharma*
Shikha Sharma**

Abstract

In this era where the technology evolves every day, there is consistent need of a new technology. Data mining is the technique which extracts important information from a humongous amount of data. Data is commonly known as the new gold. It is the most precious part of any organization and data mining or knowledge extraction are the technologies behind keeping this data up to date and usable. This is the technique which has taken the world by storm. In this paper we intend to study the various algorithms used in the past and present scenario. Additionally we also propose to study the future techniques to be used in this area.

Key Words: Sentiment Analysis, Web Mining, Bio Informatics, Scientific Computing

I. Introduction

The data stored on the Internet has been growing very rapidly. "The information age has emerged due to rapid change in electronic data". Companies collect all sorts of information about the business process such as financial, payroll, customer data, inventory etc. Similarly in other areas terabytes of data is produced. Subsequently, there is growing demand for methods to analyse large volumes of data. The field of data mining has emerged to analyse this large amount of data.

Data mining is broadly defined as the process of finding "patterns" from large amounts of data. Data mining is considered to be the most important step in the knowledge discovery process. Although the process of knowledge discovery involves several pre-mining and post-mining steps such as data cleansing and data visualization but in the present paper, the focus is on data mining.

Initially, data mining was used for well structured data such as relational data. Data mining of structured data has been successfully applied to various disciplines. However, World Wide Web (WWW) has introduced large amounts of data. Such data is also known as semi structured data [1].

Priti Sharma*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Shikha Sharma**
Department of Information Technology,
IITM, Janakpuri, New Delhi

II. Methods of Data Mining

Association

Association is the most familiar data mining technique. Here, patterns can be identified by simply making a correlation between two or more items, often of the same type. An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent. Association rules are created by analysing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. In data mining, association rules are used for analysing and predicting customer behaviour. Association rules can be used in shopping basket data analysis, catalog design, product clustering and store layout. Association rules can also be used by programmers to build programs capable of machine learning. Machine learning is a type of artificial intelligence that seeks to build programs with

the ability to become more efficient without being explicitly programmed [3].

Classification

Classification can be used to build up an idea of the type of customer, item or object by describing multiple attributes for identifying a particular class [2]. For e.g. cars can be classified into different types by identifying different attributes. Classification is a data mining technique that is used to predict group membership for data instances. For e.g by using classification you can predict whether a particular day will be sunny, rainy or cloudy. Decision trees and neural networks are popularly used classification methods.

Clustering

Clustering is grouping a set of objects in such a way that objects in the same group or cluster are more similar (in some sense or another) to each other than to those in other groups (clusters) [4]. Clustering is useful for identifying different information by considering various examples where one can see where the similarities and ranges agree. You can group individual pieces of data together by examining one or more attributes or classes. In clustering we use one or more attributes as a basis for identifying a cluster of correlating results [2].

Prediction

Prediction can be combined with other data mining techniques such as classification, pattern matching and relation. Prediction runs from predicting the failure components for identifying fraud. We can even predict company's profit through prediction data mining. Past events can be analysed to make a prediction about an event [2].

Sequential Patterns

Sequential patterns can be used for identifying trends or regular occurrences of similar events. By analysing customer data you can identify that customers buy a particular collection of products together at different times of the year. You can use this information in a shopping basket application to suggest that certain items can be added to a basket based on their past purchasing history [2].

III. Data Mining in the Past

Data mining can be applied to various fields like Artificial Intelligence, Machine learning, Database management system, Statics etc. The era of data mining applications started in 1980's. Initially, data mining algorithms worked best for numerical data collected from a single database and various data mining techniques have evolved for flat files, traditional & relational databases where data is stored in tabular form.

In early days of data mining most of the algorithms employed only statistical techniques. Later on they evolved with various computing techniques like artificial intelligence, machine learning and pattern recognition. Various data mining techniques like induction, Compression and approximation were developed. Also various algorithms were developed to mine the large volumes of heterogeneous data stored in the data warehouse [5].

IV. Data Mining in the Current Scenario

Data mining applications have now been successfully implemented in various domains like business, healthcare, telecommunications, risk analysis etc. The increasing complexities in various fields have posed new challenges in the field of data mining. The various challenges include different data formats, ever growing business challenges, different data formats etc. Advancements in data mining have shaped the present data mining applications to handle the various challenges.

V. Future of Data Mining

Data mining has become a major discipline in computer science. The future trends of data mining include:

- Web & Semantic mining
- Sentiment Analysis
- Scientific Computing
- Business Data Analysis
- Fight Against Terrorism
- Bioinformatics and care for Diseases

Table-I: Current Data Mining Techniques

Data mining type	Data Formats	Algorithms	Application Areas
Hypermedia data mining	Hypertext data	Classification & Clustering techniques	Intranet & Internet applications
Multimedia data mining	Multimedia data	Rule based decision tree classification	Audio/Video applications
Spatial data mining	Spatial data	Spatial clustering techniques, Spatial OLAP	Network, Remote sensing & GIS applications
Ubiquitous data mining	Ubiquitous data	Traditional data mining techniques drawn from statistics & Machine learning	Applications of Mobile phones, PDA's.
Time series data mining	Time series data	Rule Induction algorithms	Business & Financial applications

Table-II: Comparative Study of Data Mining Trends

Data Mining Trends	Data Formats	Algorithms	Application Areas
Past	Numerical data & structured data stored in traditional databases	Statistical, Machine Learning techniques	Business
Current	Heterogeneous data formats which includes structured, semi structured & unstructured data	Statistical, Machine Learning, Artificial Intelligence, Pattern recognition techniques	Business, Web, Medical Diagnosis
Future	Complex data objects includes high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi instance objects, Multi represented objects & temporal data etc.	Soft computing techniques like fuzzy logic, neural networks & genetic programming	Business, Web, Medical Diagnosis, Scientific & Research analysis fields, Social networking etc.

VI. Conclusion

In this paper we have reviewed different data mining techniques used in past, present and future. Data mining is a vast field and its applications are not limited to a particular area. There are still many untouched areas in this field which needs to be explored. In future we can work on different algorithms of sentiment analysis.

References

1. Svetlozar Nestorov, Data mining techniques for structured and semi structured data, June 2000.
2. A survey on data mining techniques IJARCCE Vol3, Issue 1, January 2014.
3. searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining.
4. https://en.wikipedia.org/wiki/cluster-analysis.
5. "Data Mining Past, Present & Future Scenario" International Journal of Emerging Trends & Technology in Computer Science IJETICS Volume 3, Issue 1 January-Feb 2014 ISSN 2278-6856.
6. Goele Sangeeta, Chanana Nisha "Data Mining Trend In Past, Current and Future" International Journal of Computing & Business Research ISSN (Online): 2229-6166.

Comparative Analysis of Automation Testing Tools

Charul Nigam*
Sushma Malik**

Abstract

Software testing is the process to evaluate entire software components and verifies that the software meets the customer requirements. It also bridges the gap between expected and actual results. Testing can be combined together at various project development phases in order to achieve the desired results. It also gives an overview of the tools and methodology used during the project development or project testing phase. This paper includes comparative study of three widely used automation testing tools like selenium, ranorex and test complete, based on some important features like platform, script language, programming skills etc.

Key Words: Software Testing, Automation Testing, Manual Testing, Testing Tools

I. Introduction

Software testing can be defined as a process to execute different software components with an intention to find errors in it. Testing process informs the development team about the errors and defects present in each phase of the software.

When to Start Testing

As soon as the requirements are finalized the first step is to start making test cases based on those requirements. This will minimize the cost, time to rework and development efforts so that an error free software can be delivered to the customer. The success of testing depends on the development model that is being used to develop the software.

When to Stop Testing

Software testing is a never-ending process and no one can say that any software is 100% tested. Testing is a continuous process which start as soon as the requirements are gathered and goes till the software is running in the real world. Following are the aspects which should be considered to stop the testing:

Charul Nigam*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Sushma Malik**
Department of Information Technology,
IITM, Janakpuri, New Delhi

- Deadlines of the Software Testing.
- Completion of test case execution.
- Completion of Functional and code coverage to a certain point.

II. Levels of Testing

Unit Testing

Testing the individual components of the software is termed as unit testing. It is done by the developers at their end when they are developing the software. The main objective of performing unit testing is that all the individual components of the software are working properly in isolation. If the developers are doing the testing, it does not mean that the testing team will not test the individual components of the software. The developers use test data that is separate from the test data of the testing team.

Integration Testing

The testing of combined software components to determine if they function correctly together is referred to as integration testing. There are two methods of doing Integration Testing Bottom-up Integration testing and Top-down Integration testing.

- Bottom-up integration testing begins with unit testing, followed by tests of progressively higher-level combinations of units called modules or builds.

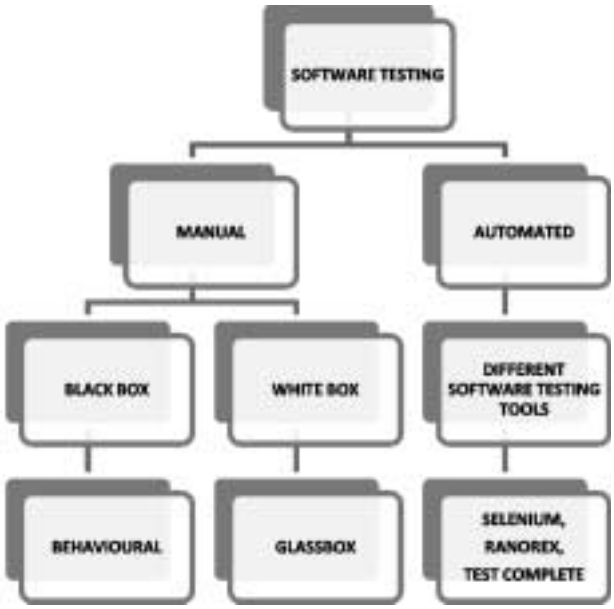


Fig. 1: Hierarchy of Software Testing

- Top-Down integration testing, the highest-level modules are tested first and progressively lower-level modules are tested after that. In a comprehensive software development environment, bottom-up testing is usually done first, followed by top-down testing.

System Testing

It is the last level in the testing and tests the system as a whole. Once all the components are integrated and tested, the entire software is tested to see that it meets the customer requirements and is developed as per the quality Standards. This type of testing is performed by a specialized testing team which may or may not different from the team which performs unit and integration testing.

III. Hierarchy of Software testing

Software testing is broadly categorized into two types:-

- Manual testing
- Automated testing

IV. Manual Testing

Manual testing is a process in which we test the software manually, i.e., without taking any help from an automated tool or any script. In this type of testing, the tester acts as an end-user and tests the software to identify any faults or failures or bugs.

Manual Testing is classified into:-

- Black Box Testing
- White Box Testing

Black Box Testing

Black-box testing is a type of software testing in which we are concerned with output based on some input conditions. We are not concerned with the internal logic of the action thus performed. Even if the tester does not have knowledge of code, he/she can perform black box testing. It is also known as functional testing. This type of testing is purely based on the software requirements as specified by the customer.

White Box Testing

White box testing is a testing technique which focuses on internal structure of the software. The tester critically examines the code, makes test cases accordingly and then performs the testing. It focuses on strengthening security, the flow of inputs and outputs through the application, and improving design and usability. It is also known as structural testing.

White-box testing follows a technique known as glass box testing technique.

V. Automation Testing

Automated testing is a method of software testing where special software tools are required to control the execution of tests and then compares actual test results with expected results of the software components. The objective of automation testing is to ease the testing effort as much as possible with a minimum set of scripts. Automated testing is applicable on large projects; where projects require rigorous testing on the same code segments again and again.

The automation software testing consists of a sequence of activities, processes and tools that processed in order to execute the test on software and to keep the record of the result of tests. The following activities include in testing process:

- Test planning
- Test design
- Implementation
- Test execution
- Test report
- Test evaluation

To improve the efficiency and flexibility of maintaining test scripts test automation interface is use. It includes Interface Engine which consist runner to execute the test scripts, the Interface Environment that consists Framework and Project Library and Object Repository are collection of application object data.

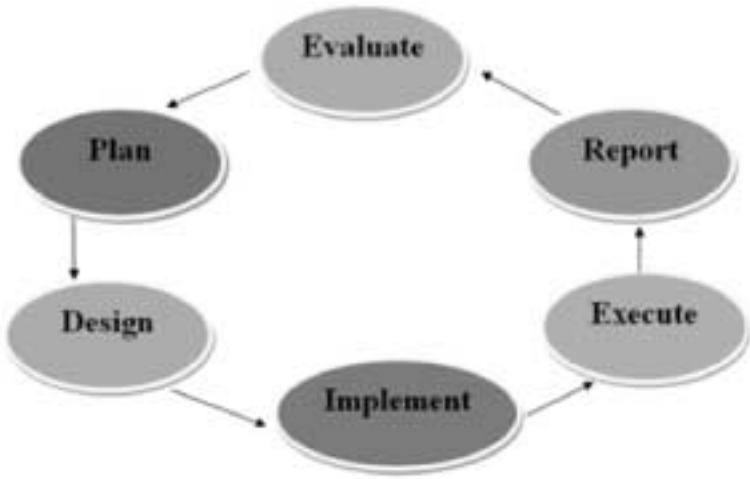


Fig. 2: Process of Automation Testing

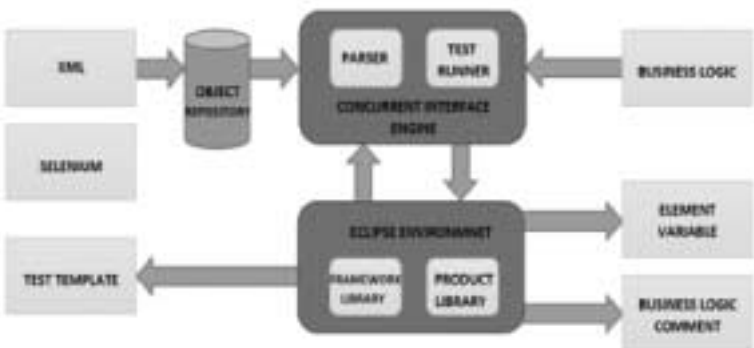


Fig. 3: Testing Interface

Why use Automation Testing?

- Manual Testing is time and cost consuming
- It is difficult to perform manual testing for multi lingual sites.
- Automation Testing does not require tester interference. Tests can be run overnight automatically.
- Speed of test execution is increased.
- Automation helps increase Test Coverage

Automation testing, which is also known as Test Automation, is when the tester writes scripts and uses software to test the product.

Automation techniques follows some software's, scripts, tools to perform testing these tools are known as automation testing tools.

Some of these automation tools are:-

- Selenium
- Ranorex
- Test Complete

VI. Automation Testing Tools

Various types of tools are used for automated testing and they can be used in different areas of testing. The selection of tool is based on the type of application which we want to test like automated web testing tools, GUI testing tools.

Selenium

Selenium is an open source web testing tool which is used to test the web browsers across different platforms. It is not just a single tool but a suite of software's, each catering to different testing needs of an organization. It works anywhere where javascript is supported. To start working with selenium you should have sound knowledge of HTML.

It is divided into four components:-

- **Selenium IDE:** It is an integrated development environment for selenium tests. It allows us to record edit and debug the tests. It also contains a log of events that are executed including any errors or warnings that may have occurred. The execution commands are run, walk and step.
- **Selenium Remote Control:** It allows users to write automated web application tests in any programming language. Selenium RC comprises of two components: 1. The selenium server which launches and kills browsers. 2. Client libraries that provide an interface between programming language and selenium server.
- **Web Driver:** It is a web automated framework that allows us to execute our tests on different browsers which was not possible with selenium IDE. Both

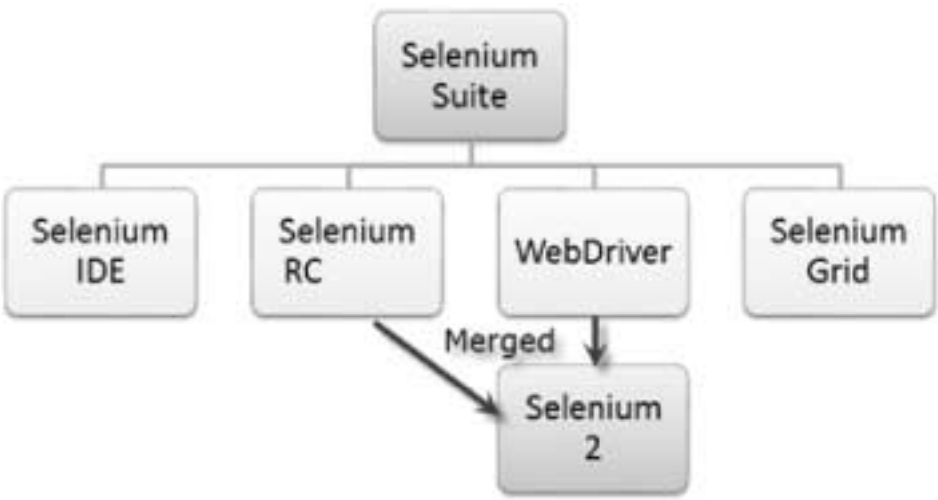


Fig. 4: Components of Selenium

selenium rc and web driver allow the use of programming language and test your script against different browsers but they differ on :-1. Architecture 2. Speed 3. Real time interaction 4. API 5. Browser Support

- **Selenium Grid:** It supports distributed test execution. There are 2 reasons for using selenium grid:1. To run your tests against multiple browsers, multiple versions of browser, and browsers running on different operating systems. 2. To reduce the time it takes for the test suite to complete a test pass.

Ranorex

Ranorex is a GUI test automation framework which provides testing of a wide range of desktop, web and mobile applications. Test cases once written can be executed on different platforms.

It simulates the user actions by record and replay tool into recording modules. Ranorex is easy to use and affordable even for small testing teams. It contains different tools that have their own characteristics which help in testing of software. All tools are integrated in the Ranorex Studio framework.

The Ranorex Studio tools allow you to test applications running the following platforms:-

- Winforms, C#, VB.NET
- Web (IExplorer, Firefox, Chrome, Safari)
- WPF, Silverlight, Qt
- SAP, Oracle Forms, MS Dynamics
- Flash/Flex
- Java
- Android
- iOS

In addition to the technologies Ranorex also supports:-

- Capture & Replay
- Keyword-Driven Testing
- Data-Driven Testing

Table-1: Comparison of Automation Testing Tools

S.No.	Tools/Features	Selenium	Ranorex	Test Complete
1.	**Pricing**	Open Source	To be purchased	To be purchased
2.	**Cross platforms.**	Windows Only	Windows Only Except XP	Window 7 and later versions
3.	**Cross –Browsers**	Chrome Firefox IE-Opera	IE Firefox Chrome Safari	IE-Firefox Opera Chrome
4.	**Record-Playback.**	support	support	support
5.	**Script-language.**	Ruby-java python php- java script	VB script	VBScript C#-Jscript
6.	**Ease of Learning.**	Experience needed	Experience needed	Experience needed
7.	**Programming skills**	Required	Partially	Required
8.	**Online-Support.**	Strong Support	Strong Support	Strong Support
9.	**Debugging support.**	Strong	Strong	Strong
10.	**Report Generation.**	HTML	HTML	HTML/XML
11.	**Application support**	Web applications only	Web, Desktop, and Mobile applications	Web, Desktop, and Mobile applications
12.	**Data-Driven**	Excel-CSV	Excel, CSV, SQL	Excel, CSV, SQL

Test Complete

Test complete has open flexible architecture for maintaining and executing automated tests for web.

This tool helps to keep the balance between quality and speed of delivery of applications at affordable cost.

With test complete tool different types of testing can be done like unit testing and GUI testing, regression testing etc.

TestComplete Platform acts as a backbone for TestComplete Desktop, Web, and Mobile.

TestComplete Platform helps the tester to find the balance between the speed of application delivery and quality of software at an within your means.

TestComplete Platform's integration with other tools ensures quick turnaround is available while releasing software in frequent and shorter release cycles.

TestComplete Platform includes support for:

- Test recording
- Automated test execution
- Object name mapping
- Data-driven testing with data generation wizards
- Object-driven testing
- Keyword-driven testing
- Test Visualizer
- Database testing support
- Developing custom extensions
- Integration with other tools

VII. Conclusion

This paper presents a study on various automated testing tools that used on different platforms. Automation testing tools helps the tester to easily automate the whole testing process. Automation testing improves the accuracy and also save time of the tester as compared to the manual testing. Complete testing is infeasible. These automation tools enable all the developers and testers to automate the complete process of software testing easily. Automated Software testing is the best method to test the software effectively and efficiently. Selenium is a framework comprises of many tools used for testing web applications. With the help of the case study, we analyze and find the testing of a web application using automation testing tool.

References

1. Richa Rattan, "Comparative Study Of Automation Testing Tools: Quick Test Professional & Selenium", Vsrd International Journal Of Computer Science & Information Technology, Vol. 3 No. 6 June 2013.
2. Manjit Kaur, Raj Kumar, "Comparative Study Of Automated Testing Tools: Test Complete And Quick Testpro", International Journal Of Computer Applications (0975-8887) Volume 24-No. 1, June 2011.
3. K. Mustafa, Rafa E, Al-Qutaish, M, I. Muhairat "Classification of Software testing tools Based on the Software Testing Methods' Second International Conference on Computer abd Electrical Engineering, Amman, Jordan, 2009.
4. Neha Bhateja "A Study on Various Software Automation Testing Tools", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 6, June 2015.
5. Shivkumar Hasmukhrai Trivedi, "Software Testing Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012
6. Jon Wiley Sons "Software Testing fundamentals, Methods, and Matrics".
7. Michael T. Grater University or Oregon "Benefits of Using Automated Software Testing Tools to Achieve Software Quality Assurance,Applied Information Technology, June 2005.
8. Richa Rattan, "Comparative Study Of Automation Testing Tools: Quick Test Professional & Selenium", Vsrd International Journal Of Computer Science & Information Technology, Vol. 3 No. 6 June 2013.
9. Manjit Kaur, Raj Kumar, "Comparative Study Of Automated Testing Tools: Test Complete And Quick Testpro", International Journal Of Computer Applications (0975-8887) Volume 24-No. 1, June 2011.
10. Whittaker J. A., "What is Software Testing? And Why Is It So Hard?" IEEE Software, January 2000, pp. 70-79.

A Review of Various Energy Optimization Routing Protocols in Wireless Sensor Networks

Nemi Rishi*
Akhil Kumar**

Abstract

In wireless sensor networks major frailty of communication between the connecting nodes is their limited energy resource. Along with increasing the efficacy of the sensor node, it is desirable to distribute the energy dissipated throughout the network in order to reduce the overhead and maximize efficiency of overall system. Any communication protocol used for synchronization of nodes incurs some overhead to set up the communication between the nodes. In this paper we have discussed and presented a comparison of various energy-efficient routing algorithms. The comparison parameters taken in consideration are the setup costs, the energy efficiency and the useful lifetime of the system. To better understand the various energy efficient algorithms, we have also presented a comparison of an optimum clustering algorithm with the existing algorithms.

The aim of this study is to present an ideal algorithm to reduce the cost of communication between the peer nodes. The algorithms are compared in terms of system lifetime, power dissipation distribution, synchronization cost, and complexity of the algorithm.

Key Words: Wireless Sensor Networks, Energy-Efficient Routing, Diffusion Routing Algorithm, Routing Protocols.

I. Introduction

In recent decade the size of computers has decreased and the processing power has increased exponentially. Due to this rapid advancement in computers, the role of computers has increased rapidly in day to day life of humans. Due to this property of scalability of computers, the idea of use of computers as a chip has come into existence, both theoretically and practically.

As wireless sensor networks is becoming a reality. Hence, the long term limitation of this technology is to be understood and addressed to remove any anomalies.

This paper address limitations of wireless sensor networks such as: limited energy capacity, energy localization, transmission cost and limited processing potentiality. These mentioned issues existing in wireless sensor networks are completely opposite to the wired networks, where energy consumption is not an issue of high concern, transmission cost is much lower, and the network nodes have high processing power. The routing approaches of past are not sufficient for this newer communication technology. Along with increasing the lifetime of the nodes, it is considered to distribute the energy dispel throughout the sensor network in order to reduce the maintenance and maximize overall system performance. Some overhead always incurs between any communication protocols. In this paper with the help of complex algorithms an attempt is made to overshadow the extra control messages each node needs to communicate. Each node could make the most informed decision regarding its communication options if they had complete knowledge of the entire network topology and power levels of all the nodes in the network.

Nemi Rishi*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Akhil Kumar**
Department of Information Technology,
IITM, Janakpuri, New Delhi

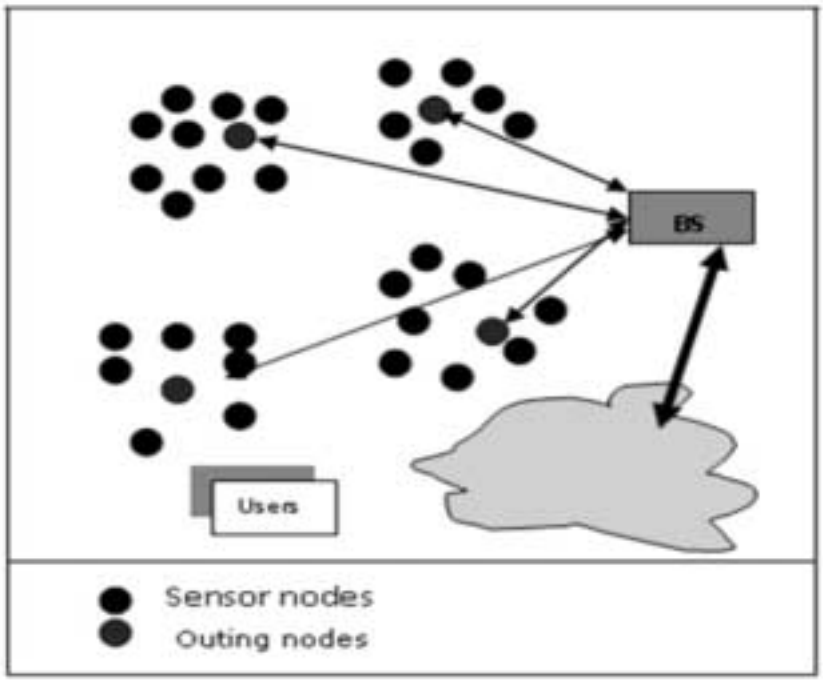


Figure-1: A wireless sensor network structure

The usual topology of wireless networks have various nodes spread over a physical area. Due to this property of not following any hierarchy, the wireless sensor networks are also known as Ad-hoc networks. These networks may perform stand-alone or in connection with a large network through a base station. In order to have an ideal communication, small hops are considered over longer communication periods.

II. The Routing Protocols:

In the next few sub-sections, we will discuss the routing protocols in detail.

Briefly, the protocols are:

1. Direct communication, in which each node communicates directly with the base station.
2. Diffusion-based algorithm utilizing only location data.
3. E3D: Diffusion based algorithm utilizing location, power levels, and node load.
4. Random clustering, similar to LEACH, in which randomly chosen group heads receive messages from all their members and forward them to the base station.
5. An optimum clustering algorithm, in which clustering mechanisms are applied after some iterations in order to obtain optimum cluster formation based on physical location and power levels.

Direct Communication

Each node is assumed to be within communication range of the base station and that they are all aware who the base station is. In the event that the nodes do not know who the base station is, the base station could broadcast a message announcing itself as the base station, after which all nodes in range will send to the specified base station. So each node sends its data directly to the base station. Eventually, each node will deplete its limited power supply and die. When all nodes are dead and the system is said to be dead.

The main advantages of this algorithm lie in its simplicity. There is no synchronization to be done

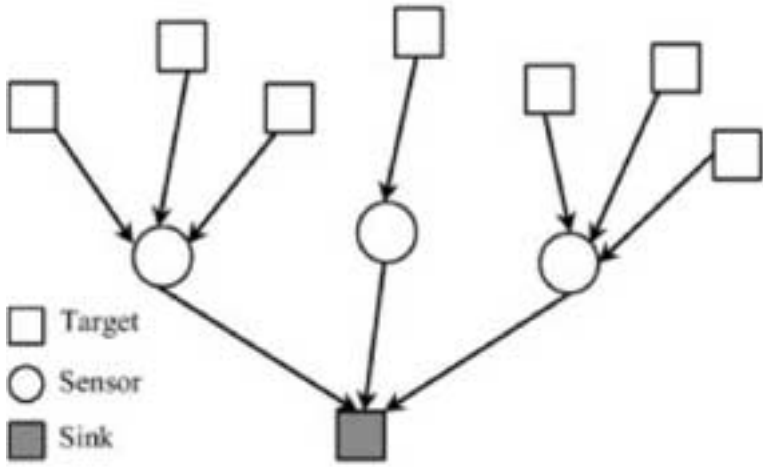


Figure-2: Direct Communication

between peer nodes, and perhaps a simple broadcast message from the base station would suffice in establishing the base station identity. The disadvantages of this algorithm are that radio communication is a function of distance squared, and therefore nodes should opt to transmit a message over several small hops rather than one big one; nodes far away from the base station will die before nodes that are in close proximity of the base station. Another drawback is that communication collision could be a very big problem for even moderate size networks.

Diffusion based algorithm using location information

Each node is assumed to be within communication range of the base station and that they are all aware who the base station is. Once the base station identity is established, the second sequence of messages could be between each node and several of their closest neighbors. Each node is to construct a local table of signal strengths recorded from each of their neighbors, which should be a direct correlation to the distance those nodes are from each other. The other value needed is the distance from each neighbor to the base station, which can be figured out all within the same synchronization messages. This setup phase needs only be completed once at the startup of the system; therefore, it can be considered as constant cost and should not affect the algorithm's performance beyond the setup phase.

In every transmission, each node sends its data that is destined for the base station, to the best neighbor. Each

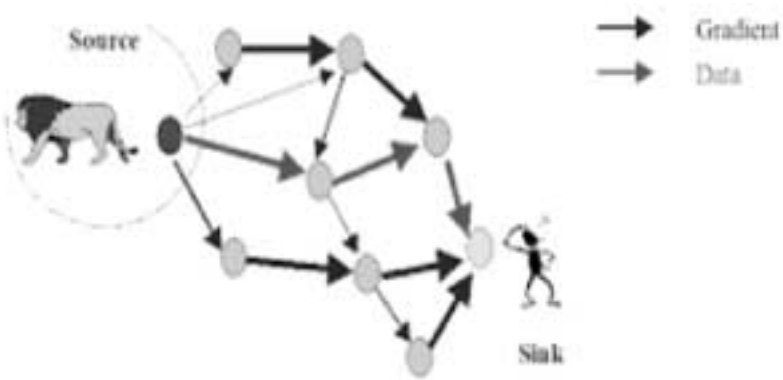


Figure-3: Diffusion Based Routing

node acts as a relay, merely forwarding every message received to its respective neighbor. The best neighbor is calculated using the distance from the sender and the distance from the neighbor to the base station. This ensures that the data is always flowing in the direction of the base station and that no loops are introduced in the system; this can be accomplished by considered not only the distance from the source to the candidate neighbor, but also from the candidate neighbor to the base station. Notice that the complete path is not needed in order to calculate the best optimal neighbor to transmit to. Since each node makes the best decision for itself at a local level, it is inferred that the system should be fairly optimized as a whole.

The main advantage of this system is its fairly light complexity, which allows the synchronization of the neighboring nodes to be done relatively inexpensive, and only once at the system startup. The system also distributes the lifetime of the network a little bit more efficiently.

The disadvantage of this system is that it still does not completely evenly distribute the energy dissipated since nodes close to the base station will die far sooner before nodes far away from the base station. Notice that this phenomenon is inversely proportional to the direct communication algorithm. It should be clear that this happens because the nodes close to the base station end up routing many messages per iteration for the nodes farther away.

E3D: Diffusion based algorithm using location, power, and load as metrics

In addition to everything that the basic diffusion algorithm performs, each node makes a list of suitable neighbors and ranks them in order of preference, similar to the previous approach. Every time that a node changes neighbors, the sender will require an acknowledgement for its first message which will ensure that the receiving node is still alive. If a time out occurs, the sending node will choose another neighbor to transmit to and the whole process repeats. Once communication is initiated, there will be no more acknowledgements for any messages. Besides data messages, there is an introduce exception messages which serve as explicit synchronization messages. Only receiver scans issue exception messages, and are primarily used to tell the sending node to stop sending and let the sender choose a different neighbor. An exception message is generated in only three instances: the receiving node's queue is too large, the receiver's power is less than the sender's power, and the receiver has passed a certain threshold which means that it has very little power left.

At any time throughout the system's lifetime, a receiver can tell a sender not to transmit anymore because the receiver's queues are full. This should normally not happen, but in the event it does, an exception message would alleviate the problem. In the current schema, once the sending node receives an exception message and removes his respective neighbor off his neighbor list, the sending node will never consider that same neighbor again. We did this in order to minimize the amount of control messages that would be needed to be exchanged between peer nodes. However, future considerations could be to place a receiving neighbor on probation in the event of an exception message, and only permanently remove it as a valid neighbor after a certain number of exception messages.

The second reason an exception message might be issued, which is the more likely one, is when the receiver's power is less than the sender's power, in which if the receiver's power is less than the specified threshold, it would then analyze the receiving packets for the sender's power levels. If the threshold was made too small, then by the time the receiver managed to react and tell the sender to stop sending, too much of its power supply had been depleted and its life expectancy thereafter would be very limited while the sending node's life expectance would be much longer due to its less energy consumption. Through empirical results, we concluded that the optimum threshold is50% of the receiver's power levels when it in order to equally distribute the power dissipation throughout the network. In order to avoid having to acknowledge every message or even have heart beat messages, we introduce an additional threshold that will tell the receiving node when its battery supply is almost gone. This threshold should be relatively small, in the 5~10% of total power, and is used for telling the senders that

their neighbors are almost dead and that new more suitable neighbors should be elected.

The synchronization cost of e3D is two messages for each pair of neighboring nodes. The rest of the decisions will be based on local look-ups in its memory for the next best suitable neighbor to which it should transmit to. Eventually, when all suitable neighbors are exhausted, the nodes opt to transmit directly to the base station. By looking at the empirical results obtained, it is only towards the end of the system's lifetime that the nodes decide to send directly to the base station.

The main advantage of this algorithm is the near perfect system lifetime where most nodes in the network live relatively the same duration. The system distributes the lifetime and load on the network better than the previous two approaches. The disadvantage when compared to of this algorithm is its higher complexity, which requires some synchronization messages throughout the lifetime of the system. These synchronization message are very few, and therefore worth the price in the event that the application calls for such strict performance.

Random Clustering Based Algorithm

This algorithm is similar to LEACH, except there is no data aggregation at the cluster heads. Random cluster heads are chosen and clusters of nodes are established which will communicate with the cluster heads.

The main advantage of this algorithm is the distribution of power dissipation achieved by randomly choosing the group heads. This yields a random distribution of node deaths. The disadvantage of this algorithm is its relatively high complexity, which requires many synchronization messages compared to e3D at regular intervals throughout the lifetime of the system. Note that cluster heads should not be chosen at every iteration since the cost of synchronization would be very large in comparison to the number of messages that would be actually transmitted. In our simulation, we used rounds of 20iterations between choosing new cluster heads. The high cost of this schema is not justifiable for the performance gains over much simpler schemes such as direct communication. As a whole, the system does not live very long and has similar characteristics to direct communication. Notice that the only difference in its perceived performance from direct communication is that it randomly kills nodes throughout the network rather than having all the nodes die on one extreme of the network.

The nodes that are farther away would tend to die earlier because the cluster heads that are farther away have much more work to accomplish than cluster heads that are close to the base station. The random clustering algorithm had a wide range of performance results, which indicated that its performance was directly related to the random cluster election; the worst case scenario had worse performance by a factor of ten in terms of overall system lifetime.

Ideal Clustering Based Algorithm

The cost of implementing this classical clustering algorithm in a real world distributed system such as wireless sensor networks is energy prohibitively high; however, it does offer us insight into the upper bounds on the performance of clustering based algorithms.

This technique is implemented using k-Means clustering (k represents the number of clusters) to form the clusters. The cluster heads are chosen to be the clustroid nodes; the clustroid is the node in the cluster that minimizes the sum of the cost metric to the other points of the corresponding cluster. In electing the clustroid, the cost metric is calculated by taking the distance squared between each corresponding node and the candidate clustroid and divided by the candidate clustroid's respective power percentage levels.

III. Conclusion

In this paper we have discussed various energy efficient routing protocols for wireless sensor networks. The comparison based of systematic review indicates that the Clustering based routing is best among the others. In contrast: 1). Direct Communication has no synchronization between peer nodes, and perhaps a simple broadcast message from the base station would suffice in establishing the base station identity, Diffusion based routing does not completely distribute the energy dissipated since nodes close to the base

station will die far sooner before nodes far away from the base station and in Random Clustering the nodes that are farther away would tend to die earlier because the cluster heads that are farther away have much more work to accomplish than cluster heads that are close to the base station.

References

1. R. Pichna and Q. Wang, "Power Control", The Mobile Communications Handbook. CRC Press, 1996, pp. 370-30.
2. R. Ramanathan and R. Hain, "Topology Control of Multihop Wireless Networks Using Transmit Power Adjustment", Proceeding Infocom 2000, 2000.
3. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan,"Energy-Efficient communication Protocol for Wireless Microsensor Networks", Hawaiian Int'l Conf. on Systems Science, January 2000.
4. S. Lindsey and C. Raghavendra, "PEGASIS: Power-Efficient GAthering in Sensor Information Systems", International Conference on Communications, 2001.
5. I. Raicu, O. Richter, L. Schwiebert, S. Zeadally, "Using Wireless Sensor Networks to Narrow the Gap between Low-Level Information and Context-Awareness", ISCA 17th International Conference on Computers and Their Applications, CATA 2002.
6. J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, K. Pister, "System architecture directions for network sensors", ASPLOS 2000.
7. A. S. Tanenbaum, Computer Networks, Third Edition, Prentice Hall Inc., 1996, pp. 276-287.
8. G Coulouris, J. Dollimore, T. Kindberg, Distributed Systems, Concept and Design, Third Edition, Addison-Wesley Publishers Limited, 2001, pp. 116-119.
9. C. Intanagonwiwat, R. Govindan and 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, Boston, Massachussetts.

Impact of 3D Printing Technology in Medical Science

Stuti Kapoor*
Aparna Bhardwaj**
Vishakha Verma***
Ankit Verma****

Abstract

In this paper we will discuss about 3D printing in medical science. Nowadays it is the most effective and major concept in medical science. This is mainly used in making organs, skin cells, bone and cartilage and cancer research.

Key Words: 3D Printing, Medical Application, Technology Advancement

I. Introduction

The 3D printing is invented by Charles hull, which is known by "stereolithography". The major advantages are this 3D printing in medical applications are amplify rapidly and are anticipated to revolutionize health care. Medical use of 3D printing, both actual as well as potential, can be systematized in to several broad categories, including customization and personalization and enhanced productivity etc. The application of 3D printing provides many advantages, contain: the customization and personalization of medical products, drugs, and equipment; cost-effectiveness; enlarge productivity; the democratization of design and manufacturing; and improve collaboration [1]. It should be wariness that despite recent significant and exciting medical advances involving 3D printing, notable scientific and inhibit challenges remain and this technology will need time to evolve.

Stuti Kapoor*
Student of BCA,
IITM, Janakpuri, New Delhi

Aparna Bhardwaj**
Student of BCA,
IITM, Janakpuri, New Delhi

Vishakha Verma***
Department of Information Technology,
IITM, Janakpuri, New Delhi

Ankit Verma****
Department of Information Technology,
IITM, Janakpuri, New Delhi

II. 3D Printing

In Three-dimensional (3D) printing the objects are made by fusing or depositing materials—such as plastic, metal, ceramics, powders, liquids, or even living cells—in layers to generate a 3D object [2]. This procedure is also relating to as additive manufacturing (AM), rapid prototyping (RP), or solid free-form technology (SFF). Some 3D printers are resembling to traditional inkjet printers; however, the difference in that is the difference in production of end product in 3D object. 3D printing is look forward in the way to revolutionize medicine and other fields, not unlike the way the printing press transformed publishing.

III. Common Types of 3D Printers

All 3D printing processes offer merits and demerits. The 3D printer which are most frequently used in medical applications are: selective laser sintering (SLS), thermal inkjet (TIJ) printing, and fused deposition modeling (FDM). A brief discussion of each of these 3D printers are follows.

- Selective Laser Sintering

In this technology printer uses powdered material as the substrate for printing new objects. in this the shape of the object drawn by the laser in the powder and fusing it together [3]. Then a new layer of powder is

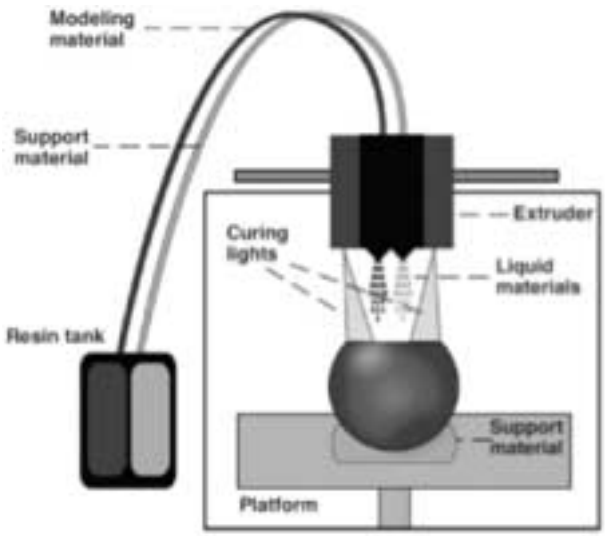


Fig. 1: Thermal Inkjet Printing (TIJ)

pass and the process reiterate, building each layer, one by one, to form the object. This can be used to create metal, plastic, and ceramic objects.

- Thermal Inkjet Printing (TIJ)

This type of printer mainly known for the "noncontact" printing technique that uses thermal, electromagnetic, or piezoelectric technology to deposit tiny droplets of "ink"TIJ printers, the process is done by heating the printhead and creates small air bubbles that go out, creating pressure pulses that eject ink drops from nozzles in volumes as small as 10 to 150 picolitres [4].TIJ printers are especially encouraging for use in tissue engineering and regenerative medicine. Because of their digital precision, control, versatility, and benign effect on mammalian cells, this technology is earlier being applied to print simple 2D and 3D tissues and organs (which is also known as bioprinting). Figure 1 shows thermal inkjet printer. TIJ printers may also uses in, such as drug delivery and gene transfection during tissue construction.

- Fused Deposition Modeling (FDM)

FDM printers are very common and reasonable than the SLS type. An FDM printer uses a similar technique of printing like to an inkjet printer. However, beads of heated plastic are released from the printhead as it moves, assemble the object in thin layers. This process is repeated again and again, allowing precise control of the amount and location of each deposit to shape each layer. Since the material is heated as it is ejected, it fuses or bonds to the layers below [5]. As each layer of plastic cools, it hardens, gradually creating the solid object as the layers build.FDM printers can use different variety of plastics. In 3D FDM printed parts are usually made from the same thermoplastics which are used in traditional injection machining, so they have similar stability, durability, and mechanical properties.

IV. Benefits of 3D Printing in Medical Applications

There are following benefits of 3D printing in medical applications:

- Customization and Personalization

The favorable advantage that 3D printers provide in medical applications is the freedom to produce custom-made medical products and equipment [6]. For example, the 3D printing is used to customize prosthetics and implants can provide great value for both patients and physicians.

- Increased Cost Efficiency

Another important aid offered by 3D printing is the capability to produce items cheaply [7]. Traditional

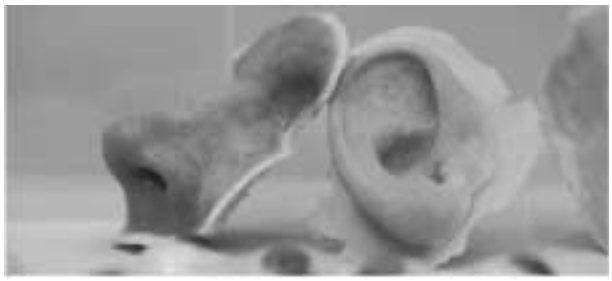


Fig. 2: Printed 3D Organs

manufacturing methods remain inexpensive for large-scale production; however, the cost of 3D printing is becoming enhanced competitive for small production runs.

- Enhanced Productivity

3D printing is a fast and effective process which results in mass production [8]. 3D printing technology is much faster than traditional methods which produces items such as prosthetics and implants, which required milling, forging, and a long delivery time.

- Democratization and Collaboration

3D printing also offers another beneficial feature which is the democratization of the design and manufacturing of goods [9]. This results in major number of people, to use little more than a 3D printer and their imaginations to design and manufacture novel products for individual or commercial use.

Fig. 3: 3D Printed Surgical Tools

V. Advancements of 3D Printing in Medical Field

Some of the most incredible uses for 3D printing are developing within the medical field are:

- Organs

Scientists have successfully printed kidney cells, sheets of cardiac tissue i.e., like a real heart and the foundations of a human liver, among many other organ tissues [10]. But scientists are still working to print human organ for transplant. Figure 2 shows printed 3D organs.

- Stem Cells

Stem cells have miraculous regenerative properties which helps them to reproduce many different kinds of human tissue. Now, stem cells are being bio printed in many university research labs [11]. Stem cell printing prints other kinds of tissues, which lead to printing cells directly into parts of the body.

- Skin

This technique can be used in printing skin grafts for burn victims, skin cancer patients and other kinds of afflictions and diseases that affect the epidermis [12]. Medical engineers in Germany have been developing skin cell bioprinting since 2010.

- Bone and Cartilage

Hod Lipson, a Cornell engineer, prototyped tissue bioprinting for cartilage [13]. Though Lipson has yet to bio print a meniscus that can withstand the kind of pressure and pounding that a real one can, he and other engineers are well on their way to understanding how to apply these properties.

- Surgical Tools

Surgical tools like forceps, hemostats, scalpel handles and clamps can be printed by this technique [14]. They come out of the printer sterile and cost a tenth as much as the stainless-steel equivalent.Figure 3 shows the 3D printed surgical tools.

- Cancer Research

Disease cells and cancer cells can also be bio printed just like tissue and some organ cells are printed and studied, in order to study how tumors, grow and develop [15]. Such medical engineering also helps for better drug testing, cancer cell analyzing and therapy development. One day this technique may help in the cure for cancer.

VI. Conclusion

A significant step forward for healthcare with huge benefits to our practice and patients, and it should be made part of the patient pathway to ensure patients get the best and safest possible treatment.

References

1. Kodama H., Automatic Method for Fabricating a Three Dimensional Plastic Model with PhotoHardening Polymer, Rev Sci Instrum, pp.1770-1773, 1981.
2. http://www.pcmag.com/slideshow_viewer/0,3253,l=293816&a=289174&po=1,00.asp.
3. Tabusca Silvia, The Internet access as a fundamental right, Journal of Information Systems and Operations Management, Vol.4, No.2/2010, Universitary Publishing House, Bucuresti, ISSN1843-4711.
4. Tabusca A., Electronic Online Advertising – The "Economic Crisis" is just the pretext of thefall, Journal of Information Systems and Operations Management, Vol. 6, No.2/2012, pp. 410-418, Universitary Publishing House, Bucuresti, ISSN 1843-4711.
5. Evans B., Practical 3D Printers: The Science and Art of 3D Printing, Apress Publisher, 2012.
6. M. Salmi, Medical applications of additive manufacturing in surgeryand dental care [Ph.D. thesis], Aalto University, 2013.
7. J. Tuomi, K.-S. Paloheimo, J. Vehvilainen et al., "A novel classification and online platform for planning and documentationof medical applications of additive manufacturing," SurgicalInnovation, vol. 21, no. 6, pp. 553–559, 2014.
8. N. Hopkinson, R. Hague, and P. Dickens, Rapid Manufacturing:An Industrial Revolution for the Digital Age, John Wiley & Sons,New York, NY, USA, 2006.
9. I. Gibson, D. W. Rosen, and B. Stucker, Additive ManufacturingTechnologies: Rapid Prototyping to Direct Digital Manufacturing,Springer, Berlin, Germany, 2010.
10. A. Makitie, K. S. Paloheimo, R. Bj ¨ orkstrand et al., "Medicalapplications of rapid prototyping–three-dimensional bodiesfor planning and implementation of treatment and for tissuereplacement," Duodecim; Laaketieteellinen Aikakauskirja, vol.126, no. 2, pp. 143–151, 2009.

11. J. Tuomi, K. Paloheimo, R. Bjorkstrand, M. Salmi, M. Paloheimo, and A. A. Makitie, "Medical applications of rapid prototyping—from applications to classification," in Proceedings ofthe 4th International Conference on Advanced Research in Virtualand Physical Prototyping (VRAP '09), pp. 701–704, Leiria,Portugal, October 2009.

12. M. Salm, K.-S. Paloheimo, J. Tuomi, T. Ingman, and A. Makitie,"A digital process for additive manufacturing of occlusal splints:a clinical pilot study," Journal of the Royal Society Interface, vol.10, no. 84, Article ID 20130203, 2013.

13. L. Joffe, "Features section: current products and practice invisalign:early experiences," Journal of Orthodontics, vol. 30, no. 4,pp. 348–352, 2003.

14. M. Salmi, J. Tuomi, R. Sirkkanen, T. Ingman, and A. Makitie,"Rapid tooling method for soft customized removable oralappliances," Open Dentistry Journal, vol. 6, no. 1, pp. 85–89, 2012.

15. J. W. Chang, S. A. Park, J.-K. Park et al., "Tissue-engineered tracheal reconstruction using three-dimensionally printed artificial tracheal graft: preliminary report," Artificial Organs, vol.38, no. 6, pp. E95–E105, 2014.

Big Data Analytics Approach using Mapreduce Algorithm on Hadoop

Harsh Arora*
Mayank Gupta**
Himanshi Kohli***

Abstract

In this fast pacing world enormous amount of data is been collected. To manage this big data different applications are implementing. However, analyzing big data is a very laborious problem today. The challenges of big data are analysis, storing, visualization etc. Hadoop is an open source framework, which is based on Java programming language that supports the storage and processing of extremely large amount data sets in a distributed computing environment. Storing and managing the collected big data can be done using "hadoop's MapReduce programming approach". But, coding and understanding java MapReduce application is difficult and hard to maintain. Hive is a relatively simple tool for analyzing big data. The paper throws light on the concept of hadoop for handling big data, analyzing of big data using MapReduce and Hive.

Key Words: Big Data, Big Data Analytics, Hadoop, Mapreduce

I. Introduction

Everyday thousand and millions of people take photos, make videos and send emails, messages. Across the world businesses collect data on consumer preferences, purchases and trends. Governments regularly collect data of all categories from census data to incident reports in police departments. This ocean of data is keep going expanding day by day. As per the record of 2013 amount of collected data was 4.4 zetabytes .And that will be continuously rising to 44 zettabytes by 2020. To put that in perspective, one zettabyte is equivalent to 44 *trillion* gigabytes.

Within last two years, organizations created 2.5 Quintilian bytes of data which is just 90% of the total data created in the world. This data is what organizations collect on daily basis [2]. The generated data is and in different formats. The term given for such huge data is **"big data".**

The need of organizing the big data into meaningful and valuable information is being realized by industries, organizations and companies. But, the obstacle comes between the handling of big data is that it's difficult to handle such large amount of data effectively and efficiently using traditional methods. Then new tools, technologies, models and methodologies are applied to handle big data. Hadoop is an open source java programming based framework that is being majorly used for processing big data. It is a prominent distributed storage and compute environment which is used for storing and processing of big data.

Hadoop uses MapReduce algorithm as it's technique which analysis big data. MapReduce has emerged as a new paradigm for large-scale data analysis due to its high scalability, fine-grained fault tolerance and easy programming model. With the combination of two distinct tasks map and reduce the term mapreduce is actually generated.

Harsh Arora*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Mayank Gupta**
Student of BCA,
IITM, Janakpuri, New Delhi

Himanshi Kohli***
Student of BCA,
IITM, Janakpuri, New Delhi

II. Big Data

Applications are generating the large volume of data in a small time interval and it is quite complex and time consuming to analyze this data for decision making. Traditional data processing applications are not suitable for handling this data, so researchers introduced the concept of Big Data and developed some frameworks to process the large scale data efficiently by considering the common issues which are following:

- How much data must be selected from a large scale data block?
- Storage Space is essential or not
- Security
- Data Validation

Big Data is a huge collection of data which is generated at an exponential rate in a wide variety of formats and has become hard to handle using traditional data management tools. The theory of big data is based on five V's:

- **Volume:** Large volume of data generated every second by individuals, organizations, machines etc.
- **Velocity:** Speed at which data is being generated.
- **Variety:** Various formats in which the data is available (text, blogs, tweets, video, barcode, databases etc.).
- **Veracity:** Correctness and accuracy of data.
- **Value:** Insights or information that may be generated by applying analytics on big data.

The interest of organizations in big data has risen due to the value it may generate for their businesses and researches. Organizations want to expand, make better business decisions and create new products and services; big data plays a major role in this. With large amounts of data spanning from user buying trends, to twitter tweets, the data holds valuable information. Proper extraction and analyzing of this data may reveal insights in future and help organizations take profitable business decisions or create actionable intelligence.

III. Astonishing Facts about Rise of Big Data Collection

- Consistently, each day purchasers make around 11.5 million installments by utilizing Paypal
- Consistently, each day Walmart (chain of markdown retail establishments) handles more than 1 million client exchanges
- 510 remarks, 293000 status and 136000 redesigns are posted on Facebook consistently
- Consistently, each day ~7000 tweets are made on Twitter.

IV. Big Data Analytics

Big Data Analytics (BDA) is the process of applying advanced analytic techniques on large varied data sets in order to gather data and find hidden patterns that may help analysts, businesses and researchers in making better decisions. Traditional analytics deals with structured, transactional data collected over a period of time, in data warehouses for performing Business Intelligence (BI). A BI analyst focuses on finding trends, generating reports and visual analysis of data. In BDA, data scientists, predictive modelers and other analytics professionals analyze large volumes of transactional, as well as, data of other forms, collected from different types of sources that may remain untapped by conventional business intelligence programs. These data forms include web server logs, internet click stream data, social media content, social network activity reports, patient's health records, text from customer emails, survey responses, mobile phone call detail records, and machine data captured by sensors connected to internet of things. BDA can be performed on different types of data like, text, image, clicks, logs and blogs to reveal insights about behavioral patterns of customers/ users/clients, optimizing performance, taking smart business decisions, predicting future values, preventing diseases, combating crime, reducing frauds, and mitigating risks.

V. Hadoop

Apache Hadoop is an open source framework which

Fig. 1: Big Data Analytics

is used by many organizations such as google, facebook, linkedin yahoo, twitter etc large datasets is being processed in distributed data with the help of this software.

It is used by the organizations to improvise their user experience, get feedback and build new services and products.

The quality of flexibility and scalable architecture makes it popular as it stores and processes big data on commodity hardware machines. It allows distributed processing of large data sets on cluster of nodes.

It supports operations in distributed environment. It consists of various modules:

1. **Hadoop Common** – Java libraries and utilities required by other Hadoop components.
2. **Hadoop Distributed File System (HDFS)** – It stores huge amount of data. It is basically, the storage component of Hadoop a distributed file system which stores huge data sets in multiple machines.

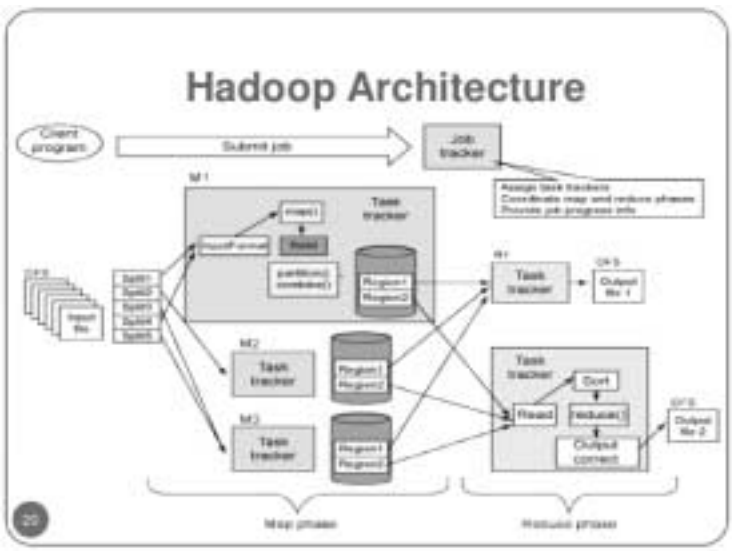


Fig. 2: Hadoop Architecture

3. **Hadoop YARN** –Resource management, process scheduling and user application management are the responsibility of Hadoop Yarn.
4. **Hadoop MapReduce** –Programming interface to process large scale data.

Hadoop Merits

It supports:

1. Scalability for data processing
2. Low cost data analysis
3. Robust Fault Tolerance

Hadoop ecosystem contains other tools like pig, sqoop, hive. These tools make it easy to address particular needs of users in an easy way.

Pig :- It is used for analyzing data with high level scripting language . it enables workers to write complex data without knowledge of java. Pig's simple SQL – like scripting language is called pig latin.

Sqoop :- efficient transfer of data b/w Hadoop & relational database is intended by this tool.

Hive :- This Tool is used for analyzing data.

The architecture of these tools can be viewed

Examples of Hadoop API's Are:-

org.apache.hadoop.io

org.hadoop.apache.mapred

org.apache.hadoop.io.lib etc.

How do technologies such as MapReduce and Hadoop help organizations harness the value of unstructured and semi structured data?

MapReduce supports distributed processing of the common map and reduction operations. In the map step, a master node divides a query or request into smaller problems. It distributes each query to a set of map tasks scheduled on a worker node within a cluster of execution nodes. The output of the map steps is sent to nodes that combine or reduce the output and create a response to the query. Because both the map and reduce functions can be distributed to clusters of commodity hardware and performed in parallel, MapReduce techniques are appropriate for larger datasets.

Apache Hadoop consists of two components: Hadoop MapReduce for parallel data processing and the Hadoop Distributed File System (HDFS) for low-cost, reliable data storage. Hadoop, the most popular open-source implementation of the MapReduce framework, can be used to refine unstructured and semi-structured data into structured formats that can be analyzed or loaded into other analytic platforms.

VI. Mapreduce

To process extremely large volumes of data in parallel mode by splitting the job into various independent tasks, a programming paradigm is designed which is popularly known as MapReduce. Mapreduce is an efficient algorithm to assemble big data in an effective way. It has majorly two functions: -

- Map () function
- Reduce () function

The job of map () function is to filter and sort operations.

The job of reduce () function is to perform a summary / aggregate operation.

The "MapReduce System" is popularly known as MapReduce "framework" or "architecture" which works in a proper framework i.e. demonstrate the processing with the distributed servers, running the various tasks in parallel, managing the communication and transfers the data between the various parts of the system, and providing for unnecessary data.

It processes coarse data, split and distribute it across huge datasets with the help of a large number of computers (nodes). The group of nodes collectively treated as a cluster, if all nodes are with similar hardware configurations working on the same local network or else the nodes are treated as a grid, if they are geographically shared and distributed with varying hardware specifications. Processing may occur on the data that is stored either in system log files (unstructured) or in a database (structured). MapReduce minimize the data transfer distance with the help of locality of data.

Map Phase: In the map phase of mapreduce, the master node takes the input, divides it into smaller sub-tasks, and distributes them to worker-nodes. A worker-node may do this again repeatedly and continuously, leading to a multi-level tree structure. The worker node processes the sub-tasks only assigned to it and then passes the intermediate result back to its master node.

Reduce Phase: During the reduce phase, the master node collects all the intermediated outputs of all the sub-tasks generated by various worker nodes and combined them to form the final output – the solution to the problem it was originally trying to solve.

a) **Input reader**: The input file is broken into appropriate sizes (in practice typically 64 MB to 512 MB as per HDFS) by the input reader and one split is assigned to one Map function by MapReduce framework. It takes input from stable storage (typically as in our case Hadoop distributed file system) and generates the output as key/value pairs.

b) **Map function**: Each Map function takes a series of key/value pairs generated by the input reader, processes each, and in turn produces zero or more output key/value pairs. The input and output types of the map can be different from each other.

c) **Partition function**: Each Map function output is assigned to a particular reducer by the application's partition function for sharing purposes. The partition function is given as input key and the number of reducers and it return the index of desired reduce.

d) **Comparison function**: The input for every Reduce is collected from the machine where the Map run and sorted using comparison function.

e) **Reduce function**: The framework calls the applications Reduce function for each unique key in the sorted/assembled order. It also iterates through the values that are associated with that key and produce zero or more outputs.

f) **Output writer**: It writes the output of the Reduce function into a stable storage, usually a Hadoop distributed file system.

Understanding Mapreduce

Map reduce is a programming approach of Hadoop which is used for distributed computation. Mapreduce algorithm splits the work already submitted by client into small parallelized map and reduced tasks. The user needs to specify a map function which the mapper class processes as a key/value pair and generates a set of intermediate key/value output pairs. The reducer class then sums up the intermediate key/value output pairs. Afterwards , it aggregates the intermediate key / value the intermediate key/value output pairs produced earlier and generates a final key/value output pairs.

Mapper Class

The mapper class takes input as a <key value> pair and the output as<key value> pair. It can be created by extending mapper classes by a predefined class called mapper with specified input and output formats. The functionality of mapper is defined in the map function. The following program syntax shows how to create user defined mapper class in Hadoop environment.

[2]Public class classname extends mapper< InputKeyDataType,InputValueDataType, Output KeyDataType,OutputValueDatatype>
{//starting of mapper class…….
……
Protected void map (InputKeyDataType argument, InputDatatypearguement)}

Reducer Class

A reducer class can be created by extending a predefined class with specified input and output formats. The functionality of reducer is defined in the reducer function. The following program shows up how to create reducer class into Hadoop environment.

Public class classname extends reducer< InputKey DataType, InputvalueDataType, OutputKey DataType, OutputValueDataType>
{//starting reducer class
……..
Protected void reduces (InputKeyDataType, InputvalueDataType) throws exception
{ //starting of reducer function
//reducer logic
}//ending reducer function
}//ending of reducer class
[2]

Table-1: Mapper and reducer input/output key/values

function	input	Output
Map	<key1,value1>	List(<k2,v2>)
reduce	<k2,list(v2)>	List<k3,v3>

Table1 indicates the input/output key/values for map and reduce functions. The input to an application is a set of keys and values processed by map function. The reducer takes the k2 and v2 values as input, processes them and generates a list of keys and values as an output, and generated a list of k2 and v2. The word count application is the best example for understanding MapReduce approach.

VII. Conclusion

This paper presents a study on Big Data, Big data Analytics, Hadoop and Mapreduce. It throws light on the Hadoop architecture and various tools used in Hadoop. It also discusses about the map() and reduce() functions in detail and how to help organizations harness the value of unstructured and semi structured data.

References

1. http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/
2. CSI communications magazine on big data, volume no.40, November 2016
3. https://www.bigdatanews.datasciencecentral.com/profiles/blogs/how-to-collect-big-data-big-data-a-new-digital-trend

A Review on Exploring the Recent Trends and Future Scope in Artificial Intelligence

Amanpreet Kaur*
Isha Shingari**

Abstract

Artificial Intelligence (AI) is the study and creation of computer systems that can perceive and act. The primary aim of AI is to produce intelligent machines so that our future becomes easier. The intelligence should never incorporate thinking, making decisions, solving problems, most important is learning. AI is an interdisciplinary field that requires knowledge in computer science, linguistics, psychology, mathematics, philosophy and many more. AI can also be defined as the area of computer sciences that deals with the ways in which computers can be made to perform intellective functions imputed to humans. But this definition does not say what functions are performed, to what degree they are performed, or how these functions are carried out using AI.

Key Words: Artificial Intelligence, Human Intelligence, Future of AI, Limitations of AI in Business, Cognitive Science, Era of automation

I. Introduction

It is the theory and development of computer systems able to perform Tasks normally requiring human intelligence such as visual perception, speech recognition, decision making and translation between languages. It is generally displayed by machines, in contrast with natural intelligence displayed by humans and other animals.

In computer sciences, AI is the study of intelligent agents any device that perceives its chances of success at same goals. AI is applied when a machine mimics "cognitive" functions that humans associate with other human minds such as learning and problem solving.

Optical character recognition is frequently excluded from AI because it became a routine technology AI has reached to the ninth cloud , it is touching heights the achievements consists of understanding human speech, strategic game competitions (we play games in our laptops, mobiles which have an option of playing with the system)autonomous cars , military simulations and interpreting complex data.

AI was founded as an academic intelligence in 1956 and in the years since has experienced several waves of optimism followed by disappointments and loss of funding which was known as "AI WINTER" but it was then followed by new approaches, success and renewed funding

The field was founded on the claim that human intelligence can be so precisely described that a machine can be made to simulate it.

Some people consider AI to danger to humanity if it progresses unabatedly. Others believes that it is primarily a risk to employment.

In 21 century AI techniques have experienced a resurgence following concurrent advances in computer power large amounts of data and theoretical understanding and AI Techniques have become an essential part of technology industry, helping to solve many challenging problems in Computer sciences. [3][6]

Amanpreet Kaur*
Student of BCA,
IITM, Janakpuri, New Delhi

Isha Shingari**
Department of Information Technology,
IITM, Janakpuri, New Delhi

AI pays heavily on following domains of study.

- Computer Science
- Cognitive Science
- Engineering
- Ethics
- Linguistics
- Logic
- Mathematics
- Natural Sciences
- Philosophy
- Physiology
- Psychology
- Statistics

II. Strong Artificial Intelligence

It deals with creation of real intelligence artificially. Strong AI believes that machines can be made sentient or self-aware. There are two types of strong AI: Human-like AI, in which the computer program thinks and reasons to the level of human-being. Non-human-like AI, in which the computer program develops a non-human way of thinking and reasoning.

III. Weak Artificial Intelligence

Weak AI does not believe that creating human-level intelligence in machines is possible but AI techniques can be developed to solve many real-life problems. That is, it is the study of mental models implemented on a computer.

IV. AI and Nature

Nowadays AI techniques developed with the inspiration from nature is becoming popular. A new area of research what is known as Nature Inspired Computing is emerging. Biological inspired AI approaches such as neural networks and genetic algorithms are already in place.

V. Challenges

It is true that AI does not yet achieve its ultimate goal. Still AI systems could not defeat even a three year old child on many counts: ability to recognize and remember different objects, adapt to new situations, understand and generate human languages, and so on. The main problem is that we, still could not understand how human mind works, how we learn new things, especially how we learn languages and reproduce them properly.

VI. Future of AI

AI is the best field for dreamers to play around. It must be evolved from the thought that making a human-machine is possible. Though many conclude that this is not possible, there is still a lot of research going on in this field to attain the final objective. There are inherent advantages of using computers as they do not get tired or loosing temper and are becoming faster and faster. Only time will say what will be the future of AI: will it attain human-level or above human-level intelligence or not.[3][5]

VII. Proposed Method And Educational Utilization

The proposed method in this paper is the electronic content of lessons is designed based on educational books content in this way, it is focused on traditional educational system in which teacher is the main factor in the classroom. The content should presented as sketch in class, the purpose is to make studying very much interesting. The main point in this method is E-content management that is important to pay special attention to content and individual differences.

VIII. Result and Proposals

Undoubtedly handling artificial intelligence through teaching and learning process is much important, of course, the quality of electronic content is of importance. By information technology progress, artificial intelligence is well used in various fields, including electronic trading. Considering individual differences and consumers needs is of course something common. Researchers have shown that, leaving school in the electronic schools is most common than traditional schools, the main reason is how to present electronic content. It is recommend, being intelligent through education, to produce good quality electronic content a unique management for feeding educational

sources which approved for each education unit, is needed learning during teaching for teachers and college professors must be planned, it is expected that, fantastic progress of information technology in the version of web 2 in our educational system beingachieved.

IX. Risk and Limitations of Artificial Intelligence in Business

Artificial intelligence (AI) involves giving machines and programs the ability to think like a human. Businesses are increasingly looking for ways to put this technology to work to improve their yield, profitability and business results.

Where there are many business benefits of AI, there are also certain loop holes to keep in mind while operating on them.

A. Limitations of artificial intelligence

One of the main limitations of AI is the cost. Creation of smart technologies can be expensive, due to their complex nature and the need for repair and ongoing maintenance.

Software programs need regular elevation to adapt to the changing business environment and in case of disintegration a risk of losing code or important data revolves around us. Restoring this is often time-consuming and extravagant.[8][7]

B. Other AI limitations relate to:

Implementation times, which are often lengthy

- Integration challenges and lack of understanding of the state-of-the-art systems
- Usability with other systems and platforms
- If we're deciding whether to take on AI-driven technology, we should also consider:
- Customer privacy
- Potential lack of transparency
- Technological complexity
- Loss of control over your business decisions and strategy[9]

C. AI and ethical concerns:

With the rapid development of AI, a number of ethical issues have cropped up. These include:

The potential of automation technology to give rise to job losses

The need to redeploy or retrain employees to keep them in jobs

Fair distribution of wealth created by machines

The effect of machine interaction on human behavior and attention

The need to eliminate bias in AI that is created by humans

The security of AI systems (e.g. autonomous weapons) that can potentially cause damage. While these risks can't be ignored, it is worth keeping in mind that advances in AI can - for the most part - create better business and better lives for everyone. If implemented responsibly, artificial intelligence has immense and beneficial potential [2][8]

X. Future Scope

During the next decade (2014–2024), mobile robots and artificial intelligence make it likely that occupations employing about half of today's US workers could be automated to some degree.

The centuries long quest to develop machines and software with human-like intelligence inches closer to reality. Scientists develop intelligent machines that can simulate reasoning, develop knowledge, and allow computers to set and achieve goals, moving closer to mimicking the human thought process. These intelligent systems improve accuracy of predictions, accelerate problem solving and automate administrative tasks bringing in an era of automation.[8][4]

XI. Artificial Intelligence in 2020

Cognitive analytics, where machines learn from experience and build associations, help develop technology systems that evolve hypothesis, draw conclusions and codify instincts and experience.

Parallel information processing, aided through chips custom designed for AI applications, help parallel processing of vast amounts of data.

Smarter gets redefined with the advances in sensor, cloud and machine learning technology, and pushes the boundary of smarter homes, cars, infrastructure and just about everything.

Deep learning approaches allow processing of raw data including images, speech and natural language; thus providing deeper insights.

Face-reading machines decipher micro facial expressions to build meaningful information on the emotional state of the user, improving human-computer interaction in areas of e-learning and e-therapy.

Intelligent automation combines automation with artificial intelligence that allows knowledge workers, from physicians to investment analysts to plant supervisors, to process, understand and use ballooning volumes of information.[4]

XII. Conclusion

We shall conclude with a hope of increment in the future scope of AI without much limitations. The field of AI gives the ability to the machines to think analytically, using concepts.

Tremendous contribution to the various areas has been made by AI techniques from the last 2 decades .AI will continue to play an increasingly important role in the various fields .This paper is based on the concept of AI, domains of AI on which the heavily pays ,AI & nature, AI & ethical concerns ,Limitations of AI in Business, what all achievements will AI make in the upcoming ERA. It also includes strong AI, weak AI, future of AI, AI WINTERS.

There is a very bright future in the analysis of network systems. We conclude that further researches in this area can be done as there are very promising & profitable results that are obtained from such techniques, countries can be safer, and cyber crimes can reduce to an extent, human efforts can be reduced by AI. It will exceed its limits in Medical Sciences too. Where scientists haven't yet realized the full potential & ability of AI. This technology & its applications will succeed & will have far reaching effects on human life in years on board.

References

1. Stuart J. Russell, Artificial Intelligence: A Modern Approach (3rd Edition)
2. M. Tim Jones, Artificial Intelligence: A Systems Approach (Computer Science)
3. Patrick Henry Winston, Artificial Intelligence
4. http://government-2020.dupress.com/driver/artificial-intelligence/
5. D. Geman and B. Jedynak, 1986. An active testing model for Tracking roads in satellite images. IEEE tran. on Pattern Analysis and Machine Intelligence, 18(1), January 1996.
6. Zandi,Bahman&MasomiFard,Marjan&MasomiFard,Mitra,," Educational planning in the context of e-learning systems". Quarterly training strategies.2012
7. Rosenberg M. (2007)"The E-Learning: Strategies for delivering knowledge in the digital age. New Jersey: Prentice Hall Publication
8. https://www.nibusinessinfo.co.uk/content/risks-and-limitations-artificial-intelligence-business
9. http://fsroundtable.org/cto-corner-artificial-intelligence-use-infinancial-services/[10]https://www.microsoft.com/en-us/research/search/?q&research_area=13556&content-type=job-opportunities.

Effective Design Prevention Approach in Software Process for Achieving Better Quality Levels

Kanika Bhalla*
Shivangi Kharbanda**

Abstract

Software is the need of the hour i.e. they have become our necessity. Examples of software are Mac OS, Linux and windows or any other software. The word software was discovered in the late 1960s to show the difference from computer hardware. Software is intangible i.e. it can be seen but cannot be touched. The preventative maintenance comes in good software that includes a well-defined backup plan, measures that ensure complete security of the software. The aim of maintenance is to reduce the occurrence of any type of failures thereby extending the system life and prevent data loss. This article focuses to present various measures that to be taken care of while creating software.

Key Words: Data Security, Software

I. Introduction

Software is created using a special type of engineering known as Software Engineering. Software engineering is used to give architectural models and programmed instructions that helps in the production of well-documented maintainable and dynamic software in a manner such that it can be predicted. Software is manufactured using a brief process cycle called Software Development Life Cycle (SDLC). SDLC is a well-described and briefed, ordered sequence of stages that should be present in software to develop a most suitable product.

II. Software Development Life Cycle

SDLC is a well-described and briefed, ordered sequence of stages that should be present in software to develop a most suitable product.

SDLC is a pre-defined strategy followed by a software developing organization to fulfill the needs of the clients. It consists of a detailed in formatted plan that accurately describes the parameters like how to develop and maintain software. The cycle also defines a brief tack-tactics for the overall development process of software.

Kanika Bhalla*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Shivangi Kharbanda**
Student of BCA,
IITM, Janakpuri, New Delhi

III. Measures to be taken Care Offwhile Developing a Software

Accurate Requirement Gathering and Analysis

It should be taken care that the requirement gathering process should be done step by step following all the necessary procedures. An organization should precisely gather all the requirements of the clients i.e. what type of software a client wants. If this process is not carried out correctly, then there are high chances of software crisis and the whole development process will become very time consuming and will incur huge cost.

Smooth Designing of Software

The designing phase begins when the requirements are well understood i.e. the requirement analysis process has been done timely and precisely. In this phase, designing team of a organization draws various UML diagrams, decides which technology is best suitable for the use for development of the particular software, Makes a list of all the required resources of both type i.e. hardware and software, Dead Line to

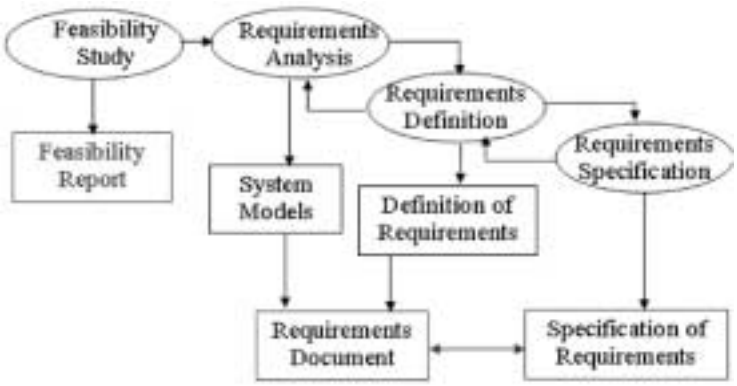


Fig. 1: Feasibility Report

Fig. 2: SDLC

serve the end product to the client. The design phase can also be referred to as the transformation phase, because this is the phase where an idea is actually transformed into a real working system. All the components of the software and security modules of the system are also determined during the design phase of the software. Rough functioning of this design process leads to various factors like over budgeting of the software.

Precise Coding and Implementation

This is the phase where the actual code is developed. Code is written following the Schema decided during taken the design phase. Developers perform Unit Testing by testing each module separately, rectifying and removing errors one by one. Integration testing is done by separate team of testers to ensure smooth functioning of the code. Once the development is completed, the product is sent to third party testers for regression testing. The basic role of coding phase is to convert designed code into implemented code using the programming language which is decided in designing phase. That is why it becomes important to carry out the processing of this phase accurately using well experienced team of developers and a well-designed layout.

Must Ensure that all types of Testing are performed Accurately

After the code is developed by the team of developers, it is tested under various vigorous parameters to ensure the product quality and to check that if it is actually

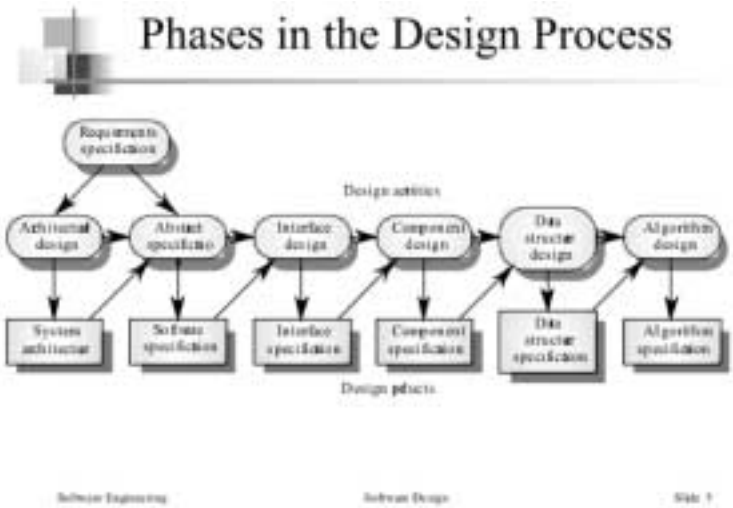


Fig. 3: Phases in Design process

made as per the users requirements. Various types of testing is performed like unit testing by the developers, integration and system testing by another team of testers within the organization, acceptance testing and regression testing by testers outside the organization(QA Engineers) to ensure a quality product as per users expectation. These engineers carry out both positive as well as negative testing's to ensure outstanding quality software. If any type of flaws is detected during this phase, then the product is resent to the development team to fix them. Once the flaws are removed, testers check it again and perform regression testing using various test cases. Once all is good, the product is ready; it is to delivered to the client. Before testing phase begins, a brief test plan is developed by the project team. This test plan includes the all types of testing that are to be performed, resources required for testing, how the software will be tested under various parameters by using test cases, who will be the testers during each phase, and test scripts i.e the instructions that are to be obeyed by each tester to test a particular software.

IV. Regression Testing

Software inevitably changes, how so ever well written and designed it may be initially. This modified software

Fig. 4: Software Testing Phases

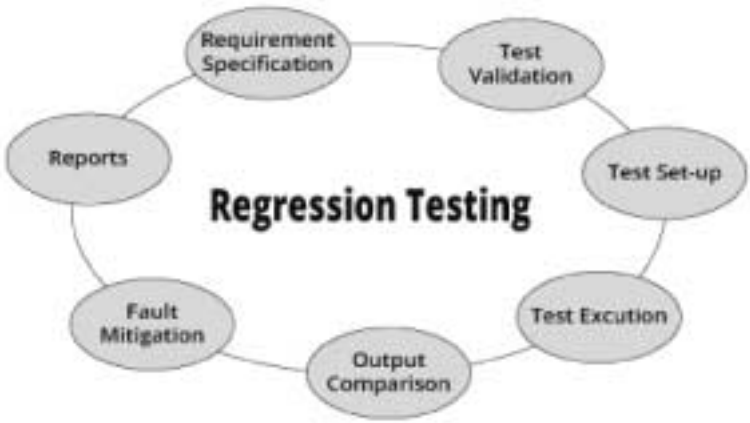


Fig. 5: Regression Testing

needs to be retested again under all types of testing to done changes that are suggested by the client or outside testers to ensure the correctness of the changes done and these changes must not adversely affected other parts of the modified software. This is just because small changes in one part of the software may affect other parts of the software as they work as a whole to ensure smooth functioning of the software.

When we develop software, we perform development testing i.e. unit testing to obtain confidence in the correctness of the software.

V. Risk Analysis

Risk is an expectation of loss, a potential problem that may or may not occur in the future. It is generally caused due to lack of information, control or time. A possibility of suffering from loss in software development process is called a software risk. Loss can be anything, increase in production cost, development of poor quality software, not being able to complete the project on time. A software risk is of two types: internal risks are within the control of the organization i.e. they occur when testing is performed and external risks are beyond the control of the organization i.e. they arise when user uses the software. Risk analysis is done to:

- Identify the risk and reduce its impact
- Reduce the chances of occurrence of risk
- Risk and performance monitoring of the software

VI. Conclusion

In this paper we reviewed and discussed the preventive measures to be taken care of in the process of software development process. Firstly we discussed what is a software and how it is developed.

Then we discussed Software Development Life Cycle i.e. SDLC(the process of software development).Then we discussed the measures to be taken care of and finally we discussed the risks that is risk analysis that is performed while designing a software.

References

1. Google.com
2. http://www.testinstitute.org/What_Is_Software_Risk_And_Software_Risk_Management.php
3. https://www.ifixit.com/Wiki/Preventative_Computer_Maintenance

A Descriptive Framework of Data Mining Approach and its Techniques in Multiple Fields

Govind Murari Upadhyay*
Shivangi Kharbanda**
Shubhangi Khandelwal***

Abstract

Managing and analyzing this data is the need of the hour. This convulsively growing huge hub of data makes our present a truly data age. Applications like security, medicine, science and engineering, remote sensing, process measuring are such examples. The storage, processing and aggregation of such data is highly difficult, complex and a very challenging task. This article focuses to present data mining techniques in different fields.

Key Words: Data, Data Mining, Data Warehouse

I. Introduction

A search engine receives hundred million of queries each day. To deal with this huge data amount, search engine uses distinct patterns to analyze the data. Mining is a vivid term featuring the process of finding a small set of precious data from a heap of raw data. Alternative phrase for data mining is data dredging. Data mining techniques are suitable for simple andstructured data sets.

Based on previous researches, because of data stream requirements, it is necessary to design new techniques to replace the old ones. The first challenge is to design fast and light mining methods for data streams. Another challenge is that the stream mining algorithms need to detect promptly changing concepts and data distribution and adapt to them.

Govind Murari Upadhyay*
Department of Information Technology,
IITM, Janakpuri, New Delhi

Shivangi Kharbanda**
Student of BCA,
IITM, Janakpuri, New Delhi

Shubhangi Khandelwal***
Student of BCA,
IITM, Janakpuri, New Delhi

II. Data Warehousing

Data warehouses are the warehouses of data that provide tools for systematic organizing, understanding and using the data to make strategies and based on that strategies it helps to take decisions. Data warehouses construction involves cleaning of data, data integration, and transformation of data. data warehouses refer to the repository of data different from operational database.

These four keywords data warehouses from the other repository systems.

SUBJECT ORIENTED: A data warehouse is generally organized around many subjects like in that of a company warehouse. Rather than focusing on operations andorganization on transactions, a data warehouse focuses on analysis of data.

INTEGRATED: A warehouse is constructed by integrating several heterogenous sources such as databases, files and previous records.

TIME VARIANT: Data is stored so that it can be processed to extract information from it. Every data warehouse contains a unique time element.

NONVOLATILE: A data warehouse is always a separate unit from the operational unit that's why it

Fig. 1: Data Mining Process

doesn't need any processing and control mechanisms. It only requires two type of operations: loading and accessing of data.

III. Data Mining Techniques

There are various major data mining techniques that are in developing phase and are used in data mining projects like association, classification, clustering, prediction, sequential patterns and decision tree, etc.

IV. Types of Data that can be Mined

Mining Sequence Data:

A sequence is a list of events that are ordered. A Sequence is classified into three things, based on the nature of the events they depict: time-series , Symbolic sequence data and Biological Sequence data.

- Time-series data (e.g. stock market data): In this type of data, sequence data consists of long numeric data recorded at equal intervals of time.it can be generated by many natural and economic processes such as stock markets, and scientific, medical, or natural observations.
- Symbolic sequence data (e.g. customer shopping sequences): It consists of nominal data or event of long sequences, which typically are not observed at equal intervals of time. For many such sequences, gaps don not matter much.
- Biological sequences (e.g. DNA and protein sequences): It includes sequences of DNA and protein. Typically, those sequences are very long and carry important, complicated, but hidden semantic meaning. Here gaps are usually important.

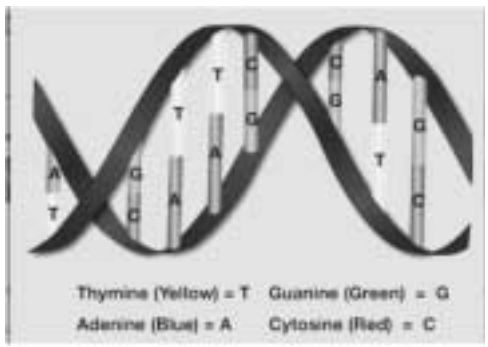


Fig. 2: DNA Sequences

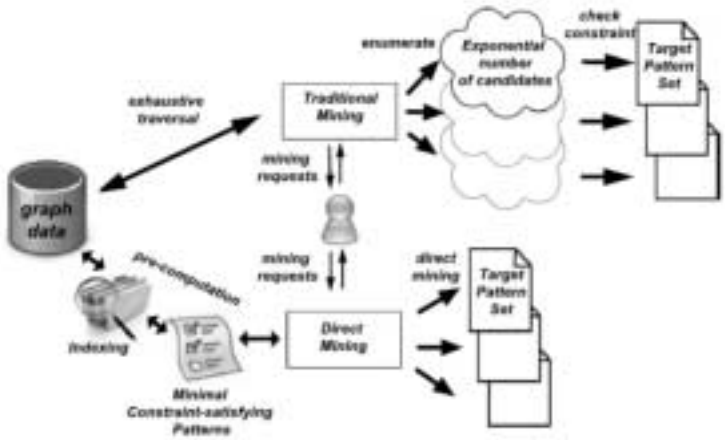


Fig. 3: Graph Pattern Mining

Graph pattern Mining:

Graph pattern mining is the mining of frequent subgraphs in one or a set of graphs. There are two methods for mining graph patterns: Apriori-based and pattern growth-based approaches. Alternatively, we can mine the set of closed graphs where a graph g is closed if there exists no proper supergraph g' that carries the same support count as g. Also, there are many diverse patterns of graph, which includes: approximate frequent graphs, coherent graphs, and dense graphs.

Spatial Data Mining:

This type of mining discovers patterns and knowledge from spatial data. Spatial data refers to the data that is stored in geospatial data repositories. We can construct spatial data cubes that contain spatial dimensions and measures, and support spatial OLAP for multi-dimensional spatial data analysis. This type of mining can be performed on spatial data warehouses, spatial databases and other geospatial data repositories.

Cyber-Physical System(CPS) Data Mining:

A CPS system consists of large quantity of interacting physical and informational components. Example of a CPS is patient care system. Data generated in CPS is inconsistent, interdependent, dynamic, volatile and noisy. This type of mining requires linking a large information base with the current situation, performing real-time calculations, and returning prompt responses.

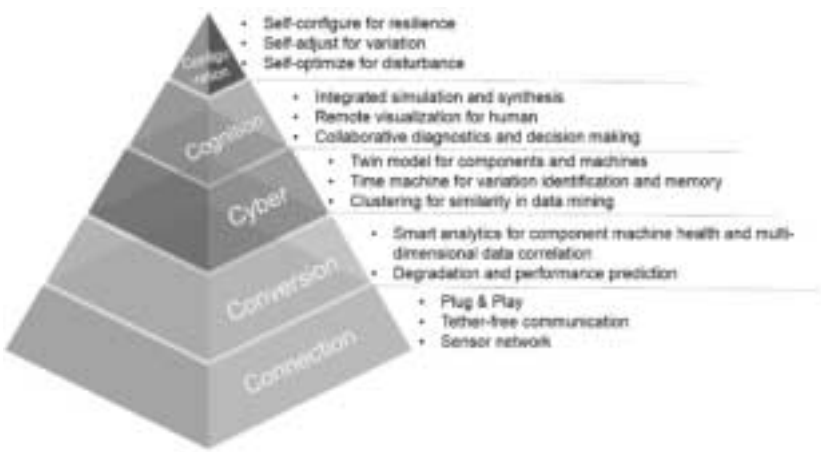


Fig. 4: Cyber Physical System Data Mining

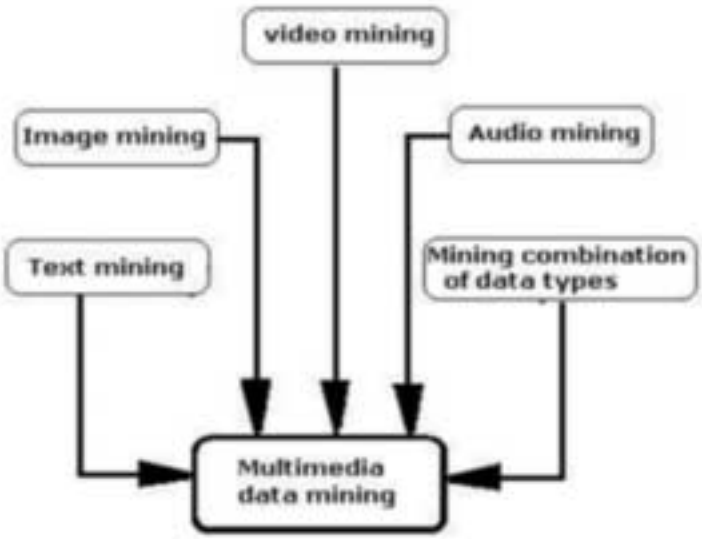


Fig. 5: Multimedia Data Mining

Multimedia Data Mining:

It is the discovery of unique fascinating patterns created from multimedia databases which stores and manages huge quantities of multimedia objects, which includes various types of data such as image, video, audio, as well as sequence data and hypertext data which contains text, markups, and linkages. It is a type of field which integrates image processing, data mining, and pattern recognition.

Text Data Mining:

Text mining is a type of field which works on information retrieval from various sources and many more fields. A small portion of information is stored in the form of various types of text such as news articles, technical papers, etc. Therefore, researches in this field of mining has always been very active. Its primitive goal is to derive high-quality information from various types of text.

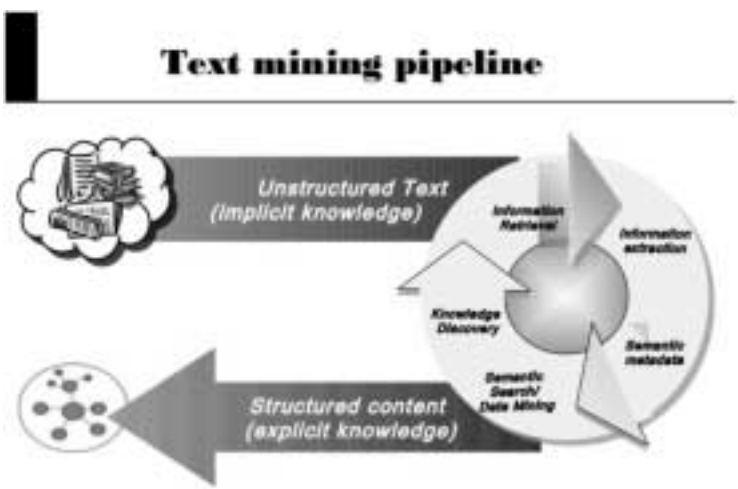


Fig. 6: Text Data Mining

Web Data Mining:

It is among one of the applications of data mining techniques that is used to discover patterns, structures, and knowledge resources from the web. According to the recent target analysis, web mining is classified into the following: web content mining, web structure mining, and web usage mining.

Web content mining analyses web content of the forms text, multimedia and structured data. This is done to understand the content of the webpages, provide scalable and informative keyword-based page indexing, matrices, and other valuable information related to web search and analysis.

Web structure mining is the process of using graph and network mining theory and methods to analyze the nodes and connection structures on the web .It extracts patterns from hyperlinks, where a hyperlink is a structural component which connects a web page to another location.

Web usage mining is the process of extracting useful information from many server logs. It finds patterns related to general or particular groups of users; understands users' search patterns, trends and associations; and predicts what users are looking for at the right time on the internet..

Stream Data Mining:

Stream data is a type of data that flow into a system in huge volume and changes very dynamically and infinitely and it contains many multi dimensional features. Stream data can't be stored in traditional data base systems. But many systems are only able to read this data only once in the sequential order. Few applications of stream data mining are: Real-time detection of anomalies in computer network traffic, text and video streams, web searches and physical cyber systems.

Statistical Data Mining:

This technique is basically derived from various computer science fields. This technique is designed for efficient handling of huge amounts of multi dimensional complex data. This technique is particularly designed for numeric data. The following mentioned are the methods used in this technique:

- Regression
- Generalized linear models
- Analysis of variance
- Mixed-effect Models
- Factor analysis
- Discriminant Analysis
- Survival Analysis
- Quality Control

Visual Data Mining:

This type of mining helps us to discover implicit and useful knowledge from huge data sets using various visualization techniques. It essentially combines the power of both the senses that is, eyes as well as brain, making it a highly attractive and efficient tool for data distribution patterns. It can be viewed as an integration of two disciplines: data visualization and data mining. The data visualization mining can be achieved in the following ways:

- Data Visualization
- Data mining result visualization
- Data mining process visualization
- Interactive visual data mining

V. Conclusion

In this paper, the concept of data mining was the hot topic of the overall discussion. Also, we learnt the concept of data stream. Later, the application of data mining in various fields of science is discussed here. In spite of the researches that have been done on data mining's application in data stream mining so far, there are still wide areas for further researches.

References

1. https://www.ibm.com/developerworks/library/ba-data-mining-techniques/index.html
2. http://www.iasri.res.in/ebook/win_school_aa/notes/spatial_data_mining.pdf
3. Data mining concepts and its techniques,Jiawei Han
4. https://en.wikipedia.org/wiki/Web_mining

National Conference on Innovative Research in Computer Science and Information Technology: "Advanced Computing Trends & Cyber Security"

The purpose of the Conference is to bring together the academicians, researchers, scientists, engineers and practitioners from various institutions, universities, colleges and industry to discuss the relevance of various futuristic technologies emerging in Networking, Big Data, Data Mining, Security and also areas related to innovations in Information Technology throughout the world.This Conference is aimed to expand its coverage in the areas specified where expert talks, young researcher's presentations will be placed in every session will be inspired and keep up your enthusiasm.

Conference Objectives

- To present the latest research and results of experimental work related to Computer Science and Information Technology
- To endow opportunities for delegates to exchange new ideas and application experiences.
- To address the real-world opportunities and challenges faced by today's IT professionals.
- To promote high level of interaction among stakeholders in theoretical, experimental and applied concepts.

Conference Themes

Topics covered by the Conference include, but are not restricted to, the following:

Advanced Computing Trends

- Computer Architecture and Engineering
- Web Security, Client and Server Security
- Applications of Cryptographic Techniques
- Self-protecting and Self-Healing System
- Artificial Intelligence and Robotics
- Soft Computing
- Software Engineering
- Human Computer Interaction
- Internet of Things (IoT)
- Social Networking
- Crowd Sourcing
- Image Processing
- Theoretical Computer Science
- Tracking and Locating Mobile Users
- Energy Saving Protocols
- Analysis of security protocols
- Communication and Networks
- Authentication and authorization of Systems
- Bots and Botnets
- Big Data Analytics

Cyber Security

- Current Trends in Ethical Hacking
- Web-Resource Protection
- Cyber Forensics and Cyber Laws
- Crypto-Systems
- Security aspects in Open Source Software
- Authentication and Authorization strategies in Networked Environment
- Management of Confidentiality and Privacy
- Securing Computing Infrastructures
- Security Breaches in Big Data Analysis
- Machine Learning applications in security
- Biometric Technologies
- Steganography
- Trojans, Vulnerability and Zombie Drones
- Emerging IT Security Models and Practices
- Intrusion detection and prevention
- Cryptography
- Diagnostics and Forensics for Security
- Information Assurance and Security
- Feature extraction and selection techniques

Institute of Innovation in Technology and Management

Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as ‘A+’ by SFRC

NAAC Accredited Institute

(ISO 9001-2008 Certified Institute)

D-27/28, Institutional Area, Janakpuri, New De1hi-110058
Tel: 011-28520890, 28520894, Te1e Fax: 011-8520239
E-mail: director@iitmjp.ac.in Website: http://www.iitmjp.ac.in