Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal

of Computer Science and Technology: C

Software & Data Engineering

Highlights

Strategy Design Pattern

Technique for Social Networking

Edgecluster based Technique

Collective Behavior Learning System

Discovering Thoughts, Inventing Future

VOLUME 14 ISSUE 6 VERSION 1.0

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Volume 14 Issue 6 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: *Open Association of Research Society*
Open Scientific Standards

Publisher's Headquarters office

Global Journals Headquarters
301st Edgewater Place Suite, 100 Edgewater Dr.-Pl,
Wakefield MASSACHUSETTS, Pin: 01880,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Integrated Editorial Board
(Computer Science, Engineering, Medical, Management, Natural Science, Social Science)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Volume

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Cost Sensitive Machine Learning Approach for Intrusion Detection

By Adamu Teshome & Dr. Vuda Sreenivasa Rao

Bahir Dar University, Ethiopia

Abstract- The problems with the current researches on intrusion detection using data mining approach are that they try to minimize the error rate (make the classification decision to minimize the probability of error) by totally ignoring the cost that could be incurred. However, for many problem domains, the requirement is not merely to predict the most probable class label, since different types of errors carry different costs. Instances of such problems include authentication, where the cost of allowing unauthorized access can be much greater than that of wrongly denying access to authorized individuals, and intrusion detection, where raising false alarms has a substantially lower cost than allowing an undetected intrusion. In such cases, it is preferable to make the classification decision that has minimum cost, rather than that with the lowest error rate.For this reason, we examine how cost-sensitive machine learning methods can be used in Intrusion Detection systems. The performance of the approach is evaluated under different experimental conditions and different models in comparison with the KDD Cup 99 winner resultsin terms of average misclassification cost, as well as detection accuracy and false positive ratesthough the winner used original KDD dataset whereas for this research NSL-KDD dataset which is new version of the original KDD cup data and it is better than the original dataset in that it has no redundant data is used.

Keywords: *intrusion detection, data mining, cost sens-itive learning.*

GJCST-C Classification : *H.1.2*

Strictly as per the compliance and regulations of:

A Cost Sensitive Machine Learning Approach for Intrusion Detection

Adamu Teshome [α] & Dr. Vuda Sreenivasa Rao [σ]

Abstract- The problems with the current researches on intrusion detection using data mining approach are that they try to minimize the error rate (make the classification decision to minimize the probability of error) by totally ignoring the cost that could be incurred. However, for many problem domains, the requirement is not merely to predict the most probable class label, since different types of errors carry different costs. Instances of such problems include authentication, where the cost of allowing unauthorized access can be much greater than that of wrongly denying access to authorized individuals, and intrusion detection, where raising false alarms has a substantially lower cost than allowing an undetected intrusion. In such cases, it is preferable to make the classification decision that has minimum cost, rather than that with the lowest error rate.For this reason, we examine how cost-sensitive machine learning methods can be used in Intrusion Detection systems. The performance of the approach is evaluated under different experimental conditions and different models in comparison with the KDD Cup 99 winner resultsin terms of average misclassification cost, as well as detection accuracy and false positive ratesthough the winner used original KDD dataset whereas for this research NSL-KDD dataset which is new version of the original KDD cup data and it is better than the original dataset in that it has no redundant data is used. For comparison of results of CS-MC4, CS-CRT and KDD winner result, it was found that CS-MC4 is superior to CS-CRT in terms of accuracy, false positives rate and average misclassification costs. CS-CRT is superior to KDD winner result in accuracy and average misclassification costs but in false positives rate KDD winner result is better than both CS-MC4 and CS-CRT classifiers.

Keywords: intrusion detection, data mining, cost sensitive learning.

I. Introduction

The field of intrusion detection has received increasing attention in recent years. One reason is the explosive growth of the Internet and the large number of networked systems that exist in all types of organizations. The increased number of networked machines has led to a rise in unauthorized activity, not only from external attackers but also from internal sources such as disgruntled employees and people abusing their privileges for personal gain [26].Since intrusions take advantage of vulnerabilities in computer systems or use socially engineered penetration techniques,intrusion detection (ID) is often used as another wall of protection. Intrusion detection includes identifying a set of malicious actions that compromise the integrity, confidentiality, and availability of information resources.

Enough data exists or could be collected to allow network administrators to detect any violations. Unfortunately, the data is so volumes, and the analysis process so time consuming that the administrators don't have the resources to go through it all and find the relevant knowledge, save for the most exceptional situations [30].Given the nature of this problem, the possible solution is data mining approach[3], [30].

Data mining approach for intrusion detection techniques generally fall into one of two categories; misuse detection and anomaly detection. In misuse detection, each instance in a data set is labeled as 'normal' or 'intrusion' and a learning algorithm is trained over the labeled data. These techniques are able to automatically retrain intrusion detection models on different input data that include new types of attacks, as long as they have been labeled appropriately. Unlike manual intrusion detection systems, models of misuse are created automatically, and can be more sophisticated and precise than manually created signatures. A key advantage of misuse detection techniques is their high degree of accuracy in detecting known attacks and their variations. Their obvious drawback is the inability to detect attacks whose instances have not yet been observed.

Anomaly detection, on the other hand, builds models of normal behavior, and automatically detects any deviation from it, flagging the latter as suspect. Anomaly detection techniques thus identify new types of intrusions as deviations from normal usage [5]. While an extremely powerful and novel tool, a potential draw-back of these techniques is the rate of false alarms. This can happen primarily because previously unseen (yet legitimate) system behaviors may also be recognized as anomalies, and hence flagged as potential intrusions. Hybrid IDS combine bothmethods and it is better one [22].

The problem with the current researches is that they try to minimize the error rate (make the classification decision to minimize the probability of error) by totally ignoring the cost that could be incurred. However, for many problem domains, the requirement is not merely to predict the most probable class label, since different types of errors carry different costs [10].

Author α: Lecturer, School of Computing and Electrical Engineering, BIT, Bahir Dar University, Ethiopia.
e-mail: adamu_teshome@yahoo.com
Author σ : Professor, School of Computing and Electrical Engineering, BIT, Bahir Dar University, Ethiopia.

For example of such problem in the case of computer network security include authentication, where the cost of allowing unauthorized access can be much greater than that of wrongly denying access to authorized individuals, and intrusion detection, where raising false alarms has a substantially lower cost than allowing an undetected intrusion. In such cases, it is preferable to make the classification decision that has minimum expected cost, rather than that with the lowest error probability [23].

Another very important case is, if class imbalanced datasets occurs but this happens in many real-world applications where the class distributions of data are highly imbalanced [23]. Again, it is assumed that the minority or rare class is the positive class, and the majority class is the negative class. Often the minority class is very small, such as 1% of the dataset. If most traditional (cost insensitive) classifiers are applied on the dataset, they will likely to predict everything as negative (the majority class) [33].

The intrusion data used for this research, KDD data set, which is publicly available and most widely used data set, has class distributions of training and test datasets with different distribution and each attack types has different costs. Statistically, the attacks of U2R and R2L are of the rarest, which makes them very hard to predict. On the other hand, they are the most dangerous types. Once an attacker gains the super user right or successfully remote login, disasters of the whole system are nothing but unavoidable [19].

Comparably, attacks of Probe are not that much dangerous. Although attacks of DOS (denial of service) are massive in the whole original dataset, they impose less danger. This is because the nature of denial of service attacks lies in that they are trying to initiate as many as possible connections to consume the network traffics and server CPU time [19].

Because of the above mentioned reasons cost sensitive learning which considers cost in decision making with acceptable accuracy is better solution for computer network intrusion detection. We used cost sensitive learning algorithms, cost sensitive classification tree CS-CRT and cost sensitive decision tree CS-MC4 algorithms. These algorithms use misclassification cost matrix to minimize the expected cost and for the detection of best prediction. Yet despite its importance, the topic is seldom addressed and researched.

II. Intrusion Detection, Cost Sensitive Machine Learning and Performance Measure

An intrusion detection system attempts to detect intrusions. In this paper, we focus on network-based systems, i.e., network intrusion detection systems (NIDS) whose primary source of data is network traffic. In contrast, there is host intrusion detection systems (HIDS) which rely on information gathered on individual hosts. Hybrid systems are both network-based and host-based [5], [22].

a) Classification of intrusion detection based on different detection method

i. *Misuse Detection*

It is also named signature-based detection, which can transform the information of attack symptom or policy disobeying into state transition-based signature or rule, and such information is stored in signature database. To judge whether or not it is attack, pre-treated case data should be first compared with the signature of signature database, and those conforming to attack signature data can be judged as attack [22]. Its advantage is high detection rate and low false alarm rate for known attacks; however, its detection capacity is low for unknown detection methods, and attack database should be renewed on a regular basis.

ii. *Anomaly Detection*

It may establish a profiles for normal behavior of users, which comes from statistics data of users in the former period; when detection is performed, the profiles is compared with actual users' data, if the offset is below threshold value, user's behavior can be considered normal, and it has no intention of attacks; if the offset is above threshold value, user's behavior can be considered abnormal [22]. Anomaly detection is based on an assumption that intruder's behavior is different from normal users' behavior. Detection rate of the method is high, and it is more likely to detect unknown attacks, but misjudgment (false positive) rate is also high.

iii. *Hybrid*

It is also possible to include both normal and attack data in the model. Such systems are referred to be hybrid detectors.

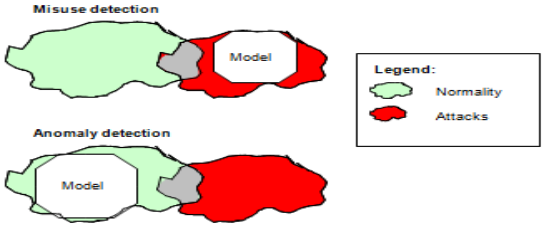


Figure 1 : The relation between misuse detection and anomaly detection.

b) Architecture

In general, a NIDS consists of different components.

1. *Collector:* Provides an interface for accessing data that is used by the detection process. For a NIDS, the primary kind of data collector is a network tap. A tap provides access to all raw network packets which cross a particular position of a network.
2. *Detector:* Conducts the actual detection process. The detector is the "brain" of the NIDS. It accesses data provided by collector and storage (see below), and it decides what should trigger an alert.
3. *User Interface:* Reports results to the user, and enables the user to control the NIDS.
4. *Storage:* Stores persistent data required by the detector or the user interface. Such data is either derived by the detector itself or provided externally.
5. *Responder:* Reacts to detected intrusions in order to prevent future damage. Active responses may include dropping the connectivity to the potential attacker or even counter-attacks. A response may be triggered automatically or manually via the user interface.

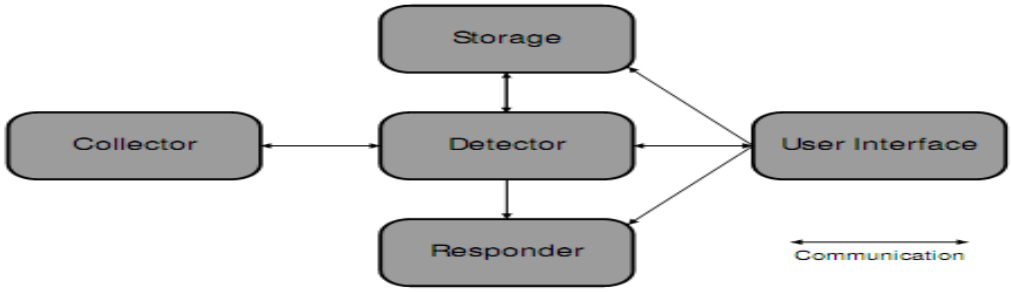


Figure 2 : NIDS architecture

III. Cost Sensitive Machine Learning

Costs are central to statistical decision theory but cost sensitive learning received only modest attention before [19]. Cost sensitive learning is a type of learning in data mining that takes the misclassification costs (and possibly other types of cost) into consideration. The goal of this type of learning is to minimize the total cost. The key difference between cost sensitive learning and cost insensitive learning is that cost sensitive learning treats the different misclassifications differently. Cost insensitive learning does not take the misclassification costs into consideration. The goal of this type of learning is to practice a high accuracy of classifying examples into a set of known classes.

Cost sensitive learning is an extensively used practice in data mining, which assigns different levels of misclassification penalty to each class. Cost sensitive technique has been incorporated into classification algorithms by taking into account the cost information and trying to optimize the overall cost during the learning process.

a) Cost Models

The cost model of IDS devises the total expected cost of the IDS. The cost model depends on the detection performance of the IDS. Misclassification costs false positive (FP, the cost of misclassifying a negative instance into positive) and false negative (FN, the cost of misclassifying a positive instance into negative), and the cost of correct classification is zero. They simply assign FP as the weight to each negative instance, and assign FN as the weight to each positive instance. That is, the weight ratio of a positive instance to a negative instance is proportional to FN/FP [9], [19].

The cost of misclassifying an instance of class i as class j is C(i, j) and is assumed to be equal to 1 unless specified otherwise(i, i) = 0 for all i . Any classification tree can have a total cost computed for its terminal node assignments by summing costs over all misclassifications. The issue in cost sensitive learning is to induce a tree that takes the costs into account during its growing and pruning phases [19]. These misclassification cost values can be given by domain experts, or learned via other approaches. In cost sensitive learning, it is usually assume that such a cost matrix is given and known. For multiple classes, the cost matrix can be easily extended by adding more rows and more columns.

Table 1: An example of cost matrix for binary classification.

	Actual negative	Actual positive
Predict negative	C(0,0), or TN	C(0,1), or FN
Predict positive	C(1,0), or FP	C(1,1), or TP

When optimizing sampling levels to improve overall cost, a logical evaluation criterion is the cost itself. Cost is calculated as follows by taking the assumption C (+|+) = C (-|-) =0.Cost = FN_{rate}x C (-|+) + FP_{rate}x C (+|-). Therefore, once classification occurs in the cross-validation phase, the wrapper or filter calculates the validation cost and uses this information to select optimal sampling levels. This approach is

dependent on a priori knowledge of the cost relationship between classes [10]. Cost sensitive learning algorithm which directly introduce and make use of misclassification costs into the learning algorithms are ICET, EG2, CS-ID3, CS-MC4 and CS-CRT. The cost sensitive learning algorithms that are used for this research are cost sensitive learning algorithms i.e. cost sensitive decision tree CS-MC4 which is the cost sensitive version of C4.5and CS-CRT the cost sensitive version of C-RT classification tree that use misclassification cost matrix to minimize the expected cost and for the detection of best prediction.

b) Performance Measures

In general, intrusion detection systems require high detection rate, low false alarm rate and lower average misclassification cost. The followings are commonly used performance measures

i. *Error Rate*

The error rate, which is only an estimate of the true error rate and is expressed to be a good estimate, if the number of test data is large and representative of the population and is defined as [68]:

$$\text{Error Rate} = \frac{TotalTestsamples - TotalCorrectlyclassifiedsamples}{TotalTestsamples} \times 100\%$$

ii. *Accuracy*

Overall Classification accuracy (OCA) is the most essential measure of the performance of a classifier. It determines the proportion of correctly classified examples in relation to the total number of examples of the test set i.e. the ratio of true positives and true negatives to the total number of examples.

From the confusion matrix, we can say that [9] accuracy is the percentage of correctly classified instances over the total number of instances in total data, namely the situation TP and TN, thus accuracy can be defined as follows:

$$\text{Accuracy} = \frac{Total\,number of correctly classified samples}{Total\,number of test samples} \times 100\%$$

Or

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

iii. *Detection Rate*

Detection rate refers to the proportion of attack detected among all attack data, namely, the situation of TP, thus detection rate is defined as follows [20]:

$$\text{DetectionRate} = \frac{no.of samples classified correctly}{no.of samples used for testing} \times 100\%$$

Or

$$\text{Detection Rate} = \frac{TP}{TP + FN} \times 100\%$$

iv. *False Positive Rate*

Another name is False Alarm Rate (FAR) measures the number of misclassified positive instances in relative to the total number of misclassified instances.

Can be expressed also the proportion that normal data is falsely detected as attack behavior, namely, the situation of FP, thus false alarm rate is defined as follows [20]:

$$\text{False Positive Rate} = \frac{total\,no.of misclassified samples}{total\,no.of test samples} \times 100\%$$

Or

$$\text{False alarm rate} = \frac{FP}{FP + TN} \times 100\%$$

Recall and Precision are two widely used metrics employed in applications where successful detection of one of the classes is considered more significant than detection of the other classes [].

v. *Average Misclassification Cost (AMC)*

$$AMC = \frac{1}{N} \sum_{i=1}^{5} \sum_{j=1}^{5} CM(I, J) * C(I, J)$$

Where CM corresponds to confusion matrix, C corresponds to the cost matrix, and N represents the number of patterns tested [10].

IV. Data Collection and Preparation

NSL-KDD data set which is the new version of KDD Cup 99 dataset and the only publicly available [21] and the widely used data set for intrusion detection have been used for the experimental purpose. The process of data cleaning and preparation is highly dependent on the specific machine learning algorithm and software chosen and algorithms used for the machine learning task. The researcher attempted to prepare the data according to the requirements of the Tanagra which is powerful, user friendly and freely available for noncommercial purpose machine learning software [11] and according to the requirements of CS-MC4 and CS-CRT algorithms by consulting different literatures.

V. Training and Testing Models

The Intrusion detection models were developed using cost sensitive classification tree CS-CRT and cost sensitive decision tree CS-MC4 algorithms on full training NSL-KDD dataset using a powerful machine learning and data mining Tanagra tool and then Full testing dataset of NSL-KDD passed through the developed models to detect the intrusions and find the detection error rates, precision, false positives rate, average misclassification cost and accuracy of the detection models but for comparison of models we used mainly average misclassification cost, false positives rate and accuracy of detection.

VI. Experimentation

We used data mining software tool known as Tanagra version 1.4.34 available freely at [11]. The software is a GUI based software and easy to use. Tanagra is capable of classifying large volumes of data within a second depending on the speed and specification of computer processor. All experiments were performed using an Intel Core 2 Duo Processor 2.16 GHz processor with 4 GB of RAM and implemented on a Vista Windows operating system.

a) NSL-KDD Intrusion Detection Dataset:

We have used NSL-KDD intrusion dataset which is available at [21] and it is a dataset suggested to solve some of the inherent problems of the original KDD Cup 99 dataset [13,20]. The NSL-KDD dataset has the following advantages over the original KDD Cup 99 dataset.

1. It does not include redundant records in the training set, so the classifiers will not be biased towards more frequent records.
2. There are no duplicate records in the proposed test sets, therefore the performance of the learners are not biased by the methods which have better detection rates on the frequent records.
3. The number of selected records from each difficulty level group is inversely proportional to the percentage of the records in the original KDD dataset. As a result the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.
4. The number of records in the training and testing sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion.
5. Statistical observations, one of the most important deficiencies in the KDD dataset is the huge number of redundant records, which causes the learning algorithms to be biased towards the frequent records and thus prevent them from learning unfrequent records which are usually more harmful to networks such as U2R and R2L attacks.

Table2 and Table3 shows the statistics of the redundant records in the KDD Cup 99 training and testing datasets.

Table 2 : Statistics of redundant records in the KDD training dataset.

Original	Records	Distinct Records	Reduction Rate
Attacks	3,925,650	262,178	93.32%
Normal	972,781	812,814	16.44%
Total	4,898,431	1,074,992	78.05%

Table 3 : Statistics of redundant records in the KDD testing dataset

Original	Records	Distinct Records	Reduction Rate
Attacks	250,436	29,378	88.26%
Normal	60,591	47,911	20.92%
Total	311,027	77,289	75.15%

a) Evaluation Metrics

We have used the cost matrix defined for the KDD Cup 99 Dataset [9] which is shown in Table 4.

Category	Normal	Probe	DOS	U2R	R2L
Normal	0	1	2	2	2
Probe	1	0	2	2	2
DOS	2	1	0	2	2
U2R	3	2	2	0	2
R2L	4	2	2	2	0

VII. Experimentation and Result

The experimentations have two major parts; the first one is experimentation on CS-MC4 and CS-CRT using all the 41 attributes and the second is experimentation on CS-MC4 and CS-CRT using information gain (IG) feature (attribute) selection method. Comparative discussions of each approaches used with the KDD 99 winner results are given.

Table 5 summarizes comparison results of detection accuracy (which refers to the proportion that a type of data is correctly classified) of normal and 4 different attacks (i.e. Probe, Dos, U2R, R2L) based on CS-MC4 and CS-CRT classifiers using 24, 30, 37, and all the attributes in comparison with KDD winner results.

Table 5 : Performance comparison of testing result for five-class classification.

	Normal		DOS		R2L		Probe		U2R	
	Cs-mc4	Cs-crt	Cs-mc4	Cs-crt	Cs-mc4	Cs-crt	Cs-mc4	Cs-crt	Cs-mc4	Cs-crt
All attributes	98.71	92.91	99.85	97.44	91.69	61.46	98.80	99.66	72.50	17.50
24 attributes	98.65	93.56	99.81	98.01	89.53	45.68	98.85	97.94	52.50	10.00
30 attributes	98.61	98.25	99.81	99.64	91.20	88.87	98.85	99.04	77.50	20.00
37 attributes	98.70	92.83	99.81	97.44	91.69	61.46	98.80	99.66	65.00	17.50
KDD winner	99.5		97.1		8.4		83.3		13.2	

It is evident from this Table that:

1. *For Normal:* CS-MC4 classifiers outperform their CS-CRT counterparts in all cases but when only 30 attributes are used, the difference between the accuracy of CS-MC4 and CS-CRT became minimal i.e. 0.36%. For this type of class, in all the cases the accuracy of the KDD winner result is better than both CS-MC4 and CS-CRT classifiers.
2. *For Probe attack:* the accuracy of CS-MC4 is better than that of CS-CRT when only 24 attributes are used, but in other circumstances CS-CRT outperforms CS-MC4. For this type of attack, in all the cases the accuracy of both CS-MC4 and CS-CRT is better than the KDD winner result.
3. *For Dos attack:* CS-MC4 classifiers outperform their CS-CRT counterparts in all cases but when only 30 attributes are used, the difference between the accuracy of CS-MC4 and CS-CRT became minimal, i.e. 0.17%. For this type of attack, in all the cases the accuracy of both CS-MC4 and CS-CRT is better than the KDD winner result.
4. *For U2R attack:* the accuracy of CS-MC4 classifiers is better than their CS-CRT counterparts in all cases. For this type of attack, in all the cases the accuracy of both CS-MC4 and CS-CRT is better than the KDD winner result except when only 24 attributes are used for CS-CRT classifier.
5. *For R2L attack:* accuracy of CS-MC4 classifiers is better than their CS-CRT counterparts in all cases. For this type of attack, in all the cases the accuracy of both CS-MC4 and CS-CRT are by far better than the KDD Cup 99 winner result.

In general, it is evident from the above table that in terms of accuracy, CS-MC4 outperformed the CS-CRT. As it can be seen from the result, the feature selection method used (i.e. information gain value) did not increase the accuracy of CS-MC4 classifier. Even though CS-MC4 classifier achieved the same result using all the attributes (41 attributes) and 37 attributes (at information gain value greater than 0), the reduction of the attributes from 41 to 37 could increase the speed of the classifier and reduce the space required. CS-CRT achieved a better result when only 30 attributes (information gain value greater than and equal to 0.011) are used relative to the other cases.

So, feature selection using information again value achieved better result for CS-CRT classifier from 41 attributes to 30 attributes but for CS-MC4 classifier from 41 attributes to 37 attributes. Results which are achieved in this work are compared with the results obtained by KDD winner results as shown in table 4. As it can be seen, the accuracy of KDD Winner is only better in normal, but it is far worse than CS-MC4 and CS-CRT in U2R and R2L attacks; this might be because of the data used for this research is NSL-KDD intrusion dataset (which is new version of the original KDD cup data and it is better than the original dataset in that it has no redundant data) but for the KDD winner the original KDD cup dataset is used.

Table 6 summarizes the efficiency of information gain value for feature selection and the performance comparison of CS-MC4, CS-CRT classifier and KDD winner result based on overall accuracy, false positives rate and average misclassification cost using 24, 30, 37 and all the attributes for 5 attack classes on NSL-KDD dataset.

Table 6 : Performance comparison of different methods.

	Overall accuracy (%)		False positives rate (%)		Average Misclassification Cost	
	CS-MC4	CS-CRT	CS-MC4	CS-CRT	CS-MC4	CS-CRT
All attributes	98.9	94.2	1.3	7.1	0.0199	0.0895
24 attributes	98.8	94.1	1.4	6.4	0.0232	0.0982
30 attributes	98.8	98.4	1.4	1.7	0.0201	0.0335
37 attributes	98.9	94.1	1.3	7.2	0.0199	0.0887
KDD winner	92.7		0.55		0.2331	

It is evident from above table that:

1. Overall accuracy of CS-MC4 classifiers is better than their CS-CRT counterparts in all cases. And in all the cases, the overall accuracy of both CS-MC4 and CS-CRT is better than the KDD winner result.
2. False positives rate of CS-MC4 classifiers are better than their CS-CRT counterparts in all cases. And in all the cases, the false positive rate of the KDD winner result is better than both CS-MC4 and CS-CRT classifiers.
3. Average misclassification cost of CS-MC4 classifiers is better than their CS-CRT counterparts in all cases. And in all the cases, the Average misclassification cost of both CS-MC4 and CS-CRT is better than the KDD winner result.
4. Attribute reduction using information again value achieved better result for CS-CRT classifier from 41 attributes to 30 attributes, from 94.2% to 98.4% accuracy, from 7.1% to 1.7% false positive rate, and average misclassification cost from 0.0895 to 0.0335; but for CS-MC4 classifier it only reduce the attribute from 41 attributes to 37 attributes.

VIII. Conclusions

This paper focuses on using cost sensitive learning techniques to existing data mining algorithms to deal with cost of different types of attacks in intrusion detection while at the same time reducing the false positives rate. The cost issue is widely ignored in the intrusion detection research. As a result, most of the research projects tried to minimize the false positives rate which may not reflect real-world scenario of dealing with ever increasing different types of attacks with different costs; and an important consideration is the fact that raising false alarms carries a significantly lower cost than not detecting attacks.

For comparison results of CS-MC4, CS-CRT and KDD 99 winner result, it was found that CS-MC4 is superior to CS-CRT in terms of accuracy, false positives rate and average misclassification costs. CS-CRT is superior to KDD winner result in accuracy and average misclassification costs but in false positives rate KDD winner result is better than both CS-MC4 and CS-CRT classifiers.

This paper proposed learning approach for network intrusion detection that performs feature selection method using information gain by selecting important subset of attributes. The performance of the proposed approach on the NSL-KDD intrusion detection dataset achieved a better accuracy from 94.2% to 98.4% for the CS-CRT classifier. It also reduced the false positives for the CS-CRT algorithm from 7.1% to 1.7%, and the average misclassification costs from 0.0895 to 0.0335; but for the CS-CM4 algorithm, it only reduced the attribute from 41 to 37 attributes.Even though feature selection method could not increase accuracy or reduce false positive rate and average misclassification cost for CS-MC4, it could increase the speed of the classifier and reduce the computation space required. The experimental results have manifested that feature (attribute) selection improves the performance of IDS.

References Références Referencias

1. Abraham, R. Jain and J. Thomas, D-SCIDS: Distributed soft computing intrusion detection system, Journal of Network and Computer Applications, vol. 30, p p. 81-98, 2007.
2. Ajith, G. Crina, C. Yuehui. Cyber Security and the Evolution of Intrusion Detection Systems, Information Management and Computer Security 9(4): 175-182, 2001.
3. AlirezaOsareh and BitaShadgar (2008). Intrusion Detection in Computer Networks based on Machine Learning Algorithms. IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.11
4. Chai, X., Deng, L., Yang, Q., and Ling,C.X. , Test-Cost Sensitive Naïve Bayesian Classification, In Proceedings of the Fourth IEEE International Conference on Data Mining. Brighton, UK: IEEE Computer Society Press, 2004.
5. D.E. Denning, An Intrusion Detection Model, IEEE Trans-actions on Software Engineering, SE-13:222-232, 1987.
6. Domingos, P, MetaCost: A general method for making classifiers cost-sensitive, In Proceedings of the Fifth International Conference on Knowledge

Discovery and Data Mining, 155-164, ACM Press, 1999.
7. Drummond, C., and Holte, R., Exploiting the cost (in) sensitivity of decision tree splitting the criteria. In Proceedings of the 17 International Conference on Machine Learning, 239- 246, 2000.
8. E. Bloedorn, Data Mining for Network Intrusion Detection: How to Get Started, MITRE Technical Report, August 2001.
9. ElkanC Results of the KDD'99 Classifier Learning Contest, 1999. Available from <http:// www- cse. ucsd.edu/users/elkan/clresults.html>
10. Elkan, C. The Foundations of Cost-Sensitive Learning, In Proceedings of the Seventeenth International Joint Conference of Artificial Intelligence, 973-978. Seattle, Washington, 2001.
11. http://eric.univlyon2.fr/~ricco/tnagra.html.
12. [12]J. Luo, Integrating Fuzzy Logic With Data Mining Methods for Intrusion Detection, Master's thesis, Department of Computer Science, Missis - sippi State University, 1999.
13. J. McHugh, Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory, ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, 2000.
14. J. P. Anderson, Computer Security Threat Monitoring and Surveillance, Technicalreport, Jam - esP Anderson Co., Fort Washington, Pennsylvania, April 1980.
15. Joseph, S and Rod, A, Intrusion detection: methods and systems. Information Management and Computer Security 11(5):222-229, 2003.
16. KDD CUP 1999 Data. The UCI KDD Archive Information and Computer Science, University of California,Irvine.http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
17. L. Ertoz, E. Eilerson and A. Lazareviv, The MINDS – Minnesota intrusion detection system, next generation data mining, MIT Press, 2004.
18. L. T. Heberlein, K. N. Levitt, and B. Mukherjee. A Method To Detect Intrusive Activity in a Networked Environment, In Proceedings of the 14th National Computer Security Conference, pages 362{371, october 1991.
19. Lee W, Stolfo SJ. A Framework for constructing features and models for intrusion detection systems, ACM TransInfSystSecurity, 2000.
20. M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, A Detailed Analysis of the KDD CUP 99 Data Set, IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.
21. Nsl-kdd data set for network-based intrusion detection Systems. Available on: http://nsl.cs.unb.ca/ NSL- KDD/, March, 2010.
22. Paul Dokas, LeventErtoz, Vipin Kumar, AleksandarLazarevic, Jaideep Srivastava, Pang-Ning Tan, Data Mining for Network Intrusion Detection ,University of Minnesota, 2001.
23. Provost, F., Machine learning from imbalanced data sets 101, In Proceedings of the AAAI'2000 Workshop on Imbalanced Data, 2000.
24. R. Heady, G. Luger, A. Maccabe, and M. Servilla. The Architecture of a Network Level Intrusion Detection System, Technical report, Department of Computer Science, University of New Mexico, August 1990.
25. R. Heady, G. Luger, A. Maccabe, and M. Servilla. The Architecture of a Network Level Intrusion Detection System, Technical report, Department of Computer Science, University of New Mexico, August 1990.
26. S. Wafa, Al-Sharafat , ReyadhSh.Naoum, Adaptive Framework for Network Intrusion Detection by Using Genetic-Based Machine Learning Algorithm, IJCS-NS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.
27. S.Selvakani, R.S. Rajesh, Genetic Algorithm for framing rules for intrusion Detection, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.11, November 2007.
28. Sheng, V.S. and Ling, C.X. , Thresholding for Making Classifiers Cost-sensitive, In Proceedings of the 21National Conference on Artificial Intelligence, 476-481, Boston, Massachusetts, 2006.
29. SimsonGarfinkel and Gene Spafford. Practical Security, O'Reilly and Associates, Sebastopol, California, 1991.
30. Sterrybrugger, Data Mining Methods for Network Intrusion Detection, University of California, June 2004.
31. Turney, P.D. Types of cost in inductive concept learning, In Proceedings of the Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning, Stanford University, California, 2000.
32. Turney, P.D., Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm, Journal of Artificial Intelligence Research 2:369- 409, 1995.
33. Zadrozny, B. and Elkan, C., Learning and Making Decisions When Costs and Probabilities are Both Unknown, In Proceedings of the Seventh Intern - ational Conference on Knowledge Discovery and Data Mining, 204-213, 2001.
34. Zadrozny, B., Langford, J., and Abe, N., Cost-sensitive learning by Cost-Proportionate instance Weighting, In Proceedings of the 3th International Conference on Data Mining, 2003.

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Recognizing Human Affection: Smartphone Perspective

By Mefta Sadat, Rasam Bin Hossain & Hasan Mahmud

Islamic University of Technology, Bangladesh

Abstract- Touch-screen Smartphone has become an obligatory segment in the lives of billions of people around the world. Understanding the human affection or emotional state of the user enables efficient human computer interaction. Smartphone is one of the most frequently used electronic devices and the number of applications developed for it is increasing day by day. Emotion recognition of the user will lead to the development of emotion aware applications. Service recommendations and intelligent user interfaces in Smartphone will be other encouraging scopes for the mobile application developers. In this paper we discuss about state-of-the-art technologies to detect human emotional states. We proposed a methodology by which three different emotional states (positive, neutral, negative) of the user can be identified using Smartphone's built-in sensors like the gyroscope, accelerometer and also additional sensors such as pressure sensor. We tried to analyse infraction log of Smartphone users, approximated different sensor values to recognize human emotions. Since the pressure values found on the existing phones are not completely accurate, we introduced the use of Force Sensitive Resistor (FSR) sensor to get more accurate pressure values.

Keywords : *emotion, affective-computing, touchscreen, pressure, smartphone, sensor.*

GJCST-C Classification : *I.5*

Strictly as per the compliance and regulations of:

Recognizing Human Affection: Smartphone Perspective

Mefta Sadat [α], Rasam Bin Hossain [σ] & Hasan Mahmud [ρ]

Abstract- Touch-screen Smartphone has become an obligatory segment in the lives of billions of people around the world. Understanding the human affection or emotional state of the user enables efficient human computer interaction. Smartphone is one of the most frequently used electronic devices and the number of applications developed for it is increasing day by day. Emotion recognition of the user will lead to the development of emotion aware applications. Service recommendations and intelligent user interfaces in Smartphone will be other encouraging scopes for the mobile application developers. In this paper we discuss about state-of-the-art technologies to detect human emotional states. We proposed a methodology by which three different emotional states (positive, neutral, negative) of the user can be identified using Smartphone's built-in sensors like the gyroscope, accelerometer and also additional sensors such as pressure sensor. We tried to analyse infraction log of Smartphone users, approximated different sensor values to recognize human emotions. Since the pressure values found on the existing phones are not completely accurate, we introduced the use of Force Sensitive Resistor (FSR) sensor to get more accurate pressure values.

Keywords: emotion, affective-computing, touchscreen, pressure, smartphone, sensor.

I. Introduction

The recognition of human affection is considered as a significant feature in many intelligent systems. In certain situations, emotion aware intelligent systems perform with greater credibility than normal systems. For this reason, the development and implementation of computer-based emotion aware systems have become an interesting research area in affective computing researches. Affective computing is the part of Human Computer Interaction (HCI) that deals with the study, recognition, interpretation & simulation of human affects.

We have seen the use of human affection recognition in desktop application as well as in web applications. Sometimes it is found as service recommendation utility in different applications. In contrast to this, it has greater promise in mobile scenario. The rapid advancement of modern science has taken us to the peak of mobility bringing powerful processor enabled mobile devices in our hand. People are using handheld devices like Smartphone and tablet computer more. Since mobile phones have become an essential part of our day to day life they represent an ideal computing platform to monitor the behavior, movement, moods of the user himself. This processor embedded powerful Smartphone will become even smarter if it can detect user emotion and act accordingly. It is an obvious matter that the emotion of the users' mental state significantly affects their social lives and interactions. As Smartphone are used more frequently than any other devices that they carry, applications that are being developed focusing the emotion will be more effective as they will not only focus on the usability or user experience but also on the interaction methodologies.

Current Smartphone usually have a virtual keyboard rather than physical keypad on its body. That means users have to directly touch the screen in order to interact with the device. One of the most significant factors affecting the usage of Smartphone is the pressure given on the screen with the fingers while typing a message, playing games, browsing, switching applications and many more. Every time a user is tapping, swiping or pinching on the screen a small number of pressure is applied. We believe that due to change of the mental state, the amount of pressure is given on the screen also changes. The variations are too little, so we need to normalize the values and create some approximated result so that we can use the resulting output to draw conclusions about the user's emotional state.

As the user's mental state changes with the context of his environment and surroundings, these may lead to a situational impairment [14]. The present phone operating systems are not that much aware of the user's emotion. We have picked Smartphone rather than desktop computers or other sensory devices because it is everyday companion. Researchers have found different touch patterns for typing in on-screen keyboards [15]. The amount of pressure applied on the screen varies with the user's emotion and substantially it increases when he is in excited or angry mood. The normal state or neutral state usually exhibits steady variations. In our system we will try to combine both of the features(pressure and the signal coming from the gyroscope, accelerometer), normalize them and make a specific relation between them so that in future we can

Author α σ : Department of Computer Science and Engineering Islamic University of Technology, Gazipur, Bangladesh.
e-mails: meftasadat@gmail.com, rasam_hossain@yahoo.com
Author ρ: Assistant ProfessorDepartment of Computer Science and Engineering Islamic University of Technology, Gazipur, Bangladesh.
e-mail: hasan.iut@gmail.com

detect the user's emotion based on these results. By examining the gyroscope signal we will be able to detect whether the phone is in a stationary state or in a mobile state. The accelerometer sensor is mostly responsible to detect the state of the user [9] that means whether the user is standing, walking, jogging, running, ascending/descending stairs or doing any kind of extra physical works.

Here we want to draw some conclusions about what we are going to do in our system: (1) we will create a system that will combine the signals coming from the built in sensors that are already available with the phone; the amount of pressure applied to touch screen using external sensor and the signals coming from the gyroscope and accelerometer. (2) The system will be less battery consuming as there will be an effective algorithm to pause or start the system based on the states of the Smartphone whether it is being used or not (3) the resulting outputs will be fetched in to a table to show the user's emotional state.

The working procedure has been divided into several sections to describe our research work precisely. In the next sections we are going to discuss about the theoretical definitions and other related terms. A brief description of our background study will also be added here and related works will be stated clearly in following sections. After the state of the art we will define the problem description briefly so that it can be easily understandable. In the last section we will state our future works.

II. Background Study

"Emotion" is an abstract term. It is an idea or expression that can define the current state or condition of human mind or behavior. Emotion can be characterized by psychological & physiological expressions, biological reactions, and mental states.

Emotion [10] is one of the fundamental components or characteristics of being human. Emotion can be derived from human experiences and expressions to represent different emotional states such as joy, hate, anger, pride, excitement, sadness and so on. The current research in psychology and technology suggest a completely different view of the various aspects of the relationship among human, computer and emotion. Emotion is no longer seen as limited to casual ebullition of rage and anger when a computer program crashes for inexplicable reasons, excitement when a video game character overcomes an obstacle or frustration when an incomprehensible error message appears on the screen. It is now understood that a wide range of emotions play a vital role in every computer related, goal oriented activity, from using drawing application and editing photos in a photo manipulation app, to browsing web pages and sending a message, to making an online purchase to playing games. The way the user carries out a task or uses an application is highly influenced by his emotional states.

Here, another important aspect is human affection which is the main focus of Affective Computing [10]. Affective Computing is one of major research areas which relates to and arises from or it adversely effects or influences human emotion or other active phenomena.
Recognition of emotion refers to the identification of emotional states. It can be done in various contexts. Facial expressions, gesture, posture, speech analysis, keystroke and pressure all of these can be used for emotion recognition.

There are several emotion recognition techniques. First, there can be emotion recognition from facial expression. It is a vast area of psychological studies for last several decades. It requires neuroscience, digital image processing, pattern recognition and a lot analysis and processing of collected data to generate probable emotion. There are also various differences in the facial expression patterns in different age groups of humans (such as children, adult, senior etc). It is also difficult to implement in smaller hardware or devices. Hence the necessity to detect emotion in some other convenient ways arises. [11]

Second, there is Vocal emotion recognition. The vocal aspect of a message or conversation carries significant variations or emotional information. If we do not consider how a sentence was verbally spoken, the meaning may change. From the input audio signals pitch, intensity, and pitch contours were estimated as acoustic features and then classified using some predefined roles and mapped to suitable emotional state. This model has some inaccuracies in case of neutral voice.

Third, another new method is Multimodal Emotion Recognition. Voice, gesture, and force-feedback etc can also be used for detecting emotion instead of using only keyboard and mouse. Sometime people say specific sentence with particular facial expression to indicate an emotional state. The multimodal approach is much more dynamic but it has some constraints and it's an ongoing research. It has to combine multiple types of inputs and finally fuse with probabilistic models. [12]

Forth, another approach is emotion recognition using pressure sensing keyboards and keystroke dynamics. [13]

The context of our work is touch screen Smartphone. People use these devices as their daily commodities. Emotion frequently changes while using such a device depending on the scenario. The input for these devices is touch. The pressure applied on the screen by the user as touch stimuli vary according to the mood or mental state of the user. So, the idea is to find out the pressure data and after further processing, it can be useful information for emotion recognition.

There are several types of built in sensors in today's smart phones. Some of them can be used to capture the pressure raw data. In most of the Smartphone, these sensors exist—accelerometer, gyroscope, ambient light sensor, proximity sensor, GPS etc. The sensors are built into handsets using Micro-electromechanical systems (MEMS).

Accelerometer: This sensor allows the device to detect the orientation of the device and adapts the content to suit the new orientation. It also measures the acceleration of the device relative to freefall. It actually controls the switching between portrait and landscape modes.

Mobile phone accelerometers are a form of three-axis micro electromechanical-system (or MEMS) and essentially consist of a series of three sensors which are fixed at right angles to each other. Each of these sensors is capable of monitoring the smallest of changes in force and pressure: as applied by gravity or movement and creating a corresponding electrical impulse. Development of this technology has allowed for the relatively cheap manufacture of increasingly accurate and sensitive accelerometers which can be easily integrated into all manner of applications.

Gyroscope: It is a device for measuring or maintaining orientation, based on the principles of angular momentum. This sensor is built into modern Smartphone mainly for navigation systems and gesture recognition systems and also for finding the position and orientation of the device.

Gyroscope is capable of measuring angular rates around one or more axes, these gyroscopes represent a fitting complement to MEMS accelerometers. Thanks to the combination of accelerometers and gyroscopes it is possible to track and to capture complete movements in a three-dimensional space. According to Wikipedia, this nifty device is able to detect angular movements such as the rotation around the X-axis, rotation around the Y-axis and the rotation around the Z-axis or also known as roll, yard and pitch in layman's terms. On the other hand, an accelerometer is only able to detect three linear axes of vectors, including left-right (X-axis), top-bottom (Y-axis) and up-down (Z-axis). Unlike a gyroscope, it measures translation of direction and cannot detect if a device made a full spin or is experiencing inertial change. Mechanical and Electronics Engineering today have managed to transform the mechanical gyroscope into a Micro electromechanical system (MEMS), also known as vibrating structure gyroscope. So, instead of having a spinning wheel inside the microchip, a vibrating mass is placed in the centre of the chip. The mass will be vibrated whenever an electrical signal goes through it. Moving the phone will cause the changes of electrical signals that are picked by the sensors. The sensors will send instructions to be interpreted by the software to provide the necessary feedback to the user.

Proximity Sensor: It detects how close the screen of the phone is to our body. This allows the phone to sense when we have brought the phone close to ear. It is used to avoid unwanted input.

When combining both the accelerometer and the gyroscope, it will be a total of 6-axis motion sensing that is able to have precise motion detection by simply moving the phone naturally.

Ambient light sensor: It is mainly used to adjust the screen brightness automatically and for power saving.
Magnetometer: This sensor measures the strength of earth's magnetic field. Result is expressed in Tesla. Smartphone provides raw magnetometer data and compass bearing. It is used as compass to rotate maps, graphics, and orientation and also as recorder to detect magnets, force fields etc.

These are the main sensors typically found in modern Smart phones. As we can see there is no dedicated pressure sensor built in to normal Smartphone. Thus the necessity to get pressure data from elsewhere such as touch screens of the device is felt.

There are many types of touch screens in Smartphone. But the most commonly used two are the Resistive and the capacitive touch screen. Resistive touch screen is used lower end Smartphone and right now almost outdated. Resistive touch screens contain two layer of conductive material with a very small gap between them which acts as a resistance. When the resistive touch screen is touched with finger (or stylus) the two layers meet at the point of touch and create a circuit. This information is recognized by the mobile's processor. This kind of screens can be operated with a finger, a fingernail, a stylus or any other object.

Capacitive touch screen technology consists of a layer of glass coated with a transparent conductor (like indium tin oxide). When a capacitive touch screen is touched by human body (finger), an interruption is created in the screens electrostatic field (which is measurable as a change in capacitance) which is detected by phone's processor or chip and which in turn instructs phone's operating system to trigger an event or action accordingly.

Haptic/Tactile touch screen is another technology used by Blackberry and Nokia targeted towards enterprise market. This technology provides a tactile feedback on a touch action on the screen thus providing an immediate and unmistakable confirmation to the user. Hepatic technology has been found to significantly improve user performance, accuracy and satisfaction while typing on a touch screen.

Retina Display is another technology mainly used by apple for their IPS LCD (with backlit LED) in iPhone4. They call it the Retina display because its

pixels cannot be individually identified by the human eye, thus making the display super sharp and brilliant.

The enhancement of the capacitive touch screens is AMOLED and Super AMOLED. AMOLED means Active-Matrix Organic Light-Emitting Diode. AMOLED displays are a type of OLED displays for mobiles and are rapidly gaining popularity in top end Smartphone segment. Super AMOLED displays are an even advanced version of AMOLED displays developed by Samsung. Super AMOLED display is built with touch sensors on the display itself, as opposed to creating a separate touch sensitive layer (as in capacitive touch screen).

III. Related Works

There are only few research works that have been done in the context of touch screen Smartphone. An emotion sensing approach has been recommended along with a proposed affective entity scoring algorithm [1]. This algorithm maintains affective scoring vectors for various entities in a mobile device. It keeps track of installed applications, multimedia contents and contacts of people and also calculates the difference between prior and posterior emotional states. Then some recommendation is proposed based on the emotional state. The device can create user's preferences from the variance in his emotional states. Then according to those preferences recommendations are generated such as a call log showing emoticons that expresses the current emotional state of the person to be called. The emotion detection is accomplished by collecting sensory data from the device and analyzing contextual information for example emails and text messages. Then affective entity score is calculated from a target application. While mapping the score, the usage pattern of the user and timer interval between different applications is also considered.

A model named OCC model (Ortony, Clore and Collins' computational emotion model) has been established [2] to provide a structure of 22 conditions which influence emotions and variables which affects the intensities of the emotion. From the study of facial expression of emotions, Ekman[3] defined six emotions, 'Joy', 'Anger', 'Fear', 'Disgust', 'Surprise' and 'Sadness' as basic emotions which has been largely used in the field of psychology and robotics.

Since today's Smartphone contains different low cost MEM sensors, the combination provides accurate measurement of the orientation of the device. An orientation estimation technique is proposed by fusing [4] different MEM sensors of the Smartphone. Orientation can be determined by the fusion of Accelerometer and Magnetometer but it is only effective when the device is stationary or not moving linearly. The device may also suffer from magnetic interference. Fusing gyroscope with the previously mentioned approach produces more accurate result of the orientation of the device. Gyroscope provides a quick response to change in angels and also does not suffer from problems like interference. But there are some bias and integration errors that can be overcome by applying a Drift and Noise removal filter. The successful estimation of orientation leads to successful development of mobile games, navigation apps, augmented reality and other kind of applications.

After the invention of pressure sensor keyboards for desktop computers some research works have been done also. One of them was biologic verification based on these keyboards [5] which was done by Hai-Rong Lv et al. and their following paper was another approach to recognize emotion by analysis the pressure sequences when any keystroke occurs [6]. Global features of pressure sequences, dynamic time warping and traditional keystroke dynamics- these three features were combined using a classifier fusion technique. Analyzing the emotion and ages of users was intended to be done in future in their paper.

The impact force on musical instrument is crucial. The built-in accelerometers, the pressure sensing capability of Android phones, and external force sensing resistors can be used to calculate this impact force on multi-touch devices such as Smartphone. Georg Essl et al. worked on these three approaches [7]. The accelerometer based approach was kind of not suitable to detect the pressure applied where the android API show some promises and FSR sensor was giving almost accurate and precise pressure values. Although it had some dynamic range it was difficult in practical use as the setup was very sensitive and calibrating with the phone and another background surface was kind of difficult. A slight movement could make the sensor unresponsive and also can give biased amount of pressure for this type of pressure sensor calibration.

Mayank Goel et al. worked on the detection of hand postures and pressure with the help of accelerometer and gyroscope. In this paper they also implemented the system in such a way that it could use the pressure applied on the screen when the vibration motor is pulsed [8]. The accuracy result was different in different cases. Their system accurately differentiates device usage on a table vs. in hand with 99.7% accuracy; when in hand it inferred the hand posture having 84.3% accuracy. They could differentiate among three types of pressure by using gyroscope sensor. Higher and lower frequencies were generated from touch-induced vibrations increase with increase in pressure. One of the limitations (as we are mainly focusing on the part of pressure sensing) they faces was they could successfully detect three levels of pressure. These three levels of pressure that they have collected were mainly to make it easy for the user with the accepted levels of accuracy. Again although the

range of the pressure was different for different types of user the system could practically detect high levels of accuracy yet we need to detect exact amount of pressure, not the levels so that the mapping with the emotion can be successfully done.

Tatsuya Shibata and Yohei Kijima demanded that not only the facial expressions but also different body parts and gestures play an affective role in non-verbal communication which they further analyzed and found relation between different body gestures and human emotion. They said in their paper that the body gestures have same structure as the facial expressions which can be calculated with the help of using a pressure sensor on a chair (that detects the amount of pressure exerted on the chair) and used some accelerometers to detect the positions of different body parts like angle of neck, leg and arms. Three types of emotional status (arousal, pleasantness and dominance) were defined in their paper but an important limitation was they could not use effective sensors to detect arm and leg positions to estimate the "dominance" or unattached sensors for body parts.

IV. Research Challenges

In this section we are focusing on the challenges against our research work in mobile perspective.

a) Absence of pressure sensor

Smartphones that are available in the market are still lacking pressure sensor embedded underneath its screen. Here, we need to mention that the biometric pressure sensor is already available but that is responsible for calculating the pressure of the atmosphere, not the tap pressure exerted on the screen of the device.

b) Resource Limitations

Mobile phones now-a-days have extremely good processing capabilities yet they are still not comparable with the powerful processors and hardware of the desktop computers. On the contrary, the battery technology has not been improved that much. The processors consume most of the battery life. It became a major issue in case of working with so many sensors.

c) Limitations with the sensors

The sensors come up with many of the mobile phones are not always absolutely accurate. For example we have seen many Smartphone with problematic gyroscopes and accelerometers. These devices do not detect the rotation of the device accurately. The quality of the camera is still not up to the mark and so face recognition technologies sometimes fail to detect correctly. So accurate algorithms should be designed so that these biased raw sensor values can be take into control.

d) Possibility of biased values

There is a chance of biased results as well. For example, when someone is stressed he can sleep, lie down on the chair or can keep working as was doing extra hours in office. These are just examples that show how people can hide their actual emotional state.

e) Privacy Concerns

Too much usage of sensors may cause serious violation of the privacy of the users. Camera photos or voice from the microphone or accelerometer data that can detect user's state or GPS data that can detect the position of the user can be sensitive data.

V. Problem Description

While using a touch screen phone the user generally does not require any button to perform any task. The user interface is built in such a way that the user will have to touch specific portions of the screen to complete different actions. For example, while typing a message a keyboard will be shown on the screen where he will touch desired letters to write the message. While using the phone, emotional state change will affect his interaction with the device. Normally larger touched area is also responsible to show a larger amount of pressure value and vice-versa. Again, due to anger he may shake the phone harder than the usual and may be remain steady and calm in case of sadness. So, the gyroscope signals will vary according to his emotional states. The accelerometer will exhibit continuous changes of the orientation of the device. Moreover, the interaction log will contain the type of usage, type of application, user tasks, time of usage and other necessary information. The challenge is to synchronize all these data, correlate them with the user's interaction log and normalize the values into a specific range. The resulting data set can be mapped with the different emotional states of the user.

VI. Proposed Approach

We have already defined that we are going to build our system for touch screen smart phones. Some of the best operating systems for the Smartphone are iOS (Apple products like Iphone, Ipad), android (an open source project of Google), windows and blackberry OS. We decided to build our system using android operating system because - (1) Android is an open source project of Google, so resources are quite easy to manage, (2) as an open source project most of the leading devices are running on this OS, leading Smartphone manufacturers like Samsung, HTC, Motorola, Sony are building their phones using this OS, so it will be easier to test our dataset.

We have considered three approaches to solve the problem we have mentioned. In the first case, the data will be collected based on the touch input and pressure applied on the screen. It is assumed that due

to different emotional states user will give different amount of pressure on the screen. So depending on the pressure we will map the emotional states later on. We are going to use the approximated tap pressure values from touch screen.

Now the second approach is we will work with an external sensor which is called Force Sensing Resistor (FSR). The need of an external sensor is obvious here since the current phones do not have any built in pressure sensor. So to collect more accurate and precise data of the pressure applied, we are focusing on this sensor also. This external sensor will provide accurate tap pressure values on the smart-phone touch screen and it will be attached with the phone externally as well. When the user is stressed or in an angry mood he will generate more pressure on the device. The pressure applied on the device will passed on the FSR and we will get some readings from that sensor. It will exhibit significant variations in the readings for other emotional states we have assumed.

Now the third case, we have to collect readings from the built-in sensors such as the accelerometer and gyroscope data and log them into files. It will provide us with a concrete idea about how these sensor values vary when orientation or movement of the device is changed in different emotional states. So we will use this data to for recognition of the emotional state of the user. Another approach is we will also collect the interaction log. Depending on the interaction log we will get the idea about what type of applications the user is using and also the amount of time that he is spending on each application. These interaction log data will be analyzed with the data collected from sensors.

VII. Conclusion and Future Works

The recognition of human affection is one of the most challenging research areas. Emotion based user interfaces or service recommendations are gaining popularity. Even though there are possibilities for emotion aware application development in desktop or web platform, we have chosen the mobile platform for its flexibility, mobility and frequent usage. As these devices have various sensors integrated in their hardware, there is a limited need for additional equipments. This encourages the study of human affection in mobile devices particularly Smartphone even more. There are various challenges such as linking the sensor values with user's emotional states and collecting user interaction log data.

A solution is proposed in our work to overcome these problems as well.

In future we will implement software in mobile operating system (e.g. Android/iOS) that will run as a background service collecting the data from the built in sensors. Another application will record the size of the area that has been touched. That software will eventually save the data locally and then we will extract the features so that that data can be mapped with different emotional states. We will collect data from the external FSR attached with the Smartphone to get exact pressure data. Then finally we will apply some machine learning approaches to the selected dataset and train our system to detect human affection.

References Références Referencias

1. Hyun-Jun Kim, Young Sang Choi "EmoSens: Affective Entity Scoring, A Novel Service: Recommendation Framework for Mobile Platform"- October 23–27, 2011, Chicago, Illinois, USA
2. Ortony, A., Clore, G., & Collins, A. (1988).The Cog - nitive Structure of Emotions. Cambridge: Cambridge University Press.
3. P. Ekman and R. Davidson, "The Nature of Emotion: Fundamental Questions", Oxford University Press, 1994
4. Shahid Ayub, Alireza Bahraminisaab, Bahram Honary , "A Sensor Fusion Method for Smart phone Orientation Estimation", PGNet, 2012
5. H. R. Lv, "Biologic verification based on pressure sensor keyboards and classifier fusion techniques,' IEEE trans .On Consumer Electronics, 52(3): 1057-1063, 2006
6. Hai-Rong LV, Zhong-Lin LIN, Wen-Jun YIN, Jin DONG, "Emotion Recognition Based on Pressure Sensor Keyboards", copyright ©2008 IEEE
7. Georg Essl , Michael Rohs , Sven Kratz "Use the Force (or something) -Pressure and Pressure-Like Input for Mobile Music Performance" - NIME2010, 15-18th June 2010, Sydney, Australia
8. Mayank Goel, Jacob O. Wobbrock, Shwetak N. Patel GripeSense: Using Built-In Sensors to Detect Hand Posture and pressure on Commodity Mobile Phones. UIST '12, October 7-10, 2012, Cambridge, Massachusetts, USA
9. Shumei Zhang, Paul McCullagh, Chris Nugent, Huiru Zheng Activity Monitoring Using a Smart Phone's Accelerometer with Hierarchical Classifi - cation
10. Scott Brave and Clifford Nass, "Emotion in Human Computer Interaction", Stanford University
11. "Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms" by Ralph Adolphs, University of Iowa College of Medicine- Behavioral and Cognitive Neuroscience Reviews, Volume 1 Number 1, March 2002 21-62
12. Nicu Sebe, Ira Cohen, Theo Gevers , and Thomas S. Huang, "Multimodal Approaches for Emotion Recognition: A Survey"
13. Clayton Epp, Michael Lippold, and Regan L. Mandryk, "Identifying Emotional States using Keystroke Dynamics", CHI 2011, May 7–12, 2011, Vancouver, BC, Canada

14. Sears, A., Lin, M., Jacko, J., and Xiao, Y. When Computers Fade Pervasive Computing and Situat - ionally-Induced Impairments and Disabilities. HCI International 2, (2003), 1298-1302.
15. Azenkot, S. and Zhai. S. Touch Behavior with Diffe - rent Postures on Soft Smartphone Keyboards. Proc. Moble HCI 2012, (2012)

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Measurement and Prediction of Software Performance by Models

By G. Kasi Reddy & Dr. D Sravan Kumar

JNTU, India

Abstract- Software Performance Engineering (SPE) provides a systematic, quantitative approach to constructing software systems that meet performance objectives. It prescribes ways to build performance into new systems rather than try to fix them later. Performance is a pervasive quality of software systems; everything affects it, from the software itself to all underlying layers, such as operating system, middleware, hardware, communication networks, etc. Software Perfor -mance Engineering encompasses efforts to describe and improve performance, with two distinct approaches: an early-cycle predictive model-based approach, and a late-cycle measurement-based approach. Current progress and future trends within these two approaches are described, with a tendency (and a need) for them to converge, in order to cover the entire development cycle.

Keywords : *SPE, performance prediction, performance measurement, UML, debugging.*

GJCST-C Classification : *H.1*

Strictly as per the compliance and regulations of:

Measurement and Prediction of Software Performance by Models

G.Kasi Reddy[α] & Dr. D Sravan Kumar[σ]

Abstract- Software Performance Engineering (SPE) provides a systematic, quantitative approach to constructing software systems that meet performance objectives. It prescribes ways to build performance into new systems rather than try to fix them later. Performance is a pervasive quality of software systems; everything affects it, from the software itself to all underlying layers, such as operating system, middleware, hardware, communication networks, etc. Software Perfor - mance Engineering encompasses efforts to describe and improve performance, with two distinct approaches: an early-cycle predictive model-based approach, and a late-cycle measurement-based approach. Current progress and future trends within these two approaches are described, with a tendency (and a need) for them to converge, in order to cover the entire development cycle.

Keywords: SPE, performance prediction, performance measurement, UML, debugging.

I. Introduction

Despite rapidly improving hardware, many recent software systems are still suffering from performance problems, such as high response times or low throughputs [1]. Hardware is often not the limiting factor as powerful multi-core and many core processors are readily available on the market and modern software systems may run in huge data centers with virtually unlimited resources. Performance problems often stem from software architectures that are not designed to exploit the available hardware. Instead, these software architectures ignore the advances of distributed computing and multi-core and many core processors.

Systematic approaches for engineering softw- are systems to achieve desired performance properties have been proposed [2, 3]. They advocate modeling software systems during early development stages, so that performance simulations can validate design decisions before investing implementation effort.

The advent of multi-core processors results in new challenges for these systematic software perfor- mance engineering (SPE) methods. Modeling software running on thousands of cores requires rethinking of existing approaches [4]. While techniques and tools for parallelizing software are evolving [5], novel methods and tools need to be created to assist software in designing systems that can exploit the capabilities for parallel execution but do not overburden software developers during implementation [6].

Author α : Research Scholar, CSE, JNTU Hyderabad, India.
e-mail : gkreddy@mgit.ac.in
Author σ: Principal and Professor of CSE, KITE women's college of Professional Engineering Sciences, Hyderabad, India.

II. Software Performance Engineering

SPE is a software-oriented approach; it focuses on architecture, design, and implementation choices. It uses model predictions to evaluate trade-offs in software functions, hardware size, quality of results, and resource requirements. The models assist developers in controlling resource requirements by enabling them to select architecture and design alternatives with accep- table performance characteristics. The models aid in tracking performance throughout the development process and prevent problems from surfacing late in the life cycle (typically during final testing).[7] SPE also prescribes principles and performance patterns for cre - ating responsive software, performance anti-patterns for recognizing and correcting common problems, the data required for evaluation, procedures for obtaining perfor - mance specifications, and guidelines for the types of evaluation to be conducted at each development stage. It incorporates models for representing and predicting performance as well as a set of analysis methods.[8]

III. Progress in Measurement, Debugging and Testing

Measurement is used by verification teams to ensure that the system under test meets its specifications, by performance modelers to build and validate models, and by designers to find and fix hot- spots in the code. Interest in the measurement of the performance of a computer system ranges back to the development of the very first systems, described in an early survey paper by Lucas [9]. Today, the state of industrial performance measurement and testing techn - iques is captured in a series of articles by Scott Barber [7] including the problems of planning, execution, instrumentation and interpretation. For performance test design, an important issue is to determine the workload under which the testing is done. An approach is to run the performance tests under similar conditions with the expected operational profile of the application in the field [9]. Briand and co-workers have pioneered the use of models to create stress tests for time-critical systems, by triggering stimuli at strategic instants [10].

Performance models are often difficult to construct, even with a live system, despite the presence of tools to actually measure performance. In the future, model building will become much more automated, and output becomes standardized, and the conversion process between measurement information and performance model becomes more practical. Ultimately, the model and measurement information will be fed back into design tools, so that performance issues are brought to the forefront early in the design process.

a) Performance Measurement- Best practices

These are practices for those responsible for measuring software performance and for performance testing. [11]

i. *Plan Measurement Experiments to Ensure That Results Are Both Representative And Reproducible*

There are two key considerations in planning performance measurements: They must be repre - sentative and reproducible. To be representative, meas - urement results must accurately reflect each of the factors that affect performance: workload, software, and computer system environment. The goal is to design your measurement experiment in a way that balances the effort required to construct and execute the measurement experiment against the level of detail in the resultant data. When unimportant details are omitted, both the design effort and the overhead required to collect the data are reduced.

Reproducibility gives you confidence in the results. In order for a measurement to be reproducible, the workload, software, and computer system environment must be controlled so that you can repeat the measurement and get the same (or very similar) results each time.

ii. *Instrument Software to Facilitate SPE Data Collection*

You instrument software by inserting code (probes) at key points to measure pertinent execution characteristics. For example, you might insert code that records the time at the start and end of a business task to measure the end-to-end time for that task. There are at least three reasons for supplementing the standard tools with instrumentation: convenience, data gran - ularity, and control.

iii. *Measure Critical Components Early and Often to Validate Models and Verify Their Predictions*

Measurements substantiate model predictions, and confirm that key performance factors have not been omitted from the models. Occasionally, software exec - ution characteristics are omitted from a model because their effects are thought to be negligible. Later, you may discover that they in fact have a significant impact on performance, as illustrated in the following anecdote: An early life cycle model specified a transaction with five database "Selects." During detailed design, "Order by" clauses were added to three of the "Selects." The developers viewed the additional clause as "insign- ificant" because only one to five records would be sorted for each "Select." Upon investigation, though, the performance analyst discovered that over 50,000 instructions were executed for each sort!

The way to detect these omissions is to measure critical components as early as possible and continue measuring them, to ensure that changes do not invalidate the models.

IV. Prediction of Performance by Models

The special capability of a model is prediction of properties of a system before it is built, or the effect of a change before it is carried out. This gives a special "early warning" role to early-cycle modeling during requirements analysis. However as implementation proceeds, better models can be created by other means, and may have additional uses, in particular

- design of performance tests
- configuration of products for delivery
- evaluation of planned evolutions of the design, recognizing that no system is ever final.

a) Performance models from scenarios

Early performance models are usually created from the intended behaviour of the system, expressed as scenarios which are realizations of Use Cases. The term "scenario" here denotes a complex behavior including alternative paths as well as parallel paths and repetition. The performance model is created by extra - cting the demands for resource services. Annotated UML specifications are a promising development.

The annotations include:

- the workload for each scenario, given by an arrival rate or by a population with a think time between requests,
- the CPU demand of steps,
- the probabilities of alternative paths, and loop counts,
- the association of resources to the steps either impl -icitly (by the processes and processors) or explicitly.

As an illustration, Figure 1 shows a set of applications requesting service from a pool of server threads running on a multiprocessor (deployment not shown). Part (a) shows the scenario modeled as a UML sequence diagram with SPT annotations, (b) shows a graph representing the scenario steps, and (c) shows the corresponding layered queueing network (LQN) model. Studies in [12] [13] use such models.

At a later stage, scenarios may be traced from execution of prototypes or full deployments, giving accurate behaviour. Models can be rebuilt based on

these experimental scenarios [14], combined with mea - sured values of CPU demands.

b) Performance models from objects and components

A performance model can be built based on the software objects viewed from a performance persp - ective. A pioneering contribution in this direction defined a "performance abstract data type" for an object [13], based on the machine cycles executed by its methods.

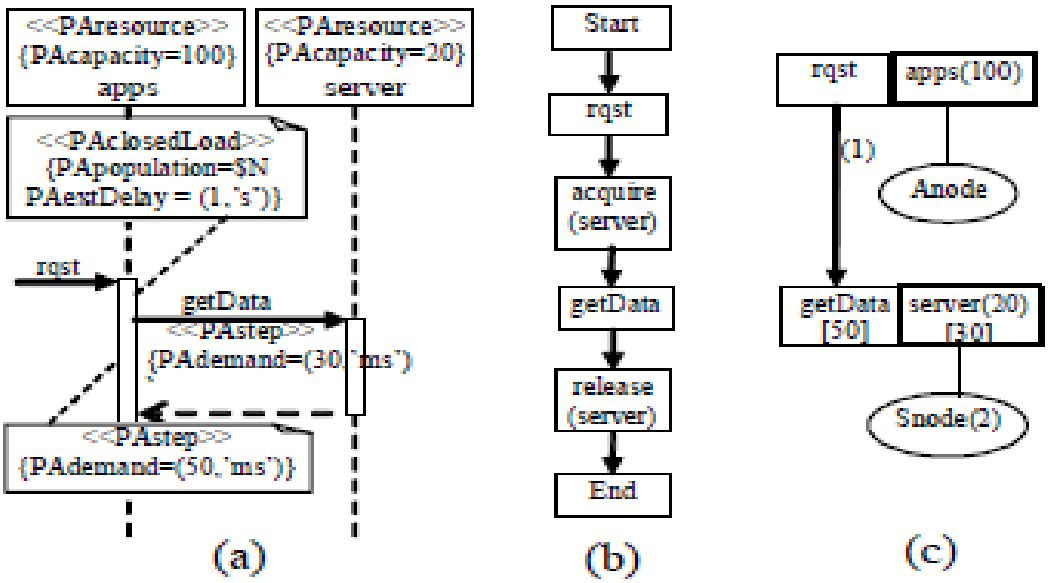


Figure 1 : Annotated UML, Scenario Model, and Performance Model

To create a performance model, one traces a response from initiation at a root object to all the interfaces it calls, proceeding recursively for each call. Queueing and layered queueing models were derived based on objects and calls in [14] and [15]. Model parameters (CPU, call frequencies) were estimated by measurement or were based on the documentation plus expertise. Object-based modeling is inherently compo - sitional, based on the call frequencies between objects. This extends to subsystems composed of objects, with calls between subsystems. In [2] an existing application is described in terms of UNIX calls, and its migration to a new platform is evaluated by a synthetic benchmark with these calls, on the new platform. This study created a kind of object model, but then carried out composition and evaluation in the measurement domain. The convergence of models and measurements is an important direction for SPE.

The object-based approach to performance modeling can be extended to systems built with reusable components. Composition of sub-models for Component-Based Software Engineering [16] was described in [17]. Issues regarding performance contr - acts between components are discussed in [18]. Components or platform layers can be modeled sepa - rately, and composed by specifying the calls between them. For example, in [18] a model of a J2EE application server is created as a component that offers a large set of operations; then an application is modeled (by a scenario analysis) in terms of the number of calls it made to each operation.

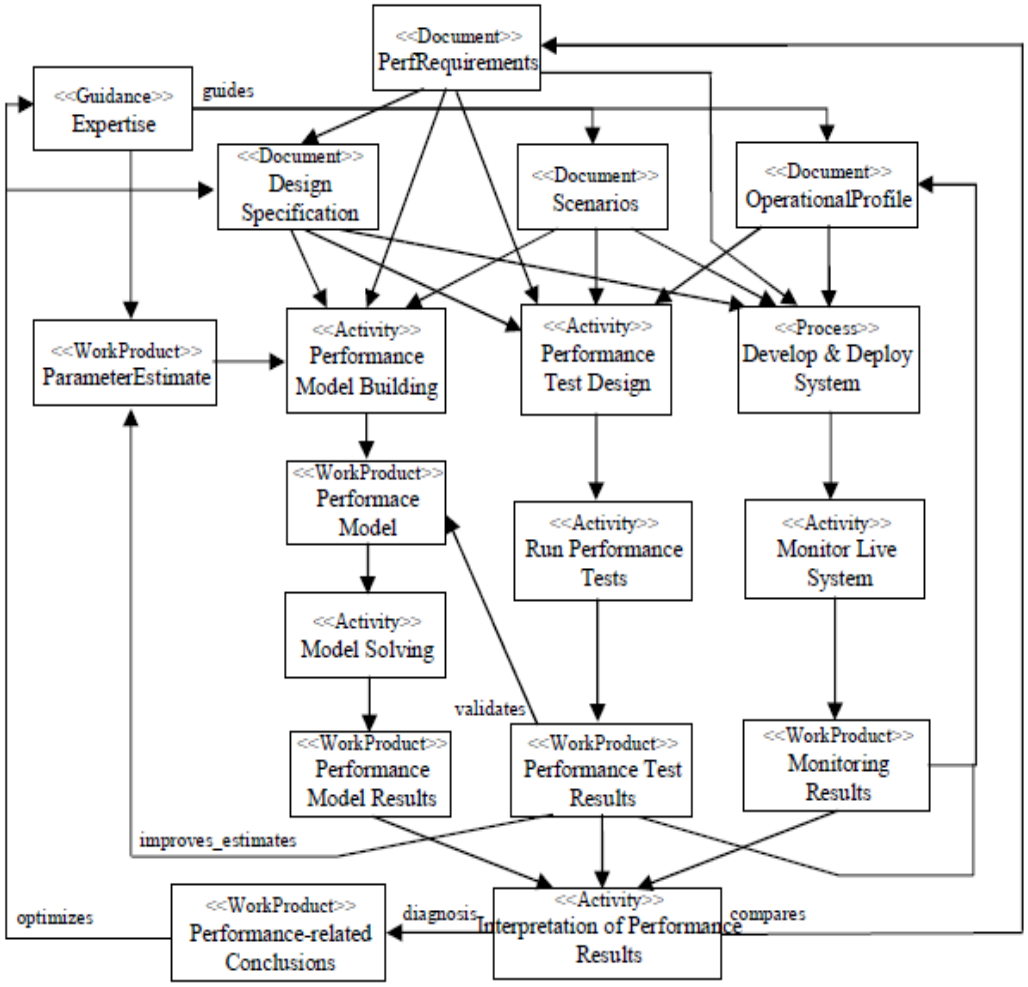


Figure 2 : Simplified domain model for a converged SPE process

The quantitative parameters of the performance model for the J2EE server - and the underlying operating system and hardware platform -were obtained by measurements for two different implementations. The main challenge regarding performance characterization of reusable components stem from the fact that the offered performance depends not only on the component per se, but also on its context, deployment, usage and load. It seems obvious that such approaches apply similarly to Generative techniques [17] and to Model-Driven Development. The *completion* of perfo - rmance models made from a software design, by adding components that make up its environment but are outside the design, is also largely based on composition of sub-models [19]. This is an aspect of Model-Driven Development.

V. Convergence of the Measurement and Modeling Approaches

The present state of performance engineering is not very satisfactory, and better methods would be welcome to all. One way forward is to combine knowledge of different kinds and from different sources into a converged process. Figure 2 outlines such a process, with the main concepts and their relationships. The notation is based on the newly adopted OMG standard Software Process Engineering Meta model (SPEM) [20]. At the core of SPEM is the idea that a software process is a collaboration between abstract active entities called *ProcessRoles* (e.g., usecase act- ors) that perform operations called *Activities* on conc - rete entities called *WorkProducts.* Documents, models, and data are examples of WorkProduct specializations. Guidance elements may be associated to different model elements to provide extra information.

Figure 2 uses stereotypes defined in [20]. Concepts related to the model-based approach appear on the left of Figure 2, and to the measurement-based approach on the right. A distinction is made between performance testing measurements (which may take place in a laboratory setting, with more sophisticated measurement tools and special code instrumentation) and measurements for monitoring live production systems that are deployed on the intended target system and used by the intended customers. The

domain model from Figure 3 is very generic. For instance, there is no indication whether different activities (such as Performance Model Building) are done automatically through model transformations or "by hand" by a performance analyst.

In a convergence of data-centric and model-centric methods, data (including prior estimates) provides the facts and models provide structure to organize and to extract significance from the facts. Our exploration of the future will examine aspects of this convergence. Models have a key role. They integrate data and convert it from a set of snapshots into a process capable of extrapolation. To achieve this potential we must develop robust and usable means to go from data to model (i.e., model-building) and from model to "data" (solving to obtain predictions). We must also learn how to combine measurement data interpretation with model interpretation, and to get the most out of both. Capabilities supported by convergence include:

- efficient testing, through model-assisted test design and evaluation
- search for performance-related bugs,
- performance optimization of the design
- scalability analysis
- reduced performance risk when adding new features,
- aids to marketing and deployment of products.

The future developments that will provide these capabilities are addressed in the remainder of this section. A future tool suite is sketched in Figure 3.

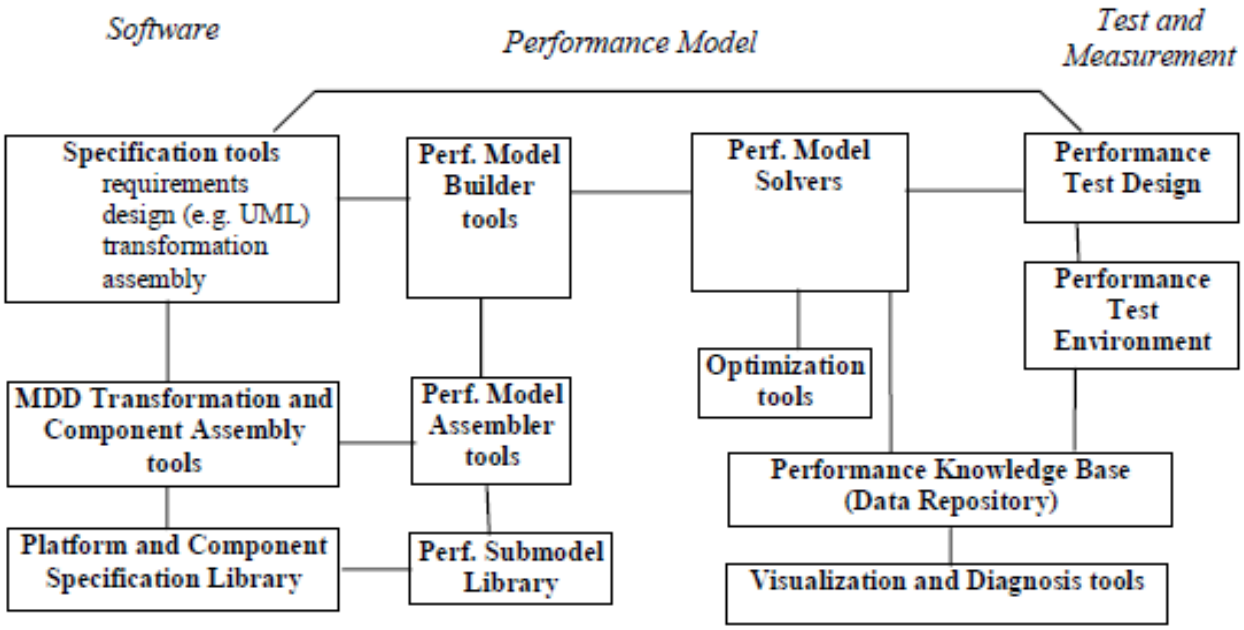


Figure 3 : Tools for a Future Converged SPE Process, linked to Software Development Tools

VI. Efficient Model-Building Tools

The abstractions provided by performance models are valuable, but some way must be found to create the models more easily and more quickly. For performance models made early in the lifecycle from specified scenarios, automated model-building has be - en demonstrated [6] and is supported by the UML profiles [21]. The future challenge is to handle every scenario that a software engineer may need to describe, and every way that the engineer can express them (including the implied scenario behaviour of object call hierarchies, and the composition of models from component designs).

The multiplicity of model formats hinders tool development, and would be aided by standards for performance model representations, perhaps building on [22]. Interoperability of performance building tools with standard UML tools is also helpful. For instance, the PUMA architecture[23] shown in Figure 6 supports the generation of different kinds of performance models (queueing networks, layered queueing networks, Petri nets, etc.) from different versions of UML (e.g., 1.4 and 2.0) and different behavioural representations (sequ - ence and activity diagrams). PUMA also provides a feedback path for design analysis and optimization. Mid and late-cycle performance models should be extracted from prototypes and implementations. Trace based automated modeling has been described in [23], including calibrated CPU demands for operations. Fut - ure research can enhance this with use of additional instrumentation (e.g. CPU demands, code context), efficient processing, and perhaps exploit different levels of abstraction. Abstraction from traces exploits certain patterns in the trace, and domain-based assumptions; these can be extended in future research.

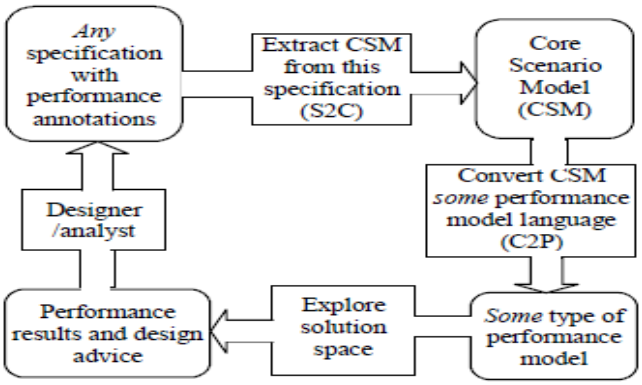


Figure 4 : Architecture of the PUMA toolset [23]

VII. Conclusion

Software Performance Engineering needs further development in order to cope with market requirements and with changes in software technology. It needs strengthening in prediction, testing and measurement technology, and in higher-level techni - ques for reasoning and for optimization.

References Références Referencias

1. M.R. Moghal, N. Hussain, M.S. Mirza, M.W. Mirza, M.S. Choudry, "Performance Evaluation and Modeling of Web Server Systems",WSEAS *Transa - ctions on Information Science and Applications* 1(1), July 2004, pp. 658-663.
2. C. U. Smith and L. G. Williams, Performance Solutions: A Practical Guide to Creating Res - ponsive, Scalable Software. Addison-Wesley, 2002.
3. M. Woodside, G. Franks, and D. C. Petriu, \The Future of Software Performance Engineering," in Proc. Future of Software Engineering (FOSE'07). IEEE Computer Society, 2007, pp. 171-187.
4. W. Hwu, S. Ryoo, S.-Z. Ueng, J. Kelm, I. Gelado, S. Stone, R. Kidd, S. Baghsorkhi, A. Mahesri, S. Tsao, N. Navarro, S. Lumetta, M. Frank, and S. Patel, \Implicitly parallel programming models for thousand-core microprocessors," in Proc. 44th ACM/IEEE Design Automation Conference (DAC '07), june 2007, pp. 754-759.
5. H. Vandierendonck and T. Mens, \Techniques and tools for parallelizing software," IEEE Softw., vol. 29, no. 2, pp. 22-25, 2012.
6. C. A. Schaefer, V. Pankratius, and W. F. Tichy, \Engineering parallel applications with tunable architectures," in Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering, Volume 1, ser. ICSE '10. ACM, 2010, pp. 405-414.
7. S. Barber, "Creating Effective Load Models for Performance Testing with Incomplete Empirical Data", in *Proc. 6 th IEEE Int. Workshop on Web Site Evolution,* 2004, pp. 51-59.
8. Bertolino, R. Mirandola. "Software performance engineering of component-based systems", *Proc. Workshop on Software and Performance,* WO - SP'2004, pp 238-242.
9. H. Lucas Jr, "Performance evaluation and monit - oring", *ACM Computing Surveys,* 3(3), Sept. 1971, pp 79-91.
10. V. Garousi, L. Briand, Y. Labiche, "Traffic-aware Stress Testing of Distributed Real-Time Systems based on UML Models", *Proc. Int. Conference on Software Engineering,* Shanghai, China, 2006, pp. 391-400.
11. V. Cortellessa, A. Di Marco, P. Inverardi, "Nonfun - ctional Modeling and Validation in Model-Driven Architecture", *Proc 6 th Working IEEE/IFIP Confer - ence on Software Architecture* (WICSA 2007), Mumbai, India, 2007.
12. C. Jittawiriyanukoon, Performance Evaluation of Parallel Processing Systems Using Queueing Network Model *WSEAS Transactions on Computers* 3(5), March 2006, pp. 612-620.
13. D.B. Petriu, C.M. Woodside, "Analysing Software Requirements Specifications for Performance", *Proc. 3rd Int. Workshop on Software and Perform - ance,* Rome, 2002.
14. M. Woodside, C. Hrischuk, B. Selic, and S. Bayarov, "Automated Performance Modeling of Software Generated by a Design Environment", *Performance Evaluation*, vol. 45, no. 2-3 pp. 107-124, 2001.
15. J. Xu, M. Woodside, and D.C. Petriu, "Performance Analysis of a Software Design using the UML Profile for Schedulability, Performance and Time", in Proc. *13 th Int. Conf. Modeling Techniques and Tools for Computer Performance Evaluation*, Urbana, USA, Sept. 2003
16. C. Szypersky, D. Gruntz, S. Murer, *Component Software: Beyond Object Oriented Programming*, Addison- Wesley, 2002.

17. X. Wu and M. Woodside, "Performance Modeling from Software Components", in *Proc.WOSP'2004,* Redwood Shores, Calif., 2004, pp. 290-301.
18. R. H. Reussner, V. Firus, S. Becker, "Parametric Performance Contracts for Software Components and their Compositionality", in *9th Int. Workshop on Component- Oriented Programming,* Oslo, June 2004.
19. M. Woodside, D.B. Petriu, K. Siddiqui, "Performan - cerelated Completions for Software Specifications", *Proc 24th Int. Conf. on Software Engineering*, Orlando, 2002.
20. Object Management Group, *Software Process Engineering Metamodel Specification*, formal/05-01-06, 2006.
21. Object Management Group, *UML Profile for Modeling and Analysis of Real-Time and Embedded systems (MARTE) RFP,* OMG doc. realtime/05-02-06, 2005.
22. C.U. Smith, C. M. Lladó, V. Cortellessa, A. diMarco, L. Williams, "From UML models to software performance results: an SPE process based on XML interchange formats", in *Proc WOSP'2005,* Palma de Mallorca, 2005, pp. 87-98.
23. M. Woodside, D.C. Petriu, D.B. Petriu, H. Shen, T. Israr, J. Merseguer, "Performance by Unified Model Analysis (PUMA)", *Proc. WOSP'2005,* Mallorca, pp 1-12.
24. Erika Corona, Filippo Eros Pani, "A Review of Lean-Kanban Approaches in the Software Development", *WSEAS Transactions on Information Science and Applications,* Vol. 10, Issue 1, pp. 1-13, January 2013.
25. Manoj Kumar Tyagi, Srinivasan M., L. S. S. Reddy, "Design of Traditional/Hybrid Software Project Tracking Technique: State Space Approach", *WSEAS Transactions on Information Science and Applications,* Vol. 11, Issue 11, pp. 345-355, November 2013.

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Review-: Benefits and Critical Factors of Customer Relationship Management

By Pooja Nahar & Vijay Pal Dhaka
Jaipur National University, India

Abstract- Customer Relationship Management (CRM) is a technical jargon which is a blend of methodologies, software and internet, which are used by a company to achieve its goal through the identification and satisfaction of customer's stated and unstated needs and wants. This software addresses customer life cycle management. This system manages company interactions with current and future customers. It involves technology to organize, automate and synchronize business processes. CRM application is an essential tool for a company to grow and help to increase the satisfaction of customers. There are many benefits of CRM; those make the market environment customer centric. In this paper, we reviewed previous studies and identify those benefits which affect customers and company both. But CRM has many problems also because of them CRM gets failure. Its failure rate is more than its success rate. We also elaborated its failure factors and along with them its critical success factors which help in making CRM a successful project for a company, however implementation of CRM is a complex task.

Keywords: *customer relationship management, information technology, software.*

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

A Review-: Benefits and Critical Factors of Customer Relationship Management

Pooja Nahar [α] & Vijay Pal Dhaka [σ]

Abstract- Customer Relationship Management (CRM) is a technical jargon which is a blend of methodologies, software and internet, which are used by a company to achieve its goal through the identification and satisfaction of customer's stated and unstated needs and wants. This software addresses customer life cycle management. This system manages company interactions with current and future customers. It involves technology to organize, automate and synchronize business processes. CRM application is an essential tool for a company to grow and help to increase the satisfaction of customers. There are many benefits of CRM; those make the market environment customer centric. In this paper, we reviewed previous studies and identify those benefits which affect customers and company both. But CRM has many problems also because of them CRM gets failure. Its failure rate is more than its success rate. We also elaborated its failure factors and along with them its critical success factors which help in making CRM a successful project for a company, however implementation of CRM is a complex task.

Keywords: customer relationship management, inform - ation technology, software.

I. Introduction

The essence of CRM, related to Information Technology (IT), resolution is the opportunity to build better relationships with customers. CRM has the ability to provide highly interactive, experienced & long term relationships than ever before. CRM is the discipline of improving the way customers are managed through changes in business processes, people, organization & technology [1, 2, 3]. The new business environment has different touch points of customers like managing enquiries, preferences, better service delivery and support. Many companies are attracting towards CRM, which is customer centric and efficient [2].

In paper [1], author first represents the introduction of CRM and after that it explores benefits and failure reasons critical components that enable the successful implementation. A recent survey reveals that investment in CRM application is very high but its success rate is very low. Author stated that without an understanding of CRM initiatives, CRM projects generally fail [1, 3]. CRM is not a merely technological application, when it fully implemented it is a customer driven, technology integrated business process management strategy [1, 3]. CRM is basically used to understand the customers, attract new customers and retain the old ones [2].

The key drivers for a CRM program within the Public Sector tend to be the improvement of services to citizens with efficiency by phone, face to face, mails, internet, SMS etc. [2]. Similarly, the CRM application link includes front office and back office functions, including internet, email, sales, telemarketing, call center etc. [1]. The central component of a CRM solution is a single shared customer database, allowing information to be collected once but used many times. CRM has the potential for achieving success & growth for organizations in this competitive environment & rapid technology development. CRM enables organizations to know their customers better to build sustainable & long standing relationships with them. In the changing market environment, Public Sectors have to use modern methods, technologies & better ideas of management. Some Public Sectors are using electronic files with M.S. Office, but this system does not enable to automate reports.

Now-a-days customers are very sensitive, demanding & they are looking for fast services. So Public Sectors too have to compete in a global market. They also have to maintain a healthy relationship with customers. The effectiveness of an organization depends on the interest, values, motives, teamwork, loyalty & the way of dealing with customers. As to maintain a healthy culture with customers & achieve values, CRM has risen. CM systems can be viewed as Information System aimed at enabling organizations realize a customer focus [4].

The product centric environment is now changed in customer centric environment, now customized product and services are provided with improved customer services [3, 4]. CRM is a business strategy to select & manage customers to optimize long term values. CRM requires a customer centric business philosophy & culture to support effective marketing, sales & service processes. CRM application can enable effective CRM, provided that an enterprise has the right leadership, strategy & culture [5].

Author referred CRM as Customer Knowledge Management and he suggested that knowledge management plays a key role in CRM initiatives. To improve service & retain customers CRM synthesizes all of a company's customer touch points. Author provides a process framework which includes 4 types of

Author α σ: Jaipur National University, Jagatpura, Jaipur, Rajasthan India. e-mail: vijaypal.dhaka@gmail.com

processes- CRM Service Processes, CRM Support Processes, CRM Analysis Processes and CRM Management Process. This paper included some case studies those enhanced CRM initiatives by applying knowledge for, from or about customers. They identified an orchestrated approach that considered strategy, processes, systems and change management aspect [6].

This paper also focuses on customer in relation to the use of CRM in the selected business subject. CRM packaged s/w supported the integration of business processes, data & technology to improve customer facing processes. CRM is used to manage interactions with customers, especially in sales prospect. CRM involves using technology to organize, automate & synthesize business processes. This paper believed that the customer knows the quality of product, if he believes that company does the best for the customers not to just profit then he will be a lifetime customer. Author also supported that companies using MS Office as an electronic file and keeping records, but those are not enabled in documents administration and automated reports [7].

Paper [8] stated that for long term business survivals, it is essential to maintain good relationship with customers. For this purpose CRM is a best option. This paper tells that Business Process Reengineering (BPR) and organizational learning are the first step in constructing CRM. BPR involves staff reward and training, organizational structure and IT tools. Organizational learning involves commitment to learning, shared visions and open -mindedness.

Public Sectors also realize that maintaining relations & providing fast services to customers are necessary in these days, because Private Sectors try to compete Public Sectors in terms of better services at reasonable cost. If Public Sectors want to retain their customers, then they also have to provide fast service delivery at low cost compared to Private Sectors. That's why they need CRM in their organizations.

II. Benefits of CRM

Adopting CRM brings the knowledge to maintain long term relationship with customers, achieving the customers, achieving the customer's satisfaction, reducing distrust, removal of complaints, etc. This paper focuses on the importance of CRM in Public Sectors. In order to gain confidence in customers & keep them, IT is a supporting tool for CRM. CRM has been recognized to redesign business processes for the improvement of performance. CRM has the advantage of collecting and analyzing data, customer behavior, timely respond with effective communication and services of the customer. Here data warehousing technology helps to provide a better understanding of customer behavior. Redundant data are removed, so that only filtered and accurate information is provided to the organization so that queries can be solved in a fast manner [1].

Paper [2, 3] stated that CRM helps in decision making while increasing profitability with decreasing cost of management. This also explains that CRM provides new ways to communicate through better services as it integrates with databases.

As paper [3] elaborated benefits of CRM from only customer's point of view and stated eight main topics, which are- improve customer service, increased personalized service, responsive to customer needs, customer segmentation, improve the customization of marketing, multichannel integration, time saving and improve customer knowledge. After using CRM, Organization can keep a customer's record for their future needs and they can provide better pricing of products. Successful CRM promotes customers to purchase more, it increases the benefits of the organization [4].

This improved customer satisfaction and more focused on customer needs. Satisfying customer means increasing revenue and income for the business. Another benefit is elaborated in paper [3] is the reduction of paper; this saves time and money both. More storage and retrieval capacity is another benefit by which system can organize each and every detail of individual. CRM helps in the segmentation of the customers like by gender, ages, geographic, etc. which is helpful in satisfying their needs [3].

CRM provides customer satisfaction, competitive advantages, profitability, reduction of distrust and cost optimization [7, 8].

CRM consolidates contacts, and center all front office operations. It reduces the cost of customer maintenance, because winning a new customer is 5-6 times higher than maintaining existing customer. It reduces the repeat calls for complaints. CRM transforms the market from a product oriented to customer oriented. CRM provides better targeted & more efficient services to its community [8, 3].

CRM increases in product time results in reducing errors & rework, due to consistency in information over compliance burden. CRM provides efficient & speedy transactions with less processing time. Queries of customers can be solved quickly, because it is directly sent to the subject matter experts. After using CRM, all the tasks and responsibilities are fixed for each employee, so it reduces staff absenteeism & turnover. It provides high employee satisfaction also.

III. Some Failure Factors

Even after recognizing so many benefits of CRM, a successful implementation is elusive to many companies, mostly because they do not understand that CRM requires a company wide, cross functional, cust -

omer focused business process re-engineering. Although a large portion of CRM is a technology, viewi - ng CRM as a technology – only solution is likely to fail. Managing a successful CRM implementation requires an integrated & balanced approach to technology, process & people. CRM implementation is a holistic and complex concept, which means that it is not merely an integration of new information technology, but everything that happens around the business process changes [1].

Many of the conversations, businesses have about CRM solutions begin and end with technology, and businesses that leave CRM in the hands of the IT department almost immediately set them up for failure. CRM implementation is a complex and challenging task, because it changes the whole process or manner of working of the business. Approximately 65% of CRM implementation results in failure. Less knowledge of CRM definition, scope, processes and technology also leads to failure. Generally, companies underestimated the complexity of CRM and invest heavily in CRM software [1, 4, 7].

CRM has its drawback also; its solutions are not with technology and cannot be work with only IT Tools [1, 2, 7].

Most CRM implementations are governed by an incomplete vision, While CRM failures are not entirely prescriptive, a relatively small number of contributing factors are cited in the majority of CRM projects gone wrong. If commitment to the customers is absent, then also CRM fails. In the software selection process generally users don't involve, but it also would be a reason of failure. Unclear set of objectives may lead to less integrity and bad quality in data, this also makes CRM unsuccessful [2].

Managers even don't know what kind of advantages CRM can provide, that's why they won't be able to implement properly [3]. Failing in selecting the optimal CRM s/w and bad selection of the project team are the biggest pain points of an organization, is a possible reason for failure of the CRM project [4].

We also stated that Software marks a paradigm shift delivery of services. There is a resistance to change against it. Management people have to be visibly and vocally committed to CRM success — Absence of such commitment is clearly associated with CRM failures. Untrained staff may be another reason of failure of CRM software; it changes the working manner from manual to computerize. Failing to involve line staff in software selection makes CRM project a big failure, because it meets the executives' requirements for information reporting, but not the users' requirements for particular feature sets, ease of use, data entry methods, data management tools and automation capabilities. One of the most common causes of CRM failure is failing in proper implementation of CRM strategy as a software or technology project. In this IT department has to play a key role in installing the system, getting it running. If IT infrastructure would not consider before purchasing software then it will result in failure. CRM implementation will fail, if software is not able to deliver to its customers or viewpoint of the customers. Data integrity and Data quality should be maintained in the software. Otherwise, this dirty, duplicate, incomplete and bogus data will be a big reason of failure of CRM software implementations. Before purchasing software, organizations should warranty a defined time, budget and customization of software. This time and budget overrun could be a reason of failing CRM software. User adoption of a software in an organization could be a reason of failure because it end user does not adopt the new working manner, then it will likely end up failing.

IV. Critical Success Factors

It is generally observed that 50%-60% of CRM projects fail. It has been found that CRM project failure is largely due to a lack of management discipline and a lack of business process definition during execution, amongst other factors. Success of CRM is not only a concentration on IT alone, but also an effective combination of skilled employees with IT solution. CRM has the greatest impact on firm performance when IT resources combine with organizational capabilities. Some report a negative relationship between IT investment & aspects of firm success, while others have demonstrated a positive relationship between IT investment & firm performance. To gain knowledge on how to turn the barriers for success into possibilities, it is important to take into account some significant success factors.

A strong management committee should be formed in favor of CRM, because if management adopts the new process effectively and participate then it motivates all the employees also. They can provide a motivation bonus to their employees or commission in against using CRM. An organization should have defined and measurable project goals, so that CRM can help in improving customer services. CRM implementation is an incremental approach, management should not think that this will create overnight results. Yes, CRM is fully technology based but still it is a business process also. So don't override the business process against technology. In making CRM project successful, an organization should maintain a right team, including experienced, technical staff, business focused and IT manager who can understand the viewpoints of both the business and the customers. Management should maintain key metrics to indicate the gaps. Implementing CRM is a big change in company, So train the employees already working in the organization. Management should not try to roll out them. Training is a good method for resistance to change [2].

To make CRM implementation a successful project, management must research in both the industry's best practices and adoption capability of their organization. So the management should strategically strong for CRM implementation. Management should have full knowledge of CRM, top management skills and commitment, cross functional communication and mandatory training for all employees [4].

Including all the above discussed success factors, we employ that motivation of employees is an important success factor for CRM implementation. Motivation has long been acknowledged as a key determinant of general behavior, acceptable behavior & work related behavior. Management should be aware of different motivational needs of the employees. Employee retention is as important as customer retention. All the purposes and goals of the organization should be properly communicated to the users. They need to understand why they need a new system of work and new processes. This will inspire the users and reduce resistance to change. The objectives need to be clearly communicated to the employees to know the expectations and to have a common guideline to work after the change. In the process of decision making, there should be an involvement of employees. After this involvement it will be easier to get people to accept the change in business operations. Participation can result in positive attention and encouragement of employees. It will lead to the success of CRM. An organization which is going to use CRM, has to invest in IT infrastructure. It includes embedded hardware and software.

V. Conclusion

In present, dynamic business environment where technology and advancement plays a pivotal role in growth, CRM finds an important place in various business systems. CRM is an old concept, but it is growing very fast nowadays, helping companies to understand the customer's perspective and matching up to their expectations. Managers are now interested in implementing it because of its benefits. This paper will help managers to know and recognize the factors of failure and critical success factors. This study will help them to make CRM success. CRM is an ongoing process which requires frequent investment in technology and training needs to bring along the customer and company on the same platform through the use of technology. Hence CRM brings together people, technology & organizational capabilities ensure connectivity between the company and its customer. It is obvious that improving satisfaction of customers, improves revenue and income of a company. Consequently, it is imperative to understand that CRM is a result of a large investment of time and resources, but it is difficult to implement without having proper research on CRM.

References Références Referencias

1. Injazz J. Chen and Karen Popovich, "Understanding customer relationship management (CRM) People, process and technology", Business Process Management Journal Vol. 9 No. 5, 2003 pp. 672-688.
2. Dr. R Jayam J Radha, "Factors Associated with Success and Failures in CRM", International Journal of Advance Research in Computer Science and Management Studies, ISSN: 2321-7782 Volume 1, Issue 6, November 2013.
3. Nastaran Mohammadhossein, Dr. Nor Hidayati Zakaria, "CRM Benefits for Customers: Literature Review (2005-2012)", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 2, Issue 6, November-December 2012, pp. 1578-1586.
4. Franka Piskar, Armand Faganel, "A Successful CRM Implementation Project in a Service Company: Case Study", Organizacija, Volume 42 Number 5, September-October 2009 DOI: 10.2478/v10051-009-0017-y.
5. Thompson B. (2002) what is CRM? www.crmguru.com (viewed 11 February, 2004).
6. Harald Salomann, Malte Dous, Lutz Kolbe, Walter Brenner, "Advancing CRM Initiatives with Knowledge Management", Journal of Information Science and The Technology, JIST 3 (2) 2006.
7. Jana Urdzikova, Martina Jakabova, Sebastian Saniuk, "The Customer Relationship Management in Terms of Business Practice in Slovakia", Research Papers Faculty of Materials Science and Technology in Trnava 2012.
8. Hsin Hsin Chang, "Critical Factors and Benefits in the Implementation of Customer Relationship Management", Total Quality Management, Vol. 18, No. 5, 483–508, July 2007.

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Enhanced Re-Ranking and Semantic Similarity Algorithm for Image Search Goals Using Click-Through Logs

By Shamali Kishor Kherdikar & Rajesh Kulkarni

University of Pune TSSM's BSCOER Narhe, India

Abstract- The objective of the proposal is to analyze the user search goals for a query which can be very useful in improving search engine relevance and user experience. Although the research on inferring user goals or intents for text search has received much attention, little has been proposed for image search with visual information. In this project, we propose a novel approach to capture user search goals in image search by exploring images which are extracted by mining single sessions in user click-through logs to reflect user information needs. Moreover, we also propose a novel evaluation criterion to determine the number of user search goals for a query. Modified re-ranking and semantic similarity algorithm are part of this proposal. Experimental results demonstrate the effectiveness of the proposed method.

Keywords : *image search, click-through log, re-ranking algorithm, semantic similarity algorithm.*

GJCST-C Classification : *I.4.1*

Strictly as per the compliance and regulations of:

Enhanced Re-Ranking and Semantic Similarity Algorithm for Image Search Goals Using Click-Through Logs

Shamali Kishor Kherdikar [α] & Rajesh Kulkarni [σ]

Abstract- The objective of the proposal is to analyze the user search goals for a query which can be very useful in improving search engine relevance and user experience. Although the research on inferring user goals or intents for text search has received much attention, little has been proposed for image search with visual information. In this project, we propose a novel approach to capture user search goals in image search by exploring images which are extracted by mining single sessions in user click-through logs to reflect user information needs. Moreover, we also propose a novel evaluation criterion to determine the number of user search goals for a query. Modified re-ranking and semantic similarity algorithm are part of this proposal. Experimental results demonstrate the effectiveness of the proposed method.

Keywords: image search, click-through log, re-ranking algorithm, semantic similarity algorithm.

I. Introduction

Search engine services are a popular means for information searching. They provide a simple and direct way of searching information for various resource types, not only textual resources, but also multimedia [1][2][3]. In image and web search applications, users submit queries (i.e., some keywords) to search engines to represent their search goals. However, in many cases, queries may not exactly represent what they want since the keywords may be polysemous or cover a broad topic and users tend to formulate short queries rather than to take the trouble of constructing long and carefully stated ones [1]–[3]. Besides, even for the same query, users may have different search goals. We find that users have different search goals for the same query due to the following three reasons considering apple as a query for search:

a) Multi-concepts

A query may represent different things. For ex, besides a kind of fruit, "apple" is established with new concepts by Apple Inc.

b) Multi-forms

Different forms can be there for a same thing. Taking Bumblebee in the film Transformers for an ex, it has two modes, car mode and humanoid mode.

Author α σ: Department of Computer Engineering, University of Pune TSSM's BSCOER Narhe, Pune, India.
e-mails: Shamli.kherdikar11@gmail.com, rkpv200@gmail.com

These two modes are the two forms of "Bumblebee".

c) Multi-representations

In image search, the same thing may be shown with different angles of view such as the query "leaf". It can be represented in a real scene or by a close-up.

Most search engines present similar interfaces allowing people to: submit a query; receive a set of results; follow a link; explore the information space; and modify a query [4][5][6]. This process is generally repe - ated during interactive searching. The popular use of search engine services has led to many investigations of general search habits on the Web. Querying behavior–query formulation and reformulation has especially been an active area of research in information retrieval. Inferring user search goals is very important in imp - roving search engine relevance and user experience. Normally, the captured user image-search goals can be utilized in many applications. For example, we can take user image- search goals as visual query suggestions [4] to help users reformulate their queries during image search. Besides, we can also categorize search results [5] for image search according to the inferred user image-search goals to make it easier for users to browse. Furthermore, we can also diversify and re-rank the results retrieved for a query [6], [7] in image search with the discovered user image-search goals. Thus, inferring user image search goals is one of the key techniques in improving users' search experience. However, although there have been many researches for text search [8]–[12], few methods were proposed to infer user search goals in image search [4], [13]. Some works try to discover user image-search goals based on textual information (e.g., external texts including the file name of the image file, the URL of the image, the title of the web page which contains that image and the surrounding texts in image search results [14] and the tags given by users [4]). However, since external texts are not always reliable not guaranteed to precisely describe the image contents) and tags are not always available (i.e., the images may not have their corresponding tags that need to be intentionally created by users), these textual information based methods still have limitations.

It should be possible to infer user image-search goals with the visual information of images (i.e., image features) since different image-search goals usually have particular visual patterns to be distinguished from each other. However, since there are semantic gaps [15] between the existing image features and the image semantics, inferring user image-search goals by visual information is still a big challenge Therefore, in this paper, we propose to introduce additional information sources to help narrowing these semantic gaps.

Digital image is nowadays the second most prevalent media in the Web only after text. Image search engines play an important role in enabling people to easily access to the desired images. A variety of search interfaces have been employed to let users submit the query in various forms, e.g., textual input, image input, and painting based input, to indicate the search goal. To facilitate image search, query formulation is required not only to be convenient and effective for users to indicate the search goal clearly, but also to be easily interpreted by image search engines. Therefore, recently more and more research attention has been paid on search interface design in developing image search engines.

II. Review of Related Works

The existing methods for image searching and re-ranking suffer from the unreliability of the assumptions under which the initial text-based image search result. However, producing such results contains a large number of images and with more number of irrelevant images.

a) TBIR – Text Based Image Retrieval

The text-based image retrieval (TBIR) can be traced back to the late 1970s. A very popular framework of image retrieval then was to first annotate the images by text and then use text-based database management systems (DBMS) to perform image retrieval. Many advances, such as data modeling, multidimensional indexing, and query evaluation, have been made along this research direction. However, there exist two major difficulties, especially when the size of image collections is large (tens or hundreds of thousands). One is the vast amount of labor required in manual image annotation. The other difficulty, which is more essential, results from the rich content in the images and the subjectivity of human perception. That is, for the same image content different people may perceive it differently. The perception subjectivity and annotation impreciseness may cause unrecoverable mismatches in later retrieval processes.

b) CBIR – Content Based Image Retrieval

The emergence of large-scale image collections, the two difficulties faced by the manual annotation approach became more and more acute. To overcome these difficulties, content-based image retrieval (CBIR) was proposed. That is, instead of being manually annotated by text-based key words, images would be indexed by their own visual content, such as color and texture. Since then, many techniques in this research direction have been developed and many image retrieval systems, both research and commercial, have been built. The advances in this research direction are mainly contributed by the computer vision community.

III. Problem Statement

Image search engines apparently provide an effortless route, but currently are limited by poor precision of the returned images and also restrictions on the total number of Images provided. While several studies reveal general characteristics of image searching based on transaction log data, little has been investigated concerning whether or not image searching behavior, especially querying behavior – query iterations and query length – differs based on a user's contextual aspects and different sources of collections on Web search engines. The existing methods for image searching and ranking suffer from the unreliability of the assumptions under which the initial text-based image search results. However, producing such results containing a large number of images gives more number of irrelevant images.

IV. System Architecture

Figure-1 describes the system model for Image Classification and Grouping based on User Query and Click through Data process. The model consists of elements as Query Handler, Query Formulation, Event Handler and Result. It has log repositories which stores user query logs and click through data. A Semantic similarity-based Matching algorithm will be implemented for classification and Grouping the search image results.

In general, when a user pose a query, the user usually navigates the entire result links list from top to bottom in a page. User generally clicks one or more result link that looks appropriate and relevance and skips those links which are not relevant. Effective information retrieval is achieved when a precise personalization approach perform re-ranking of the relevant links and place it in higher in results list. Therefore, we utilize user clicks as relevance decision measure to evaluate the searching accuracy. Since click through data can collect straightforward with less effort, it is possible to do required behavior and interest evaluation implementing this framework. Moreover, click through data shows the actual real world distribution of user search interest queries, and searching scenarios. Therefore, using click through data makes a closer real time personalization requirement cases in compare user feedback survey.

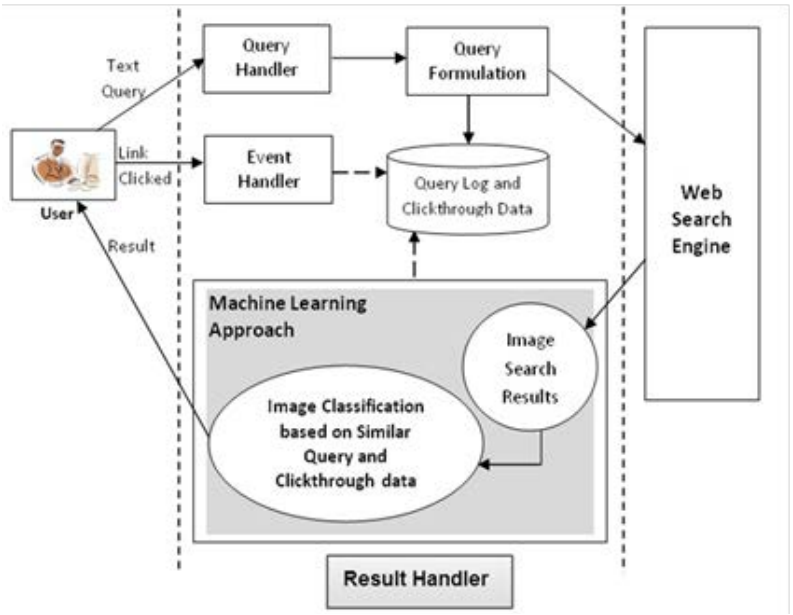


Figure 1 : System Architecture

V. Algorithm Design

A Semantic similarity-based Matching algorithm will be implemented for classification and grouping the search image results and a re-ranking algorithm is introduced to re-arrange the search results as per the click through log. Semantic similarity algorithms are data and content depended. Existing semantic similarity might tune according there data set requirements. The proposed algorithm based on existing text processing and mining techniques for effectively find the semantic similarities. We design our algorithm using Clustering Technique and Association Rule Mining. We first cluster the results based on the Density-based method from the database in relate to the user query. Density-based method implements Term frequency (TF) Similarity in the clickthrough data to decide the semantic relation of the query with the records. Secondly, based on the association rule we find the relation depth of each record in cluster in relate to the user query. These two approach combine provides the semantic similarity for our requirement. Ranking is done based on the some parameter and features of objects.

a) Semantic similarity based Algorithm

Input:
User input Query *(Q)*, and
click through data *(C)* from the database
Output: Semantically Associated and Cluster Results (S_R)
In relevance to the user queries
Begin
Create an empty cluster vector as E_C
Create keywords, *K* from Q.
For each keyword of (*K*) ***do***
Select $K(i) \rightarrow k_w$
For each click through data in $(C_j$ ***do***
Select $C(i) \rightarrow C_w$
If Compute Association $(k_w \in C_w) == true$
If Cluster E_C does not contain $k_w = true$
Add C_w to Cluster E_C
End if
End if
End for
End for
Create an empty object vector as S_R
For each objet in Cluster E_C ***do***
Object Count (O_C) ->0
Select $E_C(i) \rightarrow O_w$
For each click through data in (*C*) ***do***
Select $C(i) \rightarrow C_w$
If Compute $(O_w \in C_w) == true$
$O_C = O_C + 1$
End if
End for
Update $S_R(i) \rightarrow O_C$
End for
End

If we reprocess a ranked result for better result it is known as Re-Ranking. Support Vector Machine (SVM) is best approach for ranking any object. SVM do ranking based on multiple parameter or feature vectors in multidimensional. In our project we have only one best feature vector for ranking. So, we implement a greedy algorithm for re-ranking.

b) Re-Ranking Algorithm

Input:
Cluster Results (S_R)
Output:
Re-Ranked Results (R_{Rank})
Begin
Create an empty Object_Rank vector as, R_{Rank}
For each record in S_R ***do***
Select $S_R(i)$ -> $FR_{(i)}$
Intial_Top_Rank (T_{Rank}) -> $FR_{(i)}$
For each record in S_R ***do***
$k = i+1$
Select $S_R(k)$ -> $NR_{(k)}$
If $NR_{(k)} > FR_{(i)}$
$T_{Rank} = NR_{(k)}$
End if
End for
Update R_{Rank} -> T_{Rank}
Remove T_{Rank} object from S_R
End for
End

VI. Result Discussions

The proposed method is compared with the search result from precision and recall method.

Precision and recall are the basic measures used in evaluating search strategies. There is a set of records in the database which is relevant to the search topic. Records are assumed to be either relevant or irrelevant.

Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage. Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

Let's say an image database contains 80 records on a particular topic. An image search was conducted on that topic and 60 images were retrieved. Of the 60 records retrieved, 45 were relevant. Precision and recall scores for the search can be calculated as

- A = the number of relevant records retrieved,
- B=the number of relevant records not retrieved
- C = the number of irrelevant records retrieved.

In this example A = 45, B = 35 (80-45) and C = 15 (60-45).

Recall = (45 / (45 + 35)) * 100% => 45/80 * 100% = 56%

Precision = (45 / (45 + 15)) * 100% => 45/60 * 100% = 75%

Table 1 : Without click through Data

Query	Top Retv_ Result	No Of Assoc	Relv Result	Precision	Recall
1	10	5	2	0.2	0.133333333
2	10	7	3	0.3	0.176470588
3	10	6	3	0.3	0.1875
4	10	6	2	0.2	0.125
5	10	6	2	0.2	0.125
6	10	6	2	0.2	0.125
7	10	8	3	0.3	0.166666667
8	10	6	3	0.3	0.1875
9	10	8	3	0.3	0.166666667
10	10	7	3	0.3	0.176470588
1	10	5	2	0.2	0.133333333

Table 2 : With click through Data

Query	Top Retv_ Result	No Of Assoc	Relv Result	Precision	Recall
1	10	7	5	0.5	0.294117647
2	10	7	6	0.6	0.352941176
3	10	8	6	0.6	0.333333333
4	10	8	6	0.6	0.333333333
5	10	7	5	0.5	0.294117647
6	10	8	7	0.7	0.388888889
7	10	9	6	0.6	0.315789474

8	10	8	6	0.6	0.333333333
9	10	8	6	0.6	0.333333333
10	10	9	8	0.8	0.421052632
1	10	7	5	0.5	0.294117647

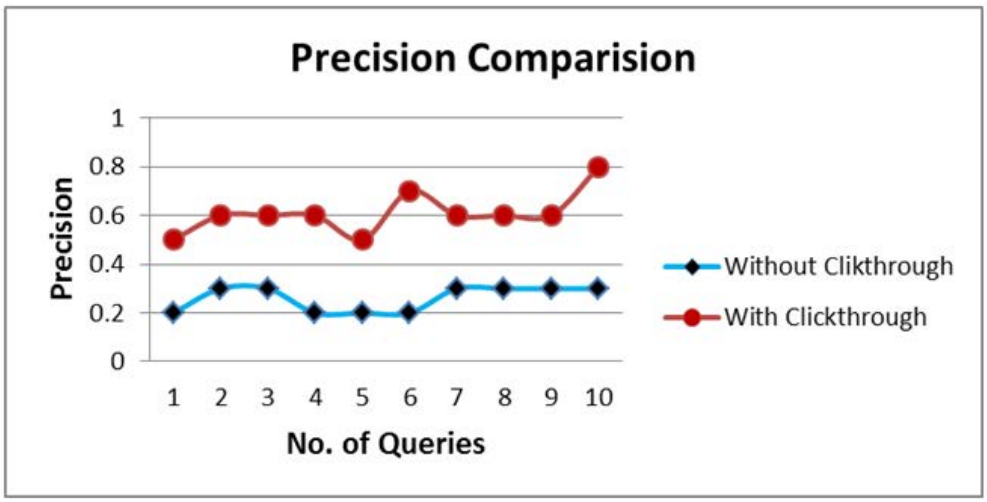


Figure 2 : Precision Comparison

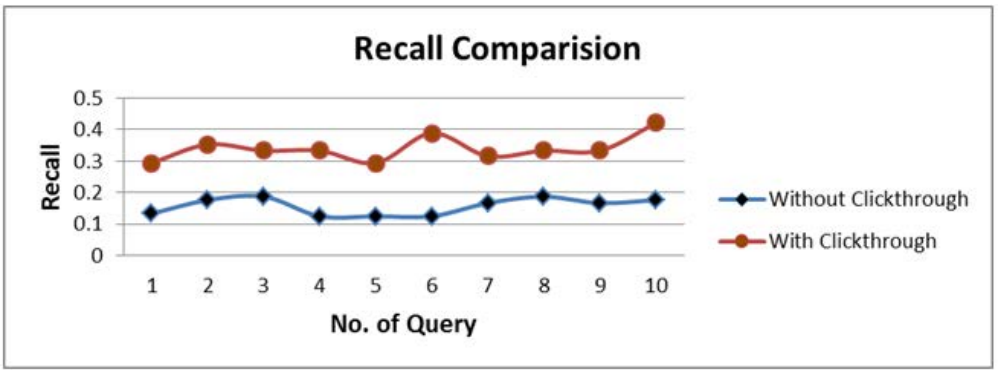


Figure 3 : Recall Comparisons

Fig2 shows the image searching method comparison with click through data and without click through data against precision. Certainly the efficiency and relevancy of image search results with click through data looks is more correct compared to image searching approach without click through data whereas Fig3 shows the image searching method uses click through data and without click through data against Recall.

VII. Conclusion

Image classification is vital field of research in computer vision. Increasing rate of multimedia data, remote sensing and web photo gallery need a category of different image for the proper retrieval of user. Existing commercial image search engines provide a textbox for users to type one or more keywords to indicate the search goal. This type of search interface is easy to use. However, besides the limitation that the associated texts may not reveal the image content, it is not easy to perform image search. We propose a machine learning approach for automatically classification and grouping similar user query for image search and to analyze the user search goal using user query and its relevant click through data in different browsing session. We classify the images semantically using density-based method on click through database related to user query and re-rank the search result based on the image classification and query similarity. We perform a query evaluation to determine the number of user search goals for a query and evaluate the performances of different user search goal. The experiments show that our method is able to infer user image-search goals effectively. It shows that inferring user image-search goals using user click information is better to meet the user image-search.

References Références Referencias

1. C. Zhang, J. Y. Chai, and R. Jin, "User term feedback in interactive text-based image retrieval,"

in Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2005, pp. 51–58.
2. F. Mahmoudi, J. Shanbehzadeh A. Eftekhari-Moghadam and H. Soltanian- Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," IEEE Trans. on Pattern Recognition, vol. 36, pp. 1725-1736, 2003.
3. G. P. Nguyen and M. Worring. Optimization of interactive visual-similarity-based search. TOM - CCAP, 4(1), 2008.
4. J. Cui, F. Wen, and X. Tang. Intentsearch: interactive on-line image search re-ranking. In MM '08, pages 997–998, 2008.
5. L. Chen, D. Xu, I. W. Tsang, and J. Luo, "Tag-based web photo retrieval improved by batch mode re-tagging," in Proc. IEEE Comput. Soc. Conf. Com - put. Vis. Pattern Recogn., 2010, pp. 3440–3446.
6. R. Yan, A. Natsev, and M. Campbell. Multi-query interactive image and video retrieval -: theory and practice. In CIVR '08, pages 475–484, 2008.
7. U. Lee, Z. Liu and J. Cho, "Automatic identification of user goals in web search," WWW, pp. 391-400, 2005.
8. W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking via information bottleneck principle," in Proc. 14th ACM Int. Conf. Multimedia, 2006, pp. 35–44.
9. X. Li, Y-Y. Wang and A. Acero, "Learning query intent from regularized click graphs," SIGIR, pp. 339-346, 2008.
10. Xiao gang Wang, Ken Liu and Xiao Tang „Query-Specific Visual Semantic Spaces for Web Image Re-ranking.. In Proceeding of the 14th ACM Internat - ional Conference on Multimedia,(2011) .
11. X. Wang and C-X. Zhai, "Learn from web search logs to organize search results," SIGIR, pp. 87-94, 2007.
12. Y. Luo, W. Liu, J. Liu, and X. Tang. Mqsearch: image search by multi-class query. In CHI '08, pages 49–52, 2008.
13. Y. Liu, D. Xu, I. W. Tsang, and J. Luo, "Textual query of personal photos facilitated by large-scale web data," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 1022–1036, May 2011.
14. Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in Proc. IEEE Int. Conf. Image Process., 1997, pp. 815–818.
15. Z.-H. Zhou and H.-B. Dai, "Exploiting image contents in web search," in Proc. 20th Int. Joint Conf. Artif. Intell. , 2007, pp. 2928–2933.
16. Z-J. Zha, L-J. Yang, Z-F. Wang, T-S. Chua and X-S. Hua, "Visual query suggestion: towards capturing user intent in internet image search," ACM Trans. On Multimedia Comput. Commu. Appl. 6, 3, Article, August 2010.
17. Z.-J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang. Visual query suggestion. In MM '09, pages 15–24, 2009.

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Strategy Design Pattern

By Mrs. Renu Bala & Mr. Kapil Kumar Kaswan

CDLU SIRSA, India

Abstract- Design patterns usually describe abstract systems of interaction between classes, objects, and communication flows. So, a description of a set of interacting classes that provide a generalized solution framework to a generalized/specific design problem in a specific context can be said as a design pattern. There are many design patterns that can be used to solve real-life problems, but it remains very difficult to design, implement and reuse software for complex applications. Examples of these include enterprise system, real-time market data monitoring and analysis system. Design patterns provide an efficient way to create more flexible, elegant and ultimately reusable object-oriented software. Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice". The solutions of the given problems are expressed in terms of objects and interfaces. Among 23 design patterns, Strategy pattern defines an interface common to all supported algorithms. Context uses this interface to call the algorithm defined by a Concrete Strategy. In accounting framework one thing is mostly needed that is tax calculation. To solve this problem author in the current study has chosen the strategy pattern.

Keywords: *design pattern, context, strategy, object, concretestrategy.*

GJCST-C Classification : *I.5*

STRATEGYDESIGNPATTERN

Strictly as per the compliance and regulations of:

Strategy Design Pattern

Mrs. Renu Bala [α] & Mr. Kapil Kumar Kaswan [σ]

Abstract- Design patterns usually describe abstract systems of interaction between classes, objects, and communication flows. So, a description of a set of interacting classes that provide a generalized solution framework to a generalized/specific design problem in a specific context can be said as a design pattern. There are many design patterns that can be used to solve real-life problems, but it remains very difficult to design, implement and reuse software for complex applications. Examples of these include enterprise system, real-time market data monitoring and analysis system. Design patterns provide an efficient way to create more flexible, elegant and ultimately reusable object-oriented software. Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice". The solutions of the given problems are expressed in terms of objects and interfaces. Among 23 design patterns, Strategy pattern defines an interface common to all supported algorithms. Context uses this interface to call the algorithm defined by a Concrete Strategy. In accounting framework one thing is mostly needed that is tax calculation. To solve this problem author in the current study has chosen the strategy pattern.

Keywords: design pattern, context, strategy, object, concretestrategy.

I. Introduction

A design pattern is a generic solution that has been observed in multiple instances to help resolve a particular problem within a known context. Design patterns provide an efficient way to create more flexible, elegant and ultimately reusable object-oriented software. Design methods are supposed to promote good design, to teach new designers how to design well and to standardize the way designs are developed. Typically, a design method comprises a set of syntactic notations usually graphical and a set of rules that govern how and when to use each notation. It will also describe problems that occur in a design, how to fix them, and how to evaluate a design. Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice [1]. The solutions of the given problems are expressed in terms of objects and interfaces. Design patterns are being increasingly used in software design. Design patterns are a good means for recording design experience as they systematically name, explain and evaluate important and recurrent designs in software systems. They describe problems that occur repeatedly, and describe the core of the solution to that problem, in such a way that this solution can be used many times in different contexts and applications. A good design is a good solution regardless of the technology. And no matter how good the technology may be, it is only as good as its design, and specifically the implementation of that design. In fact, a great design with older technology may still be good, but a bad design with new technology is usually just bad. A design pattern is a form of design information and the design that worked well in past will be used in future in any application similar to existing application which uses these designs. These design information can help both the experienced and the novice designer to recognize situations in which these designs can be reused. There are three categories of design patterns: Creational, structural and Behavioral.

Author α σ: Department of Computer Science and Application Chaudhary Devi Lal University Sirsa, Haryana, India.
e-mails: renubala.gghsodhan@gmail.com,

II. Net Framework

A .net is a new software platform for the desktop and the Web. The .NET Framework is an integral Windows component that supports building and running the next generation of applications. The .NET Framework has two main components: the common language runtime and the .NET Framework class library. The common language runtime is the foundation of the .NET Framework [2]. The .NET Framework is designed to fulfill the following objectives:

- To provide a consistent object-oriented program - ming environment whether object code is stored and executed locally, executed locally but Internet-distributed, or executed remotely.
- To provide a code-execution environment that minimizes software deployment and versioning conflicts.
- To provide a code-execution environment that promotes safe execution of code, including code created by an unknown or semi-trusted third party.
- To provide a code-execution environment that eliminates the performance problems of scripted or interpreted environments.
- To make the developer experience consistent across widely varying types of applications, such as Windows-based applications and Web-based applications.

- To build all communication on industry standards to ensure that code based on the .NET Framework can integrate with any other code [2].

III. Strategy Pattern

Strategy - defines an interface common to all supported algorithms. Context uses this interface to call the algor - ithm defined by a ConcreteStrategy.

ConcreteStrategy - each concrete strategy implements an algorithm.

Context

- contains a reference to a strategy object.
- may define an interface that lets strategy accessing its data.

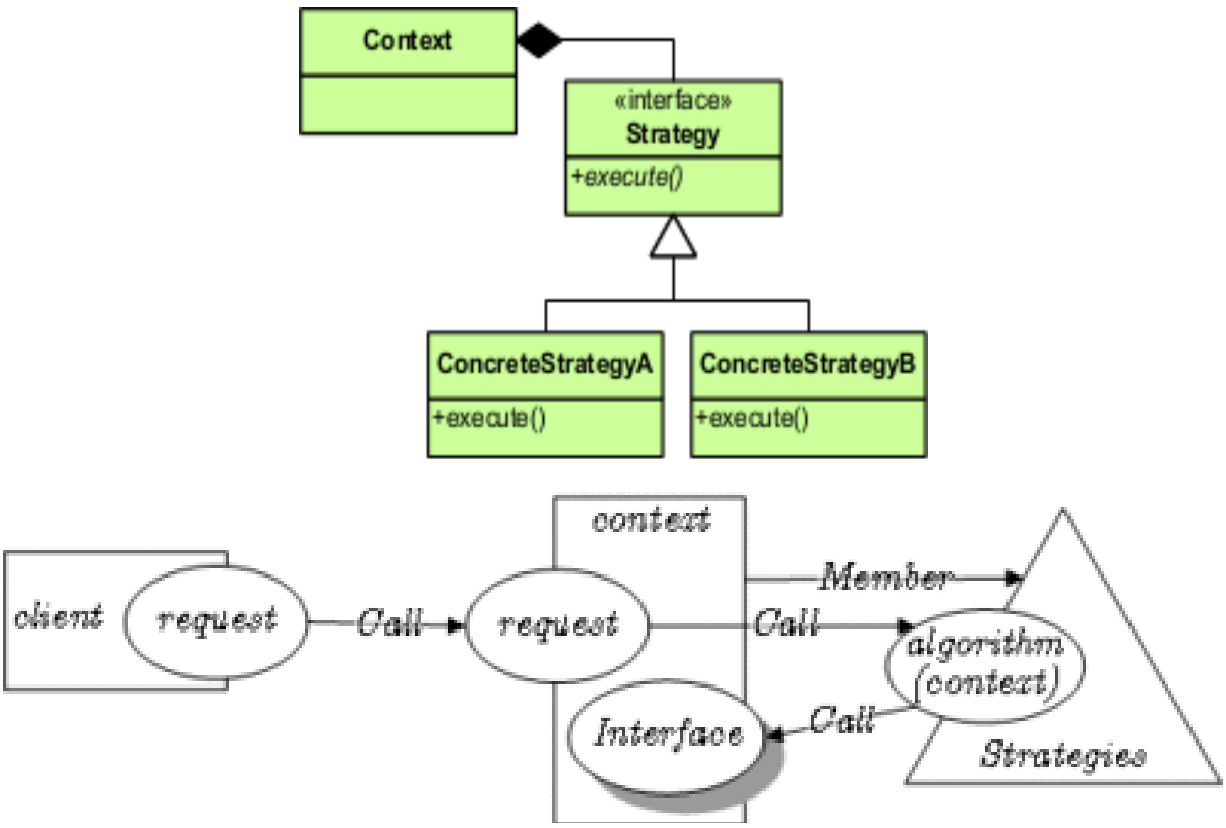
The Context objects contains a reference to the ConcreteStrategy that should be used. When an operation is required then the algorithm is run from the strategy object. The Context is not aware of the strategy implementation. If necessary, addition objects can be defined to pass data from context object to strategy.

The context object receives requests from the client and delegates them to the strategy object. Usually the ConcreteStrategy is created by the client and passed to the context. From this point the clients interact only with the context. In other words, it defines a family of algorithms, encapsulate each one and make them interchangeable. In computer programming, the strategy pattern also known as the policy pattern is a software design pattern that enables an algorithm's behavior to be selected at runtime. The strategy pattern

- defines a family of algorithms,
- encapsulates each algorithm, and
- makes the algorithms interchangeable within that family.

Strategy lets the algorithm vary independently from clients that use it. Strategy is one of the patterns included in the influential book Design Patterns by Gamma et al. that popularized the concept of using patterns in software design. For instance, a class that performs validation on incoming data may use a strategy pattern to select a validation algorithm based on the type of data, the source of the data, user choice, or other discriminating factors. These factors are not known for each case until run-time, and may require radically different validation to be performed. The validation strategies, encapsulated separately from the validating object, may be used by other validating objects in different areas of the system (or even different systems) without code duplication.The essential requirement in the programming language is the ability to store a reference to some code in a data structure and retrieve it. This can be achieved by mechanisms such as the native function pointer, the first-class function, classes or class instances in object-oriented programming languages, or accessing the language implementation's internal storage of code via reflection.

Structure



IV. Related Work

There are various design patterns that can be used to solve any of the industrial application. Here in this paper work, strategy pattern is used to build a framework. In accounting framework, one thing is mostly needed that is tax calculation. To solve this problem author in the current study has chosen the strategy pattern. Using these patterns, design solution of the industrial problem will be described. The father of the pattern concept, proposed a description template stating nine essential pattern elements. These patterns element describes the design patterns effectively; also describe how these patterns are useful to solve the problem. Industrial applications typically require different kinds of interfaces to the data they store and the logic they implement are data loaders, user interface and integration gateways and others. Instead of using for different purpose, these interfaces often need common interactions with the application to access and manipulate its data and invoke its business logic. These interactions may be complex, involving transaction across multiple resources and the coordination of several responses to an action. These interfaces decide the interaction between different layers of the application; how user interacts with middleware layer and the database layer. The framework is implemented in .Net. As we are using the design patterns to build this framework hence the developer can use this framework to build any kind of industrial application and can implement it in any other programming language using object-oriented concepts. Using the concept of design patterns. There are various classes with their methods and properties [5].

V. Analyze Strategy Pattern by Example

Strategy pattern is used when we have multiple algorithm for a specific task and client decides the actual implementation to be used at runtime.

Strategy pattern is also known as Policy Pattern. We defines multiple algorithms and let client application pass the algorithm to be used as a parameter. One of the best example of this pattern is Collections.sort() method that takes Comparator parameter. Based on the different implementations of Comparator interfaces, the Objects are getting sorted in different ways [8].

VI. Simulation Strategy Design Pattern

One common usage of the strategy pattern is to define custom sorting strategies e.g. to sort a list of strings by length in Java, passing an anonymous inner class (an implementation of the strategy interface) [7]:

```
List<String> names = Arrays.asList("Anne", "Joe",
"Harry");
Collections.sort(names, new Comparator<String>() {
public int compare(String o1, String o2) {
return o1.length() - o2.length();
}
});
Assert.assertEquals(Arrays.asList("Joe", "Anne",
"Harry"), names);
```

VII. Conclusion

Although the belief of utilizing design patterns to create better quality software is fairly widespread, there is relatively little research objectively indicating that their usage is indeed beneficial. In this paper we try to reveal the connection between design patterns and software maintainability. It is very hard to understand better what the patterns are and how they relate to each other. At this point there is a fundamental picture as reacting to an event to produce accounting entries. We used our probabilistic quality model for estimating the maintainability. We found that every introduced pattern instance caused an improvement in the different quality attributes. Moreover, the average design pattern line density showed a very high, 0.89 Pearson correlations with the estimated maintainability values. Design patterns are outstanding communication tool and help to make the design process faster. This allows solution providers to take the time to concentrate on the business implementation. Patterns help the design to make it reusable. Reusability not just applies to the component, but also the stages of the design that must go from a problem to final solution. The ability to apply a pattern that provides a repeatable solution is worth the little time spent learning formal patterns. There are some promising results showing that applying design patterns improve the different quality attributes according to our maintainability model. In addition, the ratio of the source code lines taking part in some design patterns in the system has a very high correlation with the maintainability. However, these results are only a small step towards the empirical analysis of design patterns and software quality [4].Design patterns shall support reuse of a software architecture in different application domains as well as reuse of flexible components [6].

References Références Referencias

1. http://blogs.infragistics.com/blogs/ux/archive/2009/02/03/what-is-a-design -pattern-and-why-use-them-for-quince.aspx.
2. Bertrand Meyer, Karine Arnout, Componentization: The Visitor Example, to appear in Computer (IEEE), 2006.
3. http://www.oodesign.com/abstract-factory-pattern.html
4. P´eter Heged˝us, D´enes B´an, Rudolf Ferenc, and Tibor Gyim´othy University of Szeged, Department of Software Engineering ´Arp´ad t´er 2. H-6720 Szeged,Hungary {hpeter,zealot,ferenc,gyimothy}-@inf.u-szeged.hu

5. Meyer, Bertrand "Componentization: The Visitor Example". *IEEE computer* (IEEE) 39 (7): 23–30.
6. Jurgen Dorn and Tabbasum Naz,Institute of Information Sysytems 184/2 Technicla University Vienna ,Favoritenstrabe 9-11, Vienna A-1040, Austria {dorn/naz}@dbai.tuwien.ac.at
7. http://stackoverflow.com/questions/370258/realworld-example-of-the-strategy-pattern
8. http://www.journaldev.com/1827/java-design-patterns-example-tutorial#strategy-pattern.

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Extended Edgecluster Based Technique for Social Networking Collective Behavior Learning System

By Umesh B. Shingote & Dr. Setu Kumar Chaturvedi

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India

Abstract- Growing interest and continuous development of social network sites like Facebook, Twitter, Flicker, and YouTube etc.turn to several researchers for research study, planning and rigorous development. Exact people behavior prediction is the most important challenge of these on-line social networking websites. This research focus to learn to predict collective behavior in social media networks. Particularly provided information about some person, how can we collect the behavior of unobserved persons in the same network? These tremendous growing networks in social media are of massive size, involving large number of actors. The computational scale of these networks makes necessary scalable learning for models for collective collaborative behavior prediction. This scalability issue is solved by the proposed k-means clustering algorithm which is used to partition the edges into disjoint distinct sets, with each set is showing one separate affiliation. This edge-centric structure represents that the extracted social dimensions are definitely sparse in nature. This model idealized on the sparse natured social dimensions, shows efficient prediction performance than earlier existing approaches The proposed approach can effectively able to work for sparse social networks of any growing size.

Keywords : *edge clustering, scalability, social dimension, social network.*

GJCST-C Classification : *K.4.2*

Strictly as per the compliance and regulations of:

Extended Edgecluster Based Technique for Social Networking Collective Behavior Learning System

Umesh B. Shingote[α] & Dr. Setu Kumar Chaturvedi[σ]

Abstract- Growing interest and continuous development of social network sites like Facebook, Twitter, Flicker, and YouTube etc.turn to several researchers for research study, planning and rigorous development. Exact people behavior prediction is the most important challenge of these on-line social networking websites. This research focus to learn to predict collective behavior in social media networks. Particularly provided information about some person, how can we collect the behavior of unobserved persons in the same network? These tremendous growing networks in social media are of massive size, involving large number of actors. The computational scale of these networks makes necessary scalable learning for models for collective collaborative behavior prediction. This scalability issue is solved by the proposed k-means clustering algorithm which is used to partition the edges into disjoint distinct sets, with each set is showing one separate affiliation. This edge-centric structure represents that the extracted social dimensions are definitely sparse in nature. This model idealized on the sparse natured social dimensions, shows efficient prediction performance than earlier existing approaches The proposed approach can effectively able to work for sparse social networks of any growing size. The important advantage of this method is that it easily grows upon to handle networks with large number of actors while existing methods was unable to do. This scalable approach effectively used over of online network collective behavior on a large scale.

Keywords: edge clustering, scalability, social dimension, social network.

I. Introduction

Social networking sites like Facebook Orkut, Blog Catalog, Twitter, etc are making powerful build up of social networks or different connections among persons who are interested to make friends, share valuable interests and innovative in activities. By using social networking sites, people can interact with each other, sharing and discussing information using many types of media such as their photos, videos, and always doing several of activities provided by these social networking sites. Social media networks having many forms of, including kind of blogs, user forums, tags bookmarks social networking, review process, online content sharing, etc. people can lead to use every, different activity for their needs. Social media becoming a very effective and crucial part of our living. These networks are now a useful for maintaining relationships of each priority. Social media turned new way for communication and interaction, sharing cooperation, thinking with other people. In social media networks, the connections and relationships of the same network are heterogeneous i.e. people are belonging to different groups. These relations are mixed with number of connections. e. g one user can connect to his family or friend [1].

These different connections have some limitation of effectiveness. A model is proposed to get solution on this issue of heterogeneity. This framework put the idea of social dimensions for extraction on the view of network information based on connectivity. The advantages of this model over other methods are studied efficiently on some of the social network data. Social media produces large amount of data which gives many opportunities, challenges to acquire collective behavior learning on a large area. In this research, predict collective behavior in social media network is the first goal. By observing behavior information about some person, how can get the behavior of unobserved person which is in the same network

II. Material and Methodology

a) Bi -connected Components

In previous researches one approach to find edge partitions is bi-connected components (BiComp - onents) is presented [2]. Bi-connected components of a graph are the sub types or parts of vertices and if this point is removed, its component cannot be disconnected. In a bi-connected component of a any graph two nodes are connected sub graph and connected by two paths at least. It is nothing but cut vertices in a connected graph, after removing must result in an increase number of connected components.bi-connected components are connected by various cut vertices. Each bi-connected component is considered as community, and converted into one social dimension for this process [1].

Algorithm to find Biconnected component - Firstly divide a graph into its different disconnected components. It performs a depth-first search (DFS) on

Author α σ: Department of Computer Science and Engineering, Technocrats Institute of Technology, Bhopal.
e-mails: umeshshingote62@gmail.com, dr.setuchaturvedi@gmail.com

the edges of the graph for this. After reaching new point, it is inserted on a stack and for each and every point a record is updated of the lowest point, where it is connected by a path of points which are not in a stack. When a new point not held from the top of the stack, the top point is get removed. When the stack is full, a search of a connected component is performed [16].

i. *Advantages*

- BiComponents separates edges into disjoint sets hence sparse social dimension is obtained.
- BiComponents is very effective and scalable in network.

ii. *Disadvantage*

- BiComponents gives results of highly imbalanced community structure.
- It is unable to extract informative social dimension for large classification.
- This technique gives very poor performance.
- Updating is becomes difficult whenever the network changes for new connection.

b) Node Clustering

In NodeCluster method, Social dimensions allow one actor to play role in multiple groups which are called as affiliations. The case when each actor is in only one group affiliation is verified. A social dimension on the basis of suitable node partition of network is constructed. A similar idea is given in a latent group model [5] for efficient inference. Previous researches gives idea of k-means clustering used to partition nodes (actors) of a network into different disjoint sets result into a consistent set of social dimensions. Support Vector Machine is used for discriminative learning.

Advantage

- Support Vector Machine is used for discriminative learning process.

Disadvantage

- Each actor is performing role in only one group affiliation this result in poor performance than EdgeCluster technique.

c) Edge Clustering

As network research grows, scalable approach is developed for large-scale networks without an excessive memory requirement; this method is called edge clustering. EdgeCluster, An edge-centric cluste - ring method to obtain sparse social dimensions is used effectively [6].Using this method, it can update the of social dimension needs to be extracted. K-means

social dimensions efficiently when new nodes or new edges arrive. In a large network in which large number algorithm is used to divide edges different sets. These sets are used for the information extraction. Edge clustering method preferred because there is no overlapping of edges, which the main disadvantage of node clustering method as node overlaps with each other.

It is important to develop scalable method that can handle large-scale networks efficiently without large memory requirements. Next, an extended edge centric clustering scheme to obtain sparse social dimensions is explained. With such a method, we can also update the social dimensions efficiently when new nodes or new edges arrive. In a huge network, large number of social dimension needs to be extracted.

III. Proposed System

Extending the existing approach of learning of collective behavior by presenting the approach of heterogeneity is useful for further research. Proposed framework consists of two parts: 1.Accurate social dimension extraction 2.Learning Process from extracted dimension.

In previous researches, the k-means clustering algorithm is utilized to partition the edges of a network into sets which are disjoint. Proposed efficient k-means variant is useful to take advantage of Sparsity problem, this algorithm is able to handle the clustering of large number of edges very efficiently. A model based on social dimensions is useful to be effective in this heterogeneity issue. The previous approaches, however, is not scalable to handle networks of large sizes because the extracted social dimensions are closely populated, dense. Social media network contains huge number of actors. With these huge numbers of actors, extracted dense social dimensions having problem in residing in memory, resulting in a serious computing problem and high challenges. Sparsifying social dimensions can be efficient to solve scalability problem. In this work, an extended edge-centric approach to extract sparse social dimensions is proposed. Using this proposed approach, Sparsity issue of social dimensions is achieved successfully.

Architecture:

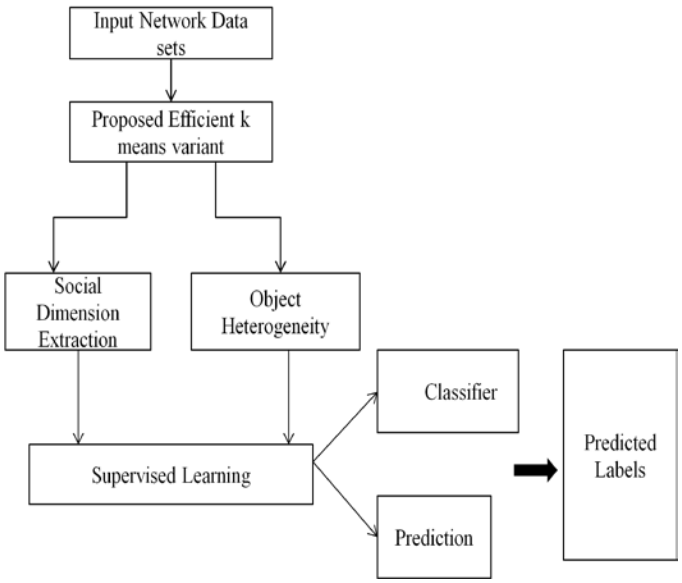


Figure 1 : Architecture of Scalable Learning of Collective Behavior [1]

Figure 1 show proposed system architecture, Firstly data sets of different social networking sites are entered as input. Various extraction techniques are used to achieve supervised learning. After social dimension extraction, discriminative and prediction occurs. The outcome of this, gives predicted labels in social networks.

In proposed work, first analyzed the results of extended edge-centric based method for the extraction of social dimensions .Large social network datasets are used for this [1]. As per the problem stated, Existing edge-centric clustering approach is extended to change the object heterogeneity. Proposed approach improves the prediction performance for social networks (multimode networks) [19].

a) Efficient K means variant algorithm

This algorithm explains design of the model, the input datasets and implementation of efficient k-means variant for social dimension extraction based on idea of extended edge centric approach. As a simple k-means is used to extract social dimensions, it is very simple and fast to update social dimensions if a given network modifies.

Proposed efficient K-means variant algorithm:

Step 1: Randomly select k centers in problem space.

Step 2: Make Suitable Partition of the data into k clusters by grouping similar data that are closest to the k centers.

Step 3: Use the mean of these clusters to find the new center.

Step 4: Update the centroid after arrival of new connection request

Step 5: Repeat steps 2 an3 until centers do not change.

Figure 2 : Algorithm efficient k-means variant

Input for the algorithm: Social Network Dataset like flicker, YouTube.

Output of the algorithm: desired social dimensions of this dataset of social media. For example, following figures showing Input and Output of this module:

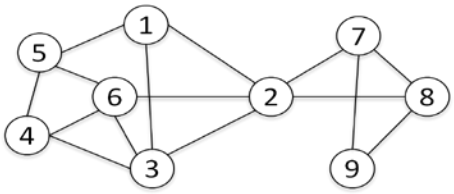


Figure 3 : Connected Nodes in a network

The Edge-centric structure of the network data is given as below:

Table 1 : Edge View Structure of Network

	Node Features								
Number of Edge	1	2	3	4	5	6	7	8	9
e(1,2)	1	1	0	0	0	0	0	0	0
e(1,3)	1	0	1	0	0	0	0	0	0
e(1,5)	1	0	0	0	1	0	0	0	0
……	.	.	.	.	.	.	.	.	.

Partition this information (edges) into the different sets as per given in below figure 4:

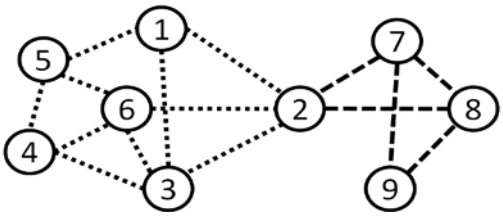


Figure 4 : Cluster formation

On the basis of edge clustering scenario, social dimensions can be obtained as per given in below table II and this is the final output of this algorithm:

Table 2 : desired social dimension(s)

Actor/Individual	Output of edge partition	
1	1	0
2	1	1
3	1	0
4	1	0
5	1	0
6	1	0
7	0	1
8	0	1
9	0	1

b) Effective Learning and Prediction

By taking input as this social dimensions as features to next algorithm, learning and prediction carried out. This algorithm is idea on linear SVM [20]. The discriminative learning procedure will find out related social dimension with the behavior and then gives proper label. One observation shows that actors of the same interests eager to connect with each other [12]. For instance, it is reasonable to expect people of the same department to interact with each other more frequently. Hence, to observe actors' latent affiliations; this research aim is to find out a group of people who interact with each other frequently than their random behavior.

Algorithm for Collective Behavior in network

Input: Datasets, labels of some people, social dimensions;

Output: labels of unlabeled people or nodes (actors).

1. First obtain edge-centric view of desired network.
2. Perform proposed extended edge clustering.
3. Determine social dimensions on the basis edge partition.
4. Apply regularization technique to social dimensions.
5. Form suitable classifier based on social dimensions of labeled actors.
6. Use this classifier to predict information about labels of unlabeled ones on their social dimension.

Data sets, social dimension and label information is provided as an input and predicted labels are the outcomes of this algorithm. Utilization of efficient k means variant algorithm to use extended edge clustering is works effectively. The regularization and SVM are applied after the formation of social dimension. The regularization parameter is used the regularize communities in a network. Finally using classifier prediction of labels is achieved.

IV. Comparative Analysis and Results

a) Prediction Performance

Micro-F1 and Macro-F1 measures for given social media network data in 10 runs are performed. Precision and recall is obtained from evaluation of different data. Table III shows the Performance on Blog Catalog network; Table IV shows Performance on Flickr network, Table V shows Performance on YouTube network.

Comparative analysis of all three table shows that proposed extended edge cluster is having best evaluation results with proportion of labeled nodes.

By observing table III, IV and V, it is note that the prediction performance on the social media data is average for F1 measure. The reason is the large number of distinctive labels in the data. Other reason is that only the network information is showed here.Extended Edge-centric clustering shows comparable performance to Edge Cluster, Node cluster and Bi-Components on Blog Catalog, Flickr and YouTube network. From the results our proposed method is the winner most of the time. Clearly, using these sparse social dimensions, it is easy to obtain best performance as dense social dimensions.The Node Cluster scheme in which each actor to be involved in only one group, showing poor performance compared with Edge Cluster. In table VI, VII and VIII, Edge Cluster-x denotes edge centric clustering for the construction of x dimensions. Time is denoted by the total time (seconds) to obtain the social dimensions; Space represents the memory required for the social dimensions; Density is the non-zeros entries in the dimensions; Upper bound is the density upper bound computation. Max-Aff and Ave-Aff denote the maximum and average number of affiliations one user in network. The computation time, the memory usage of social dimensions, density and other related statistics on all three data sets are carried out. The computation time of Extended Edge Cluster does not change much with clusters. The computation time of Extended Edge Cluster is of the same order; it does not depend on number of clusters. This is due to effectiveness of the proposed efficient k-means variant as for the memory utilization, sparse social dimension does superiority over dense. When the number of clusters k is small, the upper bound of the density not closely separated. As k increases, the bound is getting close.

Table 3 : Performance on Blog Catalog Network

Proportion of Labeled Nodes		10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1 (%)	Extended EdgeCluster	37.43	40.91	37.15	41.97	41.20	41.24	41.75	42.55	43.54
	EdgeCluster	27.94	30.76	31.85	32.99	34.12	35.00	34.63	35.99	36.29
	BiComponents	16.54	16.59	16.67	16.83	17.21	17.26	17.04	17.76	17.61
	NodeCluster	18.29	19.14	20.01	19.80	20.81	20.86	20.53	20.74	20.78
Macro-F1 (%)	Extended EdgeCluster	18.51	22.31	21.27	23.52	26.27	28.70	30.86	32.72	34.41
	EdgeCluster	16.16	19.16	20.48	22.00	23.00	23.64	23.82	24.61	24.92
	BiComponents	2.77	2.80	2.82	3.01	3.13	3.29	3.25	3.16	3.37
	NodeCluster	7.38	7.02	7.27	6.85	7.57	7.27	6.88	7.04	6.83

Table 4 : performance on flickr network

Proportion of Labeled Nodes		10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1 (%)	Extended EdgeCluster	37.43	30.78	29.83	30.21	31.09	32.24	33.55	34.96	36.42
	EdgeCluster	25.75	28.53	29.14	30.31	30.85	31.53	31.75	31.76	32.84
	BiComponents	16.45	16.46	16.45	16.49	16.49	16.49	16.48	16.55	16.55
	NodeCluster	22.94	24.09	25.42	26.53	28.18	28.32	28.58	28.70	28.93
Macro-F1 (%)	Extended EdgeCluster	20.00	22.31	21.27	23.52	26.27	28.70	30.86	32.72	34.41
	EdgeCluster	10.52	14.10	15.91	16.72	18.01	18.54	19.54	20.18	20.78
	BiComponents	0.45	0.46	0.46	0.46	0.46	0.46	0.46	0.47	0.47
	NodeCluster	7.90	9.99	11.42	11.10	12.33	12.29	12.58	13.26	12.79

Table 5 : performance on youtube network

Proportion of Labeled Nodes		10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1 (%)	Extended EdgeCluster	36.99	40.00	44.36	42.00	46.00	46.20	45.81	46.31	47.07
	EdgeCluster	23.90	31.68	35.53	36.76	37.81	38.63	38.94	39.46	39.92
	BiComponents	23.09	24.51	24.80	25.39	25.20	25.42	25.24	24.44	25.62
	NodeCluster	20.83	24.57	26.91	28.65	29.56	30.72	31.15	31.85	31.29
Macro-F1 (%)	Extended EdgeCluster	20.54	26.31	29.27	30.52	32.27	35.70	36.86	36.72	38.41
	EdgeCluster	19.48	25.01	28.15	29.17	29.82	30.65	30.75	31.23	31.45
	BiComponents	6.80	7.05	7.19	7.44	7.48	7.58	7.61	7.63	7.76
	NodeCluster	17.91	21.11	22.38	23.91	24.47	25.26	25.50	26.02	26.44

b) *Sparsity Comparison*

Table 6 : sparsity comparison on blog catalog network

Methods	Time	Space	Density	Upper Bound	Max-Aff	Ave-Aff
Ext.EdgeCluster-108	0.39	12	0.04314	0.04796	4	0.3259
Ext.EdgeCluster-162	0.35	12.1	0.04222	0.04814	5	0.3555
Ext.EdgeCluster-216	0.37	12.3	0.07277	0.05703	5	0.5833
Ext.EdgeCluster-270	0.32	12.6	0.09203	0.06203	5	0.7000
Ext.EdgeCluster-324	0.37	12.7	0.09166	0.06212	5	0.7000

Table 7 : sparsity comparison on flickr network

Methods	Time	Space	Density	Upper Bound	Max-Aff	Ave-Aff
Ext.EdgeCluster-108	0.328	12.20	0.04314	0.04800	4	0.3240
Ext.EdgeCluster-162	0.391	12.40	0.04259	0.04816	5	0.3556
Ext.EdgeCluster-216	0.360	12.70	0.07277	0.05699	5	0.5851
Ext.EdgeCluster-270	0.453	12.44	0.09203	0.06200	5	0.6981
Ext.EdgeCluster-324	0.516	12.67	0.09300	0.06214	5	0.7000

Table 8 : sparsity comparison on youtube network

Methods	Time	Space	Density	Upper Bound	Max-Aff	Ave-Aff
Ext.EdgeCluster-108	0.360	13	0.04259	0.04259	4	0.3243
Ext.EdgeCluster-162	0.454	13.45	0.04277	0.04796	5	0.3333
Ext. EdgeCluster-216	0.406	13.56	0.04185	0.04800	5	0.3314
Ext.EdgeCluster-270	0.437	13.67	0.04111	0.04797	5	0.3351
Ext.EdgeCluster-324	0.453	13.98	0.04285	0.07203	9	0.9333

c) *Performance Graph of Blog Catalog, Flickr and YouTube*

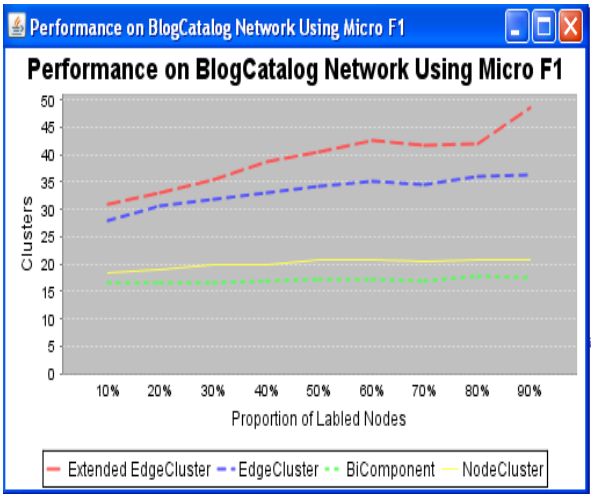


Figure 5 : Blog Catalog network (micro F1)

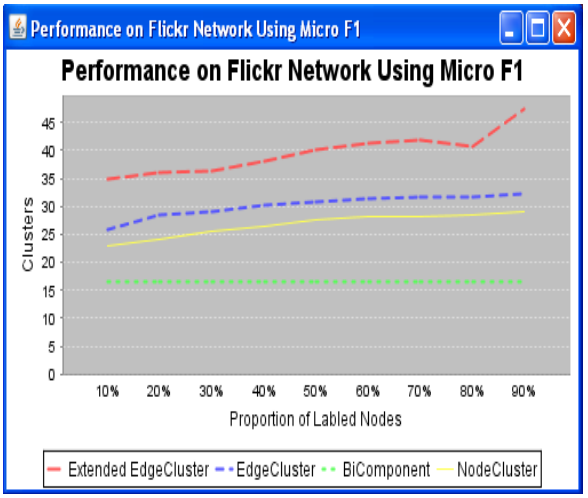


Figure 7 : Flickr network (micro F1)

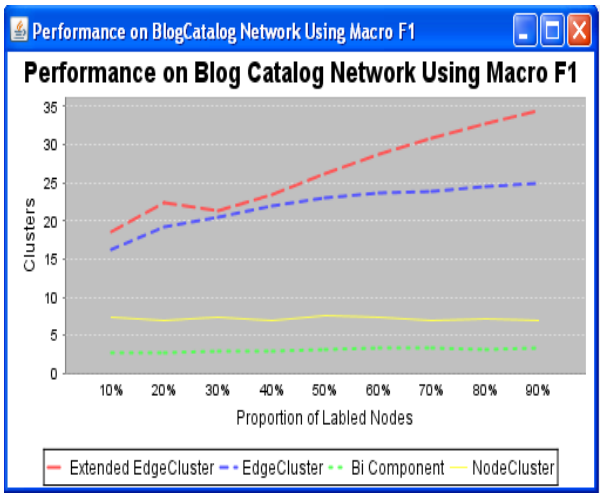


Figure 6 : Blog Catalog Network (macro F1)

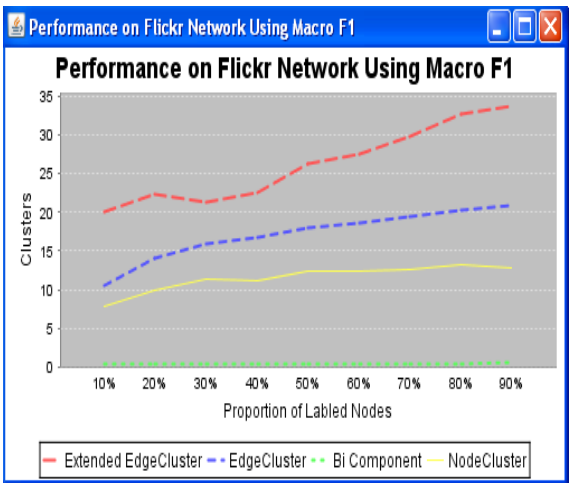


Figure 8 : Flickr network (macro F1)

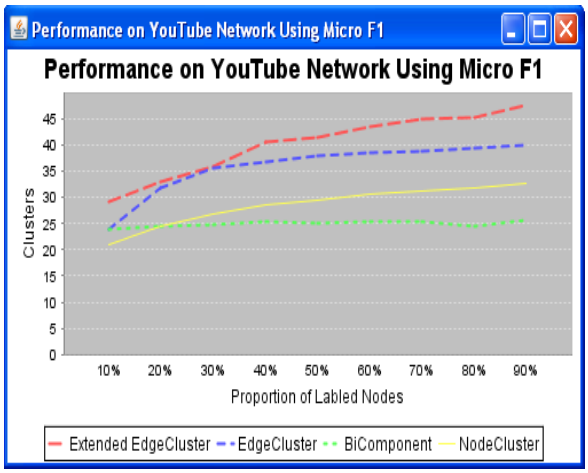


Figure 9 : You Tube network (micro F1)

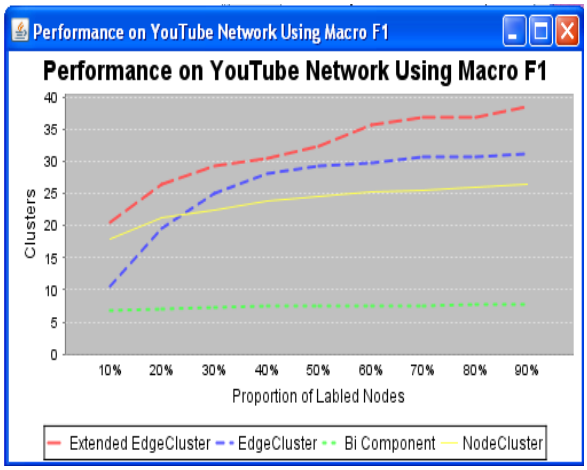


Figure 10 : You Tube network (macro F1)

In the experimental studies, it is observed that, proposed model shows best performance over existing one. Figure 5,7 and 9 shows Performance graph on Blog Catalog, Flickr, YouTube Network using Micro F-1 measure respectively and Figure 6,8 and 10 shows Performance graph on Blog Catalog, Flickr and YouTube Network using Macro F-1 measure respectively. Graph clarify that Extended Edge cluster method is superior to other given methods. In proposed model of Social Dimension extraction, extended idea of Edge Clustering works very efficiently and scalability is obtained consistently. To show best performance in the comparison to previous methods, the percentage label nodes for different extraction techniques is calculated [17].

Performance of Node Cluster and Bi Compnents is poor as scalability concern; their approach is Limited. Edge Cluster showing average performance in our model. By achieving better performance in accurate behavior prediction, scalability bottleneck, object heter – ogeneity, etc, Proposed approach Extended Edge-Cluster is proving for best results on issue discussed earlier. Perofrmance Graph for Blog Catalog, Flickr and You Tube network using Micro F1 and Macro F1 measure is given by following figures [18].

V. Conclusion and Future Work

a) Conclusion

In this research, an approach is presented to address the object heterogeneity in the networks. This idea is extended the scalability approach by heterogeneity in social network. Object heterogeneity is means same user involving in many activities in same social network. There are multiple modes of operation those are executed by one actor or one user. In previous methodology of collective behavior this approach was not used. In this research, mathematical model is designed and what's going to be our expected results. This research is aim to predict the results of collective behavior given a social network and the behavioral information of some people. Scalable learning of collective behavior even large numbers of individuals are involved in the network is carried out. This method follows a social-dimension learning model. These social dimensions are obtained to represent the potential affiliations of actors when discriminative learning occurs. For this scalability issue an extended edge-centric clustering scheme is proposed to obtain social dimensions and an efficient k-means variant for edge clustering algorithm. In this, each edge is referred as one data instance, and the connected nodes are the related features. Then, the proposed efficient k-means clustering algorithm can be applied to partition the edges into sets which are disjoint, with each set showing one possible group of activity. It is proved that using this edge-centric view, the obtained social dimensions are definitely sparse after extraction.

b) Future Work

In social media, multiple actors can be working in the same network, called a multimode network. YouTube, users, videos, tags, and comments are mixed with each other in co-existence. Extending the edge-centric clustering scheme to solve this object heterogeneity issue useful in future direction. Since the proposed Edge Cluster model is sensitive to the number of social dimensions. Future research focus is needed to determine a suitable dimensionality. It is useful to extract other behavioral features (e.g., user activities and temporal spatial information) from social media, and join them with social networking information to improve prediction performance.

References Références Referencias

1. "Scalable Learning of Collective Behavior", Lei Tang, Xufei Wang, and Huan Liu, IEEE 2012 Transactions on Knowledge and Data Engineering ,volume 24,Issue:6,2012.

2. Y.Y.Ahn, J.P.Bagrow, S.Lehmann, "Link communities reveal multi-scale complexity in networks", 2009.
3. M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," Annual Review of Sociology, volume. 27, pp. 415–444, 2001.
4. S. A. Macskassy and F. Provost, "Classification in networked data: A toolkit and a univariate case study", J. Mach. Learn. Res., volume. 8, pp. 935–983, 2007.
5. J. Neville and D. Jensen, "Leveraging relational autocorrelation with latent group models", in MRDM '05: Proceedings of the 4th international workshop on Multirelational mining. New York, NY, USA: ACM, pp. 49–55, 2005.
6. L. Tang and H. Liu, "Toward predicting collective behavior via social dimension extraction," *IEEE Intelligent Systems,* vol. 25, pp. 19–25, 2010.
7. L. Getoor and B. Taskar, Eds., "Introduction to Statistical Relational Learning". The MIT Press, 2007.
8. Z. Xu, V. Tresp, S. Yu, and K. Yu, "Nonparametric relational learning for social network analysis", in KDD'2008 Workshop on Social Network Mining and Analysis, 2008.
9. "Relational learning via latent social dimensions in KDD '09": Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, pp. 817–826, 2009.
10. M. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E, volume. 69, pp. 026113, 2004.
11. K. Yu, S. Yu, and V. Tresp, "Soft clustering on graphs", in NIPS, 2005.
12. Y.Y. Ahn, J.P. Bagrow, S. Lehmann, "Link comm. - unities reveal multi-scale complexity in networks", 2009.
13. "Relational learning via latent social dimensions," in KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, pp. 817–826 2009.
14. L. Tang, S. Rajan, and V. K. Narayanan, "Large scale multilabel classification via metalabeler," in WWW '09: Proceedings of the 18th international conference on World Wide Web. New York, NY, USA: ACM, pp. 211–220, 2009.
15. L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Community evolution in dynamic multi-mode networks," in KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, pp. 677–685, 2008.
16. J. Hopcroft and R. Tarjan, "Algorithm 447: efficient algorithms for graph manipulation,"Commun. ACM, volume 16, no. 6, pp.372–378, 1973.
17. L. Tang, S. Rajan, and V. K. Narayanan, "Large scale multilabel classification via metalabeler," in WWW: Proceedings of the 18th international confer - ence on World Wide Web. New York, NY, USA: ACM, 2009.
18. Vincent Van Asch, Macro- and micro-averaged evaluation measures", 2003.
19. L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Com - munity evolution in dynamic multi-mode networks," in KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009.
20. Michael P. S. Brown William Noble Grundy, David Lin, Nello Cristianini, Charles Sugnet, Manuel Ares, Jr., David Haussler. "Support Vector Machine Classification of Microarray Gene Expression Data", 1997.

Global Journals Inc. (US) Guidelines Handbook 2014

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards "FARSC" title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.

- The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

The following benefits can be availed by you only for next three years from the date of certification:

FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA).The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.

You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.

The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website https://associationofresearch.org which will be helpful to upgrade the dignity.

The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.

The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.

The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.
The “MARSC” is a dignified ornament which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.

MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

The following benefitscan be availed by you only for next three years from the date of certification.

MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.

Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.

AUXILIARY MEMBERSHIPS

Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as "Institutional Fellow of Open Association of Research Society" (IFOARS).

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

The Institute will be entitled to following benefits:

The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.

The author fees of such paper may be waived off up to 40%.

The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.

The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of "Open Association of Research Society, U.S.A (OARS)" so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.

The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

After nomination of your institution as "Institutional Fellow" and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.
The board can also take up the additional allied activities for betterment after our consultation.

The following entitlements are applicable to individual Fellows:

Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.

Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.

We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth $ 2376 USD.

Other:

The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.

- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10%discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in–depth understanding of the application of suitable techniques to a particular area of research practice.

Note :

"

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of "Difference of Opinion [if any]" among the Board members, our decision will be final and binding to everyone.

"

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

1. Choosing the topic: In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. Pick a good study spot: To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. Think technically: Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.

- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350