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

Global Journal of Computer Science and Technology: C

Software & Data Engineering

Highlights

- Simulation of Reliability
- Analysis of Data Mining
- Decision Tree Technique
- Algorithms and Applications

Discovering Thoughts, Inventing Future

VOLUME 13 ISSUE 13 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 13 Issue 13 (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: *investers@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

Contents of the Volume

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Analysis of Data Mining Classification with Decision tree Technique

By Dharm Singh, Naveen Choudhary & Jully Samota
Maharana Pratap University of Agriculture and Technology, India

Abstract- The diversity and applicability of data mining are increasing day to day so need to extract hidden patterns from massive data. The paper states the problem of attribute bias. Decision tree technique based on information of attribute is biased toward multi value attributes which have more but insignificant information content. Attributes that have additional values can be less important for various applications of decision tree. Problem affects the accuracy of ID3 Classifier and generate unclassified region. The performance of ID3 classification and cascaded model of RBF network for ID3 classification is presented here. The performance of hybrid technique ID3 with CRBF for classification is proposed. As shown through the experimental results ID3 classifier with CRBF accuracy is higher than ID3 classifier.

Keywords: data mining, classification, decision tree, ID3, attribute selection.

GJCST-C Classification : H.2.8

Strictly as per the compliance and regulations of:

Analysis of Data Mining Classification with Decision Tree Technique

Dharm Singh [α], Naveen Choudhary [σ] & Jully Samota [ρ]

Abstract- The diversity and applicability of data mining are increasing day to day so need to extract hidden patterns from massive data. The paper states the problem of attribute bias. Decision tree technique based on information of attribute is biased toward multi value attributes which have more but insignificant information content. Attributes that have additional values can be less important for various applications of decision tree. Problem affects the accuracy of ID3 Classifier and generate unclassified region. The performance of ID3 classification and cascaded model of RBF network for ID3 classification is presented here. The performance of hybrid technique ID3 with CRBF for classification is proposed. As shown through the experimental results ID3 classifier with CRBF accuracy is higher than ID3 classifier.

Keywords: data mining, classification, decision tree, ID3, attribute selection.

I. Introduction

With the rapid development of information technology and network technology, different trades produce large amounts of data every year. The data itself cannot bring direct benefits so need to effectively mine hidden information from huge amount of data. Data mining deals with searching for interesting patterns or knowledge from massive data. It turns a large collection of data into knowledge. Data mining is an essential step in the process of knowledge discovery (Lakshmi & Raghunandhan, 2011).The data mining has become a unique tool in analyzing data from different perspective and converting it into useful and meaningful information.Data mining has been widely applied in the areas of Medical diagnosis, Intrusion detection system, Education, Banking, Fraud detection.

Classification is a supervised learning. Prediction and classification in data mining are two forms of data analysis task that is used to extract models describing data classes or to predict future data trends. Classification process has two phases; the first is the learning process where the training data sets are analyzed by classification algorithm. The learned model or classifier is presented in the form of classification rules or patterns. The second phase is the use of model for classification, and test data sets are used to estimate the accuracy of classification rules.

Authors α σ ρ: Department of Computer Science and Engineering, College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan, India. e-mails: dharm@mpuat.ac.in, naveenc121@yahoo.com, jullysamota304@gmail.com

With the rising of data mining, decision tree plays an important role in the process of data mining and data analysis. Decision tree learning involves in using a set of training data to generate a decision tree that correctly classifies the training data itself. If the learning process works, this decision tree will then correctly classify new input data as well. Decision trees differ along several dimensions such as splitting criterion, stopping rules, branch condition (univariate, multivariate), style of branch operation, type of final tree (Han, Kamber & Pei, 2012).

The best known decision tree induction algorithm is the ID3. ID3 is a simple decision tree learning algorithm developed by Ross Quinlan. Its predecessor is CLS algorithm. ID3 is a greedy approach in which top-down, recursive, divide and conquer approach is followed. Information gain is used as attribute selection measure in ID3. ID3 is famous for the merits of easy construction and strong learning ability. There exists a problem with this method, this means that it is biased to select attributes with more taken values, which are not necessarily the best attributes. This problem affects its practicality. ID3 algorithm does not backtrack in searching. Whenever certain layer of tree chooses a property to test, it will not backtrack to reconsider this choice. Attribute selection greatly affects the accuracy of decision tree (Quinlan, 1986).

In rest of the paper, a brief introduction to the related work in the area of decision tree classification is presented in section 2. A brief introduction to the proposed work is presented in section 3. In section 4 we present the experimental results and comparison. In section 5, we conclude our results.

II. Related Work

The structure of decision tree classification is easy to understand so they are especially used when we need to understand the structure of trained knowledge models. If irrelevant attribute selection then all results suffer. Selection space of data is very small if we increase space, selection procedure suffers so problem of attribute selection in classification. There have been a lot of efforts to achieve better classification with respect to accuracy.

Weighted and simplified entropy into decision tree classification is proposed for the problem of multiple-value property selection, selection criteria and property value vacancy. The method promotes the efficiency and precision (Li & Zhang, 2010).A comparison of attribute selection technique with rank of attributes is presented. If irrelevant, redundant and noisy attributes are added in model construction, predictive performance is affected so need to choose useful attributes along with background knowledge (Hall & Holmes, 2003). To improve the accuracy rate of classification and depth of tree, adaptive step forward/decision tree (ASF/DT) is proposed. The method considers not only one attribute but two that can find bigger information gain ratio (Tan & Liang, 2012). A new heuristic technique for attribute selection criteria is introduced. The best attribute, which have least heuristic functional value are taken. The method can be extended to larger databases with best splitting criteria for attribute selection (Raghu, Venkata Raju & Raja Jacob, 2012).Interval based algorithm is proposed. Algorithm has two phases for selection of attribute. First phase provides rank to attributes. Second phase selects the subset of attributes with highest accuracy. Proposed method is applied on real life data set (Salama, M.A., El Bendary, N., Hassanien, Revett & Fahmy, 2011).

Large training data sets have millions of tuples. Decision tree techniques have restriction that the tuples should reside in memory. Construction process becomes inefficient due to swapping of tuples in and out of memory. More scalable approaches are required to handle data (Changala, R., Gummadi, A., Yedukondalu & Raju, 2012).An improved learning algorithm based on the uncertainty deviation is developed. Rationality of attribute selection test is improved. An improved method shows better performance and stability (Sun & Hu, 2012).

Equivalence between multiple layer neural networks and decision trees is presented. Mapping advantage is to provide a self configuration capability to design process. It is possible to restructure as a multilayered network on given decision tree (Sethi, 1990).A comparison of different types of neural network techniques for classification is presented. Evaluation and comparison is done with three benchmark data set on the basis of accuracy (Jeatrakul & Wong, 2009).

The computation may be too heavy if no preprocessing in input phase. Some attributes are not relevant .To rank the importance of attributes, a novel separability correlation measure (SCM) is proposed. In input phase different subsets are used. Irrelevant attributes are those which increase validation error (Fu & Wang, 2003).

III. Proposed Method

The input processing of training phase is data sampling technique for classifier. Single layer RBF networks can learn virtually any input output relationship (Kubat, 1998). The cascade-layer network has connections from the input to all cascaded layers. The additional connections can improve the speed. Artificial neural networks (ANNs) can find internal representations of dependencies within data that is not given. Short response time and simplicity to build the ANNs encouraged their application to the task of attribute selection.

IV. Process Method

1. Sampling of data from sampling technique
2. Split data into two parts training and testing part
3. Apply CRBF function for training a sample value
4. Using 2/3 of the sample, fit a tree the split at each node.
   For each tree
   - Predict classification of the available 1/3 using the tree, and calculate the misclassification rate
     = out of CRBF.
5. For each variable in the tree
6. Compute Error Rate: Calculate the overall percentage of misclassification
   - Variable selection: Average increase in CRBF error over all trees and assuming a normal division of the increase among the trees, decide an associated value of feature.
7. Resulting classifier set is classified
   - Finally to estimate the entire model, misclassification.
8. Decode the feature variable in result class

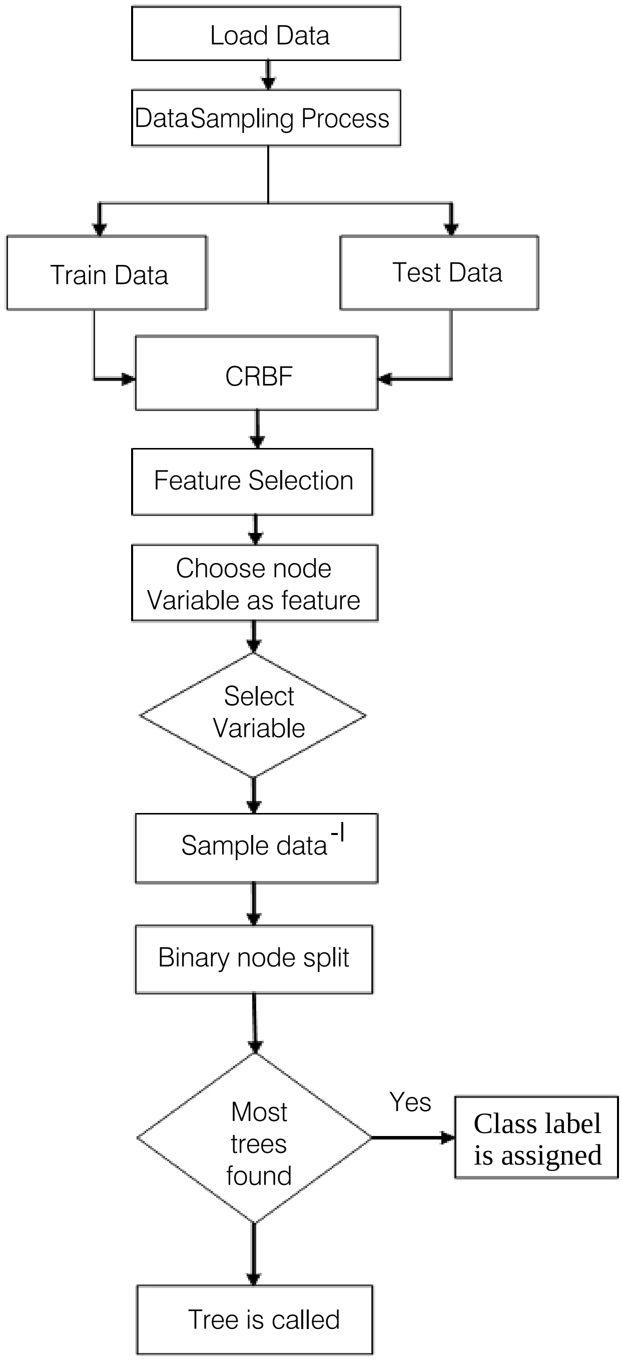


Figure 1 : Process block diagram of modified ID3-CRBF

V. Experimental Results

For the performance evaluation cancer dataset from UCI machine learning repository is used.

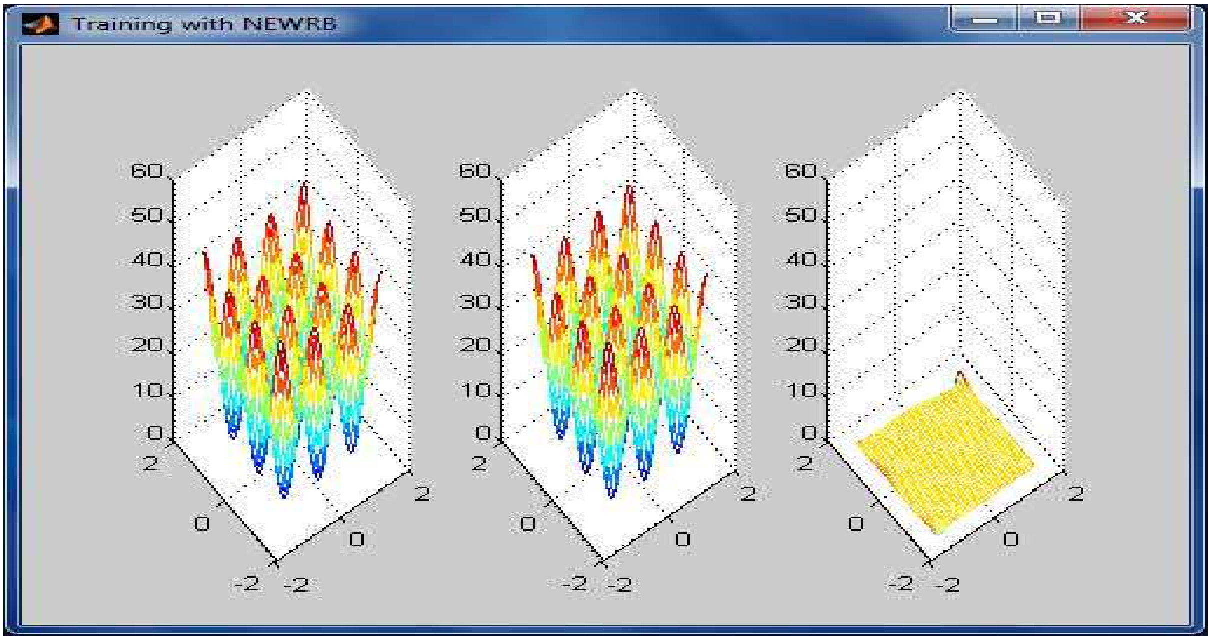


Figure 2 : Trained data, test data and unclassified region classification

Accuracy based on cross fold ratio of classifier.

Table 1 : Comparison of Accuracy

Cross Fold Ratio	Accuracy	
	ID3	ID3_CRBF
5	86.18	97.18
6	81.95	92.95
7	81.95	92.95

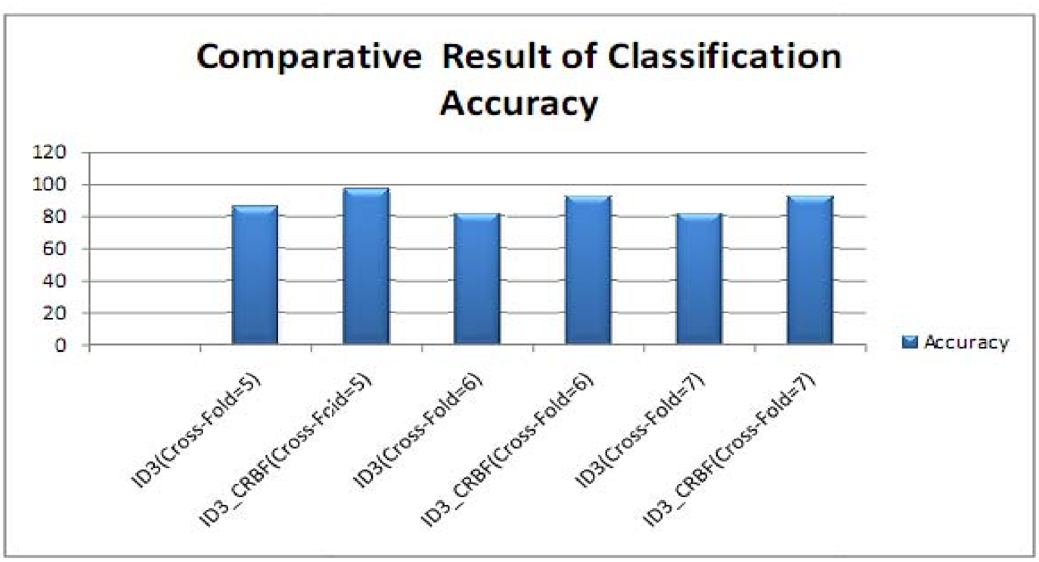


Figure 3 : Graphical representation of Table 1

VI. Conclusion

In this paper, we have experimented cascaded model of RBF with ID3 classification. The standard presentation of each attribute on selected ID3 is calculated and the Classify the given data. We can say from the experiments that the cascaded model of RBF with ID3 approach provides better accuracy and reduces the unclassified region. Increased classification region improves the performance of classifier.

References Références Referencias

1. Lakshmi, B.N., Raghunandhan, G.H. (2011). A conceptual overview of data mining. Proceedings of the National Conference on Innovations in Emerging Technology, 27-32.
2. Han, J., Kamber, M., Pei, J. (2012) DATA MINING: Concepts and Techniques. Elsevier. pp. 327-346.
3. Li, L., Zhang, X. (2010). Study of data mining algorithm based on decision tree. International Conference On Computer Design And Applications Vol. 1.
4. Quinlan, J.R. (1986). Induction of Decision Trees. Machine Learning, 1 (1): 81-106.
5. Hall, M.A., Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. IEEE transactions on knowledge and data engineering, Vol. 15, No.6.
6. Tan, T.Z., Liang, Y.Y. (2012). ASF/DT, Adaptive step forward decision tree construction. Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition.
7. Raghu, D., Venkata Raju, K., Raja Jacob, Ch. (2012). Decision tree induction-A heuristic problem reduction approach. International Conference on Computer Science and Engineering.
8. Salama, M.A., El-Bendary, N., Hassanien, A.E., Revett, K., Fahmy, A.A. (2011). Interval-based attribute evaluation algorithm. Proceedings of the Federated Conference on Computer Science and Information Systems, 153-156.
9. Changala, R., Gummadi, A., Yedukondalu, G., Raju, U. (2012). Classification by decision tree induction algorithm to learn decision trees from the class-labeled training tuples. International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2. Issue 4.
10. Sun, H., Hu, X. (2012). An improved learning algorithm of decision tree based on entropy uncertainty deviation.
11. Sethi, I.K. (1990). Layered neural net design through decision trees.
12. Jeatrakul, P., Wong, K.W. (2009). Comparing the performance of different neural networks for binary classification problems. Eighth International Symposium on Natural Language Processing.
13. Kubat, M. (1998). Decision trees can initialize radial-basis function networks. IEEE transactions on neural networks, vol. 9, no. 5.
14. Fu, X., Wang, L. (2003). Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. IEEE transactions on systems, man, and cybernetics vol. 33, no. 3.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

TPA Ensuring Data Integrity in Cloud Environment

By Jaspreet Kaur & Jasmeet Singh
Mandi Gobindgarh/ Punjab Techincal University, India

Abstract- Cloud computing is an internet based computing where virtual shared servers provide software, infrastructure, platform and other resources to customers on a pay-as-you-use basis. In one of the services offered by cloud viz. Storage as a Service, users outsource their data to cloud without having direct possession or control on it. Storage of large data in cloud reduces costs and maintenance. But the customer is unaware of the storage location. Here risk involved is modification of data or tampering of data. In this paper we propose a data correctness scheme in which Third Party can audit the data stored in the cloud and assure the customer that the data is safe. Hence we implemented a scheme for verifying integrity of data. Such verification systems prevent the cloud storage archives (storage) from misrepresenting or modifying the data stored by using frequent checks on the storage archives.

Keywords: *audit, cloud, integrity, station to station protocol, SHA-2, third party auditor, XOR.*

GJCST-C Classification : *C.2.1*

TPAENSURINGDATAINTEGRITYINCLOUDENVIRONMENT

Strictly as per the compliance and regulations of:

TPA Ensuring Data Integrity in Cloud Environment

Jaspreet Kaur [α] & Jasmeet Singh [σ]

Abstract- Cloud computing is an internet based computing where virtual shared servers provide software, infrastructure, platform and other resources to customers on a pay-as-you-use basis. In one of the services offered by cloud viz. Storage as a Service, users outsource their data to cloud without having direct possession or control on it. Storage of large data in cloud reduces costs and maintenance. But the customer is unaware of the storage location. Here risk involved is modification of data or tampering of data. In this paper we propose a data correctness scheme in which Third Party can audit the data stored in the cloud and assure the customer that the data is safe. Hence we implemented a scheme for verifying integrity of data. Such verification systems prevent the cloud storage archives (storage) from misrepresenting or modifying the data stored by using frequent checks on the storage archives.

Keywords: audit, cloud, integrity, station to station protocol, SHA-2, third party auditor, XOR.

I. Introduction

Cloud Computing is emerging as the next evolution of computing for its numerous contribution to the IT enterprise. In contrast to traditional solutions, Cloud Computing has a number of essential characteristics, such as: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid elasticity, and measured service.[3]

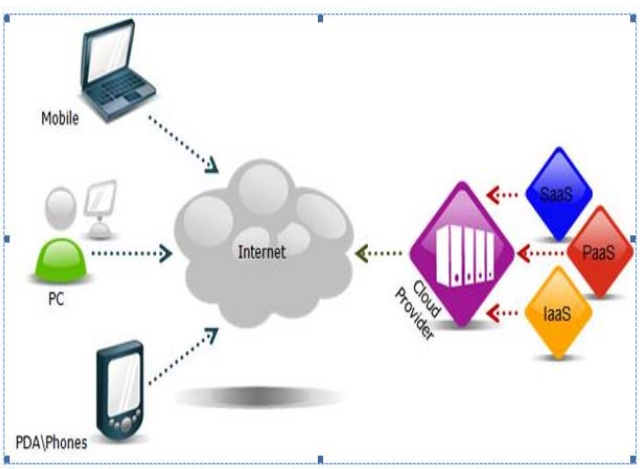


Figure 1 : Cloud Computing

Author α: Jaspreet Kaur Lecturer, Thapar Polytechnic College M.TECH (Computer Science and Engineering) RIMT-IET, Mandi Gobindgarh, Punjab India. e-mail: kkhushi12@yahoo.co.in
Author σ: Jasmeet Singh Asst. Professor(CSE Department) RIMT-IET, Mandi Gobindgarh, Punjab India. e-mail: jasmeetgum@gmail.com

Data outsourcing, one of the fundamental components of Cloud Computing, centralizes users' data to the cloud server (CS). Users including both individuals and enterprises move their data or store their data in cloud storage centers to minimize the costs and enjoy benefits, such as: release the pressure from storage management, universal data access with independent geographical locations.[4]

Despite these distinct advantages outsourcing brings, there are also many security issues.

1. Firstly, although the cloud service provider (CSP) can provide more powerful and reliable infrastructures than users, a huge mass of data storing in the CS makes it more vulnerable to active attack.
2. Secondly, towards the cloud users, the CSP may deliberately distort the status of users' outsourced data for some benefits. For example, the CSP may discard the data that users rarely or never access to save costs, or even hide data loss incidents for the sake of reputation .[4]

So we can see, although data outsourcing can bring advantages and convenience to users, it can not ensure the integrity of data. But integrity monitoring is essential in cloud storage for the same reasons that data integrity is critical for any data center. Data corruption can happen at any level of storage and with any type of media. The truth is that data corruption can happen anywhere within a storage environment. Data can become corrupted simply by migrating it to a different platform, i.e., sending your data to the cloud. Cloud storage systems are still data centers, with hardware and software, and are still vulnerable to data corruption. [13]

Generally speaking, users like to outsource a huge mass of data in the CS, so simply downloading the data to verify the integrity is not a feasible solution. To solve the security problems of data outsourcing mentioned above, researchers proposed auditing protocols to ensure the correctness of the outsourced data. The integrity of data should be guaranteed in a relatively low computation and communication overhead through an efficient auditing protocol.[4]

So he appoints a Third Party Auditor to check the availability of data and its correctness without devotion of their computation resources. [1] TPA checks the correctness of data stored in the cloud and

Global Journal of Computer Science and Technology (C) Volume XIII Issue I Version I Year 2013

Global Journal of Computer Science and Technology (C) Volume XIII Issue XIII Version I Year 2013

communicates this with the client. Whenever the client needs the data the cloud returns the data with full guarantee of delivery, availability and correctness. As TPA verifies for its correctness and availability he considers the data is safe.

Figure 2 represents data outsourcing in cloud environment where user delegates task of monitoring integrity to third party auditor(TPA).

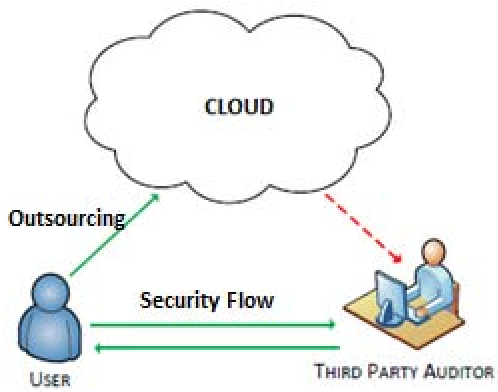


Figure 2 : Data Outsourcing in Cloud

II. Problem Definition

In Cloud Computing, it is difficult to maintain data integrity because the user usually has no control over the security mechanisms that are used to protect his/her data. User cannot trust cloud service provider to handle the data by himself as he himself can modify or delete the original data and integrity may be lost. If any intruder attacks and steals the data and modifies it then in some cases the modification is not even noticed by the cloud server or data loss or corruption is intentionally hidden. So, user can rely on a trusted third party auditor to check for the integrity of his data. To trust third party entity authentication is needed. For auditing (on user's request or at regular intervals), strong and secure cryptographic hash function is required to check for integrity of cloud data and informs the user about data corruptions or loss if any.

III. Methodology

We propose a data correctness scheme which involves verifying integrity of data with the help of third part auditor as shown in figure 3.

For ensuring the integrity of the data we will be using combination of three approaches-

1. Station-to-Station protocol (based on Diffie-Hellman key exchange algorithm) generates mutual key which is known to both user and auditor. It also provides entity authentication to both.
2. Exclusive-OR (XOR) to perform a xor operation between the message and the key generated using Station-to-Station protocol.
3. Secure Hashing Algorithm (SHA-2) to generate a digest by passing the original message to the hash function. This is done by both the user and the auditor and the value obtained from the hash function by both of them is compared and hence the data integrity is verified.

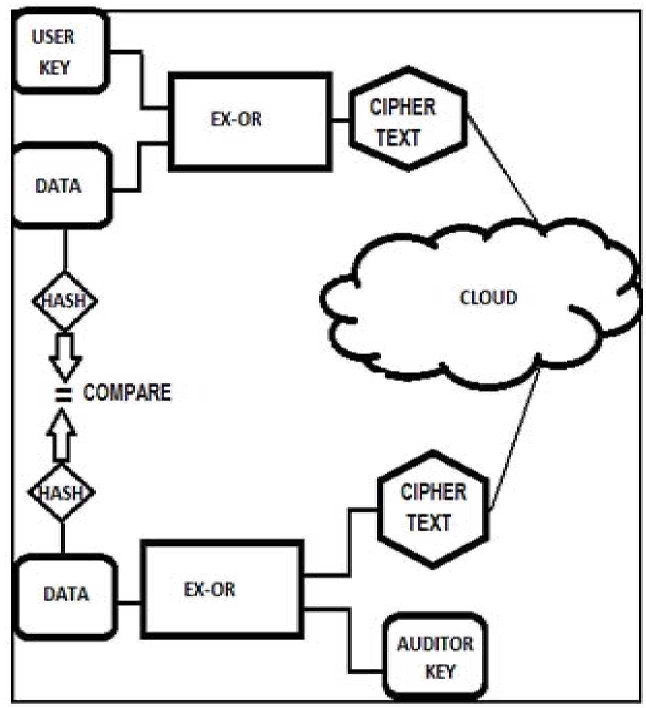


Figure 3 : Methodology

IV. Implementation

Using java netbeans IDE and XAMPP, we have implemented methodology in which TPA ensures integrity of data outsourced by user in the cloud storage and thus reduces overhead of user.

We have created three pages(forms)-

1. client side(figure 4)
2. cloud server (figure 5)
3. auditor side(figure 6)

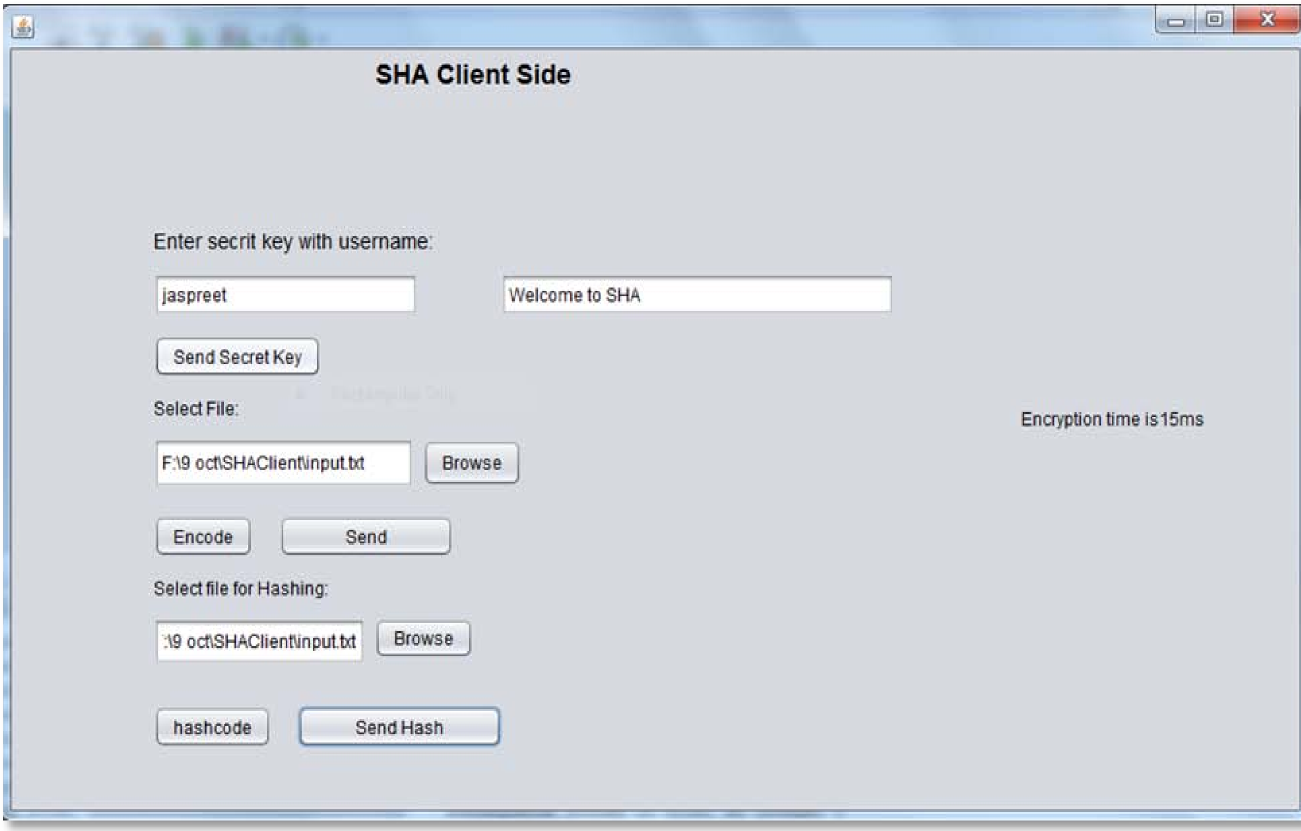


Figure 4 : Client page

Figure 5 : Cloud page

Global Journal of Computer Science and Technology (C) Volume XIII Issue XIII Version I Year 2013
9

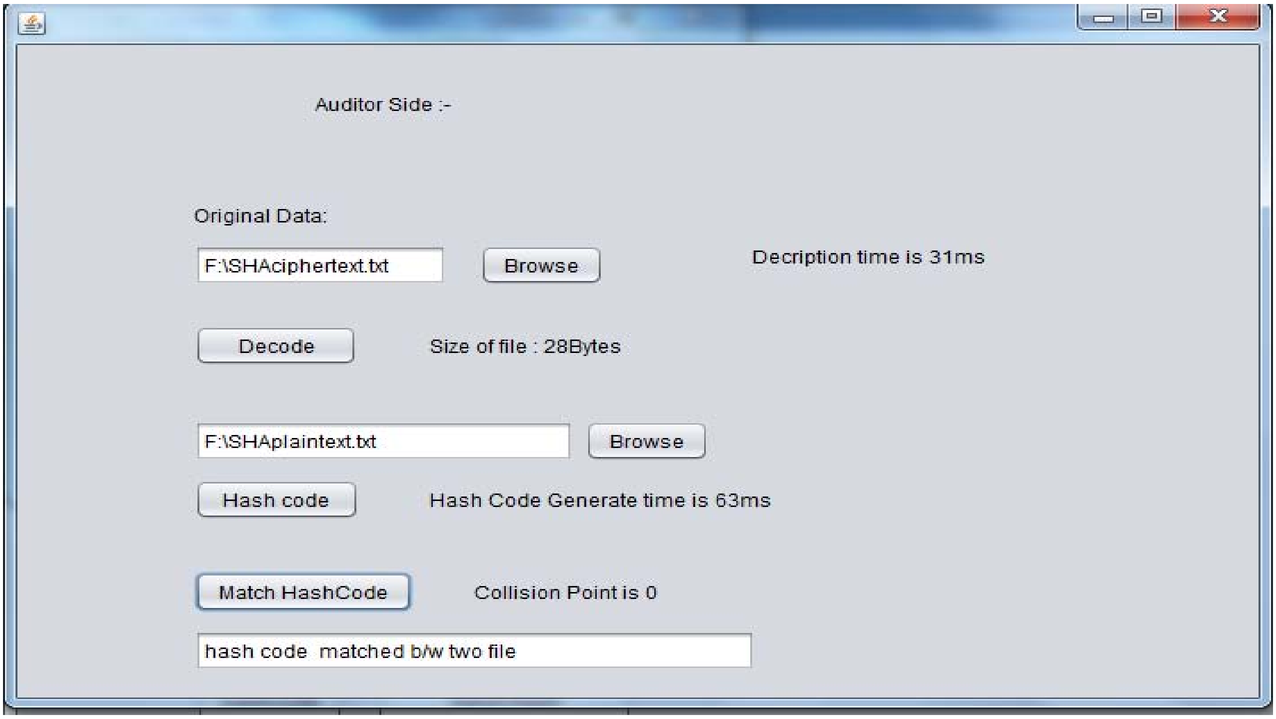


Figure 6 : Auditor Page

The steps of implementation are as follows-

1. A secret key generated using STS protocol (that is known to both user and auditor. Also, mutual authentication is done using this protocol.
2. XOR operation is done between the data and the key generated to create cipher text which is stored in cloud.
3. Separately the data is passed in a hash function (using SHA-2) and the hash value is obtained by the user.
4. Auditor gets the cipher text from the cloud and performs an XOR operation with the secret key generated by the station to station protocol and gets a plain text.
5. Auditor passes this plain text to the same hash function (using SHA2) and obtains a hash value.
6. He then compares this hash value with the hash value received from the user .If both the values are identical then the data integrity is maintained else data is tampered.

V. Results

The results of the above mentioned system are shown in Table 1 and Figure 7.

Table 1 : Result Analysis

	File Size(in Bytes)			
	28	99	124	205
Encryption Time	16	94	141	211
Decryption Time	31	141	187	265
Hash Time	63	78	93	108
Collision Point	0	0	0	0

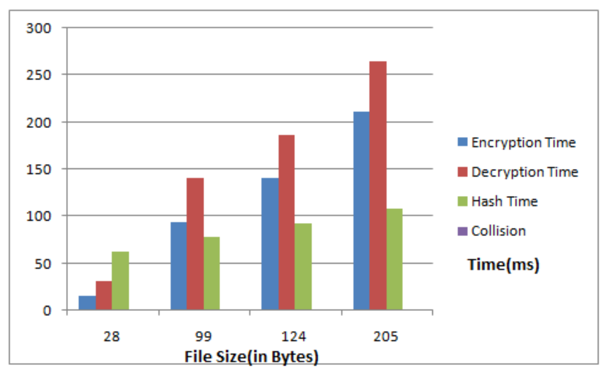


Figure 7 : Graph Showing Results of Encryption, Decryption, Hash Time and Collision

If hash of both files matches, collision point is 0.

If hash of both files do not match , collision point is 1.

VI. Conclusion and Future Work

This paper focuses on auditing mechanisms (using hash function) to ensure data integrity where users can safely delegate the integrity checking tasks to Third Party Auditors and be worry-free to use the cloud storage services. This scheme reduces the computational and storage overhead of the client as well as the computational overhead of the cloud storage server.. We are trying to improve the scheme for auditing multiple files from multiple clients simultaneously as with the increasing development of Cloud Computing technologies, it is believed that more and more users will prefer to store their data in the cloud.

References Références Referencias

1. Malathi, Murugesan, "A Scheme for Checking Data Correctness in the Cloud", International Conference on Information and Network Technology (ICINT 2012).
2. Miss. M.Sowparnika1, Prof. R. Dheenadayalu2, "Improving data integrity on cloud storage services", International Journal of Engineering Science Invention(ISSN), Volume 2, Issue 2 ,February. 2013.
3. N.Madhuri, T.V.Suneetha, A.Haritha, P.V.S.Lakshmi, "A Protocol for Ensuring Data Integrity in Cloud Environment", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.
4. XU Chun-xiang, HE Xiao-hu, Daniel Abraha, "Cryptanalysis of auditing protocol proposed by Wang et al. for data storage security in Cloud Computing", School of Computer Science and Engineering, University of Electronics Science and Technology of China, 2006.
5. K.Govinda, E.Sathiyamoorthy, "Data Auditing in Cloud Environment using Message Authentication Code", International Conference on Emerging Trends on Advanced Engineering Research (ICETT), 2012.
6. K.Govinda, V.Gurunathprasad, H.Sathishkumar, "Third Party Auditing for Secure Data Storage in Cloud through Digital Signature using RSA", International Journal of Advanced Scientific and Technical Research, Vol. 4, Issue 2, August 2012.
7. Abhishek Mohta, Lalit Kumar Awasthi, "Cloud Data Security while using Third Party Auditor", International Journal of Scientific and Engineering Research, Vol. 3, Issue 6, June-2012.
8. Akkala Saibabu, T.Satyanarayana Murthy, "Security Provision in Publicly Auditable Secure Cloud Data Storage Services using SHA-1 Algorithm", International Journal of Computer Science and Information Technologies (IJCSIT) Vol. 3(3), 2012.
9. Reenu Sara Georeg, Sabitha S," Survey on Data Integrity in Cloud Computing", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 2, Issue 1, January 2013.
10. Hiren Patel and Dhiren Patel,"Achieving Secure Cloud Data Storage without using a Trusted Third Party Auditor- A Review", International Journal of Computer Applications,Nov.2012.
11. S.Jeyadevan, Dr.S.Basavarij Patel, "A DATA INTEGRITY VERIFICATION IN CLOUD COMPUTING", International Conference on Computing and Control Engineering(ICCCE 2012),12 & 13 April,2012.
12. [7]Dalia Attas, Omar Batrafi," Efficient integrity checking technique for securing client data in cloud computing", International Journal of Electrical & Computer Sciences (IJECS-IJENS), Vol: 11, No: 05.
13. http://www.wwpi.com/index.php?option=com_content&id=12800:data-integrity-in-the-cloud&Itemid=2701018
14. http://en.wikipedia.org/wiki/Cloud_computing.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Simulation of Reliability of Software Component

By Dr. P. K. Suri & Er. Karambir
University Institute of Eengg and Technology, India

Abstract- Component-Based Software Engineering (CBSE) is increasingly being accepted worldwide for software development, in most of the industries. Software reliability is defined as the probability that a software system operates with no failure within a specified time on specified operating conditions. Software component reliability and failure intensity are two important parameters that Estimates the reliability of system after integration of component. The estimation of reliability of software can save loss of time, life and cost. In this paper, software reliability has been estimated by analyzing the failure data. The Imperfect Software Reliability Growth Models (SRGMs) model have been used for simulating the software reliability by estimating the number of remaining faults and the model parameters of the fault content rate function. We aim for simulating software reliability by connecting the imperfect debugging and Goel-Okumoto model. The estimation of reliability gives the time of stopping the unending testing of that component or time of release of software component.

Keywords: component based software engineering (cbse), software reliability growth model (srgm), reliability, goel- okumoto model.

GJCST-C Classification : D.2.4

SIMULATIONOFRELIABILITYOFSOFTWARECOMPONENT

Strictly as per the compliance and regulations of:

OPEN ASSOCIATION OF RESEARCH SOCIETY, USA
RESEARCH | DIVERSITY | ETHICS

Simulation of Reliability of Software Component

Dr. P. K. Suri [α] & Er. Karambir [σ]

Abstract- Component-Based Software Engineering (CBSE) is increasingly being accepted worldwide for software development, in most of the industries. Software reliability is defined as the probability that a software system operates with no failure within a specified time on specified operating conditions. Software component reliability and failure intensity are two important parameters that Estimates the reliability of system after integration of component. The estimation of reliability of software can save loss of time, life and cost. In this paper, software reliability has been estimated by analyzing the failure data. The Imperfect Software Reliability Growth Models (SRGMs) model have been used for simulating the software reliability by estimating the number of remaining faults and the model parameters of the fault content rate function. We aim for simulating software reliability by connecting the imperfect debugging and Goel-Okumoto model. The estimation of reliability gives the time of stopping the unending testing of that component or time of release of software component.

Keywords: component based software engineering (cbse), software reliability growth model (srgm), reliability, goel- okumoto model.

I. Introduction

With the popularity of the web and networked computers are finding their way into a wide and spread range of working environments. This new computing model have made a competition of early development, reliable and distributed software components that communicate with one another across the underlying networked and extendable infrastructure as per the requirement of different user. A distributed software component can be plugged into distributed applications and can be used for some specific purpose. The intention of most of the developers behind is reuse or slight modification of old or reliable component and this makes more reliable by using distributed software components to build new systems. Even though, it is also important for developer to know the functionality of distributed or compatible software component in any system. The design of component and requirement specification should clearly document the functional input, output with conditions and moreover it is the reliability percentage wise. Software reliability has been defined as the probability that a software system operates with no failure for a specified time on specified operating conditions. In other words, by estimating or predicting the reliability [1] of component, the quality of software product can be estimated. The satisfaction of customers is directly dependent on the quality of that software. The analysis report that is commonly used to describe software reliability has been derived from observed or failure intensity. Failure intensity is defined as the number of failures observed per unit time period. Failure intensity is a also good measure for reflecting the user perspective of software quality. As Computer applications are going more diverse and spreading through almost every area of everyday life then reliability factor becomes a very important characteristic of software or component systems. The reliable component is a base of system and part of system i.e. client, administrator and working environment. Since it is a matter of cost and performance to produce a system having documented and estimated reliability [2] of system. Therefore, it is necessary to measure its reliability before releasing any software. When reliability reaches at threshold level then the software component can be released for further use. To do this, a number of models [3] have been proposed and has been being developed. Software modeling is a statistical estimation [4] method applied to failure data collected or simulated the software component or system developed after integration of software component by different approach of joining in software engineering .This can be one after a component testing has been executed so that failure data are available. The implementation of newly developed and modified models tries to make system better and help in predicting the reliability in a accurate way. The most important parameter of any software product are level of quality, time of delivery, and final cost of the product. The time of delivery and cost should be quantitative and pre decided, whereas these attributes is difficult to define Quantitatively. Reliability is one, and probably the most Software reliability is related directly to operation and performance instead of designing of a component.

Therefore, software reliability is estimated by analyzing the observed failure data [5] of component and then applying Goel –Okumoto Model [6][7] , rather than the number of remaining faults in a component. So, estimation of reliability of system is more useful than finding the number of remaining fault. The uncertainty involved in the estimation for a specific interval expressed in terms of confidence interval and estimation of parameter used. This paper evaluates the estimates the reliability of component by using the Goel- Okumoto

Author α: Dean (R &D), Professor & Chairman (CSE/IT/MCA) HCTM Technical Campus, Kaithal (Haryana) India.
e-mail: pksurikuk@gmail.com
Author σ: Assistant Professor, Department of Computer Science and Engineering, University Institute of Engg and Technology, Kurukshetra Univeristy, Kurukshetra (Haryana) India.
e-mail: bidhankarambir@rediffmail.com

model on the set of failure data taken from simulating the real software applications. This should be done before any component release .The reusability of that component enhances the overall reliability of system and gives accurate estimation of value of reliability. The result shows that the proposed model has a technical point for improving software reliability and providing additional metrics for development project evaluation, management and time of delivery of new developed system.

II. Goel Okumoto Model : NHPP SRGM Exponential Model

Non Homogeneous Poisson Process (NHPP) based software reliability growth models are generally classified into two groups. The first group contains models, which use the machine execution time or calendar time [8] as a unit of fault detection/removal period. Such models are called continuous time models. The second group contains models, which use the number of test occasions/cases as a unit of fault detection period. Such models are called discrete time models [9], since the unit of software fault detection period is countable. A Goel Okumoto Model also known as exponential NHPP model is based on the following assumptions:(a) All fault in a component are independent from the failure detected. (b)The number of failures detected at any time is proportional to the current number of fault in a component. (c) The fault is removed immediately as soon as the failure happens, no new faults are introduced during the removal of fault.

The following differential Equation1 include the above assumptions. Where m(t) is expected number of component failures by time t, a is Total fault content rate function, i.e., the sum of expected number of initial software faults and introduced faults by time t and b is Failure detection rate per fault at time t.
important, aspect of software quality. S

$$\frac{\partial m(t)}{\partial t} = b[a - m(t)] \qquad (1)$$

The mean value solution of above differential equation is given by Equation 2 where tn is time of nth failure occurrence

$$m(t) = a\left(1 - e^{-bt_n}\right) \qquad (2)$$

Failure intensity function is given by Equation 3 as follows

$$\lambda(t) = abe^{-bt_n} \qquad (3)$$

III. Estimation of Parameter

The different a and b parameter also reflect different assumptions of the software testing processes. In this section, we derive a new NHPP model for an interrelationship dependent function between a and b parameter by a common parameter from a generalized class of model. The most common method for the estimation of parameters is the Maximum Likelihood Estimation (MLE) method. MLE method of estimation of a broad collection of software reliability for grouped data is discussed in detail. To estimate a and b for a sample of n units, first obtain the likelihood function: take the natural logarithm on both sides. The equation for estimation of a and b is given in Equation 4 where

$$a = \frac{y_n}{(1-e^{-bt_n})} \qquad (4)$$

Where yn is actual value of nth failure observed at time t. The parameter a can be estimated using MLE method based on the number o failures in a particular interval. Suppose that an observation interval {0, tk} is divided into set of sub intervals (0,ti],(t1,t2]………(tk-1,tk] , Equation 5 was used to determine the value of b .

$$\frac{y_n t_n e^{-bt_n}}{1-e^{-bt_n}} = \sum_{k=1}^{n} \left(\frac{(y_k - y_{k-1})(t_k e^{-bt_k} - t_{k-1} e^{-bt_{k-1}})}{(e^{-bt_{k-1}} - e^{-bt_k})} \right) \qquad (5)$$

The number of failures per subinterval [8] is recorded as ni(i=1,2,3..,k) with respect to the number of failures in(ti-1,ti].The parameters a and b are estimated using iterative Newton Raphson Method, which is given as in Equation 6 Equation 7 and Equation 8 .

$$b = b_0 - \frac{f(b_0)}{f'(b_0)} \qquad (6)$$

$$f(b) = \frac{y_n t_n e^{-bt_n}}{1 - e^{-bt_n}} - \sum_{k=1}^{n} \left(\frac{(y_k - y_{k-1})(t_k e^{-bt_k} - t_{k-1} e^{-bt_{k-1}})}{(e^{-bt_{k-1}} - e^{-bt_k})} \right) = 0 \qquad (7)$$

$$f'(b) = \sum_{k=1}^{n} \left(\frac{(y_k - y_{k-1})(t_k - t_{k-1})^2 e^{-b(t_k + t_{k-1})}}{(e^{-bt_{k-1}} - e^{-bt_k})^2} \right) \qquad (8)$$

IV. Model Analysis and Results

a) Data and Model Criteria

Once the analytical expression for the mean value function m(t) is derived, in this paper, the model parameters to be estimated in the mean value function can then be obtained with the help of a developed octave program based on the least squares estimate (LSE) method. Goel and Okumoto described failure detection as a non-homogeneous Poisson process with an exponentially decaying rate function .It is a simple non-homogeneous Poisson process model. The data of failure of 25 days have been observed here for estimating the reliability [10]. In table 1, the data of 25 days failure and cumulated failure have been shown here.

The two function of Reliability and Remaining fault function can be used to find the release of date or the additional testing time is required to reach ready state. After simulation the result of 25 days of testing were observed. Based on these data and using the MLE

Global Journal of Computer Science and Technology (C) Volume XIII Issue XIII Version I Year 2013

method, the estimated values for the two parameter are given in the table. Each data set provides the cumulative number of faults by each week up to 25 weeks. The Fig. 1 represents the cumulative number of faults versus the cumulative system test hours at the end of each The Phase 2 data set is given in Table 2. We developed a Octave program to perform the analysis and all the calculations for LSE estimates. The parameter a is the number of initial faults in the software and the parameter b is related to the failure detection rate during testing process.

Table 1 : Number of Failure Observed

Days	Failure Observed	Cumulative Failure	Days	Failure Observed	CumulativeFailure
1	32	32	14	5	127
2	23	55	15	5	132
3	11	66	16	6	138
4	10	76	17	3	141
5	11	87	18	5	146
6	7	94	19	1	147
7	2	96	20	1	148
8	5	101	21	3	151
9	6	107	22	1	152
10	2	109	23	2	154
11	4	113	24	1	155
12	7	120	25	1	156
13	2	122			

The software reliability R(x/t) is defined as the probability of a failure free operations of a complete software for a specified time i.e. interval (t, t +x) in a specified environment in Equation 9. The interval methods of estimation are explained by applying the results to the software failure data .The set of software errors analyzed here is borrowed from a simulated data (an 1 days interval). where R(s|t) is reliability of component during (t, t+s) time.

$$R(xIt) = e^{-a(e^{-bt} - e^{-b(t+x)})} \quad (9)$$

Table 2 : Remaining Fault,Reliability, Failure Intensity

Day	A	B	Remaining Fault	Reliability	Failure Intensity
15	138.38	0.1333	16	80.5	2.1844
16	133.71	0.1432	12	84.66	1.6769
17	141.25	0.1274	14	83.48	1.8162
18	139.72	0.1304	12	85.91	1.5286
19	138.85	0.1322	10	87.86	1.3030
20	140.34	0.1290	9	88.71	1.2052
21	140.10	0.1295	8	90.09	1.0495
22	141.91	0.1255	8	90.60	0.9929
23	142.03	0.1252	7	91.62	0.8801
24	142.3154	0.1246	6	92.48	0.7869
25	141.13	0.1275	5	93.71	0.6538

b) *Analysis*

In fig 1 the cumulative failure observed have been shown as per number of days of testing. It is obviously seen that the number of fault observed is decreasing with days. The number of fault is initially is more but with time the number of fault is decreasing. The estimation of remaining fault decreases rapidly then it becomes straight and it is at low level in fig 2. The remaining fault is never zero as per fig 2. The reliability graph 3 shows that the growth is initially is fast but with time it is also decreasing. It reaches above of 90% after 21 days. The graph of comparision of reliability versus remaining fault describes that the as the reliability is not inversely proportional to the remaining fault as shown in fig 4. The failure intensity of component is slightly decreasing with the number of days of testing from fig 5. As per the solution of problem definition mentioned in 25 days of additional testing is required to get an benchmarks level of reliability and acceptable number of remaining faults so that the software can be released for final delivery.

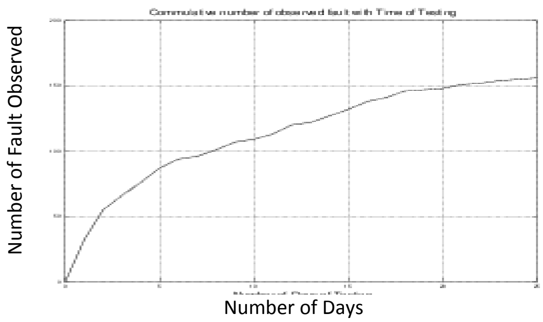


Figure 1 : Number of Fault Observed Wrt Number of Days of Testing

For example the component can be released the software if the expected reliability is greater than the threshold value i.e 90.09778 %and above 90%.

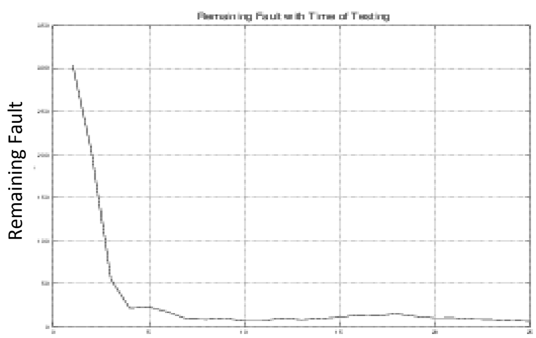


Figure 2 : Remaining Fault wrt number of Days of testing

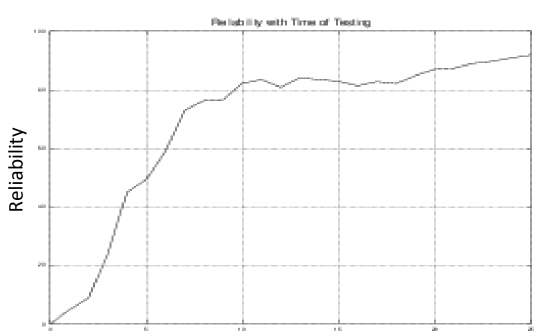


Figure 3 : Reliability wrt Number of Days of Testing

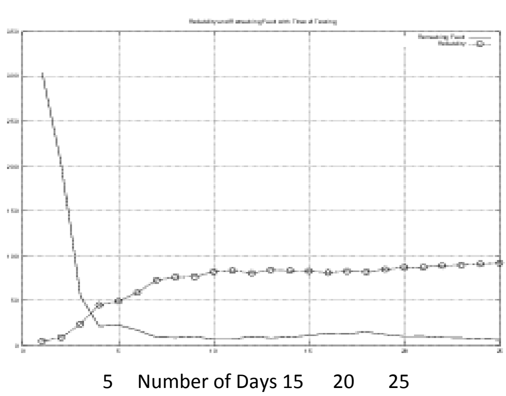


Figure 4 : Comparision of Remaining Fault and Reliability wrt Number of Days of Testing

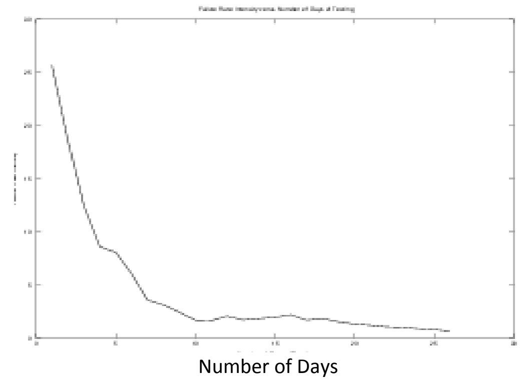


Figure 5 : Failure Intensity wrt Number of Days of Testing

V. Conclusion

This work has proposed a method of estimating the reliability of reusable architecture which can be used to build a software by using Goel-Okumoto Software Reliability Growth Model. The data available from an exponential distribution are grouped and the this model used to illustrate the parameter estimation problem. The measurement of reliability decides the quality and level of reliability decides the time of delivery of any software the reliability is increased with testing time but the reliability never becomes 100% even when the observed fault is close to zero. As per the other criteria in the above analysis, the best estimate for the remaining fault is less than 10 and then the component can be released. The integration of more reliable component can make the system more reliable. This solution will help the developer of third party component to predict the release the component with the specified marked reliability.

References Références Referencias

1. H. Pham. "System Software Reliability", Springer Series in Reliability Engineering, Springer, London, pp.149-149, 2006.
2. X. Teng, H. Pham. "A New Methodology for Predicting Soft-ware Reliability in the Random Field Environments". IEEE Transactions on Reliability, vol. 55, no. 3, pp. 458-468, 2006.
3. L. Pham, H. Pham. "Software Reliability Models with Time dependent Hazard Function Based on Bayesian Approach", IEEE Transactions on Systems, Man, and Cybernetics Part A, vol. 30, no. 1, pp. 25-35, 2000.
4. H. Pham. "Software Reliability Assessment: Imperfect Debugging and Multiple Failure Types in Software Development", EG and G-RAAM-10737, Idaho National EngineeringLaboratory, 1993.
5. H. Pham. "A Software Cost Model with Imperfect Debugging, Random Life Cycle and Penalty Cost", International Journal of Systems Science, vol. 27, no. 5, pp. 455-463, 1996.

6. A. L. Goel, K. Okumoto. “Time-dependent Fault-detection Rate Model for Software and Other Performance Measures”, .IEEE Transactions on Reliability, vol. 28, pp. 206-211, 1979.
7. S. Yamada, S. Osaki., “Software Reliability Growth Modeling: Models and Applications”, IEEE Transactions on Software Engineering, vol. 11, no. 12, pp. 1431-1437, 1985.
8. Dr. R.Satya Prasad, Bandla Srinivasa Rao, Dr. R.R. L. Kantham, “Assessing Software Reliability using Inter Failures Time Data”, International Journal of Computer Applications Volume18-No-7, March 2011
9. X. Zhang, X. Teng, H. Pham. “Considering Fault Removal Efficiency in Software Reliability Assessment”, IEEE Transactions on Systems, Man, and Cybernetics - Part A, vol.33, no. 1, pp. 114-120, 2003.
10. H. Pham, X. Zhang,” An NHPP Software Reliability Models and its Comparison”, International Journal of Reliability, Quality and Safety Engineering, vol. 4, no. 3, pp. 269-282, 1997.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 13 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Periodic Pattern Mining – Algorithms and Applications

By G.N.V.G. Sirisha, M. Shashi & G.V. Padma Raju

S.R.K.R. Engineering College, India

Abstract- Owing to a large number of applications periodic pattern mining has been extensively studied for over a decade. Periodic pattern is a pattern that repeats itself with a specific period in a give sequence. Periodic patterns can be mined from datasets like biological sequences, continuous and discrete time series data, spatiotemporal data and social networks. Periodic patterns are classified based on different criteria. Periodic patterns are categorized as frequent periodic patterns and statistically significant patterns based on the frequency of occurrence. Frequent periodic patterns are in turn classified as perfect and imperfect periodic patterns, full and partial periodic patterns, synchronous and asynchronous periodic patterns, dense periodic patterns, approximate periodic patterns. This paper presents a survey of the state of art research on periodic pattern mining algorithms and their application areas. A discussion of merits and demerits of these algorithms was given. The paper also presents a brief overview of algorithms that can be applied for specific types of datasets like spatiotemporal data and social networks.

Keywords: periodic pattern mining; time series; sequence mining; dna sequences; spatiotemporal data, approximate periodic patterns, partial periodic patterns; statistically significant patterns, recommender systems, animal migrations, gene mutations.

GJCST-C Classification : D.2.11

PERIODICPATTERNMININGALGORITHMSANDAPPLICATIONS

Strictly as per the compliance and regulations of:

Periodic Pattern Mining – Algorithms and Applications

G.N.V.G. Sirisha [α], M. Shashi [σ] & G.V. Padma Raju [ρ]

Abstract- Owing to a large number of applications periodic pattern mining has been extensively studied for over a decade. Periodic pattern is a pattern that repeats itself with a specific period in a give sequence. Periodic patterns can be mined from datasets like biological sequences, continuous and discrete time series data, spatiotemporal data and social networks. Periodic patterns are classified based on different criteria. Periodic patterns are categorized as frequent periodic patterns and statistically significant patterns based on the frequency of occurrence. Frequent periodic patterns are in turn classified as perfect and imperfect periodic patterns, full and partial periodic patterns, synchronous and asynchronous periodic patterns, dense periodic patterns, approximate periodic patterns. This paper presents a survey of the state of art research on periodic pattern mining algorithms and their application areas. A discussion of merits and demerits of these algorithms was given. The paper also presents a brief overview of algorithms that can be applied for specific types of datasets like spatiotemporal data and social networks.

Keywords: periodic pattern mining; time series; sequence mining; dna sequences; spatiotemporal data, approximate periodic patterns, partial periodic patterns; statistically significant patterns, recommender systems, animal migrations, gene mutations.

I. Introduction

Sequence is an ordered list of elements from any domain. The order among the elements of a sequence may be implied by time order as in stock market data or by physical position as in DNA or protein sequences [1]. If the order is implied by time order the sequences are called event sequences. Sequences where the order is implied by physical position are called biological sequences. Frequently occurring subsequences are referred to as sequential patterns. Sequential patterns with high support extracted from sequence databases are called frequent sequential patterns while the regularly repeating patterns found in a lengthy sequence are called periodic patterns [51]. Periodic analysis is often performed over time-series data which consists of sequences of values or events typically measured at equal time intervals [5]. Periodic patterns are classified based on different criteria. Periodic patterns are categorized as frequent periodic patterns and statistically significant patterns based on the frequency of occurrence. Frequent periodic patterns are in turn classified as perfect and imperfect periodic patterns, full and partial periodic patterns, synchronous and asynchronous periodic patterns, dense periodic patterns, approximate periodic patterns. Periodicity is of two types. The first type is fixed or known periodicity and the patterns with fixed periodicity are daily traffic patterns, daily sales patterns in super market. The second type is unpredictable periodicity and the patterns with this type of periodicity are found in datasets like power consumption, telecommunication traffic, maintenance of vehicles, web click streams, seasonal changes in climate, data sent by sensors, biological sequences. Finding periodicity should be part of mining process for these patterns. Applications of periodic patterns include prediction of drought and flood, stock market price prediction, earthquake prediction, repeat detection in DNA sequences, detection of recurrent illnesses and prediction of fraud and other outlier patterns. [10]. Generalized Sequential Patterns (GSP) [2] proposed by Srikant and Agrawal, Sequential PAttern Discovery using Equivalence Class(SPADE) [3] proposed by Zaki, Prefix and Suffix Projection (PrefixSpan) [4] proposed by Pei et al. are a few algorithms for finding frequent sequential patterns. This paper presents a survey of state of art research on periodic pattern mining.

II. Classification of Periodic Patterns

a) Full and Partial Periodic Patterns

Periodic patterns can be classified as full periodic (complete periodic) or partial periodic. Full periodic pattern is a pattern where every position in the pattern exhibits the periodicity. Periodic patterns in which one or more elements do not exhibit the periodicity are called partial periodic patterns. if {a}{b}{c}{b}{c}{b}{c}{a}{c}{d} is an input sequence {b}{c} is a full periodic pattern with period 2. It is all also called as full periodic pattern because every position in the pattern exhibits the periodicity. The sequence {a}{b}{c}{a}{d}{c}{a}{c}{c} contains a partial periodic pattern {a}{*}{c} with period 3 where the second element is not exhibiting the periodic behavior. Partial periodicity is a looser kind of periodicity than full periodicity and its application is more general because of the mixture of periodic events and non-periodic events in real world data.

Author α: Assistant Professor, Dept. of CSE, SRKR Engineering College, Bhimavaram, A.P, India. e-mail: sirishagadiraju@gmail.com
Author σ: Professor, Andhra University College of Engineering, Andhra University, A.P. ,India. e-mail: smogalla2000@yahoo.com
Author ρ: Professor & HOD, SRKR Engineering College, Bhimavaram, A.P., India. e-mail: gvpadmaraju@gmail.com

b) Perfect and Imperfect Periodic Patterns

Periodic patterns can also be classified as perfect and imperfect periodic patterns. A pattern X is said to satisfy perfect periodicity in sequence S with period p if starting from the first occurrence of X until the end of S every next occurrence of X exists p positions away from the current occurrence of X.{a}{b}{*} is perfect periodic pattern with period 3 in the sequence {a}{b}{d}{a}{b}{v}{a}{b}{f}{a}{b}{c}. (a}{b}{*} has occurred for 4 times starting from its first occurrence till the end of the sequence. It is possible to have some of the expected occurrences of X missing and this leads to imperfect periodicity. {a}{b}{*} is an imperfect periodic pattern in the sequence {a}{b}{c}{d}{b}{f}{a}{b}{g}{a}{b}{v} and hence the confidence on its periodicity is 3/4. The occurrences of {a}{b}{*} is missed in one of its expected positions. Perfect periodic patterns are first proposed by Ozden et al [6]. They developed an algorithm to find cyclic association rules that reoccur in every cycle for the entire time series. This algorithm finds cyclic association rules from time stamped transactional data. An association rule is called cyclic with cycle $c=(l,o)$ if the association rule holds in every l^{th} time unit starting with time unit t_o Here t_i corresponds to the time interval $[i.t,(i+1).t]$ where ***t is*** the time unit referring to time granularity given as input by the user. For example if the time unit is an hour and if a person buys milk and newspaper daily in the interval 7A.M-8A.M., then newspaper->milk or milk->newspaper has the cycle $c=(24,7)$ (e. starting at 7A.M every 24 hours we can observe the purchase of milk and newspaper together). Given a fixed length period as input Han et al. mined segment-wise periodic patterns by using data cube, bit-array and apriori mining techniques [17]. Segment-wise periodic patterns are full or partial periodic patterns whose number of occurrences in a time series database satisfy the minimum confidence threshold. Cyclic association rules repeat perfectly in the time series with 100% confidence.

Later on Han el al. developed an algorithm called max-subpattern hit set [13] which created a maxsubpattern tree to mine full imperfect periodic patterns and partial imperfect periodic patterns in two scans of the time series. The algorithm can finds the periodic patterns occurring with a given period in two scans of the sequence. It is also extended to detect the patterns for a set of periods (period range is given as input). For a given period the time series is segmented in to different segments called period segments where the length of each period segment is equal to the given period. The time series is scanned once and all *1-patterns* which are found to be frequent in the segmented time series are reported. A *1- pattern* is a pattern where one position in the pattern is defined, for example $(a,*,*)$ or $(*,a,*)$ or $(*,*,a)$. These frequent *1- patterns* are unioned to form max pattern. Now the time series is scanned for the second time. During the second scan of the time series each period segment is intersected with the maxpattern. The result of the intersection called max sub-pattern is either inserted into the tree if it is not already present or its corresponding node count is incremented if it is already present in the tree. Finally the max subpattern tree is then traversed and all patterns whose count is greater than *min_conf* are output as frequent periodic patterns. Max sub-pattern hit set algorithm for multiple periods works in a similar way where max sub pattern trees for all periods are constructed parallely with two scans of sequence. If the number of *1-patterns* mined in the first scan of the sequence is very large, the resultant max-pattern constructed is also very complicated. This results in a very large max-subpattern tree and a lot of time is wasted traversing the tree for finding the frequent periodic patterns.

Aref et al. extended the max-subpattern hit set by introducing algorithms for incremental, on-line and merge mining of partial periodic patterns [7]. Incremental version of max subpattern hit set algorithm allows the users to mine partial periodic patterns in case of insertion and deletion updates of time series database. It requires at most one additional scan of old database. Online version of the algorithm allows users to modify the thresholds used in mining algorithm while the algorithm execution is in progress. Merge mining aims at merging the patterns that are mined from two or more databases. This merging process helps us in discovering the patterns from combined database without having to reapply the mining algorithm.

S-S.Chen et al. proposed an algorithm [26] to find periodic patterns of given periodicity using the encoded period segments database and Frequent Pattern tree. As in max-subpattern hit set algorithm, for a given period *'p'* the given long sequence is segmented into segments where the length of each segment is *'p'*. For example if the given sequence $S_D=\langle a\{b,c\}\{b,e\}aebacb\rangle$, and period p = 3, S_D is segmented into 3 segments. The first period segment $S_1=\langle a\{b,c\}\{b,e\}\rangle$, the second period segment $S_2=\langle aeb\rangle$ and the third period segment $S_3=\langle acb\rangle$. In S_1 *a* is at position 0, ***b and c*** are at position 1 and b and e are at position 2. So, S_1 can be encoded as $\langle(a0),(b1),(c1),(b2),(e2)\rangle$. Similarly S_2 $S_2=\langle aeb\rangle$ is encoded as $\langle(a0),(e1),(b2)\rangle$ and $S_3=\langle ace\rangle$ can be encoded as $\langle(a0),(c1),(e2)\rangle$. After encoding all the period segments in this way, S_1,S_2,S_3 can be treated as separate sequences of a sequence database and frequent pattern tree with multiple minimum supports is used to find all the frequent periodic patterns. The main advantage of this algorithm is in its encoding step. After encoding the period segments in the way illustrated above, any frequent pattern mining algorithm can be used for mining the frequent periodic patterns. Instead of using single minimum support to determine the significance of a pattern, the user can specify different minimum supports for different events based on their real life occurrence frequency. This allows the algorithm to find the rare but important patterns from the data. Recently K.J. Yang et al. proposed projection based partial periodic pattern mining [9] that can efficiently mine all partial periodic patterns that occur with the given period. On giving a period value this algorithm divides the given sequence into segments whose length is equal to period. It then encodes these segments into event tuples. This encoding is similar to the encoding technique discussed above[26]. The usage of event tuples together with recursive projection based algorithm speeds up the mining of partial periodic patterns. Only one minimum support is used to find the significance of the pattern.

All the above algorithms can be applied on datasets with predictable periodicity. However as the algorithms divide a long sequence into equal size segments where the size is equal to given period, the above algorithms can only detect patterns occurring within a segment possibly missing the patterns that overlap across the segments.

c) Perfect and Imperfect Periodic Patterns

All the above periodic pattern mining algorithms require the user to give period as input. These algorithms are appropriate for all the applications where the data consists of natural periods like hour, day, week, month, quarter and year. Some data sets may have patterns that repeat with unexpected periods. In such cases we need algorithms that can extract potential periods and the patterns that appear with these periods.

Pioneering work in this direction is done by Ma el al. to find all partial periodic patterns with unknown periods [28]. Two algorithms were developed to find partially periodic patterns. The former also called period first finds potential periods first and then for each period discovers the associations between events occurring with that period. The potential periods are detected using chi-squared test. A level-wise algorithm is then used to mine partial periodic patterns that appear with these periods. The later algorithm also called association first finds associations between events first. For each temporal association pattern it then discovers its periodicity. These algorithms can find partial periodic patterns even in the presence of replacement, insertion and deletion noise. One of the drawbacks of this work is that some valid periods may not be identified as it considers the difference between adjacent time instants only as potential periods.

People belonging to various areas such as economics, digital signal processing and statistics have also proposed various methods for periodicity search in time series. These methods include the usage of autocorrelation function, FFT, DWT for finding the periodicity of the time series. They focused on finding the periodicity only and did not explore mining of patterns that occur with the detected periodicity. Unlike these methods Berberidis et al. [19] have used autocorrelation to detect all candidate periods from a discretized time series and applied max subpattern hit set algorithm to find the patterns. This algorithm scans the time series once to create a binary vector of size *n* for every symbol in the alphabet of the time series. For each symbol in the alphabet it computes the circular autocorrelation vector that contains the frequency count of each period. Among the periods from 1 to *n/2*, periods whose occurrence frequency is less than minimum confidence threshold are filtered out and the rest are kept as candidate periods. For each candidate period all the symbols occurring significantly with that period are unioned to form max pattern and Han's max subpattern hit set algorithm is applied to mine all frequent partial periodic patterns that occur with that period. The drawback of this method is that for any period, circular autocorrelation produces unexpected frequency values when the length of time series is not an integral multiple of that period. It also produces wrong frequency values for periods when successive occurrences of a symbol are repeated frequently periodically. The overall complexity of the algorithm using FFT as a filter is $O(AN\log N)$ where is A is the size of the alphabet and N is size of the time series.

Usage of different networking resources has lead to the generation of different time series data in telecommunications and network applications. These include total number and duration of calls, number of bytes or electronic mails sent out from one ISP to another, amount of web traffic at a site etc [18]. Sale of a specific item over time also represents time series data. Time series are approximately periodic (noisy signals). Application of FFT on such data involves filtering of data before processing to guarantee accuracy. Even with very good filtering techniques and with no phase shifts in data, FFT cannot find all periodicities that are present in the data. Inorder to identify periodicities in noisy time series, Indyk et al. have proposed an algorithm named representative trends [18] to detect the candidate periods called relaxed periods in $O(n\log^2 n)$ time. This algorithm identifies periodicities but not the patterns that repeat with those periodicities so we need to use a separate periodic pattern mining algorithm that uses the detected candidate periods as input to generate the periodic patterns. Convolution based periodicity detection algorithm [15], WARP [16] and STAGGER [22] are other algorithms that can detect potential periods and periodic patterns in one pass of the time series. P. Huang el al. used an algorithm based on frequent partial periodic pattern tree for wireless spectrum occupancy prediction [48]. This prediction helps unlicensed users to use licensed wireless spectrum bands that are under-utilized i.e. the unlicensed users can use the available slots that are left unused by licensed users. This in turn helps in improving the channel utilization and reduce collision rate.

Dutta et al. have proposed algorithms to find perfect and imperfect calendar based periodic patterns [30][31][32]. These algorithms are used to mine yearly, monthly, daily, hourly periodicities from discrete as well as continuous time series datasets. They find calendar based periodicities of an interval based temporal pattern. An interval based temporal pattern is a pattern that occurs across a sequence of time-intervals in either a discrete or continuous domain. These algorithms have a time complexity $O(n \log n)$ for a continuous domain and only $O(n)$ for a discrete domain.

G. Lee at al. applied parallel computing to efficiently mine frequent partial periodic patterns at all valid periods [34]. They are able to speed up the process of mining partial periodic patterns by partitioning the periods into independent subsets based on independence property of prime numbers. Each independent subset of periods will be assigned to a processor for detection of periodic patterns. Generation of redundant patterns was avoided thus increasing the efficiency of mining process as well as effectiveness and interpretability of generated patterns. A periodic pattern is redundant if the knowledge provided by it is obtained from other periodic patterns. For example in the following patterns 'Stock A's price increases every 2 days', 'Stock A's price increases every 4 days', the second pattern is redundant. Partial periodic correlation is a set of offsets within a particular period such that the data at these offsets are correlated with certain user desired strength. These partial periodic correlations are identified in [35] using principal component analysis method.

d) Synchronous and Asynchronous Periodic Patterns

A pattern which occurs periodically without any misalignment is called as synchronous periodic pattern. In the sequence {a}{b}{c}{a}{d}{c}{a}{c}{c}{d}{e}{f}, {a}{*}{c},{a}{*}{*},{*}{*}{c} are the synchronous partial periodic pattern. These patterns have repeated for three times consecutively in the sequence with a period 3. Asynchronous periodic patterns are patterns with some disturbance between the repetitions of the pattern. Disturbance is allowed not only in terms of missing occurrences but also as insertion of random noise events. {a}{*}{c}, {a}{*}{*},{*}{*}{c} are asynchronous periodic patterns in the sequence {a}{b}{c}{a}{c}{c}{c}{b}{a}{b}{c}{a}{d}{c}. The above patterns have appeared for four times in the sequence where there is a disturbance between second and third occurrences of the patterns. A number of algorithms like LSI [8], SMCA [10], OEOP [11] and E-MAP [12] were developed to find these asynchronous periodic patterns in a single long sequence. Longest Subsequence Identification (LSI) is the pioneering algorithm to mine asynchronous periodic patterns. For each asynchronous periodic pattern the algorithm detects the longest subsequence containing it. LSI is an iterative level wise search algorithm. It works in three phases. In the first phase it detects all potential periods for all events. This requires single scan of the sequence. In the second phase all candidate *1-patterns* are validated. A *1-pattern* is a pattern where one position is the pattern is defined and rest of them are *'s. For example {a}{*}{*},{b}{*} are 1-patterns with periodicity 3 and 2 respectively. An i-pattern is a pattern where I positions in the pattern are defined for e.g. {a}{b}{*} is a 2-pattern with periodicity 3 where 2 positions out of 3 are defined. In the third phase an iterative level wise approach is used where in the *i th* iteration candidate *i-patterns* are formed from *(i-1)-patterns*. Validation of these i-patterns require single scan of the sequence. The second and third phases of LSI require multiple scans of the sequence. Simple Multiple Complex and Asynchronous periodic pattern miner (SMCA) is a four phase algorithm that detects all the subsequences containing asynchronous periodic patterns. It can work with sequence of event sets also. A sequence of event sets is a sequence where each time instant contains one or more events. The first algorithm SPMiner finds *1-patterns.* The second algorithm MPMiner finds *1-patterns* containing simultaneously occurring events. The third algorithm CPMiner finds complex patterns(*ipattern for* $i \geq 2$). For all the *1-patterns, 1-patterns* with simultaneously occurring events and complex patterns the fourth algorithm APMiner finds subsequences where their appearance is significant. A pattern is said to appear significantly if the number of occurrences of the pattern in a subsequence is greater than a threshold *min_rep.* One Event One Pattern (OEOP) algorithm uses a linked list structure to detect single event one patterns in a single scan of a sequence. OEOP can be used to replace the first phase of SMCA for data sets like data streams. It can mine all *1- patterns* of all the events at all periods in a single scan of the sequence. E-MAP was proposed as a further improvement over SMCA model to mine all patterns (*1-patterns, 1-patterns* with simultaneously occurring events, complex patterns) in a single step and single scan of the sequence. Among these algorithms LSI can only handle sequence of events where as the other algorithms can handle sequence of event sets. All these algorithms can handle replacement, insertion and deletion noise that is present in the sequence. These algorithms can detect asynchronous periodic patterns that appear with periods in the range max $2 \leq p \leq L_{max}$. L_{max} is the maximum periodicity that the user wants to check in the sequence. These algorithms produce many redundant patterns and they can be extended to mine only closed and maximal periodic patterns. Inorder to capture the hierarchical nature of asynchronous periodic patterns a meta-pattern model is proposed by Yang et al. [27].

e) Patterns with Symbol, Sequence and Segment Periodicity

Some authors have classified periodic patterns as symbol periodicity or singular patterns, sequence periodicity or partial periodic patterns and segment or full-cycle periodic patterns. A sequence is said to have symbol periodicity if at least one symbol is repeated periodically. For example in the sequence {x}{y}{y}{x}{u}{i}{x}{t}{r}{x}{k}{l}, symbol {x} is periodic with periodicity 3. A pattern consisting of more than one symbol repeating with same periodicity in a sequence leads to sequence periodicity. {x}{y} (*) is partial periodic pattern in {x}{y}{z}{y}{z}{z}{x}{y}{y}{x}{y}{x} If the whole sequence can be mostly represented as a repetition of a pattern or segment then that type of periodicity is called segment or full-cycle periodicity. For example {x}{y}{z} is a full-cycle periodic pattern in the sequence {x}{y}{z}{x}{y}{z}{x}{x}{x}{x}{y}{z}. The periodicity and confidence of the pattern {x}{y}{z} are 3 and ¾ respectively.

Rasheed et al. proposed suffix tree based noise resilient algorithm (STNR) [14] to detect symbol, sequence and segment periodicities that occur in the entire sequence or a subsection of it. Suffix tree is a data structure used in string processing. Suffix trees can be used to solve exact string matching problem, the substring problem, longest common substring of two strings etc. A suffix tree represents all the suffixes of the string for which it is constructed. For each suffix of the string there is a distinguished path from the root to corresponding leaf node in the suffix tree. Given a long sequence of symbols, a suffix tree for the sequence helps us in finding the number of repetitions of all substrings (or subsequences) and the occurrence positions of these subsequences in the sequence. For detecting periodicity, periodicity detection algorithm is applied on the occurrence vector of each subsequence (substring) whose occurrence frequency satisfies the *min_conf* threshold. The major limitation of this algorithm is that it cannot detect patterns like {a}{*}{b}, {a}{b}{*}{c}{d} where the elements represented by {*} do not exhibit the periodicity. It also misses some valid periods because it only considers adjacent time instants in calculating the periods. For example if 0, 3, 5, 6, 10, 13, 15 are time instants at which an event occurs. STNR algorithm considers the difference between every pair of consecutive time instants as candidate period. In the above example 3,2,1,4,3,2 are considered as candidate periods. The period 5 which is present in the data will be unidentified. This algorithm being noise resilient can work with sequences having replacement, insertion and deletion noise. While some of the existing algorithms [15][16][17][18] can only detect periods that span through the entire sequence, STNR can detect periodic patterns that span through the entire sequence as well as a subsection of it. This algorithm cannot be used on eventset sequences. The worst case time complexity of the algorithm is $O(k.n^2)$.

M.G. Elfeky et al. have developed a convolution based algorithm to detect symbol and segment periodic patterns [15]. The algorithm scans the time series once to convert it into binary vector according to proposed mapping. It applies modified convolution on the binary vector to find the underlying periodicities and corresponding symbol and segment periodic patterns in single pass of the time series. To reduce the time complexity, convolution is computed using FFT. The time complexity of the algorithm $O(n \log n)$. It works well in mining synchronous periodic patterns. It fails in the presence of insertion and deletion noise.

Noisy data exists in almost all real world databases due to different reasons like errors in data entry, data transmission errors. In streaming applications data elements may be dropped purposefully for processing reasons. We need noise resilient algorithms in such cases. Elfeky at al. developed an algorithm named WARP [16] that uses time warping to detect the underlying periodicities and periodic patterns. It uses time warping to extend or shrink time axis at various locations to optimally remove noise. WARP is tolerant to insertion and deletion noise. The time complexity of WARP is $O(n^2)$. Though it is tolerant to noise, it can only detect segment periodicities and cannot find symbol or sequence periodic patterns. An online version of WARP called Online WARP was also developed to work with datastreams.

For mining periodic patterns from data streams Elfeky el al. have proposed a one-pass, online and incremental algorithm named STAGGER [22]. STAGGER discovers potential periodicities using multiple expanding sliding windows. With the continuous flow of the streamed data the sliding windows expand in length in order to cover the whole stream. Short length windows help us in identifying the short periods early and in real time. Larger length windows help in identifying the longer periodicities. For each discovered periodicity, STAGGER incrementally maintains a max-subpattern tree to find the corresponding periodic patterns.

f) Dense Periodic Patterns

Most of the periodic pattern mining algorithms can mine the periodic patterns only if they appear in the entire time series. A pattern which occurs in small segments of time series will not be detected by such algorithms because of lack of sufficient support or confidence. STNR can detect periodic patterns that appear in the entire time series or a part of it, but the end user has to give the range i.e. start and end positions of the segment of the time series from which he wishes to mine the periodic patterns. This is a difficult task. Sheng el al. has proposed an algorithm [25] to mine dense periodic patterns from time series database. For each unique symbol the algorithm finds dense fragments from the time series. A dense fragment is a segment of time series where the distance between every pair of consecutive occurrences of the symbol is less than the distance threshold d_{max}. All the dense fragments of all the symbols whose lengths are greater than *min_len* are retained. Then a max-subpattern tree is used to generate the partial periodic patterns using only the dense fragments of the symbols of the alphabet. This algorithm also prunes some periods by finding the lower bound period for each symbol in its each dense fragment. All the periods that are less than lower bound period can be safely pruned.

III. Approximate Periodic Patterns

Noise and imprecision exists in most of the real world data e.g. gene mutations in DNA sequence, so there is a need to mine approximate patterns. The algorithms that we have discussed previously which tolerate replacement, insertion and deletion noise mine approximately repeating periodic patterns. Here approximation is allowed in occurrence position of the patterns. In this section we are going to discuss a few more algorithms that allow approximation in occurrence positions as well as structure of the patterns. Y.L. Zhu el al. has mined approximate periodic patterns from hydrological time series [23]. Multi-event asynchronous periodic patterns were discovered from long sequence of hydrological time series. A modified suffix tree representation and traversal together with dynamic candidate period intervals adjusting method was used to mine all patterns without omitting any periods. Another approach to mine approximate periodic patterns was proposed by Amir et al. [24]. Given an input sequence and a parameter $\varepsilon \in [0,1)$ this algorithm finds for each unique event the longest ε - relative error periodic pattern i.e. the longest subsequence $s_1, s_2, s_3, \ldots, s_k$ of S for which there exists a number *p* such that the absolute difference between any two consecutive occurrences of event in the subsequence is atleast *p* and atmost $p(1+\varepsilon)$. *P* is the approximate period of the relative error periodic pattern. Zhang et el. have proposed an algorithm to find periodic patterns with gap requirement from DNA sequences [33]. The algorithm generates all periodic patterns of the form $a_1 g(N,M) a_2 g(N,M) \ldots a_{l-1} g(N,M) a_l$ where a_i for $1 \le i \le l$ is any symbol from the sequence, $g(N,M)$ is the gap between symbols and N,M are two user supplied numbers that specify the minimum and maximum gap sizes between every pair of consecutive symbols in the pattern.

IV. Surprising Periodic Patterns

In some applications like genomic data, system traces, credit card transactions and seismic data, patterns that occur less frequently yet whose occurrence frequency is beyond prior expectation are considered interesting. But all the periodic pattern mining algorithms discussed so far prefer mining periodic patterns that occur frequently. J. Yang el al. had proposed an algorithm named InfoMiner [29] to mine rare but statistically significant periodic patterns. A pattern is considered significant if its occurrence frequency exceeds its expected frequency by a large margin. Inorder to identify rare and statistically significant patterns a new model called as information model is used instead of support model. Information gain measure is used to identify the degree of surprise of pattern in a data sequence. Information gain of a pattern $G(p)$ is given $G(p) = I(p) * (Support(p) - 1)$. $I(p)$. s the information carried with *p* or surprise of *p*. m is less for a frequent pattern than that of a rare event. In the context of mining surprising periodic patterns a pattern that never repeats is of no use. So *I(p)* is multiplied by $Support(p) - 1$ to identify the true periodic patterns with statistical significance. To improve the efficiency of the mining process, pattern pruning based on bounded information gain is used.

The occurrence positions of a pattern are not taken into account by Infominer. For example in $S_1 = a_1, a_2, a_3, a_3, a_2, a_3, a_1, a_2, a_4, a_5, a_1, a_2$ and $S_2 = a_1, a_2, a_1, a_2, a_1, a_2, a_5, a_3, a_2, a_3, a_4, a_5$ pattern (a_1, a_2) has the same information gain. In the first sequence (a_1, a_2) is scattered where as in the second sequence it repeats perfectly in first half of . In some application like repeat discovery in bio-informatics, consecutive repetitions of a pattern are more significant than the scattered repetitions [20]. Some penalty is associated with the gap between pattern repeats. So we need a new measure that distinguishes between consecutive repeats and scattered repeats of a pattern. J. Yang el al. proposed an algorithm called InfoMiner+ [20] which makes this distinction. It uses generalized information gain (GIG) measure. It detects all partial periodic patterns that are significant in one or more subsequences of an event sequence. For each pattern the subsequence that maximizes the pattern's generalized information gain is reported. It can deal with replacement noise. This algorithm detects all statistically significant periodic patterns whose periodicity is less than or equal to the maximum period bound L_{max}.

Periodic patterns not only help in predicting future events but also help in identifying anomalies in data. F. Rasheed et al. have developed a framework for periodic outlier pattern detection in time series sequences [44]. Frequent periodic patterns are the periodic patterns that satisfy the user specified *min_sup* or *min_conf* . Though *min_sup* and *min_conf* measures can help us in identifying frequent periodic patterns they cannot be used to find outlier periodic patterns. The outlier periodic patterns repeat less often, the repetitions may not be strictly periodic, and the period is generally larger than the regular frequent periodic patterns. Rasheed et al. proposed a new measure based on the relative frequency of the pattern and its coverage area to estimate the surprise of the pattern. Then they used STNR [14] to find the periodicity of these surprising or outlier periodic patterns. As this framework uses STNR in periodicity mining, it can work with noisy data where periodic repetitions are not strict.

If $f(x)$ represents the number of repetitions of the pattern *'X'*, and $segLen(X)$ represents the segment length of the repetitions of *'X'*, then X is candidate outlier pattern if $f(x) < \mu(f(x_i))\ \mathrm{AND}\ segLen(x) > \min SegLen$; $\forall \mathrm{i}$ such that $|x_i| = |x|$ where $\mu(f(x_i))$ is mean of frequency of all patterns of length exactly the same as that of pattern *X*. The measure of surprise of a pattern *X* is defined as one minus the ratio of the frequency of *X* over the average frequency of all patterns with same length as *X*. $Surprise(X) = 1 - \frac{f(x)}{\mu(f(x_i))}$; $\forall \mathrm{i}$ such that $|x_i| = |x|$.

A candidate outlier pattern *'X'* is an outlier periodic pattern *iff* $surprise(X) > surprise_{\min}$ and $Conf(X, i_{st}, i_{end}, p) > Conf_{\min}$ where $Conf(X, i_{st}, i_{end}, p)$ is confidence of pattern *'X'* repeating with period *p* within the segment starting and ending at i_{st} and i_{end} respectively; $Conf_{\min}$ and $Surprise_{\min}$ are respectively the minimum confidence and minimum surprise values provided by the user. STNR-Out mines small number of non redundant outlier periodic patterns.

V. Mining Periodic Patterns from Spatiotemporal Databases and Social Networks

Periodicity is a commonly seen phenomenon in moving objects as well as social networks. People usually go to workplace every day through more or less same route, birds and animals migration from one place to another also show yearly periodicity. Interactions between people and topic discussion in social media also show periodicity, for example delivering news about different budgets in news sites and posting of information corresponding to different events like annual conferences in blogs and websites are done periodically.

The periodic pattern mining algorithms designed for event/symbol sequences cannot be used to find spatiotemporal periodic patterns. This is because the repetitions of spatial locations $\langle x, y \rangle$ may not be identical. Even if objects follow same route regularly, they may not appear at exactly the same location regularly. For example a person may go from home to office in the same route every day between 9.00 A.M. to 10.00 A.M. But it is less likely that he will be at the same location on his route everyday at 9.30 A.M. [37].

Cao el al. have proposed an algorithm to find maximal periodic patterns from spatiotemporal data [37]. This algorithm mines periodic *1-patterns* using spatial clustering. Then bottom-up and top-down mining techniques are used for generating longer patterns. They have also proposed algorithms which can mine periodic patterns that appear in only a time interval instead of whole time span. Another algorithm which can mine shifted or distorted instances of patterns is also proposed.

Algorithm for mining and indexing of spatiotemporal periodic patterns from historical spatiotemporal data was developed by N. Mamoulis el al [42]. Z. Li el al. define a periodic behavior as repeating activities at certain locations with regular time intervals. A two stage algorithm called Periodica [38] is used to detect periods and find the periodic behaviors. In the first stage reference spots which are dense regions that are frequently visited in the movement are identified. Then the periods associated with each reference spot are identified using Fourier transform and autocorrelation [45]. Since every period is associated with a reference spot, if we find the periods associated with each reference spot we can guarantee that all the periods in the movement are detected. A period may be associated with two or more reference spots. In the second stage all the reference spots associated with a period are considered together for mining the periodic behaviors. A periodic behavior is a statistical description of the periodic movement for a period. It is a probability distribution matrix which gives the probability that the object is at each reference spot at different sub intervals of a period. For example when we mine periodicities of a school student we may observe 24 hours period at three reference spots namely class room and play ground and house at different hours of the day. The periodic behavior for the student gives the probability with which the student is at different reference locations namely class room, play ground and house in each hour of the day. In [46] the authors have extended this work by explaining how periodic behaviors can be used for missing data interpolation and future movement prediction.

Z. Li et al, have developed probabilistic model to mine periodic patterns from noisy and incomplete observations [36]. Mobile, GPS and Sensor technologies help us to track human and animal movements. But the data collected by these devices have a large portion of missing and noisy data. The data is also unevenly sampled and the rate of sampling in some cases like data sent by sensors attached to animal is very low. Fourier transforms and autocorrelation generally require evenly sampled data collected at high sampling rate. So these methods fail in this case. So a new probabilistic model is proposed to mine periodic patterns in these types of datasets.

T. Jindal el al. have proposed a method to mine spatiotemporal periodic patterns from traffic data [41]. These patterns can be used to summarize the speed distribution of traffic along any road like inter-state highways, state highways or a local road.

L. Riotte-Lambert el al. have used Fourier and Wavelet analysis to extract periodic patterns from time series of presence/absence, arrival/departure from areas of interest [47]. They have also introduced reliable null models for assessing the statistical significance of the periods detected. This work is an extension of Z. Li el al.'s work [38]. They declare that some of the valid periods may be missed or false periods may be detected if we use a single threshold to find significant frequencies from a periodogram as in [38]. To avoid this, different thresholds were used for different frequencies when finding significant frequencies from a periodogram. Wavelet analysis makes it possible to determine both the periodicities and time intervals during which the different periods are found to appear in the data. Though it is theoretically possible to use only wavelet analysis for periodicity detection the authors suggest to first perform fourier analysis as it is very fast and convenient way to determine the frequencies that are potentially relevant and then apply wavelet analysis. This two stage approach helps in easy interpretation of wavelet power spectrum. Wavelet analysis helps in identifying periodic events that occur in whole time series or only a part of it.

M. Lahari el al. discussed periodic subgraph mining for dynamic social networks [43] and applied it on four real world dynamic networks like Enron e-mails, Plains Zebra. K. Ishida had presented an algorithm to find periodically discussed topics in blogs, news sites and spam [39].

M. Lv et al. have used user's long term activity regularity to find mobile user similarity for location based social network services [50]. A new branch of social network services called location based social network services (try to understand users activities and preferences) recommends places, activities, friends and other geo-related information to the users based on the knowledge of users activities and preferences. Finding user similarity based on their long-term activities is important to achieve this. In [49] the authors have proposed a method to find users similarity based on user's long term activity regularity using GPS trajectory data. So by mining periodic patterns we will be able to develop recommender systems.

S. Parthasarathy el al. have presented robust periodicity detection algorithms to mine periodicities from continuous time series datasets [50]. These algorithms combine short time fourier transform and autocorrelation analysis for finding the periodicity. Various algorithms were developed to find periodicity in stationary, non-stationary, noisy and incomplete datasets. These can also find multiple interleaved periods and discover hidden relationships among attributes in multidimensional time series.

VI. Periodic Patterns Versus Frequent Episodes and Frequent Continuities

Frequent episodes and Frequent Continuities are two classes of patterns that are closely related to periodic patterns. K. Huang el al. has given the following similarities and dissimilarities between them [40]. An episode is defined to as total or partially ordered collection of events in a specific time window. An episode considers only order among events instead of the actual positions of events in a time window bound. Mannila et al. presented a framework for discovering frequent episodes from a single long sequence through a level-wise algorithm WINEPI [21].

Continuity as defined in [40] is a kind of causal relationship which describes a definite temporal factor with exact position between the records. It is similar to a periodic pattern, but without the constraint on the contiguous and disjoint matches. Frequent episodes are a loose kind of frequent continuities since they consider only the partial order between events, while periodic patterns are a strict kind of frequent continuities with constraints on the subsequent matches.

VII. Conclusions and Future Work

Periodic pattern is a pattern that repeats itself with a specific period in a sequence. Periodic patterns can be mined from datasets like biological sequences, continuous and discrete time series data, spatiotemporal data and social networks. Different criteria are used to classify the periodic patterns. This paper presents an overview of different types of periodic patterns and their applications along with a discussion of the algorithms that are used to mine these patterns. We have also discussed about efficiency, user interaction needed and noise resilient nature of these algorithms.

Ideas for future research directions in this area are the following. Major limitation of all the frequent pattern mining algorithms is that they mine a large number of patterns when *min_sup* is set very low. Like any frequent pattern mining algorithm, frequent periodic pattern mining algorithms also generate a large number of periodic pattern when *min_sup* or *min_conf* is set low. In order to reduce the redundancy of the generated output and to improve the efficiency of mining process there is a need for the development of algorithms that mine closed and maximal periodic patterns. Approximate periodic pattern mining is a new sub area of periodic pattern mining and a few algorithms were developed recently to mine such patterns. It needs to be extended to eventset (multidimensional) sequences.

Many of the existing periodic patterns mining algorithms mine patterns from *1* dimension sequences (event sequences). These algorithms were not applicable on multi dimensional sequences. These have to be extended for multi dimensional sequences so that more hidden patterns can be found from such sequences also.

With the growing size of the datasets, development of incremental and distributed periodic pattern mining algorithms has become a necessity. Usage of periodic patterns in constructing classification/prediction and recommender systems should be further explored.

References Références Referencias

1. G. Dong, J. Pei, *Sequence Data Mining, Advances in Database Systems,* Springer science, ISBN-13: 978-0-387-69936-3, 2007
2. R. Srikant, R. Agarwal, "Mining sequential patterns: Generalizations and Performance Improvements", In Proc. 5th Int. Conf. Extending Database Technology", pp. 3-17, Avignon, France, 1996.
3. M. J. Zaki "SPADE: An Efficient Algorithm for Mining Frequent Sequences", Journal of Machine Learning, Vol. 42, Issue 1-2, pp. 31-60, 2001.
4. J. Pei, J. Han, B.Mortazavi-Asl, J. Wang, H.Pinto, Q. Chen, U. Dayal and M.-C.Hsu, "Mining sequential patterns by pattern growth: The prefixspan approach", IEEE Trans. Knowledge and Data Engineering, pp. 1424-1440, 2004.
5. J. Han, M. Kamber, "Data Mining: Concepts and Techniques", 2nd ed., Morgan Kaufmann, 2006.
6. B. Ozden, S. Ramaswamy, A. Silberschatz, "Cyclic association rules", In Proc. 14th Intl. Conf. Data Engineering.
7. W. G. Aref, M. G. Elfeky, A.K. Elmagarmid, "Incremental, online and merge mining of partial periodic patterns in time series databases", IEEE Trans. Knowledge and Data Engineering, 16(3), pp. 332-342, 2004.
8. J. Yang, W. Wang, and P. S. Yu, "Mining Asynchronous Periodic Patterns in Time Series Data", *IEEE Trans. on Knowledge and Data Engineering,* Vol. 15, Issue 3, pp. 613-628, 2003
9. K. J. Yang, T. P. Hond, Y.M. Chen, G. C. Lan, "Projection-based partial periodic pattern mining for event sequences", Expert Systems with Applications, 40, pp. 4232-4240, Elsevier., 2013
10. K. Y. Huang, C.H. Chang, "SMCA: A General Model for Mining Asynchronous Periodic Patterns in Temporal Databases", *IEEE Trans. On Knowledge and Data Engineering,* Vol. 17, Issue 6, pp. 774-785, Jun, 2005.
11. J.S. Yeh, S.C. Lin, "A New Data Structure for Asynchronous Periodic Pattern Mining", *Proc. 3rd Int'l Conf. Ubiquitous Information Management and Communication,* pp. 426-431, 2009.
12. F. Maqbool, S. Bashir, and A.R. Baig, "E-MAP: Efficiently Mining Asynchronous Periodic Patterns", *Int'l Journal of Computer Science and Network Security,* Vol. 6, no. 8A, Aug, 2006.
13. J. Han, W. Gong, Y. Yin, "Efficient mining of partial periodic patterns in time series databases", *In Proc. 15th Intl. Conf. Data Engineering,* pp. 106-115.
14. F.Rasheed, M. Alshalalfa, R. Alhaji, "Efficient Periodicity Mining in Time Series Databases Using Suffix Trees", *IEEE Trans. On Knowledge and Data Engineering,* Vol.23, Issue 1, pp. 79-94, Jan, 2011.
15. M.G. Elfeky, W.G. Aref and A.K. Elmagarmid, "Periodicity Detection in Time Series Databases", IEEE Trans. Knowledge and Data Engineering, Vol. 17, no. 7, pp. 875-887, July 2005.
16. M.G. Elfeky, W.G. Aref, A.K. Elmagarmid, "WARP: Time Warping for Periodicity Detection," Proc. 5th IEEE Int'l. Conf. Data mining, Nov. 2005.
17. J. Han, W. Gong, Y. Yin, "Mining Segment-Wise Periodic Patterns in Time Related Databases," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining, pp. 214-218, 1998.
18. P. Indyk, N. Koudas, S. Muthukrishnan, "Identifying representative trends in massive time series datasets using sketches," Proc. Int'l Conf. Very Large Data Bases, Sept. 2000.
19. C. Berberidis, W. Aref, M. Atallah, I. Vlahavas, A. Elmagarmid, "Multiple and Partial Periodicity Mining in Time Series Databases," Proc. European Conf, Artificial Intelligence, July 2002.
20. J. Yang, W. Wang, P. Yu, "InfoMiner+: Mining Partial Periodic Patterns with Gap Penalties," Proc. Second IEEE Int'l Conf. Data Mining, Dec. 2002.
21. H. Mannila, H. Toivonen, and A. I. Verkamo, "*Discovering* frequent episodes in sequences", Proc. First Int'l Conf. Knowledge Discovery and Data Mining (KDD'95), 1995, pp. 210–215.
22. 1. M.G. Elfeky, W.G. Aref, A.K. Elmagarmid, "STAGGER: Periodicity Mining of Data Streams using Expanding Sliding Windws", Proc. Int'l Conf. Data Mining, 2006. Pp. 188-199.
23. Y. L. Zhu, S.J. Li, N.N.Bao, D.S. Wan, "Mining approximate periodic pattern in hydrological time series", Geophysical Research Abstracts, Vol. 14, EGU2012-515-1, 2012, EGU General Assembly, 2012.
24. A. Amir, A. Apostolico, E. Eisenberg, G.M. Landau, A. Levy, N. Lewenstein, "Detecting approximate periodic patterns", Proc. First Mediterranean Conf. Design and Analysis of Algorithms, 2012. pp.1-12 , springer-verlag, Berlin.
25. C.Sheng, W. Hsu, M.L. Lee, "Mining dense periodic patterns in time series data", Proc. 22nd Int'l Conf. Data Engineering, April, 2006.
26. S.S. Chen, T.C.K. Huang, Z.M. Lin, "New and efficient knowledge discovery of partial periodic patterns with multiple minimum supports" , Journal

of Systems and Software, Vol. 84, Issue 10, Oct, 2011, pp. 1638-1651.
27. J. Yang, W. Wang, P.S. Yu, "Discovering High-Order Periodic Patterns", Int'l Journal of Knowledge and Information Systems, Vol. 6, Issue 3, ,May 2004, pp.243-268
28. S. Ma, J. Hellerstein, "Mining Partially Periodic Event Patterns with Unknown Periods", Proc. 17th Int'l Conf. Data Engineering, Heidelberg, Germany , 2001.
29. J. Yang, W. Wang, P.S. Yu, "InfoMiner: Mining Surprising Periodic Patterns", Proc. Seventh ACM Int'l Conference on Knowledge Discovery and Data Mining, 2001, pp. 395-400.
30. A.K. Mahanta, F.A. Mazarbhuiya, H.K. Baruah, "Finding calendar-based periodic patterns", Journal of Pattern Recognition Letters, Vol. 29, Issue 9, 2008. Pp. 1274-1284.
31. M. Dutta, A.K. Mahanta, "Mining Calendar-Based Periodicities of Patterns in Temporal Data", In the Proc. Third Int'l Conference on Pattern Recognition and Machine Intelligence. LNCS 5909, Springer-Verlag. Pp. 243-248.
32. M. Dutta, A.K. Mahanta, "Detection of Calendar-Based Periodicities of Interval-Based Temporal Patterns", Int'l Journal of Data Mining and Knowledge Management Process, Vol. 2, Issue 1, Jan, 2012.
33. M. Zhang, B. Kao, D. W. Cheung, K. Y. Yip, "Mining Periodic Patterns with Gap Requirement from Sequences", Journal of ACM Transactions on Knowledge Discovery from Data, Vol. 1, Issue 2, Aug 2007.
34. G. Lee, W. Yang, J.M. Lee, "A Parallel Algorithm for Mining Multiple Partial Periodic Patterns", Journal of Information Sciences, Vol. 176, 2006, pp. 3591-3609.
35. Z. He, X.S. Wang, B.S. Lee, A.C.H. Ling, "Mining partial periodic correlations in time series", In Journal of Knowledge and Information Systems, Vol. 15, Issue 1, 2008, pp. 31-54.
36. Z. Li, J. Wang, J. Han, "Mining Event Periodicity from Incomplete Observations", Proc. 18th ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining, 2012, pp. 444-452.
37. H. Cao, N. Mamoulis, D.W. Cheung, "Discovery of periodic patterns in spatiotemporal sequences", IEEE Trans. Knowl. Data. Eng., 19(4), 2007.
38. Z. Li, B. Ding, J. Han, R. Kays, P. Nye, "Mining periodic behaviors for Moving Objects", In. Proc. 16th ACM SIGKDD Int'l conference on Knowledge discovery and data mining, 2010.
39. K. Ishida, "Periodic Topic Mining from Massive Amounts of Data", In Proc. Int'l Conf. on Technologies and Applications of Artificial Intelligence, 2010.
40. K.-Y. Huang, C.-H. Chang, "Efficient Mining of Frequent Episodes from Complex Sequences", Journal of Information Systems, Volume33, Issue 1, March 2008, pp. 96-114.
41. T. Jindal, P. Giridhar, L.-A. Tang, J. Li, J. Han, "Spatiotemporal Periodical Pattern Mining in Traffic Data", In Proc. UrbComp'13 , August, 2013.
42. N. Mamoulis , H. Cao , G. Kollios , M. Hadjieleftheriou, Y. Tao , D.W. Cheung, " Mining, indexing, and querying historical spatiotemporal data" In Proc. 10th ACM SIGKDD Int'l conference on Knowledge discovery and data mining,2004, pp. 236-245.
43. M. Lahiri, T.Y. Berger-Wolf, " Periodic Subgraph Mining in Dynamic Networks", Journal of Knowledge and Information Systems, vol.24, No. 3, Sep.,2010, pp. 467-497.
44. F. Rasheed, R. Alhajj, " A Framework for Periodic Outlier Pattern Detection in Time Series Sequences", IEEE transactions on Cybernetics, vol.PP, No.99
45. M. Vlachos, P. S. Yu, and V. Castelli, "On periodicity detection and structural periodic similarity", In proc. of SIAM Int'l Conference on Data Mining, 2005
46. Z. Li, J.Han, B.Ding, and R. Kays, "Mining periodic behaviors of object movements for animal and biological sustainability studies", Journal of Data Mining and Knowledge Discovery, 2011.
47. L. Riotte-Lambert, S. Benhamou, S. Chamaille-Jammes, "Periodicity analysis of movement recursions", Journal of Theoretical Biology, 2013, pp. 238-243
48. P. Huang, C-J Liu, Li Xiao, J. Chen, "Wireless Spectrum Occupancy Prediction Based On Partial Periodic Pattern Mining", IEEE 20th Int'l symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2012, pp. 51-58.
49. M. Lv, L. Chen, G. Chen, "Mining user similarity based on routine activities", Int'l journal of Information Sciences, vol. 236, pp. 17-32, July 2013.
50. S. Parthasarathy, S. Mehta, S. Srinivasan, "Robust Periodicity Detection Algorithms", In Proc. 15th ACM Int'l conference on Information and Knowledge Management, 2006.
51. G.N.V.G. Sirisha, M. Shashi, G.V. Padma Raju, :Performance Analysis of Asynchronous Periodic Pattern Mining Algorithms", ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol II Advances in Intelligent Systems and Computing Volume 249, 2014, pp 87-95.

Global Journal of Computer Science and Technology (C) Volume XIII Issue XIII Version I Year 2013

Global Journals Inc. (US) Guidelines Handbook 2013

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.

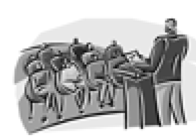
- The “FARSC” is a dignified title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

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.

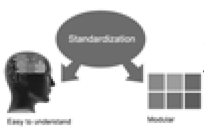


You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

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.

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).

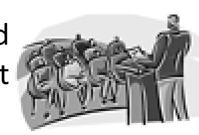
The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

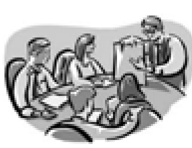
The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

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.



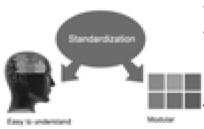


The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

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.

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