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

Global Journal of Computer Science and Technology

DISCOVERING THOUGHTS AND INVENTING FUTURE

Technology 11 Reforming Ideas

October 2011

Pinnacles

- Control Routing Models
- Frequent Pattern Mining
- Remote Sensing Images
- Symmetric Mixture Model

The Volume 11
Issue 17
Version 1.0

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 11 Issue 17 (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 Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Open Association of Research Society, Marsh Road, Rainham, Essex, London RM13 8EU
United Kingdom.

Packaging & Continental Dispatching

Global Journals, 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)

Editorial Board Members (HON.)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Contents of the Volume

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Decisive Analysis of Current State of The Art in Congestion Aware and Control Routing Models In Ad Hoc Networks

By T. Suryaprakash Reddy
Krishna Chaitanya Institute of Technology and Science Markapuram

Abstract - An important aspect that portrays a crucial position in the ad hoc network routing is congestion. Almost every research analysis is en-route in adapting this key factor in addressing congestion. This problem cannot be totally addressed by the regular TCP protocol based networks, keeping in view the special assets which include multi hop sharing etc, which is difficult to ascertain in ad hoc networks. Many attempts have been made and are in progress by researchers to provide unique solutions to the above mentioned problems. This paper projects a vital study on jamming aware and different routing standards that have been dealt with in recent times.

Index Keys : *ad hoc network, routing, congestion-aware, Mobile Ad-Hoc Networks, Wireless Multihop Networks, Congestion Control, TCP .*

GJCST-F Classification : *C.2.2, C.2.1*

Strictly as per the compliance and regulations of:

Decisive Analysis of Current State of The Art in Congestion Aware and Control Routing Models In Ad Hoc Networks

T. Suryaprakash Reddy

Abstract - **An important aspect that portrays a crucial position in the ad hoc network routing is congestion. Almost every research analysis is en-route in adapting this key factor in addressing congestion. This problem cannot be totally addressed by the regular TCP protocol based networks, keeping in view the special assets which include multi hop sharing etc, which is difficult to ascertain in ad hoc networks. Many attempts have been made and are in progress by researchers to provide unique solutions to the above mentioned problems. This paper projects a vital study on jamming aware and different routing standards that have been dealt with in recent times.**

Index Keys : *ad hoc network, routing, congestion-aware, Mobile Ad-Hoc Networks, Wireless Multihop Networks, Congestion Control, TCP .*

I. Introduction

Wireless straits and node portability project a negative authority on data packet broadcasting which paves way to intrusion, which ultimately affect the load balancing. Similarly, the ever increasing claim on real time data broadcasting leads to failure in data transmission and bandwidth results in loss of valuable energy and time. Some provision can be made for congestion awareness which can ensure the broadcasting without any intricacies. Time-varying routes can be adapted to put up with through various wireless IEEE protocols and MANET transmission supports. IEEE 802.11 DCF assures extended equity data packets in a network which may face a working problem such as decrease in performance of the high pace nodes [22]. B. Awerbuch et a [23] insists on usage of multiple paths for data flow and data transmission. A multi-hop replica deems decrease in low pace nodes that ensure a rise in rate of data linkages. Diminishing the usage of excessive nodes with an overall low data rate ensures escalation in network performance. A medium metric time has been proposed by the B. Awerbuch et al[23], which enables us to select connections with low data rate, i.e, those links which involve the most number of data paths will be picked up ultimately ensuring soaring data velocity. This can be achieved only when the data packets are broadcasted through the chosen data routes [24].Increase in jamming of networks poses the problem of access conflicts [25]. Channel admission setback usage was boosted as a recommendation to the MTM affords routing avoiding restricted regions. The gain in the case of elevated data speed is that the links are basically petite and they can be easily accessed via any path of the network chosen. There is a possibility of over-crowding if there is a rise in the number of connections [26]. Then the channel access impediment, which took utmost care of avoiding jamming in severe areas of the network. Channel tenure, buffer loading and data packet loss velocity should furnish precise capacity of link jamming levels. Path possessing various links with a variety of data types will positively direct multiple networks of various rates.

If by chance, the paths which have low data pace pursues higher data pace links, data packets mount up leading to long tail. Link consistency also poses the issue of jamming, which causes in re-broadcasting.

II. Taxonomy

a) What is Congestion?

Data rate used by each dispatcher on the network is essential to regulate so as to avoid jamming of networks which involve shared resources. There are instances when data packets are declined access to traverse through networks and are left midway. As a result, circumstances may arise wherein extreme number of data packets arrive and are evidently plunged. These extreme number of data packets gobble up a considerable amount of shared utilities while travelling in a network and usually cause broadcasting to be done again, which ultimately indicates entry of even more number of packets than the ones already present in the network resulting in worsening of network throughput. If suitable jamming control method are not implemented then it may happen that the entire network gets jammed and no data is sent to the destination fully and successfully. TCP control jamming procedure was established when such a situation took place in the early Internet.

b) Congestion Factors:

A network routing track can be jammed even if the actual load is reduced in the case of aggressive

Author : Professor , Department of Computerscience and Engineering, Krishna Chaitanya Institute of Technology and Science Markapuram, Prakasam (Dist.), A.P., INDIA.
E-mail : surya.prakash12@rediffmail.com

retransmissions of networks. Collapse occurs when throughput is reduced keeping in check the input load as stable and put. This happens due to some congestion factor and event termed as congestion collapse takes place. Some routers that act as mediators ensure packet decline in case of over-flooding hoping that the end points resurface the dropped packets back.

c) Consequence factors in congestion aware and control routing models

Route failures: Data packet and acknowledgement broadcast accounts for a substantial break because of loopholes in the paths of a network. This state forces a TCP sender to lessen his window volume resulting in jamming.

Wireless losses: There is a wrong elucidation that accidental packet losses due to other causes in a network excluding congestion is overcrowding which projects an unconstructive impact on jamming protocols.

Shared medium: The major confrontation faced by jamming control standards is recognizing real situations which involve dropping packets n a shared segment. The ultimate and essential regulation is compacting jamming at an early level that minimizes further damage to a network.

Acknowledgement traffic: This factor presents a situation where in a data packet and an ack packet ensure transmission in reverse directions which raises intra flow disputes. This factor also sometimes is responsible for deceiving jamming control standards.

III. Predominant Citations

a) Citations that deal with rout failure flaws:

TCP-Feedback (TCP-F)[1]: Non jamming related fatalities and timeout states are resulted due to route failures which possibly immobilize TCP jamming control procedures.

Explicit Link Failure Notification (ELFN)[2]: This feature involves intimation via feedback through lower layers which takes the responsibility of notifying TCP procedures about any incidents of routing failures. Such a state enables a user to take to the stand-by mode which is similar to the snooze condition of the TCP protocols. In such a case, 2 options are available. The author attempts to make use of an ICMP host or resend the intimating signals back to the network. It is familiar with researchers and used in many of the approaches.

TCP with Buffering Capability and Sequence Information (TCP-BuS) [3]: Supplementary procedures are suggested that makes a point in escalating TCP performance, which helps in extending the ELFN. An unswerving liberation scheme is offered which makes certain the delivery of specific notices. Unequivocal warnings and probe data packets are scrutinized in order to establish and check re-established tracks.

Fixed RTO scheme [4]: This is in effect with the ELFN scheme and involves sending of data packets episodically, This scheme is intrinsically altered according to the requirements when compared to the ELFN scheme again.

Enhanced Inter Layer Communication and Control (ENIC)[5]: This scheme is a combination of ELFN-like route failure handling and the TCP SACK and DACK procedures. This does not involve separate notifications like ELFN but instead they are reused.

TCP-ReComputation (TCP-RC)[6]: This focuses on re-establishment of route altering route properties. This is an extension to ELFN which considers and deals with re-computation of TCP window size on the basis of the assets of the new track (cwnd) and slow start threshold (ssthresh) constraints.

TCP for mobile ad hoc networks (ATCP)[7]: This state makes use of TCP state procedures with the involvement of TCP-ELFN like packet probe procedure. In a case when no ECN message is obtained, the loss that occurs is considered to be non-jamming related. An additional layer called ATCP layer is built in below the transport layer that takes care of minimizing the dealings with respect to TCP.

Cross-layer information awareness[8]: Two procedures have been introduced here, the first Early Packet Loss Notification (EPLN) and secondly Best-Effort ACK Delivery (BEAD). EPLN takes the responsibility of intimating TCP sender the series numbers of the lost data packets that were unable to be retrieved. The sender then immobilizes and the re-sends the respective data packets when the path is re-stacked again. Intermediate nodes receive BEAD loss warnings and then they are forwarded towards the TCP destination, thus thwarting BEAD loss.

TCP with Detection of Out-of-Order and Response (TCPDOOR)[9]: The TCP jamming control procedures may be suspended momentarily by the sender when dis-ordered packets are perceived. This involves a scheme called Instant Recovery where in effects of the jamming methods are recovered. The anticipated effect is parallel to that of the freezing technique after the reset changes to TCP and the connections operate as if no route change has occurred.

Signal strength based link management [10]: A record of node remoteness that is approximated RSSI values is kept in routing standards. The routing standard gets an intimation when the link breaks and a new track to the receiver is investigated. The author observes that 802.11 MAC fails to recognize break links accurately in connection to jamming. Two incidents are recognized where in links move out of the transmission arena and a jammed path does not grant access to successful RTS.

b) The citations those deal with wireless losses in ad hoc networks:

TCP/RCWE [11]: TCP with Restricted Congestion Window Enlargement is again formed on te basis of the ELFN procedure where the bond between nodes break and losses occur. A mechanism called RTO ie. Retransmission timeout is introduced which takes care of random losses. This value can be calculated on the basis of the observations that if the RTO value is higher, then the TCP window size is low and vice-versa.

ADTCP [12]: These showcase and throw light on the major issue of end-to-end transport standards in movable ad hoc networks is the unruliness of dimensions of indicators for some network occurrences. There are two metrics that are proposed to recognize network jamming. The inter-packet hindrance difference at the destination gets high when jamming occurs. Furthermore, the short-term throughput decreases. These two metrics join together to form a secure and strong jamming indicator. Route changes and channel errors are distinguished by the advent of out-of-order packets and packet loss proportions.

ADTFRC [13]: ADTFRC scheme is similar to TFRC, just as ADTCP is to TCP. An identical amalgamation of the metrics and general procedure are used for differentiating loss types and supply receiver-based feedback. ADTFRC contributes many of its advantages to ADTCP.

Edge based approach [14]: Medium loss detection triggers TCP jamming control response. When the destination fails to receive data packets after waiting for a long time and paves way to timeout, a track failure occurs. TCP gets into ELFN like "probe mode" where data packets are broadcasted at equal intervals of time which in turn senses a re-established track.

c) The citations those deal with shared medium issues:

Link RED (LRED) & adaptive pacing [15]: The most favorable TCP window size subsists for the given topology and traffic prototype other than the fact TCP fails to find it. Larger windows cause link-layer disputation which authorizes research in wireless multi-hop jamming control momentously. LRED makes sure that TCP flows normalize their window size closer to the best possible region of MANET's.

Neighborhood RED (NRED)[16]: In this scheme every node makes a guess as to how many nodes are lined up in succession in total. All the prevalent packets form an implicit and allotted queue. If the queue exceed the set protocol, then the data packets are dropped and lost with the maximum prospects.

Contention-based Path Selection (COPAS) [17]: This scheme mainly concentrates on the TCP issue in MANET known as the capture problem. Nodes can detain mediums unfairly and have an edge over others. The annex for reactive tracking standards is COPAS. When COPAS is in progress, all the links from the source to the receiver are taken into consideration. Then upstream TCP transfer and downstream acknowledge-ments are promoted by two of the selected tracks which then ignores after-effects when one of the directed medium gets arrested.

Congestion Aware Routing (CAR) [18]: This concentrates on TCP flows and jamming control procedures. There are few credentials that deal with Acknowledgement transfer problems:

Dynamic delayed ACK [19]: There is a blend of two conventional ACKS's, where in data packets are forwarded to the receiving end only after a predetermined timed-out sequence or when a specified number of segments are turned over.

Dynamic adaptive acknowledgment [20]: This scheme is taken care of by the application of TCP hypothesis of their jamming procedures to MANET's by a dynamic timeout ACK that is estimated on the basis of the arrival time of the data packet at the destination.

Preferred ACK retransmission [21]: This model amalgamates two schemes explained earlier mainly, ELFN messages and DACK for TCP present on Flexible Radio Network (FRN). This FRN is viable and easily accessible MANET system powered by FUJI ELECTRIC. What is noteworthy here is the fact that it is in no way connected to the IEEE standards and makes use of predetermined instances on the modes.

IV. Current State of The Art

Elastic Routing Table with Provable Performance for congestion Control in DHT Networks [27]: In this proposal authors discussed the Elastic routing Table mechanism for query load balancing; this mechanism is based on the fact that high-degree nodes are more efficient in handling traffic loads in DHT networks which suffer from a serious drawback of load balancing problem. In the ERT mechanism each heterogeneous load has a routing table, the size of which is a function of corresponding node capacity.

The DHT network works by establishing a functional relationship between the nodes and utilizes a routing protocol to identify the node responsible for the key. When the network is heterogeneous comprising of time-variable popular files the problem of load handling turns even more serious. Moreover, this problem results in the phenomenon "Bottleneck" which is nothing but the engagement of a node with too many queries at a time. Even though the DHT networks have advantages of good lookup efficiency, robustness, scalability and deterministic data location but the inherent loading balancing problems as the continuous occurrence of hashing results into imbalance degree of keys between the networks nodes. Thus the design of such a DHT protocol which can overcome the congestion control is a challenge.

The paper proposes the strategy to cope with the node heterogeneity, skewed queries and churn in

DHT networks which can overcome the flaws of past strategies, namely the requirement of same and constant DHT degrees. The ERT technique is fundamentally derived from the principle of power law networks.

The methodology adopted to resolve the issue of congestion control protocol, serving the rationale of avoiding bulky nodes in query routings and to allocate query loads among the nodes according to the capacities of the nodes is that the load share of every node is proportional to its "fair load share", which is denoted as s_i and is defined as ,

$$S_i = \frac{(l_i / \sum_i l_i)}{(c_i / \sum_i c_i)}$$

Where the variables stands as, c_i : capacity of the node i to handle number of queries in time T.

L_i: load of node i, i.e. the number of queries it receives and transmits to it neighborhood in time T.

The ideal situation is attained at S_i=1, at this stage the node is not overloaded and the distribution is fair. The paper suggests two methods to attain this ideal condition, they are:

- Periodical assessment of load l_i and transmit the queries accordingly.
- Application of the principle that a high-degree node would be the most probable to experience the high query load.

The authors have utilized the second techniques out of the above mentioned, by describing a new variable d_i which represents the indegree of a node. The authors have assumed that node and file queries are uniformly distributed and under this assumption it has been found that the l_i is directly proportional to the d_i . However, the authors have intelligently reversed the relationship for the determination of node's outdegree by choosing an appropriate value of indegree. Further, the authors have normalized the node capacity so that the mean of capacities is 1 i.e. $\Sigma_i c_i = n$. And if the churn is eliminated the load is directly proportional to indegree of node.

The fair load share s_i *expression* transforms to

$$S_i = \frac{(d_i / \sum_i d_i)}{(c_i / \sum_i c_i)} \text{ and } d_i \approx c_i \frac{\sum_i di}{n}$$

Where $d_i \approx c_i \frac{\sum_i di}{n}$ here the multiplier is a constant and is declared as α , which is a system parameter and is defined as a functions of different metrics in system experience. Experimentally, it has been found that the high-load system have small alpha whereas the low-load system have high alpha. For practical uses with initial degrees of nodes as €d_i where € is predefined % of the application purpose which is defined in pursuance of actual system carrying load capacity. The authors have taken into consideration Gnutella network and have used the bidirectional links similar to the Gnutella network in the DHT network by maintaining a backward outlink for every inlink, in order to get the acquaintance with the nodes forwarding the queries to the later to reduce the load and non-uniform distribution tendency. As a result, of which a double link is maintained for each routing table neighborhood.

The initial indegree assignment is not strong enough to restrict the load which is a function of time. The authors have practically examined that the nodes join and leave DHT overlays consistently and the files in the system have non-uniform and time-dependent popularity. Hence the congestion control protocol should be designed in such a way that it controls the query flow towards the nodes with sufficient capacity and it can change the query rate and time-varying file popularity as well as network churn. The authors have designed a periodic indegree adaption algorithm which facilitate every node to fiddle it's indegree at regular frequencies pursuant with the peak load encountered by the node. To perform this task each node records its load in a specified time T at regular frequency and checks whether it is overloaded or unloaded, this efficiency is recorded by a constant ¥$_i$. However, even the in-degree of adaption may be impotent to deal with query load imbalance. The authors have proposed a complementary randomized query forwarding algorithms to help forward queries towards light nodes in order to further condense the lookup latency.

The authors conducted simulation tests for ERT over various aspects and describes the distinguished properties of the ERT based congestion control protocol. The parameters chosen by the authors to define the performance of the ERT mechanism are as follows:-

- Metric of 99th percentile maximum congestion to measure the network congestion.
- Query distribution share S_i which represents the fair load distribution. The paper has put light on the reasons of unfair load distribution in DHT networks, which are:-
  i. Difficult to record the data for loads and capacity.
  ii. DHT is a dynamic system with continuous joined and departure of nodes.
  iii. The load changes with the file popularity and churn.
- Query processing time which is determined by two factors lookup path and number of heavy nodes encountered.

Observation: Authors have conducted experiments on cycloid networks without congest control (base) and with ERT based congestion control. The results due to the virtual server are also included for better understanding. In the test representing the

congestion control efficiency, it has been found that ERT leads to better efficiency and much lower congestion rates, apart from this in the test of lookup efficiency which is determined by two factors, namely, lookup path length and query processing time in each node along the path the test result proved that the ERT/AF leads to much higher lookup efficiency in comparisons with others. At last, the authors have described the effect of skewed lookup, effect of churn and adaption & query forwarding.

Unified Approach To Congestion Control And Node-Based Multipath Routing [28]: Fernando Paganini et al., takes into account a TCP/IP-style networking along with flow-control at end-systems based on Congestion feedback and routing decisions at network nodes on a per-destination basis. The key aspect is to allow the routers to split their traffic in a controlled way between the out-going links by instilling the global optimization criteria. A concrete implementation of various algorithms is presented, based on queuing delay as congestion price. In order to develop a multi-path variant of the RIP, we make use of a TCP-FAST, which is clearly demonstrated through the ns2-simulations.

It is already known that the amount of traffic by the transport layer and the routes chosen by the network layer greatly influences the congestion present in the packet-switched-network. But, adapting the TCP mechanism, which is slow to instantaneous congestion results in routing instabilities. So, this solution opts to Multi-path routing which can more easily attain the required equilibrium. In terms of Math, the basic difference between the single path and the multi-path is evident when the optimization of a convex congestion cost is considered to serve the matrix of end-to-end demands. The interpretation of the optimization particularly proves to be useful when the combination of multi-path routing and congestion control are mixed up. A multi-path proposal by Kelly, though is mathematically perfect, implies to transfer functionality from the network to transport layer, which needs to be aware of the network paths present in the anatomy. The work has come up with a more node-centric and scalable alter, i.e. to have routers in-charge of the multi-path function by monitoring the traffic speed to their destinations among their outgoing tasks.

As far as the formulation aspect of the problem is concerned, an N nodes network is considered which has L direct links. The network hare supports various flows between the source-destination pairs of nodes. For unique indications of flow, the traffics are allowed to follow multiple paths between source and destination. By calculating the incoming flow, the outgoing flow, the total flow in a link 'l' is formulated. But, in this thesis of formulation, optimization problems occur. These problems are further diluted into 1) Welfare Problems which outputs the maximum achievable utility over all sources if the traffic is allowed to follow multiple paths between source and destination and 2) Surplus problems. Also, by apt re-definition of variables, the equivalence of the above problems can be showed. In order to account for the reasons of scalability as well as preserving layer separation, a pre-defined set of variables is used. Another aspect that arises is the Feedback Signals. The primary feedback signal is a congestion measure for each individual link l out of the links L. it is well assumed that there is no 'service differentiation' between the commodities. Different paths from the source to the destination will have their own 'prices' at any time. To be aware of such paths, routers are not required in specific; rather they can work with the local and neighbor information in order to infer their price-to-destination. If the link prices are given, under mild conditions, the unique solutions can be obtained to the recursive equations.

Another important aspect of concern is the Anticipative Control of Traffic Splits and its Stability. All the limitations of multipath routing based on the gradient of congestion price are revealed. Just by slowing down the adaptation, oscillatory instabilities that appear cannot be overcome. If the feedback factor B is reduced, the frequency of the oscillations may be reduced but the oscillations remain. The key idea presented in the paper top introduce damping into the system is to use "proportional derivative control", i.e. to introduce some anticipation of future prices in the control of the routing splits.

All the theory thus described, can further be installed into a more than one implementation, depending on the choice of the link congestion price, the source utility function, and the methodology for sharing congestion information between the routers and with traffic sources.

1. Routing Protocol and Node Price Formation: unlike the prevailing computation methods, routers disseminate metric information for computation, either globally or to neighbors'.
2. Update of Split Ratios, Blocking and Forwarding: the methodology for avoiding the formation of routing loops is discussed here.
3. Communication of Prices from routers to Sources: Considering the time-scales involved, the issues pertaining to the multipath algorithm is discussed here

Another major part discussed in the paper is of Simulations. This part elaborates and contrasts about the Gradient and Anticipatory control with a dynamic example of the 4 Node Topology. The simulation results thus depicted contain split ratios and metrics (prices) for the denoted nodes. The various processes such as the initialization process wherein all nodes first discover the direct route to the destination. The default route is highlighted here in this process. The TCP-FAST sources react to the lowering in the average queuing delay by

increasing the rate. It is noted that as the traffic increases, the congestion increases, thereby the traffic is reduced through the route. Each of the nodes used have a number of routing links through which they send the traffic.

By and large, a framework has been successfully designed and proposed by the authors, containing details about multipath routing and congestion control which together are helpful in pursuing a common objective: the maximization of aggregate utility or surplus over the network. It is also noted that the control of input rates and routing splits is purely decentralized, completely relying on the sole factor of a common congestion "currency" for enhancing its decisions. A detailed mathematical study has been done on the equilibrium and dynamic properties of various control laws, in particular, a new anticipatory control of traffic splits has been proposed that stabilizes the maximum welfare allocation when it is combined with the dual congestion control. The persistent TCP flows have been assumed in the theory. Provided the relatively slow dynamics of the routing is given, it becomes very important to drag on the work to take into account the effect of finite TCP flows that come in and out of the network. This concept is kept open for the future research. Another key topic that has been presented is combining the network control and transport layers with the lower layers, particularly for the wireless networks, possibly offering alternatives to the back-pressure scheduling approach.

Observation: A detailed packet implementation based on queuing delay has been presented in the work. The delay pertains to congestion price wherein the routers measure local prices exchange information with neighbors', thereby following a multipath variant of a distance vector routing protocol. FAST TCP sources clearly estimate this particular delay from their RTT measurements in real time, calibrating their propagation delay through regular periodic interactions with the respective IP layer. The expected behavior from the theory is verified from the ns2 simulations. An alternate consideration is that one can as well consider implementations based on the loss of the marking as a congestion price.

Congestion-Constrained Layer Assignment For Via Minimization In Global Routing [29]: Multi layer Routing, which is one of the advanced technologies is usually done in two methods one being Detailed Routing and other global routing. The Present Paper focused on global routing. This Methodology uses full- fledged one layer routers and produces one layer routing result. Though this is an effective method, an advanced assignment algorithm is still required to bring in effective routing . Though the layer assignment the wire length and topology of the initial one layer routing result can be saved and delay, cross talk and vias can be reduced.

In Multilayer routing, of the net different layers are connected through the technology 'via'. But extensive usage of the via degrades the effectiveness of Routing. As such to minimize the via cost several new bench marks were released by ISPD '07. This paper also projects a new algorithm for via cost Minimization in layer assignment.

To minimize the via cost in the layer assignment two problems must be resolved. One is to see that the same routing topology is maintained and the other is total overflow constraint and maximum overflow Constraints are resolved. The present paper presents a layer assignment algorithm which can bring a solution for these congestion constraints.

The algorithm introduced is known as Congestion – Constrained Layer Assignment (COLA). COLA has two major procedures one is net order determination and the other is Single – Net layer assignment notifying congestion problems.

Net order determination is very important and decides the routing resources. These resources are very few as such net order is required for the maximum usage of the given resources. As such a new method is introduced by COLA to achieve the best utilization. Here in this methodology, for each one layer routed net, three factor are utilized to calculate score so as to decide the net order. The use specified parameters in the net order after the calculations of the score will be &,B and Y. In the new method proposed & and B are set equal and Y will be lesser than & and B. This produced a better layer assignment result.

For the single-Net layer assignment considering congestion problems a new algorithm which combines SOLA and APEC is newly presented in this paper. The proposed algorithm removes a group of edges from each cycle to simplify the procedure. SOLA which is a programming based algorithm is very effective in resolving the layer assignment problem where congestion problems are not taken into account. And APEC (Accurate and Predictable Examination of Congestion Constraints) is very useful in reproducing an edge during the layer assignment for a net. Which congestion constraint violation does not happen? COLA a new algorithm presenter in this paper combines the two, SOLA and APEC to resolve Single-net layer assignment problems.

COLA was experimented with standard ANSI C++ and tested on a Linux work station with AMD Dual Core Opteron Processor 2. 2 – GHZ CPU and 8-GB Memory. The first published multilayer global routing benchmarks, ISPD 07 were used in the experiments. COLA was tested to see if it could be improve the results of Maize Router, Box Router, and FGR and was implemented on a straight forward greedy algorithm to see the effectiveness of the COLA algorithm. Though the greedy algorithm a feasible layer assignment result for each benchmark is possible but the total wire length

results will be worst. Whereas COLA was able to bring down the total wire lengths and improved the Via wire lengths more than Maize Router, Box Router and FGR Router could do.

From the experiments it was found that COLA did not take more than 2 min for each benchmark which is as good as greedy algorithm. Whereas CPU time taken by COLA is much lesser when compared to greedy algorithm.

Net order Determination method introduced by COLA was also experimented. The Via wave length results by COLA are much better when compared to the existing methodologies. From these experiments it was found that net order determination method and single-net layer assignment method are equally important for the success of COLA algorithm.

COLA was also tested on the compressed one-layer results of Box Router. And results show that COLA was still able to improve the results of Via wave length and the total wire length on each benchmark for Box Router 2. 0. The other note worthy point is that it does not give raise any parallel edges. In the case of Net with parallel edges, COLA was able to produce either better or worse layer assignment results in terms of the Via wire length.

In the case of layer assignment also COLA was able to achieve better results. in the Layer assignment situation where the parallel edges were not allowed COLA decreased the via wire length. But when the parallel edges were disallowed COLA could not perform well and rose the via wire length. The success of COLA was seen in the improvement of total wire length results and reduction of total overflow results. The advance step of COLA is COLA_R which can be used to iteratively refine a multilayer global routing solution in a net-by-net manner. COLA_R was tested and found that it was able to improve Via wire length total wire length results. And it was able to reduce total overflow.

Observations: The VIA capacity constraint must be considered in a better practical layer assignment problem so that a routable result can be given to the detailed router. But this paper does not produce any solution to Via capacity constraint and limits itself to the layer assignment problem.

Interference Minimized Multipath Routing with Congestion Control in Wireless Sensor Network for High Rate Streaming [30]: Focuses on dealing with and providing solutions for estimating the track attributes and appropriate resolution for multi track load complimenting. Jenn-Yue Teo, Yajun Ha, and Chen-Khong Tham recommended an Interference-Minimized Multipath Routing (I2MR) protocol that concentrates on mounting throughputs by ascertaining zone disjoint tracks for harmonizing weight. Handling and directing overcrowding proposals is also another feature that is performed and supposedly done. Basically of importance mainly in armed services wherein wireless intersection joints of minimal control are involved which prepares a base for underground WSNs. These underground WSNs mostly help in transmitting data of sharp bandwidth with the employment of Unmanned Aerial Vehicles (UAVs).

Harmonizing load in several tracks is fundamentally elucidated with stationary wireless systems in which a network can be represented with the help of a connectivity grid. Illustration of interfering procedure copy or rather a substantial copy of interfering networks can reproduce wireless intrusions in wireless arrangement. The basic number of linkages that aids in linking two paths can be traced with the correlation factor concerning two node-disjoint tracks. This correlation factor metric effectively depicts the quantity of wireless intrusions, thereby stating that the two connections are correlated and if in case where there is an absence of connections, it is reasoned that the two mentioned tracks are distinct. There is also the involvement of a conflict graph whose vertices conflict directly to a series of intended links. To ascertain this, various notions are made up that proposes few beliefs such as network concerned is wireless and stationary, a consistent broadcast and intrusion series for all connections, distinct basis and target connection and so on.

The instant end information is constantly confined via the antenna by means of EO through a solitary path along with controlled energy WSNs which provide associations to the UAV resilient scheme concerning the adjoining access nodes, hence concentrating on the competent multihop dispatching of information. The crisis is basically addressed as multipath and multihop steering wherein the foundation node endeavors assembling of three sector disjoint trails namely primary, secondary and backup lanes to the end node which are termed as primary, secondary and backup destinations correspondingly. The primary and the secondary trails are mostly in parallel use for harmonizing load and swaps itself to the backup state whenever a failure occurs, hence ensuring trail unearthing slide reduction. The underlying principle in usage of two trails is that there is literally little or no expansion in terms of cumulative throughput while using more than two trails. There are some suppositions supposedly made which state that the WSN connections are stationary, recognized sites for the basis and the entryway, basis and entryway sites are quite less energy-controlled, the start-target duo are duly placed at equal intervals, provision of sharp competence and non-intrusive data linkages to bond several entryway sites to the resilient intersection point etc. There is commencement of the track detection process from the basis to the ultimate target site. The process of track detection covers three fundamental measures, Primary path discovery, Interference-zone marking and Secondary and backup path discovery. Primary path

discovery aids in building unswerving tracks which reduces intrusion with respect to tracks present in the interiors starting from the basis and ending at the primary target. Interference-zone marking grades one and two hop neighbors of primary tracks involving transitional intersections of least possible transparency. A maximum number of two hop neighbors are taken into consideration due to the base of the meddling scope measured twice the number of the communication scope. Interference-zone marking marks involvement of three undemanding moves, i.e. Sector marking, Broadcast Zone-marker Potential (BZP) assignment and Zone marking. Sector marking engages nodes the length of the primary tracks and aids in categorizing the corresponding neighbors into individual divisions. Broadcast Zone-marker Potential (BZP) assignment takes charge of allocating various BZPs to distinct areas concerning distinct sections. Zone marking in scripts a maximum of two-hop neighbors of respective Sector Heads (SHs) into Interference-Zone1 or Interference-Zone2.

Jamming Control schemes basically for I2MR (Interference Minimized Multipath Routing) consigns tracks at pre-defined pace which necessitates the need of three essential steps. Firstly, it engrosses in perceiving extended tenures of track jamming arenas. Secondly, there is a requirement of notifying basis about track jamming areas. Lastly, there is a want for plummeting the basis's loading velocity. Tentative intentions are firstly, I2MR for dissimilar network concentrations and the pace at which packet failures occur by making a contrast between track detection proposals compared with the other respective track detection proposals. Four different plans are made use of for analyzing performance of various track-sets mainly Aggregate throughput, Average end-to-end delay, Total energy consumed and Packet delivery ratio. The trials performed are based on the researches carried out making usage of GloMoSim network simulator. Stationary nodes are positioned homogeneously wherein the region is partitioned into a considerable number of chambers.

Observation: It is proved that I2MR is comparatively efficient in terms of the total track detection time, total control bytes broadcasted and the sum of the intact energy devotedly used up during the process of track detection, but NDMR seemingly devours more amount of energy due to the considerable dimension of the control packet as compared to I2MR. Even if the track-set performances for managing and organizing jamming are taken into consideration, I2MR plays a remarkable role with maximum throughput and significant gains when compared with I2MR, I2MR50, NDMR, and AODV. Another substantial finding worth mentioning is that I2MR consumes lowest amount of energy when compared with other multiple track proposals which makes it even more appealing. It can hence, be deduced as the modus operandi used for multiple track load balancing is effective enough to encapsulate the consequences of both intra and inter wireless intrusions, the projected I2MR procedure can showcase a considerable gain in throughput and the projected jamming control plan can increase throughput by loading effective tracks at the peak rate which in turn helps in reducing long-term overcrowding of tracks. The most probable restraint caused by I2MR protocol is that there is a vital need for spacing the source and the target that is required to be done by the wireless intrusions present between various adjoining track-sets. Future work has a proposal for an extension of the projected I2MR protocol which considers the outcome of inter-path set intrusions, with the sole viewpoint of deployment of WSN.

A Low-Complexity Message-Passing Algorithm for Reduced Routing Congestion in LDPC Decoders [31]**:** An LDPC (low density parity check) is generally used and got famous for error correction and real channel capacity performances, and also generates very low errors, for which its being suggested and possess a good error performance with large code lengths. Hence, these are widely used in Wi-max and 10GBASE-T. Every LDPC is implemented with a split row threshold algorithm that improves error performance and reduces routing congestion.

There are many iterative message algorithms for an LDPC decoder. A Min-sum decoding algorithm is an iterative message passing algorithm which is very vastly used for practical decoding. A fully parallel decoder maps every row and every column of parity check to different processing units which operate parallel in the same. These have larger area and a very good capacitance and a low operating frequency and are highly energy efficient. Fully serial and partially parallel decoders have a processing core and a memory block having lower throughputs and high latencies. And finally, an early terminator for message passing algorithm, simulations are used to determine predefined set for range of SNRs and as the condition gets satisfied early termination happens and is considered error free.

We will now be discussing about split row algorithm and ability to reduce routing congestion in layout, a split row decoding algorithm reduces the interconnect complexity. Here the total congestion is minimized and is used with SPA and minsum algorithm and a routing congestion reduction in this split row algorithm reduces routing congestion by a factor. And it is to note that splitting reduces routing congestion affecting global interconnects caused by message passing.

There is also a split row threshold decoding method in which we have a split row error performance that suffers 0.4 to 0.7 dB error proportional to spn as every split row has no information about minimum value and a split row threshold algorithm increases the error

performance without reduction of effect of split row algorithm which creates a negligible hardware to check node processor and the kernel of split row algorithm is in four conditions, they are an Origin Block i.e., reg2out delay in which the path where threshold_en signal is generated in a block. This path has comparators to generate min1 and min2 , in addition to a comparison with threshold(T) , and an OR gate to generate threshold_en_out the signal going to the next partition and middle blocks i.e., in2out delay This path has middle blocks where threshold_en signal is passing through. Assuming that local min1 and min2 in all blocks are generated simultaneously, the delay in a middle block is one OR gate which generates the signal.

And finally in, destination block i.e., in2reg delay this is the path that a block updates the final check node output and is using threshold_en signal from neighboring partitions. The path which goes through the variable processor and ends at the register will be discussed in detail. The error value depends only on the threshold values and the benefit of split row threshold is partitioning of check node processing is arbitrary.

The architecture of this whole design consists of a check node processor, variable node processor and then we implement a fully parallel decoder, which check the logic gates, compared values and correction factors and its implementations while the equations of variable node processor remain unchanged for a both min-sum and a split row threshold decoder this implementation uses a 5-bit data path and finally, and in fully parallel implementations, all check and variable processor. Outputs are updated in parallel, and as in the timing.

In the C'MOS design of decoders the design flow and its implementation use standard cell based automatic place and a routine place decoder which stresses on the working, and in the delay analysis we discuss about the interconnect variable delays and delays in various blocks and methods to reduce it. And analysis about area after synthesis and layout and before synthesis and layout and its wiring and exponential problems, power and energy analysis deals with energy consumed by decoders, capacitances partitioning and its variations and methods to yield higher output and the variations on capacitances on different kinds of fabrications.

Observation: The various disadvantages of this are it requires a high performance and low power with a large number of nodes having high degree of interconnectedness and a large memory capacity with high memory bandwidth. These necessities are still due, to the message-passing algorithm which are used by the LDPC decoder. Actually, this was done with the sum–product algorithm (SPA) or min-sum algorithm. Our previous work introduced two nonstandard LDPC decoding algorithms dependant on min sum, called "Split-Row" and "multi-split" algorithms that were proven to the increase throughput up to five times, and reduce wiring and area up to three times. Split-row algorithm achieves this through the partitioning of min-sum algorithms global operation into semiautonomous localized operations. Hence, the reduction in message passing, there is a 0.3- to 0.7-dB reduction in performance, depending on the level of partitioning.

Multipath routing algorithm for congestion minimization [32]: Various track routing policies can relieve jamming of systems when divided among a number of trails. Routing proposals that are presently in use throw light on ascertaining a best possible path for routing purpose which results in a generous loss of network means. A varying methodology called multipath routing is a favorable solution which makes use of a number of superior tracks instead of a lone track. Crowded systems bring with them the issues of deprived presentation and high discrepancies which can be evaded using multipath routing. These issues have been addressed competently with the aid of two problems specified below which are first screened meticulously with calculations to find out whether there are any discrepancies involved in the process. The jamming issue is addressed with the aid of two problems that are taken into view to prove the facts mentioned in this paper, Problem RMP (Restricted Multipath) and Problem KPR (K-Path Routing). Problem RMP (Restricted Multipath) deals with reducing the network overcrowding issue with respect to the constraint set on the span of the track selected for networking. Problem KPR (K-Path Routing) steers clear the obligation of bounding the various tracks, at the same time taking into consideration reduction of network jamming issue. Problem RMP (Restricted Multipath) as stated before focuses on reducing overcrowding of networks under track quality limitations. It is statistically proved that this problem can be solved dutifully with the help of respective calculations. A pseudo-polynomial result comes into existence and an optimal estimation proposal is devised for the same. On the same lines, this problem paves a way to two additional logics, Multi commodity Extensions and End-to-End Reliability Constraints. There is a presumption in Multi commodity Extensions that only a lone source target pair is present in all occurrences of the given problem. End-to-End Reliability Constraints explains and provides an explanation to the instance which states that any kind of breakdown in any track in case of multiple tracks will ultimately affect the entire broadcast. Problem KPR (K-Path Routing) is also explored which while routing passage along K number of various paths, reduces overcrowding of networks. A clear polynomial solution is presented as the problem is NP-Hard and happens only when the limitation on the number of tracks is more than the number of connections. Accordingly, in case of the number of tracks being less than the number of connection links, a two point estimation format is invented. Similarly, Problem Integral

Routing is proposed on the basis of provision of a supplementary state which handles network jamming all the while limiting access along each path present in the network.

It is proved through thorough simulations that multipath routing proposal results acquired by the most favorable jamming reduction schemes are comparatively well-organized than single path routing proposals. Hence, two classes of arbitrary networks are created, Power-Law topologies and Waxman topologies. This is done ensuring that the link competence, bandwidth demands and the span of each connection is spread maintaining the consistency factor. There is an assessment prepared on the basis of the contrast in the jamming caused by multipath routing and jamming caused by well-known single track routing protocol.

Observation: It is syntactically verified that feasible overcrowding reduction proposals are noteworthy as compared to single path routing proposals. What is important to note here is that both the problems discussed present a distinct series of reasonable explanations. Problem KPR (K-Path Routing) limits the stream of data along every track whereas Problem RMP (Restricted Multipath) sets itself up by applying constraints to the span of the network tracks. As every coin has two sides to consider, similarly, this proposal also has some shortcomings. There is no proposition that allows the two problems i.e. Problem RMP (Restricted Multipath) and Problem KPR (K-Path Routing) to merge and provide a competent and fused result. This drawback if taken care of can result in a sophisticated and systematic manner for data transmission and also jamming concerns can be addressed proficiently. Another issue of concern is the circulated execution of algorithm RMP which can be experimented with in future and favorable solutions can be found out. It is also stated that multipath routing also grants other areas for cumulative investigation such as network safety issues, energy competence etc and extensive studies for the same are in progress.

Towards Robust Multi-Layer Traffic Engineering: Optimization of Congestion Control and Routing [33]: This after some justification is further being resolved by Jiayue He, Bresler, Mung Chiang, and Jennifer and mainly discusses about traffic engineering and how to effectively control the congestion by TCP (transmission control protocol). Traffic engineering is a method of optimizing the performance of a telecommunications network by dynamically analyzing, predicting and regulating the behavior of data transmitted over that network. The hosts increase or decrease the data sending depending on the network congestion. The study of traffic engineering is done using two approaches. First is the bottom-up approach that compares interaction between TCP congestion control and conventional traffic engineering practices and top down approach which creates a new multi layered dynamic distributed algorithm based on the following points.

- Stability – Are congestion control and routing in equilibrium?
- Optimality – does the equilibrium maximize the aggregate user utility, over the routing parameters and source rates?

For having both performance and robustness top- down approach is preferred with the following goals

- Distributed
- Robust
- Implementable
- Efficient

The TE model (traffic engineering model) and DATE algorithm (Distributed Adaptive)

Traffic Engineering are used in implementing traffic engineering.

The paper also discusses about network model which deals with focus on routing and congestion control.

A network can be taken as a set of L bidirectional links with finite capacities $c = (cl, l = 1, \ldots, L)$, which are shared by a set of N source-destination pairs, indexed by i. a source destination pair can be represented as "source i." . The traffic engineering practices can be known by given routing configuration as follows using $ul = Rlixi/cl$. Where matrix Rli represent the current routing that captures the fraction of i's flow that traverses each link l. TCP congestion control based on reverse engineering was also discussed with a goal of maximize aggregate user utility by varying x. Simulation of the TE model can be done by using a combination of the Matlab and MOSEK environments. Different experiments were conducted by the authors by evaluating two variants of TCP congestion control: $\alpha = 2$ (*e.g.*, TCP Reno) and $\alpha = 1$ (*e.g.*, TCP Vegas) for the cost function $f(ul)$. Initial experiments evaluate a simple N - node ring topology. The experiments were done using N nodes, N sources and N nodes, 1 destination, Access-Core topology and Abilene topology.

In the first experiment the capacity of link 1 was varied and plots the gap in aggregate utility for ring topologies with three, five, and ten nodes, where each node communicates with its clockwise neighbor. The two graphs were plotted for $\alpha = 1$ and $\alpha = 2$. In the second experiment graphs were plotted for Aggregate utility gap for the N - node, 1-destination ring and the results were noted as Traffic pattern can have a significant effect and TCP variants give the same trend in both the experiments with $\alpha = 1$. This paper also discusses analysis of TE model with theorems. The design goals for multi layered traffic engineering are

- *Satisfying User and Operator Objectives*
- *Adapting Routing on a Smaller Timescale*

To conclude distributed adaptive traffic engineering was discussed with DATE algorithm and stability and optimum results. There are also some cross layer studies which resembles TE model which share congestion control model or routing model. The DATE algorithm bears similarity to MATE [10], TeXCP [11] and REPLEX [14]. The difference is that other schemes do not consider congestion control explicitly.

V. Conclusion

A decisive analysis has been conducted on current state of the art. It is clear from our discussions that the algorithms stated are aiming to handle the congestion raised due to vast payloads on network, which may be due to flooding of packets or may be due to repeat requests on the basis of error correction techniques. Based on on the decisive analysis conducted in this paper we can state that congestion control models must handle the negative influences raised because of one or more factors like link failures, wireless losses, shared media issues and ACK packet load issues. This is clear from the investigations that all of the existing solutions are not robust to handle congestion raised, since the negative impact of above stated factors are vary from one to other network topologies. This leads us to conclude that new set of solutions are needed to overcome the limits in existing models in the view of robustness and adaptation.

References Références Referencias

1. K. Chandran, S. Raghunathan, S. Venkatesan, and R. Prakash. A Feedback Based Scheme for Improving TCP Performance in Ad-Hoc Wireless Networks. In ICDCS '98: Proceedings of the 18th International Conference on Distributed Computing Systems, pages 472–479, May 1998.
2. G. Holland and N. H. Vaidya. Analysis of TCP Performance over Mobile Ad Hoc Networks. In MobiCom '99: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pages 219–230, Aug. 1999. DOI: http://doi.acm.org/10.1145/313451. 313540.
3. D. Kim, C.-K. Toh, and Y. Choi. TCP-Bus: Improving TCP Performance in Wireless Ad-Hoc Networks. In ICC '00: Proceedings of the IEEE International Conference on Communications, volume 3, pages 1707–1713, June 2000. DOI: 10.1109/ICC.2000. 853785.
4. T. D. Dyer and R. V. Boppana. A Comparison of TCP Performance over Three Routing Protocols for Mobile Ad Hoc Networks. In MobiHoc '01: Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing, pages 56–66, 2001. DOI: 10.1145/ 501422.501425.
5. D. Sun and H. Man. ENIC – An Improved Reliable Transport Scheme for Mobile Ad Hoc Networks. In GLOBECOM '01: Proceedings of the IEEE Global Telecommunications Conference, volume 5, pages 2852–2856, Nov. 2001. DOI: 10.1109/GLOCOM. 2001.965950.
6. J. Zhou, B. Shi, and L. Zou. Improve TCP performance in Ad hoc network by TCP-RC. In PIMRC '03: Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, volume 1, pages 216–220, Sept. 2003. DOI: 10.1109/PIMRC. 2003. 1264264.
7. J. Liu and S. Singh. ATCP: TCP for Mobile Ad Hoc Networks. IEEE Journal on Selceted Areas in Communications (J-SAC), 19(7):1300–1315, July 2001. DOI: 10.1109/49.932698.
8. X. Yu. Improving TCP Performance over Mobile Ad Hoc Networks by Exploiting Cross-Layer Information Awareness. In MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking, pages 231–244, New York, NY, USA, 2004. ACM Press. DOI: http://doi.acm.org/10.1145/1023720.1023743.
9. F.Wang and Y. Zhang. Improving TCP Performance over Mobile Ad-Hoc Networks with Out-of-Order Detection and Response. In MobiHoc '02: Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing, pages 217–225, 2002. DOI: http://doi.acm.org/10. 1145/513800. 513827.
10. F. Klemm, Z. Ye, S. V. Krishnamurthy, and S. K. Tripathi. Improving TCP Performance in Ad Hoc Networks using Signal Strength based Link Management. Elsevier Ad Hoc Networks, 3(2):175–191, 2005. DOI: 10.1016/j.adhoc.2004.07.005.
11. M. G¨unes¸ and D. Vlahovic. The Performance of the TCP/RCWE Enhancement for Ad-Hoc Networks. In ISCC '02: Proceedings of the 7th IEEE International Symposium on Computers and Communication, pages 43–48, 2002. DOI: 10.1109/ISCC.2002.1021656.
12. Z. Fu, B. Greenstein, X. Meng, and S. Lu. Design and Implementation of a TCP-Friendly Transport Protocol for Ad Hoc Wireless Networks. In ICNP '02: Proceedings of the 10th IEEE International Conference on Network Protocols, pages 216–225, Nov. 2002.
13. Z. Fu, X. Meng, and S. Lu. A Transport Protocol for Supporting Multimedia Streaming in Mobile Adhoc Networks. IEEE Journal on Selected Areas in Communications, 21(10):1615–1626, Dec. 2003. DOI: 10.1109/JSAC.2003.816461.
14. R. de Oliveira, T. Braun, and M. Heissenb¨uttel. An Edge-based Approach for Improving TCP in Wireless Mobile Ad Hoc Networks. In DADS '03: Proceedings of the Conference on Design, Analysis and Simulation of Distributed Systems, Orlando, USA, Mar. 2003.

15. Z. Fu, P. Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla. The Impact of Multihop Wireless Channel on TCP Throughput and Loss. In INFOCOM '03: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, volume 3, pages 1744–1753, Apr. 2003.
16. S. Xu and T. Saadawi. Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc network. IEEE Communications Magazine, 39(6):130–137, June 2001. DOI: 10.1109/35.925681.
17. C. de M. Cordeiro, S. R. Das, and D. P. Agrawal. COPAS: Dynamic Contention-Balancing to Enhance the Performance of TCP over Multi-hop Wireless Networks. In Proceedings of the 10th International Conference on Computer Communication and Networks (IC3N), pages 382–387, Miami, FL, USA, Oct. 2002.
18. Z. Ye, S. Krishnamurthy, and S. Tripathi. Use of Congestion-Aware Routing to Spatially Separate TCP Connections in Wireless Ad Hoc Networks. In Proceedings of the 1st International Conference on Mobile Ad hoc and Sensor Systems (MASS), Fort Lauderdale, FL, USA, Oct. 2004.
19. E. Altman and T. Jim´enez. Novel Delayed ACK Techniques for Improving TCP Performance in Multihop Wireless Networks. In PWC '03: Proceedings of the IFIP-TC6 8th International Conference on Personal Wireless Communications, pages 237–250, Sept. 2003. DOI: 10.1007/b12004.
20. R. de Oliveira and T. Braun. A Dynamic Adaptive Acknowledgment Strategy for TCP over Multihop Wireless Networks. In INFOCOM '05: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, volume 3, pages 1863–1874, Mar. 2005. DOI: 10.1109/INFCOM.2005.1498465.
21. M. Sugano and M. Murata. Performance Improvement of TCP on a Wireless Ad Hoc Network. In VTC '03-Spring: Proceedings of the 57th IEEE Vehicular Technology Conference, volume 4, pages 2276–2280, Apr. 2003. DOI: 10.1109/VETECS.2003.1208794.
22. M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," Proc. of INFOCOM, vol. 2, pp. 836–843, 2003.
23. B. Awerbuch, D. Holer, and H. Rubens, "High throughput route selection in multi-rate ad hoc wireless networks," Proc. of the First working conference on wireless on-demand network systems, pp. 251–269, 2004.
24. S. Zhao, Z. Wu, A. Acharya, and D. Raychaudhuri, "PARMA: a PHY / MAC aware routing metric for ad-hoc wireless networks with multi-rate radios," Proc. of WoWMoM 2005, pp. 286–292, Jun. 2005.
25. J. Kang, Y. Zhang, and B. Nath, "Accurate and energy-efficient congestion level measurement in ad hoc networks," Proc. of WCNC, vol. 4, pp. 2258–2263, Mar. 2005.
26. Y. J. Lee and G. F. Riley, "A workload-based adaptive loadbalancing technique for mobile ad hoc networks," Proc. of WCNC, vol. 4, pp. 2002–2007, Mar. 2005.
27. Haiying Shen; Cheng-Zhong Xu; , "Elastic Routing Table with Provable Performance for Congestion Control in DHT Networks," Parallel and Distributed Systems, IEEE Transactions on , vol.21, no.2, pp.242-256, Feb. 2010 doi: 10.1109/TPDS.2009.51 URL : http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4815223&isnumber=5372118
28. Fernando Paganini, Enrique Mallada: A unified approach to congestion control and node-based multipath routing. IEEE/ACM Trans. Netw. 17(5): 1413-1426 (2009).
29. Tsung-Hsien Lee; Ting-Chi Wang; , "Congestion-Constrained Layer Assignment for Via Minimization in Global Routing," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.27, no.9, pp.1643-1656, Sept. 2008 doi: 10.1109/TCAD.2008.927733. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4603083&isnumber=4603069
30. Jenn-Yue Teo; Yajun Ha; Chen-Khong Tham; , "Interference-Minimized Multipath Routing with Congestion Control in Wireless Sensor Network for High-Rate Streaming," Mobile Computing, IEEE Transactions on , vol.7, no.9, pp.1124-1137, Sept. 2008 doi: 10.1109/TMC.2008.24. URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4453827&isnumber=4573014
31. Tinoosh Mohsenin, Dean N. Truong, and Bevan M. Baas. 2010. A low-complexity message-passing algorithm for reduced routing congestion in LDPC decoders. Trans. Cir. Sys. Part I 57, 5 (May 2010), 1048-1061.DOI=10.1109/TCSI.2010.2046957 http://dx.doi.org/10.1109/TCSI.2010.2046957
32. Banner, R.; Orda, A.; , "Multipath Routing Algorithms for Congestion Minimization," Networking, IEEE/ACM Transactions on, vol.15, no.2, pp.413-424, April 2007 doi: 10.1109/TNET.2007.892850 URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4154742&isnumber=4154740

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Awareness of Electronic Banking In Pakistan

By Yasir Hassan, Farzan Yahya, Muazzam Amin, Umar Farooq Arshad

University of Lahore

Abstract - This research paper focuses on awareness of electronic banking in Pakistan. Electronic banking is today's need as it provides easy way to monitor an account. It reduces cost and save our precious time. There are also some limitations of e-banking as it requires skills to operate it and passwords can be hacked. This research paper discussed that people of Pakistan are well aware of electronic banking but they are not satisfied with the services provided by the bank and they also feel risk in that service that's why they trust more on employees. It contains different demographics and their association with research questions. This paper also includes recommendations in which it describes that what strategies the banks should choose to increase the satisfaction of customers.

Keywords : *e-banking, SMS banking, ATM.*

GJCST Classification : *K.4.4, K.4.2*

Strictly as per the compliance and regulations of:

Awareness of Electronic Banking In Pakistan

Yasir Hassan[α], Farzan Yahya[Ω], Muazzam Amin[β], Umar Farooq Arshad[Ψ]

Abstract - This research paper focuses on awareness of electronic banking in Pakistan. Electronic banking is today's need as it provides easy way to monitor an account. It reduces cost and save our precious time. There are also some limitations of e-banking as it requires skills to operate it and passwords can be hacked. This research paper discussed that people of Pakistan are well aware of electronic banking but they are not satisfied with the services provided by the bank and they also feel risk in that service that's why they trust more on employees. It contains different demographics and their association with research questions. This paper also includes recommendations in which it describes that what strategies the banks should choose to increase the satisfaction of customers.

Keywords : *e-banking, SMS banking, ATM.*

I. Introduction

The developments taking place in information and communication technology are increasing competition in financial institutions worldwide. Thus, the deployment of advanced technologies is essential to achieve a competitive edge. In the world of banking, the development of information technology has an enormous effect on development of more flexible payments methods and more-user friendly banking services. Recently, the banking industry was highly affected by the technology evolution that transformed the way banks deliver their services, using technologies such as automated teller machines, phones, the Internet, credit cards, and electronic cash. In line with global trends, electronic banking in Pakistan has been undergoing many changes.

Electronic banking is a term for the process by which a customer may perform banking transactions electronically without visiting a brick-and-mortar institution. Electronic banking refers to systems that enable bank customers to access accounts and general information on bank products and services through a personal computer (PC) or other intelligent device. There are many benefits of e-banking as it provides easy way to monitor an account, we can shop, pay bills, buy items at auction, and transfer money from anywhere at any time, it reduces costs, it saves time, and vice versa.

This study focuses on awareness of electronic banking in Pakistan. Our objectives of this research were to:

1. Check awareness of electronic banking in Pakistan.
2. Check why people are not aware of electronic banking.
3. Know if the people are aware of electronic banking then to how much extent.
4. Check whether banks of Pakistan are providing electronic banking or not.

This research is important because it tells bankers that how much people in Pakistan are aware of electronic banking and to how much extend. It also tells them to improve their e-banking services in Pakistan.

II. Literature Review

According to Hagel and Hewlin (1997) the Internet banking became very attractive to customers and lots of banks because the technology is being accepted by them and they can now understand and have information about the complex products. Nowadays banks are also facing a lot of competition and need a high market share and provide better services to its customers so that they can attract the new customers and old customers do not try to leave them.

The concept of electronic banking has been defined in many ways (e.g. Daniel, 1999). According to Karjaluoto (2002) electronic banking is a construct that consists of several distribution channels. Daniel (1999) defines electronic banking as the delivery of banks' information and services by banks to customers via different delivery platforms that can be used with different terminal devices such as a personal computer and a mobile phone with browser or desktop software, telephone or digital television. Electronic banking consists of any system that uses electronic signals to replace people or paper. There is already a significant amount of electronics in the traditional demand deposit system; virtually all checks are processed by computer, and banks often settle their obligations for the checks presented to them by an electronic funds transfer. Thornton and White (2001) compared several electronic distribution channels available for banks in US and concluded that customer orientation – towards convenience, service, technology, change, knowledge about computing and the Internet – affected the usage of different channels.

Author [α]: Yasir Hassan, Lecturer at Lahore Business School, the University of Lahore, (54000).
Email : yasir.uol@hotmail.com
Author [Ω]: Farzan Yahya, Student, Lahore Business School, the University of Lahore, (54000).
Email : Farzan.yahya@yahoo.com
Author [β]: Muazzam Amin, Student, Lahore Business School, the University of Lahore, (54000).
Email : muazzam_amin@yahoo.com
Author [Ψ]: Umer farooq Arshad, Student, Lahore Business School, the University of Lahore, (54000).
Email : umer_farooq_arshad@yahoo.com

Howcroft et al., (2002) found that the most important factors encouraging consumers to use online banking are lower fees followed by reducing paper work and human error, which subsequently minimize disputes (Kiang et al., 2000). It's essential for the banks to have the official bank website providing the possibility to do transactions so that banks can be qualified as providing the online banking services (Pikkarainen et al., 2004). According to Giglio (2002) and Robinson (2000) for delivering banking products the cheapest delivery channel can be done only through the Online Banking. According to Karjaluoto et al. (2002) with the help of online banking services, the branch networks of banks have reduced and also the staff for working in banks and customers are satisfied to use the online banking services as it will save a lot of time and effort to go to branch of bank and perform these transactions. So the main reason behind accepting the online banking a service is the time, cost saving and freedom from the place (Polatoglu and Ekin, 2001).

Byers and Lederer, (2001) concluded that it was changing consumer attitudes rather than bank cost structures that determines the changes in distribution channels; they added that virtual banks can only be profitable when the segment that prefers electronic media is approximately twice the size of the segment preferring street banks.

According to theorists (Walfried et al., 2005) customer evaluation of the electronic services is influenced by attributions of success and failure in inter personal service situations. The use of electronic banking has removed the banking personnel that facilitate the transactions and has placed additional responsibilities on the customers to transact with the service.

Although, electronic banking provides many opportunities for the banks, it is also the case that the current banking services provided through Internet are limited due to security concerns, complexity and technological problems (Sathye, 1999: Mols, 1999)

Hewer and Howcroft (1999) used the term trust to describe a measure of risk. Suganthi et al., (2001) viewed risk in the context of security concerns and risk in the context of trust in one's bank. Finally, a number of studies found trust and perceived risks have a significant positive influence on commitment (Bhattacherjee, 2002; Mukherjee and Nath, 2003) and ultimately leads towards overall satisfaction (Rexha et al., 2003).

III. Methodology

In this research data is conducted by using questionnaire as a data collection instrument, in which questions were asked starting from their command on computer and internet usage, following with the awareness of electronic banking and its types, about the benefits of e-banking and then some questions regarding their bank services and the security provide by these banks. The questionnaire includes close-ended questions and they were on likert scale of utilizing a five point categorical. Sample of 400 respondents was selected to complete the research using simple random sampling technique.

IV. Empirical Results

There is "table 1" which shows gender wise association with research questions:

Table 1 : Age association with research questions

No.	Question	Gender		Total	P-values
		Male	Female		
1	Aware of electronic banking	196	200	396	0.011
2	Your bank provides electronic banking facility	190	195	385	0.000
3	Aware of SMS banking	196	195	391	0.036
4	Aware of INTERNET banking	189	198	387	0.010
5	Aware of ATM card usage	193	198	391	0.054
6	Aware of PHONE banking	193	199	392	0.272
7	Satisfied with the E-banking services provide by bank	192	198	390	0.008
8	Satisfied with the security of E-banking provide by bank	191	195	386	0.068
9	Want to use e-banking facilities in future	194	197	391	0.000

Table 1 shows the different questions and their association with gender. First question is about awareness of electronic banking. Total respondents who answered this question were 396 in which 196 were male respondents and 200 were female respondents. Its result shows that awareness of electronic banking is dependent on gender. Next question was answered by 385 total respondents from which the amount of male and female were 190 and 195 respectively. Its results show that providing e-banking services by banks were

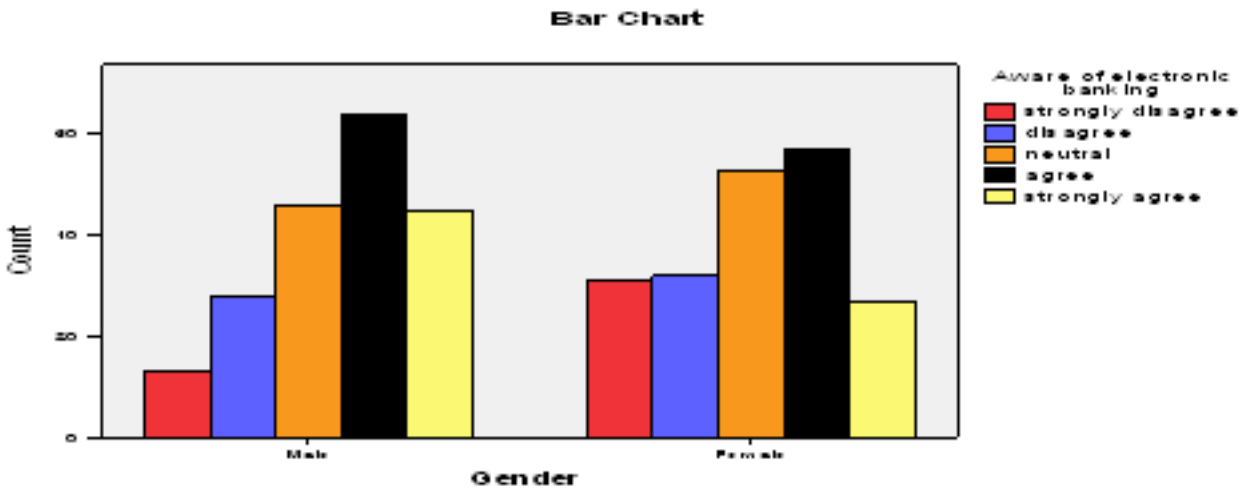
highly dependent on gender. Third one is about awareness of SMS banking which was answered by 196 male and 195 female respondents. Its result shows that awareness of SMS banking is also dependent on gender. Now the next question is about Internet banking awareness which was answered by 189 male respondents and 198 female respondents and its result shows that awareness of internet banking is dependent on gender. Next question is about awareness of ATM card usage. Total 391 respondents answered that question from which the amount of male and female were 193 and 198 respectively. It results shows that awareness of ATM card usage is not dependent on gender. Now there is a question about Phone banking awareness so total 392 respondents answered that question from which 193 were male and 199 were female and its result shows that phone banking awareness is not dependent on gender. Now the seventh question was asked by respondents about their satisfaction about e-banking services providing by their banks and total 390 answered that question and its result shows that satisfaction about e-banking services providing by banks is depends on gender. Now there is second last question which was answered by 191 male and 195 female respondents and its result shows that satisfaction about security of e-banking providing by banks is not depends on gender. Last question is about the using e-banking service in future and its results are highly dependent on gender.

Table 2

		Aware of electronic banking					
		strongly disagree	disagree	neutral	agree	strongly agree	Total
Gender	Male	13	28	46	64	45	196
	Female	31	32	53	57	27	200
Total		44	60	99	121	72	396

Now there is a table no.2 which shows the description of the question, "Aware of electronic banking". It shows the amount of respondents from strongly disagree to strongly agree. The highest figures of people are agreed about awareness of e-banking. In male respondents 64 were agree which is 33% of total male respondents. Likewise, in female respondents 57 were agreeing about it which is 29% of total females.

Chart No. 1



Now there is a simple bar chart of the table no. 2 which shows clearly that very less amount of people are strongly disagree or disagree, mostly are agree and strongly agree that they know well about e-banking. Now there is a table no. 3 which shows the association of education level with different research questions:

Table No. 3 : Education association with research questions

No.	Question	Primary	High School	Secondary	Graduate	Post Graduate	Frequency	P-values
1	Aware of electronic banking	8	7	10	195	164	384	0.000
2	Use debit or credit cards for online transactions	8	6	8	191	163	376	0.008
3	Aware of ATM card usage	8	7	10	193	163	381	0.030
4	Aware of PHONE banking	8	7	10	194	161	380	0.155
5	Feel risk about hacking of passwords	8	7	10	192	161	378	0.555
6	Satisfied with the E-banking services provide by bank	8	7	9	191	163	378	0.269
7	More trust on the bank employees with communication than E-banking	8	7	10	191	163	379	0.680
8	Want to use e-banking facilities in future	8	7	9	192	163	379	0.969

In table no. 3 there is a demographic of education whose relationship with research questions is given in above table. First there is a question about awareness of electronic banking. Total respondents were 384 from which the education of 8 respondents were primary, 7 people had just completed high school, 10 respondents complete their secondary education, 195 were graduate and 164 were post graduate. Its result shows that awareness of electronic banking is highly dependent on education. Secondly there is a question about use of debit or credit cards for online transactions. Total respondents were 376 from which respondents were from different education level which are given on above table. Next question is about awareness of ATM card usage, which is responded by 381 people and its result shows that it depends on education. Next question is about Phone banking awareness, which is answered by 380 people and it result shows that it is not dependent on education. Fifth question is about feeling of risk for hacking of passwords, 378 people responds that and its result tells that it is also not dependent on education and next there question are also not dependent on education.

Table No. 4

	Feel risk about hacking of passwords					Total
Education level	strongly disagree	disagree	neutral	agree	strongly agree	
Primary	0	1	1	2	4	8
High school	0	0	2	2	3	7
Secondary	0	2	5	1	2	10
Graduate	11	26	59	57	39	192
Post graduate	6	31	39	46	38	160
Total	17	60	106	108	86	377

In this table results shows that mostly people are graduate whose percentage is 51% percentage of total respondents in which highest percentage of people have neutral point of view about that. Following is the table no. 5 which shows the association of age with different research questions:

Table No. 5 : Age association with research questions

No.	Question	Age						Frequency	P-values
		Below 19	20 - 29	30 - 39	40 - 49	50 - 59	Above 60		
1	Have full command on computer	91	241	30	16	10	3	391	0.034
2	Involved in banking transactions	91	239	30	14	14	3	391	0
3	Awareness of usefulness of electronic banking	91	238	30	16	14	3	392	0.004
4	Your Banks provides online technical assistance or 24 helpline	88	240	30	16	14	3	391	0.002
5	Use debit or credit cards for online transactions	90	237	30	16	14	3	390	0
6	Aware of INTERNET banking	91	238	29	14	14	3	389	0.443
7	Aware of PHONE banking	91	240	30	16	14	3	394	0.105
8	It provides easy way to monitor a account	91	242	30	16	14	3	396	0.019
9	Feel risk about hacking of passwords	91	239	30	16	13	3	392	0.089
10	Have easy access to ATM machine	88	240	30	16	13	3	390	0
11	Satisfied with the E-banking services provide by bank	90	240	30	16	13	3	392	0.347
12	Satisfied with the security of E-banking provide by bank	89	239	28	16	13	3	388	0.004
13	More trust on the bank employees with communication than E-banking	90	241	30	16	13	3	393	0.036

Above table shows the relationship of research questions with age. In first question it is asked about command on computer. This question was answered by total 391 respondents from which from which 91 respondents had age less than 19. The respondents from age group 20-29 were 241 which is highest amount, 30 respondents were from age group 30-39. Respondents who had age 40-49 were 16, 10 respondents' age was lied in 50-59 and only 3 respondents were above 60. Its result shows that having full command on computer is dependent on age. Next question is about involvement in banking transaction which was answered by total 391 respondents and the amount of responses by different age groups are shown in table. Its result shows that involvement in banking transaction is highly dependent on age. Third question is about awareness of usefulness of e-banking which is answered by total 392 respondents and its results shows that it depends on age. Next question is about providing of online technical assistance or 24 hr. helpline which responds by 391 respondents and its results tells that it is also dependent on age. Moreover, there is question about using of debit or credit cards for online transactions which responds by 390 people, which shows the result that it depends on age. Next questions are about awareness of internet banking and phone banking whose result shows they are not dependent on age. Further there is a question about its easy way of monitoring an account which responds by 396 people and its result shows that it depends on age.

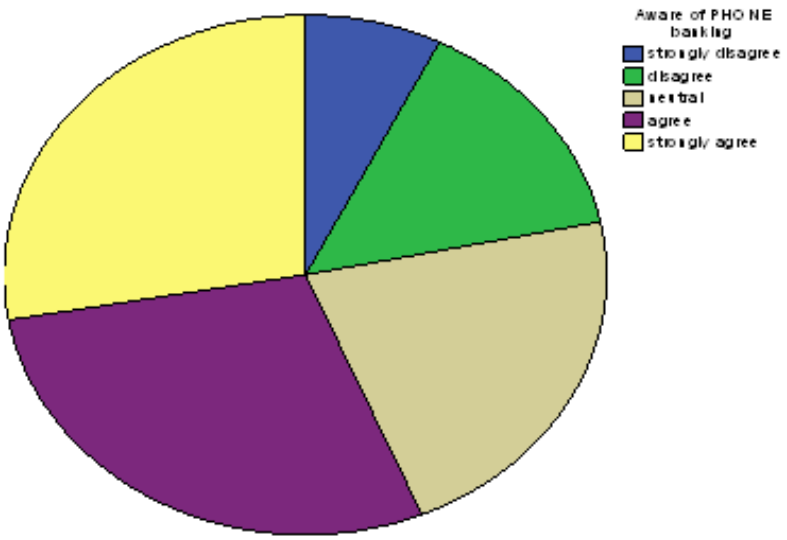
Further there is a question about feeling of risk about hacking of passwords, 392 people respond it and result shows that it is independent to age. Now there is a question about easy access to ATM machine whose result shows that it is highly dependent on age. Satisfaction of e-banking services provided by banks is independent on age. However, Satisfaction of security providing for e-banking by banks is dependent on age. Last question is about trust on bank employees more than e-banking whose result shows that it is also dependent on age.

Table 6

Age	Aware of PHONE banking					
	strongly disagree	disagree	neutral	agree	strongly agree	Total
<19	3	14	26	28	20	91
20-29	21	37	45	72	65	240
30-39	2	2	4	10	12	30
40-49	0	3	5	2	6	16
50-59	2	1	5	1	5	14
60>	1	0	2	0	0	3
Total	29	57	87	113	108	394

The table above describes about awareness of Phone banking with respect to age. In this table greatest amount of respondents are agree that they are familiar with phone banking. From age group ranges from 20 to 29 has greatest figure of 72 about knowing of phone banking which is 30% of the age group who had age 20-29.

Chart No. 2



This pie chart is of above table which shows the data in different areas which ranges from tiny to greatest. The area of agree is 29% and strongly agree is 27% of the total data. Lastly, there is a table no. 7 which shows the association and relation of income level with research questions.

Table No. 7 : Income association with research questions

No.	Question	Income						Total	P-values
		<10000	10001-20000	20001-30000	30001-40000	40001-50000	50000>		
1	Have full command on internet usage	141	57	36	39	33	31	337	0.007
2	Satisfied with the speed of internet connection	138	57	36	38	32	31	332	0.036
3	Aware of electronic banking	141	58	36	39	34	31	339	0
4	Aware of INTERNET banking	138	56	36	39	31	30	330	0.017
5	It saves time	141	58	36	39	34	28	336	0.026
6	Satisfied with the security of E-banking provide by bank	139	56	36	39	32	28	330	0.001
7	Use debit or credit cards for online transactions	138	56	36	39	34	28	331	0.462
8	Aware of SMS banking	140	57	36	39	33	30	335	0.22
9	It reduces cost	137	57	36	38	33	29	330	0.195

Above table shows the involvement of monthly income with research questions. First there is a question about command on internet usage, 337 respondents answers this question from which 141 respondents had income below 10000, 57 respondents had income between 10001-20000, 36 respondents have income between 20001-30000, 39 had between 30001-40000, 33 had income between 40001-50000 and lastly 31 respondents have income above 50000. Its result shows that command on internet usage is dependent on income. Next one is satisfaction with the speed of their internet connection, 332 people responds this and their income are given in above table respectively. Its result shows that it is also dependent on income. Third question is about awareness of e-banking, 339 responds that question and it results shows that awareness of e-banking is highly dependent on income. Next question is about awareness of internet banking, total 330 respondents answered it and it result shows that it is also dependent on income. Now there is a question that it saves time, total 336 respondents answered it and it result shows that it is dependent on income. Satisfaction with the security of e-banking services is also dependent on income.

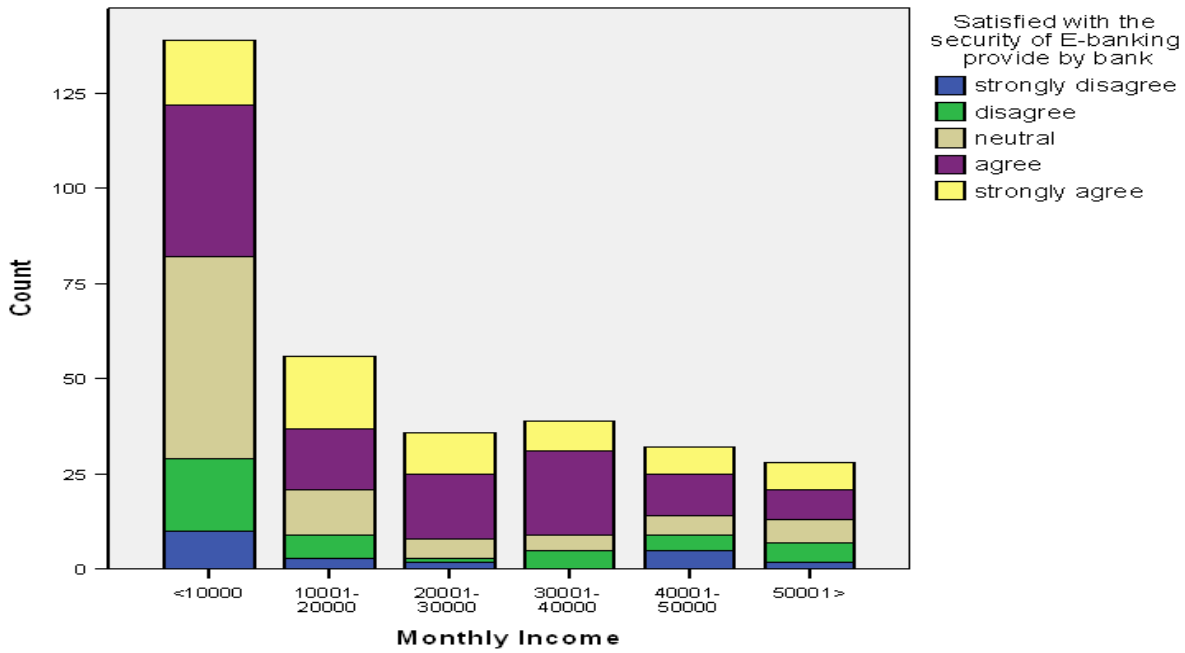
However usage of debit or credit cards for online transaction is not dependent on income. Next one is awareness of SMS banking, 335 people responds this and it results shows that it is also not dependent. Next question is about it cost reduction benefit and it result also shows that it is not dependent on income.

Table No. 8

Monthly Income	Satisfied with the security of E-banking provide by bank					Total
	strongly disagree	disagree	neutral	agree	strongly agree	
<10000	10	19	53	40	17	139
10001-20000	3	6	12	16	19	56
20001-30000	2	1	5	17	11	36
30001-40000	0	5	4	22	8	39
40001- 50000	5	4	5	11	7	32
50001>	2	5	6	8	7	28
Total	22	40	85	114	69	330

Above table shows that, to how much extent people are satisfied or unsatisfied with the security of e-banking provided by their banks. The highest percentage of people are agree about it and overall results shows that people are satisfied with the security of e-banking provided by their banks.

Chart No. 3



This is a stacked bar chart of the above table which shows that mostly respondents have income less than Rs. 10, 000. The highest stack is from first bar shows that 38% respondents have neutral point of view.

V. Conclusion and Recommendations

In conclusion it is stated that Electronic banking is very useful and today's need as it provides easy way to monitor your account. However, results of this research show that most of the people of Pakistan are well aware of electronic banking but they are not satisfied with the e-banking services providing by banks of Pakistan. They want improvement in security level and services of electronic banking. That's why they trust more on employees than e-banking.

It is recommended that in a country like Pakistan, there is need for providing better and customized services to the customers. Banks should introduce better technologies and methods that might contain little risk. There are some strategies should be applied by banks.

- Banks should ensure that electronic banking is safe and secure for financial transaction like as traditional banking.
- Banks should organize seminar and conference to educate the customer regarding uses of electronic banking as well as security and privacy of their accounts.
- Some customers are hindered by lack of computer skills. They need to be educated on basic skills required to conduct online banking.
- Banks must emphasize the convenience that online banking can provide to the people, such as avoiding long queue, in order to motivate them to use it.
- Banks must emphasize the cost saving that online can provide to the people, such as reduce transaction cost by use of online banking.

References Références Referencias

1. Daniel, E. (1999). Provision of electronic banking in the UK and the Republic of Ireland. *International Journal of Bank Marketing*. 17(2); 72-82.
2. Thornton, J. and White, L. (2001). Customer orientations and usage of financial distribution channels, *Journal of Services Marketing*. 15(3); 168-85.
3. Howcroft, B., Hamilton, R. and Hewer, P. (2002). Consumer attitude and the usage and adoption of home-based banking in the United Kingdom. *The International Journal of Bank Marketing*. 20(3); 111-21.
4. Kiang, M.Y., Raghu, T.S. and Hueu-Min Shang, K. (2000). Marketing on the Internet – who can benefit from an online marketing approach? *Decision Support Systems*. 27(4); 383-93.
5. Byers, R. and Lederer, P.J. (2001). Retail bank services strategy: a model of traditional, electronic, and mixed distribution choices. *Journal of Management Information Systems*. 18(2); 133-56.
6. Sathye, M. (1999). Adoption of Internet banking by Australian consumers: an empirical investigation. *International Journal of Bank Marketing*. 17(7); 324-34.

7. Hewer, P. and Howcroft, B. (1999). Consumers distribution channel adoption and usage in the financial services industry: a review of existing approaches. *Journal of Financial Services Marketing*. 3(4); 344-58.
8. Suganthi, B. and Balachandran, P. (2001). Internet banking patronage: an empirical investigation of Malaysia. *Journal of Internet Banking and Commerce*. 6(1).
9. Bhattacherjee, A. (2002). Individual trust in online firms: scale development and initial test. *Journal of Management and Information Systems*. 19(1); 211-41.
10. Rexha, N., Philip, R., Kingshott, J. and Shang-Aw, A., S. (2003). The impact of the relational plan on adoption of electronic banking. *Journal of Services Marketing*. 17(1); 53-67.
11. Walfried M. Lassar, Chris Manolis and Sharon S. Lassar, (2005). The relationship between consumer innovativeness, personal Characteristics, and online banking adoption. *International Journal of Bank Marketing*. 23(2); 176 – 199.
12. Hagel, J., Hewlin, T. and Hutching, T. (1997). Retail banking: caught in a Web. *The McKinsey Quarterly*. 2; 42-55.
13. Pikkarainen, T., Pikkarainen, K. Karjaluoto, H. and Pahnila, S. (2004). Consumer acceptance of online banking: an extension of the technology acceptance model. *Internet Research*, 14(3); 224-235.
14. Giglio, V. (2002). Privacy in the world of cyber banking: emerging legal issues and how you are protected. *The Secured Lender*. 14 (3); 48-60.
15. Karjaluoto, H., Mattila, M. and Pento, T. (2002). Factors underlying attitude formation towards online banking in Finland. *International Journal of Bank Marketing*, 20(6), 261-272.
16. Polatoglu, V.N. and Ekin, S. (2001). An empirical investigation of the Turkish consumers acceptance of Internet banking services. *International Journal of Bank Marketing*. 19(4); 156-165.
17. Dixit, N., and Datta, S.K. (2010). Acceptance of E-banking among Adult Customers: An Empirical Investigation in India. *Journal of Internet Banking and Commerce*. 15 (2).

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Frequent Pattern Mining With Closeness Considerations: Current State Of The Art

By Anurag Choubey, Dr. Ravindra Patel,Dr. J.L. Rana

Rajiv Gandhi Technological University

Abstract - Due to rising importance in frequent pattern mining in the field of data mining research, tremendous progress has been observed in fields ranging from frequent itemset mining in transaction databases to numerous research frontiers. An elaborative note on current condition in frequent pattern mining and potential research directions is discussed in this article. It's a strong belief that with considerably increasing research in frequent pattern mining in data analysis, it will provide a strong foundation for data mining methodologies and its applications which might prove a milestone in data mining applications in mere future.

GJCST-C Classification : H.2.8

Strictly as per the compliance and regulations of:

Frequent Pattern Mining With Closeness Considerations: Current State Of The Art

Anurag Choubey [α], Dr. Ravindra Patel [Ω],Dr. J.L. Rana [β]

Abstract - Due to rising importance in frequent pattern mining in the field of data mining research, tremendous progress has been observed in fields ranging from frequent itemset mining in transaction databases to numerous research frontiers. An elaborative note on current condition in frequent pattern mining and potential research directions is discussed in this article. It's a strong belief that with considerably increasing research in frequent pattern mining in data analysis, it will provide a strong foundation for data mining methodologies and its applications which might prove a milestone in data mining applications in mere future.

I. Introduction

Frequent patterns are itemsets, subsequences, or substructures which are often found as a repetitive pattern in any database, with rate of recurrence depending upon individual verge. For example: Itemsets are sets of item such as vegetables and spices that are often observed in a transaction data. Subsequence is a chronological pattern habitually followed, such as buying a camera, then a camera carry-bag and then lenses. Substructure denotes particular structural forms which may be pooled with data of itemsets or subsequences. Regular occurrences of a substructure in a data-base is called structural pattern. For enhancement in results for data indexing, classification, clustering and other data mining tasks, discovery of frequent pattern mining has turned out to be an essential data mining task.

Agarwal et al. (1993) was the pioneer, proposing frequent pattern mining for market basket analysis in form of association rule mining which examines customer trading behavior by ruling out relations between dissimilar items that customers buy. These statistics facilitate a retailer to be careful regarding further marketing. This relevance provided a podium for various other research publications on various kinds of extensions and applications, and hence it's time to examine what more needs to be inspected to make this expertise a keystone approach for all supplementary demanding research issues.

II. Mining Patterns with Closeness as Consideration

With large data, mining generates colossal frequent patterns satisfying min_sup threshold, principally when min_sup is set low, because with frequent patterns comes along frequent sub patterns too. Hence the principal of closed frequent pattern mining and maximal frequent pattern mining came to subsistence.

The mining of frequent closed itemsets was proposed in [2], where an Apriori-based algorithm called A-Close for such mining was presented. Other closed pattern mining algorithms include [3, 4, 5, 6, and 7]. Pattern closeness identification is the big obstacle in closed or maximal pattern mining. The usage of hashed transaction ids is on e significant approach to overcome this complication [4] and the other approach is to construct hierarchical tree structure using discovered patterns as nodes [5, 7, and 8]. Mining closed itemsets provides an interesting and important alternative to mining frequent itemsets since it inherits the same analytical power but generates a much smaller set of results. Better scalability and interpretability is achieved with closed itemset mining.

Bayardo et al [9] proposed level-wise, breadth-first search method to find max-itemset. This approach includes subspace reduction by pruning process that prunes frequency of superset and infrequency of subset for search space reduction. Transaction id list compression approach was discussed in another efficient method MAFIA [10] that targets the counting efficiency. NP-Hard is the complexity of reciting maximal itemsets.

III. Mining Patterns with Sequence as Consideration

Sequential pattern mining, the mining of frequently occurring ordered events or subsequences as patterns in set of ordered elements or events that referred as sequence dataset, was first bring out by Agarwal et al[11], and has become an important problem in data mining. We first introduce the preliminary concept about sequential patterns.

Generalized Sequential Patterns (GSP) [12] is opt to multiple pass approach and uses the down the level closeness of patterns, and other consideration is the candidature maintenance approach SPADE [13] proposed by Zaki et al is a sequential pattern mining model that works based on vertical format, which is an extension to Eclat[14] and CHARM[4].. The use of vertical data format minimizes the passes through dataset. Since the search process relies on breadth first

Author [α Ω β] *: Rajiv Gandhi Technological University, Bhopal, India*
E-mail : [α] *anuragphd11@gmail.com*

and pruning is apriori based, which causes complex candidature maintenance in order to manage longer sequences.

Prefix Span[15, 8] was developed by Pei et al that works in a divide-and-conquer way. In this model each sequential pattern considers as prefix and projects the complete set of sequential patterns and later partitioned into different subsets according to different prefixes. Finally projected datasets are mined recursively to find sub sequent patterns.Empirical study concludes that the prefixspan[15] is the best when compared to GSP and SPADE. And the study also shows that PrefixSpan has the best overall performance. The CloSpan[16] was proposed by Yan et al.The method is based on a property of called equivalence of projected databases, where two projected sequence datasets with same prefix and same size in total number of items, this model can prune the non-closed sequences from further consideration during the mining process.BIDE[17], a bidirectional search for mining frequent closed sequences was developed by Wang and Han et al, which can further perform projecting sequences in bidirectional.

IV. Mining Frequent Itemsets and no Candidature Maintenance

The Apriori principle is significant in reducing the size of candidate sets. But multiple passes through dataset and maintaining large number of candidate sets are two major obstacles of the apriori principle. In the context of these obstacles Han et al[18] proposed a model called FP-growth that mines frequent itemsets without using candidature maintenance.

FP-growth works by compressing the given dataset by ordering items as extended prefix-tree structure. FP-growth targets the decomposition of given dataset and mining activities, in this process the FP-growth endorsed by divide-and-conquer approach. It uses an incremental pattern fragmentation to eliminate the candidature maintenance (generation and testing), which is used by Apriori.

Empirical studies in literature demonstrate that fp-growth considerably minimizes the search time. The FP-growth algorithm performs well by searching for shorter frequent patterns recursively and then concatenates least frequent items as suffix. There are suitable alternative and extensions too for this approach, few of them are Agarwal et al[19], H-mine[20] Liu et al. [21, 22] and Grahne [6].

V. Current State of Art

Authors Ya-Han Hu, Fan Wu, and Tzu-Wei-Yeh [23] have presented an in-detail thesis about Market based Analysis. This dat helps in finding correlation between purchasing items in databases. This study extends RFM analysis into mining process which measures frequent patterns extending in mining. Initially a RFPM tree is prepared depending upon the RFM-patterns, to compress and store entire transactional database and later RFPM growth is developed, hence efficiently find RFM patterns.

Association rule mining, usefull for discovering relationships among items or events in various app domains, discovers complete set of frequent patterns. Here, instead of filtering patterns after discovery, constraints (specific algorithms possessing meaningfull info) are directly pushed in into process of discovering and hence desired patterns are discovered.

Constraints from RFM analysis examine uniqueness from frequent patterns into mining algorithm. A RFM pattern is one in which recency score, frequency score and monetary score which satisfies minimum thresholds. Wu et al. discussed the recency problem about the change of data distribution between the past data and the new data. Emerging patterns as introduced by Dong and Li discover significant pattern changes in datasets with different time periods. In mining process, it cant reflect requirement of RFM completely. RFM-Apriori defined the recency constraint of the last transaction that should satisfy the recency threshold. The paper proposed an algorithm for mining patterns satisfying RFM constraints using a fixed time gap. However their approach the different stripe lengths of recency measuring the importance of transaction. This method better interfaced and also measured RFM importance of patterns. . For a set I of items in a database, to measure the scores of RFM of itemsets, the database should contain the more information of a transaction X, like <tx,(a1 , qa1) (2 a , qa2)…(am , qam)>, where tx stands for the transaction time when transaction X occurs, the m pairs, (ai, qai), where $1 \leq i \leq m$ and ai Є I, denotes that there are m different items are purchased in the transaction and for item ai, the purchased quantity is qai. Given an itemset Y ( ), Y is said to be contained in X if all items in Y also occur in X.Definition 1. Assume that an itemset Y is contained in transaction X. Y's transaction recency score gained from X, denoted by trscore (Y, X), is defined as: trscore (Y, X) = (1-δ) tcurrent-tx, where δ is a user-specified decay speed (δЄ [0, 1]), tcurrent denotes the current timestamp.Definition 2.Following Definition 1, the recency score of itemset Y, denoted by Rscore(Y), is defined as the sum of transaction recency score gained from all of the transactions containing it.Definition 3.The frequency score of itemset Y in DB, denoted by Fscore(y), is the number of transactions containing itemset Y. Definition 4.Let p(ai) denotes the unit price of item ai. Assume that an itemset Y is contained in transaction X.Definition 5.Following Definition 4, the monetary score of itemset Y in DB, denoted by MscoreDB(Y), is the sum of monetary score gained from transactions containing itemset Y.Definition 5.Following the Definition 4, the transaction amount for

transaction X, denoted by ta(X), is the sum of all items' monetary score in transaction Xs.

An RFMP-tree contains RFM-header and RFMP-tree. A RFM-header contains all 1-RFT-patterns. Each node in RFMP-tree consist item-name, Rscore, Fscore, Tta, parent-link, child-link, and sibling-link as fields, where item-name registers. A complete RFMP-tree will be constructed through two database scans, in the first DB scan, all 1-RFT-patterns will be stored in the RFMheader to pool up. After that, all 1-RFT-patterns in RFM-header will be arranged by thier Fscore in descending order. In other DB scan the root node of the tree will be sorted and label it as null, is created and prunes items for every transaction if item not found in RFM-Header. In the first step it finds all RFT-patterns from the RFMP-tree. The second step scans database once to compute the Mscore of all RFT-patterns and finally discover all RFM-patterns.Inserting of transactions in RFMP-tree generates 2 possible cases. Due to branch in the transaction, scores related to transactions are updated to nodes. If no branches are observed, nodes will have to be developed and have to calculate scores for it. With new nodes, comes along creation of parent link and node link.

Observation: Three of the thresholds performs outstandingly with experiments depicting significant values which are not generalized. In my observations values should be generalized and experiments on datasets should entertain noise and uncertainity.

A deisgn determining interesting pattern, minimum support & itemset count is designed by ZalizahAwang Long et al[24]. 1.Pre-processing 2.frequent item set algorithm and 3.data re-transformation are 3 stages in it which involves 2 modules of pre-processing. The Apriori algorithm is constructed in second module. In 2nd stage minimum support and sequence length'k' is found. In third stage, output is found which is threshold in outbreak of detection task.

Experiments were conducted in 3 datasets ZOO, TIC TAC TOE (UCI data repositories) and third for WSARE implementation. Results concluded that less the support more interesting pattern observed. Depending on the results, range for minimum support is 30%-50%. Figure below depicts support applied while experiment ranged from 10%-100% graphically. Positions of high gain are detected based on mean running collected output. The graph is analyzed state wise.Hence 40%-50% of minimum support is sufficient to generate frequent itemsets. High frequency count determines number of item sites for max detection rate. This paper is so designed to detect prospective interesting pattern from sequence item sets count and hence outbreaks in pattern are identified.

Observation: Exact number of min. support and length of item set reduces false rate of outbreak detection. Approx values are considerate and investigations on detection of curve are obligatory.Based upon domain driver framework, a new approach to mine pattern among data stored as schema was proposed by CláudiaAntunes et al [25], which is performed by a D2Apriori algorithm. Problems on Onto4ARframework are centered using ontology and assumes varied problem in context of ontology. The transactions provided in transactional patterns are by knowledge base and since its tuple with ontology characterization by set of features, it corresponds to traditional items. By adapted transactional pattern mining algorithm, new context mining can be achieved, to avail knowledge represented by ontology. The applications are executed on Apriori-based algoritm-D2Apriori with described procedures performed in a two step process as:-

1. Receives the dataset as input followed by reception of knowledge base and the constraints applicable.
2. Data file is then read in the context of the knowledge base, creating the dataset with the incorporation of semantics for each read transaction. As a special feature for reduction of time spent in the discovery process, measures adopted by authors are:

- Read the set of concepts and its attributes.
- Propagate concept properties through the taxonomy.

Hence problem for pattern mining and discovery of info in data stored by framework Onto4AR is resolved by overcoming results through traditional pattern approached.

Observation: D2Apriori algorithm is inefficient although due to memory consumption.

Ya-Han Hu et al[26] have highlighted study on Sequential Pattern Frequent Mining with multiple minimum supports. This determines correlations among items or itemsets which ows to varied frequencies, specyfing single minimum support cannot accurately discover interesting patterns. Based on these solutions authors argued that they are not straight forward. So an extended version of PLWAP-tree[28](Preorder linked multiple support tree) structure has been proposed. An efficient algorithm, a extended version of PLWAP-tree, named MSCP-growth(multiple supports candidate pattern) is used for detecting set of sequential patterns with MMS.

PLMS-tree obtained from PLWAP tree-like algorithm backups necessary information in sequence database. later, MSCP-growth develops sequential pattern with MMS based on PLMS-tree. Only items with less support than minsup are contained in PLWAP-tree while in PLMS-tree not less than MIS(I). in PLMS-tree extra info is required to tree structure but not required for the later part.

Observation: In this argument minute inconveniences are detailed. The concept of MMS on sequential pattern generates interesting patterns and an algorithm named MSCP growth is described.With

advantages of several algorithms combined, And experimental results exhibiting better performances than Compressed and Arranged Transaction Sequences tree(CATS tree) and Grouping Compressed tree mining algorithm, Chuang-Kai et al [27] developed a new structure called Sorted Compression tree as mining algorithm for association rule. Inorder to overcome repeated mining due to several minimum support, many algorithms were were proposed, of which CATS tree is one of them. It was ruled out due to drawbacks and better algorithm called Sorted compression tree(SC tree) was brought in, which was efficient and less memory consumptive.

Knowledge is derived from statistical data and relationship among products. The later is obtained from "Association Rule Mining" which is usually used. Apriori algorithms and Frequent patterns growth algorithm must use CATS-tree to adjust minimum support value. Its effectiveness is improved by usage of SC-tree, which by pre-sorting dat-set SC-tree results can be reliable.

A. Frequent Pattern Growth Algorithm: The FP growth algorithm proposed by Han and et al. is the most typical algorithm which is employed. The technique avoids the scanning of the database frequently, avoiding generating candidate item sets and reducing the hunt space. It is resourceful, scalable and faster than Apriori algorithm.

B. Compressed and Arranged Transaction Sequences Tree Algorithm: In CATS tree algorithm, there is no need for reconstruction when the minimum support is altered and it preserves all elements without decreasing. When minimum support threshold is distorted, the tree structure is unnecessary to be reconstructed. CATS algorithm constructs without sorting and hence, a need to adjust the tree structure.

C. Grouping Compress Tree Algorithm: The GC-tree increases performance of CATS-tree by pre-processing and compressing data into GC data structure and hence looks out large item set depending upon complexity of data.

D. Sorted Compress Tree Algorithm: This algorithm is used for tree construction and rule mining method. It consumes less memory space by the usage of SC-tree in large databases. SC algorithm mines frequent items in 3 stages.

1. Data processing- It sorts and arranges transactions for consistency which in turn help in improving tree constructing and rule mning.
2. SC-tree constructing –SC-tree is constructed by sorting transaction database.
3. Rule Mining in SC-tree- initially nodes are sorted, so as to mine in single direction i.e bottom-up. This process builds conditional mining tree and generates frequent item set.

Observation: CATS tree and GC-tree are verified under observations of SC-tree. Focus is on execution and memory consumption in minimum support and transaction size. Under execution and memory requirements spotlight is on efficiency on tree construction and association rule mining. "Fast and memory efficient mining of frequent closed item sets"[29] was authored by Claudio Lucchese et al. It mainly cogitate an ascending algorithm to provide a distilled approach to represent transactional often item sets which are exploited from the database. For the purpose of exploiting the often item sets, it preferred the techniques such as "divide-and-conquer" and "bitwise vertical representation" of the database. To achieve this approach, they utilized the services of separation technique of the "search space". It relies on authentic theoretical model for defining the often item sets clearly. As this algorithm faces the job of managing production of similar item set, it needs to keep an account of concise the space and time complexity for calculating item sets. To satisfy this problem, they opted for an efficient and capable trimming methodology. This methodology has the advantageous edge that it need not place all the fetched data in the main memory. It also has the facility to fetch data from each divided part of search space autonomously. The fetching operation can be pipelined and also it can be done in any sequence. Various examinations were performed over the proposed algorithm mainly commonly accessible data sets which proved this algorithm as an ascendable one which is portable for algorithms as CHARM, CLOSET+ AND FP-CLOSE even for multiple ordered applications. The efficient increments are considered more important in the case of decrement of support threshold. Here, a new walkthrough across the search space which is specific for every item set is proposed. It relies on the idea that there exist producers to maintain available pattern of the often item sets. This technique produces profitable results in the form of enabling us to create autonomous subtasks which can be pipelined.

Observation: The methodology proposed in this paper is practically applied to "DCI_CLOSED". This approach adopts techniques such as "depth first traversal" and "vertical bitmap representation" of dataset. This approach satisfies all the specifications proposed in the new technique of the paper. Due to the implementation of these techniques and methodologies, the proposed algorithm has limited space and time complexity. From the conducted examinations, it can be drawn that "DCI_CLOSED" is ascendable and supports fetching of heavy data sets even with minimum support threshold.

Liping Jiet al[30] proposed a clsode pattern detection model called "Compressed Hierarchical Mining of Frequent Closed Patterns from Dense Data Sets". The common technique followed here in all algorithms is that result will be the same whether the data is entirely mined or individually mined. Here, data is

partitioned in to several blocks and mining algorithm is applied on each individual block and then combined finally. Another advantage is that this process can be done in parallel, saving time and effort with every performance measures rating good. Both the algorithms can compress the mining space and can divide data hierarchically to apply mining. The only difference between the two algorithms is the technique they follow. C-mining follows Compact-Mining technique and B-mining follows Base rows Projection. These two versions can be done in parallel successfully and is shown in this proposal. We do have many other data mining techniques like PB-mining, PC-mining, D-mining along with C and B mining algorithms.

Observation: The experimental results can be studied under several results by Varying data set density, Experiments on real micro data set, varying the number of processors and scalability. These show the success of parallelism in the implementation of both C and B mining techniques.

An Efficient Algorithm for Mining Top-K Frequent Closed Itemsets was proposed by Jianyong Wang et al[31]. Authentic way of data set fetching includes considering "minimum support" threshold. It may not be feasible to specify "minimum support" everytime. Moreover, the conventional way may demand to fetch whole set of often data sets obeying the threshold. These specifications may restrict its availability as it is hard to avail "minimum support" threshold according to the situation. Moreover, "closed" item sets are compressed than a whole set of often data sets, making it portable and more applicable. Thus, to avoid such limitations, a broader methodology called "mining top-k closed item sets" is proposed. The technique followed here is that the minimum length should be "min_l". Here, "k" represents the required count of item sets to be fetched and "min_l" stands for the minimum length to be considered. This methodology, TFP, is constructed avoiding the need of "minimum support" threshold. The methodology initializes with the "minimum support" threshold value as 0. This value along with characteristics of top-k often data sets and length limitations, we increment the value of "minimum support" gradually. Also, utilize of "node count" and "descendant sum" techniques, allows the active trimming of FP-tree even prior or posterior to its construction. Even the fetching process can be fasten with the utilization of FP-tree including "top-down" and "bottom-up" traversals, applying heap of "search space" trimming techniques, high speed 2-stepped "hash-indexed result tree" and a new item set closure checking technique. Various examining processes proved that the proposed methodology was more efficient and linear ascendable in case of database space. The methodologies of proposed approach are: incrementing "minimum support" threshold utilizing "closed node count array" and "descendant sum", fetching only required elements initially and leaving rest for the fetch during "FP-tree" fetching, fastening fetching process by embedding "search space" trimming techniques and engaging "closure verification scheme" for packing up item sets.

Observation: This algorithm can be broadened to produce association principles and to embed specified limitations. From the analysis over authentic mining, the proposed algorithm is desirable as it avoids the need for specifying "minimum support" threshold. Even many reworks have been started over various elements such as top-k often closed item sets' efficiency, tractability, environment and various sequences.

Max-Clique[32] is a top-down Graph-based Approach to Frequent Pattern Mining was proposed by Yan Xie et al. The actual working process here is unlike in traditional methods where entire data is divided into parts and each is applied algorithm and then combined. These traditional approaches can be named Bottom-up since for mining entire data each detail is to be considered which is possible only in bottom-up process. The new strategies are that the data is processed from top to down and the most frequent top K long maximal patterns are identified and are resolved. Here the other small data sets are ignored and concentrated on those large sets. The large frequent patterns identified are called Cliques technically. These cliques are matched with some predefined pattern groups. Then apply Maximal Clique Detection on such cliques. Identifying the cliques is followed by Refinement phase where cliques are transformed into true maximal patterns.

Observation: This proposed model let mine only those long and interesting patterns which improve the speed and quality of the product with out negative effects on growth of other patterns which may be considered small. In fact the idea behind and the challenging factor is the identification of those large patterns to be mapped to cliques and is done just opposite to traditional bottom-up process as top-down.

VI. Conclusion

The analysis of the pattern and itemset mining algorithms should target performance, scalability, flexibility and reusability for making generation of association rules in robust way. Based on the reviews drafted in this paper, we can conclude that the frequent pattern mining is a price high process. All of the algorithms are worthy to achieve required goals in terms of results. But much of them influenced by an obstacle called multiple pass through given dataset that requires much number of disk reads. The rest algorithms those are able to perform mining with single passes are not portray the importance of these patterns. Further research should aim the concerns like flexibility, reusability and prunes in no. of derived patterns without compromising at pattern quality. These requirements influence to gain interest in further research in Frequent Itemset and pattern mining.

References Références Referencias

1. Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993ACM-SIGMODinternational conference on management of data (SIGMOD'93), Washington, DC, pp 207–216.
2. Pasquier N, Bastide Y, Taouil R, Lakhal L (1999) Discovering frequent closed itemsets for association rules. In: Proceeding of the 7th international conference on database theory (ICDT'99), Jerusalem, Israel, pp 398–416.
3. Pei J, Han J, Mao R (2000) CLOSET: an efficient algorithm for mining frequent closed itemsets. In: Proceeding of the 2000 ACM-SIGMOD international workshop data mining and knowledge discovery (DMKD'00), Dallas, TX, pp 11–20.
4. Zaki MJ,Hsiao CJ (2002) CHARM: an efficient algorithm for closed itemset mining. In: Proceeding of the 2002SIAMinternational conference on data mining (SDM'02),Arlington,VA, pp 457–473.
5. Wang J, Han J, Pei J (2003) CLOSET+: searching for the best strategies for mining frequent closed itemsets. In: Proceeding of the 2003 ACM SIGKDD international conference on knowledge discovery and data mining (KDD'03), Washington, DC, pp 236–245.
6. Grahne G, Zhu J (2003)Efficiently using prefix-trees in mining frequent itemsets. In: Proceeding of the ICDM'03 international workshop on frequent itemset mining implementations (FIMI'03), Melbourne, FL, pp 123–132.
7. Liu G, Lu H, Lou W, Yu JX (2003) On computing, storing and querying frequent patterns. In: Proceeding of the 2003 ACM SIGKDD international conference on knowledge discovery and data mining (KDD'03), Washington, DC, pp 607–612.
8. Pei J,Han J, Mortazavi-AslB,Wang J, PintoH, ChenQ,DayalU, HsuM-C (2004) Mining sequential patterns by pattern-growth: the prefixspan approach. IEEETransKnowl Data Eng 16:1424–1440
9. Bayardo RJ (1998) Efficiently mining long patterns from databases. In: Proceeding of the 1998 ACM-SIGMOD international conference on management of data (SIGMOD'98), Seattle,WA, pp 85–93.
10. Burdick D, Calimlim M, Gehrke J (2001) MAFIA: a maximal frequent itemset algorithm for transactional databases. In: Proceeding of the 2001 international conference on data engineering (ICDE'01), Heidelberg, Germany, pp 443–452.
11. Agrawal R, Srikant R (1995) Mining sequential patterns. In: Proceedings of the 1995 international conference on data engineering (ICDE'95), Taipei, Taiwan, pp 3–14.
12. SrikantR,AgrawalR(1996) Mining sequential patterns: generalizations and performance improvements. In: Proceeding of the 5th international conference on extending database technology (EDBT'96), Avignon, France, pp 3–17.
13. ZakiM(2001) SPADE: an efficient algorithm for mining frequent sequences.Mach Learn 40:31–60.
14. Zaki MJ (1998) Efficient enumeration of frequent sequences. In: Proceeding of the 7th international conference on information and knowledge management (CIKM'98),Washington, DC, pp 68–75.
15. Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. In: Proceeding of the 2001 international conference on data engineering (ICDE'01), Heidelberg, Germany, pp 215–224.
16. Yan X, Han J, Afshar R (2003) CloSpan: mining closed sequential patterns in large datasets. In: Proceeding of the 2003 SIAM international conference on data mining (SDM'03), San Fransisco, CA, pp 166–177.
17. Wang J, Han J (2004) BIDE: Efficient mining of frequent closed sequences. In: Proceeding of the 2004 international conference on data engineering (ICDE'04), Boston, MA, pp 79–90.
18. Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: Proceeding of the 2000 ACM-SIGMOD international conference on management of data (SIGMOD'00), Dallas, TX, pp 1–12.
19. Agarwal R, Aggarwal CC, Prasad VVV (2001) A tree projection algorithm for generation of frequent itemsets. J Parallel Distribut Comput 61:350–371.
20. Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. In: Proceeding of the 2001 international conference on data engineering (ICDE'01), Heidelberg, Germany, pp 215–224.
21. Liu J, Pan Y, Wang K, Han J (2002) Mining frequent item sets by opportunistic projection. In: Proceeding of the 2002 ACM SIGKDD international conference on knowledge discovery in databases (KDD'02), Edmonton, Canada, pp 239–248.
22. Liu G, Lu H, Lou W, Yu JX (2003) On computing, storing and querying frequent patterns. In: Proceeding of the 2003 ACM SIGKDD international conference on knowledge discovery and data mining (KDD'03), Washington, DC, pp 607–612.
23. Ya-Han Hu; Fan Wu; Tzu-Wei Yen "Considering RFM-values of frequent patterns in transactional databases", 2nd International Conference on Software Engineering and Data Mining (SEDM), June 2010, pages: 422 - 427.
24. Long, Z.A. Hamdan, A.R. Bakar, A.A; Parameter setting procedure via quick parameter evaluation in

frequent pattern mining for outbreak detection, 2nd Conference on Data Mining and Optimization, 2009. DMO '09, Issue Date: 27-28 Oct. 2009, On page(s): 90 - 93.

25. Antunes, C.;Pattern Mining over Star Schemas in the Onto4AR Framework, IEEE International Conference on Data Mining Workshops, 2009, ICDMW '09, Issue Date: 6-6 Dec. 2009, On page(s): 453 - 458.
26. Ya-Han Hu;Fan Wu; Yi-Chun Liao; Sequential pattern mining with multiple minimum supports: A tree based approach, 2nd International Conference on Software Engineering and Data Mining (SEDM), Issue Date: 23-25 June 2010 On page(s): 428 – 433.
27. Chuang-Kai Chiou, Judy C. R. Tseng; Sorted Compressed Tree: An Improve Method of Frequent Patterns Mining without Support Constraint, 2nd International Conference on Software Engineering and Data Mining (SEDM), 2010, Issue Date: 23-25 June 2010, On page(s): 328 - 333.
28. Ezeife, C.I.; Min Chen; Incremental mining of Web sequential patterns using PLWAP tree on tolerance MinSupport, Database Engineering and Applications Symposium, 2004, Issue Date: 7-9 July 2004, On page(s): 465 – 469.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

E-Learning by Time Dynamic Model Using Data Mining

By Kishan Sharma, Priya Jain, Rajni katare
NRI Institute of Technology and Management

Abstract - The object of this paper is to build up Just in Time Dynamic Learner Models to analyze learners' behaviors and to evaluate learners' performance in online education systems by using rich data collected from e-learning systems. The goal is to create metrics to measure learners' characteristics from usage data. To achieve this goal we need to use data mining methods, especially clustering algorithms, to second patterns from which metrics can be derived from usage data. In this paper, we propose a six layer models(raw data layer, fact data layer, data mining layer, measurement layer, metrics layer and pedagogical application layer) to create a just in time learner model which draws inferences from usage data. In this approach, we collect raw data from online systems, latter fact data from raw data, and then use clustering mining methods to create measurements and metrics.

Keywords : *Data mining, E-learning, iHelp.*

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

E-Learning by Time Dynamic Model Using Data Mining

Kishan Sharma[α], Rajni katare[Ω], Priya Jain[β]

Abstract - The object of this paper is to build up Just in Time Dynamic Learner Models to analyze learners' behaviors and to evaluate learners' performance in online education systems by using rich data collected from e-learning systems. The goal is to create metrics to measure learners' characteristics from usage data. To achieve this goal we need to use data mining methods, especially clustering algorithms, to second patterns from which metrics can be derived from usage data. In this paper, we propose a six layer models(raw data layer, fact data layer, data mining layer, measurement layer, metrics layer and pedagogical application layer) to create a just in time learner model which draws inferences from usage data. In this approach, we collect raw data from online systems, latter fact data from raw data, and then use clustering mining methods to create measurements and metrics.

Keywords : *Data mining, E-learning, iHelp.*

I. Introduction

The research in this paper is an investigation on how to apply data mining rules [1,2,4], especially clustering algorithms, to e-learning usage data to dynamically create just in time learner models.

Make Sense of Usage Data:

E-learning systems are used for computer-based education and they have widespread use in many domains. The usage data collected from the forum for each learner includes: Messages posted in the forum, question messages posted in the forum, answering messages posted in the forum, messages accessed by the learner, messages mostly navigated by the learner.

Issues of Using Usage Data:

With rich usage data collected from e-learning systems, we try to make sense of this data by applying data mining techniques. There are some challenging issues that need to be navigated: among patterns found from data mining techniques, patterns are useful in an e-learning system, determine that a pattern is useful or not, predict learners' behaviors based on the usage data.

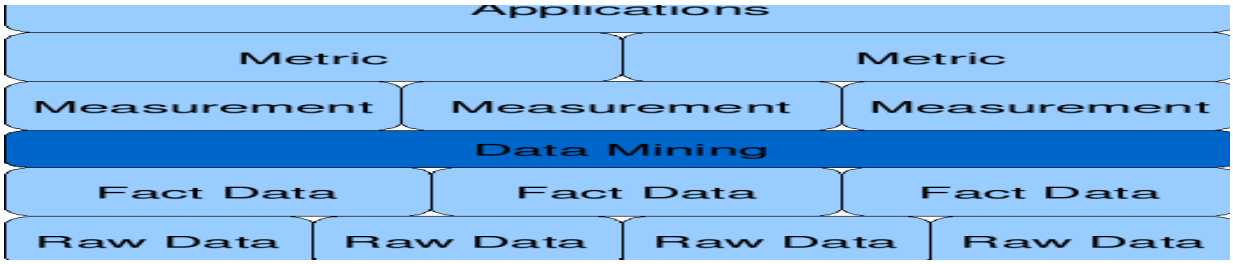


Figure 3.2 : Six layers model

II. Objective

In this paper, we did some proof of concept research to study the above issues. We studied the relationships between usage data and learner [6] characteristics and behaviors. This resulted a six layers model to create learner models. This is a dynamic model created by applying clustering techniques on the usage data collected from the real system. We implemented a test system to collect data and to create results. Two experiments have been used to evaluate and compare the results of the test system.

III. Problem Definition

Some patterns found from the usage data, also called metrics and measurements to represent learners' characteristics, seem to be clearly useful in building learner models. Other patterns show promise to describe learners' behaviors, but remain unproven. Deferent clustering algorithms produce various results. Selection and determination of data mining algorithms and associated parameters will play an important role in creating learner models. Pre-computation is necessary if anything like just in time modeling is to be achieved, and has been implemented in our test system

Author α Ω β : *NRI Institute of Technology and Management, Baraghata, Jhansi Road, Gwalior-474001, INDIA.*
E-mails : drkishansharma2006@rediffmail.com, sharmanitm.06@gmail.com , priyainfo005@gmail.com

IV. Literature Review

In order to more easily discuss the[1,2] current state of web based e-learning systems, educational data mining and my own research, it is useful to _rest look at the history that has brought educational research and data mining technologies together. It focus on web based educational theories such as adaptive intelligent learning and learner models, data mining algorithms, and educational data mining research.

a) E-Learning

E-learning is naturally associated with computer based learning, especially to be used in distance learning. The term e-learning is also called by some researchers e-training, online instruction, web-based learning, web-based training, web-based instruction, etc. The main advantages of e-learning are edibility and convenience. Learners can work at any place and at any time with an Internet connection for most e-learning environments.

b) Learning Content Management Systems

Developing a course [3]to be taught on the Internet is difficult because it requires the system to do a combination of things: publishing content on web pages, supporting tools for self learning, and providing assessments of learning performance Some good commercial LCMS systems include Blackboard (Web CT), Virtual-U and Top Class, etc. Open source LCMS include iHelp, a Tutor and Model, etc.

c) iHelp

iHelp is an e-learning system developed[5,7] by the Advanced Research in Intelligent Education Systems (ARIES). iHelp is made up of a number of web based applications designed to support both learners and instructors throughout the learning process1. The main components of iHelp are asynchronous iHelp Discussion forums, synchronous iHelp Chat rooms, the iHelp Learning Content Management Systems (also called iHelp Courses), iHelp Share and iHelp Lecture.

_ iHelp Chat: This chat room provides workspaces for learners to have synchronous communication with one another and with their instructors and teaching assistants.

_iHelp Courses: This LCMS system provides tools to support full on-line courses and is designed for distance learning. It provides learners with a portal to multimedia course content.

_iHelp Share: This is a collaborative learning tool to share information relevant to courses among learners.

_iHelp Lectures: This system provides multimedia lectures to learners so that learners can write messages and comments, make notes and tags on video clips, so that all learners can share this information.

Like other LCMS systems, iHelp collects and stores all information, such as personal information, pedagogical results, learners' interaction data, etc. into a database. These data are the source data for our project, as we will discuss in the.

V. Research Contribution and Future Direction

The goal of this research has been to show that just in time learner models can be created from analyzing learners' online tracking data. This approach consists of clustering raw data, selecting pedagogical applications and applying data mining methods. This has led to measurements and metrics that can be calculated for each individual learner to represent that learner's characteristics and behaviors.

a) General Comments on the Two Experiments

From the two experiments' results, some measurements seem to be useful in building just in time learner models;[1] some measurements only show promise to be leading in the right directions; some measurements have not found much support. The expert experiment, in which experts observed and evaluated learners as the third party, shows much more positive results compared to the self evaluation experiment, in which learners evaluated themselves.

(i)Table 5.1 and Table 5.2 show a summary of the accuracy values and correlation coefficient values for the expert experiment and the self evaluation experiment clued three measurements from the activity level metric: navigating context, read in discussion forum, chat activity; two measurements from the social tendency metric: presence and social tendency; and one measurement from the knowledge tendency metric: usage instructors to observe and evaluate learners' learning behaviors as in traditional class rooms. The results support that the instructors [5] at least will have a helpful tool to dynamically observe and evaluate learners' performance in online education environment. Six measurements had negative results in the expert experiment [4,6], with lower accuracy values or lower correlation coefficient values. Those measurements in the learning style metric especially have lower values in both accuracy and coefficient values.

Table 5.1,5.2 : Summary of expert experiment

Correlation Coefficient	Accuracy		
	<0.40	0.4-0.6	>0.6
<0.4	2	1	0
0.4-0.6	1	1	1
0.6-0.8	0	1	1
0.6-0.8	1	4	6

Correlation Coefficient	Accuracy		
	<0.40	0.4-0.6	>0.6
<0.4	3	1	1
0.4-0.6	1	2	0
0.6-0.8	0	1	0
0.6-0.8	0	0	0

b) Just In Time Model

A goal of this research has been to compute learner [3] models just in time as we need them instead of keeping static learner models. We do not use the historical learner models in our computations, instead recomposing the metrics and measurement based on the current available fact data and raw data. In this way, any changes in learners will be automatically recognized in the form of new measurements. Through two experiments, we have shown that just in time computations are possible and in some cases they lead to useful measurements.

c) Top Down

Top down computation of [2,6] measurements promises to allow the calculation of results quite rapidly when compared to computations using a pure bottom up data mining approach. In our approach, the first step is to decide the purpose of the applications such that we can figure out the necessary metrics and measurements to support the Applications. The second step is to find the raw data and retrieve the fact data to support the measurements. The last step is to mine the fact data to find patterns that can be used in creating formula that can later be used to directly calculate the measurements based on the available fact data.

VI. Future Work and Direction

The main drawback to the two experiments is that the number of evaluators is too low to get statistically significant results. It is easier to have more learners involved in the self evaluation than to get more experts in the expert evaluation. However, we need to improve and design a better self evaluation questionnaire to attract more learners to participate while at the same time get better questions. More involvement of instructors in the design of the questionnaire may be helpful. We have selected four metrics and 15 measurements in the two experiments. Then we need to show how we can use these measurements to build just in time learner models in actual pedagogical applications. Because of the relatively high correlation coefficient results in the two experiments, we can at various time points, apply classification algorithms to predict learner behaviors. A promising direction may be to keep old predictions and combine this information with the latest updated predictions to improve the accuracy of the just in time learner model.

VII. Acknowledgements

The authors are very thankful to the referees for their useful suggestions for the presentation of the paper.

References Références Referencias

1. Agapito, J. B., Sosnovsky, S., and Ortigosa, A. (2009). Detecting symptoms of low performance using production rules. In 2nd International Conference on Educational Data Mining, pages 31{40}.
2. Antunes, C. (2008). Acquiring background knowledge for intelligent tutoring systems, In 1st International Conference on Educational Data Mining, pages 18{27}.
3. Arroyo, I., Murray, T., Woolf, B., and Beal, C. (2004). Inferring unobservable learning Variables from students' help seeking behavior. Lecture Notes in Computer Science, 3220:782{784}.
4. Ayers, E., Nugent, R., and Dean, N. (2009). A comparison of student skill knowledge Estimates. In 2nd International Conference on Educational Data Mining, pages 1{10}.
5. Baker, R. (2009). Differences between intelligent tutor lessons, and the choice to go O_-task. In 2nd International Conference on Educational Data Mining, pages 11{20}.
6. Baker, R. and Carvalho, A. (2008). Labeling student behavior faster and more precisely with text replays. In 1st International Conference on Educational Data Mining, pages 38{47}.
7. Baker, R., Corbett, A., and Koedinger, K. (2004). Detecting student misuse of intelligent tutoring systems. In Proceedings of the 7th International Conference on Intelligent Tutoring Systems, pages 531{540}.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Cognitive Adaptive Mac Based on Knowledge Based Reasoning for Cognitive Radio Computer Networks

By Vibhakar Pathak, Dr K. C. Roy, Ms. Priyanka Gondiliya, Santosh Kr. Singh
Suresh Gyan Vihar University, Jaipur

Abstract - In this paper we are proposing a new concept in MAC layer protocol design for Cognitive radio by combining information held by physical layer and MAC layer with analytical engine based on knowledge based reasoning approach. In the proposed system a cross layer information regarding signal to interference and noise ratio (SINR) and received power are analyzed with help of knowledge based reasoning system to determine minimum power to transmit and size of contention window, to minimize backoff, collision, save power and drop packets. The performance analysis of the proposed protocol indicates improvement in power saving ,lowering backoff and significant decrease in number of drop packets. The simulation environment was implement using OMNET++ discrete simulation tool with Mobilty framework and MiXiM simulation library.

Keywords : *Cognitive radio, SDR, SINR ,Received Power, MAC , OMNET++.*

GJCST-F Classification : *C.2*

Strictly as per the compliance and regulations of:

Cognitive Adaptive Mac Based on Knowledge Based Reasoning for Cognitive Radio Computer Networks

Vibhakar Pathak[α], Dr K. C. Roy[Ω], Ms. Priyanka Gondiliya[β], Santosh Kr. Singh[ψ]

Abstract - In this paper we are proposing a new concept in MAC layer protocol design for Cognitive radio by combining information held by physical layer and MAC layer with analytical engine based on knowledge based reasoning approach. In the proposed system a cross layer information regarding signal to interference and noise ratio (SINR) and received power are analyzed with help of knowledge based reasoning system to determine minimum power to transmit and size of contention window, to minimize backoff, collision, save power and drop packets. The performance analysis of the proposed protocol indicates improvement in power saving ,lowering backoff and significant decrease in number of drop packets. The simulation environment was implement using OMNET++ discrete simulation tool with Mobilty framework and MiXiM simulation library.

Keywords : Cognitive radio, SDR, SINR ,Received Power, MAC , OMNET++

I. Introduction

There is a significant amount of unused space (or white space) in the licensed radio spectrum due to non-uniform spectral demand in time, frequency and space and static spectrum allocation policies widely used today. Studies sponsored by FCC in [12] show that over 65% of the allocated spectrum is not in use at any time even in a crowded area where the spectral usage is intensive. On the other hand, the remaining portion of the unlicensed spectrum (e.g. the ISM band) is being exhausted by emerging wireless services and applications, leading to the so-called spectral scarcity problem.

One solution to this problem is to allow unlicensed spectrum users to use the white space and keep the interference to licensed users below an acceptable level. This is called the dynamic spectrum access (DSA) scheme, which can be realized by cognitive radio (CR) techniques [13]–[16] A CR device monitors a swath of spectrum including those occupied by licensed services and attempts to identify the "white" space (or the spectrum hole), which is referred to as the idle period between consecutive accesses of licensed users, and exploits it for communication at that specific geographical location.

Recent development in slicon technology leads to development of smart reprogrammable circuits Using which a new class of intelligent or "COGNITIVE" radios can be develop based on Software Defined Radio(SDR).Such radio based system would be capable of dynamic physical adaptation. In recent past development of cognitive radio hardware and software, especially at the physical layer has received considerable attention., the question how one can transform a set of cognitive radio into a cognitive network is less considered by research community. Cognitive radio or agile radio is a technology to chose a wide variety of radio parameters and protocol standard in adaptive manner on observed radio link and network conditions. Such type of system is capable of handling cross layer parameter change and advice the network to change the system with view of minimum power consumption, lowering backoff and reducing rate of drop packets, and hence updating utilization of network resource. In generalized case the cognitive radio is capable of adapting modulation of wave form, OSA (opportunistically spectrum access), MAC protocols, and network protocols. The cognitive radio can make runtime change to protocols to avoid collisions by transmitting packets with minimum power utilized for hop to hop transfer. There have been many research work addressing physical layer agility of cognitive radio system based on OSA[2][3].This paper proposes novel Cross Layer aware adaptive MAC protocol based on knowledge–based reasoning for cognitive radio. In this different type of Mac protocol can be adapted by the system on basis different physical parameter, receiver power, SINR and minimum power to transmit for hop by hop packet transfer knowledge base. The candidate MAC protocols have different advantages in various network situation. The proposed C-AMAC will provide a frame work to observe and make descision to switch on between those candidate protocols. The switching between different MAC protocols will be validate and used in future by the system based on short term statistics.

The proposed system (C-AMAC) is based on the "COGNET" cognitive radio protocol architecture described in [4].The COGNET system includes the

Author[α ψ] : Associate Prof. Suresh Gyan Vihar University, Jaipur.
E-mails : vibhakarp@rediffmail.com, sksmtech@yahoo.com
Author[Ω]: Prof. Pacific University Udaipur.
E-mail : roy.krishna@rediffmail.com
Author[β]: M.Tech Scholar Suresh Gyan Vihar University, Jaipur.
E-mail : sweetangel4582@gmail.com

concept of "Global Control Plane" GCP which supports exchange of control information between networking cognitive radio devices. The MAC adaptation requires cross-layer parameters and tokens to maintain protocol consistency between radio nodes sharing the channels. The GCP is also used in many cross-layer protocols including routing in MANET.

In this paper we will focus on the simulation of C-AMAC using open source network simulation tools OMNET + + with mobility frame work. The candidate protocols used in our protocols are commonly used protocol in wireless networks, these are CSMA/CA, TDMA and multi-channel Wi-Fi protocol. CSMA is ideal for short burst source with light to medium traffic volume. In streaming source TDMA protocol do well and Multi-channel Wi-Fi protocol is ideal for real-time mission critical application. We evaluate the proposed C-AMAC protocols with various traffic types.

In next section we begin by describing related work followed by the protocol is discussed in section III finally we provide Simulation results in section IV and future work in section V.

II. Related Works

The benefit of separate control and data planes in COGNET in the COGNET architecture have been previously explained in[4]. The GCP uses low-rate radio PHY with wide coverage for robustness and may include protocol modules for topology discovery, bootstrapping, address assignment. The data plane protocol stack supports data communication via PHY MAC and routing modules specified through an API that interacts with control modules in the radio node. The GCP control architecture is used for our C-AMAC.

Many approaches have already been proposed to reduce the number of collisions by substituting the binary exponential back-off algorithm of the IEEE 802.11 by novel back-off approaches or selecting an intermediate value instead of resetting the CW value to its initial value or some random value of CW upper or lower bound. The most related work to our back-off mechanism is the determinist contention window algorithm (DCWA) in [17]. DCWA increases the upper and lower bounds instead of just doubling the CW value. In each contention stage, a station draws a back-off interval from a distinct back-off range that does not overlap with the other back-off ranges associated to the other contention stages. In addition, the back-off range is readjusted upon each successful transmission by taking into account the current network load and history (resetting the back-off ranges mechanism; see details in [17]).Among the related work concerning energy conservation, such as power saving or power control mechanisms, the power saving mechanism (PSM) is the most familiar. It is provided by the standard [1], which allows a node to go into doze mode. Power control schemes, varying the transmit power in order to reduce the energy consumption, have already been presented in many studies; for example, see [18–22]. These schemes and many others have shown that power control protocols can achieve a better power conservation and higher system throughput through a better spatial reuse of the spectrum.

One part of the research in the field of power control focuses on dependencies and tradeoffs between both the transmit power and the carrier sense threshold, while another part focuses only on the adjustment of the carrier sense threshold [23] The work in [12] investigated the tuning of the transmit power, carrier sense threshold, and data rate in order to improve spatial reuse. The authors have shown that tuning the transmit power is more advantageous than tuning the carrier sense threshold. Cross-layer protocols contributing to the enhancement of the MAC layer and the adjustment of the power level have also been presented in many papers. One of them, the power adaptation for starvation avoidance (PASA) algorithm [17], was designed following the observation from that the request-to-send/clear-to-send (RTS/CTS) collision avoidance mechanism of the IEEE 802.11 DCF cannot eliminate collisions completely. This can lead to a channel capture where a channel is monopolized by a single or a few nodes. The authors of [17] studied how to control the transmission power properly in order to offer a better fairness and throughput by avoiding a channel capture. The power level increases exponentially and decreases linearly in the PASA, while using an RTS/CTS control scheme. PASA is not applicable with the basic access scheme. It requires that a neighbor power table (NPT) is maintained by each node with information such as the minimum power that must be maintained according to the distance to the destinations, which should be obtained through some location service.

After all, maintaining the NPT table with "fresh" data is not realistic in a mobile ad hoc environment taking into account interferences, fading effects, movement of the nodes, and deaths and new entriers of nodes.

The carrier sense multiple access protocol with power back-off (CSMA/PB) has been presented in [18]. The CSMA/PB reduces the transmission power level in order to avoid collisions, following the observation that, in a smaller transmission area, interferences and contentions are expected to be reduced. Results obtained in [18] are based on an optimistic centralized power-aware routing strategy which illustrates the potential of the power back-off. The CSMA/PB protocol has been evaluated with three transmission power levels only, thus the amount of power decreases fast. Therefore, it is really important that the routing protocol takes power levels into account. Each node has to maintain the routing table with entries for each destination with corresponding power levels.

III. C-AMAC Protocol

The goal of the C-AMAC protocol is to save energy (which leads to an extension of the lifetime of nodes) and to reduce the number of collisions. However, the C-AMAC protocol does not degrade the throughput performance in terms of the throughput and data rate, while fulfilling these goals. The C-AMAC protocol tackles a couple of problems that exist in the current implementation of the standard. It does this by two means, first it concentrates on the flexible adjustment of the upper and lower bounds of the CW to lower the number of collisions. Secondly, it uses a power control scheme to limit the waste of energy and also to lower the number of collisions. Hence, it has a MAC-PHY crosslayer architecture. To tackle the inefficient use of the back-off window in the standard, we developed a MAC protocol that makes use modified Enhanced selection Bounds algorithm (EsB) [19]). The mEsB(modified Enhanced selection Bounds) adjusts the lower and upper bounds of the CW range, taking into account the number of retransmissions attempts, Received power from neighbor , and SINR sensed. Each node can predict number of active 1- hop neighbor, based on successfully detected signals Knowledge based reasoning engine. In [20] the utilization rate of the slots (slot utilization) observed on the channel by each station is used for a simple, effective and low-cost load estimate of the channel congestion level. The protocol uses knowledge based reasoning system generate collision history and predict the same during the resetting stage, the CW value is reset to a value which depends on the predicted collisions. This forms the MAC part of the C-AMAC protocol and results in a reduction of the number of collisions.

The goal is not only to lower the number of collisions, but also to save energy. If we reflect on the reason why messages collide, it becomes clear that this is because too many nodes are too close to each other. They could be positioned a few meters from each other, but their transmission range is far greater than these few meters. Hence, the nodes are too close to each other relative to their respective transmission range. This not only results in a higher number of collisions, but also in an excessive use of energy to transmit a packet.

However, not receiving an acknowledgment for a sent packet does not always mean that the packet was lost or corrupted because there was too much interference. It could also happen that the transmission power was simply too low to reach any of the surrounding nodes. Therefore, the C-AMAC protocol takes the signal-to-interference-and noise ratio (SINR) into account. If no acknowledgment has been received, but the noise level (deducted from the SINR) is low, then we assume that the transmission power was too low to reach any of the neighbors. In that case the transmission power is increased.

The C-AMAC power control part is based on this observation of received power of packet from 1-hop neighbor and it lowers its transmission power (while observing too high noise in the vicinity) when it does not get the acknowledgment that a packet has been received successfully. The final result will be that all nodes will find their optimal transmission power that ensures that they can reach their neighbors, but not interfere with other nodes. The whole information regarding SINR, minimum power to transmit ,receiver power will be communicated to whole network via GCP and a reasoning will be developed using knowledge based system to predict the behavior of network.

IV. Simulation Results

The proposed cross-layer protocol has been implemented in the OMNET++ 4.0 network simulator [10]. The simulations have been carried out for various topologies, scenarios with different kinds of traffic, and routing protocols. The following performance metrics have been used:

(i) Total packets received,
(ii) Average throughput (Mbps),
(iii) Lifetime LND (seconds),
(iv) FND: first active node died (seconds),
(v) Lifetime RCVD (seconds),
(vi) Average aggregate delay (seconds),

The first node died metric is defined as the instant in time when the active (a node transmitting/receiving) first node died. We have defined the network lifetime as the time duration from the beginning of the simulation until the instant when the active (a node transmitting/receiving) last node died, that is, there is no live transmitter-receiver pair left in the network. The Lifetime RCVD is specified as the instant in time when the last packet is received.

The average throughput has been defined as
Thr = Total numberPackets received Simulation Time [Mbps] and average sending bit rate has been defined as
Sbit = Total number Packets sent Simulation Time [Mbps].

Table 1

Number of active nodes	25, 50 (default)
Simulations area	≤ 1000 * 1000m
Topology	Random
PHY/MAC	DSSS, IEEE 802.11b
SINR thr. (dB)	22.05
Type of netwok	homo/hetero-geneous
Initial energy (J)	variable = 0.5–. . ., 5, 20
PtMAX –	250m 0.200888W
PtMAX –	100m 0.010072W
txPowerinit	250 _ 100 meters
rxPower	45% of PtMAX
idlePower	30% of PtMAX
Capture Thr.(dB)	10
Traffic model	CBR/UDP
Payload size (bytes)	2048 _ 100–8192
CWmin –CWmax (slots)	15–1023
Simulation time (s)	≤650
Routing	AODV (default), DSR
Movement	random and constant
Mobility model	turtle Model
Speed (m/s)	0 – 2_ ≤ 20; 1.5
Access scheme Basic (default) _ RTS/CTS	

Table 2 : Typical values of path loss exponent and shadowing deviation.

Environment	ρ (dB)	σ (dB)
Outdoor Free space	2.4	4 to 12
Outdoor Shadowed Urban	2.7 to 5.6	4 to 12
Indoor Line-of-sight	1.6 to 1.8	3 to 6
Indoor Obstructed	4 to 6	6.8

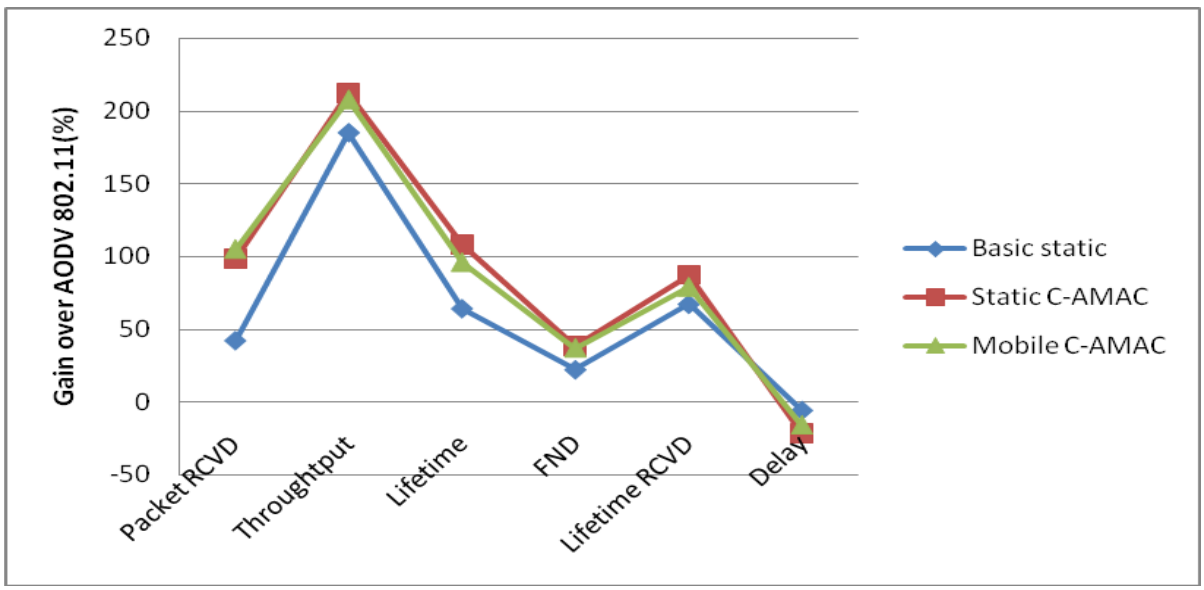


Figure 1

First, we defined a simulation scenario with 40 static nodes randomly distributed in a shadowed urban area where nodes send a CBR packet (2048 bytes payload size) from the beginning till the end of the simulation every 0.025 seconds. Figure 4 depicts the number of collisions per node in one of the simulation scenario runs (20 simulation runs in total). Notice that with the C-AMAC protocol most of the nodes have much fewer collisions, although the lifetime of the network is increased significantly (See Figure 6). Figure 5 shows the total number of packets received by the DCF standard, basic power control protocol, and C-AMAC protocol. The tuning of the C-AMAC protocol has been investigated as can be observed in the figure. The C-AMAC, the protocol outperforms the IEEE 802.11 DCF standard and basic power control protocol noticeably. The C-AMAC achieves the best performance, which means that the history of collisions experienced has an influence in a static environment.

Figure1 shows the gain in percentage over the IEEE 802.11 DCF standard obtained by the basic power control protocol in the static network and the C-AMAC protocol in both static and mobile networks. Note that, thanks to PHY (power level adjustment) and MAC

(recovery mechanism and CW resetting) layer treatment, the number of collisions can be decreased noticeably while saving lot of the energy which leads to an increase of the lifetimes (LND and lifetime RCVD) of the network and the throughput. The performance of the Lifetime RCVD is worse than the performance of the lifetime of the network, which means that some last transmitter-receiver pairs still have connections; however, the packets cannot be routed to the destination. The performance of the throughput fairness, which is improved tremendously, is explainable since nodes give others more opportunity to access a wireless channel while decreasing the transmit power level. On the other hand, by increasing the power (upon a consecutive collision and too low noise in the vicinity), their chance to get to the channel is increased since their coverage transmit area is wider. However, the average delay is degraded, because the C-AMAC protocol adjusts both the lower and upper bounds of the CW range and allows to decrease (apart from an increase) the power level, which in consequence can increase the average delay. In this work we have designed a novel cross-layer protocol, Adaptive protocol. The protocol adjusts the upper and lower bounds of the contention window to lower the number of collisions. Secondly, it uses a power control scheme, triggered by the MAC layer, to limit the waste of energy and also to decrease the number of collisions. Apart from that the protocol uses knowledge based reasoning for prediction SINR and receiver power for further optimization of contention window and further decrese in number of collision The protocol has been evaluated in three different scenarios and compared to the IEEE 802.11 DCF standard and the basic power control protocol .

References Références Referencias

1. J. Mitola III, "Cognitive Radio: An Integrated Agent Architecture for Software Radio," PhD thesis, Royal Institute of Technology (KTH),S weden, May 2000.
2. D. Maldonado, B. Le, A. Hugine, T. W. Rondeau, C. W. Bostian,"Cognitive radio applications to dynamic spectrum allocation: a discussion and an illustrative example," IEEE Proc. DySPAN, 8-11 Nov. 2005, pp. 597 - 600.
3. Jun Zhao, Haitao Zheng and Guang-Hua Yang, "Distributed Coordination in Dynamic Spectrum Allocation Networks" in IEEE DySPAN, Nov. 2005.
4. Xiangpeng Jing and Dipankar Raychaudhuri, "Global Control Plane Architecture for Cognitive Radio Networks," in Proceedings of IEEE CogNets 2007 Workshop - Towards Cognition in Wireless Networks (in conjunction with IEEE ICC), June 24-28, 2007.
5. Xiangpeng Jing, Shanmuga S Anandaraman, Mesut Ali Ergin, Ivan Seskar and Dipankar Raychaudhuri, "Distributed Coordination Schemes for Multi-Radio Co-existence in Dense Spectrum Environments: An Experimental Study on the ORBIT Testbed," in Proceedings of IEEE DySPAN 2008, Chicago, IL, Oct. 14-17, 2008.
6. D. Raychaudhuri, "ORBIT: Open-Access Research Testbed for Next-Generation Wireless Networks,"
7. Xiangpeng Jing and D. Raychaudhuri, "A Spectrum Etiquette Protocol for Efficient Coordination of Radio Devices in Unlicensed Bands," in Proceedings of PIMRC 2003, Vol. 1, pp. 172 - 176, Beijing, China, Sept. 7-10, 2003.
8. Cordeiro, Carlos; Challapali, Kiran, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks" in IEEE DySPAN, Apr. 2007
9. www.omnetpp.org
10. http://mobility-fw.sf.net/
11. FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Federal Communications Commision, Technical Report 02-155, November 2002.
12. Spectrum efficient uni- and multicast services over dynamic multiradionetworks in vehicular environments. [Online]. Available: http://www.comnets.rwth-aachen.de/odrive/index. html
13. DARPA ATO the next generation program. [Online]. Available:http://www.darpa.mil/ato/programs/XG/index.htm
14. IEEE 802.22 wireless regional area networks WRAN. [Online]. Available: http://standards.ieee.org/announcements/pr 80222.html
15. S. Haykin, "Cognitive radio: brain-empowered wireless communications (Invited Paper)," IEEE JSAC, vol. 23, no. 2, February 2005.
16. A. Ksentini, A.Nafaa, A. Gueroui, andM.Naimi, "Deterministcontention window algorithm for IEEE 802.11," in Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05), vol. 4, pp. 2712–2716, Berlin, Germany, September 2005.
17. P. Karn, "MACA—a new channel access method for packet radio," in Proceedings of the ARRL/CRRL Amateur Radio 9th Computer Networking Conference, pp. 134–140, Ontario, Canada, September 1990.
18. S. Agarwal, R. H. Katz, S. V. Krishnamurthy, and S. K. Dao, "Distributed power control in ad-hoc wireless networks," in Proceedings of the 12th IEEE International Symposium on Personal, Indoor andMobile Radio Communications (PIMRC '01), vol. 2, pp. F59–F66, San Diego, Calif, USA, September-October 2001.
19. A. Acharya, A. Mishra, and S. Bansal, "MACA-P: a MAC for concurrent transmissions in multi-hop wireless networks," IBM Research Report RC22528, IBM Corporation, Armonk, NY, USA, July 2002.
20. E. Poon and B. Li, "SmartNode: achieving 802.11 MAC interoperability in power-efficient ad hoc networks withdynamic range adjustments," in

Proceedings of the 23rd IEEE International Conference on Distributed Computing Systems (ICDCS '03), pp. 650–657, Providence, RI, USA, May 2003.

21. E.-S. Jung and N. H. Vaidya, "A power controlMAC protocol for ad hoc networks," in Proceedings of the 8th Annual International Conference onMobile Computing and Networking (Mobicom '02), pp. 36–47, Atlanta, Ga, USA, September 2002.
22. T.-S. Kim, H. Lim, and J. C. Hou, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in Proceedings of the 12th Annual International Conference on Mobile Computingand Networking (Mobicom '06), pp. 366–377, Los Angeles, Calif, USA, September 2006.

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A New Method for Impulse Noise Removal in Remote Sensing Images

By G.L.V.TataRao, M.S.Madhan Mohan, Dr.G.M.V.Prasad
B.V.C. Engineering College Odalarevu

Abstract - Existing filtering algorithms use all pixels within a window to filter out the impulse noise. They increase the size of neighboring pixels with the increase of noise density. In this paper, we propose an impulse noise removal algorithm for remote sensing images, that emphasis on few noise-free pixels. The detection map (DM) is constructed from the input noisy image, by assigning a binary value 1 for each corrupted pixel in the input image. By using the detection map, the proposed iterative algorithm searches the noise free pixels with in a small neighborhood. The noisy pixel is then replaced with the median value estimated from noise free pixels. In-order to better appraise the noise cancellation behavior of our filter from the point of view of human perception, we perform segmentation via spline regression on remote sensing image for both noisy image and filtered image. Experimental results show that the filtering performance of the proposed approach is very satisfactory providing better feature extraction in remote sensing images.

Keywords : *Impulse Noise, Image segmentation,Remote Sensing, Image Processing.*

GJCST-H Classification : *I.4.6, I.4.0*

Strictly as per the compliance and regulations of:

A New Method for Impulse Noise Removal in Remote Sensing Images

G.L.V.TataRao[α], M.S.Madhan Mohan[Ω], Dr.G.M.V.Prasad[β]

Abstract - Existing filtering algorithms use all pixels within a window to filter out the impulse noise. They increase the size of neighboring pixels with the increase of noise density. In this paper, we propose an impulse noise removal algorithm for remote sensing images, that emphasis on few noise-free pixels. The detection map (DM) is constructed from the input noisy image, by assigning a binary value 1 for each corrupted pixel in the input image. By using the detection map, the proposed iterative algorithm searches the noise free pixels with in a small neighborhood. The noisy pixel is then replaced with the median value estimated from noise free pixels. In-order to better appraise the noise cancellation behavior of our filter from the point of view of human perception, we perform segmentation via spline regression on remote sensing image for both noisy image and filtered image. Experimental results show that the filtering performance of the proposed approach is very satisfactory providing better feature extraction in remote sensing images.

Keywords : *Impulse Noise, Image segmentation,Remote Sensing, Image Processing.*

I. Introduction

Digital images are often corrupted during acquisition, transmission or due to faulty memory locations in hardware [1]. The impulse noise can be caused by a camera due to the faulty nature of the sensor or during transmission of coded images in a noisy communication channel [2].Consequently, some pixel intensities are altered while others remain noise free. The noise density (severity of the noise) varies depending on various factors namely reflective surfaces, atmospheric variations,noisy communication channels and so on.The restoration of noise-free images is carried out as a preprocessing task in a wide range of applications such as medical imaging, remote sensing images. Order-static filters are nonlinear filters [3][4]whose response is based on the ordering (ranking) the pixels contained in the image area encompassed by the filter, and then replacing the value of the center pixel with the value determined by the ranking result.The filtering should be applied to corrupted pixels only while leaving those uncorrupted ones intact. Therefore a noise detection process to discriminate the uncorrupted pixels from the corrupted ones prior to applying nonlinear filter is highly desirable.This noise detection provides the noisy density in the input image, which is often unknown in priori, will cause substantial degradation on the filtering performance.In this paper, we proposed a new iterative approach for noise removal in remote sensing images that emphasis on noise-free pixels within small neighborhood. In this scheme, first, the pixels affected by salt-and-pepper noise are detected. If we did not find a certain number of noise-free pixels within neighborhood, then the central pixel is leftunchanged. Otherwise, the noisy pixels are estimated from the noise-free pixels. The process iterates until all the noisy pixels are estimated. The paper is organized as follows: Section II presents the Impulse Noise in Digital Images, Section III presents the Noise Filtering Method, Section IV presents segmentation algorithm via Spline Regression, Section V presents the Experimental results, finally Section VI reports conclusion.

II. Impulse Noise in Digital Images

Impulse noise [5] corruption is very common in digital images. Impulse noise is always independent and uncorrelated to the image pixels and is randomly distributed over the image.Hence unlike Gaussian noise, for an impulse noise corrupted image all the image pixels are not noisy, a number of image pixels will be noisy and the rest of pixels will be noise free. There are different types of impulse noise namely salt and pepper type of noise and random valued impulse noise.

In salt and pepper type of noise the noisy pixels takes either salt value (gray level -225) or pepper value (grey level -0) and it appears as black and white spots on the images. If p is the total noise density then salt noise and pepper noise will have a noise density of *p/2*.This can be mathematically represented as

$$y_{ij} = \begin{cases} \text{Zero or 255 with probability } p \\ x_{ij} \text{ with probability } 1\text{-}p \end{cases} \quad (1)$$

Where y_{ij} represents the noisy image pixel, p is the total noise density of impulse noise and x_{ij} is the uncorrupted image pixel. At times the salt noise and pepper noise may have different noise densities $p1$ and $p2$ and the total noise density will be $p=p1+p2$.

In case of random valued impulse noise, noise can take any gray level value from zero to 225. In this case also noise is randomly distributed over the entire image and probability of occurrence of any gray level value as noise will be same. We can mathematically represent random valued impulse noise as

Author α Ω β : B.V.C. Engineering College Odalarevu.

$$y_{ij} = \begin{cases} n_{ij} \text{ with probability } p \\ x_{ij} \text{ with probability } 1\text{-}p \end{cases} \quad (2)$$

Where n_{ij} is the gray level value of the noisy pixel.

III. Noise Filtering Algorithm

In this method, the detection map is constructed from the input noise image X. In case of salt-and-pepper noise image, the maximum and minimum intensity values of the image provide information about the corrupted pixels. For 8-bit gray scale image, the maximum and minimum intensity values are 0 and 255, indicating the pixel is corrupted with salt and pepper noise image. Considering this assumption, we assign a binary value to each elements $d_{i,j} \in D$ of the detection map D. The detection map is computed from the noisy image as follows:

$$d_{i,j} = \begin{cases} 1, \text{ if } X_{i,j} = 255. \\ 1, \text{ if } X_{i,j} = 255 \\ 0, \text{ otherwise.} \end{cases} \quad (3)$$

The entries 1 and 0 in the detection map D represent the noisy and noise free pixels respectively. The noise density is calculated as follows:

$$ND = \frac{\text{sum of 1' s in Detection map (D)}}{\text{N*N (total number of pixels in input image)}} \quad (4)$$

The noise density value ranges between 0 and 1. The Filtering algorithm for noise removal is as follows:

1. We use a small window Wxy neighborhood of size 3X3 at each pixel location (x,y) of the noisy image X and the detection map D.
2. For Each iteration, we count the number of noisy pixels in the detection map D. If the value of count K is positive integer and the central pixel within 3X3 window is noisy, then an array R is populated with noise free pixels. The length of the array, depending upon the noise density varies from zero to eight within the window.
3. We estimated the value of the noisy pixel by taking the median value of all noise-free pixels in array R.
4. Update the detection map $d_{i,j}$, based on the estimated value.
5. Steps 1 to 4 are repeated until we get the image with K=0, ie; noise –free image.

IV. Segmentation Algorithm Via Spline Regression

The algorithm for Image Segmentation using Spline Regression [6] is as follows:

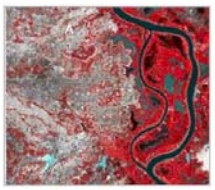
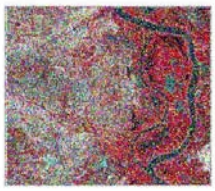
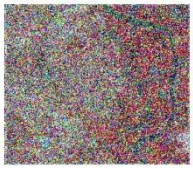
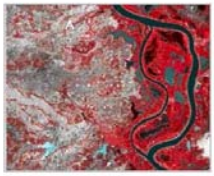
Input : The image I with n pixels to be segmented; the user specified stokes about the foreground object and its background, F and B; the number of clusters K for clustering F and B.

Output : The segmentation of I.

1. Construct the feature vector set $X=\{x_i\}$, for i=1 to n, in which $x_i = \{r,g,b,x,y\}^T$ corresponds to the feature vector of pixel p_i.
2. Construct two subsets of feature vectors according to the user specified strokes about the Foreground object and its background: $U_f=\{x_{if}\}$ where i=1 to n_f and $U_b=\{x_{ib}\}$ where i=1 to n_b.
3. if k > 0 then
4. Cluster U_f with K-means clustering algorithm, replace U_f by the k-cluster centers, and let $k= n_f$.
5. Cluster U_b with K-means clustering algorithm, replace U_b by the k- cluster centers, and let $k= n_b$.
6. end if
7. Construct the spline based on U_f and U_b.
8. Allocate an array S with n zero elements.
9. for each pixel p_i, i=1,…,n do
10. Calculate the spline regression value with $f(x_i)$.
11. Accumulate $f(x_i)$. to S[i] : $S[i]=S[i]+ f(x_i)$
12. Accumulate $f(x_i)$ to the eight neighbors of the i^{th} pixel and record them in the corresponding elements of S.
13. end for
14. for each pixel p_i, i=1,…,n, do
15. Average S[i], namely, S[i]=S[i]/9.
16. Assign class label for each pixel, ie., If S[i]>=0 then S[i]=255. Then (Foreground); otherwise, S[i]=0 , (Background).
17. end for
18. Output the binarized image S by reshaping it to be an image with the same size of source image I.

V. Experimental Results

The performance evaluation of our filtering algorithm is tested on the true color remote sensing image with 269X269 pixels. The salt-and-pepper noise is added into the image with two different noise densities 0.3 and 0.6. The images are filtered by using our proposed filtering algorithm.The performance of our algorithm is evaluated by computing segmentation using spline regression for the filtered image and the noise image.The experimental results are shown in Figure 2.

Original Image	Noisy Image (0.3)
	
Noisy Image (0.6)	Filtered image for noise density 0.3
	
Filtered image for noise density 0.6	
	

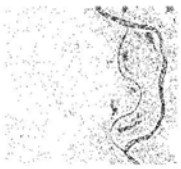
segmented Noisy Image (0.3)	Segmented Filtered Image
	

Figure 2 : Experimental Results

VI. Conclusion

In this paper we proposed a Noise filtering algorithm for removal of salt-and-pepper noise in Remote sensing images.The algorithm searches the noise free pixels with in a small neighborhood. The noisy pixel is then replaced with the median value estimated from noise free pixels.The experimental result shows that the proposed method is capable of removing salt-and-pepper noise more effectively, while preserving the fine image details and edges for the features extraction in remote sensing images.

References Références Referencias

1. Tao Chen, Kai-Kaung Ma and Li- Hui Chen, "Tri-state median filter for image Denoising", IEEE Transactions on Image Processing, Vol 8, no.12, pp.1834-1838, December 1999.
2. Reihard Berstein," Adaptive nonlinear filters for simultaneous removal of different kinds of noise in images," IEEE Trans on circuits and systems, Vol.cas-34, no 11, pp.127- 1291, Nivember 1987.
3. F.VanderHeijden, Image Based Measurement Systems. NewYork: Wiley, 1994.
4. K. Jain, Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1989.
5. S.Indu , Chaveli Ramesh, " A noise fading technique for images highly corrupted with impulse noise", Proceedings of the ICCTA07 , IEEE.
6. Shiming Xiang, Feiping Nie, Chunxia Zhang and Changshui Zhang," Interactive Natural Image Segmentation via Spline Regression", IEEE Transactiona in Image Processing , July 2009.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Knowledge Discovery from Web Logs – A Survey

By S. Chitra, Dr. B. Kalpana
Avinashilingam University for, Coimbatore

Abstract - Web usage mining is obtaining the interesting and constructive knowledge and implicit information from activities related to the WWW. Web servers trace and gather information about user interactions every time the user requests for particular resources. Evaluating the Web access logs would assist in predicting the user behavior and also assists in formulating the web structure. Based on the applications point of view, information extracted from the Web usage patterns possibly directly applied to competently manage activities related to e-business, e-services, e-education, on-line communities and so on. On the other hand, since the size and density of the data grows rapidly, the information provided by existing Web log file analysis tools may possibly provide insufficient information and hence more intelligent mining techniques are needed. There are several approaches previously available for web usage mining. The approaches available in the literature have their own merits and demerits. This paper focuses on the study and analysis of various existing web usage mining techniques.

Keywords : Web Usage Mining, Personalization, Pre -processing, Web Log, Navigation Patterns.

GJCST-C Classification : H.2.8

Strictly as per the compliance and regulations of:

Knowledge Discovery from Web Logs – A Survey

S. Chitra[α] , Dr. B. Kalpana[Ω]

Abstract - Web usage mining is obtaining the interesting and constructive knowledge and implicit information from activities related to the WWW. Web servers trace and gather information about user interactions every time the user requests for particular resources. Evaluating the Web access logs would assist in predicting the user behavior and also assists in formulating the web structure. Based on the applications point of view, information extracted from the Web usage patterns possibly directly applied to competently manage activities related to e-business, e-services, e-education, on-line communities and so on. On the other hand, since the size and density of the data grows rapidly, the information provided by existing Web log file analysis tools may possibly provide insufficient information and hence more intelligent mining techniques are needed. There are several approaches previously available for web usage mining. The approaches available in the literature have their own merits and demerits. This paper focuses on the study and analysis of various existing web usage mining techniques.

Keywords : Web Usage Mining, Personalization, Pre - processing, Web Log, Navigation Patterns.

I. Introduction

THE World Wide Web (WWW), is the current era of information explosion, has become the large source of online data, which includes text, graphics, videos, sound, etc. WWW is a comprehensive information medium in which the users can read, write and communicate through the use of computers connected to the Internet. Recent studies have estimated that the Web have more than one billion pages. It has become the powerful technology for sharing new ideas and content exchange. The impact of the Internet on everyday life is tremendous and it has changed the way of doing business, providing and receiving education, organization management, etc. The manner of information collection and sharing has changed with the advancement of hardware and communication software.

The growth has motivated the web service providers to predict the user's web usage behaviors so that, they can

- Personalize the information provided to them
- Make the websites more user friendly
- Reduce the traffic load
- Create or modify their website to suit different group of people.

The current requirement focuses on some tools which will help system analysts and business persons to learn user/consumer's needs, so that user requirements or demand can be solved immediately.

Web mining is the application of data mining techniques to web-based data for the purpose of learning or extracting knowledge. The techniques in web mining focus on providing solutions to content provider, web designer and programmers to improve their website and also to the web users with navigation assistance tools. It is a part of data mining where knowledge is gained from WWW.

Web servers trace and gather information about user interactions every time the user requests for particular resources. Evaluating the Web access logs would assist in predicting the user behavior and also assists in formulating the web structure.

Web usage mining, popularly also known as web log mining, works on the secondary data like web log file, click streams to extract knowledge with regard to web usage. It is the process which uses data mining techniques abundantly, the result of which can be used for several uses like personalization, system improvement and site modification.

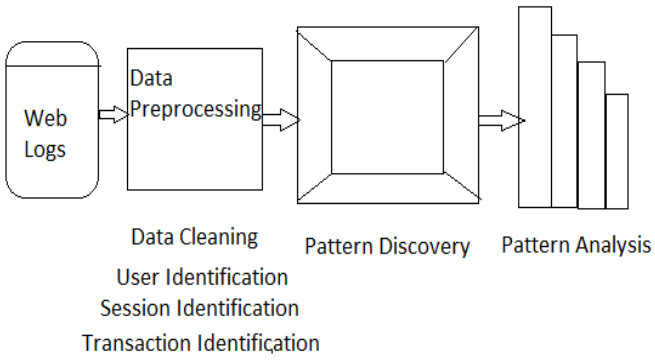
It is essential to investigate what kind of features a WUM system is estimated to have with the intention of performing effective and efficient Web usage mining, and what kind of challenges may be faced in the process of developing new Web usage mining techniques. A Web usage mining system should be able to:

- Gather useful usage data thoroughly
- Filter out irrelevant usage data
- Establish the actual usage data
- Discover interesting navigation patterns
- Display the navigation patterns clearly
- Analyze and interpret the navigation patterns correctly and
- Apply the mining results effectively.

Web usage mining comprises of three phases, specifically, preprocessing, pattern discovery and pattern analysis. Different phases of Web Usage Mining are depicted in Figure-1. There are several approaches previously available for web usage mining. This paper focuses on the study and analysis of various existing web usage mining techniques.

Author[α]: Assistant Professor, Government Arts College (Autonomous) Coimbatore - 641018.
Email : chitra.sivakumar@ymail.com
Author[Ω]: Associate Professor, Avinashilingam University for Women Coimbatore - 641043. Email : kalpanabsekar@yahoo.com

Figure 1: Phases of Web Usage Mining



II. Literature Survey

Web usage mining is a major application of data mining technology to extract the data of the Web server log file. It can find out the browsing behavior of user and a certain type of correlations among the web pages. Web usage mining offers the support for the Web site design, and also offers personalization server and further business decision making, etc. Web mining applies the data mining, the artificial intelligence and the chart technology and so on to the Web data and tracks the users browsing characteristics, and then obtains the users using pattern. Qingtian Han et al., [1] investigated on Web Mining Algorithm based on Usage Mining. And it also provides the design approach of the electronic commerce website application algorithm. This approach is easy, efficient and easy to understand, it is appropriate to the Web usage mining demand of building a low cost B2C website.

In order to enhance the Web site, it is necessary to estimate current usage of the particular website. Web usage mining and statistical analysis are two approaches to estimate usage of Web site. The integration of Web usage mining and statistical analysis provides more exact information about Web usage. With the help of Web usage mining approaches, graph mining focuses complex Web browsing patterns like parallel browsing. With the use of statistical analysis approaches, investigating the page browsing time provides precious information about Web site and its user's behavior. Heydari et al., [2] presented a Web usage mining technique which integrates Web usage mining and statistical analysis by taking into consideration of client side data. In additional way, it integrates graph based Web usage mining and browsing time examination with the consideration of client side data. It assists in rebuilding user session precisely as it has been and in accordance with these data, this Web usage patterns offers better accuracy.

Web usage mining has turn out to be very popular in different business fields associated with Web site development. In Web usage mining, frequently browsed navigational paths are obtained with the assistance of Web page addresses from the Web server visit logs, and the patterns are utilized in different applications together with recommendation. Normally the semantic information of the Web page contents is not considered in Web usage mining. Salin et al., [3] provided a structure for combining semantic information with Web usage mining. The common navigational patterns are obtained as the form of ontology instances rather than Web page addresses and the outcome is used for generating Web page recommendations to the visitor. Additionally, an assessment method is implemented with the intention of testing the accomplishment of the recommendation. Test result confirms that precise recommendations can be achieved by including semantic information in the Web usage mining.

Nasraoui at al., [4] presented a comprehensive structure and findings in mining Web usage patterns by using Web log files of a real Web site that has all the demanding aspects of real-life Web usage mining, together with evolving user profiles and external data describing an ontology of the Web content. Therefore, the authors present a technique for determining and tracing the mounting user profiles. The authors also discuss how the obtained users profiles can be improved with clear information obtained from search queries of Web log data. Profiles are also enhanced with additional domain-specific information aspects that provide a panoramic view of the discovered mass usage modes. Many experiments have been done by the author to assess the excellence of the mined profiles, especially their adaptability in the face of developing user behavior.

Web usage mining utilizes data mining methods to examine the user access of Web sites. As with any KDD (knowledge discovery and data mining) process, WUM comprises of three main phases: preprocessing, knowledge extraction, and results analysis. Tanasa at al., [5] concentrates on data preprocessing, a difficult and complicated phase. Analysts intend to find out the accurate list of users who browsed the Web site and to reconstitute user sessions-the order of actions every user carried out on the Web site. Inter-sites WUM focuses on the Web server logs from numerous Web sites, commonly belonging to the similar organization. Therefore, analysts must reconstruct the users' path through all the different Web servers that they browsed. This solution is to integrate all the log files and reconstitute the visit. Traditional data preprocessing comprises of three phases: data fusion, data cleaning, and data structurization. The author calls this solution as advanced data preprocessing. This technique comprises of a data summarization phase, which will permit the analyst to choose only the information of importance. The authors have effectively tested this technique in an experimentation with log files from INRIA Web sites.

Jianxi et al., [6] provides a Web usage mining method based on fuzzy clustering in identifying target group. Data mining is a procedure of non-trivial mining of inherent, formerly unidentified, and extremely useful data from very large quantity of data. Web mining can be defined mainly as the utilization of data mining techniques to Web data. Web usage mining is a noteworthy and fast developing area of Web mining where several researches has been carried our earlier. The author utilized the fuzzy clustering method for identifying groups that allocate comparable interests and behaviors by investigating the data gathered in Web servers.

Internet and Web technologies are extensively available, enabling it simpler for organizations to carry out business and transfer data to customers. Furthermore, they accelerate financial transactions competently by decreasing the transaction costs of commercial actions that businesses would generally incur. As a result, Internet business has generated aggressive surroundings, a flourishing organizations wanted to survive and increase a competitive advantage must offer a satisfactory package of customized services that convince customers' needs. Regardless of the Internet's apparent benefits as a novel communication medium its advertising provides the same advertising information to all customers and so has experienced from poor reactions. To increase a Web ad's usefulness, Sung Min Bae et al., [7] developed a Web ad selector with the intention of personalizing advertising information for customers according to their preferences and interests. The Web ad selection method segregates the Web site customers with comparable preferences into numerous segments through Web usage mining. It makes use of fuzzy rules that conveys customer segments' surfing patterns on the basis of specialist recommendation, and recommends suitable ads by fuzzy inference.

Wu et al., [8] recommended a Web Usage Mining technique according to the sequences of clicking patterns in a grid computing environment. Predicting user's browsing behavior is an important process of web usage mining. It can support the web designers to improve the web structure or enhance the performance of the web servers. Mining on the sequences of such clicking patterns (MSCP) can be considered as a data mining operation. MSCP is normally an expensive process because of its considerable quantity of time for computation and storage for archiving a huge quantity of information. Executing MSCP turns out to be unsuccessful or even not realistic on a computer with limited resources. The author finds out the handling of MSCP in a distributed grid computing environment and expresses its efficiency by experimental cases.

Web usage mining is a part of data mining technology to extract the data of the Web server log files. It can find out the session patterns of user and certain kinds of correlations among these Web pages. Web usage mining offers the support for the Web site design, by offering personalization server and additional business making decision. There are several session patterns accumulated in Web server log files, page attribute of the same is in Boolean quantity. With the purpose of enhancing the effectiveness of presented algorithms and decrease the time of scanning database, and consequently focusing to these aspects, Gang Fang at al., [9] proposes a double algorithm of Web usage mining in accordance with the sequence number that is appropriate for mining several session patterns. The algorithm transforms session pattern of particular user into binary, and subsequently uses up and down search approach to double generate candidate frequent itemsets. The algorithm works out support by sequence number dimension with the intention of scanning once session pattern of a particular user, which is dissimilar from conventional double search mining algorithm. in addition to this, the effectiveness of Web usage mining is competently enhanced because of this approach. The experiment result confirms that the efficiency is more rapid and more competent than the similar algorithms.

A lot of models are available and practices that analyze user behavior according to their user navigation data and use clustering algorithms to differentiate their access patterns. The navigation patterns recognized are predicted to satisfy the user's interests. Raghavendra et al., [10] modeled user behavior as a vector of the time the particular user spends at each URL, and additionally categorize a new user access pattern. The clustering and classification methods of k-means with non-Euclidean similarity measure, and artificial neural networks with consistent inputs were implemented and evaluated. Despite recognizing user behavior, this model can also be utilized as a prediction system in which it can be used to identify deviational behavior.

In Web Usage Mining (WUM), web session clustering plays a significant key part to categorize web visitors according to the user click history and comparison measure. Swarm dependent web session clustering assists in several ways to handle the web resources efficiently such as web personalization, schema modification, website modification and web server performance. Hussain et al., [11] propose a structure for web session clustering at initial level of web usage mining. The structure will cover the data preprocessing phase to organize the web log data and transform the categorical web log data into numerical data. A session vector is acquired, so that suitable comparison and swarm optimization possibly will be applied to cluster the web log data. The hierarchical cluster based technique will improve the existing web session techniques for additional structured information about the user sessions.

With the huge development of World Wide Web and e-commerce the investigation of users' navigation

patterns has developed into more significant. Predicting users' navigation behavior is a challenging subject for e-commerce enterprises. Web usage mining approaches can be utilized for modeling and predicting users' navigation patterns. In actual fact, mining users' navigation pattern is the basic approach for producing recommendations. In practice, the user interests are unpredictable, and it is complicated to follow the exact user navigation pattern. Khosravi et al., [12] proposed a technique based on naïve Bayesian method for modeling and predicting users' navigation behavior. The author has used Web server logs as source data, and carried out his experiment.

Huge volumes of data are collected automatically by Web servers and accumulated in access log files. Examination of server access data can offer important and helpful information. Web Usage Mining is the method of using data mining approaches to find the usage patterns by using the Web data and is aimed towards applications. It extracts the secondary data based on the interactions of the users throughout certain amount of Web sessions. Web usage mining contains three stages, that is, preprocessing, pattern discovery, and pattern analysis. Web usage mining has seen a huge increase of interest from the research people together with practice communities. Etminani et al., [13] applied Kohonen's SOM (Self Organizing Map) to pre-processed Web logs of Web server logs and extract frequent patterns. Experimental result of this technique confirms that this technique would be more useful for Web site owner.

In order to offer the online prediction efficiently, Shinde et al., [14] formulated a architecture for online recommendation for predicting in Web Usage Mining System. This approach provides the structural design of on-line recommendation system in Web usage mining (OLRWMS) for enhancing the exactness of classification by dealing between classifications, estimation, and provides user activities and user profile in online phase of this architecture.

Nowadays, Internet has turned out to be an essential tool for each individual, in the same way Web usage mining becomes a hotspot, which uses huge amounts of data in the Web server log and other appropriate data sets for mining analysis and achieves valuable knowledge model about usage of relevant Web site. At the moment, numerous works have to be performed with the positive association rules in Web usage mining, other than negative association rules is considerably more significant, Yang Bin at al., [15] have applied negative association rules approach to Web usage mining, in the course of the research the author have proved that the negative association rules have an additional significant role on access pattern to Web visitors, provide the mining algorithms, to resolve the deficiencies in which positive association rules are referred.

The recent development in Web technology has mounted the users and web pages at an exponential speed. The evolutionary modifications in tools have made it promising to confine the users' concentrate and communications with web applications through web server log file. Web log data is stored as text (.txt) file. Because of huge quantity amount of unrelated information in the web log, the initial log data can not be straightforwardly utilized in the web usage mining (WUM) process. As a result the preprocessing of web log data turns out to be very important. The appropriate examination of web log file is valuable to control the web sites efficiently for organizational and users' potential. Web log preprocessing is preliminary required process to enhance the quality and efficiency of the future processing of web usage mining. There are number of methods existing at preprocessing level of web usage mining. Various methods are utilized at preprocessing level like data cleaning, data filtering, and data integration. Hussain at al., [16] analyzed the existing the preprocessing methods to recognize the concerns and how WUM preprocessing can be enhanced for pattern mining and examination.

III. Problems and Directions

The main objective of web usage mining is to recognize the interesting web usage patterns. In order to recognize the interesting web usage patterns, a lot of researches are needed. They researches may focus on the following:

- To provide precise page recommendation, it is necessary to understand the browsing behavior of the user and it can be effectively done using the Machine Learning algorithms. This would definitely help in providing the accurate page recommendations according to the user needs.
- Based on the statistical analysis of the user, specifically, the amount of time user spending on a particular page, the kind of links on which the user is interested, number of visit on a particular page will help to understand the behavior of an user.
- The clustering technique can be used in grouping the user access pattern plays a significant role in determining the resemblance in used browsing sessions. Therefore, the clustering technique can be improved to enhance the performance of grouping the user sessions.

IV. Conclusion

Web log files play a significant role in the Web Usage Mining. An important knowledge that can be obtained from web log files is the user's navigation pattern. The challenge in obtaining such knowledge is that the users are constantly shifting their focus and different users have different navigational behavior with different needs associated with them. The navigation

pattern knowledge can be used to help users by predicting their future request and it will help on the personalization of websites. This paper provides the need for Web Usage Mining and various techniques which focus on the Web Usage Mining. The directions provided in this paper will assist the researchers to perform research on Web Usage Mining.

References Références Referencias

1. Qingtian Han; Xiaoyan Gao; Wenguo Wu; "Study on Web Mining Algorithm based on Usage Mining", 9th International Conference on Computer-Aided Industrial Design and Conceptual Design (CAID/CD 2008), Pp. 1121 – 1124, 2008.
2. Heydari, M.; Helal, R.A.; Ghauth, K.I.; "A graph-based web usage mining method considering client side data", International Conference on Electrical Engineering and Informatics (ICEEI '09), Vol. 1, Pp. 147 – 153, 2009.
3. Salin, S.; Senkul, P.; "Using semantic information for web usage mining based recommendation", 24th International Symposium on Computer and Information Sciences (ISCIS 2009), Pp. 236 – 241, 2009.
4. Nasraoui, O.; Soliman, M.; Saka, E.; Badia, A.; Germain, R.; "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 2, Pp. 202 – 215, 2008.
5. Tanasa, D.; Trousse, B.; "Advanced data preprocessing for intersites Web usage mining", IEEE Intelligent Systems, Vol. 19, No. 2, Pp. 59 – 65, 2004.
6. Jianxi Zhang, Peiying Zhao, Lin Shang and Lunsheng Wang, "Web Usage Mining Based On Fuzzy Clustering in Identifying Target Group", International Colloquium on Computing, Communication, Control, and Management, Vol. 4, Pp. 209-212, 2009.
7. Sung Min Bae; Sang Chan Park; Sung Ho Ha; "Fuzzy Web ad selector based on Web usage mining", Vol. 18, No. 6, Pp. 62 – 69, 2003.
8. Chih-Hung Wu, Yen-Liang Wu, Yuan-Ming Chang and Ming-Hung Hung, "Web Usage Mining on the Sequences of Clicking Patterns in a Grid Computing Environment", International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 6, Pp. 2909-2914, 2010.
9. Gang Fang; Jia-Le Wang; Hong Ying; Jiang Xiong; "A Double Algorithm of Web Usage Mining Based on Sequence Number", International Conference on Information Engineering and Computer Science (ICIECS), Pp. 1 – 4, 2009.
10. Raghavendra, P.S.; Chowdhury, S.R.; Kameswari, S.V.; "Comparative study of neural networks and k-means classification in web usage mining", International Conference for Internet Technology and Secured Transactions (ICITST), Pp. 1-7, 2010.
11. Hussain, T.; Asghar, S.; Fong, S.; "A hierarchical cluster based preprocessing methodology for Web Usage Mining", 6th International Conference on Advanced Information Management and Service (IMS), Pp. 472 – 477, 2010.
12. Khosravi, M.; Tarokh, M.J.; "Dynamic mining of users interest navigation patterns using naive Bayesian method", IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Pp. 119 – 122, 2010.
13. Etminani, K.; Delui, A.R.; Yanehsari, N.R.; Rouhani, M.; "Web usage mining: Discovery of the users' navigational patterns using SOM", First International Conference on Networked Digital Technologies (NDT '09), Pp. 224 – 249, 2009.
14. Shinde, S.K. and Kulkarni, U.V., "A New Approach for on Line Recommender System in Web Usage Mining", International Conference on Advanced Computer Theory and Engineering, Pp. 973- 977, 2008.
15. Yang Bin; Dong Xiangjun; Shi Fufu; "Research of WEB Usage Mining Based on Negative Association Rules", International Forum on Computer Science-Technology and Applications (IFCSTA '09), Vol. 1, Pp. 196 – 199, 2009.
16. Hussain, T.; Asghar, S.; Masood, N.; "Web usage mining: A survey on preprocessing of web log file", International Conference on Information and Emerging Technologies (ICIET), Pp. 1 – 6, 2010.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Studies on Image Segmentation Method Based On a New Symmetric Mixture Model with – K Means

By M.Seshashayee, K.Srinivasa Rao, Ch.Satyanarayana, P.Srinivasa Rao
GITAM University, Visakhapatnam

Abstract - In this paper we propose an image segmentation algorithm based on new-symmetric mixture model. Here the pixel intensities of the whole image are characterized through a new-symmetric mixture distribution, such that the statistical characteristics of the image coincide with that of the new symmetric distribution. Using the K-Means algorithm the number of image regions and initial estimates of the model parameters for the EM algorithm are obtained. The segmentation algorithm is proposed by component maximum likelihood under Bayesian frame work. The efficiency of the proposed method is studied with the five images taken from the Berkeley image dataset and computing the values image segmentation measures like global consistency error, probabilistic rand index and variation of information. A comparative study of the proposed model with Gaussian mixture model reveals that the proposed method performs better. The efficiency of the proposed method with respect to the image retrieval is also studied.

Keywords : *Image Segmentation, New Symmetric Mix- ture Model, Image Quality Metrics, K-means algorithm, EM algorithm.*

GJCST-H Classification : *I.4.6, I.3.3*

Strictly as per the compliance and regulations of:

Studies on Image Segmentation Method Based On a New Symmetric Mixture Model with K – Means

M.Seshashayee[α], K.Srinivasa Rao[Ω], Ch.Satyanarayana[β], P.Srinivasa Rao[ψ]

Abstract - In this paper we propose an image segmentation algorithm based on new-symmetric mixture model. Here the pixel intensities of the whole image are characterized through a new-symmetric mixture distribution, such that the statistical characteristics of the image coincide with that of the new symmetric distribution. Using the K-Means algorithm the number of image regions and initial estimates of the model parameters for the EM algorithm are obtained. The segmentation algorithm is proposed by component maximum likelihood under Bayesian frame work. The efficiency of the proposed method is studied with the five images taken from the Berkeley image dataset and computing the values image segmentation measures like global consistency error, probabilistic rand index and variation of information. A comparative study of the proposed model with Gaussian mixture model reveals that the proposed method performs better. The efficiency of the proposed method with respect to the image retrieval is also studied.

Keywords : Image Segmentation, New Symmetric Mixture Model, Image Quality Metrics, K-means algorithm, EM algorithm.

I. Introduction

Image segmentation is a preprocessing step in image analysis and understanding. Much work has been reported in literature regarding image segmentation. Pal S.K. and Pal N.R (1993), Jahne (1995), Cheng et al (2001), Mantas Paulinas and Audrius Usinskas (2007) and Shital Raut et al (2009) have discussed various image segmentation methods.

The image segmentation methods are usually classified into three categories namely (i) segmentation methods based on histogram, threshold and edge based techniques, (ii) model based image segmentation methods and (iii) image segmentation based on other methods like graph, saddle point, neural networks, fuzzy logic etc., (Caillol H. et al (1993),Tolias Y.A. and Pamas S.M (1998), Brun L. (1998), Xu Y. et al (1998)). Among these methods model based image segmentation is more efficient since it preserves the neighborhood information and characterizes the features of the image region more accurately. Hence much emphasis is given for image segmentation based on finite Gaussian mixture model (Yamazaki et .al(1998) , Lie T. et al(1993), Zhang Z.H. et al (2003) and Nasios N. et al(2006)).

In Gaussian mixture model the whole image is characterized by the collection of several image regions, where each region is characterized by a Gaussian distribution. That is the pixel intensities in each image region follow a Gaussian distribution. This Gaussian assumption serves well only when the pixel intensities in each image region are meso-kurtic and symmetric. But in some images like natural scenes the pixel intensities of the image region may not be meso-kurtic even though they are symmetric. Hence to have an accurate analysis of the images, it is needed to develop image segmentation methods based on Non-Gaussian mixture models.

In Non-Gaussian symmetric mixture models the kurtosis plays a dominant role. Based on the kurtosis the Non-Gaussian models can be classified into two categories platy-kurtic and lepto-kurtic. In general many of the natural scenes will have image regions having platy-kurtic nature. That is the kurtosis of the pixel intensities in the image regions is less than three. One such model available in literature is new-symmetric distribution given by Srinivasa Rao K. et al (1997). The new-symmetric distribution is having kurtosis 2.52 and symmetric. So it is a platy-kurtic distribution. Hence to have an efficient image segmentation algorithm for images having platy-kurtic distributed pixel intensities in the image regions, we develop and analyze an image segmentation algorithm based on new-symmetric mixture model.

For developing the image segmentation algorithm we require the number of components in the image. This is obtained from K-means algorithm. The initial estimates of the model parameters are obtained from the moment estimates. The updated equations for estimating the model parameters through the EM algorithm are derived. The segmentation algorithm is also presented by taking component maximum likelihood. The efficiency of the proposed algorithm is studied through experimentation.

Author[α] : Department of CS, GITAM University, Visakhapatnam-45. E-mail : sesha_shayee@yahoo.co.in
Author[Ω] : Department of Statistics, Andhra University, Visakhapatnam-03. E-mail : ksraoau@yahoo.co.in
Author[β] : Department of CSE, JNTU-KAKINADA, KAKINADA E-mail : chsatyanarayana@yahoo.com
Author[ψ] : Department of CS&SE, Andhra University, Visakhapatnam-03. E-mail : peri.srinivasarao@yahoo.com

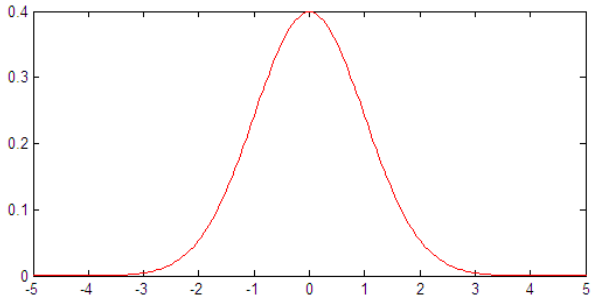
II. Finite Mixture of New Symmetric Distribution

In low level image analysis the entire image is considered as a union of several image regions. In each image region the image data is quantized by pixel intensities. For a given point (pixel) (x , y), the pixel intensity z = f(x , y) is a random variable, because of the fact that the brightness measured at a point in the image is influenced by various random factors like vision, lighting, moisture, environmental conditions etc,. To model the pixel intensities of the image region it is assumed that the pixel intensities of the region follow a new symmetric distribution given by Srinivasa Rao K. et al., (1997). The probability density function of the pixel intensity is

$$f(Z,\mu,\sigma^2)=\frac{\left(2+\left(\frac{z-\mu}{\sigma}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}}{3\sigma\sqrt{2\pi}},\ -\infty<Z<\infty,-\infty<\mu<\infty,\sigma>0 \qquad (1)$$

The probability curve of new symmetric distribution is shown in Figure 1.

Figure 1 : Probability curve of new symmetric distribution



Its central moments are

$$\mu_{2n}=\left[\frac{(n+\frac{3}{2})\Gamma(n+\frac{1}{2})}{(3/2)\sqrt{\pi}}\right]2^n\sigma^{2n} \quad \text{and} \quad \mu_{2n+1}=0 \qquad (2)$$

The kurtosis of the distribution is $\beta_2=2.52$ (3)

The entire image is a collection of regions which are characterized by new symmetric distribution. Here, it is assumed that the pixel intensities of the whole image follow a K – component mixture of new symmetric distribution and its probability density function is of the form.

$$p(z)=\sum_{i=1}^{K}\alpha_i\, f_i(z/\mu_i,\sigma_i^2) \qquad (4)$$

where, K is number of regions , $0\le\alpha_i\le 1$ are weights such that $\sum\alpha_i=1$ and $f_i(z,\mu,\sigma^2)$ is as given in equation (1). α_i is the weight associated with ith region in the whole image.

In general the pixel intensities in the image regions are statistically correlated and these correlations can be reduced by spatial sampling (Lie.T and Sewehand. W(1992)) or spatial averaging (Kelly P.A. et al.(1998)) . After reduction of correlation the pixels are considered to be uncorrelated and independent. The mean pixel intensity of the whole image is

$$E(Z)=\sum_{i=1}^{K}\alpha_i\mu_i\ .$$

.

III. Estimation of The Model Parameters by Em Algorithm

In this section we derive the updated equations of the model parameters using Expectation Maximization (EM) algorithm. The likelihood function of the observations z1,z2,z3,…,zN drawn from an image is

$$L(\theta)=\prod_{S=1}^{N}p(z_s,\theta^{(l)})\ .$$

That is $$L(\theta)=\prod_{S=1}^{N}\left(\sum_{i=1}^{K}\alpha_i\, f_i(z_S,\theta)\right)$$

$$\log L(\theta)=\sum_{s=1}^{N}\log\left(\sum_{i=1}^{K}\alpha_i\, f_i(z_s,\theta_i)\right),$$

Where $\theta=(\mu_i,\sigma_i^2,\alpha_i;i=1,2,...,K)$ is the set of parameters

$$\log L(\theta)=\sum_{s=1}^{N}\log\left[\sum_{i=1}^{K}\frac{\alpha_i\left(2+\left(\frac{z_s-\mu_i}{\sigma_i}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-\mu_i}{\sigma_i}\right)^2}}{3\sigma_i\sqrt{2\pi}}\right] \qquad (5)$$

The first step of the EM algorithm requires the estimation of the likelihood function of the sample observations. The expectation of the log likelihood function of the sample is

$$Q\left(\theta;\theta^{(l)}\right)=E_{\theta^{(l)}}\left[\log L(\theta)/\bar{z}\right]$$

Following the heuristic arguments of Jeff A. Bilmes (1997) we have

$$Q\left(\theta;\theta^{(l)}\right)=\sum_{i=1}^{K}\sum_{s=1}^{N}\left(t_i(z_s,\theta^{(l)})\left(\log f_i(z_s,\theta)+\log\alpha_i\right)\right) \qquad (6)$$

The updated equation of α for $(l+1)^{th}$ th estimate is

$$\alpha_i^{(l+1)}=\frac{1}{N}\sum_{s=1}^{N}t_i(z_s,\theta^{(l)})$$

$$= \frac{1}{N}\sum_{s=1}^{N}\left[\frac{\alpha_i^{(l)} f_i(z_s,\theta^{(l)})}{\sum_{i=1}^{K}\alpha_i^{(l)} f_i(z_s,\theta^{(l)})}\right] \quad (7)$$

The updated equation of α at $(l+1)^{th}$ iteration is

$$\mu_i^{(l+1)} = \frac{\sum_{s=1}^{N} z_s t_i(z_s,\theta^{(l)}) - \sum_{s=1}^{N} t_i(z_s,\theta^{(l)})\left(\frac{2\sigma_i^{2(l)}\left(z_s-\mu_i^{(l)}\right)}{2\sigma_i^{2(l)}+\left(z_s-\mu_i^{(l)}\right)^2}\right)}{\sum_{s=1}^{N} t_i(z_s,\theta^{(l)})} \quad (8)$$

where, $t_i(z_s,\theta^{(l)}) = \frac{\alpha_i^{(l+1)} f_i(z_s,\mu_i^{(l)},(\sigma_i^2)^{(l)})}{\sum_{i=1}^{K}\alpha_i^{(l+1)} f_i(z_s,\mu_i^{(l)}(\sigma_i^2)^{(l)})}$

The updated equation of σ_i^2 at $(l+1)^{th}$ iteration is

$$(\sigma_i^2)^{(l+1)} = \frac{2\sum_{s=1}^{N}(z_s-\mu_i^{(l+1)})^2\left(\frac{1}{2}-\frac{(\sigma_i^2)^{(l)}}{\left(2\sigma_i^{2(l)}+\left(z_s-\mu_i^{(l+1)}\right)^2\right)^2}\right)\left(t_i(z_s,\theta^{(l)})\right)}{\sum_{s=1}^{N} t_i(z_s,\theta^{(l)})} \quad (9)$$

where $t_i(z_s,\theta^{(l)}) = \frac{\alpha_i^{(l+1)} f_i(z_s,\mu_i^{(l+1)},(\sigma_i^2)^{(l)})}{\sum_{i=1}^{K}\alpha_i^{(l+1)} f_i(z_s,\mu_i^{(l+1)},(\sigma_i^2)^{(l)})}$

IV. Initialization of The Parameters by K – means

The efficiency of the EM algorithm in estimating the parameters is heavily dependent on the number of regions in the image. The number of mixture components initially taken for K – Means algorithm is by plotting the histogram of the pixel intensities of the whole image. The number of peaks in the histogram can be taken as the initial value of the number of regions K.

The mixing parameters α_i and the model parameters μ_i, σ_i^2 are usually considered as known apriori. A commonly used method in initializing parameters is by drawing a random sample from the entire image Mclanchan G and Peel D (2000). This method performs well if the sample size is large and its computational time is heavily increased. When the sample size is small, some small regions may not be sampled. To overcome this problem we use the K – Means algorithm to divide the whole image into various homogeneous regions. In K – Means algorithm the centroids of the clusters are recomputed as soon as the pixel joins a cluster.

After determining the final values of K (number of regions) , we obtain the initial estimates of μ_i, σ_i^2 and α_i for the i^{th} region using the segmented region pixel intensities with the method given by Srinivasa Rao et al.,(1997) for new symmetric distribution .The initial estimate α_i is taken as $\alpha_i = \frac{1}{K}$, where i = 1,2,...,K. The parameters μ_i and σ_i^2 are estimated by the method of moments as $\mu_i = \bar{z}$ and $\sigma_i^2 = \frac{4n}{3(n-1)}S^2$ where, S^2 is the sample variance.

V. Segmentation Algorithm

In this section, we present the image segmentation algorithm. After refining the parameters the prime step in image segmentation is allocating the pixels to the segments of the image. This operation is performed by Segmentation Algorithm. The image segmentation algorithm consists of four steps.

Step 1) Plot the histogram of the whole image.
Step 2) Obtain the initial estimates of the model parameters using K-Means algorithm and moment estimators as discussed in section 4
Step 3) Obtain the refined estimates of the model parameters μ_i, σ_i^2 and α_i for i=1,2,...,K by using the EM algorithm with the updated equations
Step 4) Assign each pixel into the corresponding j^{th} region (segment) according to the maximum likelihood of the j^{th} component L_j.
That is ,

$$L_j = \max_{j\in k}\left[(3\sigma_j\sqrt{2\pi})^{-1}\left(2+\left(\frac{z_s-\mu_j}{\sigma_j}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-\mu_j}{\sigma_j}\right)^2}\right],$$

$$-\infty < z_s < \infty,\ -\infty < \mu_j < \infty, \sigma_j > 0$$

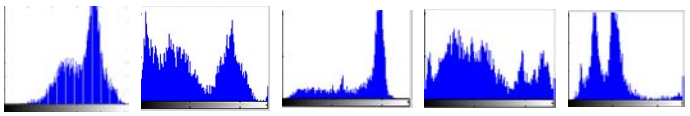
VI. Experimental Results

To demonstrate the utility of the image segmentation algorithm developed in this chapter, an experiment is conducted with five images taken from Berkeley images dataset (http://www.eecs.berkeley.edu/Research/Projects/CS/Vision/bsds/BSDS300/html).

The images HORSE, MAN, BIRD, BOAT and TOWER are considered for image segmentation. The pixel intensities of the whole image are taken as feature. The pixel intensities of the image are assumed to follow a mixture of new symmetric distribution. That is, the image contains K regions and pixel intensities in each image region follow a new symmetric distribution with

different parameters. The number of segments in each of the five images considered for experimentation is determined by the histogram of pixel intensities. The histograms of the pixel intensities of the five images are shown in Figure 2.

Figure 2 : Histograms Of The Images



HORSE MAN BIRD BOAT TOWER

The initial estimates of the number of the regions K in each image are obtained and given in Table 1.

Table 1: Initial Estimates of K

IMAGE	HORSE	MAN	BIRD	BOAT	TOWER
Estimate of K	2	4	3	4	3

From Table 1, we observe that the image HORSE has two segments, images TOWER and BIRD have three segments each and images MAN and BOAT have four segments each. The initial values of the model parameters μ_i , σ_i^2 and α_i for i = 1, 2,...,K for each image region are computed by the method given in section 3.

Using these initial estimates and the updated equations of the EM Algorithm given in Section 3 the final estimates of the model parameters for each image are obtained and presented in tables 2.a, 2.b, 2.c, 2.c, 2.d ,and 2.e for different images.

Table : 2.a

Estimated Values of the Parameters for HORSE Image

Number of Image Regions (K =2)

Parameters	Estimation of Initial Parameters		Estimation of Final Parameters by EM Algorithm	
	Regions(i)		Regions(i)	
	1	2	1	2
α_i	1/2	1/2	0.39702	0.60298
μ_i	121.47	187.91	134.09	184.97
σ_i^2	609.82	426.21	1302.8	561.41

Table : 2.b

Estimated Values of the Parameters for MAN Image

Number of Image Regions (K =4)

Parameters	Estimation of Initial Parameters				Estimation of Final Parameters by EM Algorithm			
	Regions(i)				Regions(i)			
	1	2	3	4	1	2	3	4
α_i	1/4	1/4	1/4	1/4	0.24315	0.2306	0.34648	0.17977
μ_i	63.5	20.234	184.29	106.38	64.541	23.197	183.65	103.01
σ_i^2	190.98	165.05	547.54	361.45	497.03	214.15	509.25	1074.40

Table : 2.c

Estimated Values of the Parameters for BIRD Image

Number of Image Regions (K =3)

Parameters	Estimation of Initial Parameters			Estimation of Final Parameters by EM Algorithm		
	Regions(i)			Regions(i)		
	1	2	3	1	2	3
α_i	1/3	1/3	1/3	0.13161	0.66786	0.20053
μ_i	53.491	124.05	124.05	60.691	192.85	129.81
σ_i^2	535.4	513.93	513.93	857.07	86.799	1581.2

Table : 2.d

Estimated Values of the Parameters for BOAT Image
Number of Image Regions (K =4)

Estimation of Initial Parameters					Estimation of Final Parameters by EM Algorithm			
	Regions(i)				Regions(i)			
Parameters	1	2	3	4	1	2	3	4
α_i	1/4	1/4	1/4	1/4	0.2570	0.24231	0.28458	0.22741
μ_i	34.98	216.5	81.146	131.13	41.008	212.7	81.062	128.11
σ_i^2	374.1	657.54	259.39	387.02	636.2	699.25	785.09	881.93

Table : 2.e

Estimated Values of The Parameters For TOWER Image
Number of Image Regions (K =3)

Parameters	Estimation of Initial Parameters			Estimation of Final Parameters by EM Algorithm		
	Regions(i)			Regions(i)		
	1	2	3	1	2	3
α_i	1/3	1/3	1/3	0.43267	0.051312	0.51602
μ_i	55.663	223.75	107.79	60.79	193.31	104.42
σ_i^2	276.53	1082.4	297.62	487.89	3140.4	404.79

Substituting the final estimates of the model parameters, the probability density function of pixel intensities of each image are estimated.

The estimated probability density function of the pixel intensities of the image HORSE is

$$f\left(z_s,\theta^{(l)}\right)=\frac{(0.39702)\left(2+\left(\frac{z_s-134.09}{36.0943}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-134.09}{36.0943}\right)^2}}{(36.0943)(3)\sqrt{2\pi}}$$

$$+\frac{(0.60298)\left(2+\left(\frac{z_s-184.97}{23.6941}\right)^2\right)^1 e^{\frac{-1}{2}\left(\frac{z_s-184.97}{23.6941}\right)^2}}{(23.6941)(3)\sqrt{2\pi}}$$

The estimated probability density function of the pixel intensities of the image MAN is

$$f\left(z_s,\theta^{(l)}\right)=\frac{(0.24315)\left(2+\left(\frac{z_s-64.541}{22.2942}\right)^2\right)e^{\frac{-1}{2}\left(z_s-64.541\right)^2}}{(22.2942)(3)\ \ 2\pi}$$

$$+\frac{(0.2306)\left(2+\left(\frac{z_s-23.197}{14.6339}\right)^2\right)e^{\frac{-1}{2}\left(z_s-23.197\right)^2}}{(14.6339)(3)\ \ 2\pi}$$

$$+\frac{(0.34648)\left(2+\left(\frac{z_s-183.65}{22.5666}\right)^2\right)e^{\frac{-1}{2}\left(z_s-183.65\right)^2}}{(22.5666)(3)\ \ 2\pi}$$

$$+\frac{(0.17977)\left(2+\left(\frac{z_s-103.01}{32.7780}\right)^2\right)^2 e^{\frac{-1}{2}\left(z_s-103.01\right)^2}}{(32.7780)(3)\ \ 2\pi}$$

The estimated probability density function of the pixel intensities of the image BIRD is

$$f\left(z_s,\theta^{(l)}\right)=\frac{(0.13161)\left(2+\left(\frac{z_s-60.691}{29.2758}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-60.691}{29.2758}\right)^2}}{(29.2758)(3)\sqrt{2\pi}}$$

$$+\frac{(0.66786)\left(2+\left(\frac{z_s-192.85}{9.3166}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-192.85}{9.3166}\right)^2}}{(9.3166)(3)\sqrt{2\pi}}$$

$$+\frac{(0.20053)\left(2+\left(\frac{z_s-129.81}{39.7643}\right)^2\right)^1 e^{\frac{-1}{2}\left(\frac{z_s-129.81}{39.7643}\right)^2}}{(39.7643)(3)\sqrt{2\pi}}$$

The estimated probability density function of the pixel intensities of the image BOAT is

$$f\left(z_s,\theta^{(l)}\right)=\frac{(0.2570)\left(2+\left(\frac{z_s-41.008}{25.1478}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-41.008}{25.1478}\right)^2}}{(25.14)(3)\sqrt{2\pi}}$$

$$+\frac{(0.24231)\left(2+\left(\frac{z_s-212.7}{26.4433}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-212.7}{26.4433}\right)^2}}{(26.4433)(3)\sqrt{2\pi}}$$

$$+\frac{(0.28458)\left(2+\left(\frac{z_s-81.062}{28.0195}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-81.062}{28.0195}\right)^2}}{(28.0195)(3)\sqrt{2\pi}}$$

$$+\frac{(0.22741)\left(2+\left(\frac{z_s-128.11}{29.6973}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-128.11}{29.6973}\right)^2}}{(29.6973)(3)\sqrt{2\pi}}$$

The estimated probability density function of the pixel intensities of the image TOWER is

$$f\left(z_s,\theta^{(l)}\right)=\frac{(0.43267)\left(2+\left(\frac{z_s-60.79}{22.0882}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-60.79}{22.0882}\right)^2}}{(22.0882)(3)\sqrt{2\pi}}$$

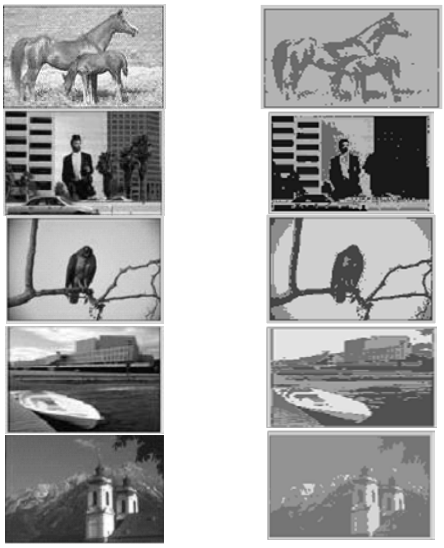
$$+\frac{(0.051312)\left(2+\left(\frac{z_s-193.31}{56.0393}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-193.31}{56.0393}\right)^2}}{(56.0393)(3)\sqrt{2\pi}}$$

$$+\frac{(0.51602)\left(2+\left(\frac{z_s-104.42}{20.1194}\right)^2\right)e^{\frac{-1}{2}\left(\frac{z_s-104.42}{20.1194}\right)^2}}{(20.1194)(3)\sqrt{2\pi}}$$

Using the estimated probability density function and image segmentation algorithm given in section 5, the image segmentation is done for the five images under consideration. The original and segmented images are shown in Figure 3.

Figure 3 : Original and Segmented Images

ORIGINAL IMAGES SEGMENTED IMAGES



VII. Performance Evalution

After conducting the experiment with the image segmentation algorithm developed in this chapter, its performance is studied. The performance evaluation of the segmentation technique is carried by obtaining the three performance measures namely, (i) Probabilistic Rand Index (PRI), (ii) Variation Of Information (VOI) and (iii) Global Consistence Error (GCE). The performance of developed algorithm using finite new symmetric distribution mixture model (NSMM-K) is studied by computing the segmentation performance measures namely PRI, GCE, and VOI for the five images under study. The computed values of the performance measures for the developed algorithm and the earlier existing finite Gaussian mixture model(GMM) with K-Means algorithm are presented in Table 3 for a comparative study.

Table 3 : Segmentation Performace Measures

IMAGES	METHOD	PERFORMACE MEASURES		
		PRI	GCE	VOI
HORSE	GMM	0.9142	0.1737	1.8643
	NSMM-K	0.9283	0.1634	1.8403
MAN	GMM	0.9228	0.3107	1.8389
	NSMM-K	0.9342	0.1734	1.7875
BIRD	GMM	0.9106	0.1369	1.7479
	NSMM-K	0.9140	0.1352	1.7259
BOAT	GMM	0.9026	0.6485	1.7882
	NSMM-K	0.9174	0.6483	1.7542
TOWER	GMM	0.9102	0.1090	1.8643
	NSMM-K	0.9246	0.0981	1.7988

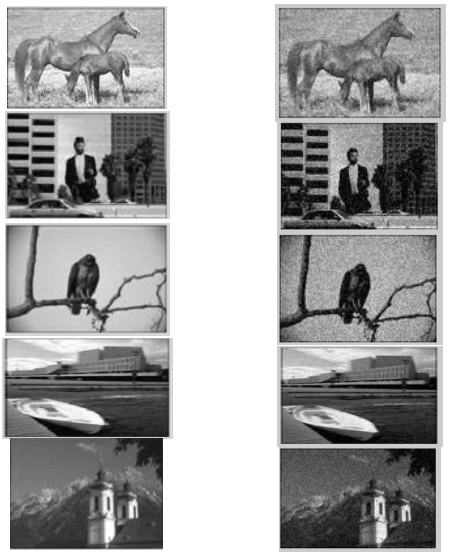
From table 3 it is observed that the PRI values of the proposed algorithm for the five images considered for experimentation are less than that of the values from the segmentation algorithm based on finite Gaussian mixture model with K-means. Similarly GCE and VOI values of the proposed algorithm are less than that of Finite Gaussian Mixture Model. This reveals that the proposed algorithm outperforms the existing algorithm based on the finite Gaussian mixture model. When the kurtosis parameter of each component of the model is zero, the model reduces to finite Gaussian mixture model and even in this case the algorithm performs well.

After developing the image segmentation method it is needed to verify the utility of segmentation in model building of the image for image retrieval.The performance evaluation of the retrieved image can be done by subjective image quality testing or by objective image quality testing. The objective image quality testing methods are often used since the numerical results of an objective measure allows a consistent comparison of different algorithms. There are several image quality measures available for performance evaluation of the image segmentation method. An extensive survey of quality measures is given by Eskicioglu A.M. and Fisher P.S. (1995). For the performance evaluation of the developed segmentation algorithm, we consider the image quality measures like average difference, maximum distance, image fidelity, mean square error, signal to noise ratio and image quality index.

Using the estimated probability density functions of the images under consideration the retrieved images are obtained and are shown in Figure 4.

Figure 4 : The Original and Retrieved Images

ORIGINAL IMAGES RETRIEVED IMAGES



The image quality measures are computed for the five retrieved images HORSE, MAN, BIRD, BOAT AND TOWER using the proposed model and FGMM with K-means and their values are given in the Table 4.

Table 4 : Comparative Study of Image Quality Metrics

IMAGE	Quality Metrics	FGMM	FNSDMM with K-Means	Standard Limits
HORSE	Average Difference	0.5011	0.44135	Close to 1
	Maximum Distance	1.0000	1.0000	Close to 1
	Image Fidelity	1.0000	1.0000	Close to 1
	Mean Square Error	0.5011	0.4414	Close to 0
	Signal to Noise Ratio	5.6542	5.9301	As big as possible
	Image Quality Index	1.0000	1.0000	Close to 1
MAN	Average Difference	0.4858	0.50021	Close to 1
	Maximum Distance	1.0000	1.0000	Close to 1
	Image Fidelity	1.0000	1.0000	Close to 1
	Mean Square Error	0.4995	0.5079	Close to 0
	Signal to Noise Ratio	5.6828	5.6251	As big as possible
	Image Quality Index	1.0000	1.0000	Close to 1
BIRD	Average Difference	0.4939	0.6573	Close to 1
	Maximum Distance	1.0000	1.0000	Close to 1
	Image Fidelity	1.0000	1.0000	Close to 1
	Mean Square Error	0.8590	0.5050	Close to 0
	Signal to Noise Ratio	5.6861	4.4842	As big as possible
	Image Quality Index	1.000	1.0000	Close to 1
BOAT	Average Difference	0.5039	0.6217	Close to 1
	Maximum Distance	1.0000	1.0000	Close to 1
	Image Fidelity	1.0000	1.0000	Close to 1
	Mean Square Error	0.7931	0.5070	Close to 0
	Signal to Noise Ratio	5.6318	4.6573	As big as possible
	Image Quality Index	1	1.0000	Close to 1
	Average Difference	0.4936	0.6640	Close to 1
	Maximum Distance	1.0000	1.0000	Close to 1
TOWER	Image Fidelity	0.9999	0.9999	Close to 1
	Mean Square Error	0.8788	0.5076	Close to 0
	Signal to Noise Ratio	5.6870	4.4347	As big as possible
	Image Quality Index	1.0000	1.0000	Close to 1

From the Table 4, it is observed that all the image quality measures for the five images are meeting the standard criteria. This implies that using the proposed algorithm the images are retrieved accurately. A comparative study of proposed algorithm with that of algorithm based on Finite Gaussian Mixture Model reveals that the MSE of the proposed model is less than that of the finite Gaussian mixture model. Based on all other quality metrics also it is observed that the performance of the proposed model in retrieving the images is better than the finite Gaussian mixture model.

VIII. Conclusion

An image segmentation algorithm based on new symmetric mixture model with K-means is developed and evaluated. This algorithm is more suitable for the images having platy-kurtic image regions. The new symmetric mixture model is capable of characterizing several natural images with kurtosis close to 2.52. The updated equations of the model parameters are derived through EM algorithm under Bayesian framework. The estimated probability density function of the pixel intensities in the whole image is useful for the image retrieval. The experimental results revealed that the proposed method out performs the existing Gaussian mixture model in both image segmentation and image retrieval.

References Références Referencias

1. Caillol H. et al (1993), "Fuzzy Random Fields and Unsupervised Image Segmentation", IEEE Transactions on Geo-Science and Remote Sensing, Vol.31, No.4, pp.801-810.
2. Cheng et al (2001) "Color Image Segmentation: Advances and Prospects" Pattern Recognition, Vo1.34, pp 2259-2281.
3. D . Martin, C. Fowlkes,D. Tal, and J.Malik, (2001) " A database of human Segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in proceedings of 8th International Conference on Computer vision, Vol.2, pp.416-423.
4. Eskicioglu M.A and Fisher P.S (1995) "Image Quality Measures and their Performance", IEEE Transactions On Communications, Vol.43, No.12.
5. Fang Qian et al (2002), "Gaussian Mixture Model For Relevance Feedback In Image Retrieval", IEEE

Transactions on Image Processing, Vol.1 , pp. 229-232.
6. Jahne (1995), " A Practical Hand Book on Image segmentation for Scientific Applications, CRC Press.
7. Jeff A. Bilmes (1997), " A Gentle Tutorial of the EM Algorithm and its application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models", Technical Report, University of Berkeley, ICSI-TR-97-021.
8. Kelly P.A. et al (1988), "Adaptive segmentation of speckled images using a hierarchical random field model". IEEE Transactions Acoust. Speech. Signal Processing, Vol.36, No.10, Pg.1628-1641.
9. Lie. T and Sewehand.W(1992), " Statistical approach to X-ray CT imaging and its Applications in image analysis", IEEE Trans. Med. Imag. Vol.11, No.1, Pg 53-61.
10. Lie T. et al.(1993), " Performance evaluation of Finite normal mixture model based image segmentation, IEEE Transactions on Image processing, Vol .12(10), pp 1153-1169.
11. Mantas Paulinas and Andrius Usinskas (2007), "A survey of genentic algorithms applications for image enhancement and segmentation", Information Technology and control, Vol.36, No.3, pp. 278-284.
12. Mclanchlan G. And Krishnan T(1997)., " The EM Algorithm and Extensions", John Wiley and Sons, New York -1997.
13. Mclanchlan G. and Peel.D(2000), " The EM Algorithm For Parameter Estimations", John Wiley and Sons, New York -2000.
14. Meila M. (2005)"Comparing Clustering – An axiomatic view," in proceedings of the 22nd International Conference on Machine Learning, pp. 577-584.
15. Nasios N., A.G.Bors,(2006) " Variational learning for Gaussian Mixtures", IEEE Transactions on Systems, Man , and Cybernatics - Part B : Cybernatics, Vol36(4), pp. 849-862.
16. Pal S.K. and Pal N.R(1993), " A Review On Image Segmentation Techniques",Pattern Recognition, Vol.26, N0.9, pp. 1277-1294.
17. Rose H. Turi (2001)," Cluster Based Image Segmentation", phd Thesis, Monash University, Australia.
18. Shital Raut et al (2009), "Image segmentation- A State-Of-Art survey for Prediction", International conference on Adv. Computer control, pp.420-424.
19. Srinivas Rao K. ,C.V.S.R.Vijay Kumar, J.Lakshmi Narayana,(1997)"On a New Symmetrical Distribution ", Journal of Indian Society for Agricultural Statistics, Vol.50(1), pp. 95-102.
20. Srinivas Y. and Srinivas Rao K. (2007), "Unsupervised image segmentat using finite doubly truncated Gaussian mixture model and Hierarchial clustering ", Journal of Current Science,Vol.93,No.4, pp. 507-514.
21. "The Berkeley segmentation dataset" http://www.eecs. berkeley.edu /Research/ Projects /CS/vision/bsds/BSDS300/html/dataset/images.html.
22. Tolias Y.A. and Pamas S.M. (1998), " Image Segmentation by a Fuzzy clustering algorithm using adaptive spatially constrained membership functions", IEEE Transactions on Image Processing, pp. 359-369.
23. Unnikrishnan R. et al (2007), " Toward objective evaluation of image segmentation algorithms," IEEE Trans.Pattern Analysis and Machine Intelligence . Vol.29, No.6, pp. 929-944.
24. Xu Y. et al (1998), "A segmentation algorithm for noisy images, design and evaluation", Pattern Recognition letters, Vol. 19, pp. 1213-1224.
25. Yamazaki T. (1998), "Introduction of EM algorithm into color image segmentation," in Proceedings of ICIRS'98, pp. 368–371.
26. Yiming WU et al (2003), "Unsupervised Color Image Segmentation Based On Gaussian Mixture Models" Proceedings Of 2003 Joint Conference At The 4th International Conference On Information, Communication and Signal Processing, Vol.1, pp. 541-544.
27. Zhang Z.H. et al(2003)., " EM Algorithms for Gaussian Mixtures with Split-and- merge Operation", Pattern Recognition, Vol. 36(9),pp. 1973-1983.

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

FAXY: Fault Aware Routing Algorithm Based On XY Algorithm for Network on Chip

By Amin Mehranzadeh, Mehdi Hoodgar
Islamic Azad University

Abstract - The performance of Network-on-Chip (NoC) largely depends on the underlying routing techniques. In this paper we present and evaluate a Fault aware routing algorithm scheme called FAXY based on XY routing algorithm. The simulation results show the effectiveness of FAXY by comparing it with XY routing schemes under different traffic patterns. Simulation results depict that the proposed routing algorithm is able to route packet even in the case of faulty links or switches in the NoC. Moreover, simulation results demonstrate the advantage of FAXY routing algorithm in terms of average packet latency, packet loss rate compared with XY routing algorithm in the presence of permanent faults. For the proposed algorithm, it can get much less average packet latency (10%) and lead to less than average 15% packet loss rate.

Keywords : *network on chip, routing algorithm, fault aware.*

GJCST-F Classification : *C.2.2*

Strictly as per the compliance and regulations of:

FAXY: Fault Aware Routing Algorithm Based On XY Algorithm for Network on Chip

Amin Mehranzadeh [α], Mehdi Hoodgar [Ω]

Abstract - The performance of Network-on-Chip (NoC) largely depends on the underlying routing techniques. In this paper we present and evaluate a Fault aware routing algorithm scheme called FAXY based on XY routing algorithm. The simulation results show the effectiveness of FAXY by comparing it with XY routing schemes under different traffic patterns. Simulation results depict that the proposed routing algorithm is able to route packet even in the case of faulty links or switches in the NoC. Moreover, simulation results demonstrate the advantage of FAXY routing algorithm in terms of average packet latency, packet loss rate compared with XY routing algorithm in the presence of permanent faults. For the proposed algorithm, it can get much less average packet latency (10%) and lead to less than average 15% packet loss rate.

Keywords : network on chip, routing algorithm, fault aware.

I. Introduction

A layered architecture called Network on Chip(NoC) [1] has been proposed for global communication in complex SoCs to meet the performance requirements. In NoC each core is connected to switch by a network interface.Cores communication with each other by sending packets via a path consisting of a series of switches and inter switch links. Fig. 1 shows an abstract view of a NOC in this architecture. As shown, each tile is composed of a resource(R) and a switch or router(S). The router is connected to the four neighboring tiles and its local resource via channels. Each channel consists of two directional point-to-point links between two routers or arouter and a local resource [2, 3].

The problem of defining communication protocols for these NoCs is not an easy matter since the resources used in traditional networks are not available on-chip. The communication in NoCs takes place via data packets, which are delivered between the communicating components.

The paths the packets are routed through the network are determined by the routing algorithm. Communication and performance of the entire system are significantly affected by the routing algorithm [4].The performance of NoC largely depends on the underlying routing technique, which chooses a path for a packet and decides the routing behavior of the switches. Routing algorithms can be generally classified into two types: deterministic and adaptive. In deterministic routing, the path is completely determined by the source and the destination address. On the other hand, a routing technique is called adaptive if, given a source and a destination address; the path taken by a particular packet depends on dynamic network conditions (e.g. congested links due to traffic variability). One main advantage of using deterministic routing is its simplicity in terms of routers design. Because of the simplified logic, the deterministic routing provides low routing latency. In this paper we present and evaluate a Fault aware routing scheme called FAXY which a packet first traverses along x dimension and then along the y dimension. The proposed routing algorithm is able to route packet even in the case of faulty links or switches in the NoC. Simulation results demonstrate the advantage of FAXY in terms of average packet latency, packet loss rate compared with XY routing algorithm in the presence of permanent faults. FAXY can get much less average packet latency and lead to less than average 15% packet loss rate.

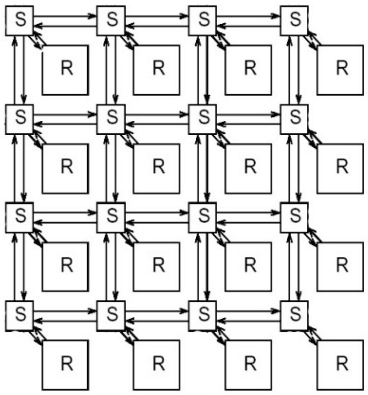


Fig.1: The typical structure of a 4*4 NoC

II. Related Work

The idea of NoC is derived from large-scale computer networks and distributed computing. However, the routing techniques for NoC have some unique design considerations besides low latency and high throughput. Due to tight constraints on memory and computing resources, the routing techniques for

Author α : Islamic Azad University, Dezful Branch, Dezful, Iran.
Email : mehranzadeh@iaud.ac.ir
Author Ω : Islamic Azad University, Dezful Branch, Dezful, Iran.
Email : hoodgar@iaud.ac.ir

NoC should be reasonably simple [5]. Several previous works have investigated different routing algorithms that are tolerant to permanent and/or temporary errors. In order to ensure successful message transmission through the network, several copies of the same packet may be sent on the network links [7].Similarly, probabilistic flooding algorithms can be employed to flood packets to the entire network, which will finally reach to the destination [8].These techniques increase network load and may create congestion when the network is heavily loaded. Different deterministic routing algorithms have been proposed to detour the faulty link(s) in the case of permanent faults [9]. Recently, a fault tolerant routing algorithm was proposed to avoid routing on the faulty links [10]. In this paper, we present a fault-aware dynamic routing algorithm for NoC applications. Our algorithm has the ability to locate the faulty links to avoid them in order to increase the overall network throughput and prevent packet losses.

III. Proposed Routing Algorithm

In proposed routing algorithm, a new fault-aware routing algorithm based on XY routing, namely fault-aware XY (FAXY), is used as a representative of deterministic routing scheme because of its simplicity and wide popularity. Obviously, XY routing is a minimal path routing algorithm and is free of deadlock and live lock [11]. As shown in Fig.3, With FAXY routing, a packet first traverses along X dimension and then along the Y dimension. When a packet traverses along the X dimension and a link is masked because permanent fault, its traverses along Y dimension in order to increase the overall network throughput and prevent packet losses. The mask is used in order to increase the overall network throughput and prevent packet losses.Our algorithm incorporates the acknowledgment signal to reroute the packets around faulty links [12].

```
FA-XY Routing Algorithm (Source, Destination, SwitchAddress) {

Compute destS, destRow, destCol, switchRow, switchCol From Source, Destination and SwitchAddress

If (destS == switchAddr) {
      Send Flit to Current Switch (local Node) and Exit
} else if (destRow == switchRow) {
      If (destCol < switchCol) {
         If (Physical Link connected to Left Switch is not Mask)
               Send Flit to Left Switch
         }
      } else if (destCol > switchCol) {
               If (Physical Link connected to Right Switch is not Mask) {
                  Send Flit to Right Switch
            }
      }

 } else if (destRow < switchRow) {
      If (Physical Link connected to Top Switch is not Mask) {
               Send Flit to Top Switch;
      } else if (destCol> switchCol) {
               If (Physical Link connected to Right Switch is not Mask)
               Send Flit to Right Switch;
            } else if (Physical Link connected to Left Switch is not Mask)
               Send Flit to Left Switch;

 } else if (destRow > switchRow) {
            if (Physical Link connected to Bottom Switch is not Mask){
               Send Flit to Bottom Switch;

            } else if (destCol> switchCol) {
               If (Physical Link connected to Right Switch is not Mask)
            Send Flit to Right Switch;
            } else if (Physical Link connected to Left Switch is not Mask)
               Send Flit to Left Switch;
            }
}
```

Fig.2 : Pseudo code of FA-XY routing algorithm

IV. Experimental Results

In order to evaluate the FAXY routing algorithm, we developed a Java based simulator, namely gpNoCsim [13]. We simulate several square mesh networks with XY and FAXY routing algorithms. The efficiency of each type of routing is evaluated through latency, throughput and average packet loss percent curves. Each simulation is run for a warm-up period of 10000 cycles. Thereafter, performance data are collected after 100,000 packets are sent. The network size during simulation is fixed to be 6*6 tiles. All of the input ports have a FIFO size of 5 flits. Fig.3 shows the network throughput and Fig.4 shows switch throughput, respectively. As shown in Fig. 3 and Fig.4, FAXY routing performs better than XY routing algorithm under different fault percent schemes. The fault percent is the number of faulty switches of network that has one link with permanent fault. The packet injection rate (i.e., the number of packets injected to the network per cycle) is fixed to 150 (packets/cycle).

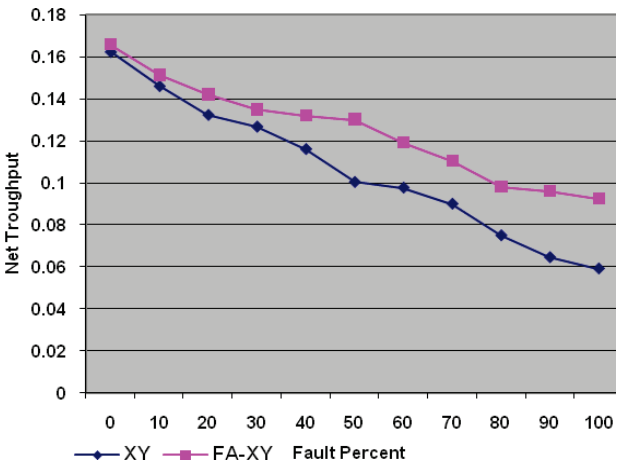


Fig.3 : Network throughput for XY and FAXY routing algorithms

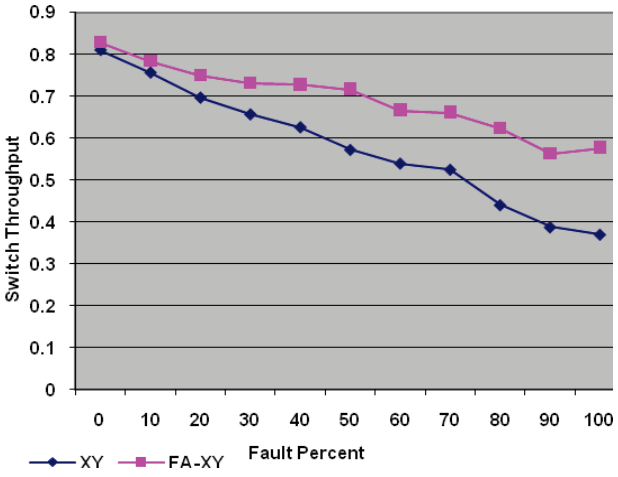


Fig.4 : Switch throughput for XY and FAXY routing algorithms

As in previous work [3, 6], the performance of the routing scheme is evaluated through latency-throughput curves. For a given packet injection rate, a simulation is conducted to evaluate the average packet latency. It is assumed that the packet latency is the duration from the time when the first flit is created at the source core, to the time when the last flit is delivered to the destination core. For each simulation, the packet latencies are averaged over 50,000 packets. Latencies are not collected for the first 5,000 cycles to allow the network to stabilize. It is assumed that the packets have a fixed length of 5 flits and the buffer size of input channels is 5 flits. As shown in Fig.5, FAXY routing performs better than XY routing algorithm. FAXY routing algorithm is able to achieve a lower packet latency rate than XY for the same traffic pattern and the injection rate (10%).

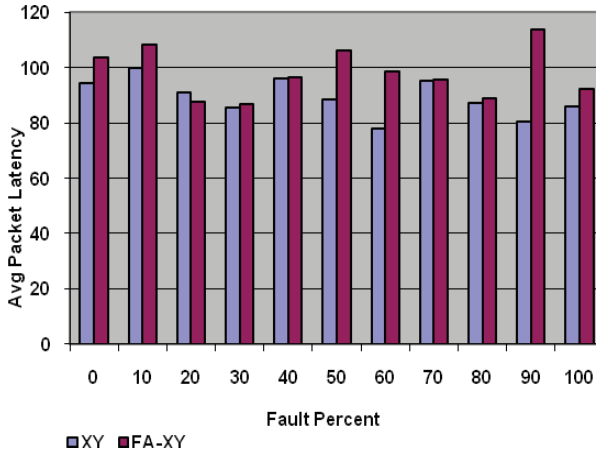


Fig.5 : Average packet latency for XY and FAXY routing algorithms

Fig.6, show the average packet loss rate. As shown in Fig.6, FAXY can achieve average 15% less packet loss rate. Similar to other work in the literature, we assume that the packet loss rate is percent of the packet is created and do not received in the destination node.

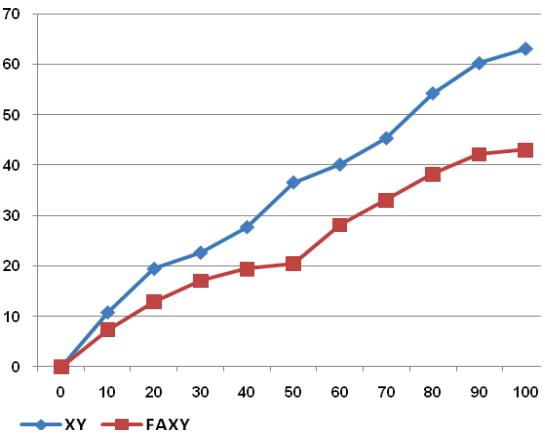


Fig.6 : Average packet loss rate for XY and FAXY routing algorithms

V. Conclusion

In this paper, we presented a new approach for fault aware routing algorithms on NoC, namely FAXY routing. We investigated the effect of permanent errors, and packet injection rates on the performance of our algorithm.Simulation results demonstrated the advantage of our routing algorithm in terms of packet loss percent and latency compared to XY routing algorithms in the presence of permanent faults. Our algorithm can achieve average 10% less latency and average 15% less packet loss rate.

References Références Referencias

1. L.Benini and G.D.Micheli, "Networks on chips: a new SOC paradigm", IEEE computer, 35:70-78, Jan 2002.
2. J.Henkel, W.Wolf, and S.Chakradhar,"On-chip networks: A scalable, communication-centric embedded system design paradigm", VLSI Design, pp.845-851, India, 2004.
3. J.Hu and R.Marculescu, "DyAD - Smart routing for networks-on-chip, " DAC, pp. 260-263, USA, 2004.
4. T.Schonwald, J.Zimmermann, O.Bringmann, "Fully Adaptive Fault-Tolerant Routing Algorithm for Network-on-chip Architectures",10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007).
5. T.T.Ye, L.Benini, and G. De Micheli, "Packetization and routing analysis of on-chip multiprocessor networks," Journal of Systems Architecture, vol. 50, pp. 81-104, 2004.
6. G.M.Chiu, "The odd-even turn model for adaptive routing", IEEE Transactions on Parallel and Distributed Systems, 2000, vol.11, pp. 729 - 38.
7. S.Manolache, P.Eles, and Z.Peng, "Fault and energy-aware communication mapping with guaranteed latency for applications implemented on noc", in Proceedings of IEEE/ACM Design Automation Conference, DAC, 2005, pp. 266-269.
8. M.Pirretti, G. Link, R. Brooks, N.Vijaykrishnan, M.Kandemir, and M. Irwin, "Fault tolerant algorithms for network-on-chip interconnect", in Proceedings of IEEE Computer society Annual Symposium on VLSI, 2004, pp. 46-51.
9. D. Greenfield, A.Banerjee, J.-G.Lee, and S.Moore," Implications of rent's rule for noc design and its fault-tolerance", in Proceedings of International Symposium on Networks-on-Chip, 2007, pp. 283-294.
10. M.Ali, M.Welzl, and S.Hessler, "A fault tolerant mechanism for handling permanent and transient failures in a network on chip", in Proceedings of International Conference on Information Technology, ITNG, 2007, pp. 1027-1032.
11. L.M.Ni and P.K.Mckinley, "A survey of wormhole routing techniques in direct networks", IEEE Int, Conference on Computer, 26:62-76, Feb. 1993.

12. A.Hosseini, T.Ragheb, Y.Massoud,"A Fault-Aware Dynamic Routing Algorithm for on-Chip-Networks", IEEE, 2008.
13. H.Hossain, M.Ahmed, A.Al-Nayeem, "GPNocSim-A General Purpose Simulator for Network-on-chip", Dhaka, Bangladesh, ICICT-2007.

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Implementation of K-Means Clustering Algorithm Using Java

By Prof. S.China Venkateswarlu, Prof. M.Arya Bhanu, Prof.Yudhaveer Katta, V.Badari

Department of CSE , JNTUH, HITS COE - HYDERABAD

Abstract - Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data pertaining to diverse fields. Conventional database querying methods are inadequate to extract useful information from huge data analysis. Cluster analysis is one of the major data analysis methods and k-means clustering algorithm Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data pertainting diverse felids. Conventional Data base methods are inadequate to extract useful information from huge data banks. Cluster analysis is one of the major data analysis methods and the k-means clustering algorithm is widely used for many practical applications. But the original k-means algorithm is computationally expensive and the quality of the resulting clusters heavily depends on the selection of initial cancroids. Several methods have been proposed in the literature for improving the performance of the k-means clustering algorithm. The k-means algorithm is computationally expensive and requires time proportional to the product of the number of data items, number of clusters and the number of iterations.This papert proposes a method for making the algorithm more effective and efficient.

Keywords : About four key words or phrases in alphabetical order, separated by commas.

GJCST-C Classification : I.5.3

Strictly as per the compliance and regulations of:

Implementation of K-Means Clustering Algorithm Using Java

Prof. S.China Venkateswarlu[α], Prof. M.Arya Bhanu[Ω], Prof.Yudhaveer Katta[β], V.Badari[Ψ]

Abstract - Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data pertaining to diverse fields. Conventional database querying methods are inadequate to extract useful information from huge data analysis. Cluster analysis is one of the major data analysis methods and k-means clustering algorithm Emergence of modern techniques for scientific data collection has resulted in large scale accumulation of data per- tainting diverse felids. Conventional Data base methods are inadequate to extract useful information from huge data banks. Cluster analysis is one of the major data analysis methods and the k-means clustering algorithm is widely used for many practical applications. But the original k-means algorithm is computationally expensive and the quality of the resulting clusters heavily depends on the selection of initial cancroids. Several methods have been proposed in the literature for improving the performance of the k-means clustering algorithm. The k-means algorithm is computationally expensive and requires time proportional to the product of the number of data items, number of clusters and the number of iterations.This papert proposes a method for making the algorithm more effective and efficient.

Keywords : About four key words or phrases in alphabetical order, separated by commas.

I. Introduction

a) Module design and organization

i. Fixed Transmission

In this module the fixed transmission is computed by retrieving each link which is limited to maximum signal ratio. Then the fixed routes are computed which give the maximum transmitted power per node and is limited to hardware constraint. Then the distance between the nodes is compute to calculate the end-to-end reliability. The end-to end reliability is a decreasing function, which can be treated as the cost metric for route selection.

ii. End to End Reliability

This module focuses on the problem of optimizing transmission power levels and route selection on an end-to end basis. This module minimizes the end-to end power for a fixed route. The end-to-end route reliability under the optimal power allocation scheme is represented as any fixed route, different power allocation schemes result in different end-to-end reliability and power consumption. The next step is to retrieve the total bandwidth then the distance between nodes are minimized and reliability routes are computed.

iii. Outage Diversity

In the module the case of a point-to-point link, is considered and the trade-off between route outage and consumed power in a network setting. This type of analysis gives insight to how fast the end-to-end outage decreases as more power is spent on the transmission. First, we look at the case that the maximum transmitted power at each link is fixed. It is observed that the route selection does not have any effect on the form of this tradeoff. By selecting the optimal route, we minimized the end to end outage probability by minimizing.

This shows that as long as we limit our approach to a single transmitter and a single receiver per link, even under optimal power allocation and route selection, the trade-off maintains the same form as in the single link case. Phase 1 of the heuristic algorithm requires a time complexity of O(nkp) for finding the initial centroids, as the maximum time required here is for computing the distances between each data point and all other data-points in the set D. In the original k-means algorithm, before the algorithm converges, the centroids are calculated many times and the data points are assigned to their nearest centroids. Since complete redistribution of the data points takes place according to the new centroids, this takes O(nkl), where n is the number of data-points, k is the number of clusters and l is the number of iterations. To obtain the initial clusters, algorithm 4 requires O(nk). Here, some data points remain in its cluster while the others move to other clusters depending on their relative distance from the new centroid and the old centroid. This requires O(1) if a data-point stays in its cluster, and O(k) otherwise. As the algorithm converges, the number of data points moving away from their cluster decreases with each iteration. Assuming that half the data points move from their clusters, this requires O(nk/2). Hence the total cost of this phase2 of the heuristic algorithm is O(nk), not O(nkl). Thus the overall time complexity of the heuristic algorithm becomes O(nkp).

In an Enhanced k-means approach[4] we are not calculating the distance of elements from each

Author[α] : Professor & HOD in Department of CSE ,JNTUH,HITS COE-HYDERABAD , 08CS5028 : +91 9030076793;
E-mail : cvenkateswarlus@gmail.com
Author[Ω] : Professor & HOD in Department of IT, VJIT,
Author[β] : Prof. Yudhaveer Katta, working as a Professor & HOD in Department of IT, VJIT,
Author[Ψ] : Assistant Professor in Dept. of CSE-HITS.

centroid. The algorithm proceeds in the following manner.Once the initial centroids are thus determined, the distance between each data point and centroids of all the clusters are determined and the data points are included in the nearest cluster. Cluster means are then recalculated to find the new centroids. As a result of this, some of the data points may become more closer to a different cluster and such points are redistributed accordingly. The entire process is repeated until no more data points cross cluster boundaries. The enhancedmethod is described as Algorithm 4. This method involves keeping track of the distance between each data point and the centroid of its nearest cluster. During the subsequent iteration, instead of computing the distance of the data point from all cluster centroids, its distance from the previous nearest cluster alone is determined. If that distance is less than or equal to the previous nearest distance, the data point.

b) Algorithm: finding the initial centroids

i. *Finding Intial Centroids*

Input:
D={d1,d2,…..dn} // set of n data items
K // number of desired cluster Output: A set of k initial centroids.
Steps:
1.Set m=1:
2. Compute the distance between each data point and all other data-points in the set D:
3. Find the closest pair of data points from the set D and form a data-point set Am(1<= m<= k) which contains these two data-points,. Delete these two data points from the set D;
4. Find the data point in D that is closest to the data point set Am. Add it to Am and delete it from D:
5. Repeat step 4 until the number of data points in Am reaches 0.75* (n/k);
6. If m< k, then m=m+1, find another pair of data points from D between which the distance is the shortest, from another data –point set Am and delete them from D, Go to step 4:
7. For each data- point set Am (1<=m<=k) find the arithmetic mean of the vectors of data points in Am, these means will be the initial centroids.

Algorithm 3 describes the method for finding initial centroids of the clusters [12]. Initially, compute the distances between each data point and all other data points in the set of data points. Then find out the closest pair of data points and form a set A1 consisting of these two data points, and delete them from the data point set D. Then determine the data point which is closest to the set A1, add it to A1 and delete it from D. Repeat this procedure until the number of elements in the set A1 reaches a threshold. At that point go back to the second step and form another data-point set A2. Repeat this till 'k' such sets of data points are obtained. Finally the initial centroids are obtained by averaging all the vectors in each data-point set. The Euclidean distance is used for determining the closeness of each data point to the cluster centroids. In the first phase, the initial centroids are determined systematically so as to produce clusters with better accuracy [12]. The second phase makes use of a variant of the clustering method discussed in [4]. It starts by forming the initial clusters based on the relative distance of each data-point from the initial centroids. These clusters are subsequently fine-tuned by using a heuristic approach, thereby improving the efficiency. The two phases of the enhanced method are described below as Algorithm 3 and Algorithm 4.

The first step in Phase 2 is to determine the distance between each data-point and the initial centroids of all the clusters. The data-points are then assigned to the clusters having the closest centroids. This results in an initial grouping of the data-points. For each data-point, the cluster to which it is assigned (ClusterId) and its distance from the centroid of the nearest cluster (Nearest_Dist) are noted. Inclusion of data-points in various clusters may lead to a change in the values of the cluster centroids. For each cluster, the centroids are recalculated by taking the mean of the values of its data-points. Up to this step, the procedure is almost similar to the original k-means algorithm except that the initial centroids are computed systematically The next stage is an iterative process which makes use of a heuristic method to improve the efficiency. During the iteration, the data-points may get redistributed to different clusters. The method involves keeping track of the distance between each data-point and the centroid of its presentnearest cluster. At the beginning of the iteration, the distance of each data-point from the new centroid of its present nearest cluster is determined. If this distance is less than or equal to the previous nearest distance, that is an indication that the data point stays in that cluster itself and there is no need to compute its distance from other centroids. This results in the saving of time required to compute the distances to k-1 cluster centroids. On the other hand, if the new centroid of the present nearest cluster is more distant from the data-point than its previous centroid, there is a chance for the data-point getting included in another nearer cluster. In that case, it is required to determine the distance of the data-point from all the cluster centroids. Identify the new nearest cluster and record the new value of the nearest distance. The loop is repeated until no more data-points cross cluster boundaries, which indicates the convergence criterion. The heuristic method described above results in significant reduction in the number of computations and thus improves the efficiency.

c) Algorithm: assigning data-points to clusters [1]

Assininig DataPoints to Centroids
Input:

D={d1,d2,…..dn} // set of n data points di, C= { c1,c2,….ck} // set of k centroids

Steps:

1. Compute the distance of each data point di (1<=i<=) to all the centroids cf (1<=j<=k) as d(di, cj);
2. For each data-point di, find the closest centroid cf and assign di to cluster j.
3. Set CllusterId[i]=j; // j: Id of the closest cluster
4. Set Nearest _ Dist[i] = d(di, cj) ;
5. For each cluster j (1<= j<= k) troids:
6. Repeat
7. For each data-point di.
8. Compute its distance is less than or equal to the present nearest cluster:
9. If this distance is < or = to the present nearest distance , the data- point satays in the cluster:
   Else
10. For every centroid cj (1 <=j<= k) Compute the distance d(di, cj); Endfor;
11. Asign the data-point di to the cluster with the nearest centroid cj
12. Set Cluster ID[i];
13. Set Nearest_Dist[i] = d(di, cf) ;
14. Endfor;
15. For each cluster j (1<= j<=), recalculate the centroids:
16. Until the convergence criterin is met.

Phase 1 of*2* the enhanced algorithm requires a time com- plexity of *O (n)* for finding the initial centroids, as the max-imum time required here is for computing the distances be- tween each data point and all other data-points in the set D. Inthe original k-means algorithm, before the algorithm converges the centroids are calculated many times and thedata points are assigned to their nearest centroids. Since complete redistribution of the data points takes place ac-cording to the new centroids, this takes *O (nkl)*, where *n* is the number of data-points, *k* is the number of clusters and *l* is the number of iterations. To obtain the initial clusters, Algorithm4 requires *O(nk)*. Here, some data points remain in its cluster while the others move to other clusters depending on theirrelative distance from the new centroid and the old centroid. This requires *O(1)* if a data-point stays in its cluster, and *O(k)* otherwise. As the algorithm converges, the number of data points moving away from their cluster decreases with eachiteration. Assuming that half the data points move from their clusters, this requires *O (nk/2)*. Hence the total cost of thisphase of the algorithm is *O(nk)*, not *O(nkl)*. Thus the overall time complex2ity of the enhanced algorithm (Algorithm 2) becomes *O (n)*, since *k* is much less than *n.*

II. Conclusion

In this section we have shown how testing is performed and different test cases are designed to test the system for its performance as well as debugging process. The validation of the test cases is also shown. The k-means algorithm is widely used for clustering large sets of data. But the standard algorithm do not always guarantee good results as the accuracy of the final clusters depend on the selection of initial centroids. Moreover, the computational complexity of the standard algorithm is objectionably high owing to the need to reassign the data points a number of times, during every iteration of the loop. This Project presents an enhanced k-means algorithm which combines a systematic method for finding initial centroids and efficient way for assigning data points to clusters.

III. Implementation and Results

a) Introduction

In the module the case of a point-to-point link, is considered and the trade-off between route outage and consumed power in a network setting. This type of analysis gives insight to how fast the end-to-end outage decreases as more power is spent on the transmission. First, we look at the case that the maximum transmitted power at each link is fixed. It is observed that the route selection does not have any effect on the form of this tradeoff. By selecting the optimal route, we minimized the end to end outage probability by minimizing

IV. Implementation and Results

Algorithm 3 describes the method for finding initial centroids of the clusters [12]. Initially, compute the distances between each data point and all other data points in the set of data points. Then find out the closest pair of data points and form a set A1 consisting of these two data points, and delete them from the data point set D. Then determine the data point which is closest to the set A1, add it to A1 and delete it from D. Repeat this procedure until the number of elements in the set A1 reaches a threshold. At that point go back to the second step and form another data-point set A2. Repeat this till 'k' such sets of data points are obtained. Finally the initial centroids are obtained by averaging all the vectors in each data-point set. The Euclidean distance is used for determining the closeness of each data point to the cluster centroids. In the first phase, the initial centroids are determined systematically so as to produce clusters with better accuracy [12]. The second phase makes use of a variant of the clustering method discussed in [4]. It starts by forming the initial clusters based on the relative distance of each data-point from the initial centroids. These clusters are subsequently fine-tuned by using a heuristic approach, thereby improving the efficiency. The two phases of the

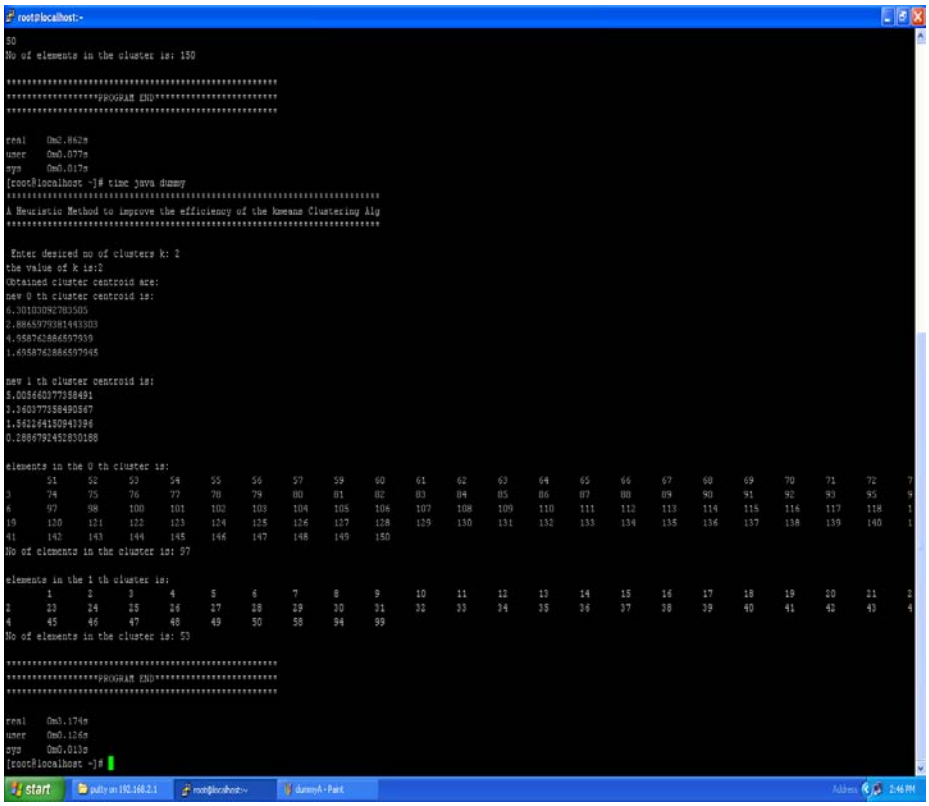
enhanced method are described below as Algorithm 3 and Algorithm 4.

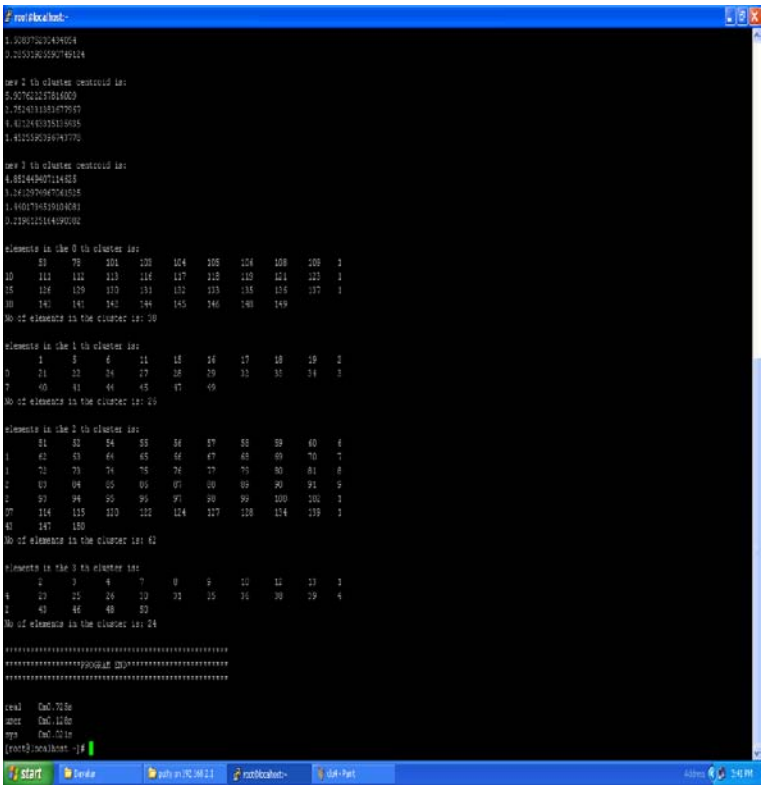
The first step in Phase 2 is to determine the distance between each data-point and the initial centroids of all the clusters. The data-points are then assigned to the clusters having the closest centroids. This results in an initial grouping of the data-points. For each data-point, the cluster to which it is assigned (ClusterId) and its distance from the centroid of the nearest cluster (Nearest_Dist) are noted. Inclusion of data-points in various clusters may lead to a change in the values of the cluster centroids. For each cluster, the centroids are recalculated by taking the mean of the values of its data-points. Up to this step, the procedure is almost similar to the original k-means algorithm except that the initial centroids are computed systematically The next stage is an iterative process which makes use of a heuristic method to improve the efficiency. During the iteration, the data-points may get redistributed to different clusters.

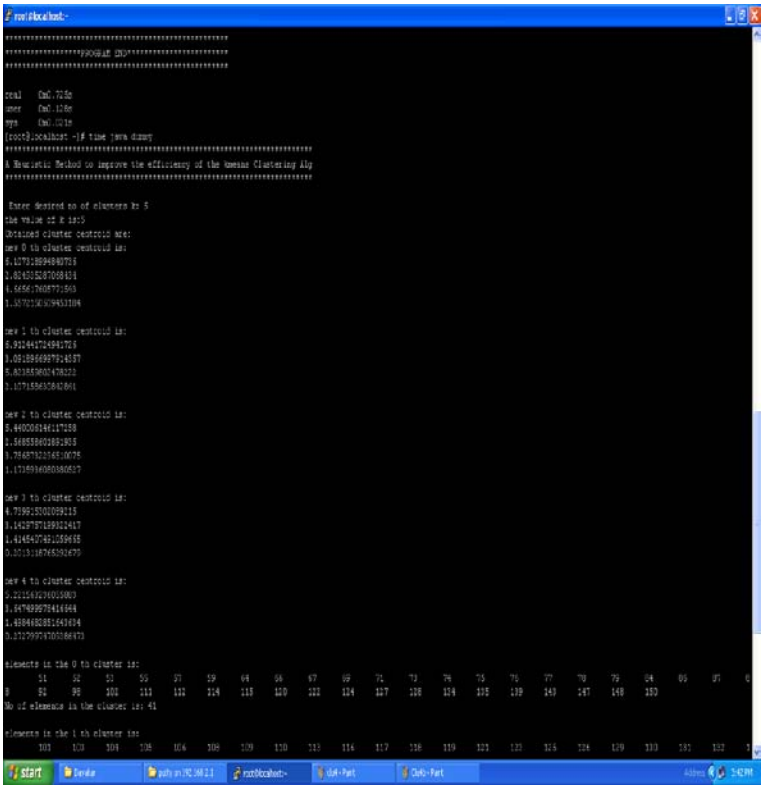
The method involves keeping track of the distance between each data-point and the centroid of its presentnearest cluster. At the beginning of the iteration, the distance of each data-point from the new centroid of its present nearest cluster is determined. If this distance is less than or equal to the previous nearest distance, that is an indication that the data point stays in that cluster itself and there is no need to compute its distance from other centroids. This results in the saving of time required to compute the distances to k-1 cluster centroids. On the other hand, if the new centroid of the present nearest cluster is more distant from the data-point than its previous centroid, there is a chance for the data-point getting included in another nearer cluster. In that case, it is required to determine the distance of the data-point from all the cluster centroids. Identify the new nearest cluster and record the new value of the nearest distance. The loop is repeated until no more data-points cross cluster boundaries, which indicates the convergence criterion. The heuristic method described above results in significant reduction in the number of computations and thus improves the efficiency.

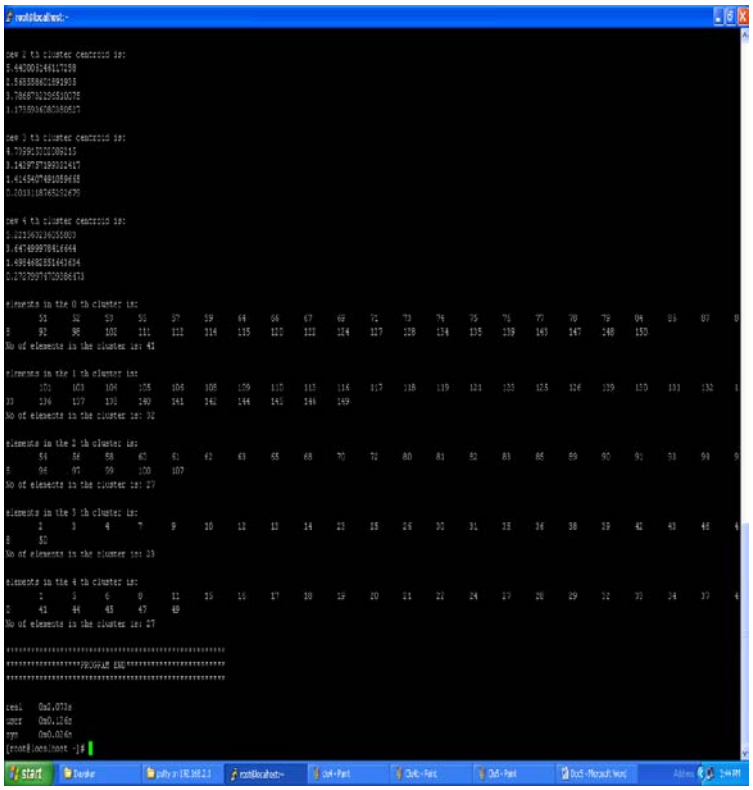
V. OUTPUT

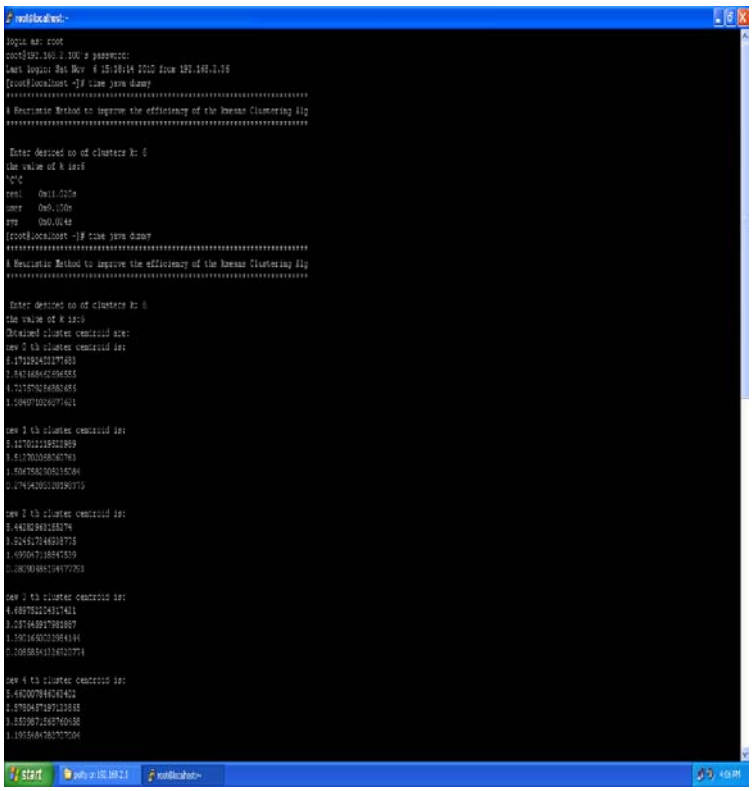
The modified algorithm is applied to multidimensional gene expression data taken from the UCI(university of california irvine) repository[7]. The input dataare the iris data[10], the breast cancer data[11], the e coli data[9], the echo cardiogram data[12], the yeast data[13] and the height-weight data obtained from the web site of disabled-world[8]. The results are compared with that of the original k-means algorithm as well as Enhanced k-means algorithm. Tables 6.1 to 6.6 show the performance comparison of the three algorithms. Figures 6.1 to 6.6 illustrate that the modified algorithm provide better accuracy and efficiency compared to the k-means and enhanced k-means methods.

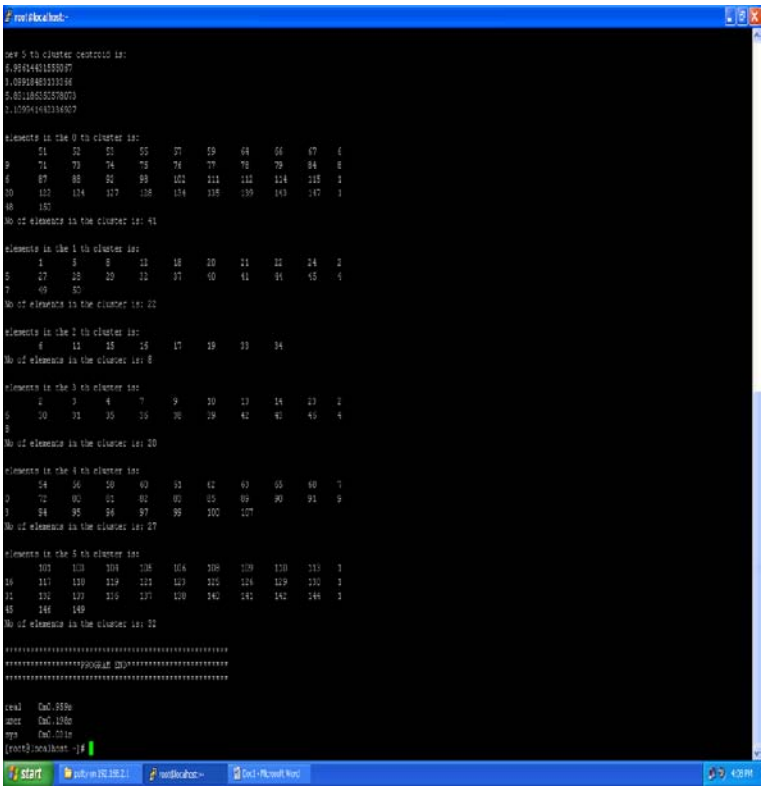












VI. Conclusion

In this section we have shown how testing is performed and different test cases are designed to test the system for its performance as well as debugging process. The validation of the test cases is also shown. The k-means algorithm is widely used for clustering large sets of data. But the standard algorithm do not always guarantee good results as the accuracy of the final clusters depend on the selection of initial centroids. Moreover, the computational complexity of the standard algorithm is objectionably high owing to the need to reassign the data points a number of times, during every iteration of the loop. This Project presents an enhanced k-means algorithm which combines a systematic method for finding initial centroids and efficient way for assigning data points to clusters.

Global Journal of Computer Science and Technology
Volume 11 Issue 17 Version 1.0 October 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Entropy of Data Compression Using Object Oriented Data Warehousing

By Dr. Mrs Pushpa Suri, Mrs Meenakshi Sharma
Kurukshetra University

Abstract - One of the hottest topics in the industry today is data warehousing and on-line analytical processing (OLAP). Although, data warehousing has been around in some form or another since the inception of data storage, people were never able to exploit the information that was wastefully sitting on a tape somewhere in a back room. Today, however, technology has advanced to a point to make access to this information an interactive reality. Organizations across the country and around the world are seeking expertise in this exploding field of data organization and manipulation. It is not a surprise, really, that business users want to get a better look at their data. Today, business opportunities measure in days, instead of months or years, and the more information empowering an entrepreneur or other business person, the better the chances of beating a competitor to the punch with a new product or service. In this paper we have discussed the efficiency of the data warehousing packages so that less CPU time and less Memory is consumed.

Keywords : *Entropy, Data warehousing, object oriented, compression.*

GJCST-F Classification : *E.4*

Strictly as per the compliance and regulations of:

Entropy of Data Compression Using Object Oriented Data Warehousing

Dr. Mrs Pushpa Suri [α], Mrs Meenakshi Sharma [Ω]

Abstract - One of the hottest topics in the industry today is data warehousing and on-line analytical processing (OLAP). Although, data warehousing has been around in some form or another since the inception of data storage, people were never able to exploit the information that was wastefully sitting on a tape somewhere in a back room. Today, however, technology has advanced to a point to make access to this information an interactive reality. Organizations across the country and around the world are seeking expertise in this exploding field of data organization and manipulation. It is not a surprise, really, that business users want to get a better look at their data. Today, business opportunities measure in days, instead of months or years, and the more information empowering an entrepreneur or other business person, the better the chances of beating a competitor to the punch with a new product or service. In this paper we have discussed the efficiency of the data warehousing packages so that less CPU time and less Memory is consumed.

Keywords : Entropy, Data warehousing, object oriented, compression.

I. Introduction

A data warehouse is a mechanism for data storage and data retrieval. Data can be stored and retrieved with a multidimensional structure - hypercube or relational, a star schema structure or several other data storage techniques. The task of transitioning from a procedural mindset to an object-oriented paradigm can seem overwhelming; however, the transition does not require developers to step into another dimension or go to Mars in order to grasp a new way of doing things. In many ways, the object-oriented approach to development more closely mirrors the world we've been living in all along: We each know quite a bit about objects already. It is that knowledge we must discover and leverage in transitioning to object-oriented tools and methodologies. Our research has been from a different point of view – our primary motivating factor is to show how existing applications can be enhanced using object –oriented Technology. Like Many new ideas, object –oriented programming does not have a universally accepted definition [1, 2]. Ideas on the subject do, however, seem to be converging the "best" definition that we have seen to date is "object-oriented = object + classes + inheritance" [3]. OOP can also be defined as an extension of the idea of abstract data type. The task of transitioning from a procedural mindset to an object – oriented paradigm can seem overwhelming: however, the transition does not require developers to step into another dimension or go to grasp a new way of doing thing. In many ways, the object oriented approach to development more closely mirrors the world we've been living in all along. [8]. we each know quite a bit about objects already. It is that knowledge we must discover and leverage in transitioning to object- oriented tools and methodology.

Author [α] : Associate professor, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India.
Author [Ω] : Assistant Professor, Department of Computer Science and Engineering, Haryana College of Technology and Management Kaithal, India. E-mail : Minny_kaushik@yahoo.com

II. Entropy in Data Compression

Data compression is of interest in business data warehousing, both because of the cost savings it offers and because of the large volume of data manipulated in many business applications. The types of local redundancy present in business data files include runs of zeros in numeric fields, sequences of blanks in alphanumeric fields, and fields which are present in some records and null in others. Run length encoding can be used to compress sequences of zeros or blanks. Null suppression may be accomplished through the use of presence bits. Another class of methods exploits cases in which only a limited set of attribute values exist. Dictionary substitution entails replacing alphanumeric representations of information such as bank account type, insurance policy type, sex, month, etc. by the few bits necessary to represent the limited number of possible attribute values. The problem of compressing digital data can be decoupled into two subproblems: modeling and entropy coding. Whatever the given data may represent in the real world, in digital form it exists as a sequence of symbols, such as bits. The modeling problem is to choose a suitable symbolic representation for the data and to predict for each symbol of the representation the probability that it takes each of the allowable values for that symbol. The entropy-coding problem is to code each symbol as compactly as possible, given this knowledge of probabilities. (In the realm of lossy compression, there is a third subproblem: evaluating the relative importance of various kinds of errors.)

For example, suppose if it is required to transmit messages composed of the four letters a, b, c, and d. A straightforward scheme for coding these messages in bits would be to represent a by \00", b by \01", c by \10" and d by \11". However, suppose if it is

known that for any letter of the message (independent of all other letters), a occurs with probability .5, b occurs with probability .25, and c or d occur with probability .125 each. Then a shorter representation might be chosen for a, at the necessary cost of accepting longer representations for the other letters. a could be represented by \0", b by \10", c by \110", and d by \111". This representation is more compact on average than the first one; indeed, it is the most compact representation possible (though not uniquely so). In this simple example, the modeling part of the problem is determining the probabilities for each symbol; the entropy-coding part of the problem is determining the representations in bits from those probabilities; the probabilities associated with the symbols play a fundamental role in entropy coding. One well-known method of entropy coding is Huffman coding, which yields an optimal coding provided all symbol probabilities are integer powers of .5. Another method, yielding optimal compression performance for any set of probabilities, is arithmetic coding. In spite of the superior compression given by arithmetic coding, so far it has not been a dominant presence in real data-compression applications. This is most likely due to concerns over speed and complexity, as well as patent issues; a rapid, simple algorithm for arithmetic coding is therefore potentially very useful. An algorithm which allows rapid encoding and decoding in a fashion akin to arithmetic coding is known as the Q-coder. The QM-coder is a subsequent variant. However, these algorithms being protected by patents, new algorithms with competitive performance continue to be of interest. The ELS algorithm is one such algorithm.

The ELS-coder works only with an alphabet of two symbols (0 and 1). One can certainly encode symbols from larger alphabets; but they must be converted to a two-symbol format first. The necessity for this conversion is a disadvantage, but the restriction to a two-symbol alphabet facilitates rapid coding and rapid probability estimation.

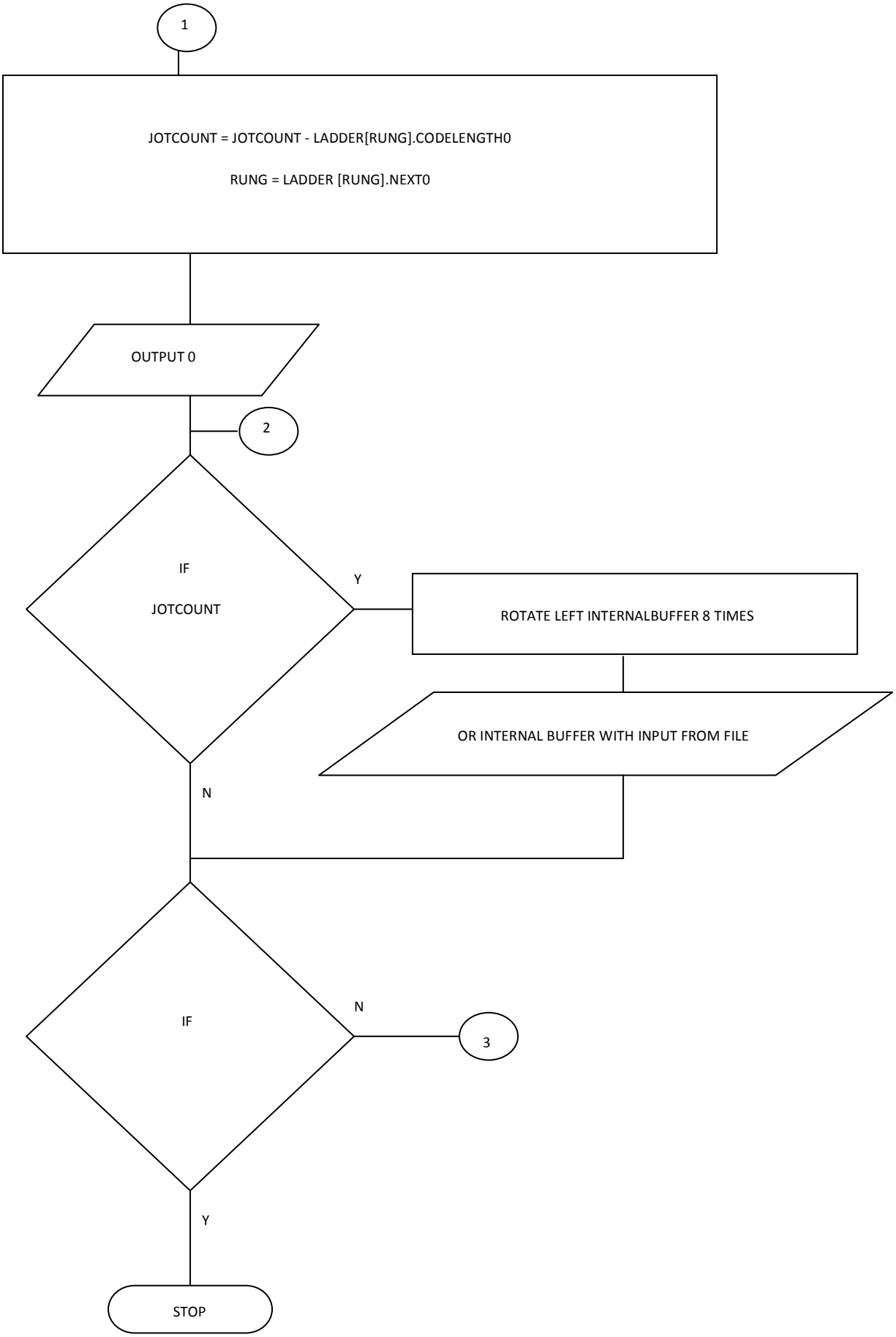
The ELS-coder decoding algorithm has already been described. The encoder must use its knowledge of the decoder's inner workings to create a data stream which will manipulate the decoder into producing the desired sequence of decoded symbols. As a practical matter, the encoder need not actually consider the entire coded data stream at one time. One can partition the coded data stream at any time into three portions; from end to beginning of the data stream they are: preactive bytes, which as yet exert no in seuence over the current state of the decoder; active bytes, which affect the current state of the decoder and have more than one consistent value; and postactive bytes, which affect the current state of the decoder and have converged to a single consistent value. Each byte of the coded data stream goes from preactive to active to postactive; the earlier a byte's position in the stream, the earlier these transitions occur. A byte is not actually moved to the external _le until it becomes postactive. Only the active portion of the data stream need be considered at any time. Since the internal buffer of the decoder contains two bytes, there are always at least two active bytes. The variable backlog counts the number of active bytes in excess of two. In theory backlog can take arbitrarily high values, but higher values become exponentially less likely.

III. Methodology

Following steps will be taken for the future work

1. Creation of different sizes of databases in oracle
2. Employment of object oriented programming for compression using datawarehousing
3. Further compression of database csv files using C++
4. Comparison of time taken and compression efficiency for different sizes of databases.

IV. Flow Chart of Data Compression



V. Conclusion

A data warehouse is an essential component to the decision support system. The traditional data warehouse provides only numeric and character data analysis. But as information technologies progress, complex data such as semi-structured and unstructured data become vastly used [2],[3]. Data Compression is of interest in business data warehousing, both because of the cost saving it offers and because of the large volume of data manipulated in many business application. The entropy is used in many areas such as image processing, document images. But in our research we used the entropy in object oriented data warehousing. Creation of different sizes of databases in oracle. Employment of object oriented programming for compression using datawarehousing. Further compression of database .csv files using C++. Comparison of time taken and compression efficiency for different sizes of databases.

References Références Referencias

1. Wei-Chou Chen; Tzung-Pei Hong; Wen-Yang Lin; , "Using the compressed data model in object-oriented data warehousing ," Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on , vol.5, no., pp.768-772 vol.5, 1999.
2. Wei-Chou Chen; Tzung-Pei Hong; Wei-Yang Lin; , "A composite data model in object-oriented data warehousing," Technology of Object-Oriented Languages and Systems, 1999. TOOLS 31. Proceedings , vol., no., pp.400-405, 1999
3. Shieh, J. C.; Lin, H. W.,"The Novel Model of Object-Oriented Data Warehouses", Workshop on Databases and Software Engineering, 2006
4. Chen Wei-Chou, Hong Tzung-Pei (2), Lin Wen-Yang "Three maintenance algorithms for compressed object-oriented data warehousing."
5. Boqiang Huang; Yuanyuan Wang; Jianhua Chen; , "2-D Compression of ECG Signals Using ROI Mask and Conditional Entropy Coding," Biomedical Engineering, IEEE Transactions on , vol.56, no.4, pp.1261-1263, April 2009.
6. Gong, Y.; Fan, M.K.H.; Huang, C.-M.; , "On entropy-constrained residual vector quantization design," Data Compression Conference, 1999. Proceedings. DCC '99 , vol., no., pp.526, 29-31 Mar 1999.
7. De, I.; Sil, J.; , "Wavelet entropy based no-reference quality prediction of distorted/decompressed images," Computer Engineering and Technology (ICCET), 2010 2nd International Conference on , vol.3, no., pp.V3-245-V3-250, 16-18 April 2010.
8. De, I.; Sil, J.; , "ANFIS tuned no-reference quality prediction of distorted/decompressed images featuring wavelet entropy," Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on , vol., no., pp.126-131, 8-10 Oct. 2010.
9. Liu, L.; Dong, Y.; Song, X.; Fan, G.; , "An entropy based segmentation algorithm for computer-generated document images," Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on , vol.1, no., pp. I- 541-4 vol.1, 14-17 Sept. 2003.
10. Tu, C.; Tran, T.D.; , "Context-based entropy coding of block transform coefficients for image compression," Image Processing, IEEE Transactions on , vol.11, no.11, pp. 1271- 1283, Nov 2002.
11. Chen, S.; Reif, J.H.; , "Using difficulty of prediction to decrease computation: fast sort, priority queue and convex hull on entropy bounded inputs," Foundations of Computer Science, 1993. Proceedings., 34th Annual Symposium on , vol., no., pp.104-112, 3-5 Nov 1993.
12. Sang Hyun Kim; Rae-Hong Park; , "A novel approach to scene change detection using a cross entropy ," Image Processing, 2000. Proceedings. 2000 International Conference on , vol.3, no., pp.937-940 vol.3, 2000.
13. Scales, A.; Roark, W.; Kossentini, F.; Smith, M.J.T.; , "Lossless Compression Using Conditional Entropy-Constrained Subband Quantization," Data Compression Conference, 1995. DCC '95. Proceedings , vol., no., pp.498, 28-30 Mar 1995.
14. Jegou, H.; Guillemot, C.; , "Entropy coding with variable length re-writing systems," Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on , vol., no., pp.1529-1533, 4-9 Sept. 2005.
15. Hua Xie; Ortega, A.; , "Entropy- and complexity-constrained classified quantizer design for distributed image classification," Multimedia Signal Processing, 2002 IEEE Workshop on , vol., no., pp. 77- 80, 9-11 Dec. 2002.

Global Journals Inc. (US) Guidelines Handbook 2011

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

- 'FARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSC" can be added to name in the following manner. eg. **Dr. John E. Hall, Ph.D., FARSC or William Walldroff Ph. D., M.S., FARSC**
- Being FARSC is a respectful honor. It authenticates your research activities. After becoming FARSC, you can use 'FARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSC will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSC. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSC member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSC will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSC member's bank account after certain threshold balance. There is no time limit for collection. FARSC member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

- 'MARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSC" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSC or William Walldroff Ph. D., M.S., MARSC
- Being MARSC is a respectful honor. It authenticates your research activities. After becoming MARSC, you can use 'MARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSC will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- MARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSC member's bank account after certain threshold balance. There is no time limit for collection. MARSC member can decide its price and we can help in decision.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJMBR for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

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

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

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

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

(II) Choose corresponding Journal.

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

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

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

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

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

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

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

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

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

Scope

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

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

2. ETHICAL GUIDELINES

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

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

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

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

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

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

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

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

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

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

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

Please mention proper reference and appropriate acknowledgements wherever expected.

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

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

3. SUBMISSION OF MANUSCRIPTS

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

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

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

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

4. MANUSCRIPT'S CATEGORY

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

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

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

Research articles: These are handled with small investigation and applications

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

5.STRUCTURE AND FORMAT OF MANUSCRIPT

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

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

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

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

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

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

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

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

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

(h) Brief Acknowledgements.

(i) References in the proper form.

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

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

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

Format

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

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

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

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

Structure

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

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

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

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

Key Words

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

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

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

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

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

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

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

Acknowledgements: Please make these as concise as possible.

References

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

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

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

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

Tables, Figures and Figure Legends

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

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

Preparation of Electronic Figures for Publication

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

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

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

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

6. AFTER ACCEPTANCE

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

6.1 Proof Corrections

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

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

(Free of charge) from the following website:

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

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

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

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

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

6.3 Author Services

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

6.4 Author Material Archive Policy

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

6.5 Offprint and Extra Copies

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

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