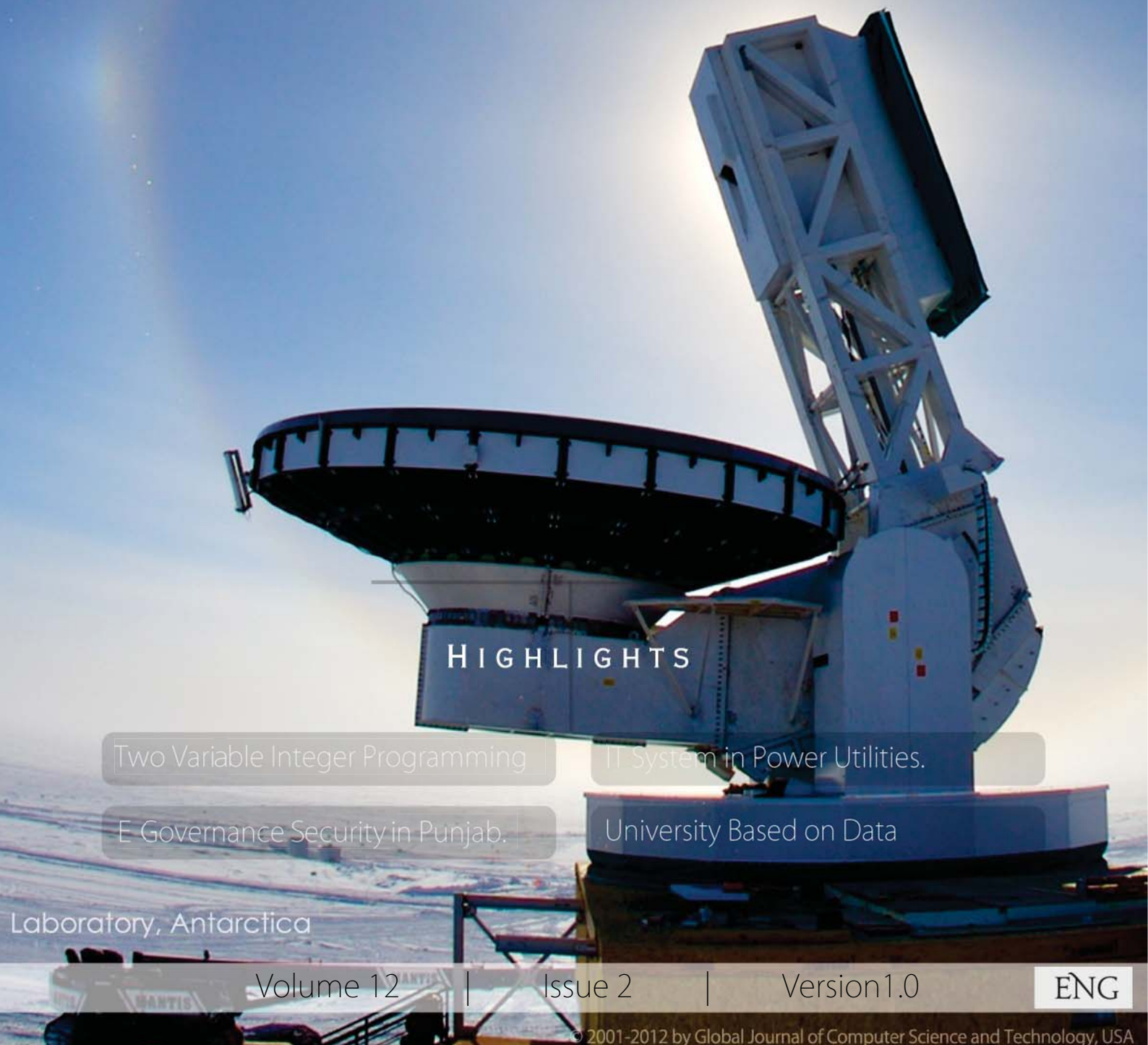
Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal of Computer Science & Technology

DISCOVERING THOUGHTS AND INVENTING FUTURE



Highlights

Two Variable Integer Programming

IT System in Power Utilities.

E Governance Security in Punjab.

University Based on Data

Laboratory, Antarctica

Volume 12 | Issue 2 | Version1.0

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 12 Issue 2 (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:Global Association of Research

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

Global Association of Research, 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 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Cultural Algorithm for the Two Variable Integer Programming Problem

By Senthil Kumar Ramadoss, I.K. Gulam Mohiddin, Ajit Pal Singh
University Adama

Abstract - A specific implementation of cultural algorithm is presented here for solving the following two variable integer programming problem with *n* constraints : Maximize or Minimize $Z = p_1 x + q x_2$ subject to constraints $\{a_1 x_1 + b_1 x_1 \le c_i;\ or\ a_1 x_1 + b_1 x_1 \ge c_i\}$, where $i = 1,2,3,...,n;\ x_1, x_2 \ge 0;\ x_1$ and x_2 are integers; a_i, b_i and c_i are positive real numbers; p and q are signed integers. A cultural algorithm consists of a population component almost identical to that of the genetic algorithm and, in addition, a knowledge component called the belief space. As the integer programming problem is a constrained optimization problem, the constraints including non-negativity and integer restrictions are availed as the knowledge component and used to build the belief space.

Keywords : *Cultural algorithm, Integer programming problem (IPP), Belief space, Genetic algorithm, Optimization.*

GJCST Classification: G.1.6

Strictly as per the compliance and regulations of:

A Cultural Algorithm for the Two Variable Integer Programming Problem

Senthil Kumar Ramadoss[α], I.K. Gulam Mohiddin[Ω], Ajit Pal Singh[β]

Abstract - A specific implementation of cultural algorithm is presented here for solving the following two variable integer programming problem with n constraints: Maximize or Minimize $Z = p_1 x + q x_2$ subject to constraints $\{a_1 x_1 + b_1 x_1 \le c_i;\ or\ a_1 x_1 + b_1 x_1 \ge c_i\}$, where $i = 1,2,3,...,n;\ x_1, x_2 \ge 0;\ x_1$ and x_2 are integers; a_i, b_i and c_i are positive real numbers; p and q are signed integers. A cultural algorithm consists of a population component almost identical to that of the genetic algorithm and, in addition, a knowledge component called the belief space. As the integer programming problem is a constrained optimization problem, the constraints including non-negativity and integer restrictions are availed as the knowledge component and used to build the belief space.

Keywords : Cultural algorithm, Integer programming problem (IPP), Belief space, Genetic algorithm, Optimization.

I. Introduction

We consider the following integer programming problem (IPP) with two decision variables and n constraints: Maximize or Minimize $Z = p_1 x + q x_2$ subject to constraints $\{a_1 x_1 + b_1 x_1 \le c_i;\ or\ a_1 x_1 + b_1 x_1 \ge c_i\}$, where $i = 1,2,3,...,n;\ x_1, x_2 \ge 0;\ x_1$ and x_2 are integers; a_i, b_i and c_i are positive real numbers; p and q are signed integers.

In 1958, Gomory (1963a) devised a method, known as the "method for integer forms", for solving integer programming problems. In 1960, Gomory (1963b) devised another method for solving all integer linear programming problems. Several computers codes using one or both methods have been written; and have successfully solved many real problems. However, their performance has not been nearly as predictable as that of ordinary linear programming codes- which are themselves rather unpredictable as regards running time (Beale, 1965). Martin et al. (1963) used a variant method of integer form called the "accelerated euclidean algorithms". Dakin (1964) and Driebeck (1964) have developed programs using a "branch and bound" method for mixed integer programming. Later Forest et al. (1974), Tomlin (1971), Driebeck (1966), Dakin (1965), Beale and Small (1965), and others improved or refined the branch and bound approach of solving integer programming problems in a number of ways. There are many survey articles and text books to describe the usage of refined and improved methods of using branch and bound approaches (Hansen, 1979; Gupta and Ravindran, 1983a, 1985b)

But the most important facet of both Gomory's cutting plan approach and branch and bound approach is that they can be applied only after obtaining non-integer solution using traditional optimization techniques like Simplex Algorithms.

But the proposed cultural algorithmic approach directly searches the integer solution in the population space which is a space obtained by narrowing the population space where the population space is obtained by the constraints of the problem.

The implementation of the cultural algorithm is presented here to find (x_1, x_2) which satisfies all the n constraints and yields the optimal value for Z. As IPP is a kind of constrained optimization problem, the constraints including non-negativity constraints and integer restrictions are availed as knowledge component to build the belief space. A cultural algorithm, introduced by Reynolds (1994), and seen as extension to genetic algorithm, is a computational model of cultural evolution process in nature where there is a knowledge component in addition to population component. The knowledge component is used to build belief space. The best individuals are selected from belief space using a fitness function. These best individuals are used to update the belief space via a vote acceptance function.

Generally cultural algorithms use five different kinds of knowledge component namely: normative, domain specific, situational, historical and topographical knowledge. The proposed cultural algorithm for solving two variable integer programming problems with n constraints uses a normative kind of knowledge gained from constraints including non-negativity and integer restrictions.

The proposed cultural algorithm for solving two variable integer programming problems uses the frame work for constrained optimization problem introduced by

Author [α Ω] : Department of Computer Science, Eritrea Institute of Technology, Asmara, The State of Eritrea, North East Africa.
E-mail : senanu.kumdev@gmail.com, ikgulam@gmail.com
Author [β] : Department of Mechanical & Vehicle Engineering, Adama Science & Technology University, Adama, Ethiopia, Africa.
E-mail : singh_ajit_pal@hotmail.com

Carlos and Licardo (2002) to identify knowledge component and to build belief space. The algorithm is also using a variant of test bed introduced by Chang and Reynolds (1996) as the backbone of computational procedure.

II. The Cultural Algorithm for Sloving Two Variable Interger Programmin Problems

- *Step 1: Initialization of Population Space*

Lets define sets of integers for x_1 and x_2 that satisfies each of n constraints and let these sets be $S_1, S_2, ..., S_n$ and $T_1, T_2, ..., T_n$ respectively. S_i and T_i can be defined as follows:

i. *Definition of Population Space for* x_1

When the constrain i is as $a_i x_1 + b_i x_1 \le c_i$ with $\le$, then S_i can be defined as $0 \le x_1 \le c_i / a_i$ irrespective of b_i and if c_i / a_i is real, it should be rounded off to the immediate lower positive integer. When the constraint *i* is as $a_i x_1 + b_i x_1 \ge c_i$ with $\ge$, then S_i can be defined as $c_i / a_i \le x_1 \le \infty$ irrespective of b_i and if c_i / a_i is real, it should be rounded off to the immediate higher positive integer. Note that if $a_i = 0$ in the constraint i, then this constraint is ignored in defining space for x_1.

ii. *Definition of Population Space for* x_2

When the constraint i is as $a_i x_1 + b_i x_2 \le c_i$ with $\le$, then T_i can be defined as $0 \le x_1 \le c_i / b_i$ irrespective of a_i and if c_i / b_i is real, it should be rounded off to the immediate lower positive integer. When the constraint i is as $a_i x_1 + b_i x_2 \ge c_i$ with $\ge$, then T_i can be defined as $c_i / b_i \le x_1 \le \infty$ irrespective of a_i and if c_i / b_i is real, it should be rounded off to the immediate higher positive integer. Note that if $b_i = 0$ in the constraint i, then this constraint is ignored in defining space for x_2.

Now our objective is to find (x_1, x_2) for which $px_1 + qx_2$ is the minimum or maximum according to the objective function, searching x_1 from $(S_1, S_2, ..., or\ S_n)$ and x_2 from $(T_1, T_2, ..., or\ T_n)$.

Here, all the population of S_i cannot satisfy all n constraints; similarly any member of T_i cannot satisfy all n constraints. Hence we identify the lowest (l) and greatest (g) member of all common components in $S_1, S_2, ..., S_n$ and we build population space for x_1, $POP(x_1) = \{l, l+1, l+2, ..., g-2, g-1, g\}$ as the range of values between lowest (l) and greatest (g) for x_1 will satisfy all the n constraints. This population space can also be stated as $POP(x_1) = \{S_1 \cap S_2 \cap, ..., \cap S_n\}$. Similarly, we build the population space for x_2 as $POP(x_2) = \{T_1 \cap T_2 \cap, ..., \cap T_n\}$.

- *Step 2: Initialization of Belief Space*

When the objective function is to maximize $px_1 + qx_2$ and if $p \rangle 0$ then the largest component of $POP(x_1)$ can only aid in achieving maximum possible value of $px_1 + qx_2$. When the objective function is to maximize and if $p \langle 0$ then the smallest component of $POP(x_1)$ can only aid in achieving maximum possible value of $px_1 + qx_2$.

Similarly, When the objective function is to maximize $px_1 + qx_2$ and if $q \rangle 0$ then the largest component of $POP(x_2)$ can only aid in achieving maximum possible value of $px_1 + qx_2$. When the objective function is to maximize and if $q \langle 0$ then the smallest component of $POP(x_2)$ can only aid in achieving maximum possible value of $px_1 + qx_2$. Using this knowledge, the *fitness function* is built to identify the best fit component (or the optimal contributor (OC) in achieving optimal $px_1 + qx_2$) to build the belief space. The fitness function which identifies and returns the optimal contributor is defined as follows:

Fitness Function:

When the objective function is to maximize $px_1 + qx_2$:

The OC for x_1 is $\max(POP(x_1))$ if $p \rangle 0$; The OC for x_1 is $\min(POP(x_1))$ if $p \langle 0$;

The OC for x_2 is $\max(POP(x_2))$ if $q \rangle 0$; The OC for x_2 is $\min(POP(x_2))$ if $q \langle 0$;

Similarly, when the objective function is to minimize $px_1 + qx_2$:

The OC for x_1 is $\min(POP(x_1))$ if $p \rangle 0$; The OC for x_1 is $\max(POP(x_1))$ if $p \langle 0$;

The OC for x_2 is $\min(POP(x_2))$ if $q \rangle 0$; The OC for x_2 is $\max(POP(x_2))$ if $q \langle 0$;

When the objective function is to maximize px_1+qx_2, the optimal contribution that x_1 can put in px_1+qx_2 is $\left[OC(x_1))^* p\right]$ and x_2 can put is $\left[OC(x_2))^* q\right]$. Here if $\left[OC(x_1))^* p\right] \geq \left[OC(x_2))^* q\right]$ then x_1 can be greater contributor than x_2, otherwise x_2 is greater contributor than x_1 in achieving optimal px_1+qx_2.

Similarly, when the objective function is to minimize px_1+qx_2, and if $[OC(x_1))^* p] \geq [OC(x_2))^* q]$ then x_2 can be greater contributor than x_1, otherwise x_1 is greater contributor than x_2 in achieving optimal px_1+qx_2.

Thus, firstly, this greater contributor (GC) must be identified. An exceptional instance occurs while identifying this GC is that $POP(x_1)$ and/or $POP(x_2)$ may be a null set or unbounded or infinite. That is there may be no common component in $S_1, S_2, ..., S_n$ and/or in $T_1, T_2, ..., T_n$. This indicates the population space of x_1 and/or x_2 is unbounded or indefinite. We adapt the following rules in such instances.

1. If the population space of both x_1 and x_2 are bounded/definite and non-null, we identify the GC as discussed earlier and make use of this in building belief space.
2. If the population space of any one variable is found to be unbounded/indefinite or null, we consider the other variable as GC in building belief space irrespective of the unbounded/indefinite or null population space of former variable.
3. If the population spaces of both variables are found to be unbounded/indefinite or null, then this is indication that there is inability in building the belief space. With such population spaces, we declare that the problem has infeasible or unbounded solution space.

When the GC is identified evidently, we believe that the $\max(POP)(GC))$ is playing the greatest role than all other components in the population space in achieving the maximum px_1+qx_2 and $\min(POP)(GC))$ is playing the greatest role than all other components in population space in achieving the minimum px_1+qx_2. Note that, here, GC is either x_1 or x_2. Thus we define the belief space $BLF(i)$, (i is initially 0), as follows:

Building Belief Space:

Case I: When x_1 is found to be GC and the objective function is to maximize px_1+qx_2 :

$BLF(i) = (x_1, x_2)$ such that:

$x_1 = \max(POP(x_1))$;

$x_2 \varepsilon B$;

where B is a set of integers which is intersection of $B_1, B_2, ..., B_n$ and B_i can be defined as $x_2 \leq (c_i - a_i x_1)/b_i$ if the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with $\leq$ and as $x_2 \geq (c_i - a_i x_1)/b_i$ if the constraint i as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$ where $x_1 = \max(POP(x_1))$.

If the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with of $\leq$, then $(c_i - a_i x_1)/b_i$ should be rounded off to the immediate lower positive integer and it is as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$, then $(c_i - a_i x_1)/b_i$ should be rounded off to the immediate higher positive integer.

Case II : When x_1 is found to be GC and the objective function is to minimize px_1+qx_2 :

$BLF(i) = (x_1, x_2)$ such that:

$x_1 = \min(POP(x_1))$;

$x_2 \varepsilon B$;

where B is a set of integers which is intersection of $B_1, B_2, ..., B_n$ and B_i can be defined as $x_2 \leq (c_i - a_i x_1)/b_i$ if the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with $\leq$ and as $x_2 \geq (c_i - a_i x_1)/b_i$ if the constraint i as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$ where $x_1 = \min(POP(x_1))$.

If the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with of $\leq$, then $(c_i - a_i x_1)/b_i$ should be rounded off to the immediate lower positive integer and it is as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$, then $(c_i - a_i x_1)/b_i$ should be rounded off to the immediate higher positive integer.

Case III: When x_2 is found to be GC and the objective function is to maximize px_1+qx_2 :

$BLF(i) = (x_1, x_2)$ such that:

$x_1 \varepsilon B$;

$x_2 = \max(POP(x_2))$;

where B is a set of integers which is intersection of $B_1, B_2, ..., B_n$ and B_i can be defined as $x_1 \leq (c_i - b_i x_2)/a_i$ if the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with $\leq$ and as $x_1 \geq (c_i - b_i x_2)/a_i$ if

the constraint i as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$ where $x_2 = \max(POP(x_1))$.

If the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with of $\leq$, then $(c_i - b_i x_2)/a_i$ should be rounded off to the immediate lower positive integer and it is as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$, then $(c_i - b_i x_2)/a_i$ should be rounded off to the immediate higher positive integer.

Case IV: When x_2 is found to be GC and the objective function is to minimize px_1+qx_2 :

$BLF(i) = (x_1, x_2)$ such that:

$x_1 \varepsilon B;$

$x_2 = \min(POP(x_2))$;

where B is a set of integers which is intersection of $B_1, B_2, ..., B_n$ and B_i can be defined as $x_1 \leq (c_i - b_i x_2)/a_i$ if the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with $\leq$ and as $x_1 \geq (c_i - b_i x_2)/a_i$ if the constraint i as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$ where $x_2 = \min(POP(x_1))$.

If the constraint i is as $a_i x_1 + b_i x_2 \leq c_i$ with of $\leq$, then $(c_i - b_i x_2)/a_i$ should be rounded off to the immediate lower positive integer and it is as $a_i x_1 + b_i x_2 \geq c_i$ with $\geq$, then $(c_i - b_i x_2)/a_i$ should be rounded off to the immediate higher positive integer.

Note that, hereafter, we use the terms greater contributor (GC) and lower contributor (LC), instead of x_1 and x_2. If x_1 is found to be GC in the earlier step then x_2 is LC, otherwise x_1 is LC and x_2 is GC.

- *Step 3: Evaluation of Space*

The moment we arrive to evaluate the current belief space, GC is fixed with one single best component and LC lies in the set B. Now lets search the best component for LC from B which may produce the optimal value of px_1+qx_2 along with the already fixed best component of GC. To find the best component for LC, We evaluate the set B as follows:

Case I: If the objective function is to maximize px_1+qx_2 and the co-efficient of LC in objective function is greater than zero or if the objective function is to minimize px_1+qx_2 and the co-efficient of LC in objective function is less than zero, then the largest component of B is considered as the best component for LC.

Case II: If the objective function is to maximize px_1+qx_2 and the co-efficient of LC in objective function is less than zero or if the objective function is to minimize px_1+qx_2 and the co-efficient of LC in objective function is greater than zero , the smallest component of B is considered as the best component for LC.

Find the value of $Z = px_1+qx_2$ using the best components of GC and LC. And lets call this value as $Z = BLF(i))$ and classify this outcome of evaluation as non-futile.

There may be an uncommon situation here while evaluating current belief space in search of the best component of LC that suits with the best component of GC which is already fixed while defining the current belief space. The unusual situation is that the set B identified for LC while defining the current belief space may be a null set. This indicates that there exist no single component in population space of LC to accept the best component chosen for GC to build the current belief space. In such situations, we assume that the mission with current belief space is failed and classify this outcome of evaluation as futile.

- *Step 4: Vote Acceptance Function*

If the $Z = BLF(i)) \langle Z(BLF(i-1))$ when the objective function is to maximize px_1+qx_2 then we stop the process and declare that $Z = BLF(i-1))$ is the optimal solution and x_1 and x_2 are the best components used for GC and LC in calculating $Z = BLF(i-1))$.

Similarly, If the $Z = BLF(i)) \rangle Z(BLF(i-1))$ when the objective function is to minimize px_1+qx_2 then we stop the process and declare that $Z = BLF(i-1))$ is the optimal solution and x_1 and x_2 are the best components used for GC and LC in calculating $Z = BLF(i-1))$.

The vote acceptance function rejects the outcome of evaluation in all other circumstances except the above explained situations. Thus the function rejects the outcome of evaluation and suggest to reproduce the population space in all the following instances:

IF WE ARE WORKING WITH INITIAL BELIEF SPACE (THAT IS $BLF(i)$ WHERE $i = 0$)

IF THE OUTCOME OF EVALUATION OF $BLF(i)$ WAS FOUND TO BE FUTILE.

IF $Z(BLF(i)) \rangle Z(BLF(i-1))$ WHEN THE OBJECTIVE FUNCTION IS TO MAXIMIZE $Z = px_1+qx_2$.

If $Z(BLF(i)) \langle Z(BLF(i-1))$ when the objective function is to minimize $Z = px_1 + qx_2$.

- *Step 5: Modification in the Belief Space*

When the vote acceptance function suggests to reproduce population space, we modify the current belief space $BLF(i)$ to get $BLF(i+1)$ by the modifying the GC. Note that the GC is one fixed component and LC lies in the set B in the current belief space. We now modify this belief space by adjusting the component fixed to the GC. Based on this modified GC, we define once again the set B where LC lies, which along with fixed modified component of GC forms new belief space $BLF(i+1)$. Thus the modification to GC is brought as follows:

Case I: If the objective function is to maximize $px_1 + qx_2$ and the co-efficient of GC in objective function is greater than zero or if the objective function is to minimize $px_1 + qx_2$ and the co-efficient of GC in objective function is greater than zero, GC is fixed with the immediate lower positive integer to the component fixed in the current belief space.

Case II: If the objective function is to maximize $px_1 + qx_2$ and the co-efficient of GC in objective function is less than zero or if the objective function is to minimize $px_1 + qx_2$ and the co-efficient of GC in objective function is less than zero, GC is fixed with the immediate higher integer to component fixed in the current belief space.

Now we define the set B based on the component fixed to GC, which together form the new belief space $BLF(i+1)$ as shown in the section *2.2.2 Building Belief Space.*

Now the new belief space $BLF(i+1)$ is (x_1, x_2) such that GC is the component codified above and LC belongs to the set B defined above. This modified belief space is then submitted to the promote influence function which will deicide the further course on whether this space is to be evaluated or not.

a. *Step 6: Promote Influence Function*

This modified belief space $BLF(i+1)$ shall be promoted for evaluation through *Step 3*, if the set B of new space, where LC belongs to, is not a null set. The set B of $BLF(i+1)$ being null indicates that there exist no component in the population space to be LC, accepting the modified GC. When B of $BLF(i+1)$ is found to be null, we stop the process and declare that $Z(BLF(i+1))$ is the optimal solution and x_1 and x_2 are components used as GC and LC in calculating $Z(BLF(i+1))$.

III. Conclusion

A specific implementation of cultural algorithm is presented using the computational model of cultural evolution process to solve two variable IPPs. Identifying formative knowledge sources in IPPs, the belief space is built in addition to the traditional framework of genetic algorithms. Whilst the existing approaches like Gomery's and Branch and Bound need non-integer solution produced by traditional optimization methods like simplex algorithms, the proposed algorithm steer clear of the computational load of solving linear programming problem relaxations. Comparing the complexity of Gomery's and Branch and Bound approaches, searching global optimal solution using proposed cultural algorithm is especially slender. This implementation can be extended for solving IPPs with any number of decision variables and constraints with signed integers in future.

References Références Referencias

1. Beale, E.M.L. (1965) Survey of integer programming. *Operational Research Quarterly*, 16(2), pp. 219-228.
2. Beale, E.M.L. and Small, R. (1965) Mixed integer programming by a branch and bound technique. In: *Proceedings of the Third IFIP Congress*, pp. 450-451.
3. Carlos, A. C. C. and Licardo, L. B. (2002) A cultural algorithm for constrained optimization. In: *MICAI-Advances in Artificial Intelligence, Second Mexican, International Conference on Artificial Intelligence*, April 22-26. Merida, Yucatan, Mexico, Proceedings, Series-Lecture Notes in Computer Science, Springer-Verlag London, UK, 2313, pp. 98-107.
4. Chang, C. J. and Reynolds, R.G. (1996) A test bed for solving optimization problems using cultural algorithms. In: John R. M. and Peter, A. (eds.) Evolutionary Programming V, A Bradford Book, MIT Press, Cambridge Massachusetts, pp. 225-236.
5. Dakin, R. (1965) A tree search algorithm for mixed integer programming problems. *Computer Journal*, 8, pp. 250-255.
6. Dakin, R.J. (1964) A MIXED-INTEGER PROGRAMMING ALGORITHM. Basser Computing Laboratory, University of Sydney, Technical Report No. 31.
7. Driebeck, N.J. (1964) A method for the solution of mixed integer and non-linear programming problems by linear programming methods. The International Symposium on Mathematical Programming, London School of Economics.
8. Driebeck, N.J. (1966) An algorithm for the solution of mixed integer programming problems. *Management Science*, 12, pp. 576-587.
9. Forest, J.J.H., Hirst, J.P.H. and Tomlin, J.A. (1974) Practical solution of large mixed integer

programming problems with umpire. Management Science, 20(5), pp. 736-773.
10. (a) Gomory, R.E. (1963) All-integer integer programming algorithm. In: J.F. Muth and G.L. Thompson (eds), Industrial Scheduling, Prentice Hall, pp. 193-206. (First issued in IBM Research Center Research Report RC-189, 1960)
11. (b) Gomory, R.E. (1963) An algorithm for integer solutions to linear programs. In: R.L. Graves and P. Wolfe (eds), *Recent Advances in Mathematical Programming*, , McGraw-Hill, pp. 269-302. (First issued in Princeton-IBM Mathematics Research Project Technical Report No. 1, 1958)
12. (a) Gupta, O. K. and Ravindran, A. (1983) Nonlinear integer programming algorithms: a survey. OPSEARCH, 20(4), pp. 189-206.
13. (b) Gupta, O. K. and Ravindran, A. (1985) Branch and bound experiments in convex nonlinear integer programming. Management Science, 31(12), pp. 1533-1546.
14. Hansen, P. (1979) Methods of nonlinear 0-1 programming. Annals of Discrete Mathematics, New York.
15. Martin, G.T., Graves, R.L. and Wolfe, P. (1963) An accelerated euclidean algorithm for integer linear programming. *Recent Advances in Mathematical Programming*, McGraw-Hill, pp. 311-318.
16. Reynolds, R.G. (1994) An introduction to cultural algorithms. In: Proceedings of the Third Annual Conference on Evolutionary Programming, February 24-26, San Diego, California, pp. 131-139.
17. Tomlin, J.A. (1971) An improved branch and bound method for integer programming. *Operations Research*, 19, pp. 1070-1075.

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Architectural principles support the continuous multimedia applications

By Dr.L.V.Reddy, J.Durga
Department of. CSE, SVEC,A.Rangampeta

Abstract - In this paper we discuss some of the architectural principles which are useful to support the continuous media applications in a microkernel environment. In particular, we discuss i) the principle of upcall-driven application structuring whereby communications events are system rather than application initiated, ii) the principle of split-level system structuring whereby, key system .functions are carried out co-operatively between kernel and user level components and iii) the principle of decoupling of control transfer and data transfer. Under these general headings a number of particular mechanisms and techniques are discussed.

GJCST Classification: H.5.1

Strictly as per the compliance and regulations of:

Architectural principles support the continuous multimedia applications

Dr.L.V.Reddy [α], J.Durga [Ω]

Abstract - In this paper we discuss some of the architectural principles which are useful to support the continuous media applications in a microkernel environment. In particular, we discuss i) the principle of upcall-driven application structuring whereby communications events are system rather than application initiated, ii) the principle of split-level system structuring whereby, key system .functions are carried out co-operatively between kernel and user level components and iii) the principle of decoupling of control transfer and data transfer. Under these general headings a number of particular mechanisms and techniques are discussed.

I. Introduction

We are interested in both communications and processing support for distributed real-lime/multimedia applications in end systems, and believe that such applications require thread-to-thread realtime support according to user supplied quality of service (QoS) parameters. Such support, depending on the level of QoS commitment required, may require dedicated, per-connection, resource allocation the CPU scheduler, virtual memory system and communication system. It may also require ongoing dynamic QoS management in all these areas[1]. Another important requirement we have imposed on ourselves is to support standard UNIX applications on the same machine as our real-time/ multimedia support infrastructure; we do not want to build a specialist real-time system that is isolated from the standard application environment. In our paper the prime consideration is the efficiency. In particular in minimizing system imposed overheads by reducing the cost and number of system calls, context switches and copy operation.

In this paper is structured as three main sections, each of which describes a key architectural principle of our design. The three principles are: i) up call driven application structuring whereby communications events are system rather than application initiated, ii) split-level system structuring whereby key system functions are carried out cooperatively between kernel and user level components and iii) decoupling of control transfer and data transfer whereby the transfer of control is carried out asynchronously with respect to the transfer of data.

Author [α] : Professor, Department of CSE, SVEC,A.Rangampeta.
E-mail : lakkireddy.v@gmail.com
Author [Ω] : Asst.professor, Department of. CSE, SVEC,A.Rangampeta.
E-mail : durgapriyadharshini@gmail.com

II. Upcall-Driven Application Structuring

In conventional designs, system APIs are mostly passive and applications are mostly active. For example, when an application needs to send or receive data, it typically invokes a system call such send() or recv(). It also provides the buffer from/to which data is to be sent/received. In contrast, our continuous media API is structured so that the system infrastructure is active and applications are passive. Application programmers attach rthandlers, which are C functions containing application code to process the real-time media, to rtports, which are globally unique units of addressing. Then, programmers establish connections with a given QoS between rtports. At connect time, the system, rather than the application, allocates buffers for connections ,and provides the thread on which the rthandlers will be executed. At data transfer time, the system decides to upcall the application to obtain/ deliver data at instants determined by the QoS specification (in terms of rate, jitter, delay etc.) provided by the application at connect time. When an application rthandler is upcalled, the address of the associated rtport's buffer is passed as an argument so that application code in the rthandler can access the buffer. Source rthandlers are expected to fill buffers with data to be sent, and sink rthandlers to use the data as provided.

In conventional designs, systems API are mostly passive and applications are mostly active. For example, when an applications needs to send or receive data, it typically invokes a system call such as send() or recv(). It also provides the buffer from/ to which data is to be sent/received. In contrast, the continuous media API is structured so that the system infrastructure is active and applications are passive.

There are three major benefits of style of an application/system interaction in our context. First, it relieves the application of the burden of explicitly creating threads and allocating buffers. Second, the system, rather than the application, can choose the timing of application code execution, and thus can optimally monitor and manage the Quality of Service (QOS) of the connection, including the execution of application code, to provide the required thread-to-thread QOS support. Third, the structuring of the API with the API rthandlers is a natural and effective model for real-time programming. Real-time programming is

considerably simplified when programmers can structure applications to react to events and delegate to the system the responsibility for initiating communication events. The programmer is still ultimately in control of event initiation but this control is expressed declaratively through the provision of QOS parameters at connect time and need not be explicitly programmed in a procedural style[2].

Along with these benefits, an efficiency gain potentially results from upcall-driven application structuring, because a single thread ca be used for both protocol and application processing. In conventional systems, application interface with communications by performing system calls which block and reschedule if the communications system is not ready to send or if data has not yet arrived. With infrastructure initiated communication, on the other hand, it is not necessary for the application and communications system to wait for each other, and thus no context switch is incurred, as the communication system always initiates the exchange and the application code is always be ready to run.

III. Split-Level System Structuring

The distributed multimedia applications will require high degree of internal concurrency. For example, it is likely that each media stream will require at least one thread of execution and it is also likely that applications will be structured as pipelines of processing stages on streams of media. Split-level structuring is used to maintain the merits of user space management while mitigating its demerits.

IV. Split Level Scheduling

The above merits and demerits are particularly evident in the case of CPU resource management through user level threads. Here the benefits are cheap user level concurrency and the drawback is that the relative urgencies of threads in different address spaces are not visible to the kernel scheduler.

In split level scheduling, a small number of virtual processors (VPs) execute user threads in each address space. The split level scheduling schemes maintains the invariant that [5]:

- Each user level scheduler (ULS) always runs its most urgent user thread, and
- The kernel level scheduler (KLS) always runs the VP supporting the globally most urgent user thread.

Split level scheduling allows many contexts switches to take place cheaply in the same address space but also ensures that the relative urgencies of threads across the whole machine are appropriately taken into account.

V. Split Level Communication

The strategy of split level communication structuring is a leave the kernel responsible for multiplexing and demultiplexing network packets to application address spaces, but the application address spaces perform transport level processing. In this way, transport protocol processing can automatically take advantages of the split level scheduling infrastructure and thus exploit cheap user level context switches.

Split level communication structuring also allows meaningful deadlines to be placed on (transport level) protocol processing activities, as the ultimate deadline of the final packet delivery is easily available in the application context. Thus, the scheduling of protocol processing need not be performed 'blind' as it is in typical kernel implementation. Another advantage is that multiple transport protocols can easily be dynamically configured in and out of applications according to their particular requirements. This is important in a multimedia context where different protocols may be appropriate for different media types.

VI. Split Level Buffer Management

The strategy of split level buffer management is for the kernel level manager to 'loan' physical, locked, buffers to per-address space managers, but to reserve the rights to reclaim the buffers if memory is urgently required or to retain the buffer longer than it has agreed to. The policy adopted in the design level is that the application is allowed to keep buffer for at least the normal duration of transport protocol processing time plus rthandlers execution time. If the period has elapsed and the application space has not returned ownership of the buffer to the kernel, the kernel may reclaim the buffer.

The semantic of 'reclaiming' locked buffers is to convert locked memory into standard swappable virtual memory. In this way, applications do not lose their data although they do lose guaranteed access latency to that data as the memory region is subject to being paged out. If the kernel does not need to reclaim buffers at the end of an rthandler execution, the user space manager may reuse buffers for other connections.

VII. Decoupling of Control Transfer and Data Transfer

In traditional systems, the transfer of control and the transfer of data are usually tightly coupled. For example, the execution of a UNIX system calls passes data to the kernel and simultaneously transfers control to the kernel.

VIII. Direct Connections

In distributed multimedia applications it is often required to receive continuous media data from the network and directly play it out on a device such as an audio card or a frame buffer (which is probably

managed by kernel level code). The application may or may not require keeping track of the transfer of individual buffers of data for synchronization purposes. The opposite scenario, where data from a local device is to be put directly onto the network, is equally common. In conventional operating system, the only way to achieve such a data flow is to route the data through an intermediate user process. But it involves significant per-buffer overheads [2].

In a direct connection, data that is to be passed directly between the network and a local device does not pass into user space at all; it is processed entirely within kernel space. When a direction connection is established, the infrastructure pre-maps the buffer associated with the connection into the output device's memory. Then, the data can be directly copied off the network card on to the device without leaving kernel space. The fragmentation/re-assembly functions of the in-kernel transport protocol.

If the user application does not need to synchronize with the delivery of buffers, no further overhead is incurred. However, if it is required to synchronize, the application can attach an rthandler to the rtport. This is up called on each buffer transfer with the usual rthandler semantic. The only difference in the API between this case and the normal case is that buffer pointer passed as an argument to the rthandler upcall will be a null pointer as the application context will not have the rights to directly access the kernel managed buffer.

IX. Asynchronous System Calls

For continuous media connections asynchronous system calls exploit the predictable periodicity of the transfer of control and data between application address spaces and the kernel. To issue an asynchronous system call (e.g. an asynchronous version of send()), user level library code:

- Places an operation identifier and parameters in the shared KLS/ULS memory bulletin board area and then
- Sets an 'operation request' bit, also in the bulletin board area.

The KLS, when it runs at the next system clock tick, notices that an operation bit is consequently passes the user's parameters to a kernel server thread which carries out the system call on behalf of the ULS. This avoids a special domain crossing for the system call at the expense a bitmap on each clock interrupt.

X. Asynchronous Software Interrupts

The implementation of software interrupts is similar to that of asynchronous system calls and similarly avoids a special domain crossing. The mechanism for kernel to VP control transfer is as follows:

- The KLS places an event identifier and parameters in the KLS/ULS bulletin board area
- The KLS alters the program counter field of the target VP's context structure points to a standard entry point in the ULS.

Thus, when the VP is next scheduled, the VP immediately enters its ULS, which picks up the event identifier and parameters, and schedules a user thread to deal with event. Asynchronous software interrupts are also provided as a service accessible from user level code. This service enables library code in one address space to cheaply notify an event to another address space on the same machine. The service also allows the sender to name a pre-existing memory segment shared between the sender and receiver address spaces so that data can be optionally transferred in the same call.

XI. User Level Pipelines

The API for pipelines of processing stages is very similar to the connection abstraction. But rather than passing a pair of rtports as arguments to the connect() primitives, we pass a list of rtports. In the case of pipelines, the delay QOS parameter applies end-to-end over the entire chain of rtports.

Intermediate processing stages in pipelines are also realized in a similar way to that described above: When data arrives at an intermediate processing stage, the rthandler returns, it is assumed that the rthandler's application code has performed some appropriate processing on the buffer whose address was passed up to it, and the data can be passed on to the next page.

As the various stages of a pipeline form part of the same application, it is typically the case that pipelines are implemented in a single address space. The data transfer mechanism in this case is as follows: When an rthandler implementing one stage of a pipeline returns, having operated on a buffer, the next stage in the pipeline is simply passed the address of the same buffer. Meanwhile, the first stage sets of work on a second buffer; and so on. At the end of the pipeline, when buffers are finished with, they are returned to a user level pool from which that can be reused by the first pipeline stage. With this implementation, intra address space pipelines incur only user level control transfers between the threads dedicated to each pipeline stage, and zero copy operations between stages. The API is the identical regardless of whether intra-address space, inter-address space or inter-machine connections or pipelines are used. Inter-address space communication on the same machine uses buffers that are statically mapped into both the source and sink address spaces for data transfer, and use asynchronous software interrupts.

XII. Proposed Scenario

The integrated use of the principles and techniques described above, the scenario is illustrated

in figure 1, involves the transfer of compressed video from a frame grabber card on a source machine to a decompress/display application on a sink machine. In figure 1, the large ovals represent user address spaces with library code below the horizontal line and application code above. The rectangles represent kernel space with the enclosed shaded regions representing devices.The send side features a direct connection, involves the video capture device and the network interface, which avoids the need for data to pass into user space. It also features the (optional) use of an rthandler to allow the sender, which is structured as an upcall-driven application, to monitor and synchronize with the progress of the connection.

On the receive side, the split level buffer management system allocates a physical buffer from the kernel buffer pool to hold incoming network packets associated with the connection. This buffer is statically mapped into both kernel space and the application address space.In the split level communication system, when a complete network level packet has been received, the application address space's ULS is notified via the conditional deadline mechanism and initiates transport level processing.

This may involve the receipt of further network packets to build a complete user level buffer.When a complete user buffer has been built, and the receiving thread has the globally earliest deadline, the ULS runs a thread which upcalls the application's rthandler with the address of the buffer as a parameter. The receive side features a user level pipeline which involves one user thread performing decompressed and another displaying uncompressed video in a window. The display is achieved by means of asynchronous system calls to display device. Context switches between the two pipeline threads are achieved at user level costs and the transmission of data from the decompressor to the displayer does not involve data copying.

XIII. Conclusions

We discussed three architectural principles useful for the support of distributed real-time multimedia application in operating system. Firstly we contended that the principles of upcall-driven application structuring leads to well structural real-time applications, relieves applications of the burden of explicit thread creation and buffer allocation, and leads to potential efficiency gains because of reduced context switches.

Secondly we argued for the principles of split-level system structuring. We suggested that it can improve efficiency by exploiting application specific knowledge (e.g. scheduling deadline or buffer requirements) in a local, user level, context where application/manager interact6ion is cheap, while relying on a kernel level manager to 'bias' resources to application address spaces of the basis of their relative needs. Active co-operation of management information between user and kernel level managers is key, but as long as an asynchronous style of communication between managers is acceptable, this can be achieved cheaply by means of a shared memory 'bulletin board'.

Finally we discussed the three principles working together in this paper. They are capable of exploited in a range of operating system environments. Similarly many of individual techniques can be useful in a stand alone fashion. When it can be compared with the distributed level, the principles which we have discussed validated in terms of direct performance measurements.

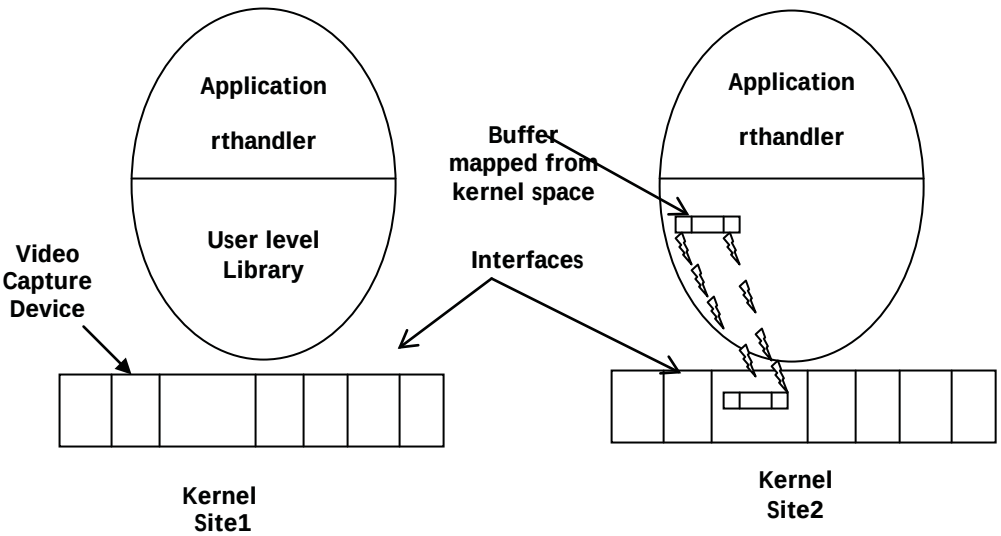


Figure 1 : An illustrative Scenario

References References Referencias

1. A.P. Black, 1983, 'An Asymmetric Stream Communication System', Proceedings of 9th ACM Symposium in Operating System Principles', pp 4-10.
2. M. Acceta, A. Tevanian and M. Young, 1986, 'Mach: A new Kernel Foundatiuon for UNIX Development', Technical Report, Camegie Mellon University.
3. Geoff Coulson and Gordon Blair, 1995, 'Architectural Principles and Techniques for Distributed Multimedia Application Support in Operating Systems', ACM SIGOPS Operating Systems Review, Vol.29 (4), pp17-24.
4. Geoff Coulson and Gordon Blair, 1994, 'Micro-kernel Support for Continuous Media in distributed Systems', Computer Networks and ISDN Systems, Vol 26 (10), pp 1323-1341.
5. Geoff Coulson, Gordon Blair, P. Robin and D. Shepherd, 1994,'Supporting Continuous Media Applications in a Micro-Kernel Environment', Proceedings of the 1st International Workshop on Architecture and Protocols for High-Speed Networks, pp 215-234.
6. Ramesh Govindan, D.P. Anderson, 1991, 'Scheduling and IPC mechanisms for continuous media', ACM SIGOPS Operating Systems Review, Vol. 25 (5), pp 68-80.
7. C.L. Liu and J.W. Layland, 1973, 'Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment', Journal of the ACM, Vol. 20(1), pp 46-61.
8. C.A. Thekkath, T.D. Nguyen, 1993, 'Implementing network protocols at user level', Proceedings on Communications architectures, protocols and applications, pp 64 -73.

Global Journal of Computer Science and Technology Volume XII Issue II Version I January 2012
12

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Enhanced Speckle Filters for Sonar Images Using Stationary Wavelets and Hybrid Inter- And Intra Scale Wavelet Coefficient Dependency

By J. Alavandan, Lt. Dr. S. Santhosh Baboo
Jawahar Science College, Neyveli, Cuddalore Dt., Tamilnadu, India

Abstract - The quality of Sonar images are often reduced by the presence of speckle noise. The presence of speckle noise leads to incorrect analysis and has to be handled carefully. In this paper, an improved non-parametric statistical wavelet denoising method is presented. The algorithm uses a stationary wavelet transformation to derive the wavelet coefficients, from which edge and non-edge wavelet coefficients are identified. Further to improve the time complexity, only homogenous regions with respect to coefficients of neighbors are considered. This method uses an ant colony classification technique. A hybrid method that exploits both inter-scale and intra-scale dependencies between wavelet coefficients is also proposed. The experimental results show that the proposed method is efficient in terms of reduction in speckle noise and speed and can be efficiently used by various sonar imaging systems.

Keywords : *Sonar Image Denoising, Stationary Wavelet Denoising, Inter-scale and Intra Scale dependency, Ant Colony Classification, Non-edge wavelet coefficients.*

GJCST Classification: *I.4.1*

Strictly as per the compliance and regulations of:

Enhanced Speckle Filters for Sonar Images Using Stationary Wavelets and Hybrid Inter- And Intra Scale Wavelet Coefficient Dependency

J. Alavandan [α], Lt. Dr. S. Santhosh Baboo [Ω]

Abstract - The quality of Sonar images are often reduced by the presence of speckle noise. The presence of speckle noise leads to incorrect analysis and has to be handled carefully. In this paper, an improved non-parametric statistical wavelet denoising method is presented. The algorithm uses a stationary wavelet transformation to derive the wavelet coefficients, from which edge and non-edge wavelet coefficients are identified. Further to improve the time complexity, only homogenous regions with respect to coefficients of neighbors are considered. This method uses an ant colony classification technique. A hybrid method that exploits both inter-scale and intra-scale dependencies between wavelet coefficients is also proposed. The experimental results show that the proposed method is efficient in terms of reduction in speckle noise and speed and can be efficiently used by various sonar imaging systems.

Keywords : Sonar Image Denoising, Stationary Wavelet Denoising, Inter-scale and Intra Scale dependency, Ant Colony Classification, Non-edge wavelet coefficients.

I. Introduction

The fact that Earth is an aquatic plant with more than 80% of the surface covered with water, has attracted many earth observers to understand what is lying below water using sonar techniques. SONAR (SOnar NAvigation and Ranging), initially used in submarines during World War II, is increasing being used in Earth observations along with various civilian applications, sea-bed imaging, depth sounding and fish-echolocation. Sonar information collected while searching for, or identifying, underwater surfaces is often presented to the operator in the form of a two dimensional images. Sonar images are created using a fan-shaped sonar beam that scans a given area by moving through the water to generate points, which forms high resolution sonar image of the given area. The Sonar images thus acquired are often disturbed by various factors like the transmission of limited range of light, disturbance of lightening, low contrast and blurring of image, color diminishing during capturing and noise. These disturbances affect image quality which often lead to incorrect analysis and has to be handled carefully.

Sonar image quality can be assessed in terms of quality parameters like contrast, distortion, blur and noise. These parameters can be altered by various factors like, lighting, movement of beam and sensitivity of the imaging devices, all of which can produce images that are difficult to interpret. Sonar Image enhancement techniques are used to enhance these quality parameters and can be achieved through the use of techniques like histogram equalization, image smoothening, image sharpening, contrast adjustment, edge or boundary enhancement and denoising. Out of these, image denoising has become a mandatory task before many processes like segmentation, feature extraction and target classification.

Sonar images are often degraded by a special kind of noise called 'Speckle Noise'. Speckle is a random, deterministic, interference pattern in an image formed with coherent radiation of a medium containing many sub-resolution scatterings. Speckle noise removal can be performed either during data acquisition stage (multi-look process) or after data is stored (spatial filtering). In both cases, the main aim is to reduce/remove speckle noise while preserving both significant image and edge features along with spatial resolution. To achieve this goal, several solutions have been suggested (Pardo, 2011; Guo et al., 2011). Examples include the usage of traditional filters like Frost, SRAD (Speckle Reducing Anisotropic Diffusion), wavelets and Non-local Means techniques. Out of these the usage of wavelets is widespread (Kaur. and Singh, 2010; Delakis et al., 2007) and is considered in this paper.

Traditional wavelet-based algorithms exploiting parametric models initially perform wavelet decomposition on the noisy image. A Bayesian estimator developed using a suitable probability density function (pdf) is then used to estimate noise-free wavelet coefficients. These estimated values are used as a prior for modeling wavelet coefficients of the image. Finally, an inverse wavelet transform is performed to construct the denoised image. The statistical models proposed for this purpose include Gaussian model, Laplacian model, generalized Gaussian model, Student-t model and

Author α : Associate Professor & Head, Department of Computer Science & Applications, Jawahar Science College, Neyveli, Tamilnadu. Ph: +91 9367645666, E-mail : alaneyveli@gmail.com
Author Ω : Reader, Postgraduate and Research department of Computer Science, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, Tamilnadu. Ph: +91 9444063888, E - mail : santhosh2001@sify.com

alpha-stable model. The major concern with these models is that the efficiency depends on the correct estimation of the prior pdf used for modeling the wavelet coefficients.

To solve the problem of estimation, Tian and Chen (2011) proposed a maximum a posteriori (MAP) estimation-based image despeckling algorithm which incorporates a non-parametric statistical model into a Bayesian inference framework. This model, referred as Tian model formulates the marginal distribution of wavelet coefficients. The Tian model uses a two-level decomposition using a Daubechies's wavelet and a novel wavelet shrinkage method called Antshrink (Tian et al., 2010) that exploits the intra-scale dependency of the wavelet coefficients to estimate the signal variance only using the homogeneous local neighbouring coefficients. This is in contrast to conventional shrinkage approaches where all local neighboring coefficients are used. Furthermore, to determine the homogeneous local neighboring coefficients, an Ant Colony Optimization (ACO) technique is used which is also used to classify the wavelet coefficients and this advanced technique is termed as AntShrink algorithm.

In this work, the Tian model is enhanced in three manners. First, the traditional wavelet transform is replaced by a more efficient wavelet transforms, Stationary wavelet (undecimated wavelet transform). Second, the AntShrink algorithm in Tian model uses intra-scale dependency of wavelet coefficients. This method is enhanced by a method that combines both intra-scale and inter-scale dependency of the wavelet coefficients. Finally, the shrinkage is applied only to the magnitude wavelet coefficients at non-edge points. For this, a simple classification algorithm based on the coefficient's statistical features is used. The rest of the paper is organized as below. Section 2 discusses the proposed despeckling algorithm. The efficiency of the proposed method is analyzed using various experiments and the results are compared with traditional despeckling algorithm and the Tian model. The experimental results are presented in Section 4. Section 5 concludes the work with future research direction.

II. Proposed Denoising Models

One common idea is to perform a logarithmic transformation to convert the multiplicative speckle noise into an additive noise (Arsenault and April, 1976; Xie et al., 2002), followed by a wavelet decomposition on the input noisy image to pack the energy of the image into a few large coefficients, then modify the noisy wavelet coefficients using certain shrinkage functions. Finally, the denoised image is reconstructed by performing an inverse wavelet transform, followed by an exponential transformation. The proposed denoising algorithm consists of four steps as listed below.

Step 1: Apply log transformation of the noisy image

Step 2: Apply stationary wavelet to the log transformed image

Step 3: Identify edge and non-edge coefficients

Step 4: Identify homogenous neighbour of non-edge coefficients

Step 5: Estimate each noise-free coefficient using hybrid intra-scale and inter-scale dependencies excluding LL subband coefficients

Step 6: Perform inverse stationary wavelet transform

Step 1: Log transformation of the noisy image

Given an image in spatial domain, the noisy pixel gi is given using Equation (1),

$$g_i = f_i + \varepsilon_i \quad (1)$$

where fi is the noise-free pixel, ei is the speckle noise and i is the pixel index. The multiplicative noise is converted to an additive one by applying the log-transformation on both sides of Equation (1)

$$g_i' = f_i' + \varepsilon_i' \quad (2)$$

where g_i', f_i', ε_i', are log transformed version of g_i, f_i and e_i respectively. Logarithmic transform shows the frequency content of an image. This transformation maps a narrow range of low gray level values in the input image into a wider range of the output level. The opposite is true of higher values of input level. This type of transformation is used to expand the values of dark pixels in an image while compressing the higher level values. The log function has the important characteristic that it compresses the dynamic range of images with large variation of pixel values. However, the histogram of this data is usually compact and uninformative. Log transformation is done in two steps. The first step requires the creation of a matrix to preserve the phase of the transform image. This will be used later to restore the phase of the transform coefficients. In the second step logarithm is taken on the modulus of the coefficients according to the following equation.

$$\hat{X}(i,j) = \ln(|X(i,j)| + \lambda) \quad (3)$$

where λ is a shifting coefficient, usually set to 1. After log transformation, the stationary wavelet transformation is performed to obtain the wavelet coefficients on the noisy image. The wavelet coefficients of the log transformed image corrupted by the speckle noise is expressed as

$$y_i = x_i + n_i \quad (4)$$

where y_i, x_i and n_i represent wavelet coefficients of g_i', f_i' and e_i' respectively.

Step 2: Stationary wavelet transformation

The wide usage of the traditional Discrete Wavelet Transform (DWT) is because of the various advantages it has for denoising images. Some of the merits offered like its multi-scale filtering property and sparse transformation, which while compressing the signal energy to a small number of wavelet coefficients also leaves the majority of the coefficients with values close to zero. However, it has a serious flaw. The DWT does not preserve translation invariance due to the subsampling performed. That is a transformed version of signal X is not the same as the original signal. To preserve the translation variance, this paper uses a Stationary Wavelet Transform (SWT) (Nason and Silverman, 1995).

Introduced by Holdschneider et al. (1989), the SWT is similar to the Discrete Wavelet Transform (DWT) in that the high-pass and low-pass filters are applied to the input signal at each level. However, in the SWT, the output signal is never subsampled (not decimated). Instead, the filters are upsampled at each level. SWT handles translation-invariance by removing the downsamplers and upsamplers in the WT and upsampling the filter coefficients by a factor of 2(j-1) in the jth level of the algorithm (Shensa, 1992). The SWT is an inherently redundant scheme, as the output of each level of SWT contains the same number of samples as the input. Thus, for a decomposition of N levels there will be N redundant wavelet coefficients. This algorithm is more famously known as "algorithme à trous" in French (word trous means holes in English) and refers to inserting zeros in the filters. The block diagram of SWT along with the filters in each level (up-sampled versions of the previous) is shown in Figure 1. (http://en.wikipedia.org/wiki/Stationary_wavelet_transform). An overview of the different names with explanation is provided by Fowler (2005).

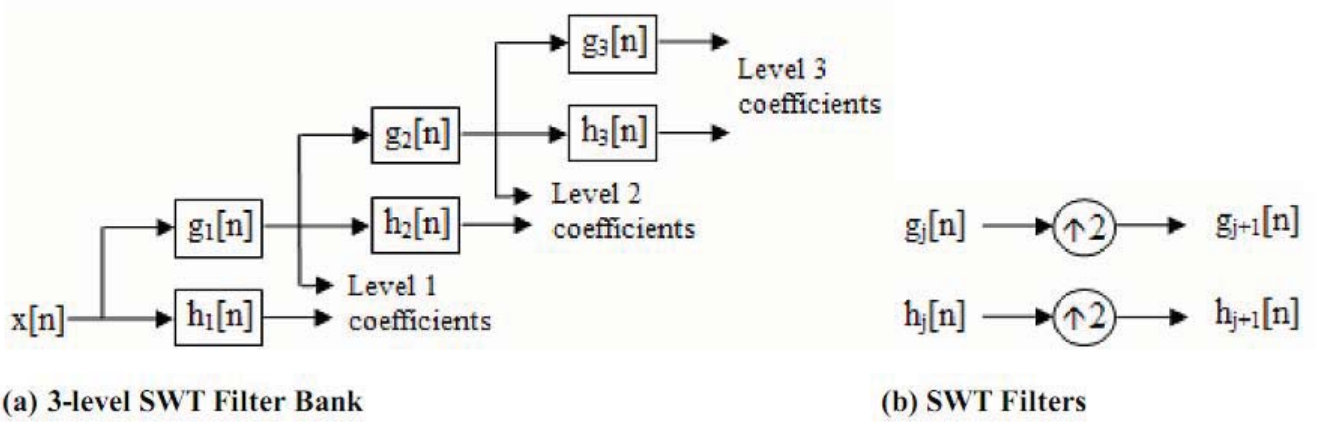


Figure-1 : Block Diagram of SWT

Step 3: Identification of Non-Edge region

The local statistics of the wavelet coefficients are used to classify a coefficient as edge or non-edge. This step is performed in order to maintain edge and line features of the image. It is a well-known fact that the coefficient of variation of edges is higher than that of smooth regions (Schulze, 1997). Using this, the coefficient of variation is calculated for the wavelet coefficients. The process of identifying edge and non-edge regions among wavelets begins by first dividing an image into 3 x 3 square windows. Four sub-images (Figure 2) are constructed for each square window for each subband that has edge details. As HL subband has vertical edge information, the subimage is created in vertical fashion, while since LH subband has horizontal edge information; the subimage is created in horizontal fashion. As the HH subband has edge details in 45^0 directions, two diagonal subwindows are used.

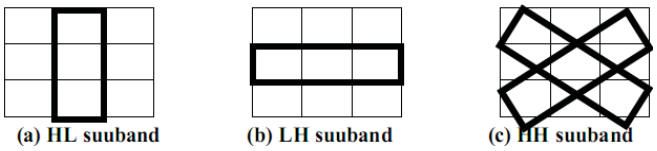


Figure-2 : 3 x 3 Window and its sub-windows Formation

The coefficient of variation is then calculated using Equation (5)

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \quad (5)$$

The edge regions are obtained based on the assumption that a coefficient is considered as an edge if its coefficient of variation measured over the sub-

window is greater than the entire window. That is, Let CS and CSW is the coefficient of variation of a window and its subwindow and if the condition CSW > CW produces a true result, then the wavelet coefficient under consideration is taken as an edge coefficient, else it is taken as a non-edge coefficient.

Step 4: Identify homogenous wavelet coefficients

The main task of this step is to classify the wavelet coefficients to find the homogenous neighbours among the non-edge coefficients using Ant Colony Optimization (Figure 3).

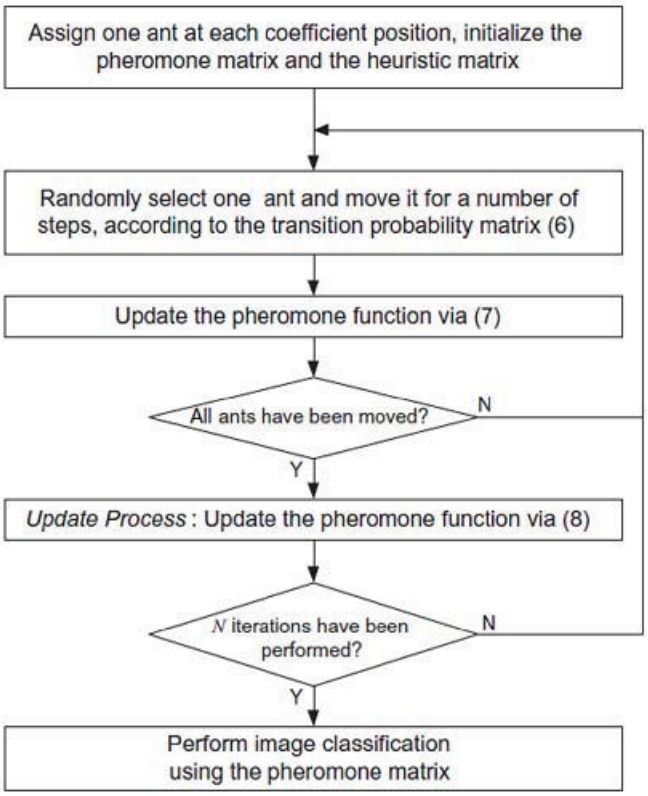


Figure-3 : ACO-based Image Classification

Step 5: Estimation of wavelet coefficient dependencies

Although a wavelet transform decorrelates images in an efficient manner, there still exist strong dependencies between wavelet coefficients. Exploitation of such dependency information with proper statistical models could further improve the performance of coding and denoising algorithms. Statistical modeling techniques that consider the dependencies between wavelet coefficients can be grouped into three categories. The first group exploits interscale dependencies while the second exploits intrascale dependencies. All techniques that exploit both inter-scale and intra-scale dependencies fall into the third category. The AntShrink algorithm belongs to the first category and this paper enhance it by converting it the third category.

While considering Inter-scale dependency, if at a given scale a coefficient is large, its correspondent at the next scale (having the same spatial coordinates) will be also large. The wavelet coefficients statistical models which exploit the dependence between coefficients give better results compared to the ones using an independent assumption (Crouse et al., 1988; Fan and Xia, 2001). That is, the estimation of coefficients in high frequency subbands based on those in lower-frequency subbands, in other words, inter-scale uses the dependency on edges. In simple terms, the correlations between the coefficients and their parents are portrayed by inter-scale dependencies. The dependencies between a coefficient and its siblings (neighborhood in the same subband) are given by the intra-scale dependencies. The various steps in the proposed hybrid algorithm is listed and explained below.

- Using the non-edge homogeneous wavelet coefficients neighbours, find parent-child relationship using inter-scale dependences of wavelet coefficients.
- Estimate local noise variance and marginal noise variance and perform denoising

Considering each cluster from ACO separately, construct a centered window and estimate the local noise variance using Equation (6).

$$\hat{\sigma}_n^2 = \left(\frac{1}{|c(i)|} \sum_{y_j \in c(y_i)} y_j^2 - \frac{\text{median}(|y_i|)}{0.6745} \right) \quad (6)$$

In this equation, the coefficient of y_i belongs to the HH band. $c(y_i)$ is defined as the coefficients within a local square window and have the same category as that is centered at the coefficient y_i. Next calculate marginal variance of noisy observations of y_1 and y_2 using Equation (7) for each wavelet coefficient.

$$\hat{\sigma}_y^2 = \frac{1}{M} \sum_{y_i \in N(k)} y_i^2 \quad (7)$$

where M is the size of the neighborhood N(k) and N(k) is defined as all coefficients within a square-shaped window that is centered at the k^{th} coefficient as illustrated in Figure 4.

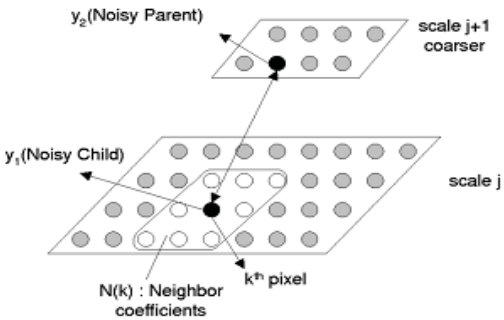


Figure-4 : Example of Neighbourhood N(k)

Then, σ can be estimated using Equation 8,

$$\sigma = \sqrt{\left(\hat{\sigma}_y^2 - \hat{\sigma}_n^2\right)_+} \tag{8}$$

where (.)+ is defined as in Equation (9),

$$(.)_+ = \begin{cases} 0 & \text{if } <0 \\ 1 & \text{Otherwise} \end{cases} \tag{9}$$

Compute the MMSE estimation for each coefficient excluding those of the LL subband, by substituting the noise variance estimated through Equation (6) into the following Equation (10),

$$S_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_n^2} y_i \tag{10}$$

Step 6: Reconstruct image

Perform the inverse wavelet transform, followed by an exponential transformation, to obtain the denoised image.

III. Experimental Results

Several experiments were conducted to evaluate the proposed model. The performance metrics used are (i) Peak Signal to Noise Ratio (PSNR) and (ii) Denoising Time. PSNR is a quality measurement between the original and a denoised image. The higher the PSNR, the better the quality of the compressed, or reconstructed image.

To compute PSNR, the block first calculates the Mean-Squared Error (MSE) and then the PSNR (Equation 11)

$$\text{PSNR} = 10 \log_{10}\left[\frac{R^2}{\text{MSE}}\right] \tag{11}$$

where, $$\text{MSE} = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M * N}$$

and M and N, m and n are number of rows and columns in the input and output image respectively.

Denoising time denotes the time taken for the algorithm to perform the despeckling procedure. Further, the proposed method was compared with Lee, Frost, SRAD, conventional Wavelet and Tian Models. Several images were used to test the proposed model. The results projected in this chapter uses the four test images shown in Figure 5(a).

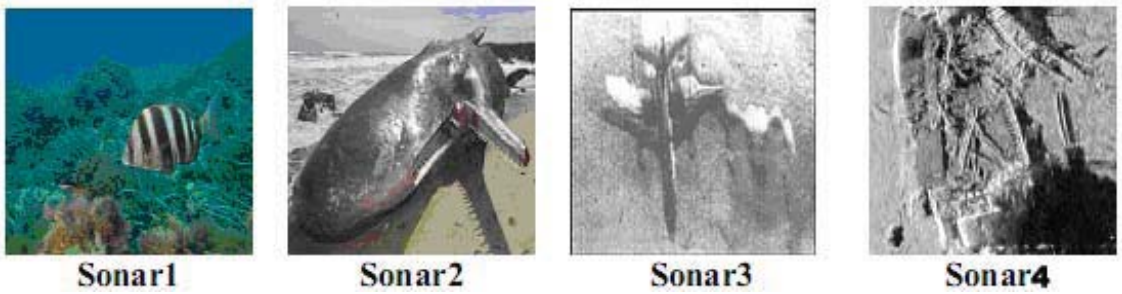


Figure-5(a) : Test Images

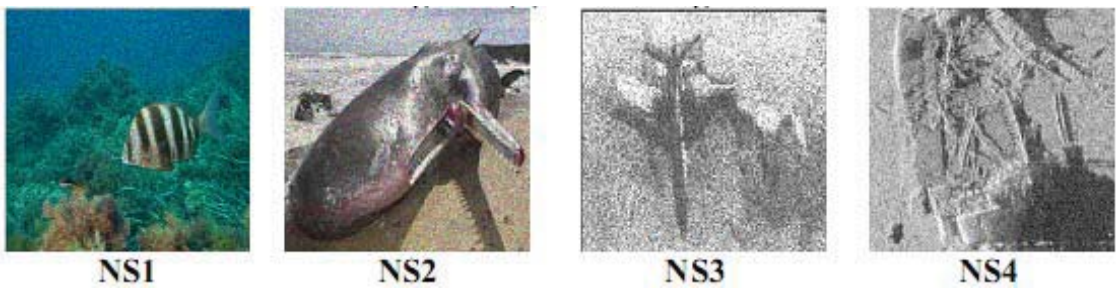


Figure-5(b) : Noisy Images

Figure 5(b) shows the speckle noisy images. The proposed model was implemented using MATLAB 7.3 and was tested on Pentium IV machine with 512 MB RAM. The PSNR values obtained for the proposed and traditional despeckling algorithms are shown in Table I.

TABLE I : PSNR (dB)

Algorithm Used	Test Images			
	NS1	NS2	NS3	NS4
Median	31	34	33	32
Kaun	30	35	32	31
Lee	30	34	31	32
Frost	32	38	34	34
SRAD	34	39	33	35
Wavelet	36	39	32	37
Tian	40	42	41	40
Proposed	42	44	44	41

From the results, it is evident that the proposed enhanced method is an improved version of the existing systems and shows significant improvement to its base method (Tian Model). The high PSNR obtained by the proposed model indicates that it is the better choice for removing Speckle noise from sonar images and produces a despeckled image whose visual quality is very near to its original noise free image. According to Venkatesan et al. (2008), an improved denoising algorithm is recognized by a high PSNR or a lower MSE. In agreement with this, the results of the proposed systems with high PSNR prove that they are an improved version over existing methods. Similarly, according to the report of Schneier and Abdel-Mottaleb (1996), a PSNR value in the range 30-40 indicates that the resultant image is a very good match to the original image. In accordance with this report, the results of all the three the proposed hybrid algorithms produce PSNR values in the range 42-44dB proving that it is an enhanced version when compared with the conventional algorithms. Figure 6 shows the time taken by the proposed and conventional filters to perform the denoising operation.

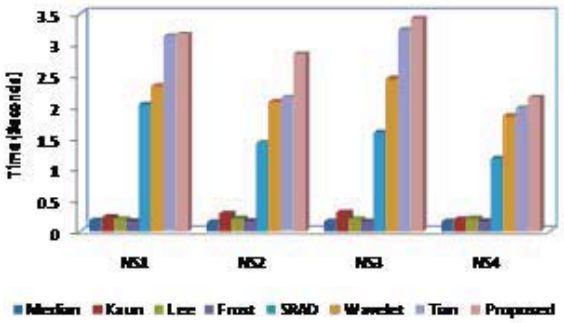


Figure-6 : Despeckling time

While considering the computational complexity, the proposed model is comparable to that of the Tian model and selected traditional despeckling algorithms. According to Müldner et al. (2005), PSNR and speed are the two most important performance factors of any denoising algorithm. From the results, it is evident that the proposed denoising algorithms is fast and produces visual quality improved images and therefore can be considered as an attractive option for several advanced applications in the field of sonar imaging. The visual comparison of the denoised image produced the proposed filters are shown in Figure 7.

Figure-7 : Despeckled Images

IV. Conclusion

The Sonar images, a type of Synthetic Aperture Radar (SAR) images, are most frequently affected with speckle noise. Speckle noise is multiplicative in nature and reduces the image quality. An important feature of sonar images is that they contain almost homogenous and textured regions and the presence of edges is relatively rare. This paper proposed a non-parametric statistical model using hybrid intra-scale and inter-scale dependencies of wavelet coefficients for removing speckle noise from speckled sonar images. First, the multiplicative speckle noise that disturbs the SONAR images is transformed into an additive noise with the aid of a logarithm computation block, after which a

stationary wavelet is applied. The inter-scale and intra-scale dependency of the wavelet coefficients are exploited during denoising. The experimental results prove that the proposed method is efficient in terms of reduction in speckle noise and speed. In future, other wavelet variants like complex wavelets, wavelet tree are to be explored.

References references referencias

1. Arsenault, H. and April, G. (1976) Properties of speckle integrated with a finite aperture and logarithmically transformed, J. Opt. Soc. Am., Vol. 66 , Pp. 1160–1163.
2. Crouse, M.S., Nowak, R.D. and Baraniuk, R.G. (1998) Wavelet-based signal processing using hidden Markov models, IEEE Trans. Signal Processing, Vol. 46, Pp. 886–902.
3. Delakis, I., Hammad, O. and Kitney, R.I.(2007) Wavelet-based de-noising algorithm for images acquired with parallel magnetic resonance imaging (MRI), Physics in Medicine and Biology, Vol. 52, No.13, Pp.3741.
4. Fan, G. and Xia, X.G. (2001) Image denoising using a local contextual hidden Markov model in the wavelet domain, IEEE Signal Processing Lett., Vol. 8, Pp. 125–128.
5. Fowler, J.E. (2005) The Redundant Discrete Wavelet Transform and Additive Noise IEEE Signal Processing Letters, Vol. 12 Issue 9, Pp.629 – 632.
6. Guo, Y., Wang, Y. and Hou, T. (2011) Speckle filtering of ultrasonic images using a modified non local-based algorithm, Biomedical Signal Processing and Control, Elsevier, Vol. 6, Pp. 129-138.
7. Holschneider, M., Kronland-Martinet, R., Morlet, J. and Tchamitchian, P. (1989) A real-time algorithm for signal analysis with the help of the wavelet transform, Wavelets, Time-Frequency Methods and Phase Space, Springer-Verlag, Pp. 289–297.
8. Kaur, A. and Singh, K. (2010) Speckle noise reduction by using wavelets, NCCI 2010 - National Conference on Computational Instrumentation CSIO, Chandigarh, INDIA, Pp. 198-203.
9. Müldner, T., Leighton, G. and Diamond, J. (2005) Using XML compression for WWW communication, Proceedings of the IADIS WWW/Internet 2005 Conference, 2005.
10. Nason, G.P. and Silverman, B.W. (1995) The stationary wavelet transform and some statistical applications, Tech. Rep. BS8 1Tw, University of Bristol.
11. Pardo, A. (2011) Analysis of non-local image denoising methods, Pattern Recognition Letter, Elsevier, In Press. doi:10.1016/j.patrec.2011.06.022.
12. Schneier, M. and Abdel-Mottaleb, M. (1996) Exploiting the JPEG compression scheme for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell., Vol.18, No. 8, Pp. 849–853.
13. Schulze, M.A. (1997) An edge enhancing nonlinear filter for reducing multiplicative noise, Proc. of SPIE, Vol. 3026, Pp. 46–56.
14. Shensa, M.J. (1992) The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms, IEEE Transaction on Signal Processing, Vol 40, No 10, Pp.2464-2482.
15. Tian, J. and Chen, L.(2011) Image despeckling usng a non-parametric statistical model of wavelet coefficients, Biomedical Signal Processing and Control, Science Direct, Vol. 6, Pp. 432-437.
16. Tian, J., Yu, W. and Ma, L. (2010) AntShrink: Ant colony optimization for image shrinkage, Pattern Recognition Letters, Science Direct, Vol. 31, Pp.1751-1758.
17. Venkatesan, M., MeenakshiDevi, P., Duraiswamy, K. and Thyagarajah, K. (2008) Secure Authentication Watermarking for Binary Images using Pattern Matching, IJCSNS International Journal of Computer Science and Network Security, Vol.8, No.2, Pp. 241-250.
18. Xie, H., Pierce, L. and Ulaby, F. (2002) Statistical properties of logarithmically transformed speckle, IEEE Trans. Geosci. Remote Sens., Vol. 40, Pp. 721–727.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

PTP-Mine: Range Based Mining of Transitional Patterns in Transaction databases

By Shyamala Pogula, Sujatha Dandu
Auroras Technological and Research Institute, JNTU

Abstract - Transaction database is a collection of transactions along with the related time stamps. These transactions are defined using some prototypes. They are called as the Transitional patterns that denote the vibrant nature of the frequent patterns in the database. The considerable high points for the transaction database are the timestamps also called as time durations. They are the points that have the alteration in the recurrence of the prototypes. There is majorly couple of stages in existing TP-Mine algorithm. The initial stage is to find out the frequent patterns and the second stage is to discover Transitional patterns or the styles and the significant milestones. These patterns consist of the two kinds of the styles likely the positive one and the negative one. In the previous time cases the effort that was made on the research was to build up the algorithms by planning the total series of the transitional patterns. In our paper we consider that the alterations made in the consequent period regarding the total concept of database is not noteworthy. So for this reason we have put forward an entirely latest transitional patterns methodology called periodical transitional pattern mining. The experimental outputs are appealing and apparent those were produced by this periodical transitional pattern mining and has high importance that when evaluated utilizing the present patterns.

Keywords : *Data mining, Association Rules, Frequent Pattern, Transitional Pattern, significant milestones*

GJCST Classification: *H.2.8*

Strictly as per the compliance and regulations of:

PTP-Mine: Range Based Mining of Transitional Patterns in Transaction databases

Shyamala Pogula [α], Sujatha Dandu [Ω]

Abstract - Transaction database is a collection of transactions along with the related time stamps. These transactions are defined using some prototypes. They are called as the Transitional patterns that denote the vibrant nature of the frequent patterns in the database. The considerable high points for the transaction database are the timestamps also called as time durations. They are the points that have the alteration in the recurrence of the prototypes. There is majorly couple of stages in existing TP-Mine algorithm. The initial stage is to find out the frequent patterns and the second stage is to discover Transitional patterns or the styles and the significant milestones. These patterns consist of the two kinds of the styles likely the positive one and the negative one. In the previous time cases the effort that was made on the research was to build up the algorithms by planning the total series of the transitional patterns. In our paper we consider that the alterations made in the consequent period regarding the total concept of database is not noteworthy. So for this reason we have put forward an entirely latest transitional patterns methodology called periodical transitional pattern mining. The experimental outputs are appealing and apparent those were produced by this periodical transitional pattern mining and has high importance that when evaluated utilizing the present patterns.

Keywords : Data mining, Association Rules, Frequent Pattern, Transitional Pattern, significant milestones

I. Introduction

In the recent time the expansively considered field in the data mining is the frequent pattern mining. The advantage of this technique is that it can be widely utilized in several methodologies such as Association rule mining [2] [3] [12], sequential pattern mining [4], structured pattern mining [7], correlation mining [5], and associative classification [9]. Even to determine these frequent patterns that are present in the transaction databases, plenty of procedures were put forward specifically Apriori [3], FP-growth [8], and Eclat [15]. The drawback of these procedures is that they can produce huge amount of the patterns on the condition of low support threshold value. As this drawback being one of the reasons, maximum procedures are not utilized for the data mining job. So, as a case of evading the above mentioned ineffectual and outmoded patterns, the latest patterns such as frequent itemsets [1] [6] [11], and closed frequent itemsets [10] were been established.

One of the common characteristic of these procedures is that they don't determine the time stamps that are related to the actions in the database. This leads to the loss of exposure of the vibrant nature of the patterns. For an illustration let us consider an electronic showroom's database. The amount of retails of refrigerator during the peak summer season is very much more when compared to that of the retails in the winter. But when we determine all the transactions in whole on an average the retails are frequent, while in actual they are frequent only in the peak season. So in order to find out such a vibrant characteristics, latest patterns have been established like Transitional patterns [13] [14]. There are two kinds in them such as the positive patterns and the negative patterns. Positive transitional patterns have the characteristic of incrementing the recurrence of the pattern's time stamp where as the negative patterns decrement the recurrence of the pattern's time stamp.

The periodic positions where the recurrence of the transitional patterns alters often are termed as the significant milestones of the transitional pattern. So to determine these transitional patterns and the respective significant milestones TP-Mine algorithm [13][14] [17] has been projected.

The two most important and chief stages of the TP-Mine[14] algorithm are as follows:

1. In the Initial stage each and every frequent pattern is produced from the present transaction database.
2. The consumer have some of the preferences, In the consequent stage based on those preferences the frequent patterns are utilized for the production of the transitional patterns and significant milestones,

By considering all the above mentioned problems we in our article are introducing an adaption to the present TP-Mine[14] algorithm with a view of decrementing the patterns that were derived in the initial stage. As a result the count of the calculations required for the production of transitional patterns gets decremented. By this there is a feasibility to develop the importance of transitional patterns by keeping in account the periodicity of the mentioned series as the milestones.

The structure of the remaining paper is designed as below: The second part of the paper illustrates the preface and defining of the terms that

Author α : Student, Department of CSE, Aurora's Technological & Research Institute uppal, Hyderabad.
Author Ω : Department of CSE, Aurora's Technological & Research Institute, uppal, Hyderabad.

were being utilized in the TP-Mine[14] algorithm. The later part that is the 3rd one illustrates the PTP-Mine Algorithm and the 4th one represents the tentative outputs of the approach. And as a final part, the 5th one terminates the approach.

II. Preface and Definitions

The following are the definitions taken from [14]. Apart from these, new definition defined in definition 3.2. The major elementary function of data mining approaches in digging out the valuable patterns that are present in the databases is excavation of the frequent patterns. Assuming a group of items is I= {i1, i2 … in} with a group of database transactions D that consists of individual transaction T. T is termed as the group of items and the count of the transactions in D is derived by ||D||.

The specified equation $X = \{i_j ... i_k\} \subseteq I(j \le k \; and \; 1 \le j, \; k \le n)$ is termed as a pattern. The pattern's X in an action D has a support value, which is termed as the count of the transactions in the D that consists of the respective pattern X. This X is defined as recurrent or simply frequent if and only if the corresponding support value is greater to the consumer mentioned the least support threshold value.

Definition 2.1: The representation of the cover for the individual itemset X in D. This is represented using cov(X, D). This is the count of the transactions that consists of the item X.

Definition 2.2: The individual itemset X for the transaction database D consists of a support value. The representation of it is sup(X, D). This is termed as the proportionate value involving the cov(X, D) and the count of transactions in D which is represented by||D||.

$$\sup(X, D) = \frac{\text{cov}(X, D)}{\| D \|}$$

Definition 2.3: Let us consider that the all the transactions in the transaction database D are arranged according to the respective time-stamps then the position of the transaction T in D which is represented using $\rho(T)$, is defined as the count transactions for which the time-stamp is lower or equivalent to the time stamp value of T. Hence $1 \le \rho(T) \le \| D \|$.

Definition 2.4: For a pattern X in D, the i^{th} transaction which is represented by $T^i(X)$ is defined as the i^{th} transaction for the cover of X i.e., $\text{cov}(X)$ that has the transactions arranged based on their positions. Here $1 \le i \le \text{cov}(X, D)$.

Definition 2.5: For a pattern X in D, the i^{th} milestone which is represented by $\xi^i(X)$, is termed as

$$\xi^i(X) = \frac{\rho(T^i(X))}{\| D \|} \times 100\%$$, where $1 \le i \le \text{cov}(X)$

Definition 2.6: For a pattern X in D, the support value preceding its i^{th} milestone in D which is represented by $\sup^i\_(X)$ is termed as follows:

$$\sup^i\_(X) = \frac{i}{\rho(T^i(X))}$$, where $1 \le i \le \text{cov}(X)$

Definition 2.7 For a pattern X in D, the support value following its i^{th} milestone in D which is represented by $\sup^i{}_+(X)$ is termed as follows:

$$\sup^i{}_+(X) = \frac{\text{cov}(X) - i}{\| D \| - \rho(T^i(X))}$$, where $1 \le i \le \text{cov}(X)$

Definition 2.8: For a pattern X in D at the i^{th} milestone the Transitional ratio is termed as follows:

$$tran^i(X) = \frac{\sup^i{}_+(X) - \sup^i\_(X)}{Max(\sup^i{}_+(X), \sup^i\_(X))}$$,

where $1 \le i \le \text{cov}(X)$

Definition 2.9: For a pattern X in D is considered as transitional pattern (TP) in D if on minimum, single milestone of $X, \xi^k(X) \in T_\xi$ must be present, so that:

$$\sup^k\_(X) \ge t_s \; and \quad \sup^k{}_+(X) \ge t_s \; and$$
$$| tran^i(X) | \ge t_t$$

here T_ξ is the series of $\xi^i(X)$ ($1 \le i \le \text{cov}(X)$) t_s and t_t are named correspondingly as pattern support threshold and transitional pattern threshold. Here X is defined as a Positive Transitional Pattern (PTP) only on condition that $trans^k(X) > 0$; and X is defined as a Negative Transitional Pattern (PTP) only on condition that $tran^k(X) < 0$.

Definition 2.10: The significant frequency-incrementing milestone for the derived positive transitional pattern X by considering the time period T_ξ is termed in form of a tuple, $(\xi^M(X), tran^M(X)), \; where \; \xi^M(X) \in T_\xi$ is considered as the M[th] milestone for a pattern X such that

$$\sup^M\_(X) \ge t_s \; and \; \forall \xi^i(X) \in T_\xi, \; tran^M(X) \ge tran^i(X)$$

Definition 2.11: The significant frequency-

decrementing milestone for the derived negative transitional pattern X by considering the time period T_ξ is termed in form of a tuple, $(\xi^N(X), tran^N(X))$, *where* $\xi^N(X) \in T_\xi$ is considered as the N[th] milestone for a pattern X such that $\sup_+^N(X) \geq t_s$ *and* $\forall \xi^i(X) \in T_\xi$, $tran^N(X) \leq tran^i(X)$

III. TP-Mine[14] Algorithm

Let us assume the below transaction database as represented in Table 1.

Table 1 : Illustration Database

TID	List of Item IDs	Time stamp
001	P1,P2,P3,P5	Nov, 2005
002	P1,P2	Dec, 2005
003	P1,P2,P3,P8	Jan, 2006
004	P1,P2,P5	Feb, 2006
005	P1,P2,P4	Mar, 2006
006	P1,P2,P4,P5,P6	Apr, 2006
007	P1,P2,P3,P4,P6	May, 2006
008	P1,P4,P6	Jun, 2006
009	P4,P5,P6	Jul, 2006
010	P1,P2,P3,P4,P5,P6	Aug, 2006
011	P1,P3,P4,P6	Sep, 2006
012	P1,P3,P5	Oct, 2006
013	P1,P2,P3,P6,P7	Nov, 2006
014	P1,P3,P4,P5	Dec, 2006
015	P1,P3,P4	Jan, 2007
016	P1,P2,P3,P5	Feb, 2007

The group of positive and negative transitional patterns including their significant milestones is produced using TP-Mine algorithm [14].

Input: A transaction database (D), milestone range (T_ξ), pattern support threshold (t_s), and transitional pattern threshold(t_t).

D is assumed as a group of transactions scheduled in Table 1, T_ξ = {25%, 75%}, t_s =0.05, t_t =0.5. As mentioned above the algorithm consists of a couple of important stages. Coming to the initial stage the production of every recurrent itemsets including the support values is done utilizing Apriori [3] or FP-growth [8] considering t_s as minimum support threshold. This stage involves the generation of n count of frequent patterns by the algorithm for which support $\geq t_s$ in the transaction database D. The consequential frequent patterns (n=87) have been represented in Table 2.

Subsequently in the second stage the support count (c_k) for each and every frequent patterns in the group of transactions starting from the initial action to latest one with the time stamp T_ξ is derived by the algorithm. Later the milestones of $P_k (1 \leq k \leq n)$ in the series T_ξ are also derived. It checks whether the derived milestones are legal or not and for the legal ones $\xi^{c_k}(P_k)$ it determines the value for the support P_k preceding $\xi^{c_k}(P_k)$ i.e., $\sup_-^{c_k}(P_k)$ and support value P_k following $\xi^{c_k}(P_k)$ i.e., $\sup_+^{c_k}(P_k)$. Then it compares whether the values are higher to that of t_s, and if they are higher, then algorithm verifies the transitional ratio P_k. It again compares whether the ratio value is more than that of the t_t. If more, then P_k is considered as the positive transitional pattern. This comparison is continued with the current maximal transitional ratio of P_k, if proved to be more, then that group of significant frequency incrementing milestones of P_k is made to consist the $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ as its only component. If proved to be less, then $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ is summed up with the significant frequency incrementing milestone of P_k. This format is repeated for the determination of negative transitional patterns and their significant milestones in the condition that $tran^{c_k}(P_k) \leq -t_t$.

Table 2 : A group of frequent patterns

FP	Sup
P1	15
P3,P5	5
P1,P6,P7	1
P1,P2,P5,P6	2
P2	10
P3,P6	4
P2,P3,P4	2
P1,P2,P6,P7	1
P3	10
P3,P7	1
P2,P3,P5	3
P1,P3,P4,P5	2
P4	9
P3,P8	1
P2,P3,P6	3
P1,P3,P4,P6	3
P5	8
P4,P5	4
P2,P3,P7	1
P1,P3,P5,P6	1
P6	7
P4,P6	6
P2,P3,P8	1
P1,P3,P6,P7	1
P7	1
P5,P6	3
P2,P4,P5	2
P1,P4,P5,P6	2
P8	1
P6,P7	1

P2,P4,P6	3
P2,P3,P4,P5	1
P1,P2	10
P1,P2,P3	6
P2,P5,P6	2
P2,P3,P4,P6	2
P1,P3	10
P1,P2,P4	4
P2,P6,P7	1
P2,P3,P5,P6	1
P1,P4	8
P1,P2,P5	5
P3,P4,P5	2
P2,P3,P6,P7	1
P1,P5	7
P1,P2,P6	4
P3,P4,P6	3
P2,P4,P5,P6	2
P1,P6	6
P1,P2,P7	1
P3,P5,P6	1
P3,P4,P5,P6	1
P1,P7	1
P1,P2,P8	1
P3,P6,P7	1
P1,P2,P3,P4,P5	1
P1,P8	1
P1,P3,P4	5
P4,P5,P6	3
P1,P2,P3,P4,P6	2
P2,P3	6
P1,P3,P5	5
P1,P2,P3,P4	2
P1,P2,P3,P5,P6	1
P2,P4	4
P1,P3,P6	4
P1,P2,P3,P5	3
P1,P2,P3,P6,P7	1
P2,P5	5
P1,P3,P7	1
P1,P2,P3,P6	3
P1,P2,P4,P5,P6	2
P2,P6	4
P1,P3,P8	1
P1,P2,P3,P7	1
P1,P3,P4,P5,P6	1
P2,P7	1
P1,P4,P5	3
P1,P2,P3,P8	1
P2,P3,P4,P5,P6	1
P2,P8	1
P1,P4,P6	5
P1,P2,P4,P5	2
P1,P2,P3,P4,P5,P6	1
P3,P4	5
P1,P5,P6	2
P1,P2,P4,P6	3

Table 3 : A group of Positive transitional patterns and the corresponding incrementing milestones.

Pattern	Incrementing Milestone	Transitional Ratio
P3	Aug2006(62.5%)	60.0%
P4	Mar2006(31.25%)	72.50%
P6	Apr2006(37.5%)	72.0%
P1, P3	Aug2006(62.5%)	60.0%
P1, P4	Mar2006(31.25%)	68.57%
P1, P6	Apr2006(37.5%)	66.66%
P3, P4	May2006(43.75%)	67.85%
P3, P5	Aug2006(62.5%)	60%
P3, P6	May2006(43.75%)	57.14%
P4, P6	Apr2006(37.5%)	66.66%
P1,P3, P4	May2006(43.75%)	67.85%
P1,P3, P5	Aug2006(62.5%)	60%
P1,P3, P6	May2006(43.75%)	57.14%
P1,P4, P6	Apr2006(37.5%)	58.33%

Table 4 : A group of Negative transitional patterns and the corresponding decrementing milestones.

Pattern	Decrementing Milestone	Transitional Ratio
P2	May2006(43.75%)	-66.66%
P6	Sep2006(68.75%)	-63.33%
P1, P2	May2006(43.75%)	-66.66%
P1, P6	Sep2006(68.75%)	-56.0%
P2, P4	May2006(43.75%)	-74.07%
P2, P5	apr2006(37.5%)	-60.0%
P4, P6	aug2006(62.5%)	-66.66%
P1,P2, P4	may2006(43.75%)	-74.07%
P1,P2, P5	apr2006(37.5%)	-60.0%
P1,P4, P6	aug2006(62.5%)	-58.33%
P2,P4, P6	may2006(43.75%)	-61.11%
P1,P2,P4, P6	may2006(43.75%)	-61.11%

There is a clear understanding from the outcome in the tables that the count for the transitional patterns that were produced in the transaction data group of magnitude 16 that come in the lower region of support and threshold value is just 26. And it is also found out that those transactions that were produced from the standard data groups that has more than ten transactions directs towards the large value complication and maximum patterns are also not significant patterns. Due to this reason the patterns must be produced in the periodical series which will help us to determine the large significant patterns. In order to reduce the complication an algorithm called Bide [16] algorithm is preferred for determining the frequent patterns and decrementing the count of the pattern.

Definition 3.1: A pattern can be sniped out if it is a resident of other patterns and their support values are equal.

Assuming that A *and* B are a couple of patterns such that

$$\sup(A) \text{ is } \omega$$
$$\sup(B) \text{ is } \omega'$$
$$\text{and } \omega \cong \omega'$$

if $A \subseteq B$ *then* A has a chance to be sniped out from the group of patterns.

Definition 3.2: The development in the impact of the transitional patterns is done by checking T_ξ for the specified timely intervals of transactions rather than checking the particular time threshold value.

Assuming that $TD = \{td_1, td_2, ...td_n\}$ is a group of dates on which the transitions are occurred,

$$TDR = \{tdr_1, tdr_2, ...tdr_s\}$$
$$here$$
$$tdr_1 = \{td_1, td_2, ...td_m\}$$
$$tdr_2 = \{td_{m+1}, td_{m+2}, ...td_{m+i}\}$$
$$tdr_3 = \{td_{m+i+1}, td_{m+i+2}, ..., td_{m+i+x}\}$$
$$.....$$
$$td_n = \{td_{m+i+x}....td_n\}$$

TD denotes the group of transitional dates

TDR denotes the group of transition interval and every interval consists of the group of transition dates.

At this instant the transition patterns threshold T_ξ checked among the interval Transition range tdr_i .

As a result there is an apparent development in the count of the transition patterns that are based on tdr_i when compared to that of transition patterns based on td_i

$$TP(tdr_i) = TP(td_{i_1}) \cup TP(td_{i_2}) \cup TP(td_{i_3}) .... \cup TP(td_{i_t})$$

Periodic Transitional Patterns

Input:

A transaction database (D), an appropriate milestone range that the user is interested (T_ξ), pattern support threshold (t_s), and transitional pattern threshold (t_t). transition range(tdr_i)

Output:

The group of transitional patterns (positive(S_{PTP}) and negative(S_{NTP})) including their significant milestones.

Algorithm:

1. Determine frequent patterns
   a. In order to preserve the individual patterns, snip out the patterns that are as subset for the superset frequent patterns and has the equal support values(definition 3.1). As a result the count of the patterns gets reduced.
2. Utilizing the Periodic Transitional Pattern –Mining methodology.
   a. Check the positive and negative transition for every pattern in a group of transitional dates tdr_i (definition 3.2).
3. In order to evaluate the supports s_{ptp} for a pattern aroused preceding tdr_i and supports s_{ntp} for the same pattern aroused following tdr_i , examine the transaction data group.
4. $S_{PTP} = \emptyset$, $S_{NTP} = \emptyset$
5. for all k = 1 to n do
6. MaxTran (P_k) = 0, MinTran ((P_k) = 0
7. $S_{FAM}(P_k) = \emptyset$, $S_{FDM}(P_k) = \emptyset$
8. end for
9. for all transactions T_i whose position satisfying T_ξ do
10. for k = 1 to n do
11. if $T_i \supseteq P_k$ then
12. $C_k = C_k + 1$
13. if $\sup_+^{c_k}(P_k) \geq t_s$ and $\sup_-^{c_k}(P_k) \geq t_s$ then
14. if $tran^{c_k}(P_k) \geq t_t$ then
15. if $P_k \notin S_{PTP}$ then
16. Add P_k to S_{PTP}
17. end if
18. if $tran^{c_k}(P_k) > MaxTran(p_k)$ then
19. $S_{FAM}(P_k) = \{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$
20. $MaxTran(P_k) = tran^{c_k}(P_k)$
21. else if $tran^{c_k}(P_k) = MaxTran(P_k)$ then
22. Add $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ to $S_{FAM}(P_k)$
23. end if
24. else if $tran^{c_k}(P_k) \leq -t_t$ then
25. if $P_k \notin S_{NTP}$ then
26. Add P_k to S_{NTP}
27. end if
28. If $tran^{c_k}(P_k) < MinTran(p_k)$ then
29. $S_{FDM}(P_k) = \{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$
30. $MinTran(P_k) = tran^{c_k}(P_k)$
31. else if $tran^{c_k}(P_k) = MinTran(P_k)$ then
32. Add $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ to $S_{FDM}(P_k)$

33. end if
34. end if
35. end if
36. end if
37. end for
38. end for
39. return S_{PTP} and S_{FAM} , (P_k) for each $P_k \in S_{PTP}$
40. return S_{NTP} and S_{FDM} , (P_k) for each $P_k \in S_{NTP}$

IV. Tentative Outputs

Input: A transaction database (D), milestone range (T_ξ), pattern support threshold (t_s), and transitional pattern threshold (t_t).

D is assumed as a group of transactions scheduled in Table 1, T_ξ = 25%, t_s =0.05, t_t =0.1, tdr_i =3months.

Table 5 : A group of frequent patterns which are generated by PTP-Mine algorithm.

FP	SUP	FP	SUP
P1	15	P1, P3, P6	4
P1, P2	10	P1, P2, P3, P5	3
P1, P3	10	P1, P4, P5	3
P4	9	P4, P6	6
P1, P4	8	P1, P2, P3, P6	3
P1, P2, P3	6	P1, P4, P6	5
P5	8	P1, P3, P4, P5	2
P1, P5	7	P1, P2, P4, P6	3
P1, P2, P4	4	P1, P3, P4, P6	3
P1, P6	6	P1, P2, P3, P8	1
P6	7	P1, P2, P3, P4, P6	2
P1, P3, P4	5	P4, P5, P6	3
P1, P2, P5	5	P1, P2, P4, P5, P6	2
P1, P3, P5	5	P1, P2, P3, P6, P7	1
P1, P2, P6	4	P1, P2, P3, P4, P5,	1
P4, P5	4		

Table 6 : some set of Positive Periodic Transitional pattern and the corresponding incrementing milestones.

Pattern	Incrementing Milestone	Transitional Ratio
P4,P5	Apr2006(37.5%)	44.44%
P1,P5	Aug2006(62.5%)	19.99%
P1,P6	Apr2006(37.5%)	66.66%
P1,P3,P4,P5	Aug2006(62.5%)	40.0%
P3,P1	Aug2006(62.5%)	60.0%
P4,P2,P1,P6	Apr2006(37.5%)	16.66%
P4,P1	Mar2006(31.25%)	68.57%
P3,P2,P1,P6	May2006(43.75%)	35.71%
P4,P3,P1,P6	May2006(43.75%)	35.71%
P4,P6	Apr2006(37.5%)	66.66%
P4	Mar2006(31.25%)	72.5%
P4,P1,P6	Apr2006(37.5%)	58.33%
P1	Aug2006(62.5%)	10.00%
P4,P6,P5	Apr2006(37.5%)	16.66%
P6	Apr2006(37.5%)	72.26
P5	July2006(56.25%)	22.22
P4,P2,P1	Mar2006(31.25%)	26.66
P4,P3,P1	May2006(43.75%)	67.85
P3,P1,P5	Aug2006(62.5%)	60.0%
P2,P1,P6	Apr2006(37.5%)	44.44%
P3,P1,P6	May2006(43.75%)	57.64%
P4,P1,P5	Apr2006(37.5%)	16.66%

Table 7 : some set of Negative Periodic Transitional Patterns and the corresponding Decrementing Milestones.

Pattern	Decrementing Milestone	Transitional ratio
P4,P2,P1,P6	May2006(43.75%)	-61.11%
P3,P2,P1,P6	Aug2006(62.5%)	-16.66%
P4,P3,P1,P6	Aug2006(62.5%)	-16.66%
P3,P2,P1,P5	Aug2006(62.5%)	-16.66%
P4,P1,P6	Aug2006(62.5%)	-58.33%
P4,P6,P5	July2006(56.25%)	-35.71%
P3,P2,P1	May2006(43.75%)	-22.22%
P4,P2,P1	May2006(43.75%)	-74.07%
P2,P1,P5	Apr2006(37.5%)	-60.0%
P2,P1,P6	Aug2006(62.5%)	-44.44%
P4,P1,P5	Aug2006(62.5%)	-16.66%
P3,P1,P6	Sep2006(68.75%)	-26.66%
P4,P5	Aug2006(62.5%)	-44.44%
P1,P5	Apr2006(37.5%)	-19.99%
P1,P6	Sep2006(68.75%)	-56.0%
P4,P1	Sep2006(68.75%)	-26.66%
P2,P1	May2006(43.75%)	-66.66%
P4,P6	Aug2006(62.5%)	-66.66%
P4	Sep2006(68.75%)	-37.14%
P1	June2006(50.0%)	-12.50%
P4,P2,P1,P6,P5	Apr2006(37.5%)	-40.00%
P4,P3,P2,P1,P6	May2006(43.75%)	22.22%
P6	Sep2006(68.75%)	-63.33%

Many tests have been organized on artificial data groups having 8 values and with transactions greater than 200. JAVA 1.6_ 20th build was utilized for execution of the PTP-Mine including TP-Mine[14] for examination. A computer unit prepared with core2duo processor, 2GB RAM and Windows XP loaded was utilized for inquiring the algorithms.

a) Dataset Characteristics

The dataset prepared is a very opaque dataset, which assists in excavating enormous quantity of recurring clogged series with a profitably high threshold

somewhere close to 90%. It also has a distinct element of being enclosed with more than 200 transactional series and 8 divergent objects. Reviewing of serviceable legacy's consistency has been made use of by this dataset.

Comparative study: In assessment with TP-Mine has made its mark as a most preferable, superior and sealed example of transitional pattern mining

Table 8 : contrast account of patterns derived below diverse supports by PTP-Mine and TP-Mine[14]

Support	TP-Mine[14]	PTP-Mine
20%	432	382
30%	411	374
40%	390	347
50%	310	292
60%	124	93
70%	35	32
80%	11	9

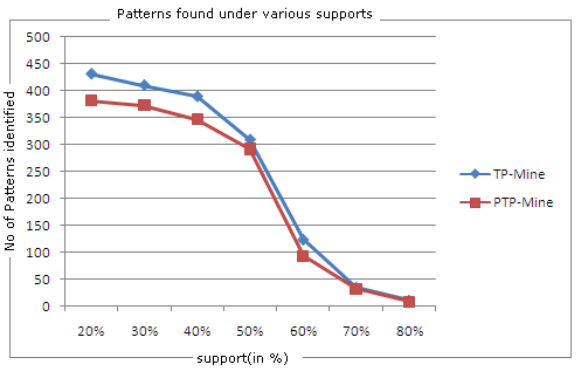


Fig1 : Count of the patterns derived by TP-Mine[14] and PTP-Mine

Table 9 : Contrast account of +ve Transitional Patterns derived below diverse supports by PTP-Mine and TP-Mine[14]

Support	TP-Mine[14]	PTP-Mine
20%	80	142
30%	63	127
40%	57	83
50%	41	69
60%	23	45
70%	7	19
80%	3	14

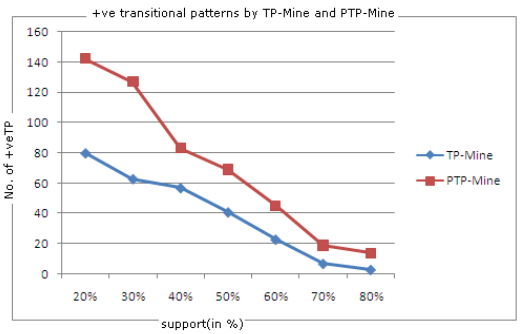


Fig 2 : Positive Transition patterns derived by TP-Mine[14] and PTP-Mine.

Table 10 : Contrast account of -ve Transitional Patterns derived below diverse supports by PTP-Mine and TP-Mine[14]

support	TP-Mine[14]	PTP-Mine
20%	234	434
30%	217	449
40%	111	493
50%	105	507
60%	223	531
70%	139	557
80%	143	562

The fig-1 represents the count of the patterns that were reduced on contrasting the PTP-Mine with the TP-Mine[14] and equally we can distinguish that the functioning of PTP-Mine in determining the +ve and –ve transitional patterns is sound. This serves as the proof for defining that the PTP-Mine subsequently works better on contrasting with the TP-Mine[14]. The PTP-Mine decreases the count of patterns in order to decrement the calculation complications and generates a large amount of transitional patterns to progress the construction of constraint dependent resolutions impact (fig 2 and fig 3).

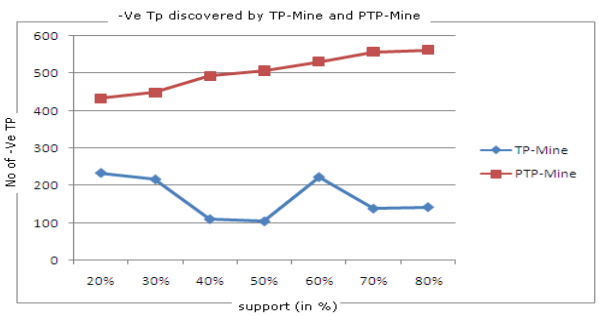


Fig 3 : Whole count of the –Ve transitional patterns derived by TP-Mine[14] and PTP-Mine

V. Conclusion

We observe that Compare to TPMine[14], PTP-Mine produces less no. of frequent patterns. TP-Mine[14] algorithm produces less number of positive and negative transitional patterns because of transitional pattern threshold that is assumed to be 0.5, in which case only the patterns which change dramatically are displayed. In PTP-Mine the transition pattern threshold is considered as 0.1 so that, even the positive and negative patterns which have less change are also exhibited. The tentative outcome delivered by our methodological algorithm that was put forward is decidedly resourceful and proficient.

References Referencias

1. R.C. Agarwal, C.C. Aggarwal, and V.V.V. Prasad, "Depth First Generation of Long Patterns," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '00), pp. 108-118, 2000.
2. R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. 1993 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '93), pp.207-216, 1993.
3. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases, pp. 487-499, 1994.
4. R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proc.11th Int'l Conf. Data Eng., pp. 3-14, 1995.
5. S. Brin, R. Motwani, and C. Silverstein, "Beyond Market Baskets: Generalizing Association Rules to Correlations," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '97), pp. 265-276, 1997.
6. D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemset Algorithm," IEEE Trans.Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.
7. J. W. Han, J. Pei, and X. F.Yan, "From Sequential Pattern Mining to Structured Pattern Mining: A Pattern-Growth Approach," J. Computer Science and Technology, vol. 19, no. 3, pp. 257-279, 2004.
8. J.W. Han, J. Pei, Y. W. Yin, and R.Y. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," Data Mining and Knowledge Discovery, vol. 8, no. 1, pp. 53-87, 2004.
9. B. Liu, W. Hsu, and Y.-M. Ma, "Integrating Classification and Association Rule Mining," Proc. Fourth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '98), pp. 80-86, 1998.
10. N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Discovering Frequent Closed Itemsets for Association Rules," Proc. Seventh Int'l Conf. Database Theory (ICDT '99), pp. 398-416, 1999.
11. R.J. Bayardo, Jr., "Efficiently Mining Long Patterns from Data-bases," Proc. 1998 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '98), pp. 85-93, 1998.
12. R. Srikant and R. Agrawal, "Mining Generalized Association Rules," Future Generation Computer Systems, vol. 13, nos. 2/3,pp. 161-180, 1997.
13. Q. Wan and A. An, "Transitional Patterns and Their Significant Milestones," Proc. Seventh IEEE Int'l Conf. Data Mining, 2007.
14. Q. Wan and A. An, "Discovering Transitional Patterns and Their Significant Milestones in Transaction Databases," IEEE Trans. on Knowledge and Data Engg., vol.21,No.12 pp.1692-1707, Dec 2009.
15. M.J. Zaki and K. Gouda, "Fast Vertical Mining Using Diffsets," Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 326-335, 2003.
16. Jianyong Wang and Jiawei Han. 2004. BIDE: Efficient Mining of Frequent Closed Sequences. In Proceedings of the 20th International Conference on Data Engineering (ICDE '04). IEEE Computer Society, Washington, DC, USA, 79-.
17. Rekharani Sidduri , T.P.Shekhar, Ganesh Kumar Nune "Mining Dynamic Behavior of Frequent Patterns using TP-Mine Algorithm" IJCTT-Volume 1 Issue 3/ July - Aug 2011.

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Improving Academic Performance of Students of Defence University Based on Data Warehousing and Data mining

By Dr. Vuda Sreenivasarao, Capt. Genetu Yohannes
Defence University College, Debrezeit, ETHIOPIA

Abstract - The student academic performance in Defence University College is of great concern to the higher technical education managements, where several factors may affect the performance. The student academic performance in engineering during their first year at university is a turning point in their educational path and usually encroaches on their general point average in a decisive manner. The students evaluation factors like class quizzes mid and final exam assignment are studied. It is recommended that all these correlated information should be conveyed to the class teacher before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant level and improve the performance of students. Statistics plays an important role in assessment and evaluation of performance in academics of universities need to have extensive analysis capabilities of student achievement levels in order to make appropriate academic decisions. Academic decisions will result in academic performance changes, which need to be assessed periodically and over span of time. The performance parameters chosen can be viewed at the individual student, department, school and university levels. Data mining is used to extract meaning full information and to develop significant relationships among variables stored in large data set/ data warehouse. In this paper is an attempt to using concepts of data mining like k-Means clustering, Decision tree Techniques, to help in enhancing the quality of the higher technical educational system by evaluating student data to study the main attributes that may affect the performance of student in courses.

Keywords : *Data base, Data warehousing, Data mining, Academic Performance, Educational data mining , Student performance analysis and K-Means clustering algorithm .*

GJCST Classification: *H.2.8*

Strictly as per the compliance and regulations of:

Improving Academic Performance of Students of Defence University Based on Data Warehousing and Data mining

Dr. Vuda Sreenivasarao[α], Capt. Genetu Yohannes[Ω]

Abstract - The student academic performance in Defence University College is of great concern to the higher technical education managements, where several factors may affect the performance. The student academic performance in engineering during their first year at university is a turning point in their educational path and usually encroaches on their general point average in a decisive manner. The students evaluation factors like class quizzes mid and final exam assignment are studied. It is recommended that all these correlated information should be conveyed to the class teacher before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant level and improve the performance of students. Statistics plays an important role in assessment and evaluation of performance in academics of universities need to have extensive analysis capabilities of student achievement levels in order to make appropriate academic decisions. Academic decisions will result in academic performance changes, which need to be assessed periodically and over span of time. The performance parameters chosen can be viewed at the individual student, department, school and university levels. Data mining is used to extract meaning full information and to develop significant relationships among variables stored in large data set/ data warehouse. In this paper is an attempt to using concepts of data mining like k-Means clustering, Decision tree Techniques, to help in enhancing the quality of the higher technical educational system by evaluating student data to study the main attributes that may affect the performance of student in courses.

Keywords : Data base, Data warehousing, Data mining, Academic Performance, Educational data mining , Student performance analysis and K-Means clustering algorithm .

I. Introduction

Data mining techniques have been applied in many application domains such as Banking, Fraud detection, Instruction detection and Communication. Recently the data mining techniques were used to improve and evaluate the engineering education tasks. Some authors have proposed some techniques and architectures for using data warehousing and data mining for higher technical education. Data mining is a process of extracting previously unknown, valid, potentional useful and hidden patterns from large data sets. As the amount of data stored in educational data bases in increasing rapidly. In order to get required benefits from such large data and to find hidden relationships between variables using different data mining techniques developed and used. Clustering and decision tree are most widely used techniques for future prediction. The aim of clustering is to partition students in to homogeneous groups according to their characteristics and abilities. These applications can help both instructor and student to improve the quality education. Analyze different factors effect a students learning behavior and performance during academic career using K-means clustering algorithm and decision tree in an higher educational institute. Decision tree analysis is a popular data mining technique that can be used to explain different variables like attendance ratio and grade ratio. Clustering is one of the basic techniques often used in analyzing data sets. This study makes use of cluster analysis to segment students in to groups according to their characteristics. Academic decisions may require extensive analysis of student achievement levels. Statistical data can also be used to see the results of important academic decisions. It is necessary to have measurements to make appropriate academic decisions on one hand; while on the other hand, there is a need to see the results of academic decisions by taking measurements. The decision, implementation, measurement and evaluation mechanisms work like a chain one leading to the other. Their relationship is shown in Fig.1.

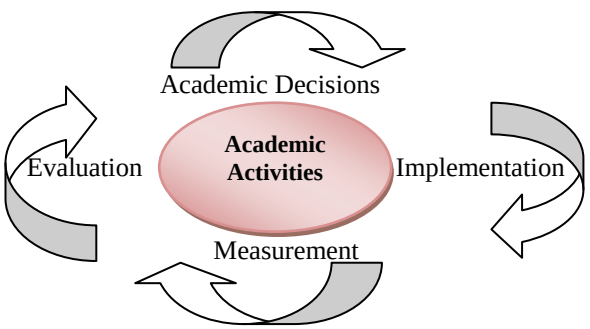


Figure 1 : Academic decision phases.

Author α : Professor, Dept of Computer & Information Technology, Defence University College, Debrezeit, ETHIOPIA.
E-mail : vudasrinivasarao@gmail.com
Author Ω : Head of Department, Dept of Computer & Information Technology, Defence University College, Debrezeit, ETHIOPIA.
E-mail : genetu77@yahoo.com

II. Related Work

a) Data Base

A data base is a collection of data usually associated with some organization or enterprise. Unlike a simple set, data in a data base are usually viewed to have a particular structure or schema with which it is associated. For example,(ID, Name, Address, Salary, Job No) may be the schema for a personal data base.

b) Data warehousing

Data warehouse is a data base devoted to analytical processing. Data warehouse to be a set of data that supports DSS and is subject-oriented, integrated, time-variant, and non-volatile. A complete repository of historical corporate data extracted from transaction systems that is available for ad-hoc access by knowledge workers. The processes of DW involve taking data from the legacy system together with corresponding transactions of the system's data base and transforming the data in to organized information in a user friendly format. The data warehouse market supports such diverse industries as manufacturing, retail, telecommunications and health care. It has access a warehouse includes traditional querying, OLAP, and data mining. Since the warehouse is stored as a data base, it can be accessed by traditional query languages.

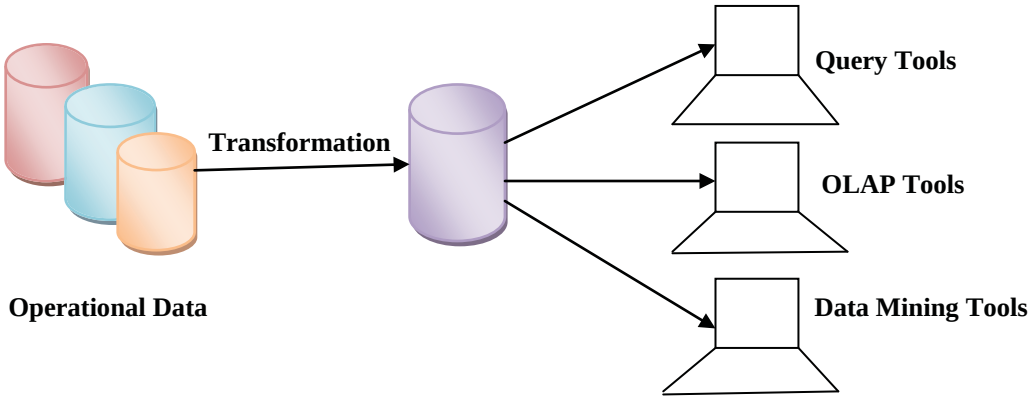


Figure 2 : Data ware house

Example of data warehousing can be defined in any of your organization. Consider the case of a Bank; a bank will typically have current accounts and saving accounts, foreign currency account etc. The bank will have an MIS system for leasing and another system for managing credit cards and another system for every different kind of business they are in . However, nowhere they have the total view of the environment from the customer's perspective. The reason being, transaction processing systems are typically designed around functional areas, within a business environment. For good decision making you should be able to integrate the data across the organization so as to cross the LoB (Line of Business) . So the idea here is to give the total view of the organization especially from a customer's perspective within the data warehouse, as shown in below figure 3.

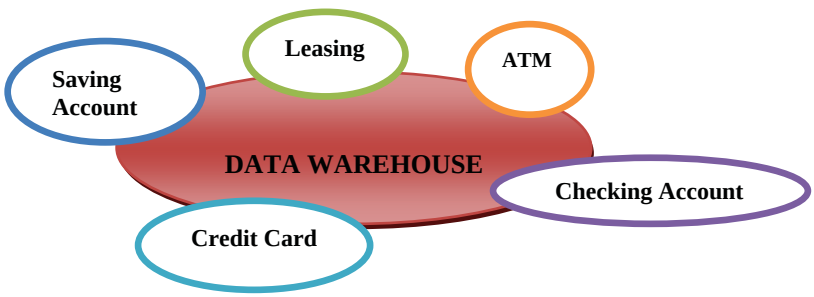


Figure 3 : A data warehouse crosses the Line of Business.

c) Data Mining

Data mining Techniques are used to extract useful and valid patterns from huge data bases. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. Large amount of data is accumulated in university students. Data mining software allow the users to analyze data from different dimensions categorize it and a summarized the relationships, identified during the mining process. Different data mining techniques are used in various fields of life such as medicine, statistical analysis, engineering, education, banking, marketing, sale etc. Data mining techniques can be differentiated by their different model functions and representation, preference criterion, and algorithms .The main function of the model that we are interested in is Classification, as normal, or malicious, or as a particular type of attack . We are also interested in link and sequence analysis. Additionally, data mining systems provide the means to easily perform data summarization and visualization, aiding the security analyst in identifying areas of concern .The models must be represented in some form. Common representations for data mining techniques include rules, decision trees, linear and non-linear functions (including neural nets), instance-based examples, and probability models.

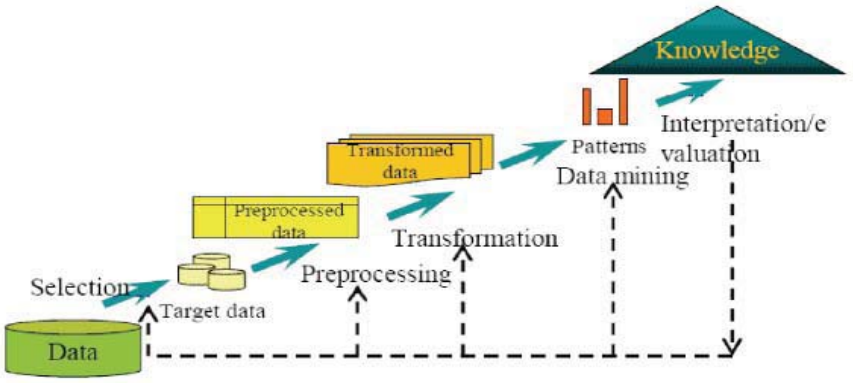


Figure 2 : The transition from raw data to valuable knowledge.

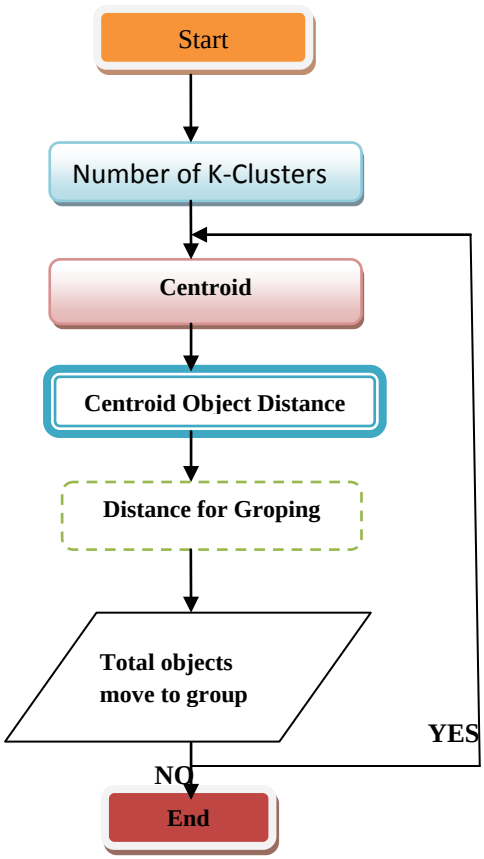
III. Clustering

Clustering is a method to group data in to classes with identical characteristics in which the similarity of intra-class is maximized or minimized. Cluster analysis used to segment a large set of data in to subsets called clusters. Each cluster is a collection of data objects that are similar to one another are placed within the same cluster but are dissimilar to objects in other clusters. A cluster of data objects can be treated collectively as one group in many applications. Cluster analysis is an important human activity. Cluster analysis has been widely used in numerous applications, including pattern recognition, data analysis, image processing, and market research. Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attributes. Current clustering techniques can be broadly classified in to three categories; partitional, hierarchical and locality-based algorithms.

Definition: Given a data base $D = \{ t_1, t_2, \text{--------} t_n\}$ of tuples and an integer value K, the clustering problem is to define a mapping **f:D** $\{1,2,3\text{--------} K\}$ where each t_i is assigned to one cluster K_j, $i \leq j \leq k$. A cluster K_j, contains precisely those tuples mapped to it; that is $K_j = \{t_i / f(t_i) = k_j, 1 \leq i \leq n$ and $t_i \in D\}$.

a) K-Means Clustering

K-Means is one of the simplest unsupervised learning algorithms used for clustering. K-means partitions "n" observations in to k clusters in which each observation belongs to the cluster with the nearest mean. This algorithm aims at minimizing an objective function, in this case a squared error function.



Flow chart: K-Means clustering.

Algorithm: K-Means Clustering.

1. **Select number of K points as the initial centroids.**
2. **Repeat.**
3. **Form K clusters by assigning all points to the nearest centroid.**
4. **Recomputed the centroid of each Cluster.**
5. **Until the centroids don't change.**

IV. Decision Tree

Decision tree induction can be integrated with data warehousing techniques for data mining. A decision tree is a predictive node ling technique used in classification, clustering, and prediction tasks. Decision tree use a "divide and conquer" technique to split the problem search space in to subsets.

A decision tree is a tree where the root and each internal node are labeled with a question. The arcs emanating from each node represent each possible answer to the associated question. Each leaf node represents a prediction of a solution to the problem under consideration.

Given a data base D = $\{t_1,t_2,$--------$t_n\}$ where t_i = $(t_{i1}$----------$t_{in})$ and the data base schema contains the following attributes {A1,A2,-------An}. Also given is a set of classes C = {C1, C2, --------Cm}. A decision tree or classification tree is a tree associated with D that has the following properties:

1. Each internal node is labeled with an attribute, Ai.
2. Each arc is labeled with a predicate that can be applied to the attribute associated with the parent.
3. Each leaf node is labeled with a class, Cj.

The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner.

Decision Tree Algorithm: generate a decision tree from the given training data

1. Create a node **N**
2. **If** samples are all of the same class, C **then**
3. Return N as a leaf node labeled with the class C;
4. **If** attribute-list is empty **then**
5. Return N as a leaf node labeled with the most common class in samples.
6. Select test-attribute, the attribute among attribute-list with the highest information gain;
7. Label node N with test-attribute;
8. **For** each known value a_i of test-attribute.
9. Grow a branch from node **N** for the condition test attribute = a_i;
10. Let Si be the set of samples for which test-attribute = a_i;
11. **If** Si is empty **then**
12. Attach a leaf labeled with the most common class in samples;
13. **Else** attach the node returned by generate-decision-tree(S_i,attribute-list-attribute);

Each internal node tests an attribute, each branch corresponds to attribute value, and each leaf node assigns a classification.

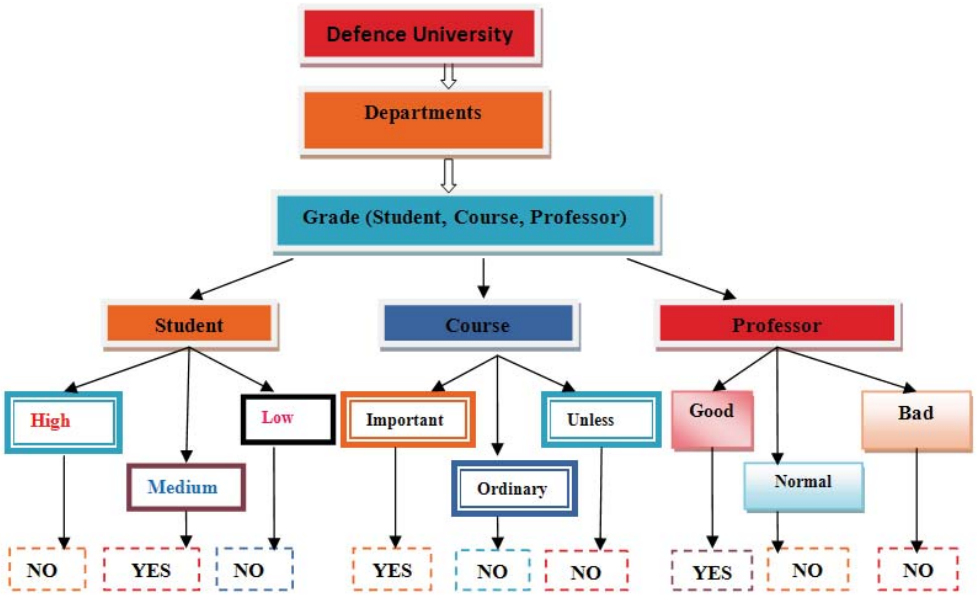


Figure 4 : Decision Tree

Table 1 shows the form of training data.1500 student score records are used for training.

Table 1 : Form of training examples.

S NO	Student	Course	Professor	Marks	Grade	Results
1	Ferede Adugna	Cryptography	Dr. Rao	86	A	YES
2	Rwibasira M	Interfacing Tech	Miss Maria	75	B	NO
3	Makuei Nyok	Operating Systems	Genetu Yohannes	85	A	YES
4	Daniel Tekif	Microprocessor	Michael	55	D	NO
5	Mesfin Dadi	Distributed Systems	Lea	95	A	YES
6	Debebe Shibeshi	Computer Networks	Genetu Yohannes	90	A	YES
7	Gidey Abrha	Network security	Dr. Srinivas	98	A	YES
8	Samuel Hagos	Web technology	Oliver	45	F	NO
9	Desta Desisa	Compiler Design	Melissa	73	B	NO
10	Tibabu Beza	Cloud computing	Praveen	50	D	NO
11	Desta Hagos	Mobile Communication	Dr.K.A.Lathiaf	70	B	NO
12	Ferede Adugna	Cryptography	Dr. Rao	86	A	YES
13	Rwibasira M	Interfacing Tech	Miss Maria	75	B	NO
14	Makuei Nyok	Operating Systems	Genetu Yohannes	85	A	YES
15	Daniel Tekif	Microprocessor	Michael	55	D	NO

Table 2 : shows Data base for Previous Semester student with effort taken from Marks taken.

Student Roll number	Marks	Effort
DEC-01	10-50	More Attention, Conducting Special Classes, Assigments, Conducting more practical classes, Daily tests and Parents and Faculty meeting.
DEC-02	51-60	Conducting Special Classes, Assignments and Conducting more practical classes.
DEC-03	61-75	Assignments and Conducting more practical classes.
DEC-04	76-85	Assignments and Conducting classes for Interviews.
DEC-05	86-100	Conducting classes for Interviews and Giving exposure for career important.

Formation of Decision tree from students marks table and applying effort depending on marks.

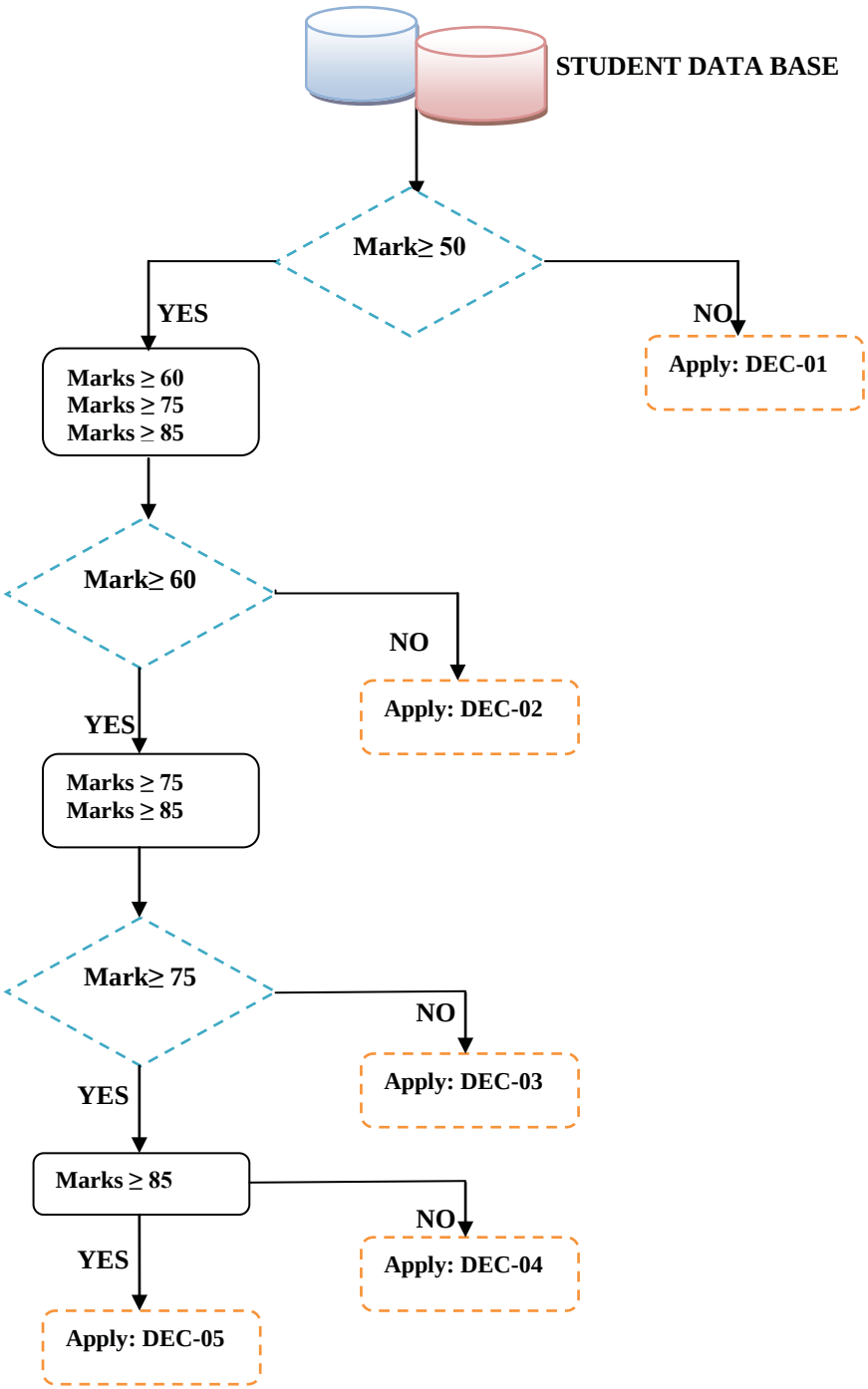


Figure 5 : Flowchart of students marks table and applying effort depending on marks.

After the pattern is classified from the decision tree we can obtain the specify knowledge discovery to form the knowledge base system. Similarly the same data mining process can be done to the professors for classifying their performance which help in improve Technical education system.

V. Results

Both K-Means clustering, Decision tree algorithm were applied on the data set.

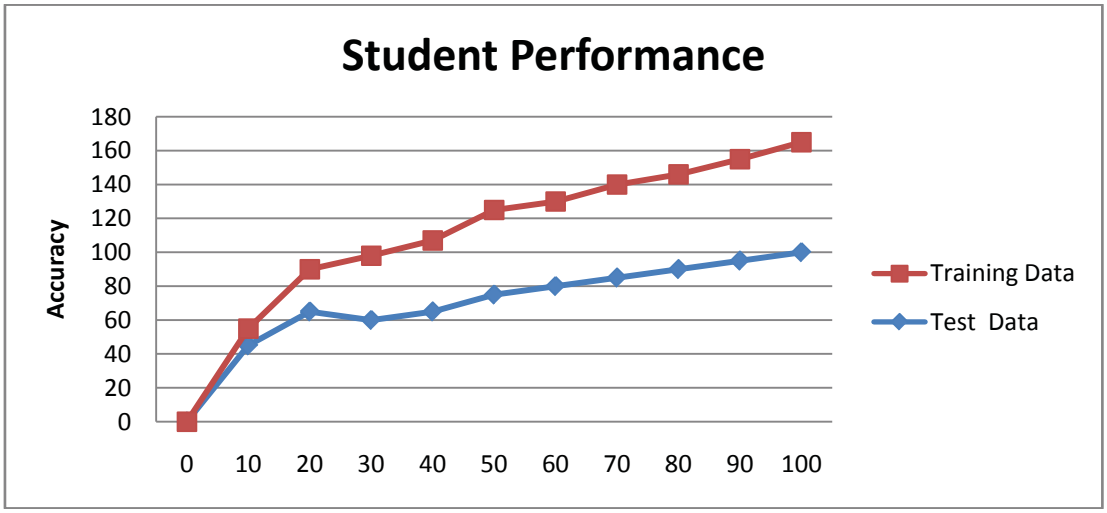


Figure 6 : Accuracy of Decision tree.

VI. Conclusion

In this study of research paper idea is a starting attempt to use data warehousing and data mining techniques to analyze and find out student academic performance and to improve the quality of the engineering system. The managements can use some techniques to improve the course outcomes according to the improve knowledge. Such knowledge can be used to give a good understanding of student's enrollment pattern in the course under study, the faculty and managerial decision maker in order to utilize the necessary steps needed to provide extra classes. Other hand, such type of knowledge the management system can be enhance their policies , improve their strategies and improve the quality of the system.

References Références Referencias

1. Vuda sreenivasarao, Dr.S.Vidyavathi, G.Ramaswamy,and Sk.Shabber, " A Research on result oriented learning process from university students based on distributed data mining and decision tree algorithm", Journal of Advance research in Computer Engineering, (Vol.4 No.2 2010), pp- 223-226 .
2. P.V.Subbareddy and Vuda Sreenivasarao, "The result oriented process for process for students based on distributed data mining", International journal of advanced computer science and applications, Vol. 1, No.5.Nov-2010, pp-22-25.
3. N.V Anand Kumar, G.V.Uma "Improving Academic Performance of Students by Applying Data Mining Technique" European Journal of Scientific Research, Volume 34, Issue 4, pp-526-534.
4. Chady EL moucary,Maric khair and Walid Zakhem " Improving student's performance using data clustering and neural networks in Foreign Language based higher education", The research bulletin of Jordan ACM, Vol 11(III),PP-27-34.
5. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza sattar and M.Inayat khan, " Data mining model for higher education system", European Journal of Scientific Research,Vol.43 No.1,pp-24-29.
6. Dervis.Z.Deniz and I brahim ersam, "An academic decision-support system based on academic performance evaluation for student and program assessment", Int Journal Ed, Vol 18, No 2, pp.236-244.
7. Bindiya M varghese, Jose Tome J, Unnikrishna A and Poulose Jacob K, "Clustering student data to characterize performance patterns", International journal of Advanced computer and applications, Special Issue,pp-138-140.
8. Data Mining: Concepts and Techniques, Han J Kamber M, Morgan Kaufmann Publishers, 2001.
9. Introduction to DATA MINING, Tan P., Steinbach M., Kumar V, Pearson Education, 2006.
10. A Machine Learning Algorithms in Java, Witten I. Frank E. WEKA, Morgan Kaufmann Publishers, 2000.
11. Hellenic Open University, Patras, Greece, Analyzing student performance in distance learning with genetic algorithms and decision trees,. Kalles D., Pierrakeas C, 2004.
12. Mining student's data to analyze E-learning behavior: A case study-Alaa el-Halees.
13. Data mining with SQL server 2005 by Zhao Hui.Maclennan.J, Wihely publishing.Inc-2005.

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Improved Vector Median Filtering Algorithm for High Density Impulse Noise Removal in Microarray Images

By V.AnjiReddy, J.Vasudevarao
GIT, GITAM University

Abstract - The digital images are corrupted by impulse noise due to errors generated in camera sensors, analog-to-digital conversion and communication channels. Therefore it is necessary to remove impulse noise in-order to provide further processing such as edge detection, segmentation, pattern recognition etc. Filtering a noisy image, while preserving the image details is one of the most important issues in image processing. In this paper, we propose a new method for impulse noise removal in Microarray images. The proposed iterative algorithm search for the noise-free pixels within a small neighborhood. The noisy pixel is then replaced with the value estimated from the noise-free pixels. The process continues iteratively until all noisy-pixels of the noisy image are filtered. The performance of the proposed method is tested using impulse noise corrupted microarray images. The experimental results show the proposed algorithm can perform significantly better in terms of noise suppression and detail preservation in microarray images than a number of existing nonlinear techniques.

Keywords : *Impulse Noise, Vector Median Filter, Noise Removal, Image Processing, Microarray.*

GJCST Classification: *I.4.1,I.4.3*

Strictly as per the compliance and regulations of:

Improved Vector Median Filtering Algorithm for High Density Impulse Noise Removal in Microarray Images

V.AnjiReddy [α], J.Vasudevarao [Ω]

Abstract - The digital images are corrupted by impulse noise due to errors generated in camera sensors, analog-to-digital conversion and communication channels. Therefore it is necessary to remove impulse noise in-order to provide further processing such as edge detection, segmentation, pattern recognition etc. Filtering a noisy image, while preserving the image details is one of the most important issues in image processing. In this paper, we propose a new method for impulse noise removal in Microarray images. The proposed iterative algorithm search for the noise-free pixels within a small neighborhood. The noisy pixel is then replaced with the value estimated from the noise-free pixels. The process continues iteratively until all noisy-pixels of the noisy image are filtered. The performance of the proposed method is tested using impulse noise corrupted microarray images. The experimental results show the proposed algorithm can perform significantly better in terms of noise suppression and detail preservation in microarray images than a number of existing nonlinear techniques.

Keywords : Impulse Noise, Vector Median Filter, Noise Removal, Image Processing, Microarray.

I. Introduction

Microarrays, widely recognized as the next revolution in molecular biology, enable scientists to analyze genes, proteins and other biological molecules on a genomic scale [1]. A microarray is a collection of spots containing DNA deposited on the solid surface of glass slide. Each of the spot contains multiple copies of single DNA sequence [2].

Microarray expression technology helps in the monitoring of gene expression for tens and thousands of genes in parallel. During the biological experiment, the mRNA of two biological tissues of interest is extracted and purified. Each of the mRNA samples are reverse transcribed into complementary DNA (cDNA) copy and labeled with two different fluorescent dyes resulting in two fluorescence-tagged cDNA (red Cy5 and green Cy3). The tagged cDNA copies, called the sample probe, are hybridized with the slide's DNA spots. The hybridized glass slides are fluorescently scanned at different wavelengths (corresponding to the different dyes used), and two digital images are produced, one for each population of mRNA. Each digital image contains a number of spots of various fluorescence intensities. The intensity of each spot is proportional to the hybridization level of the cDNAs and the DNA dots, the gene expression information is obtained by analyzing the digital images [3] [4].

The processing of the microarray images usually consists of the following three steps: (i) gridding, which is the process of assigning the location of each spot in the image. (ii) Segmentation, which is the process of grouping the pixels with similar features and (iii) Intensity extraction, which calculates red and green foreground intensity pairs and background intensities.

The evaluation of microarray images is a difficult task as the fluorescence of the glass slide adds noise floor to the microarray image. The processing of the microarray image requires noise suppression with minimal reduction of spot edge information that derives the segmentation process. Thus the task of microarray image enhancement is of paramount importance.

Non-linear filters exhibit better performance as compared to linear filters [5] when restoring images corrupted by impulse noise. Filtering techniques such as Vector Median Filter (VMF) [6], Progressive Switching Median Filter (PSMF) [7], Decision Based Algorithm (DBA) [8] etc., have been developed for removal of impulse noise. These techniques estimate noisy pixels taking into account all pixels within the window, without considering the status of (noisy/ noise-free) pixels. Consequently, the estimated noisy pixel value will not be accurate, degrading the quality of restored image.

In this paper, we proposed a new iterative algorithm for removal of impulse noise in Microarray images. The algorithm emphasis on the noise-free pixels within small neighborhood. First the pixels affected with noise are detected. If we did not find certain number of noise-free pixels within neighborhood, then the central pixel is left unchanged. Otherwise the noisy pixel is replaced with the value estimated form the noise-free pixels within neighborhood. The process iterates until all noisy pixels are estimated in the image. The main steps of the proposed filtering algorithm are shown in figure 1. The rest of the paper is organized as follows:

Section 2 presents the impulse noise models in digital images, Section 3 presents the proposed iterative algorithm, Section 4 presents the experimental results and finally Section 5 reports conclusion.

Author [α] : M.Tech Student, Department of CSE,, GMRIT, Rajam
Author [Ω] : Associate Professor, Department of CSE, GMRIT, Rajam

II. Impulse Noise In Digital Images

Impulse noise is independent and uncorrelated to the image pixels and is randomly distributed over the image. 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 two types of impulse noise namely fixed value impulse noise and random valued impulse noise.

In this paper, we focus on the detection and denoising of fixed valued impulse noise, namely salt and pepper 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 [9]. Consider a corrupted image Y of size NxM, which containing the salt and pepper noise with probability p is mathematically represented in the form:

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

Where i=1,2,….,M and j=1,2,…..N and 0<p<1. y_{ij} represents the intensity of the pixel located at position (i, j). x_{ij} and n_{ij} denote the intensity of the pixel (i, j) in the original image and the noisy image respectively.

III. The Proposed Algorithm

The proposed algorithm is divided into three stages.

Stage 1: Construction Of Binary Image

In this step, a binary image is constructed for the noisy image Y. When the gray level images is contaminated with salt-and-pepper noise, a noisy pixel takes either a maximum intensity value (I_{max}=255) or a minimum intensity value (I_{min}= 0). This dynamic range [Imax Imin] provide information about the noisy pixels in the image. The binary image b_{ij} is constructed by assigning a binary value 1, if the intensity of the pixel located at position (i, j) in the noisy image is I_{max} or I_{min}, otherwise assign a binary value 0.

The binary image B is computed from the noisy image Y as follows:

$$b_{ij} = \begin{cases} 1, \text{ if } y_{ij} = I_{max} \\ 1, \text{ if } y_{ij} = I_{min} \\ 0, \text{ otherwise} \end{cases} \quad (2)$$

where i=1,2,……N and j=1,2,……M. The entries of "1" and "0" in the binary image B represent the noisy and noise-free pixels, respectively. This binary image provides information about the noisy density in the corrupted image, which is used in the filtering process.

The Noise Density of the corrupted image is calculated as follows:

$$ND=\frac{\text{Number of } 1's \text{ in binary image}}{\text{Total number of pixels (NxM)}} \quad (3)$$

The value of the noise density (ND) ranges between 0 and 1.

Stage 2: Impulse Noise Filtering Method

Consider a window of size q x q at each pixel location (i, j) of the noisy image Y and the binary image B. We prefer to use the value of q (=3), because the larger size window may not be too efficient and effective. Larger window may also remove the edges and fine image details. By applying small window of size 3x3, we obtain the noisy image patch $Y_{i,j}$ and the binary image path $B_{i,j}$.

For each iteration, we count the number of noisy pixels in the binary map B. If the value of count K is a positive integer and the central pixel y_{ij} within the 3X3 window is noisy, then the array R is populated with noise-free pixels. The maximum length of the array R is eight, indicating all the pixels are noise free. The minimum length is zero, shows that all the pixels in the window are noisy. Depending upon the noisy density in the window, the length of the array varies from zero to eight.

We emphasize a constraint of minimum three noise-free pixels within the window, ie., the minimum length of the array R should be three. If this condition is satisfied, then we replace the central noisy pixel with the estimated value ie.,

$$g_{ij}=\begin{cases} e_s, \text{ if } b_{ij}=1 \text{ \&\& Length(R)} \geq 3 \\ y_{ij}\text{ , otherwise.} \end{cases} \quad (4)$$

Where e_s is the estimated value of the noisy pixel. Currently, we estimated the value of noisy pixels by using a suitable distance measure. The elements (noise-free pixels) in the array R are ordered on the basis of the sum of distances between each element and other elements in the array R. The sum of distances is arranged in ascending order and the same ordering is associated with the elements in the array R. The element in the array with the smallest sum of distances is the estimated value of the noisy pixel. If d_i is the sum of distances of the i^{th} element in the array R with all other elements, then

$$d_i=\sum_{j=1}^{N} \Delta(X_i, X_j) \quad (5)$$

where $1 \leq i \leq N$, X_i, X_j are the elements in the array, N is the length of the array R, and $\Delta(X_i, X_j)$, is the distance measure given by L_1 norm.

The ordering may be illustrated as

$$d_1 \leq d_2 \leq d_{3,\ldots\ldots\ldots\ldots\ldots}, d_N \quad (6)$$

and this implies the same ordering to the corresponding elements in the array R.

$$X_{(1)} \leq X_{(2)} \leq, \ldots\ldots\ldots\ldots, \leq X_{(N)}. \quad (7)$$

Where the subscripts are the ranks. Since the element with the smallest distance is the estimated noisy pixel, it will correspond to rank 1 of the ordered elements ie., $X_{(1)}$. Figure 1 shows the main steps of the proposed algorithm.

Stage 3: Update Noisy Image And Binary Image.

If the noisy pixel is estimated from the noise-free pixels within the window, the binary image B is also updated by changing the entries at the corresponding location of the image from "1" to "0". At the end of each iteration, we obtain a refined image G and updated binary image B. After a few iterations, depending upon the intensity of the salt-and-pepper noise, all entries in the binary image becomes zeros. The updating process terminated and we obtain a restored image G.

1. Take the intial noisy image Y.
2. Computation of binary map B
3. Compute the value of K that represent the noise-free
   pixels in B and assign Y→X
4. Check: If K=0, output resorted image X and stop.

else

i. Check if y_{ij} is noisy, then do
ii. Fill the array R with noise-free pixels
iii. Check if length of array R >3 , do
iv. Update b_{ij} and xij using the value estimated from noise-free pixels in R.
v. Process each y_{ij} and get updated X and B
vi. For the next iteration; assign X→Y and go to step 3.

Fig 1 : Proposed Algorithm

IV. Experimental Results

Noise removal steps of the microarray image are performed on a sample microarray slide that has 48 blocks, each block consisting of 110 spots. A sample block has been chosen and 108 spots of the block have been cropped for simplicity. The sample image is a 154*200 pixel image that consists of a total of 30800 pixels. The RGB colored image microarray image have been converted to grayscale image to specify a single intensity value that varies from the darkest (0) to the brightest (255) for each pixel shown in figure 2.

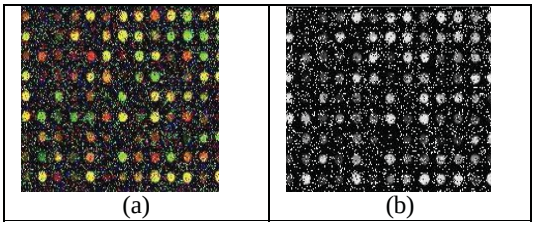


Fig2 : a) RGB Color microarray image b) Grayscale Image

First the microarray image is corrupted with varying levels of noise density from 10 to 90 using the salt-and-pepper noise. The simulation results obtained from the proposed scheme are compared with the well known salt-and-pepper filtering algorithms: AMF, PSMF and DBA. Figure 3 shows the results.

We used the image quality metric, peak signal-tonoise ratio (PSNR), to measure the quality of the restored image. The PSNR measure is defined as

$$\text{PSNR} = 10 \log 10 \frac{(255)^2}{\text{MSE}}$$

Where MSE is the mean squared error between the original noise-free image and the restored image.

Table 1 shows the simulation results, in terms of the PSNR measure, for the microarray image. In table 1, our proposed algorithm provided the best PSNR value. Among other restoration algorithms, our proposed scheme highlights the best visible quality of the restored microarray image.

V. Conclusion

In this work, we propose an iterative algorithm for removal of impulse noise in microarray images. The proposed scheme works iteratively by replacing the noisy pixel with the value estimated from the noise-free pixels within the small neighborhood for the entire image. This scheme provides superior performance in removing the noise, while preserving the fine image details and edges. The proposed algorithm provides noise suppression in the microarray image with minimal reduction of spot edge information that derives the segmentation process.

References References Referencias

1. M.Schena, D.Shalon, Ronald W.davis and Patrick O.Brown, "Quantitative Monitoring of gene expression patterns with a complementary DNA microarray", Science, 270,199,pp:467-470.
2. Wei-Bang Chen, Chengcui Zhang and Wen-Lin Liu, "An Automated Gridding and Segmentation method for cDNA Microarray Image Analysis", 19th IEEE Symposium on Computer-Based Medical Systems.
3. Tsung-Han Tsai Chein-Po Yang, Wei-ChiTsai, Pin-Hua Chen, "Error Reduction on Automatic Segmentation in Microarray Image", IEEE 2007.

4. S.Indu , Chaveli Ramesh, " A noise fading technique for images highly corrupted with impulse noise", Proceedings of the ICCTA07 , IEEE.
5. Volkan Uslan, Ihsan Omur Bucak, Clustering based Spot Segmentation of cDNA Microarray Images, IEEE 2010.
6. James C. Church, Yixin Chen, and Stephen V. Rice , "A Spatial Median Filter for Noise Removal in Digital Images", 2008 IEEE.
7. W.Zhou and D.Zhang, " Progesssive Switching median filter for the removal of impulse noise from the highly corrupted images", IEEE transactions on circuits and systems II. Analog and Digital Signal Processing, Vol 46, pp.78-80,1999.
8. K.S.Srinivasan, and D.Ebnezer," A New Fast and Efficient Decision based Algorithm for Removal of High Density Impulse Noises", IEEE Signal Processing Letters, Vol14 pp.189-192,2007.
9. Dagao Duan, Qian Mo," A Detail Preserving Filter for Impulse Noise Removal" ,Proceedings of ICCASM, IEEE 2010.

Table 1 : Performance Comparison for The Microarray Image Corrupted With Various Noise Densities

ND(in percentage)	VMF	PSMF	DBA	Proposed
10%	33.74	36.41	38.24	43.64
20%	28.47	30.27	32.44	37.47
30%	23.03	26.23	29.03	33.83
40%	18.15	20.25	21.15	26.62
50%	14.36	15.21	17.36	24.36
60%	11.61	13.06	15.61	21.61
70%	9.08	11.06	13.48	19.28
80%	7.16	9.46	12.06	13.16

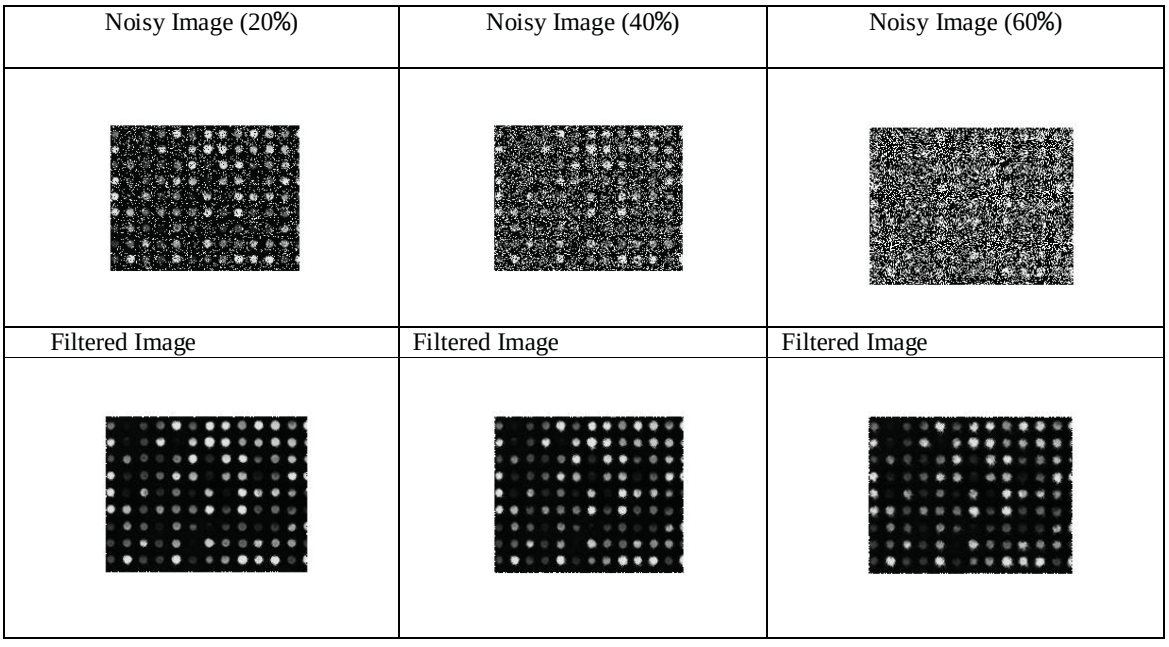


Fig3 : Experimental results of proposed algorithm on Microarray image

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Data Center Establishment to Run the IT System in Power Utilities

By Utkarsh Seetha, Rajesh Kumar Rolen & Hitesh Babu Sharma
Singhania University

Abstract - In this research we will focus the details of the IT System and business process requirements of IT Package need to be installed at Data Centers. This research details the project requirements, which are to be met by the applications and interfaces required within Data Center between different hardware and software systems. The objective of this research includes the design and development of Data Center architecture, hardware availability, proper installation and commissioning of all related networking equipments, storage devices and high end servers as per the current international standards.

Keywords : *LAN, WAN, MAN, IPS, Switch, VLAN, DC, DR, ACL, SCR, SCC, SaaS, DTMF, SAN.*

GJCST Classification: *B.4.1 , B.4.3*

Strictly as per the compliance and regulations of:

Data Center Establishment to Run the IT System in Power Utilities

Utkarsh Seetha, Rajesh Kumar Rolen & Hitesh Babu Sharma

Objective - In this research we will focus the details of the IT System and business process requirements of IT Package need to be installed at Data Centers. This research details the project requirements, which are to be met by the applications and interfaces required within Data Center between different hardware and software systems. The objective of this research includes the design and development of Data Center architecture, hardware availability, proper installation and commissioning of all related networking equipments, storage devices and high end servers as per the current international standards.

Keyword : LAN, WAN, MAN, IPS, Switch, VLAN, DC, DR, ACL, SCR, SCC, SaaS, DTMF, SAN.

I. Research Scope for Data Centre

The data centre project scope shall include the deliverables as mentioned in the document. The brief descriptions of the project scope have been described below in this document. HCL Scope as per the requirements given in the tender document as follows:

- Data Cabling
- Computing Setup with DR Solution
- WAN Connectivity

II. Overall Solution Description

The overall solution describes the required hardware need to be installed to run the applications/modules for the purpose of the Subdivision Automation of State Electricity Departments in respect to the business functionality. This solution covers the functionality as mentioned and required in the Document as this is an advanced engineered office management tool. It is developed to manage all types of useful databases, analyzes them by applying standard concepts and implement them in a manner consistent with its purpose or design the logic of electrical engineering and subdivision level management in a modernize way.

After a deep study of RAPDRP requirements and the difficulties of DISCOMs, our research has suggested the solution with additional amenities. DISCOMs related business functionality would be customized in the product on the base of the Document.

The Data Center solution has the capability to integrate with other Business Process Applications as per the requirement captured in system study and suggested by Document. The integration architecture of Data Center solution is based on SOA (Service Oriented Architecture) and due to this it is easily mapped with the integration middleware for exposing the business functionality to external systems as well as to consuming the business functionality of external systems and other future needs which will be fulfilled by the installed hardware, networking equipments and storage devices for coming decades.

Features:

- Flawless Integration of Functions and Processes
- Increase Operation Efficiency
- Process Streamlining
- Enhance Customer Service
- Business Assessment Support for Strategic Issues
- Revenue Augmentation
- Scalability, Flexibility for future application integration

III. Basic Hardware Design Considerations

In this research we will study the Basic Design Principles for the data centre and disaster recovery solution for APDRP which has been arrived at in conformance with the RFP Guidelines, Industry Best Practices, Critical Nature of the Centralized application and our experience in handling such large data centre and Multi-location projects.

We have considered the following key design considerations for architecting the Data centre -

- Scalability
- Availability
- Reliability
- Performance
- Security

a) Scalability

In Utility segment the no of consumers are bound to grow with the increase in population and the usage of IT. Also as the various facilities are extended to new geographical areas the no. of offices / no. of consumers are bound to grow. This

Author : Restructured Power Development and Reforms Programme, Rajasthan.

necessitates the scalability requirement at the Design level to ensure that there is enough room for growth to meet the future requirements.

To meet the Scalability requirements Horizontal scalability, vertical scalability or a combination of both can be used. As per the RFP guidelines we have both horizontal and vertical scalability based on the product set and the application for which it is being used.

b) Availability

This is a one of the most important design objective for the datacenter set-up, especially when all the applications are centralized in nature. Highly available Datacenter design ensures that the end users are able to operate and access the applications at all times with desired response time.

To meet this objective we have taken care of redundancies at all levels and the choice of technology is to ensure that the applications are made available to the users in any event of failure or disaster.

There are two strategies that we have followed in order to meet the above design objectives –

- High Availability within the datacenter.
  - High Availability at DC Core Network level
  - Application Load Balancing for all Web and Application Servers
  - Clustering and failover for all Database servers
- Application Availability in case of DC Site Failure
  - 100% replica of the DC site so that there is no impact on the response time of the application.
  - Continuous Replication of data to DR Site for disaster recovery
  - Transparent failover of users to the DR site in case of a Primary site failure

c) Reliability

Product reliability is a very important design criterion while designing various components of the Datacenter. Some of the key aspects which have a direct implication on product reliability are as follows:

- Provision for Redundant Power Supplies
- Provision for Redundant Cooling Fans
- Mirrored Hard Disk Drives
- Redundant Ethernet Connectivity
- Redundant Storage Connectivity
- ECC Memory

Each of the elements contributes towards improving the MTBF of the System / Component. To achieve high reliability we have offered the best in class products which passes through stringent test conditions as well as provides component level redundancy wherever needed.

d) Performance

We also understand that the systems should be designed keeping in view the performance requirement so that the desired response is met at all times. This requires capacity planning at all level to meets the performance availability criteria. Some key aspects which have been taken care in the design to meets the performance requirements are:

- All the Servers have been sized keeping in mind the appropriate concurrency at User Level
- The Servers are being load balanced to ensure that there is no performance choke due to excess load on one server
- The Interconnectivity between servers are on high performance network with Non-Blocking Architecture
- The Bandwidth Capacity planning is done keeping in mind the appropriate user load.

e) Security

It's an inevitable fact that security is required at all levels to ensure that the application is available to the users as well as the data is protected from any kind of unauthorized access. In today's environment where the users access the Datacenter from both Intranet as well as Internet it becomes a challenge to ensure that we protect the datacenter from any damage due to different attacks. We have taken care of the following design considerations on the Infrastructure side to ensure end to end security from edge location to the Datacenter -

- End Point Security – Provision of Antivirus and Antispyware.
- Gateway level Security – Multi Layered Antispam and Antivirus for Mail Messaging; AV, Content filtering and URL Filtering for Web.
- Transmission Security – IPSec for edge to DC encryption.
- Datacenter Security –
  - Perimeter security - Intrusion Prevention, Firewall and extended ACLs on required VLANs.
  - Identity and Access Management for user authentication, authorization and accounting
  - Security at OS Level by OS Hardening

IV. Basic Components of the Project

As per our research and available international standers, the overall data centre project will include the following modules:

a) Active Directory Implementation

- Validating the proper installation Windows 2008 Operating System as per best practices and basic OS hardening to be configured in New AD servers
- Design & Implementation of Active Directory Infrastructure for APDRP
- Active Directory Integrated DNS will be configured.
- Creation of users and their mailboxes (email addresses) as per defined naming convention by customer
- Configuration required for Integration with IAM (if any)
- Testing AD functionality and GPOs

b) Enterprise Messaging Setup on Exchange 2007 - DC & DR

- Validating the proper installation Windows 2008 Operating System as per best practices and basic OS hardening to be configured
- Design & Implementation of Exchange Server for Enterprise wide messaging setup
- Extending Exchange 2007 attributes in AD Schema
- Installing and configuring Exchange 2007 Active/Passive cluster
- Configuring Exchange Single Copy Cluster Replication (SCC) for primary site and Standby Cluster Replication (SCR) from Primary site to DR site
- Enforcement of mailbox quota & mail attachment size policies
- Enforcement of delivery restrictions
- Configuration of protocols - MAPI, HTTP/HTTPS and S/MIME for access by different types of clients, although the preferred client would be Microsoft Outlook 2003/2007
- Augment email services with other collaboration tools such as calendaring, scheduling, contacts, tasks
- Configuration of Deleted item recovery for end-user mail management. Customer will define the maximum period for which a mail, which has been removed from the recycle bin, be kept for the end user to recover.
- Integrate the Exchange mailing setup with Antivirus Solution to prevent it from viruses and spam.
- The system should automatically ensure that users are properly authenticated to access their mailbox.
- Configuring the web based secure mail access for the use
- Configuring Outlook on 10 Sample User Desktops for Exchange 2007.
- Testing of mail flow, routing, failover and other Exchange 2007 features

c) ISA 2006 Server – DC & DR

- Validating the proper installation Windows 2003 Operating System as per best practices and basic OS hardening to be configured
- Integrate with Windows domain for basic authentication and users directory
- Installing and configuring ISA servers in 2 Node NLB at DC & DR sites
- Configuration of reverse proxy rules on ISA server for Exchange web access & MOSS Portal
- Testing and validating web publishing

d) MS Infra Enablement for Applications - DC, DR & 3 CC Sites

- Validating the proper installation Windows 2008 Operating System as per best practices and implementing the basic OS hardening for all servers in DC, DR Site & 3 Customer Care Sites.
- Adding all Windows servers as part of Active Directory Domain
- Backend Infra –
  - Preparing the systems for Windows cluster
  - Configuring Windows 2008 Active Passive Cluster for SQL Database usage
  - Installing and configuring SQL Failover Cluster
  - Creating the SQL Instance for Application usage
  - Test and validate failover clustering
- Front end Infra -
  - Installing and configuring the Windows Network Load Balancing for the front end application (if required)
  - Test and validate Network Load Balancing
  - Publishing the Application to AD Users
- Test & Development
  - Installing and configuring Windows Hyper-V based virtualization for testing environment (if required)
  - Installing and configuring the OS for the application usage similar to Production deployment
- Setting up the Active Directory environment for test lab

e) *Enterprise Management System (EMS)*

- The IT platform shall assist in capturing and validating the energy and revenue model to gather in a transparent manner with accuracy.
- To monitor network and server infrastructure for fault and performance issues reducing outages and interruptions by proactive monitoring of infrastructure.
- To improve IT staff efficiency by enabling process-driven management, automated actions based on business policies and rapid root-cause analysis
- To improve service availability by integrating event and performance management across all domains: systems, network, storage, database and applications
- To improve IT support management by providing a thorough, versatile set of functionality that leverages ITIL principles and other practices and improves IT governance
- To collect hardware and software inventory for network devices, servers and desktops and to deploy software packages / patches remotely.
- To Setup customer care center in the towns along with supply, installation, testing and commissioning of all necessary hardware, software and managing the facilities.
- To setup Data center & Disaster Recovery center at identified location and set up the Local Area Network and Wide Area Network. Data from primary site to be replicated to DR site in a synchronized mode.

f) *Identity and Access Management Systems*

- Installing and configuring the IM, SSO and Site Minder
- Configuring the IM for enterprise identity of users
- Deploying the agents for the web server of the application
- Deploying the agents for the desktop clients
- Validating the functionality of the deployed software's

g) *Customer care center solution*

- Unified Administration – manage inbound, outbound, email, workflow and web interactions
- Unified Routing – apply unified routing strategies across your contact center
- Unified Reporting – eliminate the need to integrate reporting data from multiple point solution data sources
- Automatic Call Distribution – answer calls and intelligently route them to available agents based on the customer profile, service level goals and agent availability
- Predictive Dialing – leverage the capabilities needed to make your outbound collections, sales or telemarketing strategies successful
- Voice Portal – deliver rich voice self-service applications to your customers via standard speech enabled Voice XML or dual tone multi frequency (DTMF) applications
- Web Interaction Management – offer assisted service to customers who visit your website
- Email Management – efficiently handle the volume of email messages by providing service levels, prioritization, queuing, auto acknowledgement, auto response and reporting
- Knowledge Base – manage a repository of frequently asked questions (FAQs) and empower your agents with information to quickly respond to customer inquiries
- Contact Recording and Quality Management – whether for quality control or compliance purposes, Aspect Unified IP provides you with the tools needed to monitor, record, score and analyze the performance of your contact center
- Choice of Transport – empower IT to select their transport of choice using either open source voice over IP (VoIP), such as the Asterisk® IP-PBX, closed source VoIP or traditional voice. Migrate from traditional switching technologies to SIP-based VoIP, single-site to virtual contact center, centralized to localized management or any combination thereof
- Multi-Tenancy – take advantage of the secure partitioning and SIP-based VoIP applications that support software as a service (SaaS) models to provide hosted contact center functionality to internal and external customers

h) *Antivirus*

- Installing the Antivirus server
- Configuring the Antivirus policies
- Deployment of anti-virus clients on all the servers and desktops
- Installing and validating gateway anti-virus for SMTP
- Installing and validating gateway anti-virus for HTTP

i) *Storage and Backup*

- Configuration of Storage and the SAN switches
- Creation of storage group/LUNs for the servers
- Configuring the storage replication
- Installing the backup software
- Deploying the backup clients
- Configuring the tape library
- Configuring the backup policy
- Validating the backup and restore process

j) Network and Security

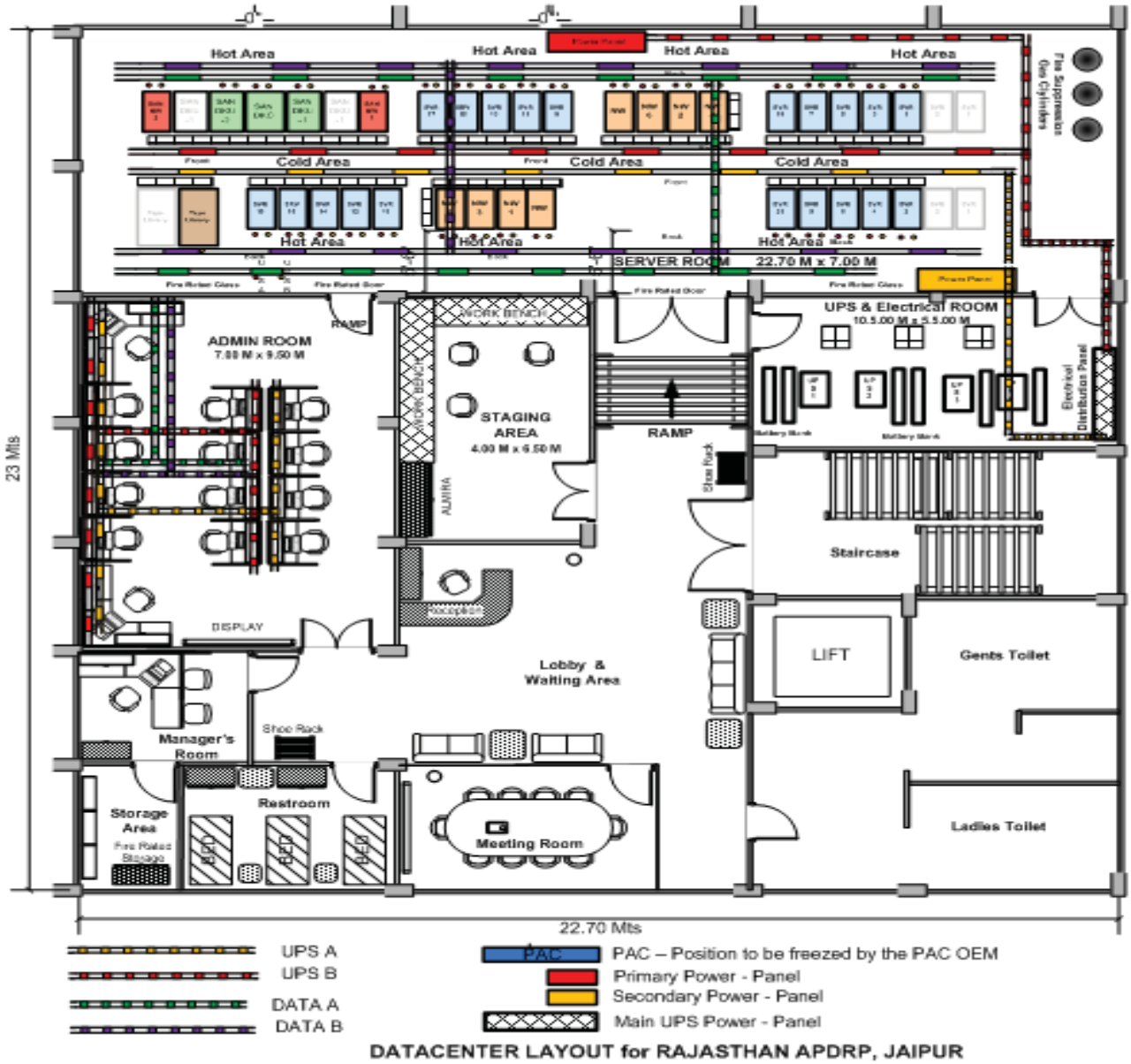
- Creating the LAN at DC, DR, CCC and other utility offices
- Creating the WAN for the DC, DR, CCC and other utility offices
- Creation of VLAN
- Creation of firewall policies
- Creation of IPS policies
- Configuring the load balancer
- Validating the functionality

V. Data Center Proposed Solution Architecture

The proposed solution diagram for datacenter has been defined below as per the scope of work.

a) Overall layout for data center

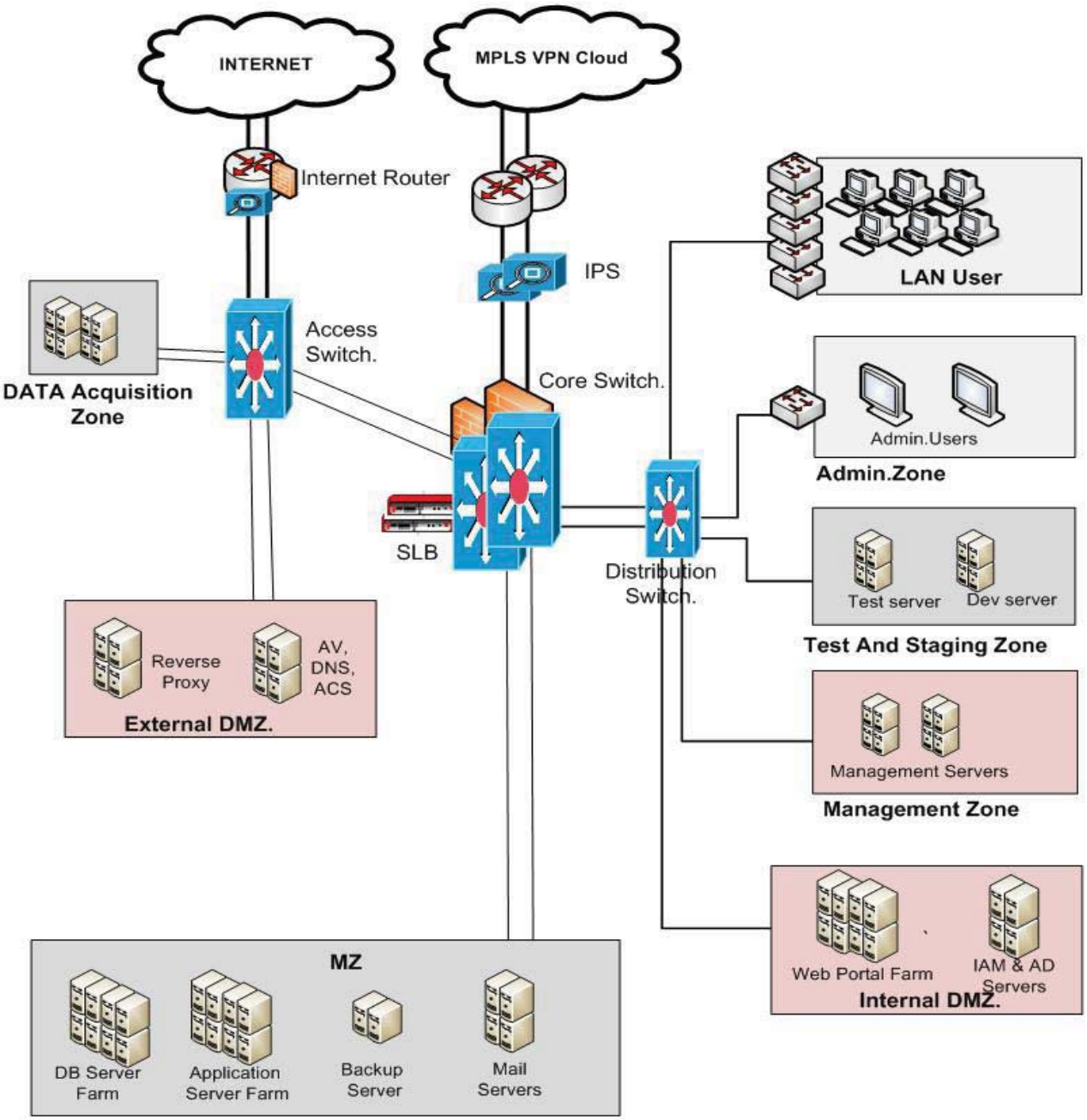
DATA Centre Floor layout



DATACENTER LAYOUT for RAJASTHAN APDRP, JAIPUR

b) Overall Architecture of the datacenter

The diagram below gives an overall snapshot of the datacenter architecture –



During Research it has been found that the DC layout will been divided into the following Zones from datacenter security perspective –

- **Meter Data Acquisition Zone** – This Zone Comprises of Servers required for Meter Data acquisition.
- **External De-Militarized Zone** – This Zone Comprises of Reverse Proxy, Antivirus, External DNS Servers, SMTP & HTTP Gateway, and Access Control server.
- **Internal De-Militarized Zone** – This Zone will comprise of Web Portal Farm, IAM servers and Active Directory Servers.
- **Militarized or Trusted Zone** – This zone comprises of the application, database, Backup, Mail DB, Integration Servers, DWH and BI servers and Storage infrastructure
- **Test and Staging Zone** – This Zone will host the Test and Staging servers. This zone will be created using the firewall Blade given in Core

Switch or using Extended ACLs feature as per the need basis.

- **Management Zone** – This Zone will comprise of Management Servers. We have created a separate Management Zone as per Industry best practices. This zone will be created using the firewall Blade given in Core Switch or using Extended ACLs feature as per the need basis.
- **Administration Zone** – One zone will be created for the Administrative users of the Data Center.
- **LAN Users** - One zone will be created for the LAN users of the Data Center.

References Références Referencias

1. R-APDRP, Power Finance Corporation of India – 2009-10.
2. Ministry of Power, Govt. of India – 2009-10.
3. JVVNL, Jaipur, Rajasthan – 2009-10.
4. ITIA, RAPDRP – 2010.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Enhanced Approach for Compress Transaction Databases

By I.Elizabeth shanthi, v.vidhya rani
Avinashilingam Deemed University For Women,Coimabtore

Abstract - Associative rule mining is defined as the task that deals with the extraction of hidden knowledge and frequent patterns from very large databases. Traditional associative mining processes are iterative, time consuming and storage expensive. To solve these processes, a way of representation that reduces this size and at the same time maintains all the important and relevant data needed to extract the desired knowledge from transaction databases is needed. This paper proposes a method that merges the transactions in the transaction database and uses FP-Growth algorithm for mining associative knowledge is presented. The experimental results in terms of compression ratio, both in terms of storage required and number of transactions, prove that the proposed algorithm is an improved version to the existing systems.

Keywords : *Associative Rule Mining, Compact Transactional Database, FP-Growth, FP-Tree, Frequent Pattern Generation, Merge Transactions.*

GJCST Classification: *H.2.8,H.2.4*

Strictly as per the compliance and regulations of:

An Enhanced Approach for Compress Transaction Databases

I.Elizabeth shanthi [α], v.vidhya rani [Ω]

Abstract - Associative rule mining is defined as the task that deals with the extraction of hidden knowledge and frequent patterns from very large databases. Traditional associative mining processes are iterative, time consuming and storage expensive. To solve these processes, a way of representation that reduces this size and at the same time maintains all the important and relevant data needed to extract the desired knowledge from transaction databases is needed. This paper proposes a method that merges the transactions in the transaction database and uses FP-Growth algorithm for mining associative knowledge is presented. The experimental results in terms of compression ratio, both in terms of storage required and number of transactions, prove that the proposed algorithm is an improved version to the existing systems.

Keywords : Associative Rule Mining, Compact Transactional Database, FP-Growth, FP-Tree, Frequent Pattern Generation, Merge Transactions.

I. Introduction

The beginning of the twenty first century has brought considerable advances in the field of computer-based information retrieval systems, where data with "hidden asset" called "knowledge" is quickly becoming the most valuable resource. Associative rule mining (Park *et al.*, 1995; Agrawal *et al.*, 1996) is defined as the task that deals with the extraction of hidden knowledge and frequent patterns from very large databases. Mining frequent patterns has become a focused topic in data mining research with the development of numerous interesting algorithms for mining association, partial periodicity, constraint-based frequent mining, associative classification and emerging patterns.

The popular area of application is the market basket analysis, which studies the buying habits of customers by searching for sets of items that frequently appear together. Associations among items of the same transaction lead to correlation and identification of frequent itemsets (Gionis *et al.*, 2007). Traditional associative mining processes are iterative, time consuming and storage expensive. Once the database becomes huge, it increases the number of Input/Output (I/O) scans and also reduces space complexity. In order to increase the performance of these applications, methods that can present data in a compact form is becoming crucial. The current need is to develop way of representation which reduces this size so that it can reside in the main memory and at the same time maintains all the important and relevant data needed to extract the desired knowledge from transactional database.

This paper analyzes algorithms that produce compact databases for knowledge discovery from large transaction databases like market basket database and web log databases. From these compact representations, association rule mining is applied to mine frequent patterns. In order to obtain a compact representation of the database Dai *et al.* (2008) proposed an algorithm called M^{2}TQT which uses a 'Merge Transactions Scheme (MTS)' to reduce storage requirement during analysis. The algorithm is efficient in two manners, namely (i) reduces database size and prunes irrelevant transactions, which saves time and (ii) Reduces the I/O time required. However, it has disadvantages namely (i) Although some rules can be mined from the new transactions, it still needs to scan the database again to verify the result. This is because the data mining step produces potentially ambiguous results. (ii) The compressed database is reversible to its original form (iii) It has the serious problem of scanning the database multiple times, which results in high cost of re-checking the frequent itemsets and (iv) Processing time is still high when compared to uncompressed mining algorithms. To solve the above difficulties, this work enhances the M^{2}TQT in the following manner.

- Develop an algorithm to recover the original transaction from the compressed database.
- M^{2}TQT uses apriori-like algorithm which is main culprit for the multiple database scans. The present research work proposes the use of FP-Tree algorithm to avoid multiple scans of the database

The rest of the paper is organized as follows. Section 2 discusses the various existing methods available to obtain a representation of transaction databases. Section 3 presents the proposed method. The experimental results are discussed in Section 4, while Section 5 concludes the work with future research directions.

II. Existing Methods

A compact representation of transaction database can be derived in three ways. They are (i) Compressing transaction database (ii) Transforming a transaction database to a compact representation (iii) Partitioning transactions in transaction database and (iv) Merging transactions in transaction database. This section briefly explains each of these techniques.

a) Compressing transaction database

Data compression is a technique that has been widely used to save storage requirement of transactional database in secondary devices. A simple characterization of data compression is that it involves transforming a string of characters in some representation (such as ASCII) into a new string (of bits, for example) which contains the same information but whose length is as small as possible. There are two fundamentally different types of data compression: lossless and lossy. A lossy compression technique removes unwanted data which eventhough degrades the output quality maintains overall important information. Lossless compression technique, on the other hand, attempts to compress data while retaining all information. Examples include, GZIP, BZIP, BOA and PKZip. File systems like NTFS are also used. The main disadvantage of these techniques is that they have to be decompressed before analysis. So, even though the problem of storage requirement is solved, the high transfer rate needed still exists. Thus the problem area still remains unsolved.

b) Transforming a transaction database to a compact representation

Another method used is to use compact data structures to represent the transactions in the database. Prefix tree (Bayardo, 1998) is an ordered tree data structure that is used to store an associative array where the keys are usually strings (items in transaction database). Unlike a binary search tree, no node in the tree stores the key associated with that node; instead, its position in the tree shows what key it is associated with. All the descendants of a node have a common prefix of the string associated with that node, and the root is associated with the empty string. Values are normally not associated with every node, only with leaves and some inner nodes that correspond to keys of interest, thus reducing same required to store a database. Another compression approach using data structure is to "unravel" the data structure into a single byte array (Germann *et al.*, 2009). This approach eliminates the need for node pointers which reduces the memory requirements substantially and makes memory mapping possible which allows the virtual memory manager to load the data into memory very efficiently. Packing the trie (Liang, 1983) is another compression approach used with prefix trees. Liang describes a space-efficient implementation of a sparse packed trie applied to hyphenation, in which the descendants of each node may be interleaved in memory. The FP-growth method uses another compact data structure, FP-tree (Frequent Pattern tree), to represent the conditional databases. FP-tree is a combination of prefix tree structure and node-links (Han *et al.*, 2000).

All these algorithms efficiently reduce the both the secondary and primary storage requirements. The major disadvantage here is the number of scans required during the construction of tree and during frequent pattern mining process.

c) Partitioning transactions in transaction database

A partition is a division of a logical database or its constituting elements into distinct independent parts. Database partitioning is normally done for manageability, performance or availability reasons. Partitioning databases increases performance of regular transactional databases which are done by either building separate smaller databases (each with its own tables, indices, and transaction logs), or by splitting selected attributes of the itemsets. Two types of partitioning algorithms exist. They are horizontal and vertical partitioning. Horizontal partitioning involves putting different rows into different tables. Vertical partitioning involves creating tables with fewer columns and using additional tables to store the remaining columns. Current high end database analysis systems provide different criteria to split the database. Some common criteria are range partitioning, list partitioning, hash partitioning and composite partitioning. All these partitioning schemes, however, require additional computations and while finding frequent patterns, the inter-relationship association between partitions might be missed.

d) Merging transactions in transaction database

Merging transaction approach to create compact transaction database is used as a preprocessing step prior to frequent pattern generation. In this technique, several transactions are merged together to create a new transactional database, which has the following desirable properties.

- Compact representation of the transactional database reduces the problem of memory requirement (both secondary and primary)
- Reduce processing time and I/O time for identifying frequent itemsets and association rules
- Frequent data mining can be performed over the compact representation without losing any mining accuracy
- Can be decompressed to get the original database at any time

As this is the method that is used by this work, the detailed description of the same is provided in the following Section.

III. Proposed Method

This section presents the M^2TQT along with the enhanced M^2TQT algorithm.

a) M^2TQT algorithm

The M^2 TQT algorithm uses a merge Transaction

Scheme (MTS) for producing a compact representation of the transaction database. The MTS is a two-step procedure. Step 1 : Preprocessing and Step 2 : Frequent Pattern Mining. The preprocessing step, uses lexical symbols, to first transform the original raw data into a new data representation. This step is based on the assumption that items in a transaction are sorted in lexicographic order. The second step of preprocessing then sorts all the transactions, groups them and then merges each group into a new transaction. The details on the algorithmic details of merging transaction algorithm, please refer to Dai *et al.* (2008).

b) Enhanced M^2tqt (E-M^2tqt) Algorithm

The general process of E-M^2TQT is shown in Figure 1. It consists of three steps, namely, preprocess, compressing database and frequent pattern mining. The proposed method focus on the problems of repeated database scans and huge number of candidate itemsets generated. The E-M^2TQT algorithm takes advantage of the FP-Tree data structure of FP-Growth algorithm to solve the first problem and the second problem is solved by using a quantification table. The usage of quantification table allows the algorithm to retrieve the original database from the compressed form and prunes irrelevant candidate itemsets, which further reduce the size of the database. Reduction of database indirectly helps to reduce the time required by the mining process. Figure 2 presents the compression algorithm.

The algorithm begins by identifying related transactions and then merging these related transactions together, for which a quantification table is constructed. A transaction T1 is said to be related to transaction T2, if T1 is a subset of T2 or if T2 is a superset of T2. The distance between T1 and T2 is calculated as the difference between the items of two transactions. For example, if T1 = {ABCD} and T2 = {ABC}, the difference (D) between T1 and T2 is 1. Similarly, if T1 = {A} and T2 = {C}, the D is given as 2.

The next step is the creation of quantification table, which is used to record details regarding the transaction relationship. Since the items in a transaction appear in a lexicographical order, the process starts from the left-most item and is termed as a prefix-item. After finding the length of the input transaction (n), for varying lengths (L = 1 .. n), the frequency of count of each itemset appearing in the transactions are recorded. The quantification table has details for each length where information regarding the prefix-item and its frequency count is recorded. Consider for example a transaction database with five items, as shown in Table I.

The quantification process begins by considering the first transaction, {ABCDE} with TID 100 and length = 5. Consider each item one by one in the transaction. Initially, For A, all the five counters L5 to L1 are incremented by 1. Second, for B, the counters L4 to L1 are incremented. Similarly, for C, counters L3 to L1 are incremented, for D, L2 to L1 is incremented and finally, for E, L1 is incremented by 1. Now considering the next transaction {CDE} with TID 200 and leng = 3.

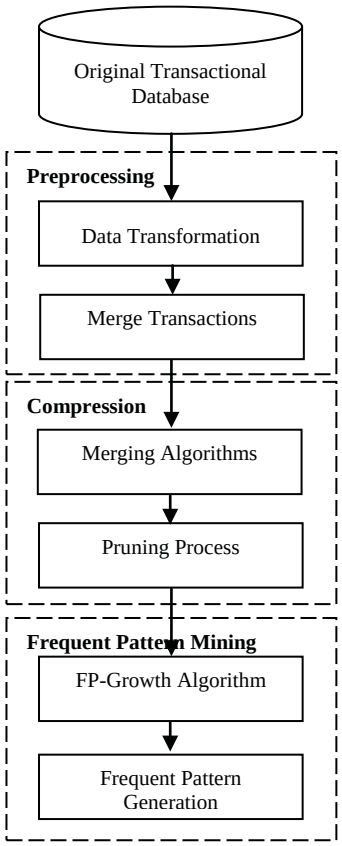


Figure 1 : Proposed Algorithm

```
For I = 1 to number of transactions
        Create Quantification Table
                Quantification table records the length of the transaction database (n) and records the
                count of each itemset appearing in the transaction under each length (L1 – has single
                length, L2 – has two itemset, …)
        Compute length of transaction using quantization table
        Calculate relation distance between transactions and relevant transaction groups
        Relation distance is defined as number of different items between two transactions.
        The smallest transaction of a relevant transaction group is grouped with the longest transaction of
        the same relevant transaction group.
        For d = 1 to n-1
                Consider a transaction, if the relation distance between transaction and merged group = a
                distance 'd', then merge the transactions.
        End for
End for
```

Figure 2 : Compression Algorithm

The counters L3 to L1 is incremented by 1 for C, L2 and L1 are incremented for D and for E, L1 is incremented. Thus, L3 has C2, L2 has D2 and L1 has E2. Finally, for the last transaction, {ACD} (with length 3 and TID 300), A1 is changed to A2 in L3, L2 and L1, C2 is changed to C3 in L2 and L1 and D2 is changed to D3 in L1. The final result is shown in Table II. Now, considering the minimum support, all the candidate itemsets whose counters is less than the minimum support can be pruned out. After pruning, the next step performs the merging process. The merging process is explained with an example below.

Table I : Sample Database

TID	Transactions
100	ABCDE
200	CDE
300	ACD

Let d = 1. Consider two transactions {BCG} and {BG}. After merging these two transactions, the relation transaction group will be {BCG=2.1.2}. Consider another transaction {B}. Compute relation distance for {BCG=2.1.2} and {B}. Since the relation distance is 1, {B} is merged into the relation transaction group and thus becomes {BCG=3.1.2}. The next step calculates the support count of items using a minimum-frequency function. This function takes as input the original transactions and merged transactions and returns the minimum number of itemsets in a transaction. For example, let C2 be {BC, AE} and transaction T* be {{AE=2.1}, {BCG=2.1.2}, {CDEG=2.3.3.1}, {ABCE}, {C}}. The minimum frequency function returns 2 for both BC (0+1+0+1+0=2) and AE (1+0+0+1+0=2), as the number of itemsets in the transactions. These values and the transaction set T* is used by FP-growth algorithm to generate frequent itemsets and association rules.

Table II : Quantification Table

L5	L4	L3	L2	L1
A1	A1	A2	A2	A2
	B1	B1	B1	B1
		C2	C3	C3
			D2	D3
				E2

c) Decompression Algorithm

The main requirement of any compression algorithm that produces a compact representation of the original transaction database is to reproduce the original database without any loss. The proposed algorithm satisfies this requirement also. The algorithm used for this purpose is shown in Figure 3.

1. Let merged transaction be expressed as <s1, s2… , sk .., sn-1,sn>= c1, c2…ck.. cn-1.cn, where s1, s2… , sk , sn-1, sn are items and c1, c2…, ck, cn-1, cn are their corresponding support counts.
2. Identify the smallest count in ci (i=1..n) whose transaction is the longest transaction. Thus the count of longest transaction is ck. Thus the transaction { s1, s2… , sk , sn-1, sn } is recovered and the merged transaction becomes <s1, s2… , sn-1, sn>= c1-ck.c2-ck… cn-1-ck. cn-ck.
3. Remove items with a zero count from the merged transaction.
4. Repeat the above process to find the next longest transaction in the merged transaction until no count left.

Figure 3 : Decompression Algorithm

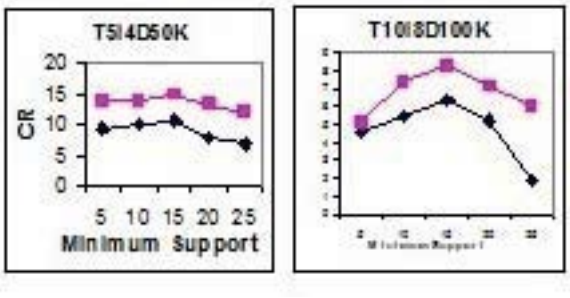
The decompression algorithm is the reverse process of compression algorithm and is explained using an example merged transaction. Consider a merged transaction <ABCD>=3.1.4.2 which has the smallest count of 1 (that is, the count of transaction <ABCD> is 1). Next, decrease the count of each item in <ABCD> by 1 to obtained <ACD>=3-1. 4-1. 2-1 = 2.3.1. Item B is removed since it has a zero count. Repeat this process, to get, <AC>==2-1.3-1=1.2. Finally, <C>=2-1=1. Combining the results we get the decompressed version of the merged transaction <ABCD>=3.1.4.2 as {ABCD}, {ACD}, {AC}, {C}.

IV. Experimental Results

The efficiency of the proposed algorithm was tested using various test data and performance metrics. The proposed model was tested using synthetic dataset as proposed by Agarwal and Srikanth (1994). These transactions mimic the actual transactions in a retail environment. Four synthetic datasets, namely, T5I4D50K, T10I8D100K, T20I10D100K and T20I12D200K were used during experimentation. The proposed algorithm was analyzed using compression ratio in terms of speed, number of transactions and storage required. The average size of the potentially large itemset is taken as 5 and minimum support is varied from 5% to 25% in steps of five. The proposed algorithm was compared with M^2TQT. The algorithms were developed in JAVA with NetBeans 5.5 as front end. All the experiments were conducted on a Pentium IV machine with 512 MB RAM.

a) Compression Ratio with Respect to Storage Space Saved

The compression result in terms of storage size is shown in Figure 4 for the selected four datasets.



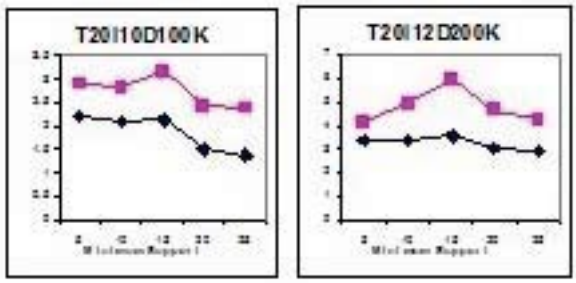
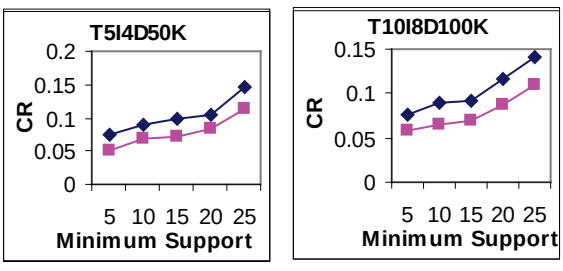


Figure 4 : Compression Ratio For Storage Space Saved

From the results, it could be seen that while both the algorithms are efficient in generating a compact version of the original database, the enhanced M^2TQT shows significant improvement in terms of storage size reduction. The results also show that the E-M^2TQT algorithm scales with different sized datasets. An interesting trend observed is that the maximum amount of compression achieved was when the minimum support is around 15% after which a slow decline in compression rate is envisaged. This means that the maximum performance of the algorithm can be achieved by setting the support value to 15%. Another pattern observed with respect to compression ratio is the relation between size of dataset and average compression achieved. As the size of the dataset increases the compression efficiency decreases. However, the efficiency gain is consistent with all databases which imply that the enhanced approach is better than the traditional algorithm. Thus, the results show that the introduction of FP-Growth algorithm to M^2TQT provides effective data compression.

b) Compression Ratio with Respect to Number of Transactions

Figure 5 shows the compactness efficiency in terms of number of transactions. These experiments were conducted to evaluate the performance of the pruning algorithm introduced in the merging transaction step of the compression algorithm.



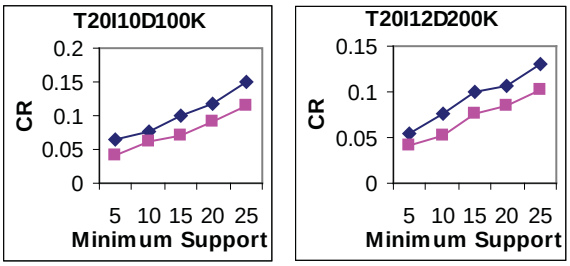


Figure 5 : Compression Ratio For No. Of Transaction

The results of compression ratio with respect to number of transactions again show that the Enhanced M²TQT algorithm is an improved version. While considering various minimum support, the results shows that the minimum support and compression ratio are directly proportional. This is evident from the increasing trend line obtained. While analyzing the scalability of the algorithm, the trend obtained shows that the efficiency is not affected by different sized datasets and remains consistent between 23% and 25%. In conclusion, it could be seen that the pruning algorithm with merging transaction scheme has improved the E-M²TQT algorithm by 25.32% on average.

c) Execution Time

Table III shows the average execution time of each dataset along with the efficiency gain obtained with respect to execution speed. The speed is calculated in seconds.

Table III : Average Execution Speed (Seconds)

DATASET	M²TQT	E-M²TQT	EFFICIENCY (%)
T5I4D50K	1.692	0.970	42.67
T10I8D100K	1.872	1.114	40.49
T20I10D200K	2.222	1.560	29.79
T20I12D300K	2.312	1.840	20.42

As with compression ratio, the E-M²TQT algorithm shows an average 33.34% speed efficiency with respect to M²TQT algorithm. This shows that the speed efficiency has improved in the enhanced version. As expected, the algorithm takes more time to execute large sized datasets than small sized datasets. The small sized datasets (T5I4D50K and T10I8D100K) shows maximum speed efficiency of 42.67% and 40.49%, while it decreased to 29.79% and 20.42% with large sized datasets (T20I10D200K and T20I12D300K).

These results indicate that the performance of E-M²TQT algorithm in terms of compactness achieved with respect to storage size, number of transactions and execution speed with different datasets and different minimum support is efficient when compared with to its existing version.

V. Conclusion

This research work proposed an enhanced transaction database compacting algorithm for frequent mining. Experiments were conducted to analyze the performance of the algorithm. The results showed high performance and there can be reliably used in various applications where pattern mining is needed. Moreover, the speed of the algorithms further makes it suitable for online applications. In future, plans for including frequent pattern-based clustering algorithms and partitioning algorithm to further improve the efficiency of the pattern-mining algorithm are considered.

References Références Referencias

1. Park, J.S., Chen, M.S. and Yu, P.S. (1995) An effective hash-based algorithm for mining association rules. Proceeding of the 1995 ACM-SIGMOD International Conference on Management of Data (SIGMOD'95), San Jose, CA, Pp 175–186
2. Agrawal R. and Shafer, J.C. (1996) Parallel mining of association rules: design, implementation and experience, IEEE Transactions Knowledge Data Engineering, Vol. 8, Pp.962–969.
3. Gionis, A., Mannila, H., Mielikainen, T. and Tsaparas, P. (2007) Assessing Data Mining Results via Swap Randomization, ACM Transactions on Knowledge Discovery from Data (TKDD), Vol. 1 , Issue 3, Article No. 14.
4. Dai, J.Y., Yang, D.L., Wu, J. and Hung, M.C. (2008) An Efficient Data Mining Approach on Compressed Transactions, International Journal of Electrical and Computer Engineering, Vol. 3, No. 2, Pp. 76-83
5. Bayardo, R.J. (1998) Efficiently mining long patterns from databases, Proceeding of the 1998 ACM-SIGMOD International Conference on Management of Data (SIGMOD'98), Seattle, WA, Pp 85–93.
6. Germann, U., Joanis, E. and Larkin, S. (2009) Tightly packed tries: how to fit large models into memory and make them load fast, too, ACL Workshops: Proceedings of the Workshop on Software Engineering, Testing and Quality Assurance for Natural Language Processing, Association for Computational Linguistics, Pp. 31–39.
7. Liang, F.M. (1983). Word Hy-phen-a-tion By Com-put-er, Doctor of Philosophy Thesis, Stanford University.
8. Han, J., Pei, J. and Yin, Y. (2000) Mining frequent patterns without candidate generation, Proceeding of the 2000 ACM-SIGMOD International Conference on Management of Data (SIGMOD'00), Dallas, TX, Pp. 1–12.
9. Agrawal, R. and Srikant, R. (1994) Fast Algorithms for Mining Association Rules, Proceedings of the 20th International Conference on Very Large Data Bases, Pp. 487-499.

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Detecting circular shapes from areal images using median filter and CHT

By Masoud Nosrati, Ronak Karimi, Mehdi Hariri
Islamic Azad University,Eslamabad-E-Gharb, Iran

Abstract - One of the challenging topics in image processing is extracting the shapes from noisy backgrounds. There are some methods for doing it from different kinds of noisy backgrounds. In this paper, we are going to introduce another method by using 4 steps to extract circular shapes from impulse noisy backgrounds. First step is applying median filter to disappear "salt and pepper" noise. This step causes edge smoothing. So, as the second step, a laplacian sharpening spatial filter should be applied. It highlights fine details and enhances the blurred edges. Using these two steps sequentially causes noise reduction in an impressive way. Third step is using Canny edge detection for segmenting the image. Its algorithm is talked during the paper. Finally, forth step is applying Circular Hough Transform (CHT) for detecting the circles in image. At the end of paper different use cases of this method is investigated.

Keywords : *Extracting circular shape • median filter • laplacian filter • Canny edge detection • Circular Hough Transform (CHT).*

GJCST Classification: *I.4,I.4.3*

Strictly as per the compliance and regulations of:

Detecting circular shapes from areal images using median filter and CHT

Masoud Nosrati [α], Ronak Karimi [Ω], Mehdi Hariri [β]

Abstract - One of the challenging topics in image processing is extracting the shapes from noisy backgrounds. There are some methods for doing it from different kinds of noisy backgrounds. In this paper, we are going to introduce another method by using 4 steps to extract circular shapes from impulse noisy backgrounds. First step is applying median filter to disappear "salt and pepper" noise. This step causes edge smoothing. So, as the second step, a laplacian sharpening spatial filter should be applied. It highlights fine details and enhances the blurred edges. Using these two steps sequentially causes noise reduction in an impressive way. Third step is using Canny edge detection for segmenting the image. Its algorithm is talked during the paper. Finally, forth step is applying Circular Hough Transform (CHT) for detecting the circles in image. At the end of paper different use cases of this method is investigated.

Keyword : Extracting circular shape . median filter . laplacian filter . Canny edge detection . Circular Hough Transform (CHT).

I. Introduction

Extracting shapes from noisy backgrounds was always been a challenging problem in image processing. Recent studies represented some solutions, but no one could give 100 percent guaranteed results. In this paper, we are going to introduce another method which is a combination of different filters and techniques. Our purpose is to achieve a better result in detecting the circular shapes within an impulse noisy background. Due to this, 4 steps are listed to be applied to image. First step is using median filter for removing impulse points in the image. These points have an absolutely different color from their neighborhood pixels, and they can have bad effects in the result of edge detecting process. So, median filter will be applied to remove them. Other implicit effect of this filter is smoothing the edges. It can make the edge detection harder and the have bad effects on the result. So, another filter that should be applied as the second step, is laplacian sharpen spatial filter. This filter causes the details become more impressive and the edges become clearer. An implicit effect of this filter is noise increasing in image, but we don't worry about it, because the next step will neutralize this effect. Now, it is time to detect the edges. Due to it, some methods were introduced like Sobel and Canny edge detection. Here, we employ Canny edge detection as the third step, because it has a better performance in detecting the thin edges and when the edges are not very sharp and clear. As the edges were detected, it is time to extract the circular shapes. So, as the fourth step, we will employ Hough transform. In general it can be used for any type shapes. But this special form of this transform is called Circular Hough Transform (CHT).

We are going to investigate all of the mentioned steps in depth during the "METHODOLOGY" section. Then, special uses of this method will be talked, and finally "CONCLUSION" is placed at the end of the paper.

II. Methodology

In this section, we are going to explain the basic concepts of techniques which are used in this paper. Also, their application on the proposed image to extract the circular shapes will be investigated.

a) Median filter

One of the most efficient solutions for disappearing "salt and pepper" noises is applying median filter. Median filter belongs to the group of order static nonlinear filters. It could be easily used for erasing impulse noise. The process is that it replaces the value of a noise pixel with the median gray levels in the neighborhood of that pixel. It causes the impulse noise in the background be disappeared, but also an extra consequence is blurring the edges. This effect could be neutralizing impressively by applying the laplacian sharpening spatial filter [1].

b) Laplacian sharpening spatial filter

The purpose of using laplacian sharpening spatial filter is for highlighting fine details and enhance them which are blurred such as edges. As you know, laplacian is a linear and rotation invariant operator. Laplacian equation can be written for each dimension independently. Partial second-order derivatives in *x* and *y* directions could be calculated by ***(I)*** and ***(II)***.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \qquad (1)$$

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \qquad (2)$$

Author α : Eslamabad-E-Gharb branch, Islamic Azad University, Eslamabad-E-Gharb, Iran. E-mail : minibigs_m@yahoo.co.uk
Author Ω : Eslamabad-E-Gharb branch, Islamic Azad University, Eslamabad-E-Gharb, Iran. E-mail : rk_respina_67@yahoo.com
Author β : Eslamabad-E-Gharb branch, Islamic Azad University, Eslamabad-E-Gharb, Iran. E-mail : mehdi.hariri@yahoo.com

$$\nabla^2 f = [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)] - 4f(x,y)$$

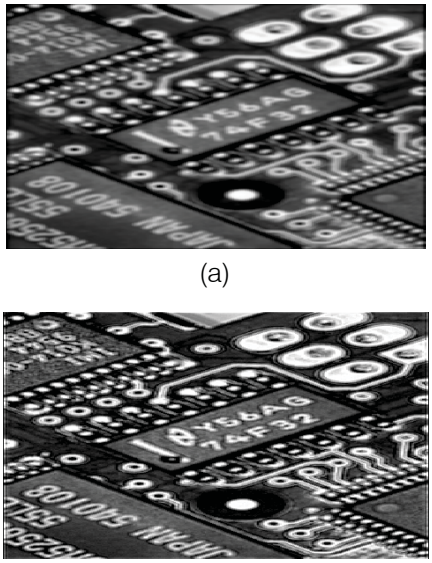
$$g(x,y) = f(x,y) \pm \nabla^2 f(x,y)$$

A 3x3 laplacian mask which could be obtained is as Fig.1.

-1	-1	-1
-1	9	-1
-1	-1	-1

Fig.1 : One pass realization of 3x3 laplacian mask

Result of its application is indicated in **Fig.2** [1].



(a)

(b)

Fig.2 : **(a)** Original image, **(b)** Sharpen image

Laplacian filter by itself causes the increasing noise in the image. But using it in consequence of median filter avoids from occurring this problem. Since the major part of noise is reduced in median filtering process.

c) Canny edge detection

Edge detection is an important topic in the research area of image analysis [2]. The purpose of edge detection in general is to significantly reduce the amount of data in an image, while preserving the structural properties to be used for further image processing. Several algorithms exists, and this worksheet focuses on a particular one developed by John F. Canny (JFC) in 1986 [3]. Even though it is quite old, it has become one of the standard edge detection methods and it is still used in research [4] [5].

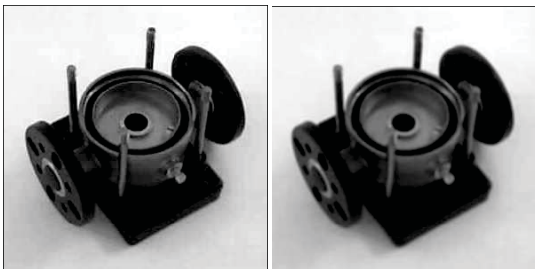
The aim of JFC was to develop an algorithm that is optimal with regards to the following criteria:

1. **Detection**: The probability of detecting real edge points should be maximized while the probability of falsely detecting non-edge points should be minimized. This corresponds to maximizing the signal-to-noise ratio.
2. **Localization**: The detected edges should be as close as possible to the real edges.
3. **Number of responses**: One real edge should not result in more than one detected edge (one can argue that this is implicitly included in the first requirement).

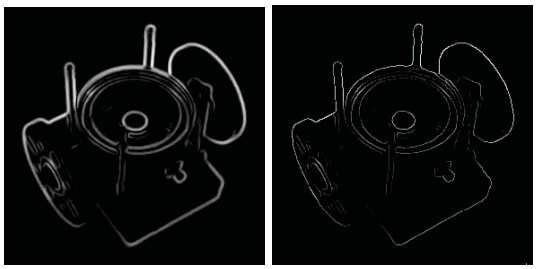
The algorithm runs in 5 separate steps:

1. **Smoothing**: Blurring of the image to remove noise.
2. **Finding gradients**: The edges should be marked where the gradients of the image has large magnitudes.
3. **Non-maximum suppression**: Only local maxima should be marked as edges.
4. **Double thresholding**: Potential edges are determined by thresholding.
5. **Edge tracking by hysteresis**: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

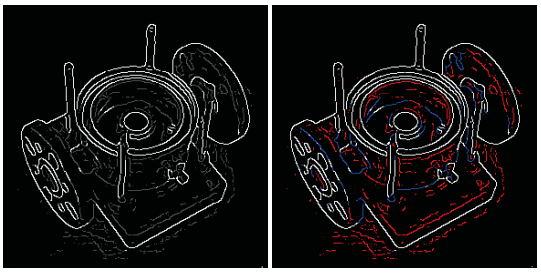
Fig.3 shows the result of applying this algorithm on an image [6].



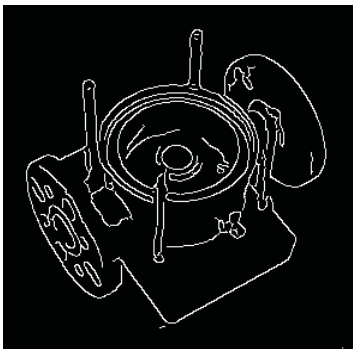
(a) Original **(b)** Smoothed



(c) Gradient magnitudes **(d)** non-maximum suppression



(e) Double thresholding (f) Edge tracking by hysteresis



(g) Final image

Fig.3 : Applying Canny edge detection algorithm to example image

Even though Canny algorithm involves smoothing process, but it doesn't dispel the need for using median filter. Because smoothing step just blur the image and it causes the little noises to be smoothen. But median filter causes impulse noise to be disappeared.

d) Circular Hough Transform (CHT)

A commonly faced problem in computer vision is to determine the location, number or orientation of a particular object in an image. For example, recognizing of roads in an aerial image or detection of circles in the image are two samples of this problem. But it can be solved using Hough transform. Often the objects of interest have other shapes than lines, it could be parables, circles or ellipses or any other arbitrary shape. The general Hough transform can be used on any kind of shape, although the complexity of the transformation increase with the number of parameters needed to describe the shape [7].

The CHT has been recognized as robust techniques for curve detection. This method can detect object even polluted by noise. The CHT was sketched by Duda et al. [8]. The CHT is one of the modified versions of the HT. The CHT aims to find circular patterns within an image. The CHT is used to transform a set of feature points in the image space into a set of accumulated votes in a parameter space. Then, for each feature point, votes are accumulated in an accumulator array for all parameter combinations. The array elements that contain the highest number of votes indicate the presence of the shape. A circle pattern is described by *(////)*.

$$(x_p - x_0)^2 + (y_p - y_0)^2 = r^2 \quad (3)$$

Where x_0 and y_0 are the coordinates of the center and r is the radius of the circle. An example of conventional CHT is shown in Fig.4.

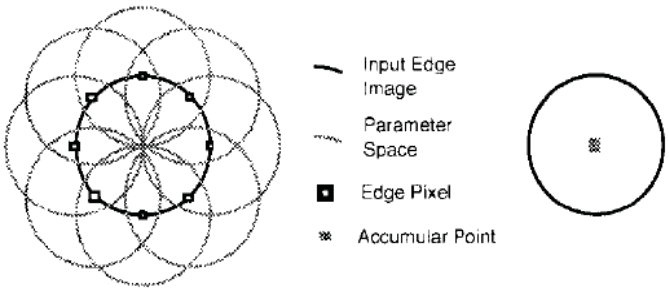


Fig.4 : The contribution of the edge points to the accumulator space

The black circles indicate a set edge points within the image. Each edge point contributes a circle of radius R to an output accumulator space indicated by the grey circles. The output accumulator space has a peak where these contributed circles overlap at the center of the original circle. Modification to the CHT has been widely implemented to either increase the detection rate or reduce its computational complexity [9] [10]. The algorithm for Circular Hough Transform can be summarized to following steps [11]:

//HOUGH BEGIN

1. For each edge point
   Draw a circle with center in the edge point with radius r and increment all coordinates that the perimeter of the circle passes through in the accumulator.
2. Find one or several maxima in the accumulator

//HOUGH END

3. Map the found parameters *(r,a,b)* corresponding to the maxima back to the original image.

After applying the median and laplacian filter, Canny edge detection and CHT, circular shapes will be achieved efficiently.

e) Use cases of represented method

A question which might be asked is "What are the uses of this method?"

Following items can indicate just a small part of use cases of proposes technique:

- Building extraction: Most of building extraction methods can specify the buildings with polygonal rooftops and they have a weak

performance in detecting circular buildings. For example look at the method which was introduced by Masoud S. Nosrati et al. [12].

- Digital filming: While taking film from sport matches like football and volleyball it can be possible to trace the ball by enhancing this technique. It means that a simple camera can take film automatically by using a developed form of this method.
- Biometric security: One way for identification of people is using their iris characteristics. For detecting the iris area this technique can be used. Eyelashes act like impulse noise and the filters which are used in this technique can make the detection process easier and more accurate.
- Astronomy: Proposed method can be used for detecting the planets surrounded by narrow clouds and recognizing the black holes in the special dim images.

Many other use cases can be found in chemistry, physics, biology, computer science and other fields of study. Also, with a little change in this method, other shapes in different noise styles can be extracted.

III. Conclusion

In this paper, we introduced a method for extracting circular shapes from impulse noisy images. Due to this, 4 steps were nominated which were:

1. Applying median filter for erasing the impulse noise.
2. Applying laplacian sharpening spatial filter for neutralizing the blurring effect of median filter and sharpen the edges in the image.
3. Applying Canny edge detection.
4. Applying Circular Hough Transform for extracting the circular shapes in the image.

At the end of paper, different use cases of this method was investigated.

References References Referencias

1. Bernhard Haim, Mathematics in Imaging, Lecture presentation, Available at: http://www.emt.jku.at/education/Inhalte/se_moderne_methoden/WS0304/Haim-Mathematics_in_Imaging.pdf
2. Soo-Chang Pei, Ji-Hwei Horng, Design of FIR bilevel Laplacian-of-Gaussian filter, (ELSEVIER) Signal Processing 82 (2002) 677 – 691.
3. John Canny. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679–698, Nov. 1986.
4. F. Mai, Y. Hung, H. Zhong, and W. Sze. A hierarchical approach for fast and robust ellipse extraction. Pattern Recognition, 41(8):2512–2524, August 2008.
5. Sergei Azernikov. Sweeping solids on manifolds. In Symposium on Solid and Physical Modeling, pages 249–255, 2008.
6. 09gr820, Canny Edge Detection, March 2009, Available at: http://www.cvmt.dk/education/teaching/f09/VGIS8/AIP/canny_09gr820.pdf
7. Simon Just Kjeldgaard Pedersen, Circular Hough Transform, Aalborg University, Vision, Graphics, and Interactive Systems,November 2007. Available at: http://www.cvmt.dk/education/teaching/e07/MED3/IP/Simon_Pedersen_CircularHoughTransform.pdf
8. Duda, R.O. and P.E Hart, 1972. Use of the Hough transformation to detect lines and curves in picture. Commun. ACM, pp: 11-15.
9. Kimme, C., D. Ballard and J. Sklansky, 1975. Finding circles by an array of accumulators. Proc. ACM, 18: 120-122.
10. Guil, N. and E.L. Zapata, 1997. Lower order circle and ellipse hough transform. Pattern Recognition, 30: 1729-1744.
11. Mohamed Rizon, Haniza Yazid, et al. Object Detection using Circular Hough Transform, American Journal of Applied Sciences 2 (12): 1606-1609, 2005.
12. Masoud S. Nosrati and Parvaneh Saeedi, A NOVEL APPROACH FOR POLYGONAL ROOFTOP DETECTION IN SATELLITE/AERIAL IMAGERIES, 2009 IEEE International Conference on Image Processing, 978-1-4244-5654-3.

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Modified Tree Classification in Data Mining

By Raj Kumar, Dr. Anil Kumar Kapil, Anupam Bhatia

Assistant Professor, Kurukshetra University P.G. Regional Centre, Jind (India)

Abstract - Classification is a data mining technique used to predict group membership for data instances [1]. There are several conventional methods for classification in data mining like Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Backpropagation and classification by Lazy Learners. In this paper we propose a new modified tree for classification in Data Mining. The proposed modified Tree is inherited from the concept of the decision tree and knapsack problem. A very high dimensional data may be handled with the proposed tree and optimized classes may be generated.

Keywords : *Classification, knapsack Problem, Modified Tree, KDD*

GJCST Classification: *E.1*

Strictly as per the compliance and regulations of:

Modified Tree Classification in Data Mining

Raj Kumar [α], Dr. Anil Kumar Kapil [Ω], Anupam Bhatia [β]

Abstract - Classification is a data mining technique used to predict group membership for data instances [1]. There are several conventional methods for classification in data mining like Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Backpropagation and classification by Lazy Learners. In this paper we propose a new modified tree for classification in Data Mining. The proposed modified Tree is inherited from the concept of the decision tree and knapsack problem. A very high dimensional data may be handled with the proposed tree and optimized classes may be generated.

Keywords : *Classification, knapsack Problem, Modified Tree, KDD.*

I. Introduction

Data mining or knowledge discovery is needed to make sense and use of data. Knowledge discovery in the data is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns of data [2]. Data mining is the core step of knowledge discovery in database(KDD) and interdisciplinary field includes database management system, machine learning, statistics, neural network, fuzzy logic etc. Any of the technique may be integrated depending on the kind of data to be mined. The research in KDD is expected to generate a large variety of systems because diversity of disciplines to be contributed. Therefore a comprehensive classification system is required able to distinguish between the systems and identify the most required by the user. The major issue involved with the classification rule mining is to identify a dataset for a small number of rules to serve as classifier for predicting the class of any new instance. The classification algorithm should be accurate, simple and efficient. The existing classification algorithm assuming that the input data is drawn from a pre-defined distribution having stationary majors. Therefore these algorithms perform poorly when used to infer real world datasets.

In this paper, we propose a modified tree classification method which is result of knapsack implementation on Decision Tree Approach.

Author α : Assistant Professor, Deptt. of Comp Sc. and Engg., Jind Institue of Engineering & Technology, Jind (India)
E-mail : rajshira@gmail.com
Author Ω : Professor and HOD, Deptt. of CSE/IT, Haryana Institute of Engineering & Technology,Kaithal(india)
Author β : Assistant Professor, Kurukshetra University P.G. Regional Centre, Jind (India)

II. Classification Methods In Data Mining

Following classification methods are used in data mining:

(1) Decision Tree Induction: Decision tree is the learning of decision trees from a class labeled training tuples. A decision tree is a flow chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label. The topmost node in a tree is the root node. The construction of decision tree classifier does not require any domain knowledge of parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees may handle high dimensional data. Their representation of acquired knowledge in tree from is intuitive and generally easy to understand by humans. The learning and classification steps of decision tree induction are simple and fast. In general, decision tree classifier has good accuracy. Decision tree induction algorithms have been used for classification in many applications areas, such as medicine, manufacturing and production, financial analysis, astronomy etc.[3]

(2) Rule-Based classification: In rule based classifiers, the learned model is represented as set of IF-THEN rules. We first examine how such rules are used for classification. Then we study the ways in which they may be generated, either from a decision tree of directly from the training data using a sequential covering algorithm. Rules are a good way of representing information or bits of knowledge. A rule-based classifier uses a set of IF-THEN rules for classification. An IF-THEN rule is and expression of the form:

IF *condition* THEN *conclusion*

The "IF" part of rule is known as the rule **antecedent or precondition**. The "THEN" part of the rule is known as the rule **consequent.** A rule R may be assessed by its coverage and accuracy. Given a tuple, X, from a class-labeled data set, D, let N_{COVERS} be the number of tuples covered by R; $N_{CORRECT}$ be the number of tuples correctly classified by R; and $|D|$ be the number of tuples in D. We may define **coverage** and **accuracy** of R as

$$\text{coverage}(R) = n_{covers} / |D|$$

$$\text{coverage}(R) = n_{corrects} / n_{covers} \quad [3]$$

(3) Classification by backpropagation: Backpropagation is a neural network learning algorithm. The field of neural networks was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogues of neurons. A neural network is a set of connected input/output units in which each unit has a weight associated with it. Neural network learning is also referred to as connectionist learning due to the connections between units. Neural networks involve long training times and are therefore more suitable for applications where this is feasible. Backpropagation is the most popular neural network algorithm. Backpropagation learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known target value. The target value may be the known class label of the training tuple or continuous value. For each training tuple, the weights are modified so as to minimize the mean squared error between the network's prediction and the actual target value. These modifications are made in the "backwards" direction, that is, from the output layer, through each hidden layer down to the first hidden layer.[3]

(4) Lazy Learners: Decision tree induction, Bayesian classification, rule-based classification, classification by backpropagation, support vector machines are the example of the eager learner. Eager learners, when given a set of training tuples, will construct a generalization model before receiving new tuples to classify. Imagine a contrasting approach, in which the learner instead waits until the last minute before doing any model construction in order to classify a given test tuple. That is, when given a training tuple, a **lazy learner** simply stores it and waits until it is given a test tuple. Only when it sees the test tuple does it perform generalization in order to classify the tuple based on its similarity to the stored training tuples. Unlike eager learning methods, lazy learners do les work when a training tuple is presented and more work when making a classification or prediction. Because lazy learners store the training tuples of "instances", they are also referred to as **instance based learners.** When making a classification of prediction, lazy learners may be computationally expensive. They require efficient storage techniques and are well suited to implementation on parallel hardware. Lazy learners naturally support incrementing learning. K-nearest neighbor classifier and case-based reasoning classifiers are the example of the lazy learners. [3]

III. Motivation

While studying the Decision Tree Classification following shortcomings are observed[6].

- Decision trees are less appropriate for estimation tasks where the goal is to predict the value of a continuous attribute.
- Decision trees are prone to errors in classification problem with many classes and relatively with many class and relatively small number of training examples.
- The process of growing a decision tree is computationally expensive. At each node, each maydidate splitting field must be sorted before its best split may be found. In some algorithms, combinations of fields are used and a search must be made for optimal combining weights. Pruning algorithms may also be expensive since many maydidate sub-trees must be formed and compared.
- Most decision-tree algorithm only examine a single field at a time. This leads to rectangular classification boxes that my not correspond well with the actual distribution of records in the decision space.

Looking at these shortcomings, there is need to modify the decision tree classification into a new structure. In this paper in later section we propose a modified tree classification, the idea for which is to implement the concepts of Knapsack approach in Decision Tree Classification.

IV. Why Knapsack

Knapsack is well- known NP-hard problem. It has many applications like budgeting, network planning, network routing, and parallel scheduling etc. Knapsack problem may be of three types

0-1 knapsack problem: In the 0-1 knapsack problem the object is either completely selected or rejected in order to find the optimal solution. In the classical 0-1 knapsack problem only one knapsack is used and binary 1 represents that the object is selected and the binary 0 represents that the item is rejected.

Fractional knapsack: In the fractional knapsack problem the objects may be selected in fractions in order to maximize the profit or in other words we may say to find the optimal solution. For example in we have to select the objects like gold or diamond. Then these type of objects may be selected in fractions.

Multiple knapsack problem(MKP): Consider m containers (knapsacks) with capacities c_1, c_2,……,c_m and a set of n items, where each item has a weight w_1,w_2,……,w_n and a profit p_1,p_2,……p_n. Packing the items in the containers to maximize the total profit of the items, such that the sum of item weights in each container does not exceed the container's capacity, and each of item is assigned to at most one container, is the 0-1 multiple knapsack problem[4].

For example consider two knapsacks having capacities c_1=10, c_2=7. Suppose we have four items with weights 9,7,6,1 and profits 3,3,7,5. Now we have the problem that which item should be placed in which knapsack so that the profit may be maximized i.e. we

want to find out the optimal solution for this MKP. The optimal solution to this problem is move item 1 and 4 to the first knapsack, and item 3 to the second knapsack, it will give us a total profit of 15. So we may say that the MKP is the generalization of the classical 0-1 knapsack problem. Let the binary decision variable x_{ij} be one if item j is placed in container i, and 0 otherwise. The 0-1 MKP may be formulated as given below.

$$\text{maximize} \sum_{i=1}^{m} \sum_{j=1}^{n} p_j x_{ij}$$

$$\text{subject to:} \sum_{j=1}^{n} w_j x_{ij} \le c_i \qquad i=1,\ldots\ldots\ldots,m$$

$$\sum_{i=1}^{m} x_{ij} \le 1 \qquad j=1,\ldots\ldots\ldots,n$$

$$x_{ij} \in \{0,1\} \qquad \text{for all i, j}$$

So in this way 0-1 MKP may be represented as a classical 0-1 knap knapsack problem [5]. After looking the 0-1 multiple knapsack problem, we have observed that in the multiple knapsack problem more than one knapsacks are used and different objects may be placed in the different knapsacks in order to find the optimal solution. That is on the basis of certain attributes objects are assigned to a knapsack. Now in the classification in data mining, constraint set is defined for a class and a object satisfying the constraint set is assigned that class. Many classes may be there and each has its own constraint set. So in multiple knapsack problem multiple knapsacks (containers) are there and if we consider that objects in one knapsack have same attributes then we may consider a knapsack as class. So in this way both approaches have the probability of combination.

V. Modified Tree Classification

In this section, we propose a new tree classification method. The modified tree is inherited from the hybrid concept of the decision tree and the knapsack problem.

As shown in the figure1, Two knapsacks are defined.

1. The topmost node of the proposed tree (i.e. the root node) consists of all data sets.
2. On the left branch of the tree, a class is defined satisfying the first set of the attribute constraints as an external node. The left branch generate a new class.
3. Now the right subtree is explored recursively. The right branch is said to explore further recursively.
4. Each internal node represents the decision node, and each leaf node holds a class label.

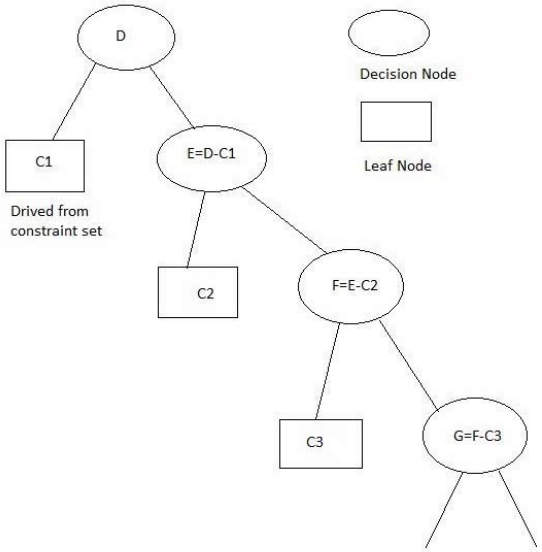


Figure1

With this approach the tree may handle and classify very high dimensional data effectively. However the complexity of the tree may be on higher side in comparison of the decision tree but the quality of the generated classes will definitely be higher. The leaf node of the proposed tree is represented by the rectangle and the internal nodes are represented by the oval.

In the figure1 D represents the all data set, on the left branch of the tree class C1 is derived from the constraint set defined for the class C1 (i.e. the objects in data set D satisfying the constraint set of C1 are placed in the class C1). So all the objects in the class C1 have the same attributes. Now the data set becomes E (E=E-C1). After this the subtree with the root node E is explored and on the left branch of the subtree class C2 is defined as the external node. Now data set becomes F(F=E-C2). Again the same method is applied to derive the class C3. In this way classes C1,C2,C3,…….and so on may be derived recursively in order to classify the data set.

This system has a number of advantages over various classified systems like decision tree induction, rule based classification and lazy learners. This technique may be used to solve a number of real world problems.

VI. Conclusion & Future Work

We proposed the modified tree for classification in data mining. The features of 0-1 Multiple Knapsack and decision tree are inherited. Greedy approach may be used to find the optimal solution of 0-1 multiple knapsack problem. Implementation of Artificial Intelligence (AI) techniques like Neural Network, Genetic

Algorithm, Simulation Annealing may be used for the above said classification method.

References Références Referencias

1. Thair Nu Phyu,"Survey of Classification Techniques in Data Mining"IMECS 2009
2. Fayyad, Piatetsky-Shapiro, Smyth and Uthurusamy,"Advances in Knowledge Discovery and Data Mining", (Chapter 1), AAAI/MIT Press 1996.
3. Jaiwei Han, Micheline Kamber,"Data MiPressning Concepts and Techniques, 2nd Edition", Morgan Kaufmann Publishers,2006 pp 289-301
4. G. Raidl, "The multiple container packing problem: A genetic approach with weighted coding," ACM SIGAPP Applied Computing Review, pp 22-32, 1999.
5. Alex S Fukunga, "A New Grouping Genetic algorithm for multiple Knapsack Problem", CEC 2008.
6. www.wikipedia.org

Global Journal of Computer Science and Technology
Volume 12 Issue 2 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Importance of Ict and E-Governance Security in Punjab

By Jaspreet Kaur, Dr. Vijay Singh Rathor

Baba Farid Group Of Institutions, Deon Bathinda

Abstract - Information and communications technology or information and communication technology, usually called ICT, is often used as an extended synonym for information technology (IT) allied with the computer and communication resources. It was treated as an electronic technique to storage, retrieval and processing on various types of data. Now IT has moved ahead towards every citizen due to its great usability and much more benefits and it also plays a vital role in e-commerce and e-business, so IT became a necessary part of the life of everybody. With the IT, various types of projects are running to provide the several types of facilities in the every area of all over the world towards the citizens. In recent digital era every Government also wants to maximum use of IT for the development of country. Indian Government also takes IT as very seriously and designs various types of projects to implement at every state level to every urban area as well as rural area. The combination of Government, IT and communication resources a concept came which known as E-governance. The purpose of this paper is to explore E-Governance in Punjab which is the richest state of India. However, it"s a typical task to explore integrated E-governance in Punjab, but this paper will try to represent the every aspect of E-governance in Punjab with security point of view. Thus, this paper will discuss from introductory definition of E-governance to implemented key projects under E-governance with security.

Keywords : *IT, e-Governance, Security*

GJCST Classification: *C.2,D.4.6,H.2.7*

Strictly as per the compliance and regulations of:

Importance of Ict and E-Governance Security in Punjab

Jaspreet Kaur[α] & Dr. Vijay Singh Rathor[Ω]

Abstract - Information and communications technology or information and communication technology, usually called ICT, is often used as an extended synonym for information technology (IT) allied with the computer and communication resources. It was treated as an electronic technique to storage, retrieval and processing on various types of data. Now IT has moved ahead towards every citizen due to its great usability and much more benefits and it also plays a vital role in e-commerce and e-business, so IT became a necessary part of the life of everybody. With the IT, various types of projects are running to provide the several types of facilities in the every area of all over the world towards the citizens. In recent digital era every Government also wants to maximum use of IT for the development of country. Indian Government also takes IT as very seriously and designs various types of projects to implement at every state level to every urban area as well as rural area. The combination of Government, IT and communication resources a concept came which known as E-governance. The purpose of this paper is to explore E-Governance in Punjab which is the richest state of India. However, it's a typical task to explore integrated E-governance in Punjab, but this paper will try to represent the every aspect of E-governance in Punjab with security point of view. Thus, this paper will discuss from introductory definition of E-governance to implemented key projects under E-governance with security.

Keywords : IT, e-Governance, Security

I. Introduction

Several dimension and related factors influence the definition of e-Governance. The word "electronic" in the term e-Governance implies technology driven governance. E-Governance is the application of Information and communication Technology (ICT) for delivering government Services, exchange of information communication transactions, integration various stand-one systems and services between Government-to-citizens (G2C), Government-to-Business (G2B), Government-to-Government(G2G) as well as back office processes and interactions within the entire government frame work.

According to the World Bank [1]:- "E-Government refers to the use by government agencies of information technologies (such as Wide Area Networks, the Internet, and mobile computing) that have the ability to transform relations with citizens, businesses, and other arms of government. These technologies can serve a variety of different ends: better delivery of government services to citizens, improved interactions with business and industry, citizen empowerment through access to information, or more efficient government management. The resulting benefits can be less corruption, increased transparency, greater convenience, revenue growth, and/or cost reductions."

Thus, the stress here is on use of information technologies in improving citizen-government interactions, cost-cutting and generation of revenue and transparency.

UNESCO defines e-Governance as [2]:- "Governance refers to the exercise of political, economic and administrative authority in the management of a country's affairs, including citizens' articulation of their interests and exercise of their legal rights and obligations. E-Governance may be understood as the performance of this governance via the electronic medium in order to facilitate an efficient, speedy and transparent process of disseminating information to the public, and other agencies, and for performing government administration activities."

This definition visualizes the use of the electronic medium in the exercise of authority in the management of a country's affairs along with articulation of citizens' interests leading to greater transparency and efficiency.

Dr. APJ Abdul Kalam, former President of India, has visualized e-Governance [3] in the Indian context to mean: "A transparent smart e-Governance with seamless access, secure and authentic flow of information crossing the interdepartmental barrier and providing a fair and unbiased service to the citizen."

E-Governance : "Information and Communication Technology (ICT) that empowers the Government, its citizens including the Government employees, weaker sections, women, people living in far flung and difficult areas and the business houses to transact businesses with government and its agencies online 24/7 "[4]. There are a numerous of definitions for e-Governance. In the framework, E-Governance is defined as a great application of Information Communication and Technology (ICT) to get the better governance and develop a healthy communication

Author α: Department of Computer Science Baba Farid Group of Institutions, Bathinda (Punjab), India
E-mail: dr.preet83farid@rediffmail.com
Author Ω : Director of Shree Karni College,Vaishali Nagar, Jaipur-302021

between the government and various parts of the society.

"E-Governance is defined as the application of electronic means in [5] the interaction between government and citizens and government and businesses, as well as [6] in internal government operations to simplify and improve democratic, government and business aspects of Governance." [5] "Electronic Governance (e-Governance) incorporates all those processes and structures by means of which the new information and communication technologies (ICTs) can be deployed by government to enable the following:

- Administration of government (eAdministration) and delivery of services to the public (eServices). This generically constitutes electronic government (abbreviated eGovernment);
- Informing, vote-enabling, representation-enabling, consulting and involving the citizenry in, among others, broad consensus making in society in matters pertinent to decision making in political, social and economic priorities in government.

This constitutes Electronic Democracy (abbreviated eDemocracy); Transacting business with its "supply chain", namely, partners, clients and the markets. This constitutes Government Electronic Business (abbreviated simply eBusiness)." [6]

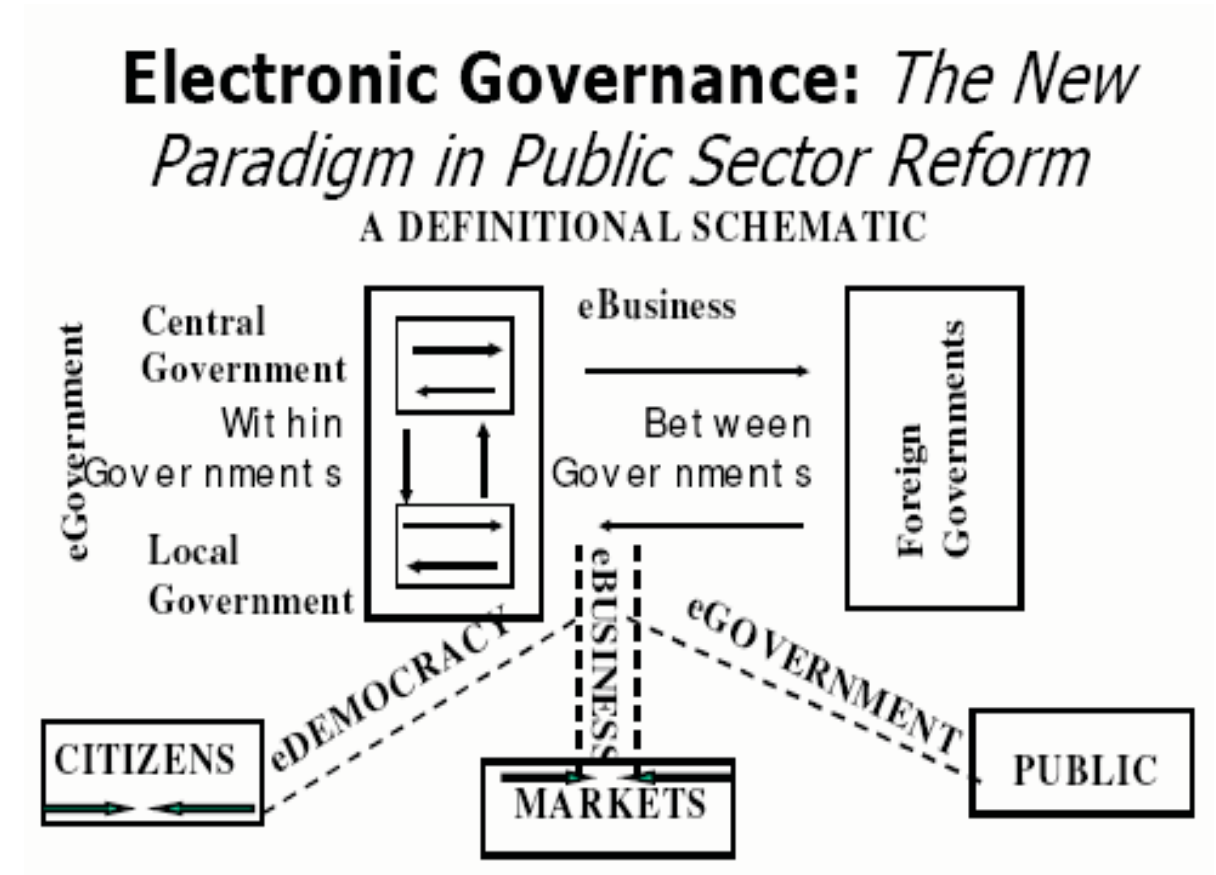


Fig . (a) A Board Definitional Schematic for e-governance [6]

A simpler translated diagram is as follows

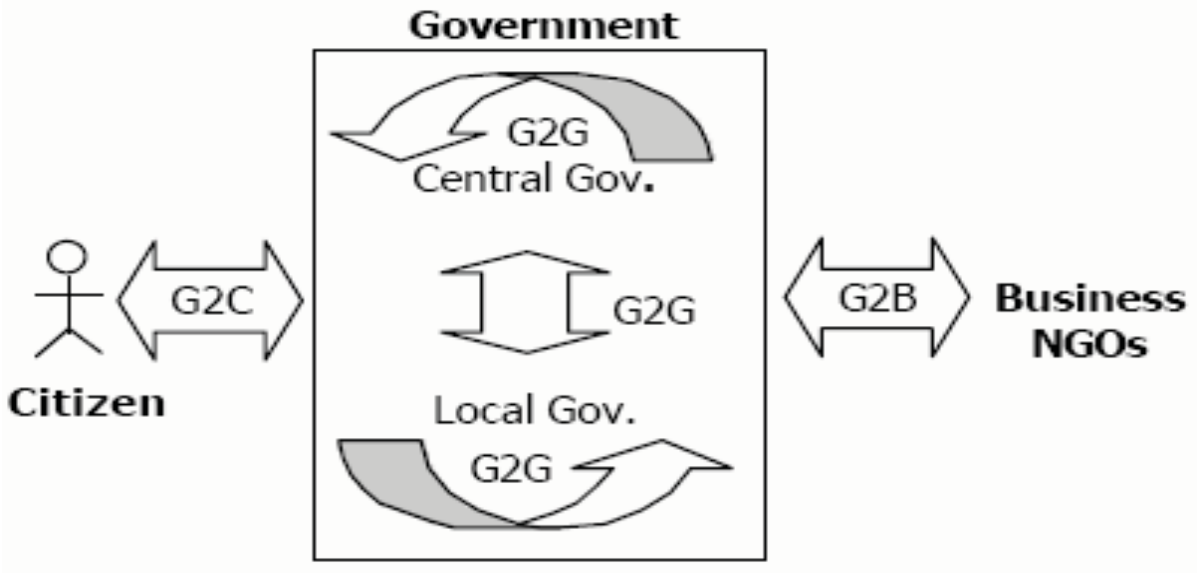


Fig . (b)

It is clearly showing that several benefits of IT which can reach the public at large as the development of e-Government services. The major components for e-Governance project are reengineering government. E-Governance is not just an ICT enabled, efficient and effective public service with enhanced revenue generation for the Government and also not just an application of certain technology but also a way of providing a great citizen services through the use of ICT.

Punjab (Land of five rivers) Punjab is located in the northwest of India surrounded by Pakistan on the west, the Indian states of Jammu and Kashmir on the north, Himachal Pradesh on its northeast and Haryana and Rajasthan to its south. It covers a geographical area of 50,362 sq. km which is 1.54 % of country's total geographical area. Punjab state is located between 29° 30' N to 32° 32' N latitude and between 73° 55' E to 76° 50' E longitude. Its average elevation is 300 m from the sea level. Chandigarh is the capital of the Punjab.

Sikhism is the predominant faith in Punjab. About 60% of the people in the state are Sikhs. The holiest of Sikh shrines, the Sri Harmandir Sahib (or Golden Temple), is in the city of Amritsar. The Sri Akal Takht Sahib which resides within the Golden temple complex is the temporal seat of Sikhs. Of the five Takhts (Temporal Seats of religious authority) of Sikhism, three are in Punjab. These are Sri Akal Takht Sahib,DamdamaSahib and Anandpur Sahib. Anandpur Sahib is where Guru Gobind Singh created the Khalsa in 1699 on the day of Vaisakhi. During major holidays on the Sikh calendar (suchas Vaisakhi, Hola Mohalla, Gurpurb and Diwali), many Sikhs gather and march in religious processions through virtually every city, town and village in Punjab.

According to India Today [7], Leading magazine in India, Punjab has been awarded best overall state since 2003, and has been able to retain the top position every year. It also affords best quality of life to its residents.

Punjab has the best infrastructure in all of India [8] . Although it has a huge shortage of electricity due to high demand, all major cities in Punjab benefit from this and have some of the lowest tariffs in India. All of Punjab's villages have been provided electricity and connected to the state electrical power grid since 1974.

II. Vision of E-Governance In Punjab

In consonance with the national objective of making India a global IT Power and a front runner in the information revolution, the government of Punjab set up the Department of Information Systems and Administrative Reforms (DISAR). The Department of Information Technology (DoIT) has been set up to execute IT policy framework in the state of Punjab. Punjab IT Policy was formulated in 2001 so as to make Punjab a favored industrial destination with a world class infrastructure, to provide citizen-centric governance, and to turn the state into a knowledge society. Punjab was the first state to implement the national e-governance plan under National common services centers (NCSC). The e-governance initiatives of the state focus on creating efficient and cost effective government by improving the internal processes of the government through administrative reforms, process re-engineering, modernization and deployment of IT for an efficient, productive and accountable government. The new Department of Information Technology (DoIT) has been entrusted with the following responsibilities [9].

- To formulate the policy on the use of Information Technology in the State.
- To formulate and implement a plan for introduction

of Information Technology in the Punjab Administration at all levels, in coordination with the concerned Government Departments; and

- To give technical advice to all the Government Departments regarding adoption of suitable information technology systems and for making appropriate arrangements to maintain the same.

Keeping in view the resources constraints, the government considered it a better strategy to prioritize its departments and agencies on rational criteria so as to computerize them in phased manner. Twenty four departments of the State Government had participated in preparation of the Roadmap and they were prioritized into three phases as below:

- First Phase: Agriculture, Excise & Taxation, Finance, Health & Family Welfare, Local Government, Revenue, Transport, e-District, Food & Civil Supplies, Secretariat.
- Second Phase: Education, Home, Information Technology, Labour & Employment, Rural Development & Panchayats, Irrigation & Power, Public Health, Social & Women Welfare& Welfare of Schedule caste & Backward Classes, Industries.
- Third Phase : Co-operation, Information & Publicity, Planning, PWD B & R, Town & Country Planning, Advocate General, Punjab, Animal Husbandry, Chief Architect, Election, Forest, Governor House, Hospitality, legal remembers, Printing and Stationery, Prosecution & Litigation, Puinjab Vidhan Sabha, Sainik Welfare, Sports, Technical Education & Industrial Training, Tourism & Culture Affairs.

III. Projects of E-Governance In Punjab

The Department of Information Technology (DoIT) prepares and executes plans in collaboration with the concerned departments to leverage the power of Information & Communication Technology (ICT) as a vehicle for improved governance and service delivery to the citizens in different departments of the State Government [9]

- Suwidha Integrated Citizen Services: (Benefits of the project: All the applications pertaining to different jobs of DC office are accepted at Single Window, Delivery time for each and every job is pre-defined, Delivery is made through SUWIDHA counter, Citizen can check the status of application through web site http://suwidha.nic.in)
- Integrated Land Records Management System: (Benefits of the project: 153 Common service centers are being opened for delivery of services relating to Land Records and Registration of Properties in an integrated manner across the State)
- Transport Services: (Benefits of the project: Able to deliver Driving licenses and Vehicle Registration Certificates in 45 minutes which earlier use to take 10-15 days)
- Excise & Taxation: (Benefits of the Project: Better Management of Records, Reduction in Tax evasion, Improvement in delivery of services to the business community through front-end windows, Reduction in defaulter's list, Zero-level in denial transactions from Information Collection Centre (ICCs), Significant plugging of tax evasion / leakage, Empowering all stakeholders through technology, Sole method of retrieving, monitoring, verification of business records and its transaction by the Staff at every level, Eliminating paper administration, Facilitates matching of Returns filled by each dealer with its transactions during a particular period)
- ICT Education Project: (Benefits of the project: Providing computer education to 11 lacs student population of Punjab out of total 13.2 lacs students in Government schools. This will give tremendous confidence and competitive advantage to the students of Government schools and especially SC's and poor students to enable them to bridge the digital divide)
- Treasury & Accounts Management System: (Benefits of project: Inter Connectivity between Secretariat and Directorate District Treasuries and Directorate and District Treasuries and Sub-Treasuries has been provided by BSNL. The Integrated Treasuries Information System of Punjab (ITISP) has been implemented at Patiala and Ropar Distt. Treasuries and the software has been replicated in all Distt. and Sub Treasuries through which data of all treasuries shall be transferred on Central Server placed at Head Quarter for generating different MIS reports. This would enable to monitor the working of Treasuries and exchange data online from the State Headquarters and vice-versa. It will also facilitate effective control over the expenditure by allocating DDO-wise allocation of budget to the Treasuries, whereby, it will not be possible for DTO to pass bills beyond the authorized appropriation. Any change in budget allocation by the HOD shall have to be intimated to the Directorate of Treasuries and Accounts from where it will pass online to the concerned DTO to enable the)
- Social Security Management System: (Benefits of the project: To streamline the functioning of the department, To reduce delays, To bring efficiency in disbursement, To keep proper accounting for audit, To have proper re-conciliation with the banks, To keep track of verifications and re-verifications resulting in savings of funds on account of weeding out the wrong beneficiaries)
- Punjab Wide Area Network: (Benefits of the project: Provide reliable, vertical and horizontal connectivity

within the state administration to make the Government more productive. Reduce communication cost. Provide a secure backbone for encouraging electronic transactions. Provide efficient service management Strengthen Disaster Management Capacity. Provide the Government agencies, the ability to leverage a robust infrastructure to provide a complete array of Government services and information. Ensure that every citizen in the state has access to Government services and information whenever and wherever they need it. Make services available in a cost-efficient manner, offering public constituencies' equivalent access at an equivalent price, regardless of their location in the State of Punjab. Move toward the provisioning of converged communication Services (voice, data and video) and the interconnection and interoperation of network platforms.)

- Common Service Centers.
- Agrinet Punjab
- Food & Civil Supplies
- Personal Management System
- e-Procurement
- State Data Centers
- e-District

IV. Security of E-Governance

The information systems give the false impression of being impregnable and secure. This impression is created because, most people are not aware of the limitation of the technology. They believe that machines are more reliable than men as they can not be corrupted. However, the incidents of cyber crimes are increasing with every passing day. The credit card frauds and fake electronic transactions are becoming the order of the day. Most people, who are aware of the cyber frauds, are quite reluctant to use technology if they believe that the IT systems are not secure. Most people are afraid of making e-payments and using credit cards to make payments.

The security of the data and information is extremely important to win the trust of the user. All the benefits of the e-government can be wiped away by a single act of security breach which may corrupt or manipulate all data of the system. It is often impossible to restore the system in the electronic environment as no copies of the original data can be extracted from the system. Therefore, the e-government projects must have highest standards of security and privacy to succeed.

It is important to note that the e-Governance security need is dynamic rather than static and depends on the operational dynamics. Thereby, the process of designing and deploying an information security infrastructure is a continuous and dynamic process. Often, the change in needs is frequent. In order to be sustainable under such frequent changes, the process has to be developed from a life-cycle approach. This observation leads to the concept of Security Engineering Life-Cycle.

Security Requirement Specification and Risk Analysis: The first phase in the Security Engineering Life Cycle collects information regarding assets of the organization that needs to be protected, threat perception on those assets, associated access control policies, existing operational infrastructure, connectivity aspects, services required to access the asset and the access control mechanism for the services.

Security Policy Specification: Security Requirement Specification and Risk Analysis Report as input and generates a set of e-Gov security policies. The policy statements are high-level rule-based and generic in nature and thereby, does not provide any insight to system implementation or equipment configuration.

Security Infrastructure Specification: This phase analyses the Security Requirement Specification and the Security Policy Specification to generate a list of security tools that are needed to protect the assets. It also provides views on the location and purpose of the security tools.

Security Infrastructure Implementation: The organization, in this phase, procures, deploys, and configures the selected security infrastructure at the system level.

Security Testing: In this phase, several tests are carried out to test the effectiveness of the security infrastructure, functionality of the access control mechanism, specified operational context, existence of known vulnerabilities in the infrastructure etc.

Requirement Validation: This phase analyses the extent of fulfillment of the security requirements for implementing e-Governance organization by the corresponding security policy and the implemented security infrastructure. Change in the service goal, operational environment, and technological advancement may lead to a fresh set of security requirements and thereby, triggering a new cycle of the Security Engineering Life Cycle.

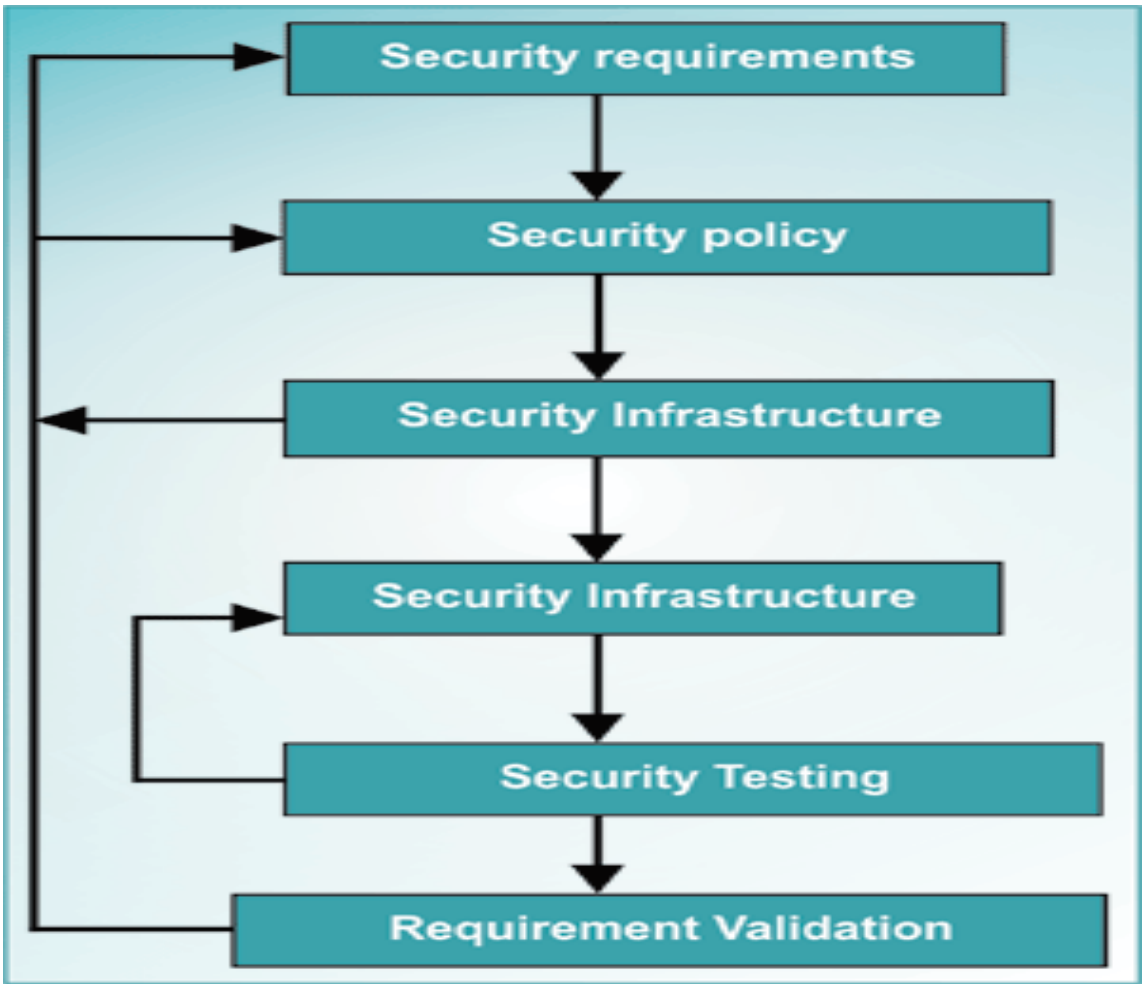


Fig.(c)

V. Conclusion

E-governance enhances the relationships between government to government, government to citizens, citizens to government, government to private sector and NGO's to government, using ICT. Thus, e-governance not merely provides information about various activities of a government but also involves citizens to participate in government's decision-making process. During the last few years, many initiatives have been taken by different state governments in India for using Information Technology as a tool in the functioning of government so as to provide better services to citizens. The initiatives of Punjab state are important. E-governance has eventually started to gain popularity in most cities of Punjab.

Many of the citizens of the developing countries like India are illiterate. Also, presently there is lack of inexpensive and easy to use security infrastructure in G2C application. Applications security issues like authenticity, accountability, confidentiality, integrity, non-repudiation, etc need also to be addressed.

Designing and implementing more effective approaches for securing E-government is an important issue, because, the governmental information is usually so sensitive. Security has an important role in trust information of citizens and their adoption of e-Governance. However proper use of e-governance certain points are still to be seen by the state government, which includes:

- Review of the progress of all the ongoing IT projects is must for all time to come.
- Sustainability of already started initiatives is must.
- Compulsory computer education from class 6th to 12th in government schools is required.
- Use of local languages in the IT implementation process. It is essential that local level databases be maintained in Punjabi language.
- Basic computer education should be given for proper use of e-governance in rural area.
- It's important to educate people at all levels about the benefits of e-governance by highlighting as to how it can save their precious time and efforts thereby making the government functioning more transparent. For this purpose, the public libraries

in the state can play a very crucial role. Hence, the state government must take measures to upgrade the public library system in the state.

References Références Referencias

1. http://go.worldbank.org/M1JHE0Z280 (extracted on 18.08.2008)
2. http://portal.unesco.org/ci/en/ev.phpURL_ID=4404&URL_DO=DO_TOPIC&URL_SECTION=201.html
3. Inaugural address at IIT Delhi during International Conference on e-Governance.
4. eGovINDIA group for TRUE eGovernance in INDIA.(http://groups.yahoo.com/group/eGovINDIA/)
5. 'E-Governance and Developing Countries, Introduction and examples" by Michiel Backus April 2001
6. "ELECTRONIC GOVERNANCE : Abridged Definitive Conceptual Framework" by Rogers W'O Okot-Uma, Common wealth Secretariat London
7. Arvind Chhabra (2009-09-18). "Punjab: Progressive by nature : India Today". Indiatoday.intoday.in. Retrieved 2010-07-18.
8. Welcome to Official Web site of Punjab, India
9. http://www.doitpunjab.gov.in/about_doit.htm
10. http://www.egovonline.net/articles/article-details.asp? Title=Information-Security-Issues-in-eGovernance & ArticalID =2037&Type=PERSPECTIVE

Global Journals Inc. (US) Guidelines Handbook 2012

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

M

N

O

P

Q

R

S

T

U

V

W

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