Framing Qualitative Bankruptcy Prediction Rules Using Ant Colony Algorithm

A.Martin,
Research Scholar,
Department of Banking
Technology,
Pondicherry University,
Puducherry, India.

Aswathy.V
PG Students,
Department of Computer Science
and Engineering,
Sri Manakula Vinayagar Engineering
College, Puducherry, India.

V.Prasanna Venkatesan,
Associate professor
Department of Banking
Technology
Pondicherry University,
Puducherry, India.

ABSTRACT

This paper is to frame qualitative Bankruptcy Prediction (BP) rules using the concept of Ant Colony Algorithm. There are various researches in the area of qualitative BP, among them Genetic Algorithm (GA) seems to more effective. But the redundancy and over lapping of the generated rules cannot be overcome by GA. In our work, we are proposing ACO for generating the rules for qualitative BP. The heuristic function and the pheromone trails are some of the features of ACO, which helps to give a positive feedback in generating rules and to avoid false rules. We are using Association Rule Miner (ARM) algorithm to cluster the generated rules and Partial Swarm Optimization (PSO) Technique to get the best rules among the generated rules. By using this we can generate better rules with more qualitative factors and redundancy and overlapping of the rules can also be avoided.

Key words

Bankruptcy Prediction, Ant Colony Algorithm, Partial Swarm Optimization, Association Rule Miner, heuristic function, pheromone trail.

1. INTRODUCTION

Ant Colony Optimization is an agent-based technique which stimulates the natural ant behavior to obtain the optimal solution. Ant colony optimization algorithms have been applied to many combinatorial optimization problems, ranging from quadratic assignment to fold protein or routing vehicles and a lot of derived methods have been adapted to dynamic problems in real variables, stochastic problems, multi-targets and parallel implementations. It has also been used to produce near-optimal solutions to the travelling salesman problem. They have an advantage over simulated annealing and genetic algorithm approaches of similar problems when the graph may change dynamically; the ant colony algorithm can be run continuously and adapt to changes in real time. This is of interest in network routing and urban transportation systems [22].

Most of the firms analyze the quantitative and qualitative data for the success of their business. Quantitative data are said as objective, i.e, based on the accounting details such us, total turnover, marginal profit, return of sales, etc. This information is available from the book of accounts or account database and based on this we can take decisions. But Qualitative data on the other hand are subjective, that is, based on the subjective

knowledge only we can take decisions. For the subjective knowledge we need help from the expert. An expert in a domain can be a person with good knowledge or good experience in a particular domain. The information that are beyond the accounting details are called as qualitative factors. Qualitative factors don't have any measurement, only the rating of the risk factors can be done by the experts based on the corresponding domain. So analysing the qualitative factors play an important role in the Bankruptcy.

Bankruptcy prediction is an analysis to ensure the stability of business. Many firms fail due to the improper analysis of their business. The success of the firm mainly depends on the decisions that are taken by the management at the proper time. So taking the appropriate decision at proper time is more important in all business. In this work we are proposing a qualitative BP model for generating decision making rules.

In our work we are using Ant Colony Algorithm for generating Qualitative BP rules. The generated rules using ACO are then clustered based on various characteristics by using the Association Rule Mining Algorithm and the best rules among this are extracted using the Partial Swarm Optimization technique. By using this we can avoid the redundancy in rules and also helps in avoiding the false rules in the prediction process.

The remaining of the work includes: Section 2 reviews the prior research; Section 3 about the proposed system, Section 4 predicted rules in qualitative BP and Section 5 discuss the conclusion and future research of the work

2. PRIOR RESEARCH

There are many researches undergone in the area of bankruptcy prediction. There are many studies regarding the analysis of qualitative and quantitative Bankruptcy Prediction. The studies include statistical techniques as well as intelligent techniques in BP. The statistical techniques in modeling of the corporate BP include invariate and Multivariate Discriminant Analysis [1][2] and the classification algorithms such as Linear Discriminant Analysis and Logistic Regression which are the linear approaches. The intelligent techniques [19] used in BP are Neural Networks (NN) [9], Fuzzy set Theory (FS), Decision Tree (DT), Rough Set (RS), Case Based Reasoning (CBR) [5], Support Vector Machine (SVM) [11][7], Data envelopment analysis and some soft computing techniques.

There are only few works under the qualitative BP [18][16][17]. Among this Genetic Algorithm (GA) [3] support more in generating rules and also in BP [6][8][10][11] and listed out six internal risk factors and the associated external factors and rated as positive, negative and average based on the risk factor. They

apply the operations of crossover, mutation and the fitness function is evaluated based on the accuracy, and coverage. And the maximum of accuracy and coverage is taken as the fittest. But the generated rules by GA fail in overcoming the overlapping and the redundancy in the generated rules. Then a financial distress early warning system is developed [13]. In this approach, they develop an early warning system for financial bankruptcy prediction. They also use almost the same risk factors that are used by the former approach. They rate the quality risk factors as white, black and gray based on the risk in BP. The white factors means that don't cause any Bankruptcy, black means they surely will cause Bankruptcy and gray means that may or mayn't cause Bankruptcy. Here they only concentrate on the gray values. And based on a grey value evaluation method they complete their work. They don't try to generate any rules.

Ant colony is already used in BP for many classification problems. There are various works that undergone ACO in various applications [4][5][12][15][14]. Ant colony shows a good sign in generating rules under various applications such as in biochips, credit card system, etc. In the case of travelling sales man problem ACO is first used and it is successfully in generating the best solution [21]. ACO based on the concept of the pheromone update and heuristic function helps in generating the best solution in travelling sales man problem and other network related application such as, routing algorithm and other optimal application.

2.1 Qualitative Factors

The qualitative factors have much influence in the success of the business because it is outside the accounting details. Based on the relevant previous experience and good knowledge on the business domain can only identify and analyse the qualitative factors regarding the business. There are various methods for collecting the qualitative factors they are questionnaires, interviews, etc. Here we are using the qualitative factors such as Industrial Risk (IR), Management Risk (MR), Financial Flexibility (FF), Credibility (CR), Competitiveness (CO), and Operational Risk (OP). These qualitative factors are said as internal Risk factors and inside each internal risk factor there are many external factors which combine to form these internal risk factors. The external factors or outer factors of Industrial Risk includes Govt. Policies & International agreements, Cyclicality, Degree of competition, The Price & stability of market supply, Product Life Cycle, The size & growth of market demand, etc; the Management Risk includes Availability and completeness of management, Stability of Management, The relation between management or owners, Human Resource Management, etc. Like this each internal factor some associated external factors.

In the next session we are discussing about the proposed work. Here we are using Ant colony algorithm concept for generating qualitative BP rules.

3. PROPOSED RSEARCH

In this proposed system, we are using Ant Colony Algorithm for generating qualitative Bankruptcy Prediction Rules. This system helps to generate more accurate rules using the properties of heuristic function and pheromone trails in the Ant Colony Optimization technique. In this work we are grouping the generated rules using Association Rule Miner (ARM) algorithm [20] and we are ranking the generated rules using Partial Swarm Optimization technique since PSO is a better technique in Swarm Intelligence to get the best among the group.

Qualitative factors used:

The qualitative risk factors that used in our work include the Problem Solving Capacity (PS), Initiative (IN), Effective Orientation (EO), Concern of Cost (CC), Concern of People (CP), Concern of Quality (CQ), Information Seeking (IS), and Assertiveness (AS)

This qualitative factors are listed and questionnaires are made based on this any are proceeded to the expert judgment. The expert's rate the parameters (qualitative risk factors) based on its influence in the bankruptcy. The judgment values are as Good, Bad and Medium based on its performance criteria in the business. The value is taken as Good means it has influence over the business, i.e, the variations in that risk factors must give high priority in the business. Means this factor must be considered when taking decisions because it can lead to bankruptcy. If the judgment rate as Medium means it is a parameter which may or mayn't cause Bankruptcy. If the judgment is Bad means this qualitative factor doesn't have much influence in the stability of business.

Some of the questionnaires are listed in table 1.

Table 1: Tables listing the questionnaires given to expert's

Table 1(a): Did *Govt. Policies and International Agreements* affect Bankruptcy?

Parameter Judgment	Explanation
Good	It has less influence over the business
Medium	It has some influence over the business
Bad	It affect the changes in business

Table 1(b): Whether the *degree of competition* affects the Bankruptcy?

Parameter Judgment	Explanation
Good	It has less influence over the business
Medium	It has some influence over the business
Bad	It affect the changes in business

The expert's rate the parameters to the questionnaires as High, Medium, and Low based on the risk factor of the parameters to the BP. These rated factors are now coded to binary forms of 0's and 1's. We code the factors as combinations of 0's and 1's. If the expert judgment to a parameter is Good, then we rate it as 100, Medium then we code as 010 and Bad code as 001.The Table 2 shows the coding of the risk factors based on the expert judgment.

The expert judged parameters values are used for generating rules. The artificial ant moves thru these parameters and when it reaches the destination class or the actual class a rule get generated. The destination class or actual class includes Bankruptcy is yes or bankruptcy is no. We code Bankruptcy is yes as 10 and bankruptcy is no as 01. Based on the expert judgment a qualitative factors have any of the three values such as, good, average or bad based on the risk factor each parameter is carrying. Based on the risk features the rules may vary. So the qualitative risk factors and their expert's judgment play an important role in bankruptcy prediction and decision making.

Table 2: Coding of risk factors based on questionnaires

Qualitative Risk Factors	Good	Medium	Average		Coding
IR	1	0	0		100
MR	0	1	0		010
FF	1	0	0		100
CR	0	0	1		001

Table 3: Coding of Destination Class

Class	Bankruptcy	Non-Bankruptcy		Coding
Actual Class	1	0		10

3.1 Generating Rules using Ant Colony Algorithm

Rule generating is done by the ACO. There are various methods for representing the data. Here we are generating rules because it is easier for understanding to human. Rules are in the form of human readable sentences having *If..Then..* conditions.

A rule may be of the form:

IF <condition1>AND<condition2>AND…<condition n>
THEN <predicted class>

A condition is in terms of <attribute=value>. For example, <gender=male>, <Nationality=Indian>, etc and the predicted class is of the form <bankruptcy=yes>, <bankruptcy=no>, etc. The Rules are easy understandable by the human because it resembles human languages with '*If and Then*' conditions.

Here, Rules are constructed by the Ant Mining Algorithms. Mining is performed through an iterative procedure that discovers one rule per time. The 3 Key steps involved in this mining are, (a) Rule construction – term selection (b) Rule pruning (c) Pheromone updating.

3.1.1 Rule construction

In this step, the rules get constructed using the ACO concept. The artificial ant starts its journey from source to destination to generate rules. Here the risk factors are the sources and the destination classes are the destinations. The destination class are Bankruptcy = yes, Bankruptcy = no, etc. Based on the probability of choosing the nodes by the artificial ant is by the heuristic function and the probabilistic trails. Each chosen nodes are attached and when reach the destination gets a rule and thus generated rules are evaluated based on various interesting measures to assure the quality.

3.1.2 Rule pruning

In this step, the generated rules are clustered using the Association Rule Miner Algorithm. Based on the support and confidence value the rules are clustered. Here the representative rules on each cluster are also formed. This avoids the redundancy among the rules. The various interesting measures used are listed in the table2.

3.1.3 Pheromone updating

The artificial ant starts its journey from one node to next based on the probability of pheromone concentration. In the ACO concept, the ant deposit the pheromone on the way of its journey and it update the path to source when reached a destination class. There are two pheromone updates, local as well as global pheromone update.

When the ant select one node to next based on probability, it deposits some amount of pheromone on the path and is called Local pheromone update. When the ant reaches a solution, the best ant updated its path back to source. And the main interesting feature of this concept is the pheromone evaporation, that is, the concentration of the deposit starts to evaporate at a time interval. This concept helps a positive feedback and helps to avoid negative feedback and to avoid the redundancy and overlapping of rules.

The pheromone trail act as a memory to the artificial ant and helps to keep track of the learning.

3.2 Ant Miner Algorithm

We are using Ant Miner Algorithm here which is a combination of Data Mining and Ant Colony Algorithm. ACO is a heuristic search algorithm used to get the optimal solution for the problem.

The artificial ant move through the risk factors and based on the probabilistic function the ant move from one factor to another. The probabilistic function is evaluated based on the heuristic and pheromone trails. The *heuristic function* evaluates the quality of the pheromone deposit on a node. The pheromone trails include the *local pheromone* as well as *global pheromone* update.

3.2.1 Rule framing using Ant Miner Algorithm

1. TrainingSet = {qualitative parameters}
2. RuleList = {}
3. **while** |TrainingSet| > Max uncovered qualitative parameters **do**
4. *ant_index* = 1
5. *cvg_index* = 1
6. Initialize probability, pheromone and heuristic values according to (1), (2) and (3)
7. **while** (*ant_index* < *no_ants*) AND (*cvg_index* < *no_same_rules_allowed*) **do**
8. Let ant run from source to destination, defining rule R*ant_index*
9. Prune rule R*ant_index*
10. Update pheromone on parameters of path according to (3)
11. **if** R*ant_index* is the same as R*ant_index*−1 **then**
12. *cvg_index* = *cvg_index* +1
13. **else**
14. *cvg_index* = 1
15. **end if**
16. Kill ant
17. *ant_index* = *ant_index* + 1
18. **end while**
19. Choose best rule R*best* among all induced rules R*ant_index*
20. Add rule R*best* to RuleList
21. TrainingSet = TrainingSet \ {parameters covered by R*best*}
22. **end while**
23. Evaluate performance on test set

In the algorithm, the Probabilistic function (1) can be calculated as follows,

The amount of pheromone trail $\tau_{ij}(t)$ associated to arc (i, j) is intended to represent the learned desirability of choosing edge j when in edge i (which also corresponds to the desirability that the arc (i, j) belong to the tour built by an ant) is given by.

$$P_{i,j}(t) = \frac{(\tau_{i,j}(t))^{\alpha} \times (\Delta\tau_{ij}(t))^{\beta}}{\sum (\tau_{i,j})^{\alpha} \times \Delta\tau_{ij}(t)^{\beta}} \quad \text{-----(1)}$$

where $\tau_{i,j}$ is the amount of pheromone on edge i,j; α is a parameter to control the influence of $\tau_{i,j}$, $\Delta\tau_{ij}(t)$ is the desirability of edge i,j (typically $1/L_{i,j}$), β is a parameter to control the influence of $\Delta\tau_{ij}(t)$

$$\tau_{ij}(t) \leftarrow (1 - \rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \qquad ----- (2)$$

where τ_{ij} is the amount of pheromone on a given edge i,j; ρ is the rate of pheromone evaporation and $\rho\ \varepsilon(0,\ 1]$ is the pheromone trail decay coefficient.

$\Delta\tau_{ij}$ is the amount of pheromone deposited, typically given by:

$$\Delta\tau_{ij}(t)= \begin{cases} 1/L_k & \text{if ant k travels on edge i,j} \\ 0 & \text{elsewhere} \end{cases} \qquad ------ (3)$$

where L_k is the cost of the kth ant's tour (typically length). The Pheromone deposit and update concept make this probabilistic and heuristic search a more efficient result. The Associated Rule Mining Algorithm [14] is used to cluster rules (pruning), so that we can avoid redundancy in rules. It uses a Rule cover based Algorithm. Based on the support and confidence value the rules will be grouped. The best among the rules can be evaluated by the Partial Swarm Optimization Algorithm.

3.2.2 Pruning of the Rules using ARM

The Association mining is used for the pruning of the rules. Based on the support and confidence value we obtain the matching patterns of rules. ARM is the most popular knowledge discovery technique. In ARM, large number of Association rules or patterns or knowledge is generated from the large volume of dataset. An Association Rule is of the form, X_i->Y; where i=1,2,…n. The Support and Confidence can be evaluated by,

$Support = P(X \cup Y) = P(XY)$

= (Number of parameters that contains both X and Y) / (Total-number of parameters in Data set) and

$Confidence = P(Y \mid X) = P(X \cup Y) / P(X) = P(XY) / P(X)$

In the ARM, many other interesting measures rather than support and confidence are used. The role of this evaluation is to get the best in the grouping of the rules and to avoid the false choosing of rules and to get more accuracy in the prediction procedure. Table3 listed out some of the interesting measures used in ARM for clustering rules.

Table 3: List of interesting Measures used in ARM

Sl. No.	Measurement	Formula
1	Support	$P(XY)$
2	Confidence/Precision	$P(Y \mid X)$
3	Coverage	$P(X)$
4	Prevalence	$P(Y)$
5	Recall / Sensitivity	$P(X \mid Y)$
6	Specificity-1	$P(\neg Y \mid \neg X)$
7	Accuracy	$P(XY) + P(\neg X \neg Y)$
8	Lift/Interest	$P(Y|X)/P(Y)$ or $P(XY)/P(X)P(Y)$
9	Leverage-1	$P(Y|X) - P(X)P(Y)$
10	Added Value / Change of Support	$P(Y|X) - P(Y)$
11	Relative Risk	$P(Y|X)/P(Y|\neg X)$

This measures help to group the rules generated into clusters based on the various features. Here the representative rules among the clustered rules are also taken so that the redundancy among the rules can be avoided. Representative rules mean that which contain all the rules in cluster but no rules will be repeated.

After the clustering of the generated rules, the rules are ranked and the best among the rules are generated using the Partial Swarm Optimization Algorithm. PSO helps to get the best among the generated rules and which helps the decision makers to take appropriate decision more easily and give more flexibility in decision making.

3.2.3 PSO Algorithm

PSO algorithm can be stated as:
1. For each Rule
2. Calculate the fitness values;
3. If the fitness value is better than the best fitness value (R*best*) in history;
4. Set the current value as the R*best*;
5. End
6. Choose the rule with the best fitness value of all the rules in the gBest
7. For each rule
8. Calculate rule velocity according to velocity equation; Update rule position according to position equation;
9. End
10. While max iteration or min error criteria are not attained;

By combining the concept of data mining, Ant Colony Algorithm, Association Rule Mining and PSO, we can generate successful rules and can also find the influence of the parameters in the decision making support systems.

4. EXPERIMENTAL DESIGN AND PREDICTED RULES

The historic data are analysed by the experts and are coded and given to the Ant Miner Algorithm. Here the ant agent starts its journey based on the high probabilistic and heuristic functions. Based on journey the rules get generated and the pheromone trails also get updated. From the generated rules, the clusters are formed based on the support and confidence value. Also the best among the rules are also evaluated by using the PSO algorithm.
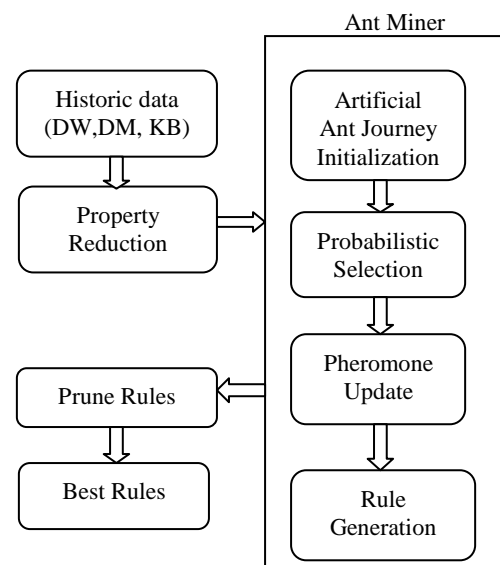

Fig.1: The experimental design of framing qualitative BP Rules

[]Some of the rules that are generated by this are described in the following table below Table 4. The generated rules help the decision makers to make efficient decision making regarding the qualitative parameters influence in the business.

Table 4: The descriptions of the rules generated from inductive learning methods

Rule	Description
Rule 1	IF PS is high and IN is high and CQ is medium and AS is high then NonBankruptcy
Rule 2	IF PS is high and CP is medium and MR is high and EO is medium THEN Non Bankruptcy
Rule 3	IF IN is low and FF is low and MR is low and CP is low then Bankruptcy
Rule 4	IF FF is low and CO is low and MR is medium THEN Nonbankrupt

These generated rules are clustered based on the support and confidence value. And also we evaluate the other interesting measures mentioned in Table 3. A support of more than 60% and a confidence of 30% above are used in clustering of generated rules.

5. CONCLUSION

In this study we are proposing a decision support system for framing qualitative Bankruptcy Prediction rules. For this we are using Ant Colony Optimization technique for the generation of the qualitative bankruptcy prediction rules. Based on the special features of heuristic and pheromone trail in ACO helps to avoid the overlapping of the generated rules and also give a positive feedback in rule generation. The generated rules can be grouped using the concept of Association Rule Mining technique and can avoid the redundancy of rules. The concept of representative rules gives more flexibility in decision making. Partial Swarm Optimization which is a better technique for generating the best among a group is used here for generating the best rules. In the future work, we can rank the parameters using this method. Here we only use qualitative factors; this work can be extended using the combination of qualitative and the quantitative parameters bankruptcy prediction.

REFERENCES

[1] Altman.E, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy". Journal of Finance, 23(3), 589–609, (1968).

[2] Altman, E. I., Marco, G., & Varet, F, " Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks", Journal of Banking and Finance, 18, 505–529, 1994.

[3] Myoung-Jong Kim, Ingoo Han , "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms" , Journal of Elsevier, Expert Systems with Applications 25, 2003, 637–646.

[4] Yannis Marinakis,Michael Doumpos,Constantin Zopounids , "Ant colony and particle swarm optimization for financial classification problem" , Published in Elsevier, 2009.

[5] Viorel Milea and Jan van den Berg , "An Ants Algorithm for Bankruptcy Prediction" ,IEEE, 2005.

[6] Kyung-Shik Shin, Yong-Joo Lee , "A genetic algorithm application in bankruptcy prediction modelling", Journal of Elsevier, Expert Systems with Applications, 2002.

[7] Ning Che, Bernardete Ribeiro, Armando S. Vieira, João Duarte and João C. Neves . "Hybrid Genetic Algorithm and Learning Vector Quantization Modeling for Cost-Sensitive Bankruptcy Prediction", IEEE, 2010.

[8] Ning Chen , Bernardete Ribeiro, Armando S. Vieira, João Duarte, João C. Neves . "A genetic algorithm-based approach to cost-sensitive bankruptcy prediction", Journal of Elsevier, Expert Systems with Applications ,2011.

[9] P. Ravisankar, V. Ravi, I. Bose , "Failure prediction of dotcom companies using neural network–genetic programming hybrids" , Journal of Elsevier, Expert Systems with Applications .2010.

[10] Hyunchul Ahn, Kyoung-jae Kim ," Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach" , Journal of Elsevier, Expert Systems with Applications , 2009.

[11] Sung-Hwan Min, Jumin Lee, Ingoo Han , "Hybrid genetic algorithms and support vector machines for bankruptcy prediction", Journal of Elsevier, Expert Systems with Applications, 2010.

[12] David Martens ,Tony Van Gestel , Manu De Backer , Raf Haesen , Jan Vanthienen , Bart Baesens , "Credit Rating Prediction Using Ant Colony Optimization" , Journal of Elsevier, Expert Systems with Applications, 2008.

[13] Jie Sun, Hui Li , "Financial distress early warning based on group decision making" Journal of Elsevier, Expert Systems with Applications, 2007.

[14] David Martens, Bart Baesens, Tom Fawcett, 'Editorial survey: swarm intelligence for data mining', Journal of Springer, Mach Learn (2011)82: 1–42.

[15] Yannis Marinakis, Magdalene Marinaki, Michael Doumpos, Constantin Zopounidis, 'Ant colony and particle swarm optimization for financial classification problems', Journal of Elsevier, Expert Systems with Applications 36, 2009. 10604–10611

[16] Ervin L. Black, F. Greg Burton and Peter M. Johnson, "Qualitative Factors as Determinants of Continued Success: An Examination of eBusiness Entrepreneurial Firms Using the New Venture Template™", The Journal of Entrepreneurial Finance, 2009, Volume 13, Issue 2, ISSN: 1551-9570.

[17] Jan Vandeweyer, "Electre Tri as a method for bankruptcy prediction", Departement Handelswetens chappen en Bedrijfskunde Strategisch KMO- en Retail management International Business, Academiejaar 2004-2005.

[18] LUAN Fugui, LI Jia, " The Construction of Assessment System: Bankruptcy Reorganization Value of Corporate", Orient Academic Forum, 2009.

[19] P.Ravi Kumar, V.Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques", Elsevier, Expert Systems with Applications 2006.

[20] S.Kannan and R.Bhaskaran, "Association Rule Pruning based on Interestingness Measures with Clustering", IJCSI International Journal of Computer Science Issues, 2009, Vol. 6, No. 1.

[21] Marco Dorigo, Luca Maria Gambardella, "Ant colonies for the traveling salesman problem", Accepted for publication in BioSystems, 1997.

[22] V.Selvi ans Dr. R.Umarani, "Comparative Analysis of Ant Colony and Partic le Swarm Optimization Techniques", International Journal of Computer Applications, Volume 5– No.4, August 2010, 0975 – 8887.