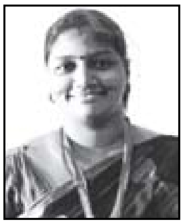
# A Framework to Develop Qualitative Bankruptcy Prediction Rules Using Swarm Intelligence

- A. Martin* - T. Miranda Lakshmi** - V. Prasanna Venkatesan***



**Abstract**

In this paper a model is proposed for qualitative bankruptcy prediction using the concepts of Ant Colony Algorithm and Particle Swarm Optimization. This work uses Ant Miner (AM) algorithm to generate qualitative Bankruptcy Prediction (BP) rules and this generated rules are pruned using the Association Rule Miner (ARM) algorithm. ARM helps to avoid overlapping and redundancy of rules. From the generated rules the best rules are categorized using Particle Swarm Optimization (PSO) algorithm. The property reduction is a technique used to find best parameters. The concept of heuristic and probabilistic features of Ant Miner Algorithm increases the prediction accuracy of Bankruptcy Prediction. The self organizing and coordination concept in swarm intelligence provides more accuracy in qualitative bankruptcy prediction.

**Keywords**: Ant Miner, Particle Swarm Optimization, Qualitative Bankruptcy Prediction, Heuristic function, Association Rule Mining Algorithm

## Introduction

Ant Colony Optimization is an agent-based technique which stimulates the natural ant behavior to obtain the optimal solution. The main idea of the ACO was inspired by the real ants. For each move, the ant constructs a solution. They have an advantage over simulated annealing and genetic algorithm approaches of similar problems when the problem space may change dynamically; the ant colony algorithm can run continuously and adapt to changes in real time [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 and based on this we can take decisions.

The non financial information is called as qualitative factors. But Qualitative data are subjective, that is, based on the subjective knowledge only we can take decisions. For the subjective knowledge we need help from the expert. Qualitative factors can be measured using expert analysis. For qualitative bankruptcy prediction very few models only are available to predict bankruptcy.

In our work we are using Ant Colony Optimization Algorithm for generating Qualitative BP rules. First the property reduction of the parameters are done using ant colony algorithm, here we get the best parameters and the removal of least important parameters can also be done here. These reduced parameters are submitted to generate rules. The rules are generated using Ant Miner algorithm and the generated rules are clustered using the Association Rule Mining Algorithm and the best rules among this are

*Research Scholar, Department of Banking Technolgoy, Pondicherry University, Puducherry, India. jayamartin@yahoo.com
**Asst. Prof., Dept. of Computer Science,St. Joseph's College (Autonomous), Cuddalore, Tamilnadu, India. cudmiranda@gmail.com
***Associate Professor, Department of Banking Technology, Pondicherry University, Puducherry, India. indiaprasanna_v@yahoo.com

extracted using the Partial Swarm Optimization technique. B y using these techniques 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 describes the proposed system, Section 4 predicts rules in qualitative BP and Section 5 discusses the conclusion and future research of the work.

## 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. These 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] used to generate bankruptcy prediction rules. These rules are generated using six internal risk factors and the associated external factors and rated as positive, negative and average based on the risk factor [6,8,10,11]. In this bankruptcy prediction the genetic operations crossover and mutation are applied and the generated rules are evaluated based on the accuracy and coverage. And the maximum of accuracy and coverage is taken as parameters of fitness function to evaluate the rules. But the generated rules b y 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 grey 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 quantitative Bankruptcy Prediction. It is applied in various classification problems. There are various works that undergone in ACO [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 is 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.

### *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 analyze 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 man y 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 has 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.

## Proposed Research

In this proposed system, the qualitative risk factors are listed out and the parameters are submitted to expert's judgment to rate the parameter. The parameter rating is done to identify the influence of the risk factors to the success of the firm. Based on the expert's judgment the parameters undergone the property reduction procedure through which the least important parameters can be avoided and the ranking of the parameters can be done. In this work we are using ant colony algorithm for property reduction. From the reduced risk factors the qualitative prediction rules has to be generated. For generating BP rules we are using ant miner algorithm and this helps to generate more accurate rules using the properties of heuristic function and pheromone trails used in the ant colony optimization technique.

### *Swarm Intelligence*

Swarm intelligence studies the collective behavior of systems composed of many individuals interacting locally with each other and with their environment. Swarms inherently use forms of decentralized control and self organizations to achieve their goals. Ants communicate only indirectly through their environment by leaving behind a substance called pheromone. This indirect communication is called as stimergy. PSO is used here since it is a better technique in swarm intelligence to get the best rules.

### *Qualitative factors used*

The qualitative risk factors that used in our work include the Industrial Risk (IR), Management Risk (MR), Financial Flexibility (FF), Credibility (CR), Competitiveness (CO), and Operational Risk (OP), Problem Solving Capacity (PS), Initiative (IN), Effective Orientation (EO), Concern of Cost (CC), Concern of People (CP), Concern of Quality (CQ), Information Seeking (IS), Cross Selling (CS) and Assertiveness (AS).

These qualitative factors are listed and questionnaires are made based on 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 more 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 may not 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 here:

a. Did *Govt. Policies and International Agreements* affect Bankruptcy?

| | |
|---|---|
| Good | It has more influence over the business |
| Medium | It has some influence over the business |
| Bad | It has less influence over the business |

b. Whether the *degree of competition* affects the Bankruptcy?

| | |
|---|---|
| Good | It has more influence over the business |
| Medium | It has some influence over the business |
| Bad | It has less influence over the 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 Table1 shows the coding of the risk factors based on the expert judgment.

**Table 1. Coding of Risk Factors Based on Questionnaires**

| Qualitative Risk Factors | Good | Medium | Bad | → | Coding |
|---|---|---|---|---|---|
| *IR* | 1 | 0 | 0 | | 100 |
| *MR* | 0 | 1 | 0 | | 010 |
| *FF* | 1 | 0 | 0 | | 100 |
| *CR* | 0 | 0 | 1 | | 001 |

| Class | Bankruptcy | Non-Bankruptcy | → | Coding |
|---|---|---|---|---|
| Actual Class | 1 | 0 | | 10 |

The proposed system is divided into five phases, while in the implementation phase. The implemented phases are listed below

- Parameter Analysis and Expert Judgment
- Parameter Reduction
- Training Set Generation
- Rule Generation
- Performance Reporting

### *Parameter Analysis and Expert Judgment*

In this phase, we listed out the risk factors from the various banks that cause bankruptcy. And the listed parameters are rated based on expert judgment. The rating is done as good, average and bad based on the influence in the bankruptcy. The expert's judgment are given as the input to the developed system based on this the prediction rules are generated.

### *Parameter Reduction*

In this phase, we are reducing the analyzed parameters. The reduction is done based on the Mean Square Error value. The parameter having the least MSE will be chosen as the best factor. The best MSE valued set is used for generating rules.

### *Training Set Generation*

In this phase, the risk factors values of the internal parameters are analyzed based on the external parameter evaluation. This is done using Ant Colony Optimization. The training set is developed based on the parameters having the least MSE. The Ant with the least MSE is selected as the best set and the training set is developed.

### *Rule Generation*

Based on ACO and the least MSE value, the rules are generated. The pheromone trail updating is done based on the Ant having the least MSE value. The Ant having the least MSE value will be the best ant and the corresponding path is taken as the rule for the particular bank.

*Performance Reporting*

In this phase, the end users such as bank persons, bank customers, banking domain expert's etc can view the generated rules and can analyze the performance in reporting. The results generated are in the form of Qualitative Bankruptcy Prediction Rules.

## Ant Colony Optimization Algorithm

**Specifications**

The specifications used in the algorithm are listed below.

F-> the original sub set containing the features $f_i$, where i-1 to n.

F = { $f_1$, $f_2$, $f_3$,.... $f_n$ }, here, n=8 features.

S-> sub set with m important features, where m<n

here, m=5

therefore, S = {$f_1$, $f_2$, $f_4$, $f_6$, $f_7$}, ( randomly chosen), $S \in F$

na -> number of ants to be search thru the features $f_i$

For each ant j, the subset,

$S_j$={$s_1$,$s_2$,…$s_m$}

Example:

F = {11, 27, 58, 34, 43, 24, 10, 65}

### *ACO Algorithm Pseudo Code*

***Step1:*** Initialization:

- Set $\tau_i$=1 => $\tau$ ={1, 1, 1,...}
  - $\Delta\tau_i$= 0; i=1,2,..n; where n=8
  - set ={0, 0, 0... 0}
- Define max. no. of iterations
  - Let us assume it as 3
- Define k, where k-> best sub sets
  - k=2
- Define p, where m-p is the number of features each ant will start within the following iterations.
  - P= $\lfloor m/2 \rfloor = \lfloor 5/2 \rfloor$ =2 ( floor function)

***Step 2:*** In first iteration

- for j=1 to na, where na->4 (no. of ant assume)
  - Randomly assign a subset of m features to $S_j$
- for ( j = 1; j <= 4; j++)

  {

  $S_1$={ 11, 34, 24, 10, 65} -> $Ant_1$

  $S_2$={27, 34, 24, 10, 65} -> $Ant_2$

  $S_3$={27, 58, 43, 24, 10} -> $Ant_3$

  $S_4$={11, 27, 58, 10, 65} -> $Ant_4$

  }
- Go to step 4

***Step 4:*** Evaluate the selected sub set of each ant

- for j-1 to na ( here, let na=4)

  {

  Find $MSE_j$ of $S_j$

  $S_1$ -> $MSE_1$ -> $ant_1$ = 1.1

  $S_2$ -> $MSE_2$ -> $ant_2$ = 1.3

  $S_3$ -> $MSE_3$ -> $ant_1$ = 1.2

  $S_4$ -> $MSE_4$ -> $ant_4$ = 0.9

  }
- Sort the subsets according to their MSE.
- Update the minimum MSE & store the corresponding subset of features.

| **Sort** | *S4* | *S1* | S3 | S2 |
|---|---|---|---|---|

Store in details as

| **MSE** | **S** |
|---|---|
| 0.9 | { 11, 27, 58, 10, 65 } |

***Step 5:*** Update trail intensity, using the feature subset of best k ants

- for j=1 to k (here k-2, (i.e. $S_4$, $S_1$)
- for (j=1; j<= 2; j++)

  {

  for ( i =1; i <=8; i ++)

  {

  Calculate $\Delta\tau_i$ ( we will get some value if $f_i \in S_j$, o otherwise)

  *[ here F = { 11, 27, 58, 34, 43, 24, 10, 65}*

  *S4 = { 11, 27, 58, 10, 65}*

  $\Delta\tau$={1, 1, 1, 0, 0, 0, 1, 1}

  $\tau$i={1.75, 1.75, 1.75, 0.75, 0.75, 0.75, 1.75, 1.75} ]

  }

  Calculate $\tau_i$= p. $\tau_i$ + $\Delta\tau_i$, where p=0.75

  }

***Step 6:*** If number of iterations is less than maximum number of iterations, initialize the subsets for next iteration and goto step 3:

- for j=1 to na

  From the features of best k ants, randomly produce m-p features subset for antj, to be used for next iteration and store it in $S_j$
- for ( j=1 ; j,= 4; j+= )

  m = 5;

  p = 2;

  m – p = 3;

  for next iterations

{

$S_1$ = { 11, 24, 65 }

$S_2$ = { 27, 34, 10 }

$S_3$ = { 27, 58, 10 }

$S_1$ = { 11, 27, 10 }

}

- Go to step 3

***Step 3:*** Select the remaining p features for each ant, ( p = 2)

Already 3 in S, select 2 features here, totally 5.

for mm=m-p+1 to m ( i.e 4 to 5)

{

for j=1 to na ( here na = 4)

{

Given $S_j$, choose feature $f_i$ that maximizez $USM_i^{Sj}$

$S_j = S_j \, U\{ f_i\}$

}

}

i.e, for ( mm = 4; mm<= 5; mm++)

{

for( j = 1; j<=4; j++)

{

for( i = 1; i <= 8; i++)

{

Calculate $USM_i^{Sj}$

}

$S_j$=$S_j$ U {fi with max USM}

Therefore, $S_j$= {4 elements} in mm=4

{5 elements} in mm=5

}

}

Replace the duplicated subsets, if any, with randomly chosen subsets.

Suppose if $S_1$=$S_3$, then any other equal case replace $S_3$ with a random set.

For example:

$S_1$ ={ 11, 24, 65, 27, 10}

$S_2$ = { 27, 34, 10, 11, 43}

$S_3$ = { 27, 58, 10, 11, 24}

$S_4$ = { 11, 27, 10, 24, 65 }

Now, $S_1$=$S_4$, So randomly chose, $S_4$ = { 58, 34, 43, 10, 65}.

ACO is used to reduce the parameters, such that this ACO algorithm yields reduced parameter set and from that qualitative bankruptcy prediction rules are generated using Ant minter algorithm.

Rule generation is implemented by Ant Miner. There are various methods available for generation of rules. 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, <IR=Good>, <MR=Bad>, etc and the predicted class is of the form <Bankruptcy=yes>, <Bankruptcy=no>, etc. The algorithm is described below,

Training set = all training cases;(Qualitative factors)

WHILE ( No. of cases in the Training set > max_uncovered_cases)

i=0;

REPEAT

i=i+1;

Anti incrementally constructs a classification rule;

Prune the just constructed rule;

Update the pheromone of the trail followed by Anti;

UNTIL

(i >= No_of_Ants) or (Anti constructed the same rule as the previous No_Rules_Converg-1 Ants)

Select the best rule among all constructed rules;

Remove the cases correctly covered by the selected rule from the training set;

END WHILE

In the algorithm, the Probabilistic function 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 node (parameter) *j* when in node *i* (which also corresponds to the desirability that the arc (*i, j*) belong to the rule built by an ant) is given by.

$$P_{i,j}(t)=((\tau_{i,j}(t))^{\alpha}(\eta_{i,j}(t))^{\beta})\backslash(\Sigma(\tau_{i,j})^{\alpha}(\eta_{i,j})^{\beta}) \quad ...(1)$$

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

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

where $\tau_{ij}$ is the amount of pheromone on a given edge *i,j;*

*ρ* is the rate of pheromone evaporation also called *evaporation constant* and *ρ* ∈ (0, 1] is the pheromone trail decay coefficient.

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

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

where $L_k$ is the cost of the $k^{th}$ ant's tour (typically length).

The probability that a parameter is chosen is a function of how close the parameter is and how much pheromone already exist on that trail. Once a cycle is completed, the pheromone deposited by each ant on the complete tour is calculated $\tau_{i,j}$. The evaporation constant value is set between (0 & 1). The pheromone evaporates more rapidly for lower values.

The Quality of the Rule can be evaluated by using the equation:

Q=[TruePos/(TruePos+FalseNeg)]*[TrueNeg/(FalsePos+TrueNeg)] ...(4)

By evaluating the quality of the rules we can avoid false negative rules.

### *Features of Proposed System*

Ant Colony Algorithm can be a better technique in Qualitative Bankruptcy Prediction. Keeping track of learning is important to remember and update previously generated rules which avoid false positive and false negative.

**Expected Benefits from Proposed System**

1. More factors – best result in the bankruptcy prediction.
2. Classify the rules effectively
3. Keep track of the Rules

By using the concept of Pheromone deposit and update concept make this probabilistic and heuristic search a more efficient result. Accuracy and Efficiency can also be increased.

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.

*(a) Rule construction*

In this step, the rules get constructed using the Ant Miner 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 of nodes by the artificial ant is by the heuristic function and the probabilistic trails. Each chosen nodes are attached and when ant reaches the destination gets the rule is generated and thus generated rules are evaluated based on various interesting measures to assure the quality.

*(b) 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.

*(c) 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. The purpose of the local pheromone update rule is to make "the visited edges less and less attractive as they are visited by ants, indirectly favoring the exploration of not yet visited edges. As a consequence, ants tend not to converge to a common path". The purpose of the global pheromone update rule is to encourage ants "to search for paths in the vicinity of the best tour found so far". 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 acts as a memory to the artificial ant and helps to keep track of the learning.

### *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.

### *Rule Framing Using the Ant Miner Algorithm*

```
Training set = all training cases; (Qualitative factors)
WHILE
    (No. of cases in the Training set >  max_uncovered_cases)
    i=0;
REPEAT
    i=i+1;
    Ant i incrementally constructs a classification rule;
    Prune the just constructed rule;
    Update the pheromone of the trail followed by Ant i;
UNTIL
    (i >= No_of_Ants) or (Anti constructed the same rule as
    the previous No_Rules_Converg-1 Ants)
```

Select the best rule among all constructed rules;

Remove the cases correctly covered by the selected rule from the training set;

END WHILE

In this 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.

$$P_{i,j}(t) = ((\tau_{i,j}(t))^{\alpha} (\eta_{i,j}(t))^{\beta}) \backslash (\Sigma(\tau_{i,j})^{\alpha} (\eta_{i,j})^{\beta}) \quad ...(1)$$

Where $\tau_{i,j}$ is the amount of pheromone on edge i,j; α is a parameter to control the influence of $\tau_{i,j;}$ $\eta_{i,j}$ is the desirability of edge i,j (typically $1/d_{i,j}$ ), β is a parameter to control the influence of $\eta_{i,j}$

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

Where $\tau_{ij}$ is the amount of pheromone on a given edge *i,j;* $\rho$ is the rate of pheromone evaporation and $\rho \in (0, 1)$ is the pheromone trail decay coefficient.

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

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

where $L_k$ is the cost of the $k^{th}$ 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.

### *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 generate the 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 tuples that contains both *X* and *Y* ) / (Total-number of tuples 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. Table 2 listed out some of the interesting measures used in ARM for clustering rules. 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.

**Table 2: 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 | $P(Y\|X)/P(Y\|\neg X)$ |

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.

### *PSO Algorithm*

For each Rule

Initialize the Rule

End

Do

For each Rule

Calculate the fitness values;

If the fitness value is better than the best fitness value (pBest) in history;

Set the current value as the pBest;

End

Choose the rule with the best fitness value of all the rules in the gBest

For each rule

Calculate rule velocity according to velocity equation;

Update rule position according to position equation;

End

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.

## Experimental Design and Predicted Rules

The historic data are analyzed 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.

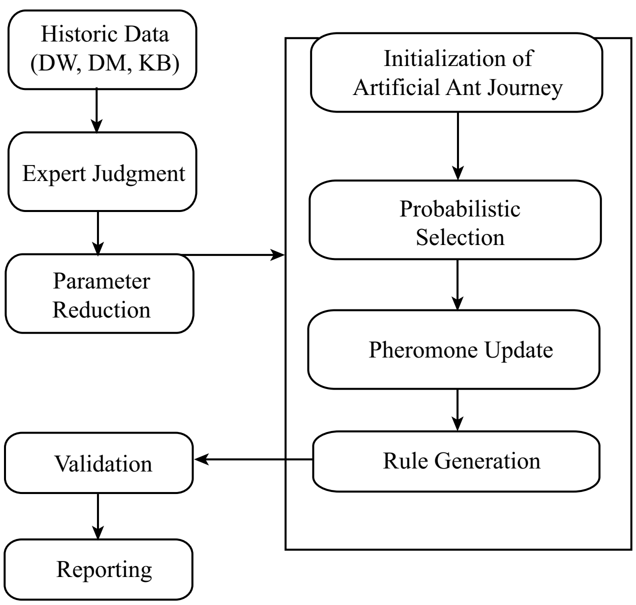



**Figure 1. Design Model of PSO to generate Qualitative Bankruptcy Prediction**

Some of the rules that are generated by this are described in the following table below Table 3. The generated rules are helping in predicting qualitative bankruptcy. These generated rules are clustered based on the support and confidence value.

| Bankid | MR | FF | CR | CC | OP | Result |
|---|---|---|---|---|---|---|
| 1253 | good | well | well | good | good | Not Possible |
| 1359 | Bad | fine | good | well | Bad | Possible |
| 1458 | good | well | good | fine | Bad | Not Possible |
| 1528 | good | good | good | good | good | Not Possible |
| 1569 | Bad | good | good | well | good | May Possible |

**Table 3: 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 Non Bankruptcy |
| 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 Non bankrupt |

From these rules qualitative bankruptcy prediction analysis is conducted for banks, the results are shown Table 3. The results are shown more 80% successful in indentifying not successful banks and 90% successful in identifying successful banks. We also evaluate the other interesting measures mentioned in Table 2. A support of more than 60% and a confidence of 30% above are used in clustering of generated rules. The prediction accuracy of the rule should be improved in by analyzing true negatives and false positives.

## 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.

## Reference

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",Elsevier, Expert Systems with Applications 25, pp637–646, 2003.

4. Yannis Marinakis,Michael Doumpos,Constrantin 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", 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", Elsevier, Expert Systems with Applications (2011)
9. P. Ravisankar, V. Ravi, I. Bose, "Failure prediction of dotcom companies using neural network–genetic programming hybrids",Elsevier, Expert Systems with Applications (2010)
10. Hyunchul Ahn, Kyoung-jae Kim," Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach", Elsevier, Expert Systems with Applications (2009)
11. Sung-Hwan Min, Jumin Lee, Ingoo Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction", 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", Elsevier, Expert Systems with Applications (2008)
13. Jie Sun, Hui Li, "Financial distress early warning based on group decision making" Elsevier, Expert Systems with Applications (2007)
14. David Martens, Bart Baesens, Tom Fawcett, 'Editorial survey: swarm intelligence for data mining', Springer, Mach Learn (2011)82: PP 1–42.
15. Yannis Marinakis, Magdalene Marinaki, Michael Doumpos, Constantin Zopounidis, 'Ant colony and particle swarm optimization for financial classification problems', Expert Systems with Applications 36, (2009), pp 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 Volume 13, Issue 2, ISSN: 1551-9570, 2009
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, Vol. 6, No. 1, 2009
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 (0975 – 8887) Volume 5– No.4, August 2010