Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 3, No. 3, August 2012), Pages: 53-64
www.jacr.iausari.ac.ir

Fuzzy Apriori Rule Extraction Using Multi-Objective Particle Swarm Optimization: The Case of Credit Scoring

Mohammad Reza Gholamian[1]✉, Seyed Mahdi Sadatrasoul[1], ZeynabHajimohammadi[2]

(1) School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
(2) Department of computer science, Amirkabir University of Technology, Tehran, Iran

Gholamian@iust.ac.ir;Sadatrasoul@iust.ac.ir; Zhajimohammadi@aut.ac.ir

Received: 2012/01/16; Accepted: 2012/07/15

Abstract

There are many methods introduced to solve the credit scoring problem such as support vector machines, neural networks and rule based classifiers. Rule bases are more favourite in credit decision making because of their ability to explicitly distinguish between good and bad applicants.In this paper multi-objective particle swarm is applied to optimize fuzzy apriori rule base in credit scoring. Different support and confidence parameters generate different rule bases in apriori. Therefore Multi-objective particle swarm is used as a bio-inspired technique to search and find fuzzy support and confidence parameters, which gives the optimum rules in terms of maximum accuracy, minimum number of rules and minimum average length of rule. Australian, Germany UCI and a real Iranian commercial bank datasets is used to run the algorithm. The proposed method has shown better results compared to other classifiers.

Keywords: *Credit scoring, Banking, Fuzzy association rules, Apriori, multi-objective particle swarm*

1. Introduction

Credit scoring is widely used in today's competitive banking industry. Every day, individual and company's records of past borrowing and repaying actions are gathered and analyzed by information systems. Banks use this information to determine the individuals and companies potential profitability. The process of lending can be divided into four main stages and depending on each stage and the different situations, different kind of scoring exists and can be summarized as follows[1]:

- **Pre application:** Response score is the main score in this stage, and it refers to probability that a potential customer will react to a marketing campaign, e.g., a direct mailing for a new product.
- **Application:** Fraud score and application (credit) score are the scores in this stage. Fraud scoring rank the applicants according to the likelihood that they are fraudulent and credit scoring refers to the assessment of the credit worthiness for new applicants; the latter score is considered throughout this paper.
- **Performance:** Performance score, behavioral score, retention score andearly-warning score are the scores in this stage, among them, Behavioral score is the

most important, and it is about computing the default probability based upon the repayment behavior of the existing bank customers.

- **Collection:** In this stage, the main score is collection scoring. Collection scoring is used to divide customers with different levels of loan repayment into different groups.

The combined information of the abovementioned scores is used for profit scoring. In this paper, we address the credit scoring problem. Credit scoring is used to answer one key question - what is the probability of default within a fixed period, usually one year.

There are many methods suggested to perform classification in the credit scoring problems include statistical and intelligent methods. Logistic regression is the most favorite statistical and traditional method used to assess the credit score [2]. Linear discriminant analysis is also applied and it's shown that it is as efficient as logistic regression [3]. There are also many intelligent methods applied in the problem include neural networks, support vector machines, Bayesian networks, case based reasoning (CBR), decision trees, etc. Among intelligent methods neural network and support vector machines are used widely and owing to their nonlinear fitness and generalization capabilities, better classified the UCI credit datasets[4-6]. Some studies have shown that neural networks, SVM, decision trees and other intelligent methods, are superior to statistical methods [7-9].

In recent years, hybrid methods are also proposed and they are the focus of many researchers. Hybrid techniques usually use different algorithm's strengths to improve the other's weaknesses. In some hybrid methods, both statistical and intelligent methods are used together. There are so many hybridization algorithms used throughout the literature. A hybrid neural discriminant technique with BP neural network and discriminant analysis proposed, and showed better accuracy than the BP neural network and discriminant analysis[10]. A two-stage hybrid procedure with artificial neural networks and multivariate adaptive regression is also proposed[11]. In a study hybrid approaches are divided into four main areas and different combination of clustering algorithms and classifiers are tested; logistic regression and neural network hybrid shown the best accuracy[12].

There are also studies which hybrid meta-heuristic methods with intelligent methods. An integration of support vector machines, genetic algorithms and F-score is studied[6]. In the last decade, using ensemble methods increased in the area, and in some cases they give better accuracy rate[13, 14]. Neural network ensemble strategies include cross validation, bagging and boosting for financial decision applications are studied and shown better accuracy rate and generalization ability[13]. Ensemble learning is still an open issue in recent year's studies[15, 16].

Because of robustness and transparency needs and also the auditing process done by regulators in different countries on the credit scoring, Banks cannot use many of the mentioned methods[17]. On the other hand, by using rule bases, banks can easily interpret the results and explore the rejecting reasons to the applicant and regulatory auditors.

In the field of rule-based credit scoring, there is actually a little literature on the area. Ben-Davide provides a new method for rule pruning and examined his method on the credit scoring data set[18]. Hoffmann et.al introduced a new learning method for fuzzy rule induction based on the evolutionary algorithms[19].Martens et. al. used the support vector machine for rule induction in the credit scoring problems[20]. Malhotra et.al. used the adaptive neuro fuzzy inference systems(ANFIS) for rule induction and showed that this method works betters from discriminant analysis on their own credit scoring dataset, which is gathered from credit unions[21], they used the back propagation method to learn their rules membership function to fit to the data. Baesens et.al. used and evaluate three neural network rule extraction techniques include Neurorule, Trepan, and Nefclass, for rule extraction in three real-life data bases include German credit database, Bene1 and Bene2 credit database. They showed Nero'srule and Trepan yield better classification accuracy compared to the C4.5 algorithm and the logistic regression. Finally, they visualize the extracted rule sets using thedecision table[22].

As mentioned, there are many works on the literature, which extract both crisp and fuzzy rules in the credit scoring area. Fuzzy rules are more attractable, easier to interpret and robust because the rules are expressed in terms of linguistic variables, which are usually used by the experts. Fuzzy apriori is a method of inducing fuzzy association rules from datasets. This study mainly focuses on inducing optimized fuzzy apriori rules using theMulti-objective particle swarm algorithm.

This study addresses the following research area; with the aim of achieving the most compact rule base in term of number of rules and average rule length at the valuable accuracy rate.These rules can be used as the rules of the thumb by banks and financial institutes, and none of the published works seen this aspect of credit scoring yet.First, the credit data are fuzzified using a fuzzification method, then fuzzy association rules are induced using apriori rule induction method. Because the quality of rule bases are measured based on the fewer number of rules, average rule length and higher performance simultaneously and always there is a tradeoff between these three, multi-objective fitness function is used to acquire the best rule base. Searching the fuzzy support and fuzzy confidence to reach the best rule base is done using continuous particle swarm algorithm. The experiments established using the Germany and Australian credit data set of UCI and a real dataset from a major Iranian bank.

The rest of this study is divided into four major parts: section 2 describes the proposed method. Section 3 introduces the data, experiments setting and results and finally study concluded in section 4.

2. The Proposed Classification Method

Particle Swarm Optimization (PSO), is a new population-based evolutionary computation technique which was first introduced in 1995 by Eberhart and Kennedy [23].PSO is an efficient global optimization algorithm and since then because of its very good results and low computational costs, has been applied to many nonlinear function optimizations include neural network, support vector machines parameters training and other algorithms. In particle swarm, a swarm of Np particles search a D dimension of solution space to find the optimum answer.

The success of single objective PSO in different problems motivates the researchers to extend the use of PSO in multi-Objective problems. There are different approaches for multi-objective PSO (MOPSO), include aggregation, lexicographic ordering, sub-population, Pareto-based and others[24]. In aggregation approach, combining all of the objective functions into a single one is considered [25]. In this paper, a conventional weighted aggregation (CWA) is applied in which a linear fixed weights aggregation is used. According to this approach, a fixed weighted sum of objectives is considered $F = \sum_{i=1}^{k} w_i f_i(x)$, where $w_i, i = 1,2 \ldots, k$ are non-negative weights. In fact, a priori knowledge is fed to the problem by defining the importance of each objective and only a single Pareto optimal point is obtained from each run[26].

In this paper, the solution space for finding fuzzy association rules is a 2-dimension space; first dimension is the fuzzy support and the second dimension is the fuzzy confidence. Particle swarm is used to find the fuzzy support and confidence, which give the optimized rule base in terms of higher accuracy, lower number of rules and average rule length.

An example of a credit decision fuzzy association rule can be explained as follows:

If $A_{\text{high}}^{\text{income}}$ and $A_{\text{medium}}^{\text{previous credit}}$ then $A_{\text{high}}^{\text{credit worthy}}$ [with CG (certainty grade)=0.79]

It means that if the income is triangular high and previous credit is triangular medium then applicant is credit worthy with the certainty grade of 0.79.The algorithm used to find the optimized fuzzy association rules is described in figure 1. A further explanation of each step is then described.

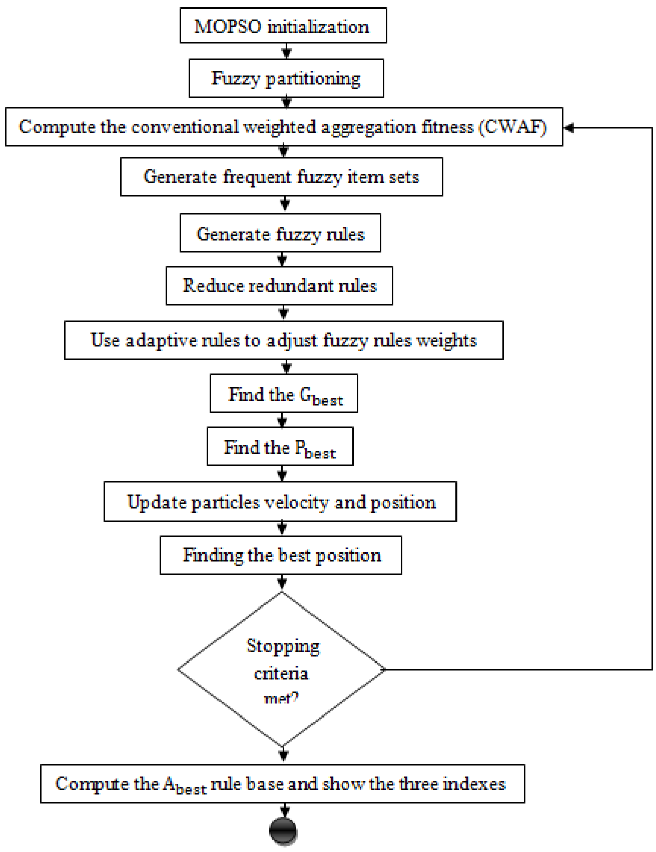

Figure1. The overall steps of the proposed credit approval decision making process.

Step1. MOPSO Initialization

Step 1.A. Initialize iteration counter t=0;

Step 1.B. Generate initial positions x_i^t for N particles and zero velocity v_i^t for each N particle. Each position is shown by an order pair $(\mathrm{fs_i}, \mathrm{fc_i})$ between zero and one. The first digit in the order shows the fuzzy support (FS) and the second digit shows the fuzzy confidence (FC). The i index shows the ith particle position in the space (i=1,2,….N);

Step 1.C. Initialize N particle best $p_{best\ i}$;

Step 1.D. Set G_{best} as the best $p_{best\ i}$.

Step 2. Fuzzy partitioning

For partitioning the data set to linguistic variables, the triangular membership functions are used. Each variable is converted to k linguistic variables. Each linguistic variable and its correspond membership function is shown as $\mathrm{A}_{\mathrm{k,i_k}}^{\mathrm{x_k}}$ and $\mu_{\mathrm{k,i_k}}^{\mathrm{x_k}}$. Where $x_k$is the linguistic variable, i_k is the i_kth linguistic variable for x_k and i_k={1,2,…,K}. In this paper it is assumed that k=3. It means that each variable is converted to three linguistic variables and can be low, medium and high. The membership degrees of attributes are computed as follows(ex. for person income attribute[27]):

$$\mu_{\mathrm{k,i_1}}^{\mathrm{income}} = \max\left\{1 - \frac{\left|\mathrm{x} - \mathrm{a}_{\mathrm{i_1}}^{\mathrm{k}}\right|}{\mathrm{b}^{\mathrm{k}}}, 0\right\} \tag{1}$$

Where $\mathrm{a}_{\mathrm{i_1}}^{\mathrm{k}} = \min_{\mathrm{income}} + \frac{(\max_{\mathrm{income}} - \min_{\mathrm{income}})(\mathrm{i_1}-1)}{(\mathrm{k}-1)}$and $\mathrm{b}^{\mathrm{k}} = \frac{(\max_{\mathrm{income}} - \min_{\mathrm{income}})}{(\mathrm{k}-1)}$.

Step 3. Compute the conventional weighted aggregation fitness (CWAF)

Decode s_i,c_i, and generate rules for current particle positions through Step 4–7. In this step conventional weighted aggregation (CWA) is applied, and a linear fixed weights aggregation is used. Compute the fitness value $f(r_i)$ for each position. Each position gives a set of rules r_i , and its fitness can be computed[27]:

$$F(r_i)=w_{Ac}.Accuracy(r_i) - w_g.n_{ri} - w_{Ar}.Ar_{ri} \tag{2}$$

Where "accuracy" shows the classification accuracy of ith iteration rule base, which evaluated using leave-one-out. w_{Ac} is the weight of classification accuracy, n_{ri} is the number of rules of ith iteration rule base and w_g is the weight of number of rules for rule base. Ar_{ri} isanaverage number of ith iteration rule base and $w_{Ar}$is its appropriate weight.

Rule base compactness is of interest, and is measured in terms of number of rules and average rule length in this paper. The length of the rules (number of conditions in rule antecedence) is an important issue and for a same accuracy rate, the lower the number of rule antecedences, the better the rule base is. So, an average of rule length index is defined to measure the quality of rule base length.

$$\mathrm{Ar_{ri}} = \frac{\text{number of conditions in rules antecedences}}{\text{total number of rules of the rulebase}} \tag{3}$$

Step4. Generate frequent fuzzy item sets

Frequent k-dim item sets are of interest in association rule induction. The frequent item sets recognized by “support” which can be computed [28]:

$$S\left(A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k}\right) = \frac{\sum_{p=1}^{n} z}{n} \tag{4}$$

Where z equals to one for the existing items and zero for the items, which doesn't. n is the total number of applicants. So "fuzzy support" is defined[29]:

$$FS\ (A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k}) = \frac{\sum_{p=1}^{n} \mu_{k,i_1}^{x_1} \times \mu_{k,i_2}^{x_2} \times \ldots \times \mu_{k,i_{k-1}}^{x_{k-1}} \times \mu_{k,i_k}^{x_k}}{n} \quad (5)$$

In the algorithm, the item sets, which fuzzy support is larger than or equal to the min FS are of interest, after finding the frequent item sets rules can be generated from them.

Step 5. Generate fuzzy rules

After generating items sets, it is the time to extract rules from the frequent item sets. Rule induction is done using theconfidence measure. The confidence measure for a crisp rule is defined[28]:

$$C(R) = \frac{S\left(A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k} \times A_{k,i_\gamma}^{x_\gamma}\right)}{S\left(A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k}\right)} \quad (6)$$

"fuzzy confidence" which used for fuzzy rule induction can be computed [29]:

$$FC(R) = \frac{FS\left(A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k} \times A_{k,i_\gamma}^{x_\gamma}\right)}{FS\left(A_{k,i_1}^{x_1} \times A_{k,i_2}^{x_2} \times \ldots \times A_{k,i_{k-1}}^{x_{k-1}} \times A_{k,i_k}^{x_k}\right)} \quad (7)$$

Using fuzzy confidence one can generate an effective fuzzy rule, whose fuzzy confidence is larger than or equal to the min FC.

Step 6. Reduce redundant rules

Fewer numbers of rules, in rule bases yield to a compact and better rule base. If there are $R_1, R_2, \ldots R_n$ rules with the same consequence, such that$R_1 \in R_2 \in \cdots \in R_n$, in antecedence of the rules then $R_2, \ldots R_n$are recognized as redundant rules, and can be pruned.

Step 7. Use adaptive rules to adjust fuzzy rules weights

Itwas shown that the performance of rule base systems could be improved by using and adjusting the certainty grade of rules. If a sample is correctly classified, then w_α is increased and if not $w_\alpha$decreased[30].

Set g to be zero.
Repeat
$g = g + 1$
For each sample do
Find the fuzzy rule which matches to the sample (R_α); if the sample is correctly classified then increase the rule weight by $w_\alpha = w_\alpha + \eta_i(1 - w_\alpha)$ else reduce the rule weight by $w_\alpha = w_\alpha - \eta_r . w_\alpha$
End
Until $g = g_{max}$

Where η_i and η_r are the increasing and decreasing learning rate.

Step 8. Find the G_{best}

Find the particle with maximum fitness in current particles positions and store it in the G_{best} .

Step 9. Find the P_{best}

Compare the fitness of each particle with its best position based on its previous positions; store the best position of each particle in $p_{best\ i}$.

Step 10. Update particles velocity and position

Update each particles velocity using following formula[23]:

$$\vec{v}_i(t) = \phi\vec{v}_i(t-1) + r_1 c_1\left(\vec{x}_{pbest_i} - \vec{x}_i(t)\right) + r_2 c_2\left(\vec{x}_{gbest_i} - \vec{x}_i(t)\right) \quad (8)$$

Where c_1 and c_2 are two positive constants, called cognitive learning rate and social learning rate respectively; r_1 and r_2 are random functions in the range [0, 1] ; $\phi$is inertia factor.
Update the position of each particle based on the predefined velocity[23]:

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \quad (9)$$

The positions of the particles are confined within [$x_{\min}$=0, $x_{\max}$=1]. If an element of positions exceeds the threshold $x_{\min}$or $x_{\max}$, it is punished and set equal to the corresponding threshold, which it has exceeded.

Step 11. Finding the best position

Compare the G_{best} with A_{best} , if it was better, replace it with A_{best} .

Step 12. Stopping criteria

If t_{max} generations have been reached, then terminate the execution of the algorithm, A_{best} represents the best particle and its related position is the best position. Compute the optimized rule base and its performance indexes using A_{best} . Otherwise replace CP with NP and go to step3.

3. Empirical Evaluation

Empirical evaluations are presented in this section. First, German, Australian and Iranian credit Data sets are introduced and their experimental setups are presented. Then, the results and discussions on these three datasets are brought for a selected number of classifiers in comparison with the proposed method. The comparisons are performed by means of some well-known performance measures including accuracy, average number of rules and average rule length.

3.1 Data Sets and Experiments Setup

Different datasets are used to evaluate the performance of the proposed algorithm. Australian and Germany Credit Data Sets from University of California at Irvine (UCI) Machine Learning Repository are applied. These datasets can be found at http://archive.ics.uci.edu/ml/datasets.html. An Iranian commercial bank real dataset is also used to evaluate the proposed algorithm. Table 1 shows the characteristics of the datasets.

Table1. Datasets description

Dataset name	Data size	Inputs variables		
		Total	Continuous	nominal
German UCI credit	1000	20	7	13
Australian UCI credit	690	14	6	8
Iranian real credit data set	222	50	42	8

Australian credit dataset has been successfully used for credit scoring and evaluation systems in many previous works, especially using intelligent methods [9, 13, 14, 19,31-33]. It includes 15 attributes, from which, eight attributes are categorical and six attributes are continuous. Australian dataset includes 690 instances ofloan applicants, and the data instances are labeled as classes one (worthy, 307 instances) and zero (unworthy, 384 instances).

Germany's credit dataset is also used in many works. For each applicant, the dataset includes 20 input variables describe the credit history, account balances, loan purpose, loan amount, employment status, and personal information. This data set consists of numeric attributes only and includes 1000 instances of loan applicants; the data instances are labeled as classes 1 (worthy, 700 instances) and 2 (unworthy, 300 instances).

Iranian real credit dataset is also used in the experiments. The initial dataset include 312 corporate applicants with41 financial and non-financial attributes in the period 2006 to 2009. First, we have a data cleaning stage. In general, data cleaning include removing redundant, outlier's data and missing values. There were a few missing values for some corporates, some of them lack financial data and others lack the result of their loans, in fact, they were in the process of debt repay. So 90 corporate were excluded. From remained 222 corporates, 177 were credit worthy and 45 were unworthy. Once the data cleaning process was completed, the categorical attributes include type of industry; type of company and type of book were converted to numerical attributes using dummy variables. The results and descriptions of the changes are shown in table (3) in appendix (1). By using dummy variables number of variables increased to 50.

The three datasets were used to train the proposed method, and the results were compared[1] with a selection of classifiers include rule's base and others such as ANFIS[2], JRip, C4.5, SVM[3] and MLP[4].

For MLP, a feed forward neural network with one hidden layer was considered and trained with error gradient descent using conjugate gradients. The C4.5 decision trees and MLP were run using the PR Tools Matlab Toolbox (http://www.prtools.org). The ANFIS was run several times with different squash factor and reject ratios, and the nearest results with the proposed method accuracy were reported. ANFIS were run using the ANFIS Matlab Toolbox. C-SVC type of SVM was run using the RFB (i.e. radial basis function) kernel type. No special parameter setting was done for JRip, and these two algorithms were run using Weka 3.6.5 (http://www.cs.waikato.ac.nz).

The proposed algorithm was run 30 times (for considering the diversification), and Different parameters were set for each run. In each run, a rule base was discovered and evaluated using the predefined fitness function. At last, the following parameters were selected, and the last runs are done using the tuned parameters.

1. The algorithms were run on a corei5 CPU and 4GB ram PC.
2. Adaptive neuro fuzzy inference system(ANFIS)
3. Support vector machine(SVM)
4. Multi-layer perceptron (MLP)

$$w_{Ac} = 20, w_g = 1, w_{Ar} = 1, N_p = 30, C_1 = 1, C_2 = 1, r_1 = 1, r_2 = 1, g = 100$$

It's suggested that the learning rates should be specified as $0 < \eta_i \ll \eta_r < 1$ for example $\eta_r = 0.001$and $\eta_i = 0.1$. Because compactness and accuracy rate are important simultaneously in the rule bases, the results analysis contains both of them.

Some square of weights in conventional weighted aggregation approach is fixed $\sum_{i=1}^{k} w_i = 1$[25]. It should be noted that apart from the first objective, the normalization method could not be used; since the min and max values are unknown. Meanwhile the importance of the weights could not be defined exactly, and hence they were tuned during the runs easily.

In order to compute the classification accuracy 10 fold cross validation was selected for all of the algorithms excluding ANFIS and specially for ANFIS, the data split into 2/3 for training and 1/3 for testing, and that was done because of the ANFIS tool restrictions for test options in Matlab.

3.2 Results and Discussion

Table (2) shows classification accuracy, number of rules and average rule length for three datasets. The best test set classification accuracy; the lowest number of rules and average rule length for each data set are bolded. A test set at the 5% level from the best performer using one-tailed t test wasrun. The accuracy rates, which have not been significant difference at 5% from the best accuracy in each dataset, are shown in italic mode.

As shown in table 2,rule based classifiers are not the best performers in terms of accuracy in any of the datasets. However, the main significance difference between rule based classifiers and other classifiers can be observed in German credit dataset. In Germany and Australian dataset, the MLP shows better classification accuracy whereas in Iran dataset, SVM shows the best results. The proposed method has not been significant difference from best performer's accuracy in Iranian and Australian datasets.

Table2. Classification accuracy of the proposed method versus a selection of well-known classification algorithms in different datasets

	German dataset			Australian dataset			Iranian dataset		
Method	**Accuracy**	**Number of rules**	**Average rule length**	**Accuracy**	**Number of rules**	**Average rule length**	**Accuracy**	**Number of rules**	**Average rule length**
Proposed Method	70	3	1.6	*85.6*	3	1.5	*79.65*	11	2
ANFIS	70	20	20	82	28	14	*79.29*	21	50
Ripper(JRip)	71	6	3.2	*85.4*	9	3.6	*79.5*	15	15
C4.5	72	-	-	*85.7*	-	-	*79.41*	-	-
Support vector machine(SVM)	70	-	-	*85.4*	-	-	80.12	-	-
Multilayer perception(MLP)	73	-	-	85.9	-	-	*79.5*	-	-

The number of rules in the proposed method is the lowest and there is a significant difference compared to other rule base classifiers. Average length of rules is better in the proposed method and there is a significant difference between it and other rule base classifiers. So the proposed method can be used to find a rule of the thumb to find the best rules at a valuable and acceptable accuracy rate to make credit decisions. The generated rule's antecedents mainly have two or three attributes on average. It seems that proposed method generates the most compact rule base at a valuable and reasonable accuracy rate.

4. Conclusion

In this paper, a new search method using multi-objective particle swarm algorithm was introduced for inducing fuzzy association rules withthe certainty degree for each rule. Itwas examined on different databases include Australian, Germany UCI and a real Iranian bank dataset. The Iranian bankdataset has not been used in other previously published works.

Using other classification methods as a benchmark, the results were measured and compared in different datasets.

For different datasets, the results were measured and compared with other classification methods. Theproposed method demonstratedgood and competitiveaccuracy rate in Australian and Iranian datasets with the fewest rules and lower average rule length in all three databases. The algorithm used the fixed membership functions and certainty grade to fit into the dataset, which was better for expert judgment. Owing to the rule base simplicity, the proposed algorithm can be used to find arule of the thumb for credit decision making in banks and financial institutions, especially in the absence of internal rating software.

Next researches can focus on using Gaussian functions to fuzzyify the data, enhancing the weighting method and the voting, using multiple minimum supports and confidences to find frequent item sets and rules for increasing the rules quality, and also using the hybrid Meta heuristic methods to find the better rule bases.

Appendix (1)

Attributes included in Iran credit dataset and their typesare shown in table 3.

Table3. List of variables in Iran commercial bank credit dataset

Variable	Variable type	Variable	Variable type
Net profit	Continuous	Type of industry: industry and mine (=1, other =0)	nominal
Activeininternal market	nominal	Type of industry: agricultural (=1, other =0)	nominal
number of countries that the company export to	Continuous	Type of industry: oil and chemical (=1, other =0)	nominal
Target market risk (from 1 to 5)	Continuous	Type of industry: infrastructure and service(=1, other =0)	nominal
Companyhistory(number of years)	Continuous	Type of book: Taxdeclaration(=1,other=0)	nominal
Mangers history	Continuous	Type of book: Audit Organization (=1,other=0)	nominal
Type of company: Cooperative (=1, other =0)	nominal	Type of book: Accreditedauditor (=1,other=0)	nominal
Type of company: Stock Exchange(LLP) (=1, other =0)	nominal	Inventorycash	Continuous
Type of company: PJS (=1, other =0)	nominal	Accounts receivable	Continuous
Type of company: Limited and others (=1, other =0)	nominal	Other Accounts receivable	Continuous

Variable	Variable type	Variable	Variable type
Type of company: Stock Exchange (=1, other =0)	nominal	Stock	Continuous
ExperiencewithBank(number of years in 5 categories)	Continuous	Currentassets	Continuous
Current periodsales	Continuous	Non-current assets	Continuous
Prior periodsales	Continuous	Totalassets	Continuous
Two-Prior periodsales	Continuous	Short-termfinancial liabilities	Continuous
Current periodassets	Continuous	Currentliabilities	Continuous
Prior periodassets	Continuous	Long-termfinancial liabilities	Continuous
Two-Prior periodassets	Continuous	Non-current liabilities	Continuous
Current periodshareholder Equity	Continuous	Totalliabilities	Continuous
Prior periodshareholder Equity	Continuous	Capital	Continuous
Two-Prior periodshareholder Equity	Continuous	Accumulatedgainsorlosses	Continuous
Current accounts creditor turn over	Continuous	shareholder Equity	Continuous
WeightedAverageCurrentAccoun t	Continuous	Sale	Continuous
Averageexportsover the pastthree years	Continuous	Grossprofit	Continuous
Last three yearsaverageimports	Continuous	Financialcosts	Continuous

5. References

[1] Van Gestel, T. and B. Baesens, *Credit risk management: basic concepts: financial risk components, rating analysis, models, economic and regulatory capital*. 2009: Oxford University Press, USA.

[2] Wiginton, J.C., *A note on the comparison of logit and discriminant models of consumer credit behavior.* Journal of Financial and Quantitative Analysis, 1980. 15(03): p. 757-770.

[3] Harrell, F.E. and K.L. Lee, *A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality.* Biostatistics: Statistics in Biomedical; Public Health; and Environmental Sciences. The Bernard G. Greenberg Volume. New York: North-Holland, 1985: p. 333–343.

[4] DESAI, V.S., et al., *Credit-scoring models in the credit-union environment using neural networks and genetic algorithms.* IMA Journal of Management Mathematics, 1997. 8(4): p. 323.

[5] Lee, Y.C., *Application of support vector machines to corporate credit rating prediction.* Expert Systems with Applications, 2007. 33(1): p. 67-74.

[6] Huang, C.L., M.C. Chen, and C.J. Wang, *Credit scoring with a data mining approach based on support vector machines.* Expert Systems with Applications, 2007. 33(4): p. 847-856.

[7] Crook, J.N., D.B. Edelman, and L.C. Thomas, *Recent developments in consumer credit risk assessment.* European Journal of Operational Research, 2007. 183(3): p. 1447-1465.

[8] Huang, Z., et al., *Credit rating analysis with support vector machines and neural networks: a market comparative study.* Decision support systems, 2004. 37(4): p. 543-558.

[9] Ong, C.S., J.J. Huang, and G.H. Tzeng, *Building credit scoring models using genetic programming.* Expert Systems with Applications, 2005. 29(1): p. 41-47.

[10] Lee, T.S., et al., *Credit scoring using the hybrid neural discriminant technique.* Expert Systems with Applications, 2002. 23(3): p. 245-254.

[11] Lee, T.S. and I. Chen, *A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines.* Expert Systems with Applications, 2005. 28(4): p. 743-752.

[12] Tsai, C.F. and M.L. Chen, *Credit rating by hybrid machine learning techniques.* Applied soft computing, 2010. 10(2): p. 374-380.

[13] West, D., S. Dellana, and J. Qian, *Neural network ensemble strategies for financial decision applications.* Computers & operations research, 2005. 32(10): p. 2543-2559.

[14] Tsai, C.F. and J.W. Wu, *Using neural network ensembles for bankruptcy prediction and credit scoring.* Expert Systems with Applications, 2008. 34(4): p. 2639-2649.
[15] Louzada, F., et al., *Poly-bagging predictors for classification modelling for credit scoring.* Expert Systems with Applications: An International Journal, 2011. 38(10): p. 12717-12720.
[16] Finlay, S., *Multiple classifier architectures and their application to credit risk assessment.* European Journal of Operational Research, 2010.
[17] Thomas, L.C., *Consumer credit models: pricing, profit, and portfolios*. 2009: Oxford University Press, USA.
[18] Ben-David, A., *Rule effectiveness in rule-based systems: A credit scoring case study.* Expert Systems with Applications, 2008. 34(4): p. 2783-2788.
[19] Hoffmann, F., et al., *Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms.* European Journal of Operational Research, 2007. 177(1): p. 540-555.
[20] Martens, D., et al., *Comprehensible credit scoring models using rule extraction from support vector machines.* European Journal of Operational Research, 2007. 183(3): p. 1466-1476.
[21] Malhotra, R. and D.K. Malhotra, *Differentiating between good credits and bad credits using neuro-fuzzy systems.* European Journal of Operational Research, 2002. 136(1): p. 190-211.
[22] Baesens, B., et al., *Using neural network rule extraction and decision tables for credit-risk evaluation.* Management Science, 2003. 49(3): p. 312-329.
[23] Kennedy, J. and R. Eberhart. *Particle swarm optimization*. 1995: IEEE.
[24] Reyes-Sierra, M. and C.A.C. Coello, *Multi-objective particle swarm optimizers: A survey of the state-of-the-art.* International Journal of Computational Intelligence Research, 2006. 2(3): p. 287-308.
[25] Parsopoulos, K.E. and M.N. Vrahatis. *Particle swarm optimization method in multiobjective problems*. 2002: ACM.
[26] Jin, Y., M. Olhofer, and B. Sendhoff, *Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how?* 2001.
[27] Hu, Y.C., R.S. Chen, and G.H. Tzeng, *Finding fuzzy classification rules using data mining techniques.* Pattern Recognition Letters, 2003. 24(1-3): p. 509-519.
[28] Han, J., M. Kamber, and J. Pei, *Data mining: concepts and techniques*. 2011: Morgan Kaufmann Pub.
[29] Ishibuchi, H. and T. Nakashima, *Effect of rule weights in fuzzy rule-based classification systems.* Fuzzy Systems, IEEE Transactions on, 2001. 9(4): p. 506-515.
[30] Nozaki, K., H. Ishibuchi, and H. Tanaka, *Adaptive fuzzy rule-based classification systems.* Fuzzy Systems, IEEE Transactions on, 1996. 4(3): p. 238-250.
[31] Wang, G., et al., *A comparative assessment of ensemble learning for credit scoring.* Expert Systems with Applications, 2010. 38(1): p. 223-230.
[32] Hsieh, N.-C., *Hybrid mining approach in the design of credit scoring models.* Expert Systems with Applications, 2005. 28(4): p. 655-665.
[33] Nanni, L. and A. Lumini, *An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring.* Expert Systems with Applications, 2009. 36(2, Part 2): p. 3028-3033.