

Credit Risk Assessment Based on Rough Set Theory and Fuzzy Support Vector Machine

Jianguo Zhou Jiming Tian

School of Business Administration, North China Electric Power University, Baoding 071003, P. R. China

Abstract

In this paper, a hybrid intelligent system, combining rough set approach and fuzzy support vector machine (FSVM), is applied to the study of credit risk assessment in commercial banks. We can get reduced information table, which implies that the number of evaluation criteria such as financial ratios and qualitative variables is reduced with no information loss through rough set approach. And then, this reduced information table is used to develop classification rules and train FSVM. The rationale of our hybrid system is using rules developed by rough sets for an object that matches any of the rules and FSVM for an object that does not match any of them.

Keywords: Rough sets, FSVM, Credit risk assessment

1. Introduction

Evaluation of the credit risk in the commercial banks has been for a long time, a major preoccupation of researchers and practitioners. Credit risk is a general term and, according to a wide spread definition, is the risk that the commercial bank can not withdraw the long-term credit loan completely or in time. All these situations may result in a discontinuing of the bank's operations and lead to the bankruptcy of the bank. The number of failing commercial banks is an important indicator for the health of the economy and it can be considered as an index of the development and robustness of the economy.

The development and use of models, able to predict loss in advance, can be very important for the commercial banks in two different ways. First, as "early warning systems", such models can be very useful for those (i.e. managers, authorities, etc) who have to prevent failure. Second, such models can be useful in aiding decision-makers of the bank in charge of evaluation and selection of customers. At the beginning, statistical methods such as univariate statistical methods, discriminant analysis, linear probability models and logit and probit analysis have been mainly used for credit risk assessment problem.

Later, the development and application of artificial intelligence led some researchers to employ inductive learning and neural networks in credit risk domain [1], [2].

In this paper, we proposed a hybrid model which is composed of rough set component and FSVM component. By rough set, some rules are extracted from the information system. Using rough set tool, we can discover knowledge in two kinds of rules: deterministic and non-deterministic. Sometimes, the rules generated by rough sets fail to predict newly entered object because of non-deterministic rules. To handle this situation, some researchers reported that reduced data is fed into BP neural network for complementing the limitation of rough sets, which finally produces full prediction of new case data. However, BP neural network has some limitations in that it is an art to find an appropriate model structure and optimal solution and it cannot acquire the ideal result with the small training data which is one of the commercial bank data's characters in China. On the other hand, SVM can captures the geometric characteristics of feature space without deriving weights of networks from the training data and it is capable of extracting the optimal solution with small training data [3]-[5].

2. Rough sets and SVM

2.1. Basic concepts of rough sets

An information system is a 4-tuple $S = (U, A, V, f)$, where U is a finite set of objects, called the universe, A is a finite set of attributes, $V = \bigcup_{a \in A} V_a$ is a domain of attribute a , and $f: U \times A \rightarrow V$ is called an information function such that $f(x, a) \in V_a$, for $\forall a \in A, \forall x \in U$. In the classification problems, an information system is also seen as a decision table assuming that $A = C \cup D$ and $C \cap D = \emptyset$, where C a set of condition is attributes and D is a set of decision attributes.

Let $S = (U, A, V, f)$ be an information system, every $P \subseteq A$ generates an indiscernibility relation $IND(P)$ on U , which is defined as follows:

$$IND(P) = \{(x, y) \in U \times U : f(y, a) = f(x, a), \forall a \in P\}. \quad (1)$$

$U / IND(P) = \{c_1, c_2, \dots, c_k\}$ is a partition of U by P , every C_i is an equivalence class. For $\forall x \in U$ the equivalence class of x in relation $U / IND(P)$ is defined as follows:

$$[x]_{U/IND(P)} = \{y \in U : f(y, a) = f(x, a), \forall a \in P\}. \quad (2)$$

Let $P \subseteq A$, $x \subseteq U$. The P-lower approximation of x (denoted by $P_*(x)$) and the P-upper approximation of x (denoted by $P^*(x)$) are defined as follows:

$$\begin{aligned} P_*(x) &= \{y \in U : [y]_{U/IND(P)} \subseteq x\}, \\ P^*(x) &= \{y \in U : [y]_{U/IND(P)} \cap x \neq \emptyset\}. \end{aligned} \quad (3)$$

Where $P^*(x)$ is the set of all objects from U which can be certainly classified as elements of x employing the set of attributes P . $P^*(x)$ is the set of objects of U which can be classified as elements of x using the set of attributes P . Let $P, Q \subseteq A$, the positive region of classification $U / IND(Q)$ with respect to the set of attributes P , or in short, P -positive region of Q , is defined as $POS_P(Q) = \bigcup_{x \in U / IND(Q)} P(x)$.

$POS_P(Q)$ contains objects in U that can be classified to one class of the classification $U / IND(Q)$ by attributes P . The dependency of Q on P is defined as:

$$\gamma_P(Q) = \text{card}(POS_P(Q)) / \text{card}(U). \quad (4)$$

An attribute a is said to be dispensable in P with respect to Q , if $\gamma_P(Q) = \gamma_{P-\{a\}}(Q)$; otherwise a is an indispensable attribute in P with respect to Q .

Let $S = (U, A, V, f)$, be a decision table, the set of attributes P ($P \subseteq C$) is a reduce of attributes C , which satisfies the following conditions:

$$\gamma_P(D) = \gamma_C(D), \gamma_P(D) \neq \gamma_{P'}(D), \forall P' \subset P. \quad (5)$$

A reduce of condition attributes C is a subset that can discern decision classes with the same accuracy as C , and none of the attributes in the reduced can be eliminated without decreasing its distrainable capability [14].

2.2. SVM and FSVM

SVM is the theory based on statistical learning theory. It realizes the theory of VC dimension and principle of structural risk minimum. The whole theory can be simply described as follows: searching an optimal

hyper plane satisfies the request of classification, then using a certain algorithm to make the margin of the separation beside the optimal hyper plane maximum while ensuring the accuracy of correct classification. According to the theory, we can classify the separable data into classes effectively. The following is the brief introduction of SVM in cases.

Suppose we are given a set of training data $x_i \in R^n$ ($i=1, 2, \dots, n$) with the desired output $y_i \in \{+1, -1\}$ corresponding to the two classes. And suppose there exists a separating hyper plane with the target functions $w \cdot x_i + b = 0$ (w represents the weight vector and b the bias). To ensure all training data can be classified, we must make the margin of separation ($2 / \|w\|$) maximum. Then, in the case of linear separation, the linear SVM for optimal separating hyper plane has the following optimization problem,

$$\text{Minimize} \quad \phi(w) = \frac{1}{2} w^T w \quad (6)$$

$$\text{Subject to} \quad y_i(x_i \cdot w + b) \geq 1, i = 1, 2, \dots, n \quad (7)$$

The solution to above optimization problem can be converted into its dual problem. We can search the nonnegative Lagrange multipliers by solving the following optimization problem,

$$\text{Maximize} \quad Q(\alpha) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j x_i^T x_j \quad (8)$$

$$\text{Subject to} \quad \sum_{i=1}^n a_i y_i = 0 \quad \alpha_i \geq 0 \quad i = 1, 2, \dots, n \quad (9)$$

The corresponding training data are the support vectors. Suppose α_i^* are the optimal Lagrange multipliers, the optimal weight vectors are

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad (10)$$

The optimal biases are

$$b^* = y_j - \sum_{i=1}^n y_j \alpha_i^* x_i^T x_j \quad (11)$$

Then, the optimal equation for classification is

$$f(x) = \text{sgn} \left\{ (w^* \cdot x) + b^* \right\} \quad (12)$$

The above discussion is restricted to the case that the training data is separable. To generalize the problem to the non-separable case, slack variable $\varepsilon_i \geq 0, i = 1, 2, \dots, n$ is introduced under the constraints of (2). The objective equation is

$$\text{Minimize:} \quad \phi(w, \varepsilon) = \frac{1}{2} w^T w + C \sum_{i=1}^n \varepsilon_i \quad (13)$$

$$\text{Subject to} \quad y_i(w^T x_i + b) \geq 1 - \varepsilon_i \quad \varepsilon_i \geq 0, i = 1, 2, \dots, n \quad (14)$$

C is the nonnegative parameter chosen by users. Solving the problem is similar to the problem of the case of linear separation. But the constraints are changed to be

$$\sum_{i=1}^n a_i y_i = 0 \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n \quad (15)$$

Due to over fitting, in SVM and least-squares SVM, the training process is very sensitive to those outliers in the training dataset which are far away from their own class. In order to decrease the effect of these outliers of noises, we assign each data point in the training dataset with a membership and sum the deviations weighted by their memberships. If one data point is detected as an outlier, it is assigned with a low membership, so its contribution to total error term decreases. Unlike the equal treatment in standard SVM or least-square SVM, this kind of SVM fuzzifies the penalty term in order to reduce the sensitivity of less important data points. The classification problem is modeled by the following programming:

$$\text{Minimize: } \phi(w, \varepsilon) = \frac{1}{2} w^T w + C \sum_{i=1}^n \mu_k \varepsilon_i \quad (16)$$

Subject to

$$y_i (w^T x_i + b) \geq 1 - \varepsilon_i \quad \varepsilon_i \geq 0, \quad i = 1, 2, \dots, n \quad (17)$$

where m_k is the membership generalized by some outlier-detecting methods.

An important step in implementing fuzzy SVM is the determination of memberships. In this study, the relationship of the sample with other samples is reflected by affinity. In the determination of the fuzzy membership by membership function based on affinity, both the distance of the sample to the centre of its class and the relationship with other samples are taken into consideration [15].

3. Experiments

3.1. Research data

The research data we employ is provided by a subsidiary of the Construction bank in China, and consists of 330 debtors from 2002 to 2004. The sample data sets consist of the equal number of every kind debtor: 110 of normality debtors, 110 of doubt debtors, 110 of loss debtors. The data set is arbitrarily split into two subsets: about 80% of the data is used for a training set and 20% for a checking set.

3.2. Index system

The index system of risk assessment is established with 8 financial ratios and 4 qualitative indexes. The

selection of the financial ratios is based upon two main characteristics: their usefulness in previous studies and the experiences from past decisions, the knowledge and the preferences of financial experts. Financial ratios we used were gathered one year before the debtor was evaluated. In other words, when a debtor received credit assessment in a certain year, this debtor is classified as a certain grade one year before is used. The data of financial indexes can be got from the financial reporting and the data of non-financial indexes can be evaluated by experts in the bank [10], [11]. As rough sets approach is concerned with discrete values, we have to transform quantitative attributes into qualitative terms according to some norms.

Number	Index	The number of category
x_1	(Current assets-current liability) / total assets	4
x_2	Undistributed profit/ total assets	3
x_3	(Total profit+ interest expense) / total assets	4
x_4	Equity capital/total liability	5
x_5	Sales income/ total assets	4
x_6	Net loan dependence rate	3
x_7	Net income to sales	4
x_8	Cash flow to total liability	3
x_9	Trade risk	4
x_{10}	Credit record	4
x_{11}	Development prospect	4
x_{12}	Level of management	4

Table 1: Index System of Credit Risk Assessment

And the norms mainly followed from the financial manager's experience and some standards of the corporate financial analysis. The index system and the number of category for every index are shown in Table 1.

3.3. Membership determination of outliers

In the determination of the fuzzy membership by membership function based on affinity, both the distance of the sample to the centre of its class and the relationship with other samples are taken in consideration. In this study, the relationship of the sample with other samples is reflected by affinity.

The membership function based on affinity is formulated by two parts

$$\mu_i = f(\mu_d(x_i), \mu_k(x_i, \bar{x})) \quad (18)$$

where, μ_i denotes the membership of x_i to its class, $\mu_d(x_i)$ denotes the distance relationship of the x_i with its class centre. $\mu_d(x_i)$ is determined by the following formulations.

$$\mu_d(x_i) = \begin{cases} 1, & d_i \leq a \\ 1 - 2[(d_i - a)/(c - a)]^2 & a \leq d_i \leq b \\ 2[(d_i - c)/(c - a)]^2 & b \leq d_i \leq c \\ 0 & d_i > c \end{cases} \quad (19)$$

Where $d_i = \|x_i - \bar{x}\|$, a, b, c are predefined parameters. where, $\mu_k(x_i, \bar{x})$ denotes the fuzzy connectedness of x_i with the centre of its class, and the affinity relationship of x_i with other samples in the same class. $\mu_k(x_i, \bar{x})$ is determined by the following formulation.

$$\mu_k(x_i, \bar{x}) = \max[\min(\mu_k(c_1, c_2), \mu_k(c_2, c_3), \dots, \mu_k(c_{m-1}, c_m))] \quad (20)$$

Where, $\rho(x_i, \bar{x})$ denotes the a path from x_i to the centre of its class \bar{x} . $c_1, c_2, c_3, \dots, c_m$ denote the points on the path, and $m \geq 2$, $c_1 = x_i, c_m = \bar{x}$.

$f(\cdot, \cdot)$ denotes the a certain function relationship. In this study, we take product relation. Thus the formulation can be described as follow

$$\mu(x_i) = \mu_d(x_i) \times \mu_k(x_i, \bar{x}) \quad (21)$$

3.4. The assessment models configuration

In experiment, the polynomial kernel and the Gaussian radial basis function are used as the kernel function of SVM and FSVM. Since SVMs do not have a general guidance for determining the upper bound C and the kernel parameter σ^2 , this study varies the parameters to select optimal values for the best prediction performance. At last, we take the choice of $\sigma^2 = 2.4$, $C = 57.7$ and $\varepsilon = 0.01$, which can produced the best possible results according to the validation set [12]. The MATLAB SVM toolbox executed these processes. For verifying the applicability of SVMs, we also design discriminant analysis (DA), BP neural network

(BPN) and standard SVM as the benchmark. The network structure was 5-9-1 for input layer, hidden layer and output layer respectively. We used sigmoid function for activation and Levenberg-Marquardt algorithm for learning. The statistical analysis and BPN were executed by Sas V9.0 and MATLAB NN toolbox. We constituted Hybrid Model I with rough sets and BP network, Hybrid Model II with rough sets and SVM, and Hybrid Model III with rough sets and FSVM [13], [14].

3.5. Experiment and results

Table 2 shows the results after rough set analysis was performed. As we can see, in 12 experiments we obtained one minimal reduct $\{x_3, x_5, x_8, x_{10}, x_{12}\}$. This minimal reduct is the result of horizontal reduction. The column named Group I shows ratio of cases that have a matching rule and Group II shows the cases which had no matching rule. Rough set analysis part of the experiment was performed by Rosetta developed by Norwegian Scientific and Technological University and Portland Warsaw University.

After rough set analysis was finished and holdout sample was separated into two groups, we tested the performance of each methodology. First, five methods which consisted of rule method, DA model, BPN, SVM and FSVM were tested with Group I holdout sample.

Next, four methods, except rule method, were tested with Group II sample. Because cases in Group II have no matching rule, rules cannot classify them. Table III shows Group I hit ratios of each method and Table IV shows Group II hit ratios.

Table 2 and 3 summary the results of comparison. We can see that rule method performs best in classifying Group I samples. And we also can see that FSVM with the membership based on affinity outperforms DA, BPN and SVM in Group II test.

Experiment	Minimal reduct	Quality of sorting	Group I			Group II		
			N ^a	D ^b	L ^c	N	D	L
1	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.89	0.85	0.80	0.87	0.15	0.20	0.13
2	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.85	0.9	0.82	0.92	0.1	0.18	0.08
3	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.91	0.85	0.71	0.84	0.17	0.29	0.16
4	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.9	0.87	0.73	0.86	0.13	0.27	0.14
5	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.88	0.96	0.9	0.9	0.04	0.1	0.1
6	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.89	0.94	0.85	0.88	0.06	0.15	0.12
7	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.9	0.88	0.79	0.79	0.14	0.21	0.21
8	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.88	0.88	0.73	0.86	0.12	0.27	0.14
9	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.85	0.93	0.82	0.79	0.07	0.18	0.21
10	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.88	0.92	0.86	0.95	0.08	0.14	0.05
11	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.87	0.91	0.82	0.83	0.09	0.18	0.17
12	$\{x_3, x_5, x_8, x_{10}, x_{12}\}$	0.897	0.89	0.8	0.85	0.11	0.2	0.15

^aNormality, ^bDoubt, ^cLoss

Table 2: Results After Rough Set Data Analysis.

Experiment	Rlue			DA			BPN			SVM			FSVM		
	N	D	L	N	D	L	N	D	L	N	D	L	N	D	L
1	95.3	96.5	93.5	68.4	88.6	69.2	83.5	86.4	80.2	88.8	92.6	83.2	90.1	92.9	85.3
2	95.6	97.6	92.4	73.3	89	71.3	83.7	87.8	81.6	86.7	95.1	84.6	89.5	96.8	87.6
3	95.3	95.8	95.4	68.2	85.9	72.6	84.7	83.1	84.2	87.1	95.8	85.7	88.2	95.9	87.8
4	95.6	95.4	94.2	69	87.7	81.3	81.6	87.7	85.6	88.4	94.5	90.2	92.1	95.5	93.2
5	93.5	96.6	93.1	69.6	91	69.3	85.9	85.4	84.7	84.8	96.6	89.3	88.9	96.7	91.4
6	95.4	96.5	96.1	75.5	89.7	68.2	85.1	87.3	80.2	87.2	91.8	92.1	92.1	92.8	93.1
7	96.8	96.1	94.6	70.8	85.5	72.1	83.7	86.8	86.3	81.4	95.1	93.5	87.1	95.2	93.7
8	95.6	98.4	93.6	70	88.3	74.6	82.2	88.5	84.5	86.4	95.8	84.6	89.5	96.8	89.6
9	95.3	96.8	95.3	73.8	89	80.4	85.9	82.9	81.7	85.7	92.7	89.6	87.5	92.8	89.8
10	94.9	96.4	94.2	71.4	90.5	74.7	81.3	85.7	83.2	88.5	89.3	88.3	91.2	90.3	88.9
11	95.7	96.4	97.6	69.8	88.1	72.3	83.9	84.5	85.4	94.3	95.2	87.6	94.1	95.3	91.6
12	94.4	97.5	95.2	66.3	86.3	68.3	80.9	87.5	81.6	85.8	92.5	93.2	85.9	93.5	92.2
Average	95.51			77.1			84.25			89.77			91.53		

Table 3: Hit Ratios of Methods in Group I Holdout Subset.

E ^a	DA			BPN			SVM			FSVM		
	N	D	L	N	D	L	N	D	L	N	D	L
1	72.7	84.2	72.3	81.8	84.2	81.9	81.8	94.7	85.6	84.9	93.7	88.7
2	66.7	83.3	71.6	80	88.9	83.4	80	94.4	86.3	86.1	91.4	86.8
3	69.2	89.7	82.1	80	86.2	86.3	86.7	93.1	90.1	83.8	94.1	90.6
4	62.5	88.9	74.6	76.9	92.6	84.4	84.6	92.6	87.4	89.6	93.6	89.4
5	71.4	90.9	75.1	75	81.8	84.6	87.5	81.8	86.2	87.4	81.7	85.2
6	64.3	85.7	79.3	85.7	85.7	80.6	85.7	92.9	88.6	85.7	90.9	88.4
7	72.7	83.3	78.1	78.6	83.3	82.2	78.6	91.7	91.3	81.6	91.9	93.3
8	60	86.4	80.2	81.1	86.4	82.4	90.9	95.5	86.7	93.9	95.3	88.7
9	70	81.3	70.2	80	87.5	81.5	80	93.8	89.5	88	98.8	89.8
10	75	86.7	74.5	79	80	83.6	80	86.7	88.6	84.2	89.7	91.6
11	63.6	73.3	72.5	87.5	86.7	84.6	87.5	93.3	85.6	89.5	94.3	88.6
12	67.2	80	71.6	81.8	85	82.1	81.8	90	86.6	84.9	93.5	89.6
A ^b	75.86			81.14			87.72			89.42		

^aExperiment, ^bAverage

Table 4: Hit Ratios of Methods in Group II Holdout Subset.

From the general view, the hybridized model of rough sets and FSVM dominates the other four classifiers, revealing the hybridized model is an effective tool for credit risk evaluation.

4. Conclusions

The experiment results show the effectiveness of rough set approach as a rule generator in data digging. And the hybridized model of rough sets and FSVM approach with membership based on affinity outperforms DA, BPN and standard SVM to the problem of systematic risk reorganization.

Our study has following limitations that need further research. We neglected the difference among industries and difference in sizes of debtors. Further, in generating rules, we strictly applied rough set theory, so there are a possibility that rule could not be generated because of just a few extraordinary objects. With more work in dealing with doubtful region we may discover more valuable knowledge.

References

- [1] J.N. Xu, B. Xi, AHP-ANN Based Credit Risk Assessment for Commercial Banks. *Journal harbin univ.sci. & tech*, 6: 94-98, 2002.
- [2] G.A. Yu, H.B. Xu, Design and Implementation of an Expert System of Loan Risk Evaluation. *Computer Engineering & Science*. 10: 104-106, 2004.
- [3] Z. Pawlak, Rough sets. *International Journal of Computer and International Sciences*, 11(3): 341-356, 1982.
- [4] B. Ahn, C. Cho, C. Kim, The integrated methodology rough set theory and artificial neural network for business failure prediction. *Expert Systems with Application*, 18(2): 65-74, 2000.
- [5] K.S. Shin, T.S. Lee, Hyun-jung Kim, An Application of Support Vector Machines in Bankruptcy Prediction Model. *Expert System with Application*, 28: 127-135, 2005.
- [6] P. Lak., Rough Sets and Intelligent Data Analysis. *Information Science*, 147(11): 1-12, 2002.

- [7] Y.S. Zhu, Y.Y. Zhang, The Study on Some Problems of Support Vector Classifier. *Computer Engineering and Applications*, 13: 66-38, 2003.
- [8] D.Y.M. Sinalingam, Narenkumar Pandia, Minimal Classification Method with Error Correlation Codes for Multiclass Recognition. *International Journal of Pattern Recognition and artificial intelligence*, 5: 663-680, 2005.
- [9] C. Gold, P. Sollish, Model Selection for Support Vector Machine Classification. *Neurcomputing*, 55: 221-249, 2005.
- [10] T. Salmi, I. Virtanen, The Generalized Association between Financial Statements and Security Characteristics. *Scand. J.Mgmt*, 13(2): 121-136, 1997.
- [11] P. Jackson, W. Perraudin, Regulatory implications of credit risk modeling. *Journal of Banking and Finance*, 24: 1-14, 2000.
- [12] N. Y. Deng, Y.J. Tian, A New Method in Data Mining: Support Vector Machines. *Science press*, Beijing, 2004.
- [13] Y.W. Wu, C.N. Zhang, A Rough Neural Network for Material Proportioning System. *Proceeding of IEEE 2002 International Conference*, pp. 1189-1193, 2002.
- [14] R. Li, Z.O. Wang, Mining classification rules using rough sets and neural networks. *European Journal of Operational Research*, 157(7): 439-448.
- [15] L.H. Chiang, M.E. Kotanchek, A.K. Kordon, Fault diagnosis based on fisher discriminant analysis and support vector machine. *Computers and Chemical Engineering*, 28: 1389-1074, 2004.
- [16] L.A. Zadeh, Soft computing and fuzzy logic. *IEEE Software*, 11(6): 48-56, 1994.