The Integrated Methodology of Rough Set Theory and Fuzzy SVM for Customer Classification

Jianguo Zhou, Tao Bai, Aiguang Zhang and Jiming Tian
School of Business Administration
North China Electric Power University
Baoding, China
dldxzjg@126.com, ncepubaitao@126.com

Abstract—**In this paper, an intelligent system that hybridized rough set approach (RS) and fuzzy support vector machine (FSVM) is applied to the study of customer classification 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 one that dose not match any of them. By applying the proposed approach to customer classification of China Construction Bank, RS-FSVM not only provides satisfactory approximation and generalization property, but also achieves superior performance to traditional discriminant analysis model (DA), BP neural networks (BPN) and standard SVM.**

Keywords—**Customer Classification, Rough Set Theory, Fuzzy Support Vector Machine**

I. Introduction

Recently, China's finance industry especially bank credit has been developed rapidly, and has made great effect on the investment, consumption and social economy. Along with the improving information degree of bank industry, the commercial banks have attained a plenty of data resource about customer's information. For this service-oriented industry, discriminating faithful customers from bad ones accurately plays an important role on bank's success. For example, commercial banks can optimize the deposit and loan strategy according to their client classification activities, and then the business risk is decreased, simultaneity the operating performance can be improved. Thus, accurate and efficient classifiers should be found to the customer classification of commercial banks [1].

The methods of customer classification can be classified into two types: statistical methods and machine learning techniques. The statistical methods include univariate approaches, linear multiple discriminant analysis (MDA) [2] [3], multiple regression [4], and logistic regression [5]. To develop a more accurate and general applicable prediction approach, data mining and machine learning techniques are employed, e.g. neural network models (NN) [6] [7], rough set theory [8], Bayesian network (BN) models, genetic programming [9][10]. Recently, new algorithms in machine learning, support vector machines (SVMs) [11], developed by Boster and Vapnik (1992) provide better solutions to design boundary than that of neural network. Since the new model was proposed, SVM has been successfully applied to numerous applications, including the bankruptcy prediction, handwriting recognition, particle identification, digital image identification and customer classification.

In this paper, we proposed a hybrid model which is composed of rough set component and fuzzy SVM [12] 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.

II. Basic Concept of Models

A. Rough Set Theory

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, $V=U_{a\in A}V_a$ 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 $C\cap D=\phi$, 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:

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

$U/\text{IND}(P)=\{c_1,c_2,\ldots,c_k\}$ 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 $U/\text{IND}(P)$ is defined as follows:

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

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:

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

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 $U/\text{IND}(Q)$ with respect to the set of attributes P , or in short, P -positive region of Q , is defined as $POS(Q)=\bigcup_{X\in U/IND(Q)}P(x)$.

$\text{POS}_P(Q)$ Contains objects in U that can be classified to one class of the classification $U/\text{IND}(Q)$ by attributes P . The dependency of Q on P is defined as:

$$\gamma_P(Q)=\text{card}(\text{POS}_P(Q))/\text{card}(U)\,. \tag{4}$$

An attribute a is said to be dispensable in P with respect to Q , if $\gamma_P(Q)=\gamma_{P\text{-}\{a\}}(Q)$; 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 reduction of attributes C , which satisfies the following conditions:

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

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.

B. 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 $x_i\in R^n$ $(i=1,2,\cdots n)$ 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,

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

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

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,

$$\text{Maximize}\ Q(\alpha)=\sum_{i=1}^n a_i-\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n a_ia_jy_iy_jx_i^Tx_j \tag{8}$$

$$\text{Subject to}\ \sum_{i=1}^n a_iy_i=0\,,\ \alpha_i\geq 0\,,\ i=1,2,\cdots n \tag{9}$$

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

$$w^*=\sum_{i=1}^n\alpha_i^*y_ix_i \tag{10}$$

The optimal biases are

$$b^*=y_i-\sum_{i=1}^n y_j\alpha_i^*x_i^Tx_j \tag{11}$$

Then, the optimal equation for classification is

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

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

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

$$\text{Subject to}\ y_i(w^Tx_i+b)\geq 1-\varepsilon_i\,,\ \ \varepsilon_i\geq 0\,,\ i=1,2,\cdots,n \tag{14}$$

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

$$\sum_{i=1}^n a_iy_i=0\quad 0\leq\alpha_i\leq C\,,\ i=1,2,\cdots,n \tag{15}$$

Due to over fitting, in SVM, the training process is very sensitive to those outliers in the training dataset. 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, this kind of SVM fuzzifier the penalty term in order to reduce the sensitivity of less important data points. The classification problem is modeled by the following programming:

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

Subject to: $$y_i\left(w^T x_i+b\right)\geq 1-\varepsilon_i,\ \varepsilon_i\geq 0,\ i=1,2,\cdots,n \quad (17)$$

Where μ_k is the membership generalized by some outlier-detecting methods.

III. EXPERIMENT

A. 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.

B. 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. As rough sets approach is concerned with discrete values, we have to transform quantitative attributes into qualitative terms according to some norms.

TABLE I. INDEX SYSTEM OF CREDIT RISK ASSESSMENT

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

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.

C. Membership Determining of Outliers

An important step in implementing fuzzy SVM is the determination of memberships. In the fuzzy membership determination 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.

The membership function based on affinity is formulated by two parts:

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

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

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

Where $d_i=\left\|x_i-\bar{x}\right\|$, a,b,c are predefined parameters. Where, $\mu_k\left(x_i,\bar{x}\right)$ 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. $\mu_k\left(x_i,\bar{x}\right)$ is determined by the following formulation.

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

Where, $\mu_k\left(x_i,\bar{x}\right)$ denotes a path from x_i to the centre of its class $\bar{x}$. $c_1,c_2,\cdots,c_m$ denotes the points on the path, and $m>2$, $c_1=x_i, c_m=\bar{x}$.

$f\left(\mu_d\left(x_i\right),\mu_k\left(x_i,\bar{x}\right)\right)$ denotes the certain function relationship. In this study, we take product relation. Thus the formulation can be described as follow:

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

D. The assessment Models Configuration

In experiment, 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 fuzzy SVM.

The feed-forward back-propagation neural network (BPN) applied to the experimental sample includes 5 input neurons in the input layer, 9 neurons in the hidden layer, and 1 neuron in the output layer. This study constructed a three-layer network and employed the "TRAINLM algorithm", "LEARNGDM", and "MSEREG" as the training function, the adaptive learning function, and the performance function, respectively. The transfer function was set to the "TANSIG function" and the "PURELIN function" for hidden layer and out put layer. The number of epochs was set to 300 and the learning rate was set to 0.05 in each epoch.

In machine learning theories, popular kernel functions, such as the Gaussian kernel function, have been found to provide good generalization capabilities. Accordingly, the Gaussian kernel function is used as the kernel function of SVM and FSVM. The Gaussian kernel function is given in Section Ⅱ. Since SVMs don't 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.

E. Experiment and Results

Table Ⅱ shows the results after rough set analysis was performed. As we can see, in 12 experiments we obtained one minimal reduction $\{x_3,x_5,x_8,x_{10},x_{12}\}$. This minimal reduction 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.

TABLE II. RESULTS AFTER ROUGH SET DATA ANALYSIS

Experiment	Minimal reduction	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

[a]Normality, [b]Doubt, [c]Loss

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, standard SVM and FSVM were tested with Group Ⅰ holdout sample. The results were shown in Table III.

TABLE III. HIT RATIOS OF METHODS IN GROUP Ⅰ HOLDOUT SUBSET

experiment	Rule			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	90.2	93.8	85.1
2	95.6	97.6	92.4	73.3	89	71.3	83.7	87.8	81.6	86.7	95.1	84.6	89.1	95.5	86.4
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	89.4	96.6	87.6
4	95.6	95.4	94.2	69	87.7	81.3	81.6	87.7	85.6	88.4	94.5	90.2	90.3	95.4	91.5
5	93.5	96.6	93.1	69.6	91	69.3	85.9	85.4	84.7	84.8	96.6	89.3	87.6	96.3	90.8
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	90.1	93.2	92.9
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	85.3	96	96.7
8	95.6	98.4	93.6	70	88.3	74.6	82.2	88.5	84.5	86.4	95.8	84.6	89.8	95.9	86.5
9	95.3	96.8	95.3	73.8	89	80.4	85.9	82.9	81.7	85.7	92.7	89.6	87.9	94.1	91.1
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	90.6	91.2	91.8
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	95.2	96.4	89.6
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	88.5	94.7	95.4
Average	95.51			77.1			84.25			89.77			91.63		

Next, four methods, except rule method, were tested with Group Ⅱ sample. Because cases in Group Ⅱ have no matching rules, rules cannot classify them. Table III shows Group Ⅰ hit ratios of each method and Table V shows hit ratios of Group Ⅱ. After testing each methodology, we calculated performance of evaluated models as shown in Table IV.

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

From the general view, as it is shown in Table Ⅴ, the hybridized model of rough sets and FSVM dominates the others, revealing the hybridized model Ⅲ is an effective tool for credit risk evaluation.

TABLE IV. HIT RATIOS OF MODELS IN GROUP II 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	86.4	95.2	88.6
2	66.7	83.3	71.6	80	88.9	83.4	80	94.4	86.3	84.6	96.1	90.2
3	69.2	89.7	82.1	80	86.2	86.3	86.7	93.1	90.1	88.2	92.6	91.4
4	62.5	88.9	74.6	76.9	92.6	84.4	84.6	92.6	87.4	87.9	92.2	89.3
5	71.4	90.9	75.1	75	81.8	84.6	87.5	81.8	86.2	90.2	89.2	87.2
6	64.3	85.7	79.3	85.7	85.7	80.6	85.7	92.9	88.6	88.1	94.5	89.5
7	72.7	83.3	78.1	78.6	83.3	82.2	78.6	91.7	91.3	82	93.4	90.2
8	60	86.4	80.2	81.1	86.4	82.4	90.9	95.5	86.7	90.6	96.3	87.6
9	70	81.3	70.2	80	87.5	81.5	80	93.8	89.5	85.2	94.2	91.4
10	75	86.7	74.5	79	80	83.6	80	86.7	88.6	84.8	90.1	91.1
11	63.6	73.3	72.5	87.5	86.7	84.6	87.5	93.3	85.6	90.5	95.7	87.8
12	67.2	80	71.6	81.8	85	82.1	81.8	90	86.6	84.7	92.9	88.2
A[b]	75.86			81.14			87.72			89.95		

TABLE V. HIT RATIOS OF MODELS

Experiment	DA	BPN	SVM	Hybrid I	Hybrid II	HybridIII
1	75.9	83	87.7	88.8	91.2	92.3
2	75.8	84.2	87.8	89.6	91	92.6
3	77.9	84	89.7	89.8	92.1	92.8
4	77.3	84.8	89.6	89.8	91.6	92.5
5	77.8	82.9	87.7	87.4	89.7	91.7
6	77.1	84.1	89.7	90	92.5	92.6
7	77	83.4	88.6	88.6	91.5	92.2
8	76.5	84.1	89.9	89.5	93.4	93.5
9	77.4	83.2	88.5	89.4	91.7	92.4
10	78.8	82.1	86.9	88	90.1	91.3
11	73.2	85.4	90.5	91.4	92.6	93
12	73.2	83.1	88.3	89.3	90.9	92.1
Average	76.5	83.7	88.7	89.3	91.6	92.4

IV. CONCLUSION

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] W. E. Henley and D. J. Hand, "A k-NN classifier for assessing consumer credit risk," *Statistician*, vol. 45, pp. 77–95, Jan. 1996.

[2] F. Glover, "Improved linear programming models for discriminant analysis," *Decision Sci.*, vol. 21, pp. 771–785, 1990.

[3] T. S. Lee, C. C. Chiu, C. J. Lu, and I. F. Chen, "Credit scoring using the hybrid neural discriminant technique," *Expert Syst. Appl.*, vol. 23, pp. 245–254, Oct. 2002.

[4] S. Canbas, A. Cabuk, S.B. Kilic, "Prediction of commercial bank failure via multivariate statistical analysis of financial structure: The Turkish case," *European Journal of Operational Research,* vol. 16, no.6, pp. 528–546, 2005.

[5] S. Jones, D.A. Hensher, "Predicting firm financial distress: A mixed logit model," *Accounting Review*, vol. 79, no. 4, pp. 1011–1038, 2004.

[6] R. Malhotra and D. K. Malhotra, "Evaluating consumer loans using neural networks," *Omega*, vol. 31, pp. 83–96, Apr. 2003.

[7] R. Li, Z.-o. Wang, "Mining classification rules using rough sets and neural networks", *European Journal of Operational Research*, Vol.157, NO. 7, pp439–448.

[8] Ahn. B, Cho. C, Kim. C, "The integrated methodology rough set theory and artificial neural network for business failure prediction", *Expert Systems with Application,* vol.18, no. 2, pp65-74, 2000.

[9] K-S. Shin, Y-J Lee, "A genetic algorithm application in bankruptcy prediction modeling," *Expert Systems with Applications,* vol.23, no. 3, pp. 321–328, 2002.

[10] J. Wallrafen, P. Protzel, H. Popp, "Genetically optimized neural network classifiers for bankruptcy prediction," in: Proceedings of the 29 the Annual Hawaii International Conference on System Sciences, 1996.

[11] Carl Gold, Peter Sollish, "Model Selection for Support Vector Machine Classification. Neurcomputing", vol. 55, pp. 221-249, 2005.

[12] H. P. Huang and Y. H. Liu, "Fuzzy support vector machines for pattern recognition and data mining," *Int. J. Fuzzy Syst.*, vol. 4, pp. 826–835, Sep. 2002.

[13] Jianan Xu, Bao Xi, "AHP-ANN Based Credit Risk Assessment for Commercial Banks",*Journal harbin univ*. sci & tech. 6 (2002) 94-98.

[14] Timo Salmi, and Ilkka Virtanen, "The Generalized Association between Financial Statements and Security Characteristics", *Scand. J.Mgmt*, Vol. 13, NO. 2, pp121-136, 1997.

[15] R. Malhotra and D. K. Malhotra, "Evaluating consumer loans using neural networks," *Omega*, vol. 31, pp. 83–96, Apr. 2003.

[16] Jackson P, Perraudin W, "Regulatory implications of credit risk modeling", *Journal of Banking and Finance*, Vol. 24, pp1-14, 2000.