Customer’ Credit Sale Risk Classification Based on Support Vector Machine and Rough Sets

Yuping Wu
School of Economic and Management, Henan Polytechnic University
Jiaozuo, 454000, P. R. China
wyp79055@163.com

Abstract—Aiming at the shortages of the existing data-mining model for classification of customer’s credit sale risk, a new classification model based on rough sets and support vector machine presents is put forward in this paper. First, the theory of rough set is applied to pick up and reduce the index attributes. Then, the training samples are sent to the support vector machine to train and learn. After that, the sorts of the customers’ credit sale risk in test samples are differentiated. The test results indicate that the new classification model based on rough sets and support vector machine shows higher forecast precision than the traditional ones and it is more efficient and practical.

Keywords—credit sale; rough set; SVM; multi-classification; statistical learning theory;

I. INTRODUCTION

Credit sale risk is the uncertainty of the future generated by credit sale. Companies always seek help to their partners in supply chain to solve the problem of funds and resources shortage, so credit sale has become more and more popular [1]. Credit customers generally do not provide substantive guarantees, so credit enterprises face increasing credit sale risk. In recent years, some customers in trouble make other companies in the supply fall into crisis because the payment of goods sold by credit can not withdraw. These examples are proves. Therefore it is very important to explore and set up an effective credit risk assessment method.

Methods for assessing banks’ credit risks are classified into three categories: the quantitative methods based on statistics, the qualitative rating methods based on expert judgment and the rating methods combined quantitative with qualitative analysis. As to the assessment methods, multiple discriminate analysis (MDA) [2], logistic regression analysis [3], neural network analysis [4], support vector machine (SVM) [5], etc, are widely used in commercial bank credit risk. SVM is used to evaluate the credit sale risk in the paper. Reasons are as follows: (1) SVM is a kind of methods with the learning function and it can take full advantage of credit enterprise and credit customers’ related data; (2) it can better solve the problems with small sample, high and non-linear dimension.

So classification model of customers’ credit sale risk based on rough sets and support vector machine is constructed combing rough set theory with support vector machine theory in the paper, which is a new classification method to analyze credit sale risk. Because support vector machine is usually used in two-class classification and it is easy to will fall into over-learning, local minimum points with the model, which is a new classification model based on rough sets and support vector machine is put forward in this paper. First, the theory of rough set is applied to pick up and reduce the index attributes. Then, the training samples are sent to the support vector machine to train and learn. After that, the sorts of the customers’ credit sale risk in test samples are differentiated. The test results indicate that the new classification model based on rough sets and support vector machine shows higher forecast precision than the traditional ones and it is more efficient and practical.

Keywords—credit sale; rough set; SVM; multi-classification; statistical learning theory;

I. INTRODUCTION

Credit sale risk is the uncertainty of the future generated by credit sale. Companies always seek help to their partners in supply chain to solve the problem of funds and resources shortage, so credit sale has become more and more popular [1]. Credit customers generally do not provide substantive guarantees, so credit enterprises face increasing credit sale risk. In recent years, some customers in trouble make other companies in the supply fall into crisis because the payment of goods sold by credit can not withdraw. These examples are proves. Therefore it is very important to explore and set up an effective credit risk assessment method.

Methods for assessing banks’ credit risks are classified into three categories: the quantitative methods based on statistics, the qualitative rating methods based on expert judgment and the rating methods combined quantitative with qualitative analysis. As to the assessment methods, multiple discriminate analysis (MDA) [2], logistic regression analysis [3], neural network analysis [4], support vector machine (SVM) [5], etc, are widely used in commercial bank credit risk. SVM is used to evaluate the credit sale risk in the paper. Reasons are as follows: (1) SVM is a kind of methods with the learning function and it can take full advantage of credit enterprise and credit customers’ related data; (2) it can better solve the problems with small sample, high and non-linear dimension.

So classification model of customers’ credit sale risk based on rough sets and support vector machine is constructed combing rough set theory with support vector machine theory in the paper, which is a new classification method to analyze credit sale risk. Because support vector machine is usually used in two-class classification and it is easy to will fall into over-learning, local minimum points with the increase of redundant data, rough set method is used as the front system of support vector machine in this article. The evaluation index attributes are reduced effectively through the rough set method and the penalty parameter is increased the prediction accuracy of the model in order to improve the prediction accuracy of the model.

II. BASIC MODEL

A. Rough Set Theory

Rough set theory was proposed by Pawlak in 1892 and as new mathematical tool to portray incomplete and uncertain information; it has become the field of artificial intelligence, a new hot spot [6].

1) Information system and decision table

An information system $S$ can be expressed as follows:

$$S = (U, A, V, f)$$

Where: $U = \{x_1, x_2, \cdots, x_n\}$ is a limited set of objects, that is universe of discourse $A = C \cup D = \{a_1, a_2, \cdots, a_n\}$ ($C$ is the set of condition attributes and $D$ is the set of decision attributes) a limited set of the set of condition attributes; $V = \bigcup_{a \in A} V_a$ ($V_a$ is the value set of attribute $a$) is the value set of attributes; $f : \bigcup A \rightarrow V$ is an information systems [7].

The expression form of information system is relationship table and the knowledge expression system with condition attributes and decision attributes is the decision table. In the decision table, column represents attributes, s line represents objects and each line indicates information of the object. Thus, information system is also called information table or decision table.

2) Reduction of decision table

Reduction can be understood as using the most simple condition attribute set to represent the conclusion attributes of the information system without affecting the classification. The process of extracting rules from the information tables by rough set theory is also the process of finding reduction in decision tables. Through reduction, can condition attributes...
and decision-making rules have been simplified and can data
be used more efficient [8]. Reduction steps are as follows:
The first: delete repeated examples in information table;
The second: strike nuclear attribute of condition attribute in
relation to condition attribute;
The third: remove the redundant attributes according to
nuclear attribute and strike the minimum of condition
attribute and remove repeated examples;
The fourth: strike value nuclear of each example’s
attribute values.
The fifth: delete redundant attribute values and strike the
minimum simplification for each instance.
The last: delete the duplicate instance and get the
minimum simplification table.

B. SVM Theory

The idea of SVM classification is to map the samples in
the input space into high dimension feature space with
nonlinear transformation and seek the optimal linear
classification surface in the dimensional feature space [9].
Standard SVM algorithm is to find the optimal classification
surface which not only can separate two different samples
rightly but also have the largest interval [10].

Assumed that the known observation sample set is
\((x_1, y_1), \ldots, (x_n, y_n)\), where: \(x_i \in \mathbb{R}^d\) is
the put vector, \(y_i \in \{+1, -1\}\) is the output vector or
category label. If the training set can be partitioned linearly
by hyper-plane, then the equation of separating hyper-plane
indimensional space is:
\[
g(x) = \langle w, x \rangle + b = \sum_{i=1}^{n} w_i x_i + b = 0 \tag{2}
\]

Where \(w\) is the weight vector. To get the optimal
classification surface is to seek the maximum of
classification interval. The classification interval is \(\frac{1}{2} \|w\|^2\),
that is, the maximum of classification interval is the
minimum of \(\|w\|^2\); Classification line is required to
separate all the samples correctly, that is, it is required to meet:
\[
y_i \left[\langle w, x_i \rangle + b \right] - 1 \geq 0 \tag{3}
\]
The classification surface which meets the above
conditions and makes \(\|w\|^2\) minimize is the optimal
classification surface. The composition of the optimal
classification surface is shown in Fig.1.

The circle and square in Fig.1 denotes two kinds of
training samples. \(H\) is the optimal classification surface,
which separates the two kinds of training samples without
any mistakes and makes the classification gap minimize.
\(H_1\) and \(H_2\) are the hyper surface that crosses the nearest
points to the classification surface and runs parallel with the
optimal classifier. The distance between \(H_1\) and \(H_2\) called
interval classification (margin). The training samples on \(H_1\)
and \(H_2\) make the equation (2) founded and called support
vector machine (SVM). They support the optimal
classification surface and are denoted as circles in Figure 1.

The solution of the optimal surface can be changed to the
optimization function \(\phi(w)\) under the constraints of
equation (4):
\[
\phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} \langle w, w \rangle \tag{4}
\]

According to optimization theory, the optimal
classification function in linear partition condition is as follows:
\[
f(x, \alpha^*, b^*) = \text{sgn} \left\{ \langle w^*, x \rangle + b^* \right\} = \text{sgn} \left\{ \sum_{i=1}^{n} \alpha_i y_i \langle x_i, x \rangle + b^* \right\} \tag{5}
\]

Where: \(\langle x_i, x \rangle\) is the dot product of input vector \(x\) and
each support vector; \(\alpha_i\) is the coefficient corresponding
to \(x_i\); \(\text{sgn}(\cdot)\) is the symbol function, \(b^*\) is the domain of
classification, \(n\) is the number of support vector. The
calculating complexity depends on the support vector's
number because there are only the support vector’s inner
product operation and summation operations in the
discriminate function. Kernel function is used to seek the
optimal classification discriminate function when the data is
linearly inseparable:
\[
f(x, \alpha^*, b^*) = \text{sgn} \left\{ \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b^* \right\} \tag{6}
\]
Where \( K(x, y) \) is kernel function SVM maps the samples in the input space into high dimension feature space with different kernel function. So different kernel functions can be used to construct different support vector machines.

### III. CLASSIFICATION METHOD BASED ON ROUGH SETS AND SVM

#### A. Evaluation Index Attribute

The characteristics of credit sale risk must be assured firstly. The risk factors are from external macro environment, financial situation, credit conditions, the ownership structure, the main decision-making personnel’s quality and customer relationship. 17 indexes are given from these risk factors, which are the local GDP per capita, local residents prices bear ability, lending rate, local finance income, liquidity ratio, quick ratio, profit rate on sales, asset liabilities ratio, contract fulfills ratio, compensate ratio of lend money, ownership structure, personnel economic situation, personnel education level, age, relation frequency, relation level and Intimate degree [11].

#### B. Data Discretization

As the rough set theory can only deal with discrete data, the data must be discreted without losing classification capability of original data.

1) Data discretization

The discretization of continuous attributes is to divide them into several intervals and each interval represented by different codes expresses property values. The dividers method is used to discrete data in the paper.

2) Building decision table

A two-dimensional table is formed according to the discrete condition attributes and decision attribute. Each line of the table describes an object and each column corresponds to an attribute of the object.

#### C. Attributes Reduction

There are redundancy and dependency among the condition attributes of the evaluation indexes such as personal monthly income, the basis of trading volume, cross-purchase volume, repeat purchase rate, etc. So the attribute set which include the smallest attributes must be found to make the decision rule sets derived from the decision table be more closer to man’s natural reasoning. The genetic algorithm is applied to make the condition attribute reduction of the evaluation index in this paper. The fitness \( e \) is defined as follows:

\[
f(B) = (1 - \eta) \times \frac{\cos t(C) - \cos t(B)}{\cos t(C)} + \eta \times \min \left\{ \epsilon, \frac{\left| \sin U / U \cap B \neq \Phi \right|}{|U|} \right\}
\]

Where: \( U \) is the set of training samples; \( \eta \) is a subset of values and weights between Hitting Fraction; \( B \) is a subset gotten from \( C \) through evolutionary search algorithm; the value of property set is defined by the cost function. A feature attribute is subtracted gradually at the beginning of the entire feature attribute to achieve the purpose of attribute reduction in the genetic algorithms.

#### D. Classification by SVM

The standard SVM is two-class and how to make it be multi-class still is the research focus. There are two kinds of multi-class SVM methods: (1) combining two or more SVM; (2) using one two-class classifier to analyze all data sets.

BSVM is one of methods (2). In the classification and regression analysis with certain restrictions, it deals with these restrictions in the way of gradual decomposition. In addition, BSVM uses a simple set to select methods to have a faster convergence speed when classification is difficult [12]. Those are very effective and important for the credit risk analysis, so BSVM is applied in the paper.

The appropriate kernel function \( K(x, y) \) and penalty parameter must be determined in each support vector machine after the training set is given. The RBF function is used as kernel function and selection criteria of parameters is 10-fold cross validation in the paper. So the training process of SVM is to seek the lowest classification error rate point by 10 - fold cross-validation method and determine \( C \) and radial basis function parameter [13], \( x, xK_i \).

### IV. EMPIRICAL ANALYSIS

#### A. Evaluation Set and Samples

1) Evaluation set

Supposeset \( V = \{ V_1 \} = \{ V_1, V_2, V_3, V_4 \} \) is the evaluation set. Where \( V_1, V_2, V_3, V_4 \) is respectively the comment of customer’s credit evaluation. The customers’ credit sale risk is divided into four categories according to the practice in the enterprises: \( V_1 \) represents the customer with minimal risk, \( V_2 \) represents the customer with lower risk, \( V_3 \) represents the customer with higher risk, \( V_4 \) represents the customer with highest risk. The four grades of credit customers’ characteristics are shown in Table I.

2) Samples

426 customers in eastern, central and western regions are selected as samples in an enterprise’s credit sale database. All data is the relevant data of 2007. 290 (approximately 2 / 3) customers in all kinds of risk grades are selected randomly as training samples and the remaining 136 (approximately 1 / 3) are tested samples.

#### B. Index Attribute Abbreviations Reduction

The original samples should be discrete to apply the rough set theory. The 17 evaluation indicators in the paper are both quantitative and qualitative, so their discretization methods are different. As to the quantitative evaluation indicators, the dividers method is used to discrete data in the paper. And the discrete rules for the three qualitative indicators are shown in Table II.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>CREDIT SALE RISK GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_1 )</td>
<td>Customer with minimal risk</td>
</tr>
<tr>
<td>( V_2 )</td>
<td>Customer with lower risk</td>
</tr>
<tr>
<td>( V_3 )</td>
<td>Customer with higher risk</td>
</tr>
<tr>
<td>( V_4 )</td>
<td>Customer with highest risk</td>
</tr>
</tbody>
</table>
After the original samples are pretreated and discrete, Rosetta software is used to reduce the samples. At last 10 reduction indicators are got, which are the local GDP per capita, liquidity ratio, quick ratio, profit rate on sales, asset liabilities ratio, contract fulfills ratio, compensate ratio of lend money, ownership structure, relation frequency and level.

### C. Classification by SVM

Sample set $U(x, y)$ is constructed according to the reduction indicators above, where $y(y=1,2,3,4)$ represents the sample and $x$ is 10 dimension vector, and is the attribute of the sample’s class, $4321,\ldots,321$. SVM is trained and tested by these samples. The curve of classification error rate of different combinations of parameter $(C, \sigma)$ is shown in Fig.1. Compared the classification error rate in all situations, the optimal values of parameters $C$ and $\sigma$ are 1000 and 2 respectively. As shown in Fig.1, classification performance is closely related to parameter and. The bigger $C$ is, the shorter the classification interval is, that is to say, the lower the classification error rate is. The smaller $C$ is, the more samples SVM neglects and the longer the classification interval is, that is to say, the higher the classification error rate is. And the smaller $\sigma$ is, the lower the classification error rate is. On the contrary, the bigger $\sigma$ is, the higher the classification error rate is. So the best parameters can make SVM classifier reach the best performance.

136 samples are used to test classification accuracy rate of SVM with the optimal parameter $C$ and $\sigma$. The overall prediction accuracy is shown in Table III. It can be seen from Table III that the overall correct rate is 89.0%, which shows that the model is effective and feasible.

### References

[10] V. Vapnik and N. Ladimir, The nature of statistical learning theory,
  and medium-sized enterprises’ default risk”, Management World. Vol
  radius margin bound and iterative algorithm”, IEEE Transactions on

**TABLE III. OVERALL PREDICTION ACCURACY**

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Accuracy rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(87.5%)</td>
<td>(87.70%)</td>
<td>(88.2%)</td>
<td>(90.9%)</td>
<td>(89.0%)</td>
</tr>
</tbody>
</table>

Note: The figure outside brackets is the number of samples classified wrongly and figure in brackets is the proportion of rightly classified samples to the respective total samples.