Abstract: The accuracy of credit risk prediction in SC financing is critical for many enterprises, based on machine learning algorithms can be good for SME credit risk assessment research, for this reason, this paper establishes a combinatorial model that can improve credit risk prediction, using support vector machine (SVM) and particle filtering to achieve credit risk classification and prediction, and we introduce information gain (IG) to extract feature variables that contribute significantly to the prediction results and optimize model feature inputs. Compared with the benchmark model, the prediction accuracy of the model in this paper is 97.62%, which is 8.97% higher than that of SVM, and the performance of IG with feature optimization improves the prediction accuracy by another 3%.

Keywords: SC financing; Information gain; Support vector machine; Credit risks; Classification prediction.

I. INTRODUCTION

As a new financing mode, supply chain (SC) finance has developed rapidly in different countries in recent years [1]. As the connection node of capital flow in the SC, commercial banks can provide extended services up or down, so that the capital flow of manufacturers, logistics enterprises, retailers or final consumers can achieve a virtuous circle within the banking system, thus opening up new customer groups and reducing marketing costs while improving marketing efficiency [2]. In addition, the development of SC finance business also helps to improve the profit model of commercial banks and expand the income sources of intermediary business [3]. Therefore, SC finance has become the focus of attention and the focus of profit growth of domestic and foreign commercial banks, financial companies, and even logistics enterprises.

As SMEs themselves generally have weak credit, little collateral and difficulties in capital turnover, coupled with products with characteristics such as high price fluctuations, long production cycles and perishability, it makes it difficult for them to obtain financing from commercial banks [4]. The emergence of SC finance has provided a new way to solve the problem of difficult financing for SMEs. SC finance takes the whole SC as the object of examination, changing the traditional risk management model, shifting the risk management for individual enterprises to risk management for the whole SC [5]. Due to information asymmetry, commercial banks do not have complete information about the operation and profitability of SMEs, and the financing process has greater uncertainty, which easily leads to credit risk [6][7]. Therefore, how to effectively improve the credit risk assessment level of SC finance and reduce the occurrence of loan risks is the key to the healthy development of agricultural SC finance.

In recent years, there has been more global research on credit risk issues in the SC finance model. There are mainly two categories: One is the research on the causes, characteristics and risk prevention measures of credit risk in SC finance. [8] have studied the risks and manifestations faced by banks under the SC finance model, [9] through the study of the risk model of the accounts receivable financing model, pointed out that the risk avoidance mechanism relied on by SC finance still has the possibility of failure, and obtained several key factors affecting the credit risk under the SC finance model. Another category is the research on credit risk assessment issues in SC finance, in which [10] proposed an index system for credit risk assessment in SC finance and used a multi-level grey comprehensive evaluation method to select a single enterprise as the assessment object, but this method relies too much on expert scoring and is too subjective; [11] improved the credit risk evaluation method under the SC finance model, using principal component analysis and logistic analysis. The use of principal component analysis and logistic regression method to establish the credit risk evaluation model, to a certain extent, overcomes the

1 College of Finance and Economics, Zhengzhou University of Science and Technology, Zhengzhou450000, Henan, China;

*Corresponding author: fansongwei37@126.com;

Copyright © JES 2024 on-line: journal.esrgroups.org
shortcomings of too subjective expert evaluation and improves the objectivity of the evaluation, but the logistic regression method requires a large sample size and the prediction accuracy is not high.

Therefore, to address the limitations of the logistic regression method in terms of sample size and prediction accuracy, this paper uses SVM combined with particle filtering approach and then accelerated IG to assess the credit risk of SMEs under the SC finance model and conducts a comparative study of different methods.

II. FEATURE SELECTION AND CLASSIFIER PARAMETER CO-OPTIMIZATION

A. SC finance risk evaluation index system

The SC finance model is a form of logistics finance in the period of logistics evolution to SC, where all parties in the SC maintain a contractual cooperation relationship [12]. Compared with the traditional credit method, its financing model emphasizes mutual benefit, and downplays financial analysis and access control, thus circumventing financing barriers. The bank's assessment of SC members is based on a series of factors such as the macro environment, the SMEs that are the subject of financing, the core SC enterprises that are doing business with, and the condition of the SC. This paper constructs an evaluation index system based on the characteristics of the SC financing model, divided into 3 primary indicators (F1 to F3), 14 secondary indicators (S1 to S14) and 41 tertiary indicators (T1 to T41), as shown in Figure 1 Figure 2 Figure 3 respectively, in which the financial situation of the financing enterprise itself (T1 to T21) is used as the traditional credit risk assessment index, and the financing enterprise, the core enterprise, the SC The situation (T1 to T41) is used as the credit risk assessment index of SC finance. The set of credit risk indicators of SC financing enterprises is shown in Figure 1.

Fig.1 Subset of evaluation characteristics of SC financing enterprises

The set of credit risk characteristics indicators of core enterprises in the SC is shown in Figure 2.

The set of characteristic indicators for product SC performance evaluation of financing enterprises is shown in Figure 3.

Fig.2 Subset of evaluation characteristics of SC core enterprises
B. Description of the algorithm

The SC finance model contains both quantitative and qualitative indicators for the evaluation of participation risk, and the number is large, and the correlation between attributes is strong and redundant, these will certainly reduce the classification accuracy and speed of the evaluation model [13]. The evaluation model must retain as many indicators with high information content as possible and eliminate redundant and noisy attribute values in order to reduce the computational complexity of classification and thus improve the classification accuracy of the model [14]. This section uses particle filtering algorithm for evaluation index selection and SVM parameter co-optimization to construct a SC financial risk evaluation model.

C. Determination of the fitness function

The effectiveness of the selected indicators is evaluated by the classification performance of the SVM classifier, i.e., the classification accuracy of the classifier is used as the evaluation criterion. The particle filter is used to collaboratively optimize the set of evaluation indicators and SVM parameters, and the quality of the set of evaluation indicators and SVM parameters is estimated by the evaluation function, and the optimal set of evaluation indicators and SVM parameters are output as the result of the search for the best.

D. Particle coding scheme

The particles cover two parts, the feature vector and the SVM parameter values. The core of feature selection is to select attribute values from $A$ attributes to form a subset of attributes $B \leq A$. Therefore, the first part of the particle is encoded in a discrete binary variable, with each of the $A$ attributes corresponding to a one-dimensional binary space. For each particle, if bit is 1, then the attribute is selected; if it is 0, then the attribute is not selected. For example, particle $K = (110010001)$ means that of the 10 attributes, the selected attributes are 1, 2, 6, 10.

The rest of the attributes are not selected. The second part of the particle is the kernel parameters of the SVM. In this paper, the radial basis kernel function is chosen, and the parameters include the kernel function parameter $C$, the penalty parameter $g$, which is optimized using the continuous PSO algorithm. In the iterative process, the discrete PSO algorithm and the continuous PSO algorithm produce different credit feature subsets and parameter values, and the algorithm uses the SVM classification accuracy as the evaluation criterion, and the feature subsets and SVM parameters obtained when the accuracy is the highest are the desired ones. (See Figure 4).

![Fig.4 Particle representation of feature subset and SVM parameters](image-url)
**E. Algorithm implementation process**

1) Initialize the particle swarm, each particle consists of a subset of credit features, penalty parameter and kernel function parameter. Initialize the particle swarm parameters, including setting the learning factor, particle length, maximum number of cycles, etc.

2) Initialize particle swarm speed.

3) Transforming the values of each part of each particle into the corresponding selected credit feature subset mask according to the particle coding scheme and obtaining the parameter values, and calculating the fitness of each particle according to the selected feature subset and parameter values.

4) Update $P_i$ and $P_g$ according to the value of particle fitness;

5) Update particle velocity $V_i$ and position $X_i$;

6) If the iteration reaches the maximum number of iterations, continue iterating.

7) Output the current optimal feature subset, parameter $c_i$, $g$ and classification accuracy.

**III. INFORMATION GAIN**

Information gain is an important metric for feature selection to measure the extent to which information uncertainty is reduced [15]. The clearer the influence of a feature's information on the classification result, the greater the contribution of the feature in the classification decision, and the greater the corresponding information gain value. The so-called information is entropy.

The process of constructing a credit risk classification and prediction model for SC financing based on this paper is shown in Figure 5. In the process of designing the network operation, the complete SVM model was finally determined after several experiments and debugging.

![Fig.5 The proposed classification prediction](image-url)
The detailed steps of the model run are as follows.

Step 1 Construct a credit risk evaluation index system from five aspects: the SME itself, the core enterprise, the financing project, the trade SC and the macro environment.

Step 2 Obtain data on financial and non-financial indicators of enterprises through the National Stock Transfer System, a commercial bank management system in Tianjin and questionnaire distribution respectively to form the original data set \( \{D_1, D_2, \ldots, D_m\} \).

Step 3 The original data set \( \{D_1, D_2, \ldots, D_m\} \) is feature selected by IG to obtain a ranking of the information gain values of each indicator, and the best indicator is selected to form a new group data set \( \{I_1, I_2, \ldots, I_m\} \).

Step 4 Normalize the experimental data, the result of normalization is that the original data is regularized to the range of \([0,1]\) to obtain the dataset \( \{G_1, G_2, \ldots, G_m\} \), formula \( x' = \frac{x_i - \min_i}{\max_i - \min_i} \), where \( \max_i \) is the maximum value and \( \min_i \) is the minimum value in the sample data. The SVM model is trained using the new set of data sets and the classification prediction model \( f(\{I_i\}) \) is obtained by repeating the experiment with parameter search.

Step 5 Inputs the risk evaluation indicators of the unknown company into the \( f(\{I_i\}) \) prediction model and obtains the prediction results.

IV. EXPERIMENTAL ANALYSIS

A. Indicator system

By summarizing and concluding the assessment indicators used in previous literature and combining the characteristics of credit risk in pharmaceutical SC finance\[16\]. Therefore, the variables were first subjected to factor analysis to extract the variables with the main resolving power, and then the resulting variables were used for empirical analysis.

B. Parameter settings and performance evaluation metrics

The experimental model involved in this paper was implemented by programming in the Python language under Linux. The data used are new data after normalization, and the new data are highly compact with the original data. Any function that satisfies the Mercer condition can be used as a kernel function. In this paper, the four most representative kernel functions are selected: radial basis kernel function and Sigmoid kernel function to compare the classification results of the model and determine the optimal classification kernel function. The detailed parameters are shown in Table 1.

<table>
<thead>
<tr>
<th>SVM Kernel function</th>
<th>RBF; LKF; PF; Sigmoid kernel function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consume</td>
<td>0.26,266,1025</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.0626,0.26,5</td>
</tr>
</tbody>
</table>

For the purpose of evaluating the classification performance of a model, this paper uses the Confusion matrix to characterize the classification effectiveness of a classifier, and the results are characterized by the classification correctness (Accuracy, Acc).

Let TP be the number of predict actual defaulters as defaulters, TN be the number of predict actual compliance firms as compliance firms, FP be the number of predict actual compliance firms as defaulters, and FN be the number
of samples that predict actual defaulters as compliance firms. The confusion matrix representation is shown in Table 2.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Forecast enterprise default</th>
<th>Forecast enterprise performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual enterprise default</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Actual enterprise performance</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

C. Data collection and allocation

SC financing business is different from traditional bank financing products in that it selects well-qualified SMEs as credit targets based on the credit guarantee of core enterprises in the SC [17]. Given the unique attributes of SC financing business, the sample enterprises should be selected from SMEs in the same industry with obvious SC patterns. This paper examines the credit status of 137 SMEs in the electronic technology industry in Beijing, Tianjin, Shanghai and Shenzhen from 2015 to 2017 to examine their repayment of short-term loans and accounts payable. The financial data of the sample enterprises were obtained from the National Stock Transfer System. Qualitative indicators were obtained in the form of questionnaires, which were distributed to senior executives, including department managers of the target enterprises, through each city’s high-tech zone committee [18]. A total of 400 questionnaires were distributed in this study and 376 were returned, with 357 valid questionnaires. The data was collected and collated to obtain a total of 357 sample points, including 72 default samples and 287 compliance samples.

Usually, a certain type of sample tends to be overwhelmed by the small sample size, resulting in reduced model stability, therefore, in this paper, a balanced sampling of the sample data is carried out, balancing the ratio of defaulting and performing enterprises to approximately 1:1 [19]. On the basis of an equal number of samples in both categories, the sample data are divided into a training set for modelling and a test set for extrapolation testing in a number ratio of 8:2. In order to expand the sample base and avoid the occurrence of chance, three sets of data were randomly selected from the overall sample for the experiment to make the experimental model more stable and extrapolative. This was done by: Firstly, 72 samples were randomly selected from the 287 compliance samples to match the number of default samples 72, forming a new overall sample. Then 56 samples were randomly selected from the default and compliance samples of the new sample set as training samples to construct the SVM classification prediction model, while the remaining 16 default samples and 16 compliance samples formed the test samples to test the model performance. This operation was repeated three times. The final weighted average was used to reflect the classification performance of the model (Table 3).

D. Results and Discussion

In order to explore the kernel functions suitable for this model, different kernel functions were substituted into the SVM model one by one to compare the classification prediction effect, and the results were characterized by the Acc values of the sample set. The empirical results are shown in Table 4.

<table>
<thead>
<tr>
<th>Kernel function</th>
<th>Training value/%</th>
<th>Test value/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigmoid kernel function</td>
<td>74.79</td>
<td>63.22</td>
</tr>
<tr>
<td>PF</td>
<td>75.58</td>
<td>70.52</td>
</tr>
<tr>
<td>KF</td>
<td>78.36</td>
<td>68.21</td>
</tr>
</tbody>
</table>
As can be seen from Table 4, the models constructed using the Sigmoid kernel function, the Polynomial kernel function and the Linear kernel function do not differ significantly in their correct classification rates for both the training and test samples, indicating that the difference in model performance between these three kernel functions is small, while the Acc values of the SVM classification prediction models constructed with the RBF kernel function are higher, with the training sample set Acc value of 84.12% and the test sample set value of 81.72% [20]. In addition, the Sigmoid kernel function is ineffective in taking certain parameters, and the selection of parameters is difficult. In addition, the RBF kernel function has the following advantages: it can effectively deal with non-linear problems and has a moderate number of parameters, which makes it easier and faster to debug the parameters. Therefore, in this paper, the RBF function is chosen as the kernel function for building the model.

The model was used to predict the classification of the test sample, with the new data set as the input variable and the creditworthiness of the enterprise as the output variable (+1 for defaulting enterprises; -1 for performing enterprises), and the correct classification rate was 97.63% [21]. In order to more intuitively portray the efficacy of the model in this paper, it is proposed to compare the classification prediction results of this paper with those of other common classifiers in the same context, and the evaluation results are distinctive by the correct classification rates of the default and compliance samples. The specific implementation results are shown in Table 5.

Table 5 Comparison of classification efficiency of different models

<table>
<thead>
<tr>
<th>Model</th>
<th>Training sample</th>
<th>Test sample</th>
<th>Default sample</th>
<th>Performance sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.9241</td>
<td>0.8866</td>
<td>0.8748</td>
<td>0.8656</td>
</tr>
<tr>
<td>KPCA-SVM</td>
<td>0.9448</td>
<td>0.8352</td>
<td>0.8451</td>
<td>0.8108</td>
</tr>
<tr>
<td>IG-SVM</td>
<td>0.9826</td>
<td>0.9763</td>
<td>0.9322</td>
<td>0.9153</td>
</tr>
<tr>
<td>BP</td>
<td>0.8121</td>
<td>0.7660</td>
<td>0.8872</td>
<td>0.8660</td>
</tr>
<tr>
<td>KPCA-BP</td>
<td>0.8080</td>
<td>0.7776</td>
<td>0.8430</td>
<td>0.8290</td>
</tr>
<tr>
<td>IG-BP</td>
<td>0.8458</td>
<td>0.7906</td>
<td>0.9067</td>
<td>0.8912</td>
</tr>
<tr>
<td>SGD</td>
<td>0.9404</td>
<td>0.7560</td>
<td>0.7138</td>
<td>0.7110</td>
</tr>
<tr>
<td>KPCA-SGD</td>
<td>0.9345</td>
<td>0.7331</td>
<td>0.7656</td>
<td>0.7022</td>
</tr>
<tr>
<td>IG-SGD</td>
<td>0.9805</td>
<td>0.8458</td>
<td>0.7805</td>
<td>0.8207</td>
</tr>
<tr>
<td>DT</td>
<td>0.9551</td>
<td>0.8108</td>
<td>0.8342</td>
<td>0.7598</td>
</tr>
<tr>
<td>KPCA-DT</td>
<td>0.9436</td>
<td>0.7952</td>
<td>0.8439</td>
<td>0.8110</td>
</tr>
<tr>
<td>our method</td>
<td>0.9708</td>
<td>0.8470</td>
<td>0.8817</td>
<td>0.8606</td>
</tr>
</tbody>
</table>

The samples in this paper can draw the following conclusions from the data results in Table 5.

(1) Among the classification correct rate of the single models of SVM, BP, SGD and DT, the classification correct rate of the SVM model for the training samples is 92.4%, which is slightly lower than the classification correct rate of the SGD and DT models for the training samples, but the classification correct rate of the SVM model for the test samples is the highest, up to 88.66% [22]. The purpose of model building is to predict the credit risk of a company, so the correct classification rate of the test sample should be relatively important. In addition, the SVM model and the BP model are more prominent in the correct classification rate of the default sample and the performance sample, with comparable classification effectiveness. The combined comparison results show that the SVM model outperforms the other three single models in terms of overall classification prediction.

(2) For the classification results after applying the IG model and the KPCA model respectively, the classification correct rates of this paper, IG-BP, IG-SGD and our method models are higher than those of the corresponding KPCA-SVM, KPCA-BP, KPCA-SGD and KPCA-DT models, both for training samples, testing samples and for default samples and performance samples correct rates [23]. This indicates that the IG model is more effective than the KPCA model in feature selection and can effectively optimize feature selection.
Using the IG model for feature selection of the input variables, the classification correctness of the new model was improved in different degrees compared with the original single model. Among them, this model has the highest classification accuracy, with 98.26% for training samples and 97.63% for test samples, which is 5.86% and 8.98% higher than that of the single SVM model, respectively [24]. Also, the excellent classification performance of the model in this paper is reflected in the comparison with the single SVM model on the classification correctness of default and compliance samples, with the model correctness improved by 5.75% and 4.98% respectively after applying IG. Similarly, comparing the classification prediction performance of the IG-BP, IG-SGD and our method models with the corresponding BP, SGD and DT single models, the correct classification rates were improved by 3.38% and 2.47%, 4.02% and 8.99%, 1.58% and 3.63% in the training and test samples, respectively; in the default and performance samples, the correct classification rates were improved by 1.96% and 2.53%, 6.68% and 10.98%, 4.76% and 10.09% respectively [25]. This further demonstrates that the use of IG for feature selection not only reduces the dimensionality of the model operations, but also effectively improves its classification effectiveness.

In summary, it can be seen that: when the feature selection method is fixed, the classification accuracy and extrapolation ability of the SVM model are better than those of the BP, SGD and DT models; when the basic classification model is fixed, the optimize of feature variables using the IG model is better than the optimize selection of feature variables using the KPCA model. Therefore, the combined model proposed in this paper has better predictive value and can meet the needs of commercial banks for accurate decision-making.

**E. Feature selection and parameter optimize evaluation**

In order to verify the performance of the given model, the following experiments are conducted in this paper. Comparing the traditional radial basis SVM (SVM-RBF), feature extraction using principal component analysis (PCA) (PCA-SVM-RBF) and particle filter feature extraction and parameter co-optimize (BPSO-SVM-RBF) for the classification performance of three classification models.

The results of KMO and Bartlett's test are listed in Table 6. It can be seen that the KMO test result is 0.797, which is greater than the lowest value of 0.6 suitable for factor analysis; meanwhile, the Bartlett's spherical test significance level value is 0 (<0.01). The above results all show that the group of data is suitable for principal component extraction. According to the total variance explained, the eigenvalues greater than 1 are selected as principal components, for the traditional credit risk assessment index system there are 6 principal components, for the SC finance credit risk assessment index system, it includes 11 principal components. The comparison results of the three evaluation models in different assessment index systems are listed in Table 7. From Table 7, it can be seen that: the obtained PCA-SVM-RBF model is more suitable than the SVM-RBF model in terms of default classifier parameters. -RBF model compared to the SVM-RBF model in the case of default classifier parameters \(c = 1, g = 0.5\) The classification accuracy is improved from 84.76% to 86.95%. The method uses particle filtering to select features, and the number of "1" is 9, which reduces the number of features from 41 to 9. The BPSO-SVM was used to optimize the values of SVM parameters \(c\) and \(g\), and the penalty parameter 54.43 and the kernel function parameter 0.1 were obtained [26,27]. This shows that the values of the SVM classifier parameters have a great influence on the classification accuracy. In addition, redundant features interfere with the performance of the classifier to a certain extent, and the use of particle filtering algorithm to select feature values is better than the traditional PCA dimensionality reduction method.

<table>
<thead>
<tr>
<th>KMO measurement coefficient</th>
<th>Bartlett sphericity test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi-square distribution</td>
</tr>
<tr>
<td>0.797</td>
<td>4682.010</td>
</tr>
</tbody>
</table>
Table 7.3 Comparison of evaluation models in different evaluation index systems

<table>
<thead>
<tr>
<th>Evaluation model</th>
<th>Traditional credit risk assessment index system</th>
<th>Credit risk evaluation index system of SC finance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total number of features</td>
<td>Number of selected features</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>PCA-SVM-RBF</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>our method</td>
<td>22</td>
<td>8</td>
</tr>
</tbody>
</table>

The credit risk assessment index system of SC finance and the traditional credit risk assessment index system are evaluated by three models respectively. The set of credit risk characteristics of the SC financing enterprises discussed in the traditional credit risk assessment index system has 21 attributes (T1-T21). The comparison shows that the classification accuracy of our method using the traditional credit risk assessment index system is 81.72%, while the classification accuracy of BPSO-SVM-RBF using SC finance credit risk assessment index system has a classification accuracy of 91.44%, an improvement of 9.73%. Meanwhile, the classification accuracy of both SVM-RBF and PCA-SVM-RBF models improved by 6.33% and 6.77%, respectively. The experimental results show that the SC finance credit risk assessment index system helps to improve the overall performance of the classifier.

V. CONCLUSION

The evaluation and selection of collaborative relationships among SC members is a very complex problem, and it is of great practical importance to establish a practical, scientific and operable method. In this paper, a particle filtering combined with SVM criterion, unaided by information gain, was used to establish an analysis and evaluation method with wide applicability. The binary particle algorithm is used to achieve simultaneous optimize of feature attribute selection and SVM key parameters, which effectively solves the impact of high-dimensional and redundant feature attributes and inaccurate classifier parameters on the classification model. The experimental results show that the risk evaluation model in this paper has better performance in solving the credit risk evaluation problem of SC finance.

ACKNOWLEDGMENT

This work no funding supported.

REFERENCES


