Credit Risk Assessment of Receivable Accounts in Industry Chain based on SVM

Huan Sun[1]

School of computer and information, Hohhot Vocational College
Inner Mongolia, 010051, China
E-mail: `sunhhvc@163.com`

Industrial chain finance is a financial innovation that can effectively integrate the financial needs of the upstream and downstream enterprises. The rapid development of industry chain finance has put forward higher requirements for corresponding risk management. Among them, how to evaluate the credit risk of industrial chain receivable accounts is a key link of industrial chain financial risk management. For the small sample size in a single industry chain, a support vector machine (SVM) suitable for small sample learning is introduced to assess the credit risk. The credit risk evaluation index system for the industrial chain accounts receivable is constructed by principal component analysis of the financial data. V-SVR (Support Vector Regression) model is employed to evaluate principal component of the receivable accounts in telecom industry chain. By comparing the evaluation results with historical data, it is found that the model can well predict the credit risk of commercial banks, and has higher prediction accuracy than Logistic model.

[1]Speaker

1.Introduction

Industrial chain Finance is an extension and expansion of supply chain finance. Supply chain finance, simply put, is to connect the core enterprises with upstream and downstream enterprises to provide a financing model for the flexible use of financial products and services. The industrial chain finance is not only limited to the horizontal upstream and downstream of the various links, design standardization, personalized financial services products, but also the vertical reorganization of mergers and acquisitions and other financial activities. For the enterprises in the industrial chain, it is helpful to improve the overall operation efficiency of the industrial chain and promote the transformation and upgrading of the industry because of obtaining more abundant financial support.

The establishment and vigorous development of industrial chain finance has aroused extensive attention of many scholars. For example, some researchers have focused on an important branch of the industry chain Finance, supply chain finance, from the supply chain finance strategy [1], financial optimization [2] and supply chain financial management [3], and so on. Of course, the development of the industry chain Finance also put forward the higher request to the corresponding risk management technology, among them, the credit risk assessment of the industrial chain receivables is an important problem that is needed to be solved before the financial institution. At present, the most influential credit risk measurement model in the world has KMV Company's KMV model, J. P Morgan's metrics model and the McKinsey Portfolio view model. However, the number of the samples of the industrial chain accounts receivable is relatively small, and the related credit status of the effective data loss, resulting in these models cannot be used in China's commercial banks credit risk assessment of the field of effective application. In view of the above characteristics of the credit risk assessment of the industrial chain receivables, this paper will try to evaluate the small sample learning algorithm using support vector machine.

Since the support vector machine algorithm is suitable for the classification of small sample problems and has high precision, it has been favored by researchers at home and abroad, and some scholars have gradually applied it to the field of credit risk assessment of commercial banks since its related theories were founded. For example, for the Turkish commercial banking system, Melek et al. respectively use neural network technology, support vector machine and multivariate statistical methods to predict the possibility of bank insolvency, the different financial indicators to the data set to improve the predictive performance. The results show that the SVM model can be regarded as the most successful model to predict the financial bankruptcy of commercial banks [4]. Support vector regression machine has been applied in some fields of financial research and has high prediction precision. Harris compared support vector machine (SVM) based credit-scoring models built using Broad and Narrow default definitions, and the results of empirical revealed that credit risk evaluation at the Barbados based institution could be improved if quantitative credit risk models are used as opposed to the current judgmental approach.[5]. After nearly ten years of development, support vector regression machine has been derived from many types, including ε-SVR, C-SVR, v-SVR and so on. Among them, v-SVR is a kind of support vector machine with high accuracy, and it has been applied gradually in financial field. In the prediction of gas price, by using the ε-SVR, v-SVR model and artificial neural network model, RBF network model and asymmetric GARCH

model, Hu and Trafalis found that the prediction accuracy of v-SVR model with sigmoid kernel function is highest [6].

Therefore, aiming at the measurement of the credit risk of receivables accounts in industrial chain, it is necessary to forecast the indexes with high accuracy, and the sample quantity is few, this paper attempts to use the v-SVR model to estimate the credit risk of the industrial chain receivables, and compare the result with the forecast result of Logistic model to verify the predictive effect of the model.

2.Support Vector Machine Model

Support vector machine (Support Vector Machine, SVM) [7] is a theory developed on the basis of statistical learning theory, which specializes in the study of machine learning in small samples, and is an important development and supplement of traditional statistics.

2.1Support Vector Classifier Algorithm

On the basis of the classification method, Vapnik [8] proposed ε-Support Vector Regression. He added and subtracted ε from y value of each training point in training set T, and obtained two sets of positive and negative points respectively. Thus the hard ε-band hyperplane is equivalent to the hyperplane whose training set is correctly partitioned completely. The regression problem can be transformed into a classification problem. On the basis of ε-SVR, Schölkopf et al. [9] proposed the v-SVR model, and improved the support vector regression machine. The specific algorithm is as follows:

Step 1: Given training set $T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (R^n \times R)^l$, where $x_i \in R^n, y_i \in R, i = 1, 2, \cdots, l$;

Step 2: Select the appropriate kernel function $K(x, x')$, parameter $v \in (0, 1]$ as well as C>0;

Step 3: Construct and solve the convex quadratic programming of optimization problems:

$$\min_{\alpha^{(*)}} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\alpha_i^s - \alpha_i)(\alpha_j^s - \alpha_j) K(x_i, x_j) - \sum_{j=1}^{l} (\alpha_j^s - \alpha_j) y_j,$$

$$s.t. \sum_{i=1}^{l} (\alpha_i^s - \alpha_i) = 0, 0 \leqslant \alpha^{(*)} \leqslant C/l, i = 1, 2, \cdots, l, \sum_{i=1}^{l} (\alpha_i^s + \alpha_i) \leqslant Cv \qquad (2.1)$$

get solution: $\overline{\alpha^{(*)}} = (\overline{\alpha_1}, \overline{\alpha_1^s}, \cdots, \overline{\alpha_l}, \overline{\alpha_l^s})$;

Step 4: Calculate $\overline{b}$: select the two components $\overline{\alpha_j}$ and $\overline{\alpha_k}$ of $\overline{\alpha^s}$ in the open interval. Accordingly,

$$\overline{b} = \frac{1}{2}[y_i + y_k - [\sum_{i=1}^{l} (\overline{\alpha_i^s} - \overline{\alpha_i}) K(x_i, x_j) + (\overline{\alpha_i^s} - \overline{\alpha_i}) K(x_i, x_k)]] \ . \qquad (2.2)$$

Step 5: Construct decision function

$$g(x) = \sum_{i=1}^{l} (\overline{\alpha_i^s} - \overline{\alpha_i}) K(x_i, x) + \overline{b} \ . \qquad (2.3)$$

The kernel function $K(x, x')$ transformed the nonlinearity into a linear classification in high dimensional space, so the kernel function plays an important role in the nonlinear classification. Considering the randomness of training samples, the training samples are non-

linearly separate, thus using an inner kernel function algorithm. A Gauss radial base function (RBF) is selected in this paper, namely:

$$K(x, x') = \exp(\|x - x'\|^2 / \sigma^2) \quad . \tag{2.4}$$

Compared with the traditional ε-SVR, ν-SVR uses a quantitatively meaningful parameters ν to replace the parameter ε, where ν is the upper bound of the percentage of the error training points and the lower bound of the number percentage of support vector machines. In addition, under certain conditions, when the number of training points $l \rightarrow +\infty$ in training set T, ν is close to the ratio of the number of support vectors to the number of training points as the probability of 1. This design allows the model to be optimized and to achieve higher computational accuracy. Therefore, this paper mainly focuses on the training in ν-SVR model to test its effect in the credit risk forecast of the industry chain accounts receivables.

2.2 Input and Output Indexes

Level 1 indicators	Level 2 indicators	Level 3 indicators
Basic situation of enterprises	Basic quality of enterprises	Enterprise management status
		Enterprise scale
		Enterprise status
		Financial disclosure quality
Financial situation of company	Profitability	ROE
		Ratio of income as a percentage of sales
	Operating capacity	Turnover capacity
	Short term debt service	Quick ratio
	Long term debt service	Interest protection multiples
		Asset-liability ratio
		Long-term asset fitness ratio
	Development potential	Sales revenue growth rate
		Net profit growth rate
		Total assets growth rate
Assets under financing	Accounts receivable	Accounts receivable turnover ratio
	Stock	Inventory turnover ratio
	Prepayment turnover ratio	Prepayment turnover ratio
Supply chain operation condition	Industrial chain management	Basic situation of upstream enterprises
	Operation of logistics enterprises	Process risk
	Compliance	Industrial chain compliance risk
Industry status	Macro environment	Macroeconomic situation
		Legal policy environment
	Industry development prospect	Industry development stage
		Industry competition intensity
		Industry environment

Table 1: Input Index

The indicators include qualitative indicators and quantitative indicators, in which the quantification of qualitative indicators is divided into 0-5 values according to the five-level quantification standard and the specific values are determined according to the relevant influencing factors.

According to the principle of systematicness and comprehensiveness, flexibility and practicability, conciseness and scientificalness and so on, the paper analyzes the influencing factors of bank credit risk, and then draws on the relevant indexes of literature at home and abroad. This paper constructs the evaluation system of credit risk index of accounts receivable in industrial chain, selecting 14 financial indicators which are closely related to the credit risk as model input indicators. The contents of the input indicators and their contents are shown in Table 1.

Since the data units in this paper are not uniform, the data is normalized first: in the sample data, the maximum value is $max\{x_i\}$, the minimum value is $min\{x_i\}$. The original data is normalized as new data：

$$\overline{x_i} = (x_i - min\{x_i\})/(max\{x_i - min\{x_i\}\}) \quad . \tag{2.5}$$

After the data is normalized, the new data can be used as the input data of the v-SVR model, and the risk level of the small and medium-sized enterprises is the output of the model. In order to avoid the high correlation and the overlapping of information between the financial indicators including bank credit risk information, and taking into account the efficiency of the support vector regression machine in training, this paper uses the principal component analysis.

Based on the bad debt provision of accounts receivable as the output of the measure of Y, $y \in \{0,1\}$. Y=0 means the company's bad debts are small. To a certain extent, it reflects less bad debt reserves in accounts receivable, which means the smaller risk level. Here are the "credible" enterprises. The same way, Y = 1 indicating that the company's bad debts, to a certain extent, reflects more accounts receivable need for bad debts, that is, high level of risk. Here are "incredible" enterprises.

3.Support Vector Machine Model

3.1Data Preprocessing and Principal Component Analysis

Component	Initial Eigenvaluesa			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	2.972	18.576	18.576	2.972	18.576
2	2.616	16.347	34.923	2.616	16.347
3	2.214	13.839	48.762	2.214	13.839
4	1.659	10.369	59.131	1.659	10.369
5	1.297	8.104	67.235	1.297	8.104
6	.991	6.196	73.431	.991	6.196
7	.866	5.415	78.845	.866	5.415
8	.769	4.809	83.654	.769	4.809

Table2: Eigenvalue and variance contribution rate

In order to evaluate the credit risk of the receivables in the industrial chain more accurately, this paper takes the small and medium-sized enterprises with more frequent accounts receivable in China's electronic industry as the research object. Samples are from listed small and medium-sized enterprises whose market value is under 20 billion in 2016 electronics industry The data of 75 eligible small and medium-sized enterprises is from the company disclosed annual statements, JUYUAN database and TONGHUASHUN software.

The cumulative contribution rate of the extracting main component is up to 80%. According to it, so many original dimensions of the indicators are reduced to a small number of comprehensive indicators. The results of principal component analysis of financial indicators using SPSS 17.0 software are as follows:

Considering the principle that the eigenvalues are more than 1 and the cumulative contribution rate is more than 80%, the first eight factors are selected as the principal components, and the accumulated information reaches 83.6%.

3.2 Logistic Regression Analysis and Results

First, the logistic regression model is used to evaluate the credit risk of accounts receivable so as to compare with the classification effect of v-SVR model. From the above, this paper uses the principal component analysis method to deal with the 16 indexes related to the credit risk of the industrial chain financial receivables, and selects the first eight factors as the main component. Furthermore, Logistic regression is performed on the eight principal components. The regression method uses forward stepwise regression (Wald method). Finally the first main factor enters the equation, the others are removed.

		B	S.E.	WALD	df	Sig	Exp(B)
	FAC1_1	.918	.406	5.11	1	.024	2.51
Step1a	Constant	1.31	.298	19.440	1	.000	3.71

a. The variable entered on step 1：[%1:, 1:]

Table 3: Variables in equations

The results of the logistic regression available from Table 3 are as follows：

$$P = \frac{e^A}{1 + e^A} \quad,$$

where

$$A = 1.3 + 2.8x_1 + 0.24x_2 + 0.67x_3 + 0.87x_4 + 0.71x_5 + 0.20x_6 + 0.15x_7$$
$$+ 0.012x_8 + 0.0784x_9 + 0.16x_{10} + 0.034x_{11} + 0.79x_{12} + 0.023x_{13} + 0.12x_{14}$$
$$+ 0.092x_{15} + 0.11x_{16} + 0.052x_{17} + 0.043x_{18} \quad.$$

Observed values		Predictive value			
		y			Correct percentage
			0	1	
Step 1	y	0	2	15	11.8
		1	2	56	96.6
	Overall percentage				77.3

a. The split value is 500

Table 4: Logistic regression results classification

According to the above formula, the classification results calculated by SPSS software are shown in Table 4. The accuracy rate of the logistic regression model is 77.3%, and the accuracy rate of identifying "incredible" enterprises of the small and medium-sized in the electronic

industry is 96.6%, and the accuracy rate of identifying "credible" enterprises is 11.8%; that is, the first type of error rate of 3.4% and the second type of error rate of 88.2%.

Among them, the first type of error refers to the judgment of "incredible" small and medium financing enterprises (1) as "credible " enterprises (0). The second type of error refers to the judgment of "credible" small and medium financing enterprises (0) as "incredible " enterprises (1).

3.3 Analysis of v-SVR Model Predicted Results

In this paper, Matlab2016a software is used for experimental analysis, and the v-SVR model is trained in an SVM toolbox. Considering training accuracy and time, the parameters C and v in the main models of the v-SVR model and the parameter σ in the kernel function are optimized. Appropriate precision is selected, and traversal searches are performed respectively at the steps of 0.1, 0.01, and 0.01, so as to achieve the best training results.

A sample set is constructed according to the foregoing analysis; $x \in R^8$ is an input index, that is, an eight-dimensional vector; and $y \in R$ is an output index of the sample, that is, provision for bad debts of accounting receivable. Eight principal components Factot1-Factot8, which are calculated by using the principal component analysis are used as input variables of the v-SVR model. "0" is used to indicate that small and medium-sized financing enterprises in the electronics field belong to the "Trusted" credit risk customer type, "1" is used to indicate that the small and medium-sized financing enterprises belong to the "Untrusted" credit risk customer type, and {0,1} is the output variable of the model.

The v-SVR model derives from the machine learning method in the statistical learning theories, and the method needs to divide the samples into two categories, that is, the training sample and the test sample. First, the training sample is input and output, and the model automatically learn and get the law. Then, the law is tested during test sample input and output, so as to determine whether the results of model training can be well promoted, that is, generalization capability of model. In order to avoid the occurrence of contingency results, this paper draws the two categories of samples ten times by random selection, and obtains 10 groups of different training sample sets and test sample sets. The sample set distribution and the test results of the ten tests are as follows respectively:

Sample sets	Scale of the sample sets	Number of the "Trusted" samples	Number of the "Untrusted" samples
Training samples	57	12	45
Test samples	18	0	18
All samples	75	12	63

Table 5: Distribution of the first group of sample sets

The distribution of the other nine sample sets is the same.

Sample number	Training samples		Test samples		Entire samples
	Type I	Type II	Type I	Type II	Type I

1	3.5%	8.3%	0%	0%	2.8%
2	0%	0%	4.7%	20%	2.8%
3	0%	0%	0.9%	11.1%	2.8%
4	0%	0%	2.3%	14.3%	2.8%
5	1.4%	11.1%	0%	0%	2.8%
6	0%	0%	4.6%	25%	2.8%
7	1.8%	7.1%	0%	0%	2.8%
8	0%	0%	3.2%	14.3%	2.8%
9	2.5%	7.7%	0%	0%	2.8%
10	0%	0%	4.1%	20%	2.8%
Average	2.3%	3.42%	3.3%	10.47%	2.8%

Table 6: Test results of the ten samples

It can be seen from the table above that in the entire samples, the error rate of Type Ⅰ is 2.8%, the error rate of Type Ⅱ is 5.9%. This means the accuracy rate of distinguishing the "Untrusted" small and medium-sized financing enterprises is 100% and the accuracy rate of distinguishing the "Trusted" small and medium-sized financing enterprises is 94.1%, which shows that the model is remarkably effective. The empirical results indicate a high accuracy rate of the credit risk assessment that is performed, by using the PCA-SVM model, on the financial accounting receivable in industrial chain of small and medium-sized financing enterprises.

3.4 Comparison Between v-SVR and Logistic Regression Results

The results of the credit risk assessment performed on the accounting receivable by using the two models are compared as follows:

Industrial chain assessment method	Entire samples Accuracy rate of overall classification	Entire samples The first type of error rate
Logistic regression	77.3%	3.4%
v-SVR model	94.1%	2.8%

Table 7: Comparison between the two models

As can be seen from the table above, under the index system of credit risk assessment of financial accounting receivable in the industrial chain, the accuracy rate of the overall classification of the entire samples by using Logistic regression is as high as 77.3%, with the first type of error rate of 3.4% and the second type of error rate of 88.2%, and the overall classification of the entire samples by using the SVM model has an accuracy rate of 94.1%, with the first type of error rate of 2.8% and the second type of error rate of 29%. Therefore, the SVM model is superior to the Logistic model, both from the overall prediction accuracy rate and from the error rates of the two types.

For financial institutions such as trust, the first type of classification errors is more important. It can be seen from this paper that the first type of error rate in a v-SVM evaluation model is lower than the first type of error rate in a Logistic regression model. Therefore, the v-SVM evaluation model more meet the requirements of trust and other financial institutions to forecast the credit risk of small and medium-sized financing enterprises.

4. Conclusion

Support Vector Machine (SVM) is a general learning algorithm based on small sample learning theory. It has a rigorous theoretical basis and can solve practical problems such as small samples, non-linearity, high dimension and local minimum points. After years of development,

the support vector machine has got various improved forms, and v-SVR is a kind of support vector regression machine with high prediction accuracy. In view of the low number of listed companies in the telecom industry, this paper introduces the v-SVR to assess the credit risk of the accounting receivable in the industrial chain. The assessment index system of credit risk of the accounting receivable is established by using the principal component analysis, and the credit risk assessment model of the credit risk of the accounting receivable in the industrial chain is constructed based on v-SVR. Compared with the Logistic model, it is found that the v-SVR model has higher accuracy in training and prediction. Therefore, the credit risk assessment model of the accounting receivable in the industrial chain based on v-SVR provides a more useful auxiliary tool for the credit risk management of commercial banks, and can be promoted to more areas for risk assessment.

References

[1] M Christopher, L Ryals. *Supply chain strategy: its impact on shareholder value*[J]. The International Journal of Logistics Management, 1999, 10(1): 1-10.

[2] H C Pfohl, M Gomm. *Supply chain finance: optimizing financial flows in supply chains*[J]. Logistics Research, 2009, 1(3-4): 149-161.

[3] M L Gomm. *Supply chain finance: applying finance theory to supply chain management to enhance finance in supply chains*[J]. International Journal of Logistics: Research and Applications, 2010, 13(2): 133-142.

[4] A B Melek, K Yakup, K B Omer. *Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods*[J]. Expert Systems with Applications, 2009, 36(2): 3355-3366.

[5] T Harris. *Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions*[J]. Expert Systems with Applications, 2013, 40(11): 4404-4413.

[6]Y Hu, T B Trafalis. N*ew kernel methods for asset pricing: application to natural gas price prediction*[J]. International Journal of Financial Markets and Derivatives, 2011, 2(1-2): 106-120.

[7] C Cortes, V Vapnik. *Support-vector networks*[J]. Machine Learning, 1995, 20(3): 273-297.

[8]V N Vapnik. *The Nature of Statistical Learning Theory. Springer-Verlag*, New York, 1995.

[9] B Schölkopf, R C Williamson, P L Bartlett. *New support vector algorithms*[J]. Neural Computation, 2000, 12(5): 1207-1245.