SVM-based Abnormal Account Monitoring Model of Bank

Xue-Zhi QIN[1,a], Jing-Yi LI[2,b,*], You-Qun HU[3,c]

[1]College of Business Administration, Dalian University of Technology, 116024,China

[2]College of Business Administration, Dalian University of Technology, 116024,China

[a]qinxuezhi@gmail.com, [b]lijingyi0128@aliyun.com, [c]huq308@163.com

*Corresponding author

Abstract. The account transaction flow of bank loan customers reflects the capital transaction of customers and represents their management level. Normal account transaction flow shows regularity to some extent, while abnormal account transaction flow manifests itself differently in transaction amount and transaction frequency et al. When compared with those of normal account transaction flow, abnormal account usually shows abnormal change of fund flow caused by money laundering, hot money, private lending or bad management. In this study, based on identification approach between normal account and abnormal account, the characteristics of different industry transaction flow are drew out, and then the classification for normal and abnormal accounts of different industry is well established by the Support Vector Machine model. Thus, the loan monitoring and risk management can be conducted more appropriately and efficiently.

Introduction

The capital flow of bank account is needed for normal demand of the business activities. However, many abnormal or illegal capital flows may be mingled, which may bring challenges to perfecting finical supervision system and improving supervision method.

Until to now, some studies have been conducted on monitoring finical capital flow: Liu et al.[1] distinguished account transaction types through scan statistic method. This method had an effect on distinguishing abnormal transaction. The accuracy of the model was strongly relative with the parameter, and the model parameter setting is strongly subjective, Yang[2] studied the credit card fraud detection model based on Radial-Basis Function Support Vector Machine (RBF-SVM). This model has some flaw that the performance of the model is unstable, Through the RBF (Radial-Basis Function) neutral network technique, Lv et al.[3] proposed a RBF neutral network model for anti-money laundering based on APC-III clustering algorithm and RLS (Recursive Least Square) algorithm. The sensitivity of the accuracy of the model to the parameter variation was strong. For commercial banks, Wang[4] suggested to use the theory of decision tree to identify suspicious customers. Results showed that this method is effective for generating anti-money laundering rules from customer information, while the setting of rules is relatively subjective. Zhu[5] proposed the EMD method for the decision of the abnormal capital flow based on historical information in the same group. The accuracy and confidence level for the model was not calculated. Taking Bank of the Malay Archipelago as an example, Aspalella[6] studied the influence of setting reporting system for suspicious transactions on bank risk monitoring. Pramod et al.[7] structured the regulation frame of anti-money laundering for the bank, and they considered that the information system based on the process control played a critical role in finding fraudulent transactions.

Current techniques effectively used in the monitoring of abnormal capital are the Support Vector Machine (SVM) model and neutral network technique et al[8]. Some drawbacks of the neutral

network algorithm are discovered: (1) the model assumption is too strict, for example, uniform distribution of the trained samples is usually required, (2) dimension disaster: overabundant characteristic numbers may cause the algorithm poorly efficient or inefficient, (3) the accuracy of the results: the results from the neutral network algorithm are highly dependent on the prior knowledge and experience of the user.

The advantages of using SVM model in the monitoring of bank account flow are as follows: (1) SVM model can solve multidimensional sample issues through linear method. This feature is fit for the multidimensional characteristics of bank account. (2) The assumed conditions of SVM model is not as strict as that of the neutral network technique: the sequence of samples is not required by SVM model. (3) Compared with the neutral network technique, fewer parameters are needed for SVM model. Users can obtain satisfactory classification results from raw data provided that they have some prior knowledge and experience.

SVM Model of linearly separable sample set

Assume $g(x) = w_0^T x + b_0$, thus $g(x_p) = w_0^T x_p + b_0 = 0$,

$$g(x) = w_0^T x + b_0 = w_0^T x - w_0^T x_p = r \|w_0\|$$

(1)

where $g(x)$ represents the distance of sample x to the separable hyperplane. The decision function is:

$$f(x) = \text{sgn}(g(x))$$

(2)

It is vital to find the maximum geometrical distance r from the support vector to the separable hyperplane. The normal way for this issue is to fix the absolute value of $g(x)$ to be 1, and then minimize $\|w_0\|$.

The determination of the separable hyperplane can be boiled down to the quadratic programming problems (denoted as (QP)) as follows:

$$\min_{w,b} \ \frac{1}{2} w^T w$$
$$s.t. \ \ d_i(w^T x_i + b) \geq 1, \ i = 1, 2, ..., N$$

(3)

where N is the number of samples.

The Lagrange function for (QP) can be expressed as follows:

$$L(w,b) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i d_i (w^T x_i + b) + \sum_{i=1}^{N} \alpha_i$$

(4)

where α_i is Lagrange coefficients.

Let the partial derivative of Eq. 4 with respect w and b equal 0, then the question can be transformed into a simple dual question.

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i d_i x_i = 0$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i d_i = 0$$

(5)

Substituting Eq. 5 into the Lagrange function, the issue is transformed into a dual optimization problem[9]:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j d_i d_j (x_i^T x_j)$$

$$s.t. \quad d_i[(\sum_{j=1}^{N} \alpha_j d_j x_j^T) x_i + b] - 1 \geq 0$$

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$\alpha_i \geq 0, \quad i = 1, 2, ..., N$$

(6)

According to Karush-Kuhn-Tucker(KKT)[10],this optimization problem has to satisfy the following condition:

$$\alpha_i \{d_i[(\sum_{j=1}^{N} \alpha_j d_j x_j^T) x_i + b] - 1\} = 0, \quad i = 1, 2, ..., N$$

(7)

Thus, α_i equals zero for most samples. The samples satisfying Eq. 6 when $\alpha_i \neq 0$ is referred to as support vectors (SVs).

The support vector machine transforms the original problem into a dual problem. Once the (QP) problem is solved, the classification decision function can be expressed as follows:

$$f(x) = \text{sgn}(w^{*T} x + b^*) = \text{sgn}(\sum_{i=1}^{N} \alpha_i^* d_i (x_i^T x) + b^*)$$

(8)

The summation in Eq. 7 is only for support vectors. Furthermore, based on $\alpha_i \{d_i(w^T x_i + b) - 1\} = 0$, $i = 1, 2, ..., N$, we can obtain,

$$b^* = \frac{1}{d_i} - w^T x_i = d_i - w^T x_i = d_i - (\sum_{x_j \in SV} \alpha_j^* d_j x_j^T) x_i \qquad (9)$$

For a reliable calculation, b* is calculated for all standard support vectors, then the average value of b* is determined:

$$b^* = \frac{1}{N_{SV}} \sum_{x_i \in SV} [d_i - (\sum_{x_j \in SV} \alpha_j^* d_j x_j^T) x_i] \qquad (10)$$

where N_{SV} is the number of standard support vectors, SV is the assemble of support vectors.

Thus, the decision classification function can be expressed as follows:

$$f(x) = \text{sgn} \{ \sum_{j=1}^{N} \alpha_j^* d_j (x_j^T x) + \frac{1}{N_{SV}} \sum_{x_i \in SV} [d_i - (\sum_{x_j \in SV} \alpha_j^* d_j x_j^T) x_i] \} \qquad (11)$$

Account Flow Monitoring Model for Bank Loan Based on Support Vector Machine

Modelling Rules

(1) Quantitative criterion for the identification of abnormal capital flow: according to "Management Method of Large Trade and Suspicious Transaction Report for Financial Institution" [11] released by People's Bank of China on the year 2006, a single trade amount by the corporation, other organization or individual business above 2 million RMB is regarded as large trade. In this study, large trade with low frequency (less than 3 times in one year) is regarded as suspicious transaction.

(2) Strength criterion for the identification of abnormal capital flow: the strength criterion refers to the strength and velocity characteristics of the capital during capital's transfer process[12]. Burst capital flow refers to the fact that capital flow of account enlarges abruptly with unknown reasons, frequent capital flow refers to that the speed of capital flow increased significantly. According to the general experience, if the ratio of capital flow of the current month to the capital flow of last month is equal to 2 or exceeds 2, or the ratio of capital flow times of the current month to capital flow times of the last month is equal to 2 or exceeds 2, the capital flow is regarded as abnormal capital flow.

(3) Capital flow obviously inconsistent with operational business: The characteristics of one enterprise's capital flow are deemed to reflect the company scale. If the capital flow is obviously inconsistent with the company scale, the legality of operational business can be suspected, and the transaction can be categorized to suspicious transaction[13]. If the transaction flow is significantly greater than operating income, money laundering behavior is considered to be existing. "The classified standard for large, medium, small and micro enterprise on statistics"[14] released by China State Statistical Bureau on the year of 2011 provides standards on operating income limit for different industries and different types of enterprises. According to general experience on estimated proportion between capital flow and operating income, if the ratio of the annual capital flow to the annual operating income limit exceeds 11, this capital flow is regarded as abnormal capital flow.

Based on these three rules, the samples of normal capital transaction and abnormal capital transaction can be obtained.

Model construction

Research object: the capital flow of one customer's bank accounts.

Assumption: all bank accounts and all bank cards are integrated in one bank account.

Sample set 1: the capital flow record of one bank loan account. The type of industry is wholesale for vegetables and fruits, the number of capital flow is 665. For the capital flow of the same month and same flow direction, flows with the transaction amount variable rate lower than 10% are aggregated together. The number of capital flow is 160 after this treatment, among which 80 flows are used as training sets, and the other 80 flows are used as test sets (please refer to appendix). According to modelling rule 3, the annual operating income limit is 50 million RMB of this type of customer.

Sample set 2: the capital flow record of one bank loan account. The type of industry is agricultural and sideline products processing, the number of capital flow is 280. For the capital flow of the same month and same flow direction, flows with the transaction amount variable rate lower than 10% are aggregated together. The number of capital flow is 160 after this treatment, among which 80 flows are used as training sets, and the other 80 flows are used as test sets (please refer to appendix). According to modelling rule 3, the annual operating income limit is 20 million RMB of this type of customer.

Variables: Type of account holder (individual or enterprise), single capital inflow (outflow) amount/average single capital inflow (outflow) amount, transaction frequency, mark of account flow (normal or abnormal).

Interpretation of variables: the transactions can be differentiated into different types according to the type of account holder (individual or enterprise). In general, the transaction amount for enterprise transaction is large, thus the corresponding threshold value for abnormal transaction is large. On the contrary, the transaction amount for individual transaction is small, thus the corresponding threshold value for abnormal transaction is small. In this study, the type of bank account is enterprise. Through the three variables: single capital inflow (outflow) amount/average single capital inflow (outflow) amount and transaction frequency, transaction flow can be differentiated according to transaction amount and transaction frequency. Large trade with low frequency is regarded as abnormal transaction. The mark of account flow represent the attributes of the flow, with the value +1 for normal and -1 for abnormal.

Model: support vector machine model for linearly separable sample set.

Substituting the data of sample set 1 into Eq. 3~ Eq. 11, we obtain, $\sum_{i=1}^{N}\alpha_i=38.51, b^*=4.69$ and the maximum value of object function (Eq. 6) is 30267.26.

Substituting the data of sample set 2 into Eq. 3~ Eq. 11, we obtain, $\sum_{i=1}^{N}\alpha_i=28.15, b^*=10.85$ and the maximum value of object function (Eq. 6) is 4763.95.

Substituting the training sets and test sets according to these parameters into Eq. 8, we can draw the conclusion in Table 1:

Table 1 SVM model results

sample type	status	true value	predicted value	mistaken number of samples	type I mistaken number of samples	type II mistaken number of samples	the probability of mistaken samples	the probability of type I mistake	the probability of type II mistake
training set 1	normal	79	67						
	abnormal	1	13	12	12	0	15%	15.19%	0%
test set 1	normal	78	64						
	abnormal	2	16	14	14	0	17.5%	17.9%	0%
training set 2	normal	79	66						
	abnormal	1	14	13	13	0	16.3%	16.5%	0%
test set 2	normal	80	65						
	abnormal	0	15	15	15	0	18.8%	18.8%	0%

It can be seen from the table that for the training set 1, the accuracy of the model is 85%, and 82.5% for test set 1, 83.7% for training set 2, 81.2% for the test set 2. The performance of the model is convincing to some extent. Besides, The SVM model shows the satisfactory classification effect for samples with different industries and different sample capacities. The accuracy of SVM for wholesale of vegetables and fruits is better than that of agricultural and sideline products processing. The accuracy of SVM is proportional to the capacity of samples.

Summary

This paper presents a study on the application of SVM model on the monitoring of abnormal capital flow of bank account. SVM model can give real-time detection for abnormal capital flow, thus can help decision makers to take measures, such as freezing Quota and freezing capital. This application of SVM model enhances monitoring ability for suspicious transaction and money laundering transaction. Through establishment of SVM model for different industries, the classification rules for different industries are determined. So SVM model uncovered risk's degree of different industries, thus the industry with low risk can be given credit extension and the industry with high risk can be alerted and prevented. With the application of SVM model, banks' risk management level can be improved and finical stability can be promoted.

Acknowledgement

References

[1] Liu Xuan, Zhang Peng zhu, Li Jia, et al, Suspicious Transaction Recognition in Commercial Banks, Journal of Systems & Management. 5(2013)379-384.

[2] Yang Xi, Research on Credit Card Fraud Detection Based on Support Vector Machine, Sichuan Normal University, Chengdu, 2008.

[3] Lv Lin tao,Ji Na,Zhang Jiu long, Suspicious transaction detection model based on Radial Basis Function Neural Net-work, Computer Engineering and Applications. 46(2010)207-210.

[4] Wang Sunan, Yang Jiangang, A money laundering risk evaluation method based on decision tree, International Conference on Machine Learning and Cybernetics, 2007.

[5] Zhu Tianqing, Suspicious Financial Transaction Detection Based on Empirical Mode Decomposition Method, Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing, 2006.

[6] Aspalella A. Rahman, The impact of reporting suspicious transactions regime on banks: Malaysian experience, Journal of Money Laundering Control. 16(2013)159-170.

[7] Vandana Pramod, Jinghua Li, Ping Gao, A framework for preventing money laundering in banks, Information Management &Computer Security. 20 (2012)170-183.

[8] Ji Na, Research of Indentify Money-Laundering Transactional Behavior Base on RBF Neural Net work, Xi'an University of Technology, Xi'an, 2009.

[9] Peng Lu, Research on Classification Algorithm of Support Vector Machine and Its Applications, Hunan University, Changsha, 2007.

[10] V.Vapnik, The Nature of Statistical Learning Theory, NY:Springer-Verlag, 1995.

[11] Information on http://www.pbc.gov.cn.

[12] Xue Yaowen, The Study of Monitoring of capital abnormal flow in financial network, China Financial Press, Beijing, 2009.

[13] Liu Xuan, Research on Pattern Recognition and Intelligent Financial Supervision of Capital Flow in Financial Netwok, Shanghai Jiao Tong University, Shanghai, 2010.

[14] Information on http://www.stats.gov.cn/tjsj/tjbz/201109/t20110909_8669.html.

Appendix

Table 2 Training set 1(After integration)

Month	Transaction amount(CNY)	Frequency	Single capital flow amount/Average single capital flow amount	Mark of abnormal
201205	-1400000	1	3.05	1
201205	-1080000	3	2.35	1
201205	-900000	7	1.96	1
201205	-750000	4	1.63	1
201205	-700000	5	1.53	1
…	…	…	…	…

Table 3 Test set 1(After integration)

Month	Transaction amount(CNY)	Frequency	Single capital flow amount/Average single capital flow amount	Mark of abnormal
201206	-450000	5	0.98	1
201206	-430000	1	0.94	1
201206	-330000	2	0.72	1
201206	-270000	4	0.59	1
201206	-250000	1	0.54	1
…	…	…	…	…

Table 4 Training set 2(After integration)

Month	Transaction amount(CNY)	Frequency	Single capital flow amount/Average single capital flow amount	Mark of abnormal
201212	1521027.61	1	11.53	1
201301	-2000000	1	10.12	1
201301	-700000	1	3.54	1
201301	-500000	2	2.53	1
201301	-400000	2	2.02	1
…	…	…	…	…

Table 5 Test set 2(After integration)

Month	Transaction amount(CNY)	Frequency	Single capital flow amount/Average single capital flow amount	Mark of abnormal
201211	-200000	2	1.01	1
201211	-150000	2	0.76	1
201211	-100000	2	0.51	1
201211	-4320	1	0.02	1
201211	-3500	1	0.02	1
…	…	…	…	…