

# Research on Multi-Classification of Credit Rating of Small and Medium-Sized Logistics Companies Based on Ordinal Regression Support Vector Machine

Ying Chen (Corresponding author)

Sydney Institute of Language & Commerce, Shanghai University  
Chengzhong road 20, Jiading district, Shanghai 201800, P.R. China  
E-mail: dorothychen@staff.shu.edu.cn

Hong Chen

Sydney Institute of Language & Commerce, Shanghai University  
Chengzhong road 20, Jiading district, Shanghai 201800, P.R. China  
E-mail: tracychen@shu.edu.cn

Received: November 2, 2011

Accepted: November 22, 2011

Published: February 1, 2012

doi:10.5539/ijbm.v7n3p127

URL: <http://dx.doi.org/10.5539/ijbm.v7n3p127>

## Abstract

Small and medium-sized logistics companies are playing an increasingly important part in social life. We selected the data of small and medium-sized logistical companies in Beijing, Shanghai and Guangzhou, reformulated ordinal regression support vector machine method so that different input points could make different contributions to decide hyper plane, to analyze multi-classification of credit rating problem, and divided them into four different categories to demonstrate the good performance. The effectiveness of this improved method is verified in multi-classification of credit rating of small and medium-sized logistical companies and the results show that our method is promising and can be used to other multi-classification problems.

**Keywords:** Support vector machine for ordinal regression method, Small and medium-sized logistical companies, Credit rating

## 1. Introduction

Small and medium-sized enterprises play an important role in social life. The statistical data demonstrated that small and medium-sized enterprises account for more than 99% of total number of enterprises, the contribution to GDP is more than 60% and the tax contribution is more than 50%, providing nearly 70% of the import and export trade amounts, creating about 80% of urban jobs in China in 2009. Similar to logistical industry, large-scale logistical enterprises have well-equipped facilities, sufficient capital, perfect internal management control system, strong external financing capacity, a fixed customer base and stable source of income. Simultaneously small and medium-sized logistical companies also occupy an important place in national economy. Millions of small and medium-sized logistical companies provide logistical services to people and meet people's daily delivery service need. However, they also have their own weaknesses. To start with, resources in small and medium-sized logistical companies are limited; they are restricted by customers, and can only provide the relatively lower level of logistical services. Second, due to little extra profits and slow business growth, they are more easily replaced by other companies. Third, in order to win orders, small and medium-sized logistical companies tend to accept customers' harsh requirements. As a result, they may encounter due balance burden, vulnerable to bad credit or financial problems caused by customers. In the end, they come into a debt crisis and bankrupt for money chain tension.

Currently, the qualitatively appraised method used in the practice is mainly the artificial expert analytic method, called the classical credit analysis method using 5 fields to analyze the debtor's credit condition (called 5C method): Character, Capacity, Capital, Collateral and Condition. The quantitative assessment method is

statistical methods used vary widely, such as Multi-linear Discriminate Analysis, Logistic analysis model, Fisher model analysis and artificial analysis method.

## 2. Literature review

### 2.1 Research on small and medium-sized enterprises' credit rating

Xu Zhi-chun (2008) stochastically selected 12 financial indexes and 11 non-financial indexes (external environment changes, guarantee outside, investment increased, appraised value of mortgages decreased, misbehaviors, negative news such as tax or environmental pollution, investigated by government authorities, net cash flow decreased, not coordinating to bank, manager changed) of small and medium-sized enterprises' as samples in the CMS database from some provinces, establishing logistic regression model to forecast the borrowers' default risk. In doing so, early warning ability has been remarkably enhanced and early-warning model classification leads to better results than random classification. Dai Fen (2009) proposed the small and medium-sized enterprise credit rating index system which combined the non-financial indexes with the financial indexes, using ant group neural network to evaluate the samples to five levels, the results of which indicated that the ant group neural network is better than the traditional BP neural network.

The enterprise's credit rating from the view of credit sales in enterprises: Steven Finlay (2010) selected the selling data from a British large-scale catalog retailer and described the customers' default behavior with the binary continually financial behavior forecasting model. Finlay proved that this model is a feasible taxonomic approach and that people can even may also further calculate the profit optimization boundary if using genetic algorithm.

### 2.2 Support vector machines

Support vector machines (SVM) was first proposed by Corinna Cortes and Vapnik in 1995 that has been used very well to solve problems with a few such samples as nonlinear and high-dimensional pattern recognition, and can applies to other machine learning question, just as the function fitting. Support vector machine can solve regression problem (time series analysis) and the pattern recognition problem (classified question analysis), and also is also used in comprehensive evaluation and forecasting problems. Support vector machine is established on statistical learning theory, making up of VC dimension and structural risk minimization, which, based on the limited sample information, seek the balance between complexity of the model (i.e. specific learning accuracy of training samples) and learning ability (i.e. the ability of error-free identification to any sample) in order to obtain the best generalization ability. The classification process is a machine learning process. In the space of  $N$  hyper plane, VC is  $N + 1$ . These data in VC dimension is  $n$ -dimensional space points, these points separated by a  $n-1$  dimensional hyper plane, which is usually referred to as a linear classifier. Training samples of VC dimension in limited circumstances, when the sample size  $n$  is fixed, the higher VC dimension of learning machine, the more complicated of learning machine, VC dimension of learning reflects the learning ability of function set. The greater VC dimension is, the more complicated machine learning is (larger capacity is larger). Structural risk minimization is to ensure the classification accuracy (empirical risk), while reducing VC dimension of learning machine, and control expected risk of learning machine on the entire sample set, and then find the best classification plane, which can make the two different types data have the largest interval, also known as the largest interval hyper plane. SVM maps the vectors to a higher dimensional space, in this space to establish a maximum interval of hyper plane, then to build two parallel hyper plane on both sides of hyper plane which separates data each other, and maximize the distance between the two parallel hyper plane which have the right direction. The larger that the distance between these parallel hyper planes is, the smaller the classifier has total errors, support vector refers to the training sample points at edge of the interval. Support vector machine and neural network have the similar learning mechanisms, but SVM differ from neural network to use the mathematical methods and optimization techniques.

Kernel function of SVM is the key. Set of low-dimensional vector space is usually difficult to divide, if mapping them to high-dimensional space, which will increase the computational complexity, citing the concept of kernel function can solve such problems. Choosing the appropriate kernel function, high-dimensional space will be classified by function, using different kernel functions will lead to different SVM algorithms. Zhou Qifeng, et al (2005), selected more than 1,000 industrial enterprises as samples from a commercial bank in 2003, made the rates of AAA, AA, A and A- as the output. By experiments, several multi-classification strategies assessing performances were compared and directed acyclic graph SVM and embedded space SVM were more suitable for credit risk assessment, which had a faster learning speed and higher forecast accuracy for commercial banks as credit risk assessment modeling. Wun-Hwa Chen et al (2006) used SVM to classify commercial banks which were qualified distributors in Taiwan's capital market, selected the index of stock market information,

government financial support, major stockholder financial support and other variables analysis, and got the results which had the ratio higher accuracy than that of neural network. Xiao Wenbing et al (2007) used support vector machines to establish individual credit evaluation model, and compared the results with that of three fully connected BPN methods. The results showed that SVM in identifying potential loan applicants is better than that of neural network. Cheng-Lung Huang (2007) used mixed method combining genetic algorithms with support vector machine to analysis credit data in Australia and Germany, and obtains a more ideal result. Shu-Ting Luo, et al (2008) used support vector machines and clustering algorithm to rate German credit data, and received more satisfied results.

Atish P. Sinha, et al (2008) compared the results of Naive Bayes method, logistic regression, decision trees, decision tables, neural networks, K nearest neighbor method, support vector machine method in the borrower's credit evaluation and considered that in the certain fields and data limited it is important that to stress synergy between data mining and professional knowledge. Wu Chong, et al (2008) used integrated fuzzy integral-based support vector machine method to assess customers' credit in e-commerce environment and found good robustness and generalization ability.

Wan Fuyong (2009) using Gaussian kernel SVM established the financial risk assessment model for listed companies, selected from 13 key indicators of listed companies, established 42 kinds of financial risk assessment models and compared the prediction accuracy of results. Wan proved support vector machines and statistical methods are better, and that based on numerical simulations Gaussian kernel of support vector machines in financial risk assessment of listed companies is superior. Wen Jing Chen (2010) used similar Lp standard support vector machines to analyze 6,000 records of major U.S. credit card banks, solving smoothing problems by conjugate gradient method, and the results showed that this method was effective.

### 3. Research methodology

In the supply chain system, analysis of small and medium sized logistics companies can make relatively objective comprehensive evaluation and favorable credit ratings. Core enterprises in supply chain establish long-term cooperation relationship with small and medium sized logistics companies according to their own capital and customers. These can enhance the overall competitive strength and credit levels of small and medium sized logistics companies, such as petrochemical industry and retail websites. Meanwhile systemic risk of industry also needs consideration.

In terms of small and medium sized logistics companies, to comprehensively assess their financial situation firstly it is necessary to consider basic financial information, total assets, liabilities, net profit, and the ratio of asset management efficiency, profitability, operational efficiency and solvency (the composition of assets and liquidity of the assets). Because small and medium sized logistics enterprises have limited fixed assets and a large proportion of receivables in current assets, short-term cash flow deficit and inability to pay off short-term debt and payables will happen affected by payment time, resulting in their own liquidity occupied by other enterprises to bankruptcy. Secondly, what needs to consider also include operational ability and services of small and medium sized logistics companies, transportation services, warehousing services or value-added services. If the service provided by enterprises is single and value-added services are weak, revenues and income will be limited. Otherwise, all companies of the industry will suffer for the homogenous competition. If the small and medium sized logistics companies cannot provide much better service to maintain the existing resources, comprehensive evaluation and credit status will be affected in the future. Thirdly, it is also important to analyze internal control system, the status of asset management, quality management of service, calculation of expense and special case settlement in small and medium sized logistics companies.

We selected the small logistics companies of chemical industry in Beijing, Shanghai and Guangzhou to distribute survey forms aiming at access to detailed information. We collected companies' names and addresses, legal representatives, information of people in charge (such as age and education), and stockholder's capital, business type (including restructuring state-owned or foreign owned, joint venture, private, etc.), date of establishment, location of company, bank information, business licenses, staff, total assets, net plant assets, total liabilities, debt ratio, total revenue, gross profit, net profit; logistics services: transportation services (including private railways of rail transportation, consignment shipment station and other ways, road transport and sea transport of general cargo transportation, container transportation and dangerous liquid goods transportation), storage services (including solid storage of general cargo, dangerous solid goods storage, ordinary liquid production storage, dangerous liquid goods storage and liquefied gas storage), other services (including port handling, cargo agency, railway projects agency), value-added services (including logistical project design and logistics resource integrated services); transportation services facilities (including road transport vehicles of company's own -

vehicles of general cargo transport, vehicles of dangerous goods transport, and transportation vehicles in cooperated companies - general cargo transport vehicles and dangerous goods transport vehicles); storage facilities (self-processed solid storage facilities, self-processed liquid storage facilities, warehouse and facilities with loading tank supports, self-processed water transport vessels, berthing facilities, railway transport - loading truck lines, rail tank car).

71 valid samples were obtained. Because some small and medium-sized logistical companies did not have storage capacity and transport conditions, default data is 0. Since sample data obtained mostly were qualitative variables, quantitative variables were less. And the selected indicators will duplicate information, such as total assets, total liabilities and debt ratio. Therefore, principal component analysis is used to dimension reduction analysis, which obtained factors through computing process: company's overall conditions factor, asset condition factor, profitability factor, capacity of general goods transportation factor, capacity of dangerous goods transportation factor, capacity of general goods storage factor, capacity of dangerous goods storage factors, other services factors, value-added services, factors, facilities of general goods transport factor, facilities of dangerous goods transport services factor, facilities of solid storage factor, facilities of liquid storage factor, facilities of water storage factor, ability rail transportation and warehousing factor.

Because this problem belongs to small sample classification question, it selects support vector machine of multi-classification algorithm on small sample learning. Assigns training set:

$$T = \{(x_1, x_2), \dots, (x_{l^j}, x_2)\} \in (\mathbb{R}^n \times Y)^l, \quad x_i \in \mathbb{R}^n, y_i \in Y = \{1, -1\}, i = 1, \dots, l \tag{1}$$

Support vector machine solving multi-classification problem is to construct a series of problems into two classifications, and establish corresponding two classified machines, to determine which category input  $x$  belongs to based on two classified machines. One versus one method, one versus the rest method, and Crammer-Singer multi-classification support vector machines. It is selected ordinal regression support vector machine, consider that  $M$  category of input is sequential from 1 to  $M$  in space  $\mathbb{R}^n$ , which has identified adjacent relationship, that is, class  $j$  is the adjacent class of  $j-1, j+1$ , and class  $j-1$  and class  $j+1$  are not adjacent class, and space can be separated by  $M-1$  parallel hyper plane.

$$T = \{x_i^j\}_{i=1,2,\dots,l^j}^{j=1,2,\dots,M} \tag{2}$$

Here,  $x_i^j$  is training input, superscript  $j$  expresses the category of corresponding training,  $l^j$  is the number of  $j$ th

kind training data.

The primitive question is:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \tag{3}$$

$$s.t. \quad (w \cdot x_i^j) - b_j \leq -1, j = 1, \dots, M, i = 1, \dots, l^j \tag{4}$$

$$(w \cdot x_i^j) - b_{j-1} \geq 1, j = 1, \dots, M, i = 1, \dots, l^j \tag{5}$$

Here  $b = (b_1, \dots, b_{M-1})^T, b_0 = -\infty, b_M = +\infty$

According to the training regulations and training set, introduce to slack variable  $\xi^{(x)} = (\xi_1^1, \dots, \xi_{l^1}^1, \dots, \xi_1^M, \dots, \xi_{l^M}^M, \xi_1^{*1}, \dots, \xi_{l^1}^{*M}, \dots, \xi_{l^M}^{*M})^T$  and penalty parameter  $C > 0$ , primitive question

transforms to convex quadratic programming problem :

$$\min_{w,b,\xi^{(x)}} \frac{1}{2} \|w\|^2 + C \sum_{j=1}^M \sum_{i=1}^{l^j} (\xi_i^j + \xi_i^{*j}), \tag{6}$$

s.t.

$$(w \cdot x_i^j) - b_j \leq -1 + \xi_i^j, j = 1, \dots, M, i = 1, \dots, l^j$$

$$(w \cdot x_i^j) - b_j \geq -1 + \xi_i^{*j}, j = 1, \dots, M, i = 1, \dots, l^j$$

$$\xi_i^j \geq 0, \xi_i^{*j} \geq 0, j = 1, \dots, M, i = 1, \dots, l^j$$

And

$$b = (b_1, \dots, b_{M-1})^T, b_0 = -\infty, b_M = +\infty$$

#### 4. Conclusion

Using support vector machine classified credit rating of small and medium sized logistics companies into AAA, AA, A, B four category. According to the experimental results, accuracy rate of training set were: 95.8%, 93.9%, 92.1%, 91.9%, accuracy rate of testing set were: 93.1%, 92.7%, 91.1%, 87.9%. The results can be explained that level AAA is the enterprises with capacities of dangerous goods storage and transportation, level AA is the enterprises with capacities of dangerous goods transportation, level A is the enterprises with capacities of ordinary goods transportation and warehousing, level B is the enterprises with ordinary goods transportation, the errors can be explained that small and medium-sized logistical companies usually only have capacities of ordinary goods transportation and warehousing, however, these enterprises owned many transportation vehicles or vehicles in affiliated companies, these raised their level of credit rating. For these exceptional phenomena, it is manually adjusted by artificial expert method in practice, and achieves the better results of judgments. Thus, it is certified that we reformulated ordinal regression support vector machine method so that different input points could make different contributions to decide hyper plane, to analyze multi-classification problems, and divided them into different categories to demonstrate the good performance. The effectiveness of this improved method is verified in multi-classification of credit rating of small and medium-sized logistical companies and the experiment results show that our method is promising and can be used to comprehensive evaluation, sequencing problems, and other multi-classification problems. Our future direction of the research will focus on how to improve the accuracy of multi-classification. We believe that more suitable parameters and variables selection will affect and improve the performance of generalization. Extending the multi-class classification to solve other problems is also our future research work.

#### References

- Atish P. Sinha, & Huimin Zhao. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems*, 46, 287–299. <http://dx.doi.org/10.1016/j.dss.2008.06.013>
- Cheng-Lung Huang, Mu-Chen Chen & Chieh-Jen Wang. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 33, 847 – 856. <http://dx.doi.org/10.1016/j.eswa.2006.07.007>
- China Small and medium-sized enterprise quantity ultra thousands of household. [Online] Available: <http://www.chinanews.com.cn/cj/cj-gncj/the news/2010/05-14/2283411.shtml>, (May 14, 2010)
- Dai Fen, et al. (2009). Application of ant colony neural network to credit evaluation of small and middle enterprises. *Computer Technology and Development*, 10, 218-221
- Deng Naiyang, et al. (2009). *Support vector machines: theory, algorithm and development*. Chinese scientific publishing house
- Nello Cristianini & John Shawe-Taylor. (2004). *Introduction to support vector machines(Chinese edition)*. Chinese electronics industry publishing house.
- Shi Xiufu, et al. (2009). SVM prediction method for assessing the financial credit rating of the listed companies in stock market. Doctoral thesis in East China Normal University
- Shu-Ting Luo, Bor-Wen Cheng, & Chun-Hung Hsieh. (2009). Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications*, 36, 7562 – 7566
- Steven Finlay. (2010). Credit scoring for profitability objectives. *European Journal of Operational Research*, 202,528–537. <http://dx.doi.org/10.1016/j.ejor.2009.05.025>

- Wen Jing Chen & Ying Jie Tian. (2010). Lp-norm proximal support vector machine and its applications. *Procedia Computer Science*, 1, 2411–2417. <http://dx.doi.org/10.1016/j.procs.2010.04.272>
- Wu Chong, et al. (2008). Customer credit evaluation model based on support vector machine in e-business environment. *China Management Science*, 10,368-373.
- Wun-Hwa Chen & Jen-Ying Shih. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, 30,427–435. <http://dx.doi.org/10.1016/j.eswa.2005.10.003>
- Xiao Wen-bing, et al. (2007). Based on support vector machine model for credit assessment and risk. *Chinese science abstract*, 22, 284.
- Xu, Zhi-chun & Wang, Zong-jun, et al. (2008). Introducing the non-financial factor empirical study on early warning model of credit risks of the small and medium-sized enterprises. *Financial Theory and Practice*, 6,3-6.
- Zhou Yi-feng, et al. (2005). A comparison of multi-classification methods for credit risk assessment in commercial banks. Proceedings of the 24th Chinese control conference, 1734-1737.