Consumer Credit Evaluation Model in Tmall Using SVM Method

Shuang Chen[1, a, *], Dan Li[1, b], HongYun Gao[1, c] and FanYun Meng[2, d]

[1] Information and Engineering College of Dalian University, Dalian, 116622, China

[2] School of Computer, Qingdao Technological University, Qingdao, 266033, China

[a] chenshuang0707@163.com, [b] 279675738@qq.com, [c] gaohongyun@dlu.edu.cn,

[d] mengfanyundlut@163.com

Keywords: E-commerce, consumer credit evaluation, B2C.

Abstract. Reducing mistrust between businesses and consumers is a central motivating force of e-commerce. This research identifies a number of key factors related to consumer credit in the B2C context and investigates a method which named support vector machines(SVM)to hedge consumer credit evaluation. Our experimental results of consumer credit evaluation based on data sets from Tmall show that SVM can enhance the separation of different consumers.

1. Introduction

E-commerce has been a phenomenal growth since the development of the internet and there is a growing interest from many businesses to use it as a way to reach a wider customer base. A major factor influencing the successful proliferation of E-commerce, identified by major corporations, the government and the business bureau, is people's trust in Internet vendors. Credit evaluation should be even more important in e-commerce than in traditional commerce because of the paucity of rules and customs in regulating e-commerce and because online services and products typically are not immediately verifiable. Therefore, credit evaluation become a very challenging and important problem in the domain of e-commerce.

In business-to-consumer (B2C) e-commerce, In recent years a lot of work focus on consumer trust [1] or initial trust [2] [3] which has a great influence on transaction. However, there is very little research work on consumer credit evaluation which is unreasonable. In recent years, we survey merchants in some e-commerce websites and operators often complain that some consumers always give negative feedback or return goods unreasonably. What is more, they threaten seller with bad reviews for money or competition. These kinds of reasons would have a adverse effect on e-commerce business, especially for new sellers. Therefore, build mutual trust between merchants and consumers is very necessary. We need to establish a consumer credit evaluation model to distinguish “good” or “bad” consumers. This should remind consumer being a civilized buyer and not threatening sellers with bad reviews.

A way of solving credit scoring problem is the support vector machines(SVM) method. This method has advantages of using flexible objectives and constraints to fit a decision function for separation of different classes. SVM is introduced by V. Vapnik et al. in literatures [4] and [5-7]and is an approach for classification that was developed in the computer science community in the 1990s and that has grown in popularity since then. In this paper, we used SVM method for consumer credit evaluation in B2C e-commerce. This application also illustrates new aspects of the applicability of SVM in credit evaluation and data mining.

The rest of this paper is organized as follows: Section 2 we establish consumer credit indexes system. Then the SVM method are illustrated in Section 3. The experiment on consumer credit evaluation and the results are demonstrated in Section 4. Finally, conclusions will be given in Sections 5.

2. Determine indexes of consumer credit evaluation

The Delphi technique, mainly developed by Dalkey and Helmer (1963) at the Rand Corporation in the 1950s, is a widely used and accepted method for achieving convergence of opinion concerning real-world knowledge solicited from experts within certain topic areas.

Delphi method is a structured communication technique, originally developed as a systematic, interactive forecasting method which relies on a panel of experts. The experts answer questionnaires in two or more rounds. After each round, a facilitator provides an anonymous summary of the experts' forecasts from the previous round as well as the reasons they provided for their judgments. Thus, experts are encouraged to revise their earlier answers in light of the replies of other members of their panel. It is believed that during this process the range of the answers will decrease and the group will converge towards the ``correct" answer. Finally, the process is stopped after a pre-defined stop criterion and the mean or median scores of the final rounds determine the results.

In this paper, we determine evaluation indexes using Delphi method. We visit 48 experts from e-commerce and related field. We use SPSS 20.0 for reliability analysis and factor analysis. Combining expert decisions with Cronbach α coefficient and factor loadings, we determine final evaluation indexes (see Table 1).

Table 1 Index system of consumer credit

First grade	Second grade
Customer equity	The number of linked bank card (A1)
Transaction records	The accumulative amount of transactions (A2) The cumulative number of transactions(A3) The number of return(A4) The number of bad review which the consumer gives (A5)
Personal information	Customer age (A6) Customer gender (A7)

3. Classification methods

In recent years, SVM has been successfully applied to a wide range of applications and now we use it to classify consumer credit. In order to find an optimal separating hyperplane the margin between the two supporting hyperplanes should be maximized, in the meantime, the total errors of misclassifications should be minimized. The supporting hyperplanes corresponding with two classes are denoted as the following formulation. The nonnegative slack variables ξ_i are introduced so as to deal with the case of constraint violations.

$$y_i(\mathrm{w}^T \mathrm{x}_i + \mathrm{b})\xi_i \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{1}$$

Therefore, the optimization problem of the SVM classifier is expressed as:

$$\begin{aligned} & \min \ \mathrm{f}(\mathrm{w}, \mathrm{b}) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i \\ & s.t. \quad y_i(\mathrm{w}^T \mathrm{x}_i - \mathrm{b}) \geq 1 - \xi_i \\ & \qquad\ \ \xi_i \geq 0, \forall i. \end{aligned} \tag{2}$$

Let $\overline{\alpha} = (\alpha_1, ..., \alpha_n)$ be the Lagrange multipliers and the Lagrange dual problem of (2) can be expressed as

$$
\begin{aligned}
&\min\ g(\overline{\alpha}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=i}^{n}\alpha_i\alpha_j y_i y_j (\mathrm{x}_i \cdot x_j) - \sum_{i=1}^{n}\alpha_i \\
&s.t.\ \ \sum_{i=1}^{n}\alpha_i y_i = 0,\ 0 \le \alpha_i \le C
\end{aligned}
\tag{3}
$$

where C is a upper bound of the α_i. For the optimal solution α^* of the problem (3) and its Karush-Kuhn-Tucker (KKT) conditions, the input points x_i associated with $0 < \alpha^* < C$ are called

as support vectors which lie on the supporting hyperplanes. In the case of α^* the corresponding points are correctly classified while the input points with $\alpha^* = C$ are misclassified. Thus, during testing we may use the decision function to predict the class label of an input point $\hat{x}$ as below

4. Numerical experiments

Our algorithm code is writing in MATLAB 2009b. The experiment environment: Intel Core I5 CPU, 2 GB memory. The MATLAB optimal tools and [] are employed to solve optimization problems related to this paper. In our experiments, two credit data sets are used for credit risk analysis: Consumer credit data set of T mall and consumer credit data set of Tao Bao. They are presented separately as follows.

Tmall.com currently features more than 70,000 international and Chinese brands from more than 50,000 merchants and serves more than 180 million buyers. Tmall.com ranked number one among all Chinese B2C retail websites for 2010 in terms of transaction volume, with a gross merchandise volume of RMB30 billion-about three times the amount facilitated by 360buy, its closest competitor. The data set which we get from T mall consists of 201 instances of Good credit consumers and 238 instances where consumers are Bad credit. Each instance contains 7 numeric attributes and 1 class attribute (male or female). There are also a few missing values in this data set.

To protect customer privacy, the attributes names and values have been changed to meaningless symbolic data. For the consumer credit data set of T mall, by using the 10-fold CV method the SVM model is trained on training subsets and validated on validation subsets in order, the classifiers are obtained and tested on the independent test set. The results are listed in Table 2. In Table 2, we list test results of every single index A_i and all indexes, where Iter, obj, rho, n SV, n BSV, Total n SV and Accuracy denote iterations, the minimum of quadratic programming, constant b in decision function (4), the number of support vectors, the number of support vectors at boundary, the total number of support vectors and classification accuracy.

Table 2 Index system of consumer credit

	Iter	obj	rho	nSV	nBSV	Total nSV	Accuracy
All indexes	64	-25.9491	-0.3123	43	37	43	99.54%
A1	114	-155.9231	-0.2919	170	168	170	89.97%
A2	73	-86.4915	-0.3134	98	96	98	92.48%
A3	106	-142.9890	-0.3624	160	158	160	89.52%
A4	67	-107.1365	-0.0552	124	124	124	91.79%
A5	53	-81.7900	-0.3107	144	102	144	93.84 %
A6	179	-324.7713	-0.0708	350	348	350	64.46%
A7	98	-193.5819	-0.0000	196	194	196	78.13 %

In the second line of Table 2, the classification accuracy of consumer credit reached a very high level 99.54 when we consider all 7 indicators. However, owing to the inevitable missing information and the huge amount of computation in practice, we do not want to use all of the indicators. Therefore, we hope to work out the importance of each index to credit classification, then we can get an acceptable classification accuracy (such as 90% or 95%) only using one or two indicators. According to Table 2, we analyze each index then rank them by what's most important to the classification of customer credit. The order is A5>A2>A4>A1>A3>A7>A6.

The main contribution of this paper is to establish a consumer credit evaluation model and use SVM method to solve it. At the same time, SVM was tested with real world data sets from Tmall. The experimental results show that SVM is a more effective classifier for consumer credit evaluation in B2C e-commerce, and has great potential as a prospective classification approach for other applications in e-commerce.

Acknowledgments

This work was financially supported by the Natural Science Foundation of China, Grant 11626051, 11626052, 11501074

References

[1]. Winch G, Joyce P. Exploring the dynamics of building, and losing, consumer trust in B2C eBusiness [J]. International Journal of Retail & Distribution Management, 2006, 34 (7): 541-555.

[2]. Yaobin L, Tao Z. An empirical analysis of factors influencing consumers' initial trust under B2C environment [J]. Nankai Business Review, 2005, 8 (6): 96-101.

[3]. Yaobin L, Tao Z. An empirical analysis of factors influencing consumers' initial trust under B2C environment [J]. Nankai Business Review, 2005, 8 (6): 96-101.

[4]. Niklis D, Doumpos M, Zopounidis C. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines [J]. Applied Mathematics and Computation, 2014, 234: 69-81.

[5]. Brengman M, Karimov F P. The effect of web communities on consumers' initial trust in B2C e-commerce websites [J]. Management Research Review, 2012, 35 (9): 791-817.

[6]. Gefen D, Straub D W. Consumer trust in B2C e-commerce and the importance of social presence: experiments in e-products and e-services [J]. Omega, 2004, 32 (6): 407-424.

[7]. Hong, I B, Cho H. The impact of consumer trust on attitudinal loyalty and purchase intentions in B2C e-marketplaces: Intermediary trust vs. seller trust [J]. International Journal of Information Management, 2011, 31 (5): 469-479.