

# Consumer credit evaluation model in C2C e-commerce using MCOC methods

Shuang Chen<sup>1, a, \*</sup>, Hongyun Gao<sup>1, b</sup>, Dan Li<sup>1, c</sup> and Fanyun Meng<sup>2, d</sup>

<sup>1</sup> Information and Engineering College of Dalian University, Dalian, 116622, China

<sup>2</sup> School of Computer, Qingdao Technological University, Qingdao, 266033, China

<sup>a</sup> chenshuang0707@163.com, <sup>b</sup>

gaohongyun@dlu.edu.cn, <sup>c</sup> 279675738@qq.com,

<sup>d</sup> mengfanyundlut@163.com

**Keywords:** E-commerce, multi-criteria optimization classifier, consumer credit evaluation, B2C, C2C.

**Abstract:** In this paper, we investigate a method named multi-criteria optimization classifier (MCOC) to hedge consumer credit evaluation in consumer-to-consumer (C2C) e-commerce. Consumer credit is one of the key obstacles to vendors succeeding on the internet medium and a lack of consumer credit is likely to discourage online businesses from participating in e-commerce. Our experimental results of consumer credit evaluation based on data sets from TaoBao show that MCOC can enhance the separation of different consumers, the efficiency of credit evaluation, and the generalization of predicting the credit rank of a new consumer.

## 1. Introduction

Consumer-to-consumer (C2C) platforms such as eBay, Amazon's Marketplace, and Taobao have become a major engine of growth in Internet commerce. With rapid development of C2C e-commerce, fake evaluation and harmless consumers have rushed into market and greatly affected the normal market order, fair competition and business confidence. Consumer credit evaluation of C2C e-commerce has more and more aroused people's concern. It is a tool of disseminating reputation, the online credit evaluation system facilitates building trust, encouraging trustworthy behavior, and deterring participation by those who are unskilled or dishonest. Thus consumer credit evaluation contribute to reducing the transaction risks, increasing the rate of successful transactions.

The research of MCOC approaches to classification problems is first introduced by Freed and Gl over in literature [1] and has found its path to the real business world at 1998 in literature [2]. The MCOC has been used in various applications within different fields of science, such as credit scoring in bank's consumers or enterprises credit [3] medical treatment [4], biology [5] and analyze the behavior of VIP E-mail users [6].

The rest of this paper is organized as follows: Section 2 we establish consumer credit indexes system. Then the MCOC 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. Classification methods

In this paper we visit 50 experts from e-commerce and related field and use Delphi method for determining evaluation indexes. We use SPSS 22 for reliability analysis and factor analysis. Combining expert decisions with Cronbach  $\alpha$  coefficient and factor loadings, we determine final evaluation indexes (see Table 1).

In this section, we use MCOC to classify consumers' credit. First we want to determine the coefficients  $w = (w_1, w_2, \dots, w_n)$  for an appropriate subset of the variables  $x = (x_1, x_2, \dots, x_n)$  and a

boundary value  $b$  to separate two consumers' groups: ``Good" credit and ``Bad" credit. Therefore, we give a independently and identically distributed training set

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

each input point  $x_i \in R^n$  belongs to either of the two classes with a label  $y_i \in \{1, -1\}$ .  $x_i$  denotes the consumer who need to be classified, and  $y_i = 1$  denotes that  $x_i$  belong to ``Good" credit consumers and  $y_i = -1$  denotes  $x_i$  belong to ``Bad" credit consumers. The separating hyperplane is defined as:

$$w^T x - b = 0.$$

Next we will establish MCOC models respectively to classify these  $x_i$ . In order to separate the consumers of good credit and bad credit, we choose the two measures for any input point like literature [7]. The first is the overlapping degree of deviation from the separating hyperplane. The second is the distance between input points and the separating hyperplane. In the first case an input point located the wrong side of the hyperplane is misclassified, while in the second case an input point positioned in the right side of the hyperplane is correctly.

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 give (A5)
Personal information	Customer age (A6)
	Customer gender (A7)

Let  $\alpha_i \geq 0$  be the distance where an input point  $x_i$  deviates from the separating hyperplane, and the sum of the distance  $\alpha_i$  is characterized by the function  $f(\alpha) = \|\alpha\|_p^p$ ,  $p \geq 1$  which should be minimized with respect  $\alpha_i$ . Let  $h_i \geq 0$  be the penalties for  $\alpha_i$  which denotes the cost of misclassification. Then we have

$$\begin{aligned}
 \min f(\alpha) &= \|h\alpha\|_p^p \\
 \text{s.t. } w^T x_i - b &\geq -\alpha_i, y_i = 1, \\
 w^T x_i - b &\leq \alpha_i, y_i = -1, \\
 \alpha_i &\geq 0, \forall i.
 \end{aligned} \tag{1}$$

Similarly, let  $\beta_i$  be the distance where the input point  $x_i$  departs from decision hyperplane, then the sum of  $\beta_i$  is characterized by the function  $f(\beta) = \|\beta\|_q^q$ ,  $q \geq 1$  which should be maximized with respect to  $\beta_i$ , and we get

$$\begin{aligned}
 \max f(\beta) &= \|\beta\|_q^q \\
 \text{s.t. } w^T x_i - b &\geq -\beta_i, y_i = 1, \\
 w^T x_i - b &\leq \beta_i, y_i = -1, \\
 \beta_i &\geq 0, \forall i.
 \end{aligned} \tag{2}$$

A hybrid model that combines the above models can be:

$$\begin{aligned} \max h(\alpha, \beta) &= \|h\alpha\|_p^p - \|\beta\|_q^q \\ \text{s.t. } w^T x_i - b &= y_i(\beta_i - \alpha_i), \\ \alpha_i, \beta_i &\geq 0, \forall i. \end{aligned} \quad (3)$$

In (3) if  $p = 1, q = 1$ , a linear MCOC with the linear objective function which can be rewritten

$$\begin{aligned} \max h(\alpha, \beta) &= h \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \beta_i \\ \text{s.t. } w^T x_i - b &= y_i(\beta_i - \alpha_i) \\ \alpha_i, \beta_i &\geq 0, \forall i. \end{aligned}$$

Similarly, If  $p = 2, q = 2$ , we get a quadratic MCOC with the quadratic objective function and the linear constraints.

During testing we may use the decision function  $l(x)$  to predict the class label of an input point  $x$  as  $l(x) = w^T x - b$  (5). If  $l(x) \geq 0$ , then  $x$  belong to "Good credit" consumer set, else  $x$  belong to "Bad credit" consumer set.

However, we can't guarantee that the data set of consumer credit information is linear separability. Furthermore the cases of nonlinearly separable data is very common for classification and we often use kernel method  $\phi(\bullet)$  to deal with these problems. The non-linear function  $\phi(\bullet)$  maps the input space where the data set is not linearly separable to a high (possibly infinite) dimensional feature space where the data set may be linearly separable. We replace  $\phi^T(x) \cdot \phi(y)$  by the kernel function  $K(x, y) = \phi^T(x) \cdot \phi(y)$ . For the kernel function  $K(\bullet, \bullet)$  one typically has the following choices:

- 1)  $K(x, y) = x^T y$ , (linear kernel);
- 2)  $K(x, y) = (1 + x^T y / c)^d$  (polynomial kernel of degree  $d$ );
- 3)  $K(x, y) = \exp\{-\|x - y\|_2^2 / \sigma\}$  (RBF kernel).

The above types of the kernel functions will be used in our paper. In consumer credit classification, expect for the data set may not be linear separability because of noise, we need to face another difficult. Data sets of consumer credit information often contain anomalies, class imbalance, nonlinear separability or other uncertainties, and these lead to the predictive performance may be rapidly degenerate. Similarly, to [8] which deal with credit risk evaluation for bank, we also use fuzzification method and class-imbalanced learning method to deal with these problems.

Following the above illustrations, the classification process of consumer credit can be summarized into three steps as follows:

Step1. Solving (3) and getting the optimal solution  $\alpha_i, i = 1, \dots, n$  based on the training set;

Step2. Constructing a decision function (5) according to the separating hyperplane  $w^T x - b$  respect  $\alpha_i$  with the training set;

Step3. Testing an unknown consumer  $\bar{x}$  by decision function: If  $\delta(\bar{x}) > 0$ , the consumer  $\bar{x}$  is classified as the "Good" credit; otherwise, the consumer  $\bar{x}$  is predicted as the "Bad" credit.

### 3. Numerical experiments

Our algorithm code is wrote in MATLAB 2009b. The experiment environment: Intel Core I5 CPU, 2 GB memory. TaoBao is founded by Alibaba Group on May 10, 2003, Taobao Marketplace facilitates consumer-to-consumer (C2C) retail by providing a platform for small businesses and individual entrepreneurs to open online stores that mainly cater to consumers in China, Hong Kong, Macau and Taiwan. We also get a data set of consumer credit evaluation from TaoBao. This data set also contains 7 numeric attributes and 1 class attribute, and consists of 273 instances of "Good" credit

consumers and 249 instances where consumers are ``Bad" credit.

For each credit data set mentioned above, we randomly select samples from each class of the respective data set, and form the training set. The remainder is used for the independent testset. Then the 10-fold cross-validation (CV) method is used to train MCOC on the training subset, and the averages of predictive performance with the independent test set are calculated and reported.

In the second line of Table 2, the classification accuracy of consumer credit also reached a very high level 97.85 when we consider all 7 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

$$A_5 > A_4 > A_2 > A_3 > A_1 > A_7 > A_6.$$

Table 2 Index system of consumer credit

	Iter	obj	rho	nSV	nBSV	Total nSV	Accuracy
All indexes	85	-68.7556	0.1640	101	93	101	97.89%
A1	226	-331.2593	0.9697	341	338	341	71.83%
A2	175	-180.9931	1.1222	192	189	192	86.78%
A3	183	-237.0296	1.2296	249	158	249	83.71%
A4	82	-128.1040	-0.0304	150	124	150	91.76%
A5	57	-82.6896	-0.2653	106	102	106	95.40 %
A6	277	-419.2666	0.1546	424	348	350	59.38%
A7	184	-365.58	-0.0000	368	194	196	65.13 %

## Acknowledgments

This work was financially supported by the Natural Science Foundation of China, Grant 11501074,11701061.

## References

- [1] Zhang Z, Shi Y, Gao G. A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis[J]. Expert Systems with Applications, 2009, 36(5): 8932-8937.
- [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] Qi Z, Tian Y, Shi Y. Regularized multiple criteria linear programming via linear programming[J]. Procedia Computer Science, 2012, 9: 1234-1239.
- [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] Zhang Z, Gao G, Shi Y. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors[J]. European Journal of Operational Research, 2014, 237(1): 335-348.
- [7] Zhang P, Zhu X, Zhang Z, et al. Multiple criteria programming models for VIP E-Mail behavior analysis[J]. Web Intelligence and Agent Systems, 2010, 8(1): 69-78.
- [8] Zhang Z, Gao G, Yue J, et al. Multi-criteria optimization classifier using fuzzification, kernel and penalty factors for predicting protein interaction hot spots[J]. Applied Soft Computing, 2014, 18: 115-125.