European Journal of Operational Research 237 (2014) 335–348

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors

Zhiwang Zhang [a,*], Guangxia Gao [b], Yong Shi [c,d]

[a] School of Information and Electrical Engineering, Ludong University, Yantai 264025, China
[b] Shandong Institute of Business and Technology, Yantai 264005, China
[c] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[d] College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA

ARTICLE INFO

Article history:
Received 1 April 2012
Accepted 20 January 2014
Available online 28 January 2014

Keywords:
Data mining
Fuzzy set
Kernel function
Multi-criteria optimization
Classification
Credit risk

ABSTRACT

With the fast development of financial products and services, bank's credit departments collected large amounts of data, which risk analysts use to build appropriate credit scoring models to evaluate an applicant's credit risk accurately. One of these models is the Multi-Criteria Optimization Classifier (MCOC). By finding a trade-off between overlapping of different classes and total distance from input points to the decision boundary, MCOC can derive a decision function from distinct classes of training data and subsequently use this function to predict the class label of an unseen sample. In many real world applications, however, owing to noise, outliers, class imbalance, nonlinearly separable problems and other uncertainties in data, classification quality degenerates rapidly when using MCOC. In this paper, we propose a novel multi-criteria optimization classifier based on kernel, fuzzification, and penalty factors (KFP-MCOC): Firstly a kernel function is used to map input points into a high-dimensional feature space, then an appropriate fuzzy membership function is introduced to MCOC and associated with each data point in the feature space, and the unequal penalty factors are added to the input points of imbalanced classes. Thus, the effects of the aforementioned problems are reduced. Our experimental results of credit risk evaluation and their comparison with MCOC, support vector machines (SVM) and fuzzy SVM show that KFP-MCOC can enhance the separation of different applicants, the efficiency of credit risk scoring, and the generalization of predicting the credit rank of a new credit applicant.

1. Introduction

Credit risk evaluation is a very challenging and important data mining problem in the domain of financial analysis. Credit scoring models have been extensively used to evaluate the credit risk of consumers or enterprises, and they can classify the applicants as either accepted or rejected according to their demographical and behavioral characteristics. Over the past three decades, a large number of methods have been proposed for credit risk decision (Lando, 2004; Thomas, Crook, & Edelman, 2002). These methods mainly include logistic regression (Bolton, 2009; Wiginton, 1980), probit regression (Grablowsky & Talley, 1981), nearest neighbor analysis (Henley & Hand, 1996), Bayesian network (Baesens, Egmont-Petersen, Castelo, & Vanthienen, 2002; Pavlenko & Chernyak, 2010), artificial neural network (Jensen, 1992; West, 2000), decision tree (Bastos, 2008; Zhang and Zhang, et al., 2010; Zhang and Zhou, et al., 2010; Zhang and Zhu, et al., 2010), genetic algorithm (Abdou, 2009; Ong, Huang, & Tzeng, 2005), multiple criteria decision making (Shi, Peng, Xu, & Tang, 2002; Shi, Wise, Luo, & Lin, 2001), SVM (Bellotti & Crook, 2009; Gestel, Baesens, Garcia, & Dijcke, 2003; Huang, Chen, & Wang, 2007; Martens, Baesens, Van Gestel, & Vanthienen, 2007; Schebesch & Stecking, 2005), and so on. Among these classification methods, logistic regression is considered to be the most popular statistical approach and has been more widely used than the others in practice. Neural network credit scoring models have high accuracy but some modeling skills are required – for instance, to design proper network topologies – and it is difficult to explain the results of credit scoring to users clearly. SVM based models indicated promising results in credit risk evaluation, but the SVM classifier needs to solve a convex quadratic programming problem which is very computationally expensive in real world applications.

Recently increasing interests in the synergies of optimization and data mining can be observed (Olafsson, Li, & Wu, 2008; Meisel & Mattfeld, 2010; Corne, Dhaenens, & Jourdan, 2012). As optimization

* Corresponding author. Tel.: +86 015053532980.
E-mail address: zzwmis@163.com (Z. Zhang).

http://dx.doi.org/10.1016/j.ejor.2014.01.044
0377-2217/

techniques, SVM based on statistic learning theory and optimization recently grew in popularity (Cortes & Vapnik, 1995; Vapnik, 1995, 1998), mainly owing to its higher generalization power than that of some traditional methods. The main idea of the SVM algorithm is to separate instances from different classes by fitting a separating hyperplane that maximizes the margin among the classes and minimizes the misclassification simultaneously. For the linearly separable case, the hyperplane is located in the input space. Besides, for the nonlinearly separable case, kernel techniques are used to map the data from the input space into the feature space, and the hyperplane is positioned in the feature space (Cristianini & Shawe-Taylor, 2000; Hamel, 2009).

When SVM is employed to solve classification problems, each input point is treated equally and assigned certainly to one of different classes. However, in many practical applications, SVM is very sensitive to noise, outliers and anomalies in data so that the separating hyperplane severely deviates from the right position and direction. Thus several methods have been proposed to solve the problem by introducing a proper fuzzy membership function to SVM model (Abe, 2004; Abe & Inoue, 2002; Jiang, Yi, & Jian, 2006; Lin & Wang, 2002, 2004; Takuya & Shigeo, 2001; Tovar & Yu, 2008; Tsujinishi & Abe, 2003). Additionally, in real world applications, it is very usual that one class is more important than others, and that class distribution is imbalanced, resulting in a rapidly degenerating classification precision and accuracy. In order to effectively deal with the class imbalance problem, penalty techniques based on cost-sensitive learning are used (Koknar-Tezel & Latecki, 2009, 2010; Tang, Zhang, & Chawla, 2002; Yang, Wang, Yang, & Yu, 2008; Yang, Yang, & Wang, 2009; Zeng & Gao, 2009).

Another optimization method mentioned above is MCOC which is used to solve classification problems in data mining and machine learning (Shi et al., 2001). The classifier mainly uses a trade-off between the overlapping degree of different classes and the total distance from input points to the separating hyperplane, with the former to be minimized and the latter to be maximized simultaneously based on the idea of linear programming for classification (Freed & Glover, 1981; Glover, 1990). Then MCOC has been used in various applications within different fields of science, ranging from credit scoring to bioinformatics. Subsequently, a linear MCOC based on a compromise solution was proposed for the behavior analysis of credit cardholders (Shi et al., 2002). A multiple phase fuzzy linear programming approach was provided for solving classification problem in data mining (He, Liu, Shi, Xu, & Yan, 2004). And a penalized MCOC using weight of the target class was proposed for solving the class-imbalanced classification problem in credit cardholder behavior analysis (Li, Shi, & He, 2008). Then a quadratic MCOC was proposed and used for credit data analysis (Peng, Kou, Shi, & Chen, 2008). A rough set-based MCOC was put forward and used for the medical diagnosis and prognosis (Zhang, Shi, and Gao, 2009; Zhang, Shi, and Tian, 2009), and a MCOC with fuzzy parameters was used to improve the generalization power of MCOC, where an appropriate fuzzy membership function is introduced to MCOC, and the objective functions and the constraints were transformed into the fuzzy decision set, then the new MCOC with fuzzy parameters was constructed (Zhang, Shi, and Gao, 2009; Zhang, Shi, and Tian, 2009). A kernel-based MCOC was given just like the use of kernel methods in SVM (Zhang and Zhang, et al., 2010; Zhang and Zhou, et al., 2010; Zhang and Zhu, et al., 2010). Additionally, MCOC was used to analyze the behavior of VIP E-mail users (Zhang and Zhang, et al., 2010; Zhang and Zhou, et al., 2010; Zhang and Zhu, et al., 2010). The above rough set-based MCOC was also used to predict protein interaction hot spots (Chen et al., 2011). In these applications, MCOC outperformed some traditional methods in data mining (Shi, 2010). A number of models and algorithms related to MCOC gradually developed to powerful tools for solving classification, regression and other problems.

However, in many real world applications such as bioinformatics, language information processing and credit risk evaluation, quality problems like noise, outliers and anomalies within the data set are very common; additionally the set may be class-imbalanced, nonlinearly separable, and uncertain. Because of these uncertainties, some input points are difficult to correctly classify as one of predefined classes. Consequently, when we train MCOC using these data sets with uncertainties, MCOC will degenerate into an inefficient, instable, and inaccurate classifier. In other words, MCOC lacks the capacity of effectively dealing with noise, outliers, anomalies, class imbalance, nonlinear separable cases and other uncertainties in data.

To improve the performance of MCOC, we reformulate the model by introducing a kernel function to constraints, a fuzzy membership degree to each input point and penalty factors to objective functions of imbalanced classes in MCOC. Thus, the proposed new method (KFP-MCOC) can improve performance of the original MCOC approach in stability, efficiency and generalization, which reduces the effects of anomalies, class imbalance and nonlinearly separable problems significantly.

The rest of this paper is organized as follows: Section 2 describes basic principles of MCOC. Then the new KFP-MCOC is illustrated in Section 3. The experiment on credit risk evaluation and the results are demonstrated in Section 4. Finally, discussion and conclusions will be given in Sections 5 and 6 respectively.

2. MCO classifiers

Compared to many traditional methods in data mining, the multi-criteria optimization (MCO) approach based on optimization techniques was only recently introduced to practical applications. This is partly due to SVM being successfully applied to various domains at first, and eventually MCO approaches now are paid more attention. The two methods share common advantages of using flexible objectives and constraints to fit a decision function for separation of different classes.

Thus, a general classification problem using MCOC can be described as follows: For a binary classification problem, given the training data $T = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$, each input point $\boldsymbol{x}_i \in R^d$ belongs to either of the two classes with a label $y_i \in \{-1, 1\}$, $i = 1, \ldots, m$ for $y_i = -1$; $i = m + 1, \ldots, n$ for $y_i = 1$, where d is the dimensionality of the input space, and n is the sample size. In order to separate the two classes, e.g. Freed and Glover have chosen the two measures for any input point: The overlapping degree of deviation from the separating hyperplane, and the distance between input points and the separating hyperplane (Freed & Glover, 1981). 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 classified. Subsequently, Glover took the above two factors into consideration when building the classification models (Glover, 1990).

Let α_i ($\alpha_i \geqslant 0$) be the distance where an input point $\boldsymbol{x}_i$ deviates from the separating hyperplane, and the sum of the distance α_i is characterized by the function $f(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_p^p$ ($p \geqslant 1$) which should be minimized with respect to α_i, we have

$$\begin{aligned} &\min f(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_p^p \\ &\text{subject to } \boldsymbol{w}^T\boldsymbol{x}_i - b \geqslant -y_i\alpha_i, \qquad \alpha_i \geqslant 0, \quad \forall i. \end{aligned} \tag{1}$$

Obviously if $\alpha_i = 0$, the input point x_i is correctly classified. If $\alpha_i > 0$, the input point x_i is misclassified. Where the input point x_i is given training data, the weight vector w and the offset term b are unrestricted variables, $i = 1, 2, \ldots, n$.

Similarly, let β_i ($\beta_i \geqslant 0$) be the distance where an input point x_i departs from the separating hyperplane, then the sum of the