Applied Soft Computing 65 (2018) 139–151

Contents lists available at ScienceDirect

Applied Soft Computing

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

Dynamic ensemble classification for credit scoring using soft probability

Xiaodong Feng[a], Zhi Xiao[a,*], Bo Zhong[b], Jing Qiu[a], Yuanxiang Dong[c]

[a] School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China
[b] College of Mathematics and Statistics, Chongqing University, Chongqing 400044, PR China
[c] School of Management Science and Engineering, Shanxi University of Finance and Economics, Taiyuan, Shanxi 030006, PR China

A R T I C L E I N F O

Article history:
Received 15 September 2017
Received in revised form 15 January 2018
Accepted 16 January 2018

Keywords:
Credit scoring
Dynamic ensemble classification
Selective ensemble
Soft probability
Machine learning

A B S T R A C T

In recent years, classification ensembles or multiple classifier systems have been widely applied to credit scoring, and they achieve significantly better performance than individual classifiers do. Selective ensembles, an important part of this group of systems, are a promising field of research. However, none of them considers the relative costs of Type I error and Type II error for credit scoring when selecting classifiers, which bring higher risks for the financial institutions. Moreover, earlier dynamic selective ensembles usually select and combine classifiers for each test sample dynamically based on classifiers' performance in the validation set, regardless of their behaviors in the testing set. To fill the gap and overcome the limitations, we propose a new dynamic ensemble classification method for credit scoring based on soft probability. In this method, the classifiers are first selected based on their classification ability and the relative costs of Type I error and Type II error in the validation set. With the selected classifiers, we combine different classifiers for the samples in the testing set based on their classification results to get an interval probability of default by using soft probability. The proposed method is compared with some well-known individual classifiers and ensemble classification methods, including five selective ensembles, for credit scoring by using ten real-world data sets and seven performance indicators. Through these analyses and statistical tests, the experimental results demonstrate the ability and efficiency of the proposed method to improve prediction performance against the benchmark models.

1. Introduction

With the rapid development of the personal consumer finance market, people can get a credit card more easily without enough credit evaluation. The number of personal credit defaults has increased rapidly, especially during the financial crisis. As a useful tool for consumer credit risk assessment, credit scoring has received much attention from both the business and the academic worlds [1]. The models for credit scoring are developed to help the financial institutions to decide whether to grant credit to customers who apply to them or increase their credit limit [2]. Even a 1% improvement in the classification accuracy of the credit scoring model would greatly reduce losses from potential bad debt and increase the profit of financial institutions [3,4]. Lately, the BCBS [5] (Basel Committee on Banking Supervision) claims that banks and financial institutions should build strict and complex credit scoring systems to improve their credit risk levels and capital allocation.

As an important part of credit scoring, application scoring is used to estimate the applicant's probability of default in the future, which is the classification problem addressed in this paper. Over the last couple of decades, many classification methods for credit scoring have been built, and they fall into the following three main classifier families: individual classifiers, homogenous ensembles and heterogeneous ensembles [6]. The ensemble classification methods are proved to perform better than the individual classifiers do, especially the heterogeneous ensembles [7–9]. Selective ensembles, an important component of these ensembles, are gradually becoming a promising research area because of their promising performance [10]. Selective ensembles usually involve the following three steps: generating base classifiers, selecting a classifiers subset and combining the selected classifiers [6]. The methods for selecting the classifiers subset are usually based on the classifiers' classification accuracy, diversity or sensitivity [11,12], but none of them selects the classifiers based on the relative costs of Type I error

* Corresponding author.
E-mail addresses: dongcqu@126.com (X. Feng), xiaozhicqu@163.com (Z. Xiao), zhongbo@cqu.edu.cn (B. Zhong), 20160204132@cqu.edu.cn (J. Qiu), fox_dyx@163.com (Y. Dong).

https://doi.org/10.1016/j.asoc.2018.01.021
1568-4946/

and Type II error. That leads to a high Type I error of the selective ensembles, and brings a higher cost for commercial applications in credit scoring. To improve the classification ability, some studies consider dynamic ensemble selection methods, which are usually built based on the base classifiers' performance in the training set or validation set [13,14]. Few of them make the dynamic selection based on the classifiers' behaviors in the testing set, which decreases the generality of the model. Thus, a new dynamic ensemble classification method for credit scoring needs to be proposed.

Soft probability was proposed by Molodtsov [15]. Based on this, he built a novel approach to financial portfolio control and noted that soft probability has many advantages over classical probability [16,17]. (1) Soft probability is defined via immediate measurements over a statistical base, similar to basic physical notions, such as length and distance. (2) Soft probability has a parametric family of intervals as its values, which give bounds for the average values of the function for each pair of statistical point and length. (3) By definition, soft probability is a dynamic object, since in general, soft probability changes when new statistical data appear. (4) Soft probability is widely applicable, not only to stochastically stable phenomena. (5) The formalism of soft probability is much simpler than the formalism of classical probability.

Inspired by the above, we build a dynamic ensemble classification method based on soft probability (DECSP) for credit scoring. The reasons are as follows. First, the base classifiers of the selective ensembles can help to determine the statistical base, point and length of soft probability based on their classification ability and relative costs of Type I error and Type II error in the validation set. Second, the statistical point helps to select the classifiers subset, and the statistical length helps to determine interval probabilities of default for the samples in the testing set, whose lower and upper values are calculated by combining different classifiers based on their classification results. Lastly, the process of selecting and combining classifiers can be completed by only using soft probability, and this process is simpler than other state-of-the-art selective ensembles, which usually employ one or more algorithms for selecting and combining classifiers, respectively [11,12]. Therefore, soft probability provides a novel theory framework for building a dynamic ensemble classification method.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 explains the definition of soft probability and the proposed method in detail. Section 4 gives the experimental setup carried out for verifying the DECSP. Section 5 shows the results of the experiment. Section 6 presents some conclusions of this paper.

2. Literature review

The main idea of credit scoring is to build a quantitative model for estimating an applicant's credit worthiness based on a set of explanatory variables [18]. Estimating the probability of default (PD) is its main task, and it can be considered as a general population classification task [19]. During the last two decades, various classification algorithms have been exploited for credit scoring based on traditional statistical techniques or machine learning techniques. They are categorized as individual classifiers, homogenous ensembles and heterogeneous ensembles [6].

2.1. Individual classifiers

Individual classifiers mainly employ one statistical method or machine learning method to build credit scoring models. The statistical methods include linear discriminate analysis (LDA) [20], multiple discriminant analysis (MDA) [21], logistic regression (LR) [3], and Bayesian network [22]. However, many studies have indicated that machine learning methods have a higher prediction accuracy than the traditional statistical methods [7,18,23]. These methods include neural network (NN), decision trees (DT), support vector machine (SVM), genetic algorithm (GA), and naïve Bayes (NB) [2]. DT has been widely applied in building classification models, since it closely resembles human reasoning and is easy to understand [24]. Zhang et al. [25] employed vertical bagging to generate multiple decision trees and combine them, which shows outstanding prediction accuracy for credit scoring. SVM does not need many prior assumptions about the input data and can solve the problem with high-dimensional data, which has been widely used for credit scoring [26]. Harris [27] employed the clustered support vector machine (CSVM) for credit scoring and achieved a higher level of classification performance than other methods. Neural networks achieve a better discriminatory power than LR and other statistical methods by using multiple-layer networks and nonlinear transfer functions [28,29], and they show their superiority in building credit scoring models. The basic principles and optimization functions of the machine learning methods are diverse. Therefore, considering both the diversity and accuracy, this study employs multiple machine learning methods to generate the base classifiers for credit scoring.

2.2. Ensemble classification methods

The ensemble classification methods usually combine multiple diverse, unstable and good classifiers in some way. These classifiers are solving the same problem and collectively achieve a forecasting result with higher stability and accuracy [30,31]. Homogenous ensembles usually employ one of the abovementioned individual classification methods with various samples or parameters to build base classifiers [6]. Then, they use a majority voting rule or other methods to combine the classifiers. The well-known methods for generating various samples are bagging, boosting and the random subspace method [32–34]. Heterogeneous ensembles build the ensemble credit scoring models by employing more than one classification method [35,36]. Selective ensembles, an active research field of heterogeneous ensembles, are seldom employed in credit scoring, which represents a significant research gap [6].

Selective ensembles usually involve the following three steps: generating base classifiers, selecting a classifiers subset and combining them, and step two is its main task [37,38]. Hill-climbing ensemble selection (HCES) and genetic algorithm (GA) are two famous search-based methods, and they search in the space of subsets to present brilliant classification accuracy [10,37,39]. Some works regard classifiers subset selecting as an optimization problem and use some optimization-based methods to get the best classification accuracy, such as upper integrals [40], linear programming [41], and semi definite programming-based strategy [42]. The abovementioned methods select the classifiers mainly based on their classification accuracy (ACC). Moreover, some methods select the classifiers subset by ranking their diversity, precision, F-score, correlation, etc. [11,12,43], such as Kappa pruning, orientation ordering, and correlation minimization strategy [44,45]. However, none of them select classifiers based on their Type I error and Type II error with relative costs for credit scoring. Type I error indicates the number of customers with bad credit who are classified as having good credit, which brings higher costs for financial institutions than Type II error does [46]. Thus, they should focused on more when building the selective ensemble for credit scoring.

Meanwhile, some dynamic ensemble classification methods have also been studied. These methods usually select ensemble classifiers for each sample in the testing set. K-nearest-oracles (KNORA) based dynamic ensemble selection (DES) is a well-known method, which helps to search the K nearest neighbors in the validation set for each sample in the testing set and then selects the