Knowledge-Based Systems 26 (2012) 61–68

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Two credit scoring models based on dual strategy ensemble trees

Gang Wang [a,b,*], Jian Ma [c], Lihua Huang [d], Kaiquan Xu [c,e]

[a] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui, PR China
[c] Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
[d] School of Management, Fudan University, Shanghai 200433, PR China
[e] Department of Electronic Commerce, School of Business, Nanjing University, Nanjing, Jiangshu 210093, PR China

ARTICLE INFO

Article history:
Received 14 August 2009
Received in revised form 28 June 2011
Accepted 28 June 2011
Available online 13 July 2011

Keywords:
Credit scoring
Ensemble learning
Bagging
Random subspace
Decision tree

ABSTRACT

Decision tree (DT) is one of the most popular classification algorithms in data mining and machine learning. However, the performance of DT based credit scoring model is often relatively poorer than other techniques. This is mainly due to two reasons: DT is easily affected by (1) the noise data and (2) the redundant attributes of data under the circumstance of credit scoring. In this study, we propose two dual strategy ensemble trees: RS-Bagging DT and Bagging-RS DT, which are based on two ensemble strategies: bagging and random subspace, to reduce the influences of the noise data and the redundant attributes of data and to get the relatively higher classification accuracy. Two real world credit datasets are selected to demonstrate the effectiveness and feasibility of proposed methods. Experimental results reveal that single DT gets the lowest average accuracy among five single classifiers, i.e., Logistic Regression Analysis (LRA), Linear Discriminant Analysis (LDA), Multi-layer Perceptron (MLP) and Radial Basis Function Network (RBFN). Moreover, RS-Bagging DT and Bagging-RS DT get the better results than five single classifiers and four popular ensemble classifiers, i.e., Bagging DT, Random Subspace DT, Random Forest and Rotation Forest. The results show that RS-Bagging DT and Bagging-RS DT can be used as alternative techniques for credit scoring.

1. Introduction

The recent world financial tsunami arouses unprecedented attention of financial institutions on credit risk. A good credit risk assessment method can help financial institutions to grant loans to creditable applicants, thus increase profits; it can also deny credit for the non-creditable applicants, so decrease losses. In recent years, credit scoring has been become one of the primary ways for financial institutions to assess credit risk, improve cash flow, reduce possible risks and make managerial decisions [1,2]. The accuracy of credit scoring is critical to financial institutions' profitability. Even 1% of improvement on the accuracy of recognizing applicants with bad credit will decrease a great loss for the financial institutions [3].

Credit scoring was originally evaluated subjectively according to personal experiences, and later it was based on 5Cs: the character of the consumer, the capital, the collateral, the capacity and the economic conditions. However, with the tremendous increase in the number of applicants, it is impossible to conduct the work manually. Two categories of automatic credit scoring techniques, i.e., statistical techniques and Artificial Intelligence (AI) techniques have been studied by prior studies [4].

Some statistical techniques have been widely applied to build the credit scoring models, such as Linear Discriminant Analysis (LDA) [5,6], Logistic Regression Analysis (LRA) [7,8], Multivariate Adaptive Regression Splines (MARS) [9]. However, the problem with applying these statistical techniques to credit scoring is that some assumptions, such as the multivariate normality assumptions for independent variables, are frequently violated in reality, which makes these techniques theoretically invalid for finite samples [4].

In recent years, many studies have demonstrated that AI techniques, such as Artificial Neural Network (ANN) [8,10], decision tree (DT) [11,12], Case based Reasoning (CBR) [13,14] and Support Vector Machine (SVM) [2,15,16] can be used as alternative methods for credit scoring. In contrast with statistical techniques, AI techniques do not assume certain data distributions. These techniques automatically extract knowledge from training samples. According to previous studies, AI techniques are superior to statistical techniques in dealing with credit scoring problems, especially for nonlinear pattern classification [4]. Among all the AI techniques, DT is a widely used technique for three reasons: first, due to its intuitive representation, the resulting classification model

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China. Tel.: +86 0551 2904910; fax: +86 0551 2904910.
E-mail address: wgedison@gmail.com (G. Wang).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.06.020

is easy to assimilate by humans [17,18]. Second, DT is non-parametric. DT construction algorithms do not make any assumptions about the underlying distribution and are thus especially suitable for exploratory knowledge discovery. Third, DT can be constructed relatively fast compared to other techniques [18,19]. In despite of its merits, DT is less used as a credit scoring model because its classification accuracy is relative lower than others techniques, and easily affected by the noise data and the redundant attributes of data [20–22].

In this study, we propose two dual strategy ensemble trees: RS-Bagging DT and Bagging-RS DT, which are based on two ensemble strategies: bagging and random subspace, to reduce the affection from the noise data and the redundant attributes of data and get the relative higher classification accuracy. Ensemble is a machine learning paradigm where multiple learners are trained to solve the same problem [23]. In contrast to ordinary machine learning approaches that try to learn one hypothesis from the training data, ensemble learning tries to construct a set of hypotheses and combine them to use [24]. In order to reduce the influence of the noise data and the redundant attributes to the accuracy of DT, we introduce two ensemble strategies, i.e., bagging and random subspace. Firstly, as prior studies have proved that bagging performs better then others ensemble methods, e.g. boosting, in situations with large noise [25,26], we introduce bagging as one of ensemble strategies to reduce the influence of the noise data to DT. Secondly, as random subspace has been found to work well when there is redundant information which is dispersed across all the features [27,28], we introduce random subspace as another ensemble strategy to reduce the affection by the redundant attributes of data to DT. As we can process data with different order, i.e., firstly reduce the noise data using bagging strategy and then reduce the redundant attributes of data using random subspace strategy, or firstly reduce the redundant attributes of data using random subspace strategy and then reduce the noise data using bagging strategy, there are two dual strategy ensemble trees: RS-Bagging DT and Bagging-RS DT. For the testing and illustration purposes, two open credit datasets are used to verify the effectiveness of the proposed two ensemble methods, i.e., RS-Bagging DT and Bagging-RS DT. The experimental results reveal that DT gets the lowest average accuracy among five single classifiers, i.e., LRA, LDA, Multi-layer Perceptron (MLP) and Radial Basis Function Network (RBFN). In addition, RS-Bagging DT and Bagging-RS DT get the better results than five single classifiers and four popular ensemble classifiers, i.e., Bagging DT, Random Subspace DT, Random Forest and Rotation Forest. All these results illustrate that RS-Bagging DT and Bagging-RS DT can be used as alternative techniques for credit scoring.

The remainder of the paper is organized as follows. In Section 2, the background of DT, bagging and random subspace are presented. In Section 3, we propose two algorithms, i.e., RS-Bagging DT and Bagging-RS DT based on the bagging and the random subspace for credit scoring. In Section 4, we present the details of experiment design. Section 5 reports the experimental results. Based on the observations and results of these experiments, Section 6 draws conclusions and future research directions.

2. Background

2.1. Decision tree

A decision tree is a tree-like structure (Fig. 1) which divides a set of input samples based on some characteristics of their attributes into several smaller sets [17,29]. Unlike conventional statistical classifiers, which use all available features simultaneously and make a single membership decision for each pixel, the DT uses a multi-stage or sequential approach to the problem of label assignment. The labeling process is considered to be a chain of simple decisions based on the results of sequential tests rather than a single, complex decision. Sets of decision sequences form the branches of the DT, with tests being applied at the nodes.

DT construction involves the recursive partitioning of a set of training data, which is split into increasingly homogeneous subsets on the basis of tests applied to one or more of the attribute values [29]. These tests are represented by nodes. The univariate DT applies a test to a single attribute at a time, whereas the multivariate DT uses one or more attributes simultaneously. Labels are assigned to terminal (leaf) nodes by means of an allocation strategy, such as majority voting. In this study, we choose widely used C4.5 as base learner.

2.2. Bagging

Breiman's bagging, short for bootstrap aggregating, is one of the earliest ensemble learning algorithms [30]. It is also one of the most intuitive and simplest to implement, with a surprisingly good performance. Diversity in bagging is obtained by using bootstrapped replicas of the training dataset: different training data subsets are randomly drawn—with replacement—from the entire

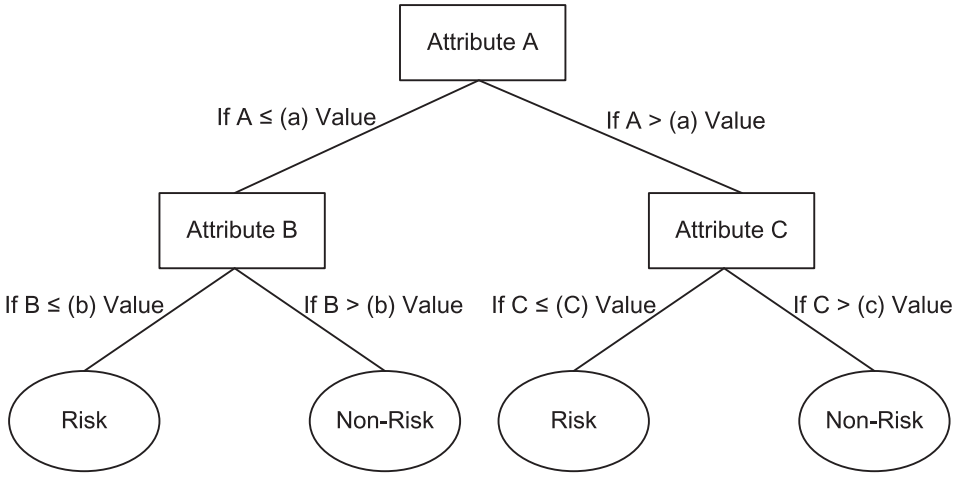

Fig. 1. An example of decision tree.