Expert Systems with Applications 39 (2012) 10916–10922

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

# Two-level classifier ensembles for credit risk assessment

A.I. Marqués [a], V. García [b], J.S. Sánchez [b,*]

[a] Department of Business Administration and Marketing, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castelló de la Plana, Spain
[b] Department of Computer Languages and Systems, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castelló de la Plana, Spain

ARTICLE INFO

*Keywords:*
Credit scoring
Classifier ensemble
Bagging
Boosting
Random subspace
Rotation forest

ABSTRACT

Many techniques have been proposed for credit risk assessment, from statistical models to artificial intelligence methods. During the last few years, different approaches to classifier ensembles have successfully been applied to credit scoring problems, demonstrating to be generally more accurate than single prediction models. The present paper goes one step beyond by introducing composite ensembles that jointly use different strategies for diversity induction. Accordingly, the combination of data resampling algorithms (bagging and AdaBoost) and attribute subset selection methods (random subspace and rotation forest) for the construction of composite ensembles is explored with the aim of improving the prediction performance. The experimental results and statistical tests show that this new two-level classifier ensemble constitutes an appropriate solution for credit scoring problems, performing better than the traditional single ensembles and very significantly better than individual classifiers.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

The recent world financial crisis has aroused increasing attention of banks and financial institutions on credit risk, which remains the most important and hard to manage and evaluate. The main problem comes from the difficulty to distinguish the creditworthy applicants from those who will probably default on repayments. One of the primary tools for credit risk management is credit scoring, which allows to assess credit risk, improve cash flow, reduce possible risks and make managerial decisions (Thomas, Edelman, & Crook, 2002). The decision to grant credit to an applicant was originally based upon subjective judgments made by human experts, using past experiences and some guiding principles. Common practice was to consider the classic five Cs of credit: character, capacity, capital, collateral and conditions (Rosenberg & Gleit, 1994). This method suffers, however, from high training costs, frequent incorrect decisions, and inconsistent decisions made by different experts for the same application.

Credit scoring is essentially a set of techniques that help lenders decide whether or not to grant credit to new applicants. Therefore, the objective of a credit scoring system is to distinguish "good" applicants from "bad" applicants, depending on the probability of default with their repayments (Hand & Henley, 1997). From a practical viewpoint, the process of credit scoring can be deemed as a prediction or classification problem where a new input sample (the credit applicant) must be categorized into one of the predefined classes based on a number of observed variables or attributes related to that sample. The input of the classifier consists of a variety of information that describes socio-demographic characteristics and economic conditions of the applicant, and the classifier will produce the output in terms of the applicant creditworthiness.

Because of the vast amount of information available, financial institutions have currently a need for advanced analytical tools that support the credit risk management processes in order to comply with the Basel regulatory requirements. As a consequence, many automatic credit scoring systems have been proposed in the literature. The most classical approaches are based on statistical models, such as logistic regression, linear discriminant analysis, and multivariate adaptive regression splines. However, the problem with using statistical techniques is that some assumptions, such as the multivariate normality for independent variables, are frequently violated, what makes them theoretically invalid for finite samples (Huang, Chen, Hsu, Chen, & Wu, 2004).

In recent years, several empirical studies have demonstrated that artificial intelligence techniques (decision trees, artificial neural networks, support vector machines, evolutionary computing) can be successfully used for credit risk management (Chi & Hsu, 2012; Huang, Chen, & Wang, 2007; Huang et al., 2004; Ince & Aktan, 2009; Martens et al., 2010; Ong, Huang, & Tzeng, 2005). Besides, an important advantage compared to statistical models is that the artificial intelligence methods do not assume any specific prior knowledge, but automatically extract information from past observations.

Although previous studies conclude that artificial intelligence techniques are superior to traditional statistical models, it is unlikely to find a single classifier achieving the best results on the whole application domain. Taking this into account, classifier ensembles have emerged to exploit the different behavior of

* Corresponding author. Tel.: +34 964 728350.
*E-mail address:* sanchez@uji.es (J.S. Sánchez).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.033

individual classifiers and reduce prediction errors. Recent practical investigations have demonstrated that classifier ensembles generally perform better than single prediction methods in most credit scoring problems (Doumpos & Zopounidis, 2007; Twala, 2010; Wang, Hao, Ma, & Jiang, 2011; West, Dellana, & Qian, 2005).

An ensemble of classifiers is efficient only if these have a minimum of errors in common (Ali & Pazzani, 1996; Bian & Wang, 2007). In other words, the individual classifiers have to make decisions as diverse as possible. Probably, using different training sets and using different attribute subsets are the two most typical strategies to generate a diverse set of classifiers. The distinction in purpose and performance between both approaches suggests a synergistic relationship between them that is worth to be explored. The idea is that, by using them in conjunction, the diversity induced by one method can be improved with the diversity produced by the other strategy in order to construct a composite ensemble approach significantly better than any single ensemble.

The focus of this paper is therefore primarily on exploring the joint use of both diversity induction strategies for the construction of composite ensembles in the scope of credit scoring. This can be viewed as a two-level ensemble that combines two single ensembles of different nature with the aim of improving the classification performance. Another point of investigation in this paper is whether the ordering of methods matters, that is, what are the practical implications of using first a data resampling algorithm followed by an attribute selection technique or vice versa?

We investigate these questions by using two resampling-based ensembles (bagging and AdaBoost) and two attribute-based algorithms (random subspace and rotation forest) in varied sequences. The details of these ensemble approaches are presented in Section 2. Section 3 introduces the proposed methodology. Section 4 provides a description of the experiments carried out, with their results in Section 5. Finally, Section 6 remarks the main conclusions and discusses directions for further research.

## 2. Classifier ensembles

An ensemble of classifiers (committee of learners, mixture of experts, multiple classifier system) consists of a set of individually trained classifiers (the base classifiers) whose decisions are combined in some way, typically by weighted or unweighted voting, when classifying new examples (Kuncheva, 2004). It has been found that in most cases the ensembles produce more accurate predictions than the base classifiers that make them up (Dietterich, 1997). Nonetheless, as already said, for an ensemble to achieve better generalization capability than its members, it is critical that the ensemble consists of highly accurate base classifiers whose decisions are as diverse as possible.

Various strategies have been developed to enforce diversity on the classifiers that form the ensemble. For instance, Kuncheva (2003) identified four basic approaches: (i) using different combination schemes, (ii) using different classifier models, (iii) using different attribute subsets, and (iv) using different training sets. These two last strategies constitute the most commonly used methods. In this context, two representative ensemble algorithms that use different training sets are bagging and boosting, whereas random subspace and rotation forest constitute two well-known examples of the ensemble methods that utilize different attribute subsets. In the following subsections, these popular approaches will be briefly described.

### 2.1. Bagging

In its standard form, the bagging (Bootstrap Aggregating) algorithm (Breiman, 1996) generates $M$ bootstrap samples $T_1, T_2, \ldots, T_M$ randomly drawn (with replacement) from the original training set $T$ of size $n$. From each bootstrap sample $T_i$ (also of size $n$), a base classifier $C_i$ is induced by the same learning algorithm. Predictions on new observations are made by taking the majority vote of the ensemble $C^*$ built from $C_1, C_2, \ldots, C_M$. As bagging resamples the training set with replacement, some instances may be represented multiple times while others may be left out.

Since each ensemble member is not exposed to the same set of instances, they are different from each other. By voting the predictions of each of these classifiers, bagging seeks to reduce the error due to variance of the base classifier.

### 2.2. Boosting

Similar to bagging, boosting also creates an ensemble of classifiers by resampling the original data set, which are then combined by majority voting. However, in boosting, resampling is directed to provide the most informative training data for each consecutive classifier.

The AdaBoost (Adaptive Boosting) algorithm proposed by Freund and Schapire (1996) constitutes the best known member in boosting family. It generates a sequence of base classifiers $C_1, C_2, \ldots, C_M$ by using successive bootstrap samples $T_1, T_2, \ldots, T_M$ that are obtained by weighting the training instances in $M$ iterations. AdaBoost initially assigns equal weights to all training instances and in each iteration, it adjusts these weights based on the misclassifications made by the resulting base classifier. Thus, instances misclassified by model $C_{i-1}$ are more likely to appear in the next bootstrap sample $T_i$. The final decision is then obtained through a weighted vote of the base classifiers (the weight $w_i$ of each classifier $C_i$ is computed according to its performance on the weighted sample $T_i$ it was trained on).

### 2.3. Random subspace

The random subspace method (RSM) is an ensemble construction technique proposed by Ho (1998), in which the base classifiers $C_1, C_2, \ldots, C_M$ are trained on data sets $T_1, T_2, \ldots, T_M$ constructed with a given proportion of attributes picked randomly from the original set of features $F$. The outputs of the models are then combined, usually by a simple majority voting scheme. The author of this method suggested to select about 50% of the original features.

This method may benefit from using random subspaces for both constructing and aggregating the classifiers. When the data set has many redundant attributes, one may obtain better classifiers in random subspaces than in the original feature space. The combined decision of such classifiers may be superior to a single classifier constructed on the original training data set in the complete feature space. On the other hand, when the number of training cases is relatively small compared with the data dimensionality, by constructing classifiers in random subspaces one may solve the small sample size problem.

### 2.4. Rotation forest

Rotation forest (Rodriguez, Kuncheva, & Alonso, 2006) refers to a technique to generate an ensemble of classifiers, in which each base classifier is trained with a different set of extracted attributes.

The main heuristic is to apply feature extraction and to subsequently reconstruct a full attribute set for each classifier in the ensemble. To this end, the feature set $F$ is randomly split into $L$ subsets, principal component analysis (PCA) is run separately on each subset, and a new set of linear extracted attributes is constructed by pooling all principal components. Then the data are transformed linearly into the new feature space. Classifier $C_i$ is trained with this data set. Different splits of the feature set will lead to

متن کامل مقاله ← دریافت فوری

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات