Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 3028–3033
Expert Systems with Applications
www.elsevier.com/locate/eswa

An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring

Loris Nanni *, Alessandra Lumini

DEIS, IEIIT – CNR, Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

Abstract

In this paper, we investigate the performance of several systems based on ensemble of classifiers for bankruptcy prediction and credit scoring.

The obtained results are very encouraging, our results improved the performance obtained using the stand-alone classifiers. We show that the method "Random Subspace" outperforms the other ensemble methods tested in this paper. Moreover, the best stand-alone method is the multi-layer perceptron neural net, while the best method tested in this work is the Random Subspace of Levenberg–Marquardt neural net.

In this work, three financial datasets are chosen for the experiments: Australian credit, German credit, and Japanese credit.

Keywords: Bankruptcy prediction; Credit scoring; Ensemble of classifiers

1. Introduction

An important issue in financial decision-making is to predict, timely and correctly, business failure (Atiya, 2001; Zhang, Hu, Patuwo, & Indro, 1999) (e.g. bankruptcy prediction and credit scoring). The credit scoring models permit to discriminate between good credit group and bad credit group (Chen & Huang, 2003). The benefits obtained developing a reliable credit scoring system are (Tsai & Wu, 2007; West, 2000):

- reducing the cost of credit analysis;
- enabling faster decision;
- insuring credit collections and diminishing possible risk.

Several financial decision-making methods based on machine learning (examples of machine learning techniques used to solve the above financial decision-making problems are Atiya, 2001; Huang, Chen, Hsu, Chen, & Wu, 2004; Lee, Chiu, Chou, & Lu, 2006) use the multi-layer perceptron (MLP) (Haykin, 1999) as classifier. Other tested classifiers are the Decision Tree and the Support Vector machine. We want to stress that these studies show that the machine learning based systems are better than the traditional (statistical) methods for bankruptcy prediction and credit scoring problems (Huang et al., 2004; Ong, Huang, & Tzeng, 2005; Vellido, Lisboa, & Vaughan, 1999; Wong & Selvi, 1998). In Tsai and Wu (2007) the authors compare a single MLP classifier with multiple classifiers and diversified multiple classifiers on three datasets. However, they conclude that there is no an exact winner.

In Table 1, several machine learning based methods are compared.

The main drawbacks of the machine learning based methods proposed in the literature are (Tsai & Wu, 2007).

- In several works only one dataset is used to validate the proposed system.
- In several works only the accuracy is used to validate the proposed system (examples of exceptions are Lee et al., 2006; Lee, Chiu, Lu, & Chen, 2002; Tsai & Wu, 2007).

* Corresponding author.
E-mail address: lnanni@deis.unibo.it (L. Nanni).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.01.018

Table 1
Several machine learning based methods are compared (from Tsai and Wu (2007))

Paper	Classifier	Datasets	Accuracy	Type I/II error
Fan and Palaniswami (2000)	MLP	Australian	Yes	No
West (2000)	MLP	Australian/German	Yes	No
Atiya (2001)	MLP	US	Yes	No
Lee et al. (2002)	MLP + LDA	Taiwan	Yes	Yes
Huang et al. (2004)	MLP	Taiwan/US	Yes	No
Min and Lee (2005)	MLP	Korea	Yes	No
Shin et al. (2005)	MLP	Korea	Yes	No
West et al. (2005)	MLP ensembles	Australian/German	Yes	No
Lee et al. (2006)	MLP	Taiwan	Yes	Yes
Tsai and Wu (2007)	MLP ensembles	Australian/German/Japanese	Yes	Yes

- Few papers study ensemble of classifiers (Frosyniotis, Stafylopatis, & Likas, 2003; Ghosh, 2002; Kang & Doermann, 2003; Roli, Kittler, & Windeatt, 2004) on both credit scoring and bankruptcy prediction problems (Tsai & Wu, 2007; Wong & Selvi, 1998).

In this paper, we improve the results of Tsai and Wu (2007), we have made a deep study of the ensemble of classifiers for bankruptcy prediction and credit scoring. We have tested four different methods for creating an ensemble of classifiers and we have tested four different classifiers. We show that the Random Subspace method (RS) (Ho, 1998) improve the performance of the classifiers. In our test the best stand-alone classifier is MLP (as reported in several papers) but the best system is a Random Subspace of Levenberg–Marquardt neural nets.

The paper is organized as follows: in Section 2, a brief literature of the multiple classifiers is reported, in Section 3, the experimental results are presented. Finally, in Section 4, some concluding remarks are given.

2. Ensemble of classifiers

To improve the performance of the single classifier approaches the combination of multiple classifiers has been proposed in the field of machine learning (Fierrez-Aguilar, Nanni, Lopez-Penalba, Ortega-Garcia, & Maltoni, 2005; Bologna et al., 2002; Melville & Mooney, 2005). The multiple classifier systems are based on the combination of a pool of classifiers such that the their fusion achieves higher performance than the stand-alone classifiers. Hence, an ensemble of classifiers is a set of classifiers, whose individual classification decisions are combined in some way (Lumini & Nanni, 2006).

The key idea of several methods for building ensemble of classifiers is to modify the training data set (generating NK new training sets), build classifiers on these NK new training sets (in this paper $NK = 50$), and then combines them into a final decision rule. The ensemble of classifiers methods tested in this paper are

- *Bagging (BA)* (Breiman, 1996): Given a training set S, it generates NK new training sets $S_1, \ldots, S_{NK}$; randomly extracting a subset of the training patterns, each new set S_i is used to train exactly one classifier.
- *Random Subspace (RS)* (Ho, 1998): Each stand-alone classifier uses only a subset of all features for training and testing. In this work the percentage of the features retained in each training set is 50%.
- *Class Switching (CW)* (Martınez-Muñoz & Suárez, 2005): Each new training set is obtained randomly switching the classes of the training examples.
- *Rotation Forest (RF)* (Rodriguez, Kuncheva, & Alonso, 2006): The feature set is divided in subsets of dimension 3, for each subset a subset of classes is randomly selected, a bootstrap of samples is drawn and the principal component analysis (PCA) is applied on the features that belong to subset and on the patterns that belong to a given subset. These projections are combined to build a projection matrix, named "rearranged rotation matrix" (see Rodriguez et al., 2006 for details) that builds a modified training set.

The best results are obtained by the Random Subspace, a scheme of the RS is reported in Fig. 1.

Several decision rules can be used to combine the classifiers that build an ensemble of classifiers; in this paper we test the "Sum rule" (Kittler, Hatef, Duin, & Matas, 1998). The sum rule selects as final score ($score(\mathbf{s}, c)$) the sum of the scores of the pool of NK classifiers

$$score(\mathbf{s}, c) = \sum_{j=1,\ldots,NK} sim(\mathbf{s}, c)$$

where $sim(\mathbf{s}, c)$ is the similarity, of the pattern $\mathbf{s}$ to the class c, obtained by the classifier j.

We have tested the following classifiers:

- Levenberg–Marquardt neural net with five hidden units[1] (LM) (Duda, Hart, & Stork, 2000).
- Multi-layer perceptron neural net with five hidden units[1] (MLP).
- Radial Basis function Support Vector Machine (RV)[2] (Duda et al., 2000) (parameters: $Gamma = 1$, $C = 1$);
- The old 5-nearest neighbour.[1]

3. Experiments

In our experiments we have used the same dataset[3] used in Tsai and Wu (2007):

[1] Implemented as in PrTools 3.1.7 ftp://ftp.ph.tn.tudelft.nl/pub/bob/prtools/prtools3.1.7.

[2] Implemented as in the OSU svm toolbox.

[3] These datasets are available from the UCI repository.