

Chapter 4.19

Empirical Evaluation of Ensemble Learning for Credit Scoring

Gang Wang

Fudan University, PR China & City University of Hong Kong, Hong Kong

Jin-xing Hao

City University of Hong Kong, Hong Kong

Jian Ma

City University of Hong Kong, Hong Kong

Li-hua Huang

Fudan University, PR China

ABSTRACT

Credit scoring is an important finance activity. Both statistical techniques and Artificial Intelligence (AI) techniques have been explored for this topic. But different techniques have different advantages and disadvantages on different datasets. Recent studies draw no consistent conclusions to show that one technique is superior to the other, while they suggest combining multiple classifiers, i.e., ensemble learning, may have a better performance. In this study, we conduct an empirical evaluation of the performance of three

popular ensemble methods, i.e., bagging, boosting, and stacking, based on four base learners, i.e., Logistic Regression Analysis (LRA), Decision Tree (DT), Artificial Neural Network (ANN) and Support Vector Machine (SVM). The experiment uses the credit dataset including 239 companies' financial records from China, collected by the Industrial and Commercial Bank of China. Results reveal that ensemble learning can substantially improve individual base learners. Stacking, in our experiments, gets the best performance in terms of all six performance indicators, i.e., type I error, type II error, average accuracy, precision, recall, and F-value.

DOI: 10.4018/978-1-60960-818-7.ch4.19

INTRODUCTION

The recent world financial tsunami arouses unprecedented attention of financial institutions on credit risk. Especially for any credit-granting institution, such as commercial banks and certain retailers, the ability to discriminate good customers from bad ones is crucial to their profit (Yu, Wang & Lai, 2008). These institutions must balance risks as well as returns. A good credit risk assessment method can help them to grant loans to more 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 credit-granting institutions to assess credit risk, improve cash flow, reduce possible risks and make managerial decisions (Huang, Chen & Wang, 2007).

The purpose of credit scoring is to classify the applicants into two types: applicants with good credit and applicants with bad credit. Applicants with good credit have great possibility to repay financial obligation; applicants with bad credit have high possibility of defaulting. For credit scoring, the accuracy is quite significant to financial institutions' profitability. For example, the accuracy of credit scoring of applicants with bad credit increases only 1% may retrieve a great loss for the financial institutions (Hand & Henley, 1997).

Many credit scoring models have been developed by credit-granting institutions and researchers for the credit admission decision. 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. But 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., traditional statistical techniques and Artificial Intelligence (AI) techniques, have been studied by prior researches (e.g., Huang, Chen, Hsu, Chen & Wu, 2004).

Some traditional statistical techniques have been widely applied to build the credit scoring models, such as Logistic Regression Analysis (LRA) (Thomas, 2000; West, 2000), Linear Discriminant Analysis (LDA) (Reichert, Cho & Wagner, 1983; Karels & Prakash, 1987), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991). Although these methods can be used to assess credit risk, the ability to discriminate good customers from bad ones is still not satisfactory (Yu, Wang & Lai, 2008).

In recent years, many studies have demonstrated that AI techniques such as Artificial Neural Networks (ANN) (Desai, Crook, & Overstreet, 1996; West, 2000), Case-Based Reasoning (CBR) (Buta, 1994; Shin & Han, 2001), and Support Vector Machine (SVM) (Baesens, Gestel, Viaene, Stepanova, Suykens & Vanthienen, 2003; Schebesch & Stecking, 2005; Huang, Chen & Wang, 2007) can be used as alternative methods for credit scoring. Some comprehensive introductions of the methods in credit scoring can be found in the three recent surveys (Rosenberg & Gleit, 1994; Thomas, 2000; Baesens, Gestel, Viaene, Stepanova, Suykens & Vanthienen, 2003; Zekic-Susac, Sarlija, & Bensic, 2004). In contrast with traditional 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 traditional statistical techniques in dealing with credit scoring problems, especially for nonlinear pattern classification (Huang, Chen, Hsu, Chen & Wu, 2004).

Recently, there is a growing interest that existing applications of single AI technique can be further improved by ensemble methods. Latest researches (Yu, Wang & Lai, 2008; Tsai & Wu, 2008; Huang & Chen, 2009) have shown that such ensemble methods have performed better than single AI technique for credit scoring. However, the application of ensemble methods in credit scoring is a relatively new and untried area. To the best of our knowledge, this may be the first

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/empirical-evaluation-ensemble-learning-credit/56189

Related Content

Investigation of Software Reliability Prediction Using Statistical and Machine Learning Methods

Pradeep Kumar and Abdul Wahid (2017). *Handbook of Research on Machine Learning Innovations and Trends* (pp. 251-271).

www.irma-international.org/chapter/investigation-of-software-reliability-prediction-using-statistical-and-machine-learning-methods/180949

An Ensemble of Hybrid Search-Based Algorithms for Software Effort Prediction

Wasiur Rhmann (2021). *International Journal of Software Science and Computational Intelligence* (pp. 28-37).

www.irma-international.org/article/an-ensemble-of-hybrid-search-based-algorithms-for-software-effort-prediction/280515

Potentials of Quadratic Neural Unit for Applications

Ricardo Rodriguez, Ivo Bukovsky and Noriyasu Homma (2011). *International Journal of Software Science and Computational Intelligence* (pp. 1-12).

www.irma-international.org/article/potentials-quadratic-neural-unit-applications/60745

Finding Optimal Input Values for Desired Target Output by Using Particle Swarm Optimization Algorithm Within Probabilistic Models

Goran Klepac (2018). *Incorporating Nature-Inspired Paradigms in Computational Applications* (pp. 76-107).

www.irma-international.org/chapter/finding-optimal-input-values-for-desired-target-output-by-using-particle-swarm-optimization-algorithm-within-probabilistic-models/202192

AI-Powered Healthcare Solutions: Bridging the Medical Gap in Underserved Communities Worldwide

Leelawati Pokhrel, Mohd Arsalan, Priyanka Rani, Puneet Garg and Placido Rogerio Pinheiro (2026). *Applied AI and Computational Intelligence in Diagnostics and Decision-Making* (pp. 57-86).

www.irma-international.org/chapter/ai-powered-healthcare-solutions/390112