Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers & Operations Research 32 (2005) 2543–2559
computers & operations research
www.elsevier.com/locate/dsw

Neural network ensemble strategies for financial decision applications

David West*, Scott Dellana, Jingxia Qian

Department of Decision Sciences, College of Business Administration, East Carolina University, Greenville, NC 27836, USA Voice 252-328-6370

Accepted 29 March 2004

Available online 25 May 2004

Abstract

Considerable research effort has been expended to identify more accurate models for decision support systems in financial decision domains including credit scoring and bankruptcy prediction. The focus of this earlier work has been to identify the "single best" prediction model from a collection that includes simple parametric models, nonparametric models that directly estimate data densities, and nonlinear pattern recognition models such as neural networks. Recent theories suggest this work may be misguided in that ensembles of predictors provide more accurate generalization than the reliance on a single model. This paper investigates three recent ensemble strategies: crossvalidation, bagging, and boosting. We employ the multilayer perceptron neural network as a base classifier. The generalization ability of the neural network ensemble is found to be superior to the single best model for three real world financial decision applications.

Keywords: Credit scoring; Bankruptcy prediction; Ensembles; Bagging; Boosting

1. Introduction

Financial credit is an immense global industry. In the United States alone the annual transactions of Visa, Mastercard, Discover, and American Express credit cards totaled $1.2 trillion from over 500 million cards in circulation. The outstanding level of consumer debt in the U.S. totals about $1.5 trillion, with high interest credit card loans comprising $568.4 billion of that total. More than 4% of credit card loans are delinquent and placed for collection every year. U.S. bankruptcy filings for the

* Corresponding author. Tel.: +1-252-3286370; fax: +1-919-3284092.
E-mail addresses: westd@mail.ecu.edu (D. West), dellanas@mail.ecu.edu (S. Dellana).

0305-0548/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2004.03.017

year 2002–2003 set a record level, totaling 1,650,279, which includes 37,182 business bankruptcy filings.

There is a clear need for accurate decision support for both the credit granting decision and the monitoring of the ongoing health of credit customers. An improvement in accuracy of even a fraction of a percent translates into significant future savings for the credit industry.

Traditional methods of financial decision support include scorecards for consumer credit [1–5] and discriminant models for assessing corporate financial health [6,4]. Both are essentially multivariate linear models that output a probability that the client will repay debt as agreed. Recent research interest has focused on more complex nonlinear models, particularly neural networks, to increase the credit decision accuracy [2,6–19]. The reader is referred to Smith and Gupta [20] for a recent survey of the application of neural networks in a diverse range of operations research problems that include financial forecasting and creditworthiness.

The focus of prior research has been to identify the "single best" model that is most accurate for a given financial decision application. This reliance on a single model may be misguided. Recent studies of ensembles (or committees) of predictors have demonstrated the potential to reduce the generalization error of a single model from 5% to 70% [21,22]. Three major strategies have been advanced for forming ensembles of predictors. The simplest is the crossvalidation (CV) neural network ensemble where all ensemble members are trained with the same data [23,16]. The second and third strategies create perturbed versions of the training set so that ensemble members learn from different variants of the original training data. Bagging ensembles create a unique training set for each ensemble member by sampling with replacement over a uniform probability distribution on the original data [21,24,25]. This creates training sets where some observations are replicated and others may be missing. Boosting is also a re-sampling strategy, with a probability distribution that is dependent on the misclassification rate for each observation [26,16]. Boosting is an iterative algorithm where the probability of the misclassified observations is increased and the corresponding probability of correctly classified observations is decreased over time. As boosting progresses, the composition of the training sets becomes increasingly dominated by hard-to-classify examples. The purpose of this research is to investigate the accuracy of ensembles of neural networks formed from these three strategies for credit granting and bankruptcy decision applications.

In the next section of this paper we review the recent theory and application of ensembles, with particular attention given to neural networks. Specific research questions are defined in this section. The research methodology is described in Section 3, and in Section 4 the comparison of generalization errors for the neural network ensemble strategies is discussed. We conclude in Section 5 with guidelines for implementing neural network ensembles for financial decision applications.

2. Ensemble strategies

The basic concept of the ensemble method is that diverse perspectives on different aspects of a problem can be combined to produce a high quality decision. For example, O'Leary [27] investigated human performance in the task of knowledge acquisition of probability estimates. He compared the relative performance of individuals versus groups of "multiple experts" (i.e., ensembles). His results suggest that knowledge acquisition from groups provided more correct probability orderings than the orderings from individuals acting alone. This finding is consistent with earlier research cited