Using supervised kernel locality preserving projections to improve classifier performance on credit rating forecasting

Shian-Chang Huang*
Department of Business Administration
College of Management
National Changhua University of Education
No.2, Shi-Da Road, Changhua 500
Taiwan, R.O.C.

Chih-Wei Lee
Graduate Institute of Finance
National Taipei College of Business
Taiwan, R.O.C.

Ming-Jen Chang
Institute of International Economics
National Dong Hwa University
Taiwan, R.O.C.

Tung-Kuang Wu†
Department of Information Management
National Changhua University of Education
Changhua
Taiwan, R.O.C.

Abstract

Kernel classifiers (such as support vector machines) have been successfully applied in numerous areas, and have demonstrated excellent performance. However, due to the high dimensionality and nonlinear distribution of financial input data in credit rating forecasting, finding a suitable low dimensional subspace is a key step to enhance classifier performance. By integrating supervised kernel locality preserving projections (SKLPP) with kernel classifiers, this study develops a novel prediction model for credit ratings forecasting. SKLPP is employed to gain a perfect approximation of data manifold and simultaneously preserve local within-class geometric structures according to prior class-label information.

*E-mail: shhuang@cc.ncue.edu.tw
†E-mail: tkwu@im.ncue.edu.tw (corresponding author)

Journal of Information & Optimization Sciences
Vol. 32 (2011), No. 1, pp. 189–204

Empirical results indicate that, compared with other dimensionality reduction methods, the performance improvement owing to SKLPP is significant. Moreover, our hybrid classifier outperforms other conventional classifiers.

Keywords: *Kernel Locality Preserving Projections, Subspace Analysis, Dimensionality Reduction, Kernel Classifier, Credit Rating*

1. Introduction

Credit rating assesses the credit worthiness of an individual, corporation, or even a country. Typically, a credit rating indicates to a lender or an investor the probability of the borrower repaying a loan. Therefore, credit ratings are important determinants of risk premiums and even the marketability of corporate bonds.

Recently, credit rating forecasting had been a critical issue in the banking industry. All banking institutes and their regulators attempt to search for a precise internal expert system to model the credit quality of their evaluation borrowers. Further, the subprime mortgage crisis in the latter half of 2007 profoundly affected the U.S. banking sector. Banks that can accurately estimate credit risk will be the most profitable. Thus, the objective of this study was to develop a reliable and accurate prediction model for risk assessment.

Corporate credit rating prediction models are studied intensively by the academic and business community. Many researchers have attempted to construct automatic classification systems using data mining methods, such as statistical and artificial intelligence. However, due to the high dimensionality of input financial data, this study combined a supervised kernel locality preserving projections with kernel classifiers to enhance forecasting performance.

Numerous classification techniques have been adopted for credit scoring. These techniques include (1) traditional statistical methods; such as discriminant analysis, logistic regression (Steenackers and Goovaerts, 1989; Stepanova and Thomas, 2001), and Bayesian network, and (2) nonparametric statistical models such as k-nearest neighbor (Henley and Hand, 1997), (3) decision trees (Yobas *et al*., 2000), and (4) neural networks (Desai *et al*., 1996; West, 2000; Yobas *et al*., 2000). Recently, the support vector machine (SVM) method (Vapnik, 1999; Cristianini and Shawe-Taylor, 2000; Schoelkopf *et al*., 1999), a special form of kernel classifiers, has become increasingly popular. The formulation of SVM embodies the