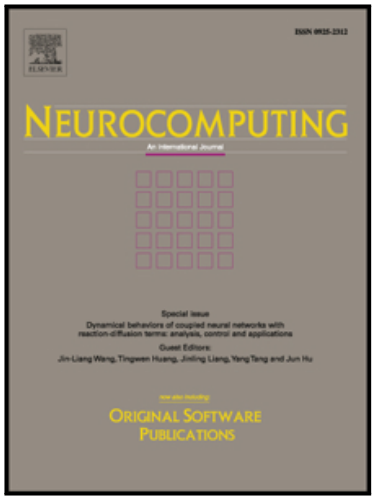
Communicated by Yukun Bao

Accepted Manuscript



An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking

Qi Zhang, Jue Wang, Aiguo Lu, Shouyang Wang, Jian Ma

PII: S0925-2312(17)31232-8
DOI: 10.1016/j.neucom.2017.07.002
Reference: NEUCOM 18690

To appear in: *Neurocomputing*

Received date: 22 July 2016
Revised date: 26 June 2017
Accepted date: 4 July 2017

Please cite this article as: Qi Zhang, Jue Wang, Aiguo Lu, Shouyang Wang, Jian Ma, An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.07.002

An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking

Qi Zhang*‡, Jue Wang*†, Aiguo Lu§, Shouyang Wang*, Jian Ma‡
*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
§Department of Applied Mathematics, Xi'an Shiyou University,Xian, 710065, China
‡Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
†University of Chinese Academy of Sciences, Beijing 100190, China

***Abstract*—With rapid development of financial services and products, credit risk assessment has recently gained considerable attention in the field of financial risk management. In this paper, an improved credit risk assessment approach is presented. Based on the credit data from China Banking Regulatory Commission (CBRC), a multi-dimensional and multi-level credit risk indicator system is constructed. In particular, we present an improved sequential minimal optimization (SMO) learning algorithm, named four-variable SMO (FV-SMO), for credit risk classification model. At each iteration, it jointly selects four variables into the working set and an theorem is proposed to guarantee the analytical solution of sub-problem. The assessment is made on China credit dataset and two benchmark credit datasets from UCI database and CD-ROM database. Experimental results demonstrate FV-SMO is competitive in saving the computational cost and outperforms other five state-of-the-art classification methods in credit risk assessment accuracy.**

Keywords: Credit risk assessment, SVM, Sequential minimal optimization (SMO), Four-variable working set

I. Introduction

The assessment of financial credit risk is emerging as an important research topic in the banking industry. The financial credit risk indicates the risk associated with financing, in other words, a borrower cannot pay the lenders, or goes into loan default. Credit risk assessment has become a particularly challenging issue for banks and financial institutions to access the performance of borrowers (customers), serving as the impetus to evaluate the credit admission or potential business failure of customers in order to make early actions. The great loss resulted from the financial distress or bankruptcy of customers usually leads to considerable criticism on the functionality of financial institutions due to the inappropriate evaluation of credit risk.

Most governments are forced to implement rescue plans for the banking systems with more effective credit risk assessment. In China, the massive credit boom poses challenge for the quality of bank assets. In fact, total bad loans reached 1.27 trillion yuan at the end of 2015, the highest since the global financial crisis, on the back of an economic slowdown and a ballooning corporate debt. An meticulous management information system is in urgent requirement. Credit risk assessment, which enables or supports an early-warning detection and fast response mechanism, is a key in this system. Since 2004, the China Banking Regulatory Commission (CBRC), which is responsible for regulation of banking industry in China, enables a reporting system for credit data collection. In recent years, CBRC has attached much importance to risk characteristics mining, custom's behavior analysis and risk assessment model.

Generally, credit risk refers to the risk that a bank borrower or a counterparty fails to meet its obligations in accordance with the agreed terms [1]. Numerous methods have been proposed in the literature to develop accurate classifier models to predict the default risk. Many statistic and optimization models are widely applied, such as linear discriminant analysis (LDA) [2], logistic regression analysis (LRA) [3], [4], multivariate adaptive regression splines (MARS) [5] and multi-criteria optimization classifier [6], [7]. However, the assumptions embedded within these statistical models, such as the multivariate normality assumptions for independent variables, are not satisfied in reality, which makes these methods theoretically invalid for finite samples [8]. Meanwhile, these models usually fail to capture enough information of nonlinear structure of real credit data. Recent studies focus on the research of artificial intelligent (AI) techniques for credit assessment, including artificial neural networks (ANN) [9], [10], radial basis function (RBF) model [11], decision tree [12], Bayesnet [13], extreme learning machine (ELM) [14], [15], support vector machine