Applied Soft Computing 11 (2011) 2906–2915

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An empirical study of classification algorithm evaluation for financial risk prediction☆

Yi Peng[a], Guoxun Wang[a], Gang Kou[a,*], Yong Shi[b,c]

[a] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, PR China
[b] CAS Research Center on Fictitious Economy and Data Sciences, Beijing 100080, PR China
[c] College of Information Science & Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA

ARTICLE INFO

Article history:
Received 13 September 2009
Received in revised form 26 July 2010
Accepted 28 November 2010
Available online 7 December 2010

Keywords:
Classification algorithm
Multiple criteria decision making (MCDM)
Financial risk prediction
Knowledge-rich financial risk analysis

ABSTRACT

A wide range of classification methods have been used for the early detection of financial risks in recent years. How to select an adequate classifier (or set of classifiers) for a given dataset is an important task in financial risk prediction. Previous studies indicate that classifiers' performances in financial risk prediction may vary using different performance measures and under different circumstances. The main goal of this paper is to develop a two-step approach to evaluate classification algorithms for financial risk prediction. It constructs a performance score to measure the performance of classification algorithms and introduces three multiple criteria decision making (MCDM) methods (i.e., TOPSIS, PROMETHEE, and VIKOR) to provide a final ranking of classifiers. An empirical study is designed to assess various classification algorithms over seven real-life credit risk and fraud risk datasets from six countries. The results show that linear logistic, Bayesian Network, and ensemble methods are ranked as the top-three classifiers by TOPSIS, PROMETHEE, and VIKOR. In addition, this work discusses the construction of a *knowledge-rich financial risk management process* to increase the usefulness of classification results in financial risk detection.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Risk is exposure to uncertain consequences, usually unfavorable outcomes [1,2]. Financial risks are uncertainties associated with any form of financing, including credit risk, business risk, investment risk, and operational risk. Take operational risk in health insurance industry as an example. At least 3% of the United States' health care expenditure, which in calendar-year 2003 alone amounted to $1.7 trillion, is lost to fraud or erroneous payment [3]. Another example is credit card debt risk. The total credit card debt at the end of the first quarter of 2002 in the US is about $660 billion [4] and the total credit card holders declared bankruptcy in 2003 are more than 1.6 million [5].

Financial risk prediction is an important and widely studied topic in the domain of financial analysis since it can help companies to detect financial risks in advance and take appropriate actions to minimize the defaults. Many financial risk prediction tasks are basically binary classification problems, which means observations are assigned to one of the two groups after data analysis [6]. The focuses of this paper are credit risk and fraud risk problems. Credit risk denotes the probability that debtors will not pay their debts and the risk prediction is to classify an account as normal or default. Fraud risk is the risk of financial loss resulting from deception and the risk prediction is to classify a data record as normal or fraud.

Classification methods bring several advantages to credit and fraud risk management. First, traditional methods of credit and fraud risk assessment rely on personal judgments, which are based on previous experience. As the business demands and the size of databases increase, the traditional approach cannot evaluate credit and fraud risks efficiently. With the development of computer power and data storage technologies, classification algorithms can be used to quickly predict credit and fraud risk and consequently expedite the investigative process. Second, classification models provide higher prediction accuracies than traditional approach (e.g., [7]), thus increasing profits and decreasing losses. Third, the decisions based on the results of classification methods are objective since they do not involve human biases. In addition, the use of classification methods for financial risk prediction produces some indirect advantages, such as fast processing of credit applications, flexible management of credit risk, fewer requirements of human

☆ A short version of this paper appeared previously at the workshop on Computational Finance and Business Intelligence in the International Conference on Computational Science, 25–27, May 2009 [74].

* Corresponding author.
E-mail addresses: pengyi@uestc.edu.cn (Y. Peng), guoxunwang@163.com (G. Wang), kougang@uestc.edu.cn, kougang@yahoo.com (G. Kou), yshi@gucas.ac.cn, yshi@mail.unomaha.edu (Y. Shi).

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2010.11.028

inspectors for fraud risk prediction, and more optimal allocation of investigative resources [8,9].

Driven by the strong business needs, many classification models have been proposed for credit and fraud risk prediction in the past few decades. These classification algorithms can be broadly summarized into statistical models, nonparametric statistical models, artificial intelligence methods, and mathematical programming methods. Examples of statistical models include discriminant analysis, logistic regression (e.g., [10]), and Bayesian classifier (e.g., [11]). Examples of nonparametric statistical models include recursive partitioning algorithm (RPA) (e.g., [6]), nearest neighbors (e.g., [12]), and goal programming (e.g., [13]). Proposed artificial intelligence methods include neural networks (e.g., [7,14,15]), decision tree (e.g., [16]), expert systems (e.g., [17]), and genetic algorithms (e.g., [18,19]). Examples of mathematical programming methods include support vector machines (e.g., [20–22]) and multi-criteria convex quadratic programming (e.g., [23–25]). Review works on classification algorithms for credit risk and fraud risk prediction are given in Srinivasan and Kim [26], Rosenberg and Gleit [8], Hand and Henley [27], Thomas [28], and Phua et al. [29].

Previous studies indicate that classifiers' performances in financial risk prediction may vary using different performance measures and under different circumstances. For example, in the application area of building credit scoring models, Desai et al. [30] and West [31] concluded that customized neural network methods outperformed linear discriminant analysis, whereas Yobas et al. [32] reported that the predictive performance of linear discriminant analysis was superior to neural networks. As the No Free Lunch (NFL) theorem states, "if algorithm A outperforms algorithm B on some cost functions, then loosely speaking there must exist exactly as many other functions where B outperforms A" [33]. There exists no single classification algorithm that could achieve the best performance for all measures.

Although there have been some studies evaluating the performance of classification methods for credit and fraud risk prediction, such as Viaene et al. [9] and Baesens et al. [34], they assess the performance of classifiers using only a couple of traditional measurements (e.g., the classification accuracy and the area under the receiver operating characteristic curve (AUC)), which have certain limitations [35]. How to provide a comprehensive assessment of classifiers and recommend an adequate classifier (or set of classifiers) is an important and understudied area in financial risk management.

In fact, the algorithm evaluation or algorithm selection problem is an active research area in many fields, such as artificial intelligence, operations research, and machine learning [36]. Rokach [37] suggests that the algorithm selection can be considered as a multiple criteria decision making (MCDM) problem and MCDM methods can be used to systematically choose the appropriate algorithm.

The main goal of this paper is to develop an approach to evaluate classification algorithms for financial risk prediction. It constructs a performance score to measure the performance of classification algorithms and introduces MCDM methods to rank the classifiers. An empirical study is designed to assess nine classification algorithms using five performance measures over seven real-life credit risk and fraud risk datasets from six countries. For each performance measure, a performance score is calculated for each selected classification algorithm. The classification algorithms are then ranked using three MCDM methods (i.e., TOPSIS, PROMETHEE, and VIKOR) based on the performance scores.

Another problem in financial risk detection is that the knowledge gap [38] between the results classification methods can provide and taking actions based on them remains large. The lack of interaction between industry practitioners and academic researchers makes it hard to discover financial risks or opportunities and hence weakens the value that classification methods may bring to financial risk detection. To deal with the knowledge gap problem, this paper combines the classification results, the knowledge discovery in database (KDD) process, and the concept of chance discovery to build a *knowledge-rich financial risk management process* in an attempt to increase the usefulness of classification results in financial risk prediction.

The contributions of this paper are twofold. The principal contribution of this paper is to introduce a performance metric and MCDM methods to study the performance of various classification algorithms for financial risk prediction. An empirical study of nine classification techniques on seven real-life financial risk datasets is conducted to validate the usefulness of the performance metric and MCDM methods in classification algorithm evaluation. These datasets include both public-domain and private data, and represent various aspects of financial risk (including credit approval, credit behavior, bankruptcy risk, and fraud risk). Second, this paper develops a conceptual process for knowledge-rich financial risk management. The process adapts concepts from KDD, the chance discovery process, and the CRISP-DM process model.

The rest of this paper is organized as follows: Section 2 describes the two-step approach designed for classification algorithm evaluation, including the construction of a performance metric and an overview of the three MCDM methods; Section 3 presents an empirical study that examines the proposed evaluation process using seven real-life financial risk datasets; Section 4 discusses the knowledge-rich financial risk management process and Section 5 concludes the paper.

2. Evaluation approach for classification algorithms

This paper develops a two-step process to evaluate classification algorithms for financial risk prediction. In the first step, a performance score is created for each selected classification algorithm. The second step applies three MCDM methods (i.e., TOPSIS, PROMETHEE, and VIKOR) to rank the selected classification algorithms using the performance scores as inputs. This section describes how the performance scores are calculated and gives an overview the three MCDM methods used in the study.

2.1. Performance score

There are a variety of measures for classification algorithms and these measures have been developed to evaluate very different things. Some studies have shown that the classification algorithm achieves the best performance according to a given measure on a dataset, may not be the best method using a different measure [35,39]. In addition, characteristics of datasets, such as size, class distribution, or noise, can affect the performance of classifiers. Hence, evaluating the performance of classification algorithms using one or two measures on one or two datasets often proves to be inadequate.

Based on these two considerations, this study constructs a performance metric that assesses the quality of classifiers using a set of measures on a collection of financial risk datasets in an attempt to give a comprehensive evaluation of classification algorithms. The basic idea of this performance metric is similar to ranking methods, which use experimental results generated by a set of classification algorithms on a set of datasets to rank those algorithms [40]. Specifically, it resembles the significant wins (SW) ranking method by conducting pairwise comparisons of classifiers using tests of statistical significance.

2.1.1. Selection of performance measures

Accuracy and error rates are important measures of classification algorithms in financial risk prediction. This work utilizes overall accuracy, precision, true positive rate, true negative rate,

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات