Knowledge-Based Systems 63 (2014) 59–67

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Financial ratio selection for business failure prediction using soft set theory

Wei Xu [a], Zhi Xiao [a,*,1], Xin Dang [b], Daoli Yang [a], Xianglei Yang [c]

[a] School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China
[b] Department of Mathematics, University of Mississippi, University 38677, USA
[c] Survey Office of the National Bureau of Statistics in Yongchuan, Chongqing 402160, PR China

ARTICLE INFO

Article history:
Received 20 September 2013
Received in revised form 13 March 2014
Accepted 15 March 2014
Available online 25 March 2014

Keywords:
Business failure prediction
Financial ratios
Logistic regression
Parameter reduction
Soft set theory

ABSTRACT

This paper presents a novel parameter reduction method guided by soft set theory (NSS) to select financial ratios for business failure prediction (BFP). The proposed method integrates statistical logistic regression into soft set decision theory, hence takes advantages of two approaches. The procedure is applied to real data sets from Chinese listed firms. From the financial analysis statement category set and the financial ratio set considered by the previous literatures, our proposed method selects nine significant financial ratios. Among them, four ratios are newly recognized as important variables for BFP. For comparison, principal component analysis, traditional soft set theory, and rough set theory are reduction methods included in the study. The predictive ability of the selected ratios by each reduction method along with the ratios commonly used in the prior literature is evaluated by three forecasting tools support vector machine, neural network, and logistic regression. The results demonstrate superior forecasting performance of the proposed method in terms of accuracy and stability.

1. Introduction

Business failure prediction (BFP) is one of the most essential problems in the field of economics and finance. It has been a subject of great interest to practitioners and researchers over decades. Being able to forecast potential failure provides an early warning system so that timely decisions are allowed to be made and appropriate adjustment in resource allocation can be taken place. There are three major tasks involved in the process of BFP, as shown in Fig. 1.

First, researchers need to determine their research objects [13]. Business failure prediction is a broad subject. Specific fields such as bank failure prediction, tourism failure prediction, small business failure prediction may take different approaches. Here we are interested in the failure prediction of Chinese listed firms from the Shenzhen Stock Exchange and Shanghai Stock Exchange.

Second, researchers need to select variables for BFP. Since the operational business environment changes quickly, BFP must be done in a timely fashion to provide early warnings. It has been shown that financial ratios have more forecasting power than other types of variables in such dynamic settings [20]. Hence we focus on financial ratio variables. Available ratios could not be used indiscriminately because some ratios could prove to be more powerful in their predictive ability than others. Predictive ability presents in two aspects. One is the forecasting accuracy (ACC). The other is the forecasting stability. If a forecasting system includes too many nonsignificant financial ratios, it will produce results low in forecasting accuracy and stability [10]. The goal of this paper is to select important financial ratios for BFP.

Third, researchers need to select forecasting models for BFP. The forecasting method, in particular the forecasting classifier used for a qualitative response, has a significant impact on the forecasting performance. Since the early empirical work on methods adopted by large USA banks, there has been a large number of literatures on the forecasting methods. Those methods include discriminant analysis [2], logistic regression (LR) [28], neural networks (NN) [3], probit method [35], rough set theory (RS) [11], support vector machine (SVM) [25], case-based reasoning [17], combination methods [7,20,34] and others. For a detailed review, one can refer to Dimitras et al. [10] and Zopounidis [36]. In this paper, we will use LR, SVM and NN to evaluate performance of variable selection methods rather than to study the selection of those forecasting models.

* Corresponding author. Tel.: +86 13808345199.
E-mail address: xiaozhicqu@163.com (Z. Xiao).

[1] Support by the National Science Foundation of China (Grant No. 71171209) and support by Major Consulting Research Project of Chinese Academy of Engineering (Grant No. 2012-ZD-12).

http://dx.doi.org/10.1016/j.knosys.2014.03.007
0950-7051/

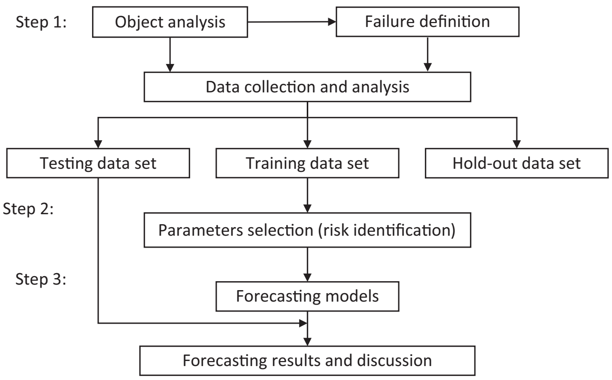

Fig. 1. The process of business failure prediction.

Differing from the development on prediction models, not much progresses have been made in variable selection for BFP. Beaver [4] selected six financial ratios including debt ratios. Altman [2] employed five financial ratios including sales to total assets. Deakin [8] made an attempt to identify variables useful in BFP. Ohlson [28] adopted nine different features. Recently, scholars [12,21,26,31,34] proposed additional financial ratios for BFP. Most popular financial ratios adopted in the prior literatures are summarized in Table 1. However, most of those financial ratios are selected either by the expert system method or by statistical approaches. Expert system relies heavily on users' knowledge and ability, which imposes difficulty to make it widely used. Statistical approaches have disadvantages for variable selection on their stringent model assumptions, which are often not met in practice. Small departures from the assumed model may make the statistical methods yielding unreliable even unacceptable results.

On the other hand, as we mentioned before, BFP must to be done in a dynamic setting. We shall include more factors or variables in the model such that information loss on the nature of firms in a dynamic operational environment is minimal. Inevitably we deal with BFP problem based on high-dimensional data. Soft set theory (SS), initiated by Molodtsov [27], has advantages to deal with high-dimension data sets. It also has been proved theoretically to be an effective tool for dimension reduction [9]. We expect SS a good performance on financial ratio selection for BFP. However, the prior literatures on SS are either purely theoretical or applied only on simple situations [24,6,16,37,14,30,23,1]. The available algorithms are rarely useful to be applied directly to the BFP problem. This motivates us to develop a novel method based on SS (NSS) to select financial ratios for BFP.

We first propose a general way to transfer the complex real-life data to 0–1 data frame so that SS or RS methods can be applied. Using LR, the importance of each variable is measured by its influence on predicting whether the firm will fail or not. A critical parameter involved in this step is determined optimally by a cross-validation procedure. Then the *uni-int* decision making on the SS is employed to obtain an optimal set of significant financial ratios. In such a way, our method utilizes the flexibility and efficiency of soft set theory and in the same time takes advantages of the statistical method without worrying about justifications of the underlying assumptions.

For comparison, principle component analysis (PCA) [15], traditional soft set (TSS) [16], rough set (RS) [29] are reduction methods included in the study of real data sets from Chinese listed firms along with evaluations of the financial ratio set proposed in previous literatures. TSS and RS use the same tabular representation data as NSS. Comparing with TSS, the *uni-int* decision making method is developed based on the redefined operations that exploits available tabular information more fully.

The remainder of this paper is organized as follows. Section 2 reviews the classical SS theory and introduces the proposed parameter reduction method. Section 3 describes the application to a real data set. In Section 4, we present the empirical results and compare performance of the proposed method with other methods. We conclude and discuss possible future work in Section 5.

Table 1
Financial ratios adopted in prior literatures for BFP.

No.	Financial ratio	No.	Financial ratio
x_1	Tax rates	x_2	Equity value per share
x_3	No-credit interval	x_4	Operating earnings per share
x_5	Equity growth ratio	x_6	EPS
x_7	Current ratio	x_8	Cash flow/total debt
x_9	Cash flow/total asset	x_{10}	Cash flow/sales
x_{11}	Debt ratio	x_{12}	Working capital/total asset
x_{13}	Market value equity/total debt	x_{14}	Current assets/total asset
x_{15}	Quick asset/total asset	x_{16}	Sales/total asset
x_{17}	Current debt/sales	x_{18}	Quick asset/sales
x_{19}	Working capital/sales	x_{20}	Net income/total asset
x_{21}	Retained earnings/total asset		
x_{22}	Earnings before interest and taxes/total asset		
x_{23}	Continuous 4 quarterly EPS (earning per share)		
x_{24}	Log (total asset/GNP price-level index)		
x_{25}	One if total liabilities exceeds total assets, zero otherwise		
x_{26}	One if net income was negative for the past 2 years, zero otherwise		
x_{27}	$(NI_t - NI_{t-1})/(\lvert NI_t\rvert + \lvert NI_{t-1}\rvert), NI_t$: Latest net income		

2. Soft set oriented parameter reduction methods

Originated by Molodtsov [27], soft set theory deals with uncertainty in a non-parametric manner. It has been extended to effectively select parameters [16]. In this section, we first review soft set theory, the *uni-int* decision making method and the traditional reduction method proposed, then propose our novel method.

2.1. Soft set theory

Let U be a non-empty initial universe of objects, E be a set of parameters to objects in U, $\mathcal{P}(U)$ be the power set of U and $A \subseteq E$.

Definition 2.1. A soft set F_A on U is defined by the set of ordered pairs

$$F_A = \{(x, f_A(x)) : x \in E, f_A(x) \in \mathcal{P}(U)\},$$

where $f_A : E \to \mathcal{P}(U)$ such that $f_A(x) = \emptyset$ if $x \notin A$.

Here f_A is called approximate function of the soft set F_A. The soft set F_A, in other words, is a parameterized family of subsets of the set U. Every set $f_A(x) (x \in E)$ from this family may be considered as the set of x-elements of the soft set F_A. Denote the collection of all soft sets on U as $\mathcal{S}(U)$.

Çağman and Engioğlu [5] redefined operations on soft sets. They defined product operations as binary operations of soft sets depending on an approximation function of two variables to exploit information of soft sets more fully. Then, they proposed the *uni-int* decision making method as follows.

Definition 2.2. If $F_A,\ F_B \in \mathcal{S}(U)$, then the $\wedge$-product (and-product) of two soft sets F_A and F_B is a soft set $F_A \wedge F_B$ defined by the approximation function

$$f_{A\wedge B} : E \times E \to \mathcal{P}(U), f_{A\wedge B}(x, y) = f_A(x) \cap f_B(y).$$

Denote $\wedge(U)$ as the set of all $\wedge$-products of the soft sets over U.