Expert Systems with Applications 39 (2012) 7041–7049

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

MCELCC*h*-FDP: Financial distress prediction with classifier ensembles based on firm life cycle and Choquet integral

Yu Cao *

Business School, Central South University, 932 Southern Lushan Road, Changsha 410083, China

ARTICLE INFO

Keywords:
Forecasting
Financial distress
Choquet integral
Firm life cycle
Ensemble

ABSTRACT

Financial distress prediction (FDP) has always been an important issue in the business and financial management. This research proposed a novel multiple classifier ensemble model based on firm life cycle and Choquet integral for FDP, named MCELCC*h*-FDP, as a new approach to tackle with financial distress. Empirical study based on Chinese listed companies' real data is conducted, and the results show that the proposed MCELCC*h*-FDP model has higher prediction accuracy than single classifiers. In order to verify the prediction capability of firm life cycle and Choquet integral in FDP model, comparative analysis is conducted. The experiment results indicate that the introduction of firm life cycle and Choquet integral in FDP can greatly enhance prediction accuracy.

1. Introduction

The problem of corporate financial distress has received much attention in the field of business and financial management in the past decades. It has serious adversary effect on which ranges from healthy development of enterprises, to the interests of credit institutions and securities investors, even to the economic security of countries which might trigger a worldwide economic crisis. The continuous acceleration of business globalization and fierce competition among the companies around the world increase the uncertainty and unpredictability of the market, which makes firm face a far more complex internal and external economic environment than before. The global financial crisis, started from 2008, has seriously hindered the pace of the development of world economy, which lifts the importance of financial distress prediction (FDP) to an unprecedented high position.

In the field of FDP, statistical techniques were firstly applied, such as single ratio analysis (Beaver, 1968), multiple discriminate analyses (MDA) (Altman, 1968), logistic regression (Logit) (Lau, 1987; Ohlson, 1980). After that, artificial intelligent ones, such as neural networks (NNs) (Salchengerger, Cinar, & Lash, 1992; Sharda & Wilson, 1996; Wilson & Sharda, 1994; Wu, Yang, & Liang, 2006; Xiao, Ye, Zhong, & Sun, 2009; Yang, Li, Ji, & Xu, 2001; Zhang, Hu, Patuwo, & Indro, 1999), decision tree (DT) (Sun & Li, 2008; Tae, Namisik, & Gunhee, 1999; Todorovski & Dzeroski, 2003), support vector machine (SVM) (Chen & Hsiao, 2008; Ding, Song, & Zen, 2008; Gestel, Baesens, & Johan, 2005; Hui & Sun, 2006; Min & Lee, 2005), case-based reasoning(CBR) (Jo, Han, & Lee, 1997; Li, Sun, & Sun, 2009; Sun & Hui, 2006; Yip & Deng, 2003) were also applied to financial distress prediction.

On the whole, pervious research on the topic of FDP mainly uses the single classifier to predict the financial conditions of enterprises, neglecting the advantages of the ensemble models of multiple classifiers. The performance of a single classifier depends on the characteristics of samples and also is uncertain because different classifiers might have different results of prediction, which makes its validity questionable.

Compared with single classifier, the combination classifier is expected to reduce the variance of estimated error and improve the whole recognition performance (Kumar & Ravi, 2007; Kuncheva, 2005; Ruta & Gabrys, 2005). So recently, financial distress prediction methods based on combination classifier become a rising trend in this field. Hua, Wang, Wang, Zhang, and Liang (2007) introduce a combination classifier by using SVM and Logistic regression for financial distress prediction. Hung and Chen (2009) integrated DT, NNS, and SVM to predict financial distress; Sun and Li (2009) put forward a financial distress prediction method based on serial combination of multiple classifiers; Cho, Hong, and Ha (2010) ensembled DT and CBR for financial distress prediction.

Fuzzy integral is a strong reasoning method under conditions of uncertainty. The fusion based on fuzzy integral can deal with the information from multiple classifiers that may agree or conflict with each other. And many empirical studies have proved that the fuzzy integral (Choquet and Sugeno integrals) performances well in multiple classifier fusion (Li et al., 2005; Verikas & Lipnickas, 2002). Recently, Narukawa and Torra (2007) reviewed the use of fuzzy measures and integrals (Choquet and Sugeno integrals) and showed their applications. They also considered their

* Tel.: +86 15873115668.
E-mail address: xiaoyu198549@126.com

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.043

computational cost showing that it is not much higher than the straightforward information fusion methods (e.g. the arithmetic mean and the weighted mean) usually used in most applications. In particular, they have shown that the computational cost of the integrals is exactly the same cost that characterizes the OWA operator has. This also provided the motivation for our research.

Therefore, this article employs Choquet fuzzy integral method, which is a tough powerful nonlinear reasoning method for uncertain environment, to integrate different single classifiers. This method could take the advantages of single classifiers together and make the information of different classifiers complementary, providing a great tool for early warning of corporate financial distress under complex environment.

Feature selection as the pre-processing step is one of the most important steps in the data mining process (Tsai, 2009; Yang & Olafsson, 2006; Yeh, Chi, & Hsu, 2010). However, previous research ignored an important variable: firm life cycle. The interior environment are significant different in different stages of firm life cycle and also the enterprises in different stages have different ways and abilities to perceive and respond to their external environment (Miller & Friesen, 1984). Many researchers have identified the firm life cycle as an important economic determinant of the value-relevance of reporting data (Anthony & Ramesh, 1992; Black, 1998; Jenkins, Kane, & Velury, 2004). Two enterprises in different stages will show different characters though they share the same ownership or management system (Agarwal, Sarkar, & Chambadi, 2002; Peteraf, 2003). So not taking the impact of firm life cycle into account, we will lose very important and effective predicative information.

This research aims to contribute to the financial distress prediction by introducing an integrated methodology to improve its prediction accuracy. We proposed a multiple classifier ensemble (MCE) model based on firm life cycle and Choquet integral for FDP, named MCELCC*h*-FDP, as a new approach to tackle with financial distress. More precisely, we firstly used firm life cycle to divide the sample into feature subsets, then the Choquet integral is applied to ensemble the results of several individual classifiers. To validate this methodology, an empirical study based on Chinese listed companies' real data is conducted. In order to verify the prediction capability of firm life cycle and Choquet integral in FDP model, comparative analysis is conducted.

This paper is organized as follows. In Section 2, we review the firm life cycle and Choquet integral. Section 3 describes the framework, methods and procedure of MCELCC*h*-FDP model. Section 4 explains the experimental design. Section 5 provides experiment results and discussion. Finally, some concluding remarks are drawn in Section 6.

2. A brief literature review

2.1. Firm life cycle

The enterprise life cycle theory suggests that enterprises, like life, experience a process from flourishing to decline and from birth to death. Haire (1959) declared that "life cycle" in biology can be used in enterprises and that the development of enterprises is also in line with the growth curve of biology. Gardner (1965) further pointed out that an enterprise also has a life cycle, but compared with people and other living things, its life cycle has its own uniqueness. After that, many scholars studied the enterprise life cycle, and the research achieves its prosperity in the 1970s, 1980s of the 20th century and emerges a new climax at the end of the 1990s.

The main contents of theoretical studies on this issue are the enterprise life cycle stages division and the differences of enterprises characteristics in each stage of the life cycle. Miller and Friesen (1984) empirical study the different characteristics in different stages, such as corporate strategy, the organizational structure, the environment and the decision-making styles. Adizes (1997) according to the flexibility and controllability, divided the enterprise into the incubation period and infancy period and other ten periods, and established a set of methods to diagnosis and treatment for enterprises and their culture. Chen (1995), according to the enterprise scale, divides the enterprise growth process into the incubation period, survival period, high-speed developing period, maturity, recession and the metamorphosis period of six stages, and described the characteristics of all stages of the enterprise, as well as discussed the forms and methods of the enterprise transformation in detail. Li (2000), on the basis of comparing the enterprise life cycle models respectively proposed by Adizes (1997), Chen (1995), analyzes the advantages and disadvantages of the current model, and then according to the enterprise sales, divided the enterprise life cycle into new-born, growth, mature and decline stages. Dickinson (2006) utilizes a firm's operating, investing, and financing cash flows in combination to assign life cycle stage.

2.2. Choquet integral

The philosophy of the Choquet integral was first introduced in capacity theory (Choquet, 1953) and used as a (fuzzy) integral with respect to a fuzzy measure proposed by Höhle (1982) and then rediscovered later by Murofushi and Sugeno (1989, 1991). Choquet integral is defined to integrate functions with respect to the fuzzy measures (Murofushi & Sugeno, 1989). As for the financial distress prediction, in the Choquet integral model, where the output of classifiers can be dependent, a fuzzy measure for classifiers is used to define a weight on each combination of classifier, thus making it possible to model the interaction existing among classifiers. Fuzzy measure was first introduced by Sugeno (1974), which make a monotonicity instead of additive property. The definition of fuzzy measures and Choquet integral are as follows (Murofushi & Sugeno, 1989).

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of classifiers, and let $P(X)$ denote the power set of X or set of all subsets of X.

Definition 1. Let X be a finite set of classifiers A discrete fuzzy measure on X is a set function $\mu : P(X) \to [0,1]$, satisfying the following conditions:

(1) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions).
(2) If $A, B \in p(X)$ and $A \subseteq B$ then $\mu(A) \leqslant \mu(B)$ (monotonicity).

$\mu(S)$ can be viewed as the grade of subjective importance of classifier set S. Thus, in addition to the usual weights on classifier taken separately, weights on any combination of classifiers are also defined.

Definition 2. Let μ be a fuzzy measure on $X = \{1, 2, \ldots, n\}$. The Choquet integral of function $f : X \to R$ with respect to μ is defined by

$$C_\mu(f) = \sum_{i=1}^{n} f_{(i)}[\mu(A_{(i)}) - \mu(A_{(i+1)})] \tag{1}$$

where (i) indicates a permutation on X such that $f_{(1)} \leqslant f_{(2)} \leqslant \cdots \leqslant f_{(n)}$. Also $A_{(i)} = \{i, \ldots, n\}$, $A_{(n+1)} = \phi$.

Since the fuzzy integral model does not need to assume independency of one classifier from another, it can be used in non-linear situations. The fuzzy integral of f with respect to μ gives the overall evaluation of an alternative. When the classifiers are independent, the fuzzy measure is additive, and the Choquet integral coincide with the weighted arithmetic mean method. That is,