Expert Systems with Applications 39 (2012) 4007–4012

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Decision tree-based technology credit scoring for start-up firms: Korean case

So Young Sohn *, Ji Won Kim

Department of Information & Industrial Engineering, Yonsei University, 134 Shinchon-dong, Seoul 120-749, Republic of Korea

ARTICLE INFO

Keywords:
Technology credit scoring model
Start-up
Small and medium enterprise
Decision tree

ABSTRACT

Various types of Technology Credit Guarantees (TCGs) have been issued to support technology development of start-up firms. Technology evaluation has become a critical part of TCG system. However, general technology credit scoring models have not been applied reflecting the special phenomena of start-ups, which are distinguishable from those of established firms. Furthermore, somewhat complicated approaches have been applied to existing models. We propose a rather simple decision tree-based technology credit scoring for start-ups which can serve as a-replacement for the complicated models currently used for general purposes. Our result is expected to provide valuable information to evaluator for start-up firms.

1. Introduction

Start-up firms are in need of an immense amount of funds for technology development (Sohn & Jeon, 2010). Many governments employ several support strategies to encourage start-up firms to accelerate economic growth and to decrease unemployment. Among many support schemes, the Technology Credit Guarantee (TCG) system is seen as one of the most important instruments (Kang & Heshmati, 2007). The TCG is a system that government gives warranty to private financial institutions such as banks which give loans to the SMEs that received the guarantee from TCG agency. This kind of system removes the risks of lending to start-up firms (Oh, Lee, Heshmati, & Choi, 2009). In order to enhance the technology competitive power of start-ups, continuous support based on the evaluation of applicant start-ups is required. Therefore, the importance of professional technology assessment cannot be over emphasized. Further, evaluation must achieve an appropriate level of precision so as to distinguish successful start-ups from those which would default the loan subsequent to funding.

Despite its importance, many existing technology credit scoring systems have not considered using a separate model for start-ups from that for established firms. In addition, existing technology credit scoring models are complicated, and it is not easy to understand the mechanism involved in.

Typically, a technology scorecard had been used to CEO's knowledge about technology and technology superiority, marketability and profitability with pre-assigned weight (Henderson & Cockburn, 1994; Korea Technology Transfer Association, 2005; Moon & Sohn, 2008a; Sohn, Kim, & Moon, 2005; Walsh & Linton, 2002). However high default rate of funded SMEs based on such scorecard has been reported (Boocock & Shariff, 1996; Cowling & Mitchell, 2003; Sohn & Kim, 2007). Therefore, accurate technology evaluation is crucial. The use of inadequate evaluation models could jeopardize the entire funding process, causing critical losses. As a preventative effort, many researchers have been studying how to select firms that will not default on their loans subsequent to technology funding.

Recently developed technology evaluation scorecards are as follows. Sohn, Moon, and Kim (2005) provided an improved version of the technology evaluation model by eliminating multicollinearity among the evaluation attributes. Kim and Sohn (2007) proposed a bivariate probit model to determine whether the rejected inference technique was more useful than existing models, which use only the history of accepted applicants. Sohn and Kim (2007) proposed a random effects logistic regression model for default prediction by considering not only the financial and non-financial characteristics of the SMEs, but also an element of uncertainty, which could not be explained by such characteristics. Sohn, Kim, and Moon (2007) further applied a structural equation model (SEM) to predict the financial performance of funded technology based SMEs by considering the relationship among various variables. Moon and Sohn (2008a) proposed a new CBR system for predicting multi-period financial performances for technology-based SMEs after funding and, Moon and Sohn (2008b) proposed a new technology-scoring model to reflect the phenomenon of total perception scoring, which occurs often in many technology evaluations. Jeon and Sohn (2008) proposed an expected loss model for the TCG fund, a model which takes into account defaults for various types of competing risks such as delay, bad credit, and bad checks, among others. Most recently, Moon and Sohn (2010)

* Corresponding author. Tel.: +82 2 2123 4014; fax: +82 2 364 7807.
E-mail address: sohns@yonsei.ac.kr (S.Y. Sohn).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.09.075

proposed new technology evaluation models, which consider not only technology-related variables, but also environmental variables such as enterprise characteristics and economic conditions.

However, in these studies, start-ups have not been separately considered for model development. Because start-up firms would behave differently from established firms, it is necessary build technology credit scoring model. Many previous studies also have focused on developing various evaluation methods, ranging from intuitive judgment to complex options models (Mitchell & Hamilton, 1996). Complex models have some drawbacks such as difficulty of delivering the meaning.

The decision tree method (classification tree) is known to be simple and appeal to lay practitioners who do not have heavy statistical backgrounds. It can also easily examine the interaction effects of predictor variables.

The main purpose of this study is to propose a decision tree based technology credit scoring model for start-up firms and identify significant predictors on technology loan default of start-ups. It is expected that the proposed model can be applied to a wide range of technology investment-related decision-making procedures.

This paper is organized as follows. In Section 2, we provide a literature review as well as the relevant theories related to credit scoring models and the data mining analysis method. In Section 3, we provide the empirical data and input variables used in this paper. In Section 4, new scoring models are proposed for start-up firms, and in Section 5, we discuss our study results and suggest areas for future research.

2. Literature review

2.1. Credit scoring model

The objective of quantitative credit scoring models is to assign credit applicants to one of two groups: a "good credit" group that is likely to repay their financial obligation or a "bad credit" group that should be denied credit because of a high likelihood of defaulting on their financial obligation (West, 2000).

Studies of credit scoring models have been performed for a long time. The linear discriminant model (Chen & Huang, 2003) was one of the first credit scoring models, and it is still commonly used today. However, linear discriminant analysis (LDA) for credit scoring has been challenged due to the categorical nature of the credit data and the fact that covariance matrices of the accepted and rejected classes are likely to be unequal (West, 2000). Practitioners and researchers have also applied statistical techniques to develop more sophisticated models for credit scoring, which involve logistic regression analysis (LRA) (Chen & Huang, 2003), k nearest neighbor (KNN) (Henley & Hand, 2007), neural network (NN) models (Desai, Crook, & Overstreet, 1996; Malhotra & Malhotra, 2002; West, 2000), genetic programming models (Ong, Huang, & Tzeng, 2005), and decision tree models (Bensic, Sarlija, & Zekic-susac, 2005).

Many researchers have indicated that the best model can be created through the comparison of various methods. Thomas (2000) and West (2000) indicated that both LDA and LRA can be used when the relationship between variables is linear. Hence, both methods may not reflect potential nonlinear relationship.

Over the last two decades, a number of studies has been conducted to compare NN model to other models such as LDA, decision tree and k-nearest. Coats and Fant (1993) suggested that the NN model was more accurate than LDA, especially, for predicting the default rate of firms in financial distress. Thus, NNs are the most popular tool used for credit scoring and it has been reported that their accuracy is superior to that of traditional statistical methods in dealing with credit scoring problems, especially with regard to nonlinear patterns (Desai et al., 1996; Mahlhotra & Malhotra, 2003). In contrast, NNs have been criticized for their poor performance when incorporating irrelevant attributes or small data sets (Feraud & Clerot, 2002; Nath, Rajagopalan, & Ryker, 1996).

In addition, Galindo and Tamayo (2000) performed a comparative analysis of CART decision-tree models, NNs, the k-nearest neighbor and probit algorithms on a mortgage loan dataset. They revealed that CART decision-tree models provided the best estimations for default.

2.2. Important variables

Performance of start-up firms has been investigated in terms of the relationship between growth and size of start-up firms (Audretsch, Santarelli, & Vivarelli, 1999). Recent researches have indicated that firm size as well as other firm-specific characteristics and managerial characteristics play important roles in growth of start-up firms. In particular, a talent for CEO of start-up firms has been shown to positively affect the growth of the firm (Almus, 2002; Wasilczuk, 2000). Also prior managerial experience of the founder has been demonstrated to positively affect firm growth and survival higher than those without prior experience (Storey, 1994).

Lynskey (2004) considered six firm characteristics for examining the determinants of innovative activity in Japanese technology-based start-up firms: 'technological capability', 'the availability of internal funds', 'the effectiveness of venture capital funding', 'joint research with universities', 'geographic location' and 'the age of the firm'. In addition to these firm characteristics, the author also considered six managerial characteristics: 'the CEO's educational background', 'the CEO's prior experience', 'whether the current CEO is the founding entrepreneur of the firm', 'the CEO's age', 'the CEO's involvement in a network of other researchers' and 'the CEO's experience of having managed other firms'.

Harada (2004) considered the firm characteristics in order to investigate the effects of productivity: 'type of industry', 'sales (million yen per month)', 'labor and capital input' and 'the age of the firm'. Also, managerial characteristics considered are: 'the CEO's age', 'previous occupational status', 'related business experience', 'the CEO's gender'. Benzing, Chu, and Kara (2009) considered firm characteristics for examining the determinants of motivation, success, and problems in Turkish start-up firms: 'How the business was established', 'average age of firms', 'average number of employees' and 'type of business'. They also considered managerial characteristics such as 'the CEO's gender', 'the CEO's age' and 'level of education'.

As summarized, many studies have considered the firm-specific characteristics and managerial characteristics for examining the determinants variables in various aspects such as growth, innovative activity, productivity, motivation, success and problems in start-up firms. Although, their purposes had some difference, variables considered are closely related to the success in start-up firms. Because increasing growth and productivity of start-up firms depend on innovation activity, also they contribute to success of start-up firms (Heunks, 1996).

In order to help success of start-up firms with a high degree of potential in technology, TCG agencies evaluate the firms which obtain high scores by a technology scorecard in Korea. The technology scorecard used in the current TCG program is as follows: the ability of a Chief Executive Officer (CEO), the level of technology, marketability of technology, and potential or realistic profitability of technology (Moon & Sohn, 2010; Sohn & Jeon, 2010; Sohn, Moon, et al., 2005). These attributes include the technology characteristics as well as the managerial characteristics. Recently, Moon and Sohn (2010) considered the firm characteristics as well as