Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 36 (2009) 3199–3207

Expert Systems with Applications

www.elsevier.com/locate/eswa

A genetic programming model for bankruptcy prediction: Empirical evidence from Iran

Hossein Etemadi, Ali Asghar Anvary Rostamy *, Hassan Farajzadeh Dehkordi

Tarbiat Modares University (TMU), No. 12, Shahid Rahnama Alley, Saidi Street, Lavasani Avenue, P.O. Box 19546, Tehran, Iran

Abstract

Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors/creditors, borrowing firms, and governments alike. Timely identification of firms' impending failure is indeed desirable. By this time, several methods have been used for predicting bankruptcy but some of them suffer from underlying shortcomings. In recent years, *Genetic Programming* (GP) has reached great attention in academic and empirical fields for efficient solving high complex problems. GP is a technique for programming computers by means of natural selection. It is a variant of the genetic algorithm, which is based on the concept of adaptive survival in natural organisms. In this study, we investigated application of GP for bankruptcy prediction modeling. GP was applied to classify 144 bankrupt and non-bankrupt Iranian firms listed in Tehran stock exchange (TSE). Then a multiple discriminant analysis (MDA) was used to benchmarking GP model. Genetic model achieved 94% and 90% accuracy rates in training and holdout samples, respectively; while MDA model achieved only 77% and 73% accuracy rates in training and holdout samples, respectively. McNemar test showed that GP approach outperforms MDA to the problem of corporate bankruptcy prediction.

Keywords: Bankruptcy prediction; Financial ratios; Genetic Programming; Multiple discriminant analysis; Iranian companies

1. Introduction

Corporate bankruptcy is a very important economic phenomenon. The health and success of the firms are of widespread concern to policy makers, industry participants, investors, and managers (O'Leary, 1998). It also is a problem that affects the economy of every country. The number of failing firms is important for the economy of a country and it can be considered as an index of the development and robustness of the economy (Zopounidis & Dimitras, 1998). The high individual, economic, and social costs encountered in corporate failures or bankruptcies have spurred searches for better understanding and prediction capability (McKee & Lensberg, 2002).

Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors/creditors, borrowing firms, and governments alike. Timely identification of firms' impending failure is indeed desirable (Jones, 1987). By this time, several methods have been used for predicting bankruptcy. Early research focused primarily on univariate models such as individual financial ratios. Among these studies Beaver (1966) is more noticeable than the others. He introduced a univariate technique for the classification of firms in two groups using some financial ratios. The ratios were used individually and a cut-off score was calculated for each ratio on the basis of minimizing misclassification. The univariate methods were later criticized, in spite of its considerable results, because of the correlation among ratios and providing different signals for a firm by ratios (Dimitras, Zanakis, & Zopounidis, 1996).

Later research turned to multivariate models. Researchers found that corporate bankruptcy can be affected by many different factors at the same time. Altman (1968) introduced a multivariable technique, multiple discriminant

* Corresponding author. Tel.: +98 21 22291235; fax: +98 21 22291279.
E-mail addresses: etemadih@modares.ac.ir (H. Etemadi), Anvary@modares.ac.ir, amin_armin@yahoo.com (A.A. Anvary Rostamy), farajzadeh@modares.ac.ir (H.F. Dehkordi).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2008.01.012

analysis (MDA), for failure prediction. Because this study made use of more than one variable for bankruptcy prediction and applied an advanced statistical technique for determining the relationship among predictor variables, it was of much interest.

MDA provides good predictions but suffers from some limitations. Hence, variety methods introduced to overcome MDA shortcomings and improving accuracy. These methods can be grouped in two categories: statistical and artificial intelligence models. First group consists of Logit (Foreman, 2002; Ohlson, 1980; Zavgren, 1985), Probit (Casey, Mc Gee, & Stinkey, 1986; Theodossiou, 1991), Linear Probability (Stone & Rasp, 1991; Vranas, 1992) Cumulative Sums (Kahya & Theodossiou, 1999), and etc. Neural Networks (Altman, Marco, & Varetto, 1994; Coats & Fant, 1993; Jo, Han, & Lee 1997), Genetic Algorithms (Shin & Lee, 2002; Varetto, 1998), Case Based Reasoning (Park & Han, 2002), Rough Sets (Dimitras, Slowinski, Susmaga, & Zopounidis 1999; McKee & Lensberg, 2002), Support Vector Machine (Min & Lee, 2005), and etc, constitute second group. Some of these models have high predictive accuracy levels but because of absence bankruptcy theory, attempts to establish a generally accepted model for bankruptcy prediction are not successful. Some studies have provided comprehensive surveys on bankruptcy prediction methods such as Dimitras et al. (1996), Jones (1987), and Kumar and Ravi 2007.

A common approach to bankruptcy prediction is to review the literature to identify a large set of potential predictive financial and/or non-financial variables and then develop a reduced set of variables, through some combination of judgmental and mathematical analysis that will predict bankruptcy (Lensberg, Eilifsen, & McKee 2006). In this study we implemented such approach for variables selection stage. After this, a relatively new technique for bankruptcy prediction, Genetic programming, constructed an accurate classification model for bankruptcy prediction. This model was benchmarked with the MDA, the most common used classification model for this subject. In the rest of the paper, first, we discuss about GP and MDA, two techniques were used for bankruptcy prediction modeling. In the section 4 we explain variable selection process and after that models development, empirical results and conclusion will be discussed.

2. Genetic programming

Genetic programming (GP) is a search methodology belonging to the family of evolutionary computation (EC). GP can be considered an extension of Genetic algorithms (GAs) (Koza, 1992). GAs is stochastic search techniques that can search large and complicated spaces on the ideas from natural genetics and evolutionary principle (Goldberg, 1989; Holland, 1975). They have been demonstrated to be effective and robust in searching very large spaces in a wide range of applications (Colin, 1994; Shin & Han, 1999). GAs has been applied in wide range financial fields such as trading system (Deboeck, 1994), stock selection (Mahfoud & Mani, 1995), bankruptcy prediction (Shin & Lee, 2002), and etc.

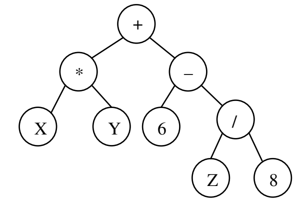

Fig. 1. Tree representation of the program (expression): $(X^*Y) + 6 - (Z/8)$.

GP is basically a GA applied to a population of computer programs (CP). While a GA usually operates on (coded) strings of numbers, a GP has to operate on CP. GP allows, in comparison with GA, the optimization of much more complicated structures and can therefore be applied to a greater diversity of problems (Sette & Boullart, 2001). Koza (1992) has extensively described GP. Because bankruptcy prediction can be considered as a classification problem, in continues, we offer a necessary description of GP with emphasis on its application in classification role.

In general, genetic programming models were inspired by the Darwinian theory of evolution. According to the most common implementations, a population of candidate solutions is maintained, and after a generation is accomplished, the population is expected fitted better for a given problem. Genetic programming uses tree-like individuals that can represent mathematical expressions, making valuable the application of GP in symbolic regression problems. Such a GP individual is shown in Fig. 1.

Three genetic operators are mostly used in these algorithms: reproduction, crossover, and mutation.

2.1. Reproduction

The reproduction operator simply chooses an individual in the current population and copies it without changes into the new population.

2.2. Crossover

Two parent individuals are selected and a subtree is picked on each one. Then crossover swaps the nodes and their relative sub-trees from one parent to the other. This operator must ensure the respect of the depth limits. If a condition is violated the too-large offspring is simply replaced by one of the parents. There are other parameters that specify the frequency with which internal or external points are selected as crossover points. Figs. 2 and 3 show an example of crossover operator.

2.3. Mutation

The mutation operator can be applied to either a function node or a terminal node. A node in the tree is