Expert Systems with Applications 39 (2012) 10140–10152

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Enhanced default risk models with SVM+

Bernardete Ribeiro [a,*], Catarina Silva [a,b], Ning Chen [c], Armando Vieira [c], João Carvalho das Neves [d]

[a] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
[b] ESTG, School of Technology and Management, Polytechinc Institute of Leiria, Portugal
[c] GECAD, Polytechnic Institute of Porto, Porto, Portugal
[d] ISEG-School of Economics and Management, Technical University of Lisbon, Portugal

ARTICLE INFO

Keywords:
Bankruptcy prediction
Default risk model
Support vector machines
Multi-task learning

ABSTRACT

Default risk models have lately raised a great interest due to the recent world economic crisis. In spite of many advanced techniques that have extensively been proposed, no comprehensive method incorporating a holistic perspective has hitherto been considered. Thus, the existing models for bankruptcy prediction lack the whole coverage of contextual knowledge which may prevent the decision makers such as investors and financial analysts to take the right decisions. Recently, SVM+ provides a formal way to incorporate additional information (not only training data) onto the learning models improving generalization. In financial settings examples of such non-financial (though relevant) information are marketing reports, competitors landscape, economic environment, customers screening, industry trends, etc. By exploiting additional information able to improve classical inductive learning we propose a prediction model where data is naturally separated into several structured groups clustered by the size and annual turnover of the firms. Experimental results in the setting of a heterogeneous data set of French companies demonstrated that the proposed default risk model showed better predictability performance than the baseline SVM and multi-task learning with SVM.

1. Introduction

The world is observing one of the most severe financial crisis ever observed. While in the past the SME (Small and Medium Enterprises) and micro companies had higher propensity of bankruptcies in the recent past an increasing number of large bankruptcies is systematically announced and the financial distress of all type of firms across all industries is in danger. Aided by technology and lower barriers to international capital flows, these crisis have shown a greater tendency to spread to markets through out the world, severely affecting the global economic activity. At the heart of the present global recession is an inappropriate evaluation of credit risk and most of governments were forced to implement rescue plans for the banking systems, including the Portuguese Government.

Given the devastating effects of the financial distress of firms, it is urgent that management and regulators are able to anticipate this kind of issues. Although credit loss is a normal cost of doing business in the case of banks, the excess of losses can be sufficiently severe to threaten their own existence. The evidence in the present situation is the need for banks to revise their models which evaluate the risk of each loan and the default rates of portfolio's loans. International rating agencies, like Moodys and Standard & Poor's, are also criticized for their models and inadequacy of quantifying and predicting the risk of insolvency of firms and banks. Moreover, these local and international rating agencies tend to analyze the risk of large companies while the financial system and banks, in particular, also need models for analyzing the risk of SME's. Banks have their own internal rating models to quantify the risk of loans but they are still in their infancy and rely on relatively simple mathematical methods with inadequate assumptions.

As a consequence, there is an ever-increasing need for fast automated recognition systems for bankruptcy prediction. The extensive recent literature shows that at the core of the business failure problem is the asymmetric information between banks and firms. Additionally, the development of analytical tools to determine which financial information is more relevant to predict financial distress has gained popularity along with the design of early warning systems that predict bankruptcy (Pena, Martinez, & Abudu, 2009).

The health of firm in a highly competitive business environment is dependent upon its capability to yield profitability and financial solvency. This means that a firm becomes unhealthy when it loses its competence to maintain profitability and financial solvency (Wu, 2010). Business failure is not only common with new startups but also with listed companies, and it can easily happen to firms of any and all sizes.

* Corresponding author. Tel.: +351 239 790 000; fax: +351 239 701 266.
E-mail addresses: bribeiro@dei.uc.pt (B. Ribeiro), catarina@dei.uc.pt (C. Silva), ningchen74@gmail.com (N. Chen), asv@isep.ipp.pt (A. Vieira), jcn@iseg.utl.pt (J. Carvalho das Neves).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.142

In Portugal, according to the Annual Survey of Insolvency and Constitutions Company Coface,[1] 4519 companies were declared insolvent in 2011, 1192 more than in the previous year (for the same period) representing an increase by 35.8%, while the number of start-ups raised only by 13.2%.

However, for this study we used a large database of French companies. This database is very detailed containing information on a wide set of financial ratios spawning over a period of several years. It contains up to three thousands distressed companies and about sixty thousand healthy ones. The financial Coface Data set (French credit risk provider) is strongly heterogenous with regards to the type of companies and their financial statuses. A great deal of research has been pursued disregarding this aspect. In this paper we focus on the improvement of financial distress decision-making by including structured information into heterogeneous groups of companies. We investigate the use of an advanced SVM+ approach by Vapnik and Vashist (2009) whose rationale is to take into account additional information in a financial setting of French companies. The firms are grouped by their category of large, medium and small sizes thus clustered by the number of employees and annual turnover. The properties resulting from these well-defined profiles unveil decisive correlations among firms. Our study shows that (a) SVM+ not only yields better prediction model than baseline SVM but also a better model as compared to a similar approach of Multi-Task Learning (MTL) and (b) the most salient data parameters per group both in the kernel decision space and in kernel correcting space are optimized, whereby the parameters and parameter ranges that shape the various firms profiles are exposed. The classification results demonstrate the prediction capability and robustness of the proposed method.

The rest of the paper is organized as follows. Next section describes relevant background knowledge on bankruptcy prediction and related work. In Section 3 we introduce SVM+ algorithm and give its mathematical foundations. In Section 4 the proposed approach is described, and further details on comparable settings are schematically illustrated for model comparison. In Section 5 we describe the database with information on healthy (and distressed) firms appropriately labeled for bankruptcy prediction model design in a case study of the French Market. We also describe the preprocessing preparation phase and the performance metrics. The experimental results including model and parameter selection, discussion and statistical hypothesis tests are performed in Section 6. Finally, in Section 7 we present the conclusions and draw further lines of work.

2. Related work

The prediction of bankruptcy is a well-researched area in finance analysis and attracts much interest to creditors, auditors and bank managers. The accurate prediction and early warning of bankrupt events has critical impact on economics to control the risk associated with wrong decisions, decrease the cost of monitoring solvency, and shorten the time of credit assessment. Bankruptcy prediction solves the important decision-making problem that identifies the potential bankrupt company based on the analysis of historical finance characteristics.

The problem is stated as follows: given a set of parameters (mainly of financial nature) that describe the situation of a company over a given period, predict the probability that the company may become bankrupt during the following year. During the years, this problem has been approached by various methods ranging from statistics to machine learning. A review of the topic of bankruptcy prediction with emphasis on neural networks (NN) is given in Atiya (2001). Also, in Ravi Kumar and Ravi (2007) there is a broad coverage of a wide range of other intelligent techniques such as fuzzy set theory (FS), decision trees (DT), rough sets, case-based reasoning (CBR), and support vector machines (SVM). More recently, in Verikas, Kalsyte, Bacauskiene, and Gelzinis (2010) a comprehensive review of hybrid and ensemble-based soft computing techniques applied to bankruptcy prediction is presented. Despite the numerous papers dealing with the problem, it is often difficult to compare the techniques due to possible differences in assumptions, data sets, time periods and failure definitions.

Fig. 1 illustrates the general framework of bankruptcy prediction using machine learning methods, composed of feature selection, dimensionality reduction using linear (e.g., PCA/KPCA) or nonlinear (e.g., ISOMAP/NMF) projection methods, followed by a machine learning (through NN, SVM etc.) process. It may be noticed that the additional information is helpful to attain better generalization of default risk models, which can be, for example in SVM+, the structured information in the data.

2.1. Neural networks

Neural Networks (NNs) are particularly suited for predicting the bankrupt probability, thus they are a strategic choice among other methods. Likewise, their properties make them often used in financial applications because of their excellent performances of treating non-linear data with self-learning capability (Fu-yuan, 2008). As a competitive learning neural network, self-organizing map (SOM) is used to determine the credit class through a visual exploration (Merkevicius, Garsva, & Simutis, 2004). Learning Vector Quantization (LVQ) is a supervised variant of SOM useful for non-linear separation problems (Kohonen, 2001). The network is composed of two levels, in which the input level is fully connected with the output level. The modeling technique is based on the neurons representing prototype vectors and the nearest neighbor classification rule. The goal of learning is to determine the weights that best represent the classes. LVQ has been employed to detect the distressed companies with satisfactory performance as in Chen and Vieira (2009) and Boyacioglu, Kara, and Baykan (2009). Carvalho das Neves and Vieira (2006) show that an enhanced version of Hidden Layer Learning Vector Quantization can enhance the performance of a multi-layer perceptron (MLP). In recent research efforts, combined techniques have been studied to optimize the learning models by evolutionary algorithms, in particular, Genetic algorithm (GA) is used by Sai, Zhong, and Qu (2007) and Huang, Kuo, and Yeh (2008) to optimize the parameters and connected weights of back-propagation neural networks.

2.2. Support vector machines

Support Vector Machines (SVMs) transform the input vectors nonlinearly into a high-dimensional feature space through a kernel function so that the data can be separated by linear models. In the literature there is an endless list of articles with SVM approaches. Min and Lee (2005) applied a grid-search technique to find out the optimal parameter settings of both polynomial and RBF kernel functions and showed that SVM outperforms techniques such as multiple discriminant analysis (MDA), logistic regression analysis (Logit), and three-layer fully connected back-propagation neural networks (BPNs).

More often, evolutionary algorithms including genetic algorithm, annealing simulation, particle swarm optimization, ant colony optimization are widely used in hybrid classification to significantly advance both the SVM single prediction model and feature selection (Lin, Shiue, Chen, & Cheng, 2009). Research efforts have been directed to combine SVM with other soft computing

[1] http://www.cofaceportugal.pt/CofacePortal/PT/pt_PT/pages/home/noticias/Estudos.