Expert Systems with Applications 41 (2014) 7235–7247

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Credit rating with a monotonicity-constrained support vector machine model

Chih-Chuan Chen [a,c], Sheng-Tun Li [a,b,*]

[a] Department of Industrial and Information Management, National Cheng Kung University, Taiwan, ROC
[b] Institute of Information Management, National Cheng Kung University, Taiwan, ROC
[c] Department of Leisure and Information Management, Taiwan Shoufu University, Taiwan, ROC

ARTICLE INFO

Article history:
Available online 12 June 2014

Keywords:
Credit rating
SVM
Monotonicity constraint
Prior domain knowledge
Data mining

ABSTRACT

Deciding whether borrowers can fulfill their obligations is a major issue for financial institutions, and while various credit rating models have been developed to help achieve this, they cannot reflect the domain knowledge of human experts. This paper proposes a new rating model based on a support vector machine with monotonicity constraints derived from the prior knowledge of financial experts. Experiments conducted on real-world data sets show that the proposed method, not only data driven but also domain knowledge oriented, can help correct the loss of monotonicity in data occurring during the collecting process, and performs better than the conventional counterpart.

1. Introduction

The main source of income for a depository financial institution is the interest paid on loans. Therefore, deciding whether borrowers are able to fulfill their obligations or not is a major issue when granting loans. Credit ratings are based on a collection of decision models and corresponding techniques that can help financial institutions make decisions with regard to granting consumer credit, answering questions such as who will get credit, how much credit they will get, and what strategies or policies will be adopted to control risk and increase profitability.

With the implementation of the Basel II requirements in 2004, this issue has become even more important for financial institutions. The Basel Committee, comprised of central banks and banking business representatives from various countries, formulates broad supervisory standards and guidelines for banks to implement. Due to changes in the banking business, risk management practices, supervisory approaches, and financial markets, in 2004 the Committee published the International Convergence of Capital Measurement and Capital Standards: a Revised Framework ("Basel Committee on Banking Supervision, Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework, Bank for International Settlements, 2004"). This new capital adequacy framework, also known as Basel II, requires financial institutions to have more flexibility and risk sensitivity. In addition, it allows banks to measure credit and operational risk using an internal rating based (IRB) approach in order to determine capital levels. Therefore, financial institutions worldwide are developing implementation strategies for risk control, as well as risk sensitivity analysis, to ensure a significant and sustainable reduction to credit risk, in compliance with Basel II (Dietch & Petey, 2002; Rosch, 2005). Of the various methods that have been suggested, a robust and systematic credit rating model or system is one of the most effective candidates to achieving an IRB approach (Dietch & Petey, 2002).

Many recent studies have developed systems to evaluate banks' business loans (e.g. du Jardin & Severin, 2011; Hájek, 2011; Min & Lee, 2005; Pendharkar, 2005; Rosch, 2005; Tsaih, Liu, Liu, & Lien, 2004), because the related data is more readily accessible to the public. However, the rising national income in some countries has resulted in the widespread use of consumer loans and a more competitive market for such products, although there has also been an increase in consumer bankruptcies (Malhotra & Malhotra, 2002; Malhotra & Malhotra, 2003). In loan management, it is thus vital to assess the creditworthiness of a borrower using credit rating systems, and current systems can be categorized as following one of two major approaches: specialized judgment and statistical modeling. The former relies on the expertise and tacit knowledge of human specialists, which can lead to fatigue, misjudgments and slow responses by financial experts, since the assessment process is usually time-consuming and laborious. In contrast, the latter approach can avoid these problems, and also

* Corresponding author at: No. 1, Ta-Hsueh Road, Tainan 701, Taiwan, ROC. Tel.: +886 6 2757575 53126; fax: +886 6 2362162.
E-mail address: stli@mail.ncku.edu.tw (S.-T. Li).

http://dx.doi.org/10.1016/j.eswa.2014.05.035
0957-4174/

reduce overheads due to its objectivity and consistency. In the past few decades, in a challenge to traditional statistical methods, artificial neural network (ANN) based techniques, such as multilayer perceptrons (MLPs) and radial basis function networks (RBFNs), have been used to extract tacit knowledge from vast amounts of data and domain experts, and have been shown to outperform conventional methods with regard to classification accuracy (Gately, 1996). The ANN scheme has been widely applied to the area of financial analysis, and the work of Vellido, Lisboa, and Vaughan (1999), which conducted a comprehensive survey of 38 articles focusing on bankruptcy prediction and credit evaluation, confirmed that this approach has shown real, practical benefits in this field. However, many parameters, such as network topology, learning rate and training methods, have to be fine-tuned before ANNs can be deployed successfully (Wallrafen, Protzel, & Popp, 1996). Evolutionary computing techniques, such as genetic algorithms and genetic programming, have been widely applied to variable selection and parameter optimization for further classification with prediction models, such as ANNs and Support Vector Machines (SVM) (Marqués, García, & Sánchez, 2013).

While ANNs have some weaknesses, such as a tendency to become trapped in a local optimum, overfitting, and requiring a large amount of time for learning (Arisawa & Watada, 1994), these have been overcome by the application of SVMs. SVMs, pioneered by Vapnik in 1995, are essentially ANNs based on statistical learning (Vapnik, 1995; Vapnik, 1998), and have attracted much attention from diverse research communities due to their outstanding performance with regard to solving classification problems and improving the generalizability of ANNs (Burges, 1998; Cristianini & Shawe-Taylor, 2000). Unlike ANNs, which minimize empirical risk, SVMs are designed to minimize structural risk by minimizing the upper bound of the generalization error rather than the training error, thus solving the problem of overfitting. Compared to ANNs, the other significant property of SVMs is that the task of training them can be mapped to a uniquely solvable linearly constrained quadratic programming problem, which produces a solution that is always unique and globally optimal. SVMs have been widely applied to many fields in the past few years, such as text categorization, handwriting recognition, speaker verification, bioinformatics, and face detection, as well as in finance, such as analyzing the risk of corporate distress (Fan & Palaniswami, 2000; Huang, Nakamori, & Wang, 2005; Huang, Chen, Hsu, Chen, & Wu, 2004; Min & Lee, 2005; Shin, Lee, & Kim, 2005; Van Gestel, Baesens, Garcia, & Van Dijcke, 2003) and undertaking consumer loan evaluations (Hsieh & Hung, 2010; Huang, Chen, & Wang, 2007; Li, Shiue, & Huang, 2006).

In domains such as finance there is a wealth of human expertise that has been developed by professionals over many years. Although data mining techniques can automatically extract patterns from data and then offer recommendations, they may not be able to reflect this expert knowledge, especially tacit knowledge that cannot be easily formalized. It is thus important to combine data and domain knowledge in order to construct decision support systems, which are not only data driven but also domain knowledge oriented.

Many classification applications utilize a priori domain knowledge to the extent that, all other things being equal, an increase in an input variable should not lead to a decrease (or increase) in class labels. For example, if loan applicants A and B have the same attribute values, except applicant A has a higher income than applicant B, then it would be surprising if applicant B got the loan while applicant A did not. Examples of other application domains in which we can have this type of knowledge are legal support systems, medicine (e.g. smoking increases the probability of vascular diseases), operations research and economics (e.g. house prices increase or decrease based on the location). In the aforementioned problems, one can see that there are some monotonic relationships between the class and some of the attributes. When taking into account this prior knowledge about the data, one needs to add some monotonicity constraints into the classification model, such as an SVM. It has been shown that a classification technique that incorporates monotonicity constraints can extract knowledge that is both more reasonable and comprehensible (Doumpos & Pasiouras, 2005; Doumpos & Zopounidis, 2009; Duivesteijn & Feelders, 2008; Evgeniou, Boussios, & Zacharia, 2005).

In the data mining literature about monotonicity constraints, there are two different approaches for dealing with problems that have prior knowledge of monotonic properties, although there only a few papers that focus on this issue. One method is to apply a relabeling technique to the data which contradicts the property of monotonicity (Duivesteijn & Feelders, 2008), while the other is to add the monotonicity constraints directly to the optimization modeling settings (Doumpos & Zopounidis, 2009; Evgeniou et al., 2005; Falck, Suykens, & De Moor, 2009). In the latter approach, Evgeniou et al. (2005) and Doumpos and Zopounidis (2009) simulated a large amount of monotonic data to formulate the constraints needed to enforce monotonicity, and the data thus simulated can increase the complexity of the problem computation-wise. Pelckmans, Espinoza, De Brabanter, Suykens, and De Moor (2005) discussed monotonic kernel regression and developed a monotonic LS-SVM regression model, and their formulation assumed that the input data follows a linear order and the bias term is omitted. However, in practice this assumption may not be valid, and the sparseness is lost in the LS-SVM model. Therefore, to deal with the limitations of these earlier studies, in this research we propose a new SVM model with monotonicity constraints that are inequalities and are based on the partial order in the input data. An index called frequency of monotonicity (FOM) is introduced to measure the monotonicity. The results of the experiment show that the proposed method, which takes into account the prior domain knowledge of monotonicity, performs better than the conventional SVM.

The rest of the paper is organized as follows. In Section 2, a review of the related literature is provided. In Section 3, we discuss the formulation of the monotonicity constrained SVM model. Section 4 presents the experimental results, while in Section 5 we present the discussion and conclusion of this work.

2. Literature review

In this section, we review the related literature to lay the foundation of this research project. The topics examined are SVM and classification with monotonicity constraints.

2.1. Support vector machines

SVMs are a state-of-the-art neural network technology based on statistical learning (Vapnik, 1995; Vapnik, 1998). They were originally designed for binary classification in order to construct an optimal hyperplane so that the margin of separation between the negative and positive data set would be maximized. If the data are linearly separated, the optimal hyperplane will separate the data without error, and the data points closest to the optimal separating hyperplane are known as support vectors. However, in practice, the data set of interest is usually linearly non-separable. In order to enhance the feasibility of linear separation, one can generally perform a non-linear transformation to move the data set into a higher dimensional space, the so-called feature space. Unfortunately, the problem of dimensionality in machine learning means that non-linear mapping is too difficult to solve. However, an SVM can overcome this by using the mechanism of an