

Available online at www.sciencedirect.com



Expert Systems with Applications

Expert Systems with Applications 33 (2007) 67-74

www.elsevier.com/locate/eswa

Application of support vector machines to corporate credit rating prediction

Young-Chan Lee *

College of Commerce and Economics, Dongguk University, Gyeongju, Gyeongbuk 780-714, South Korea

Abstract

Corporate credit rating analysis has drawn a lot of research interests in previous studies, and recent studies have shown that machine learning techniques achieved better performance than traditional statistical ones. This paper applies support vector machines (SVMs) to the corporate credit rating problem in an attempt to suggest a new model with better explanatory power and stability. To serve this purpose, the researcher uses a grid-search technique using 5-fold cross-validation to find out the optimal parameter values of RBF kernel function of SVM. In addition, to evaluate the prediction accuracy of SVM, the researcher compares its performance with those of multiple discriminant analysis (MDA), case-based reasoning (CBR), and three-layer fully connected back-propagation neural networks (BPNs). The experiment results show that SVM outperforms the other methods. © 2006 Elsevier Ltd. All rights reserved.

Keywords: Credit rating; SVM; BPN; MDA; CBR

1. Introduction

Credit ratings have been extensively used by bond investors, debt issuers, and governmental officials as a surrogate measure of riskiness of the companies and bonds. They are important determinants of risk premiums and even the marketability of bonds (Huang, Chen, Hsu, Chen, & Wu, 2004). The development of the corporate credit rating prediction model has attracted lots of research interests in academic and business community. Although of interests in accurate quantitative prediction of corporate bond rating, due to lack of scientific credit rating methodology and sufficient data accumulation to construct the model, the traditional approach produce an internal rating on the basis of credit officer's judgment to a significant extent in the real world (Shin & Han, 2001).

Several studies used statistical methods, including regression, multi-variate discriminant analysis, probit and logit models to predict bond rating (Altman & Katz,

E-mail address: chanlee@dongguk.ac.kr

1976; Ang & Patel, 1975; Baran, Lakonishok, & Ofer, 1980; Belkaoui, 1980; Bhandari, Soldofsky, & Boe, 1979; Martin, Henderson, Perry, & Cronan, 1984). McAdams (1980) employs the use of multiple discriminant analysis to design a statistical credit analysis model to assist portfolio managers to predict agency downgrades of electric utility bonds. Horrigan (1966) and Pogue and Soldofsky (1969) use multiple regression model to predict Moody's findings. Pinches and Mingo (1973) use factor analysis to screen variables for predicting bond ratings and then apply multiple discriminant analysis. Kamstra, Kennedy, and Suan (2001) improve the statistical predictive model by combining several forecasting methods to predict bond ratings in the transportation and industrial sectors. They use ordered logit method to combine forecasts and they find that combined forecasts outperform their input forecasts (Kim, 2005). Recently artificial intelligence approaches such as inductive learning (Shaw & Gentry, 1990), artificial neural networks (Dutta & Shekhar, 1996; Kim, 1992; Kwon, Han, & Lee, 1997; Maher & Sen, 1997; Moody & Utans, 1995; Singleton & Surkan, 1995), and case-based reasoning (Butta, 1994; Kim & Han, 2001; Shin & Han,

^{*} Tel.: +82 54 770 2317; fax: +82 54 770 2476.

^{0957-4174/\$ -} see front matter @ 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.eswa.2006.04.018

1999; Shin & Han, 2001) have been applied to bond rating. Although artificial intelligence approaches have several advantages over statistical methods (Salchenberger, Cinar, & Las, 1992; Tam & Kiang, 1992), the results of these studies were less than expected because the real data in application is usually unevenly distributed among classes and these approaches are limited in dealing with the ordinal nature of bond rating.

The purpose of this study is to apply support vector machines (SVMs), a relatively new machine learning technique, to corporate credit rating prediction problem and to provide a new model improving its prediction accuracy. Developed by Vapnik (1998), SVM is gaining popularity due to many attractive features and excellent generalization performance on a wide range of problems. In addition, bearing in mind that the optimal parameter search plays a crucial role to build a credit rating prediction model with high prediction accuracy and stability, this study employs a grid-search technique using 5-fold cross-validation to find out the optimal parameter values of RBF kernel function of SVM. To evaluate the prediction accuracy of SVM, this study also compares its performance with those of multiple discriminant analysis (MDA), case-based reasoning (CBR), and three-layer fully connected back-propagation neural networks (BPNs).

2. Support vector machines

Support vector machines (SVMs) use a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high-dimensional feature space. The linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, an optimal separating hyperplane (OSH) is constructed. Thus, SVM is known as the algorithm that finds a special kind of linear model, the maximum margin hyperplane. The maximum margin hyperplane gives the maximum separation between decision classes. The training examples that are closest to the maximum margin hyperplane are called support vectors. All other training examples are irrelevant for defining the binary class boundaries (Cristianini & Shawe-Taylor, 2000; Gunn, 1998; Hearst, Dumais, Osman, Platt, & Scholkopf, 1998; Vapnik, 1998).

SVM is simple enough to be analyzed mathematically since it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. In this sense, SVM may serve as a promising alternative combining the strengths of conventional statistical methods that are more theory-driven and easy to analyze, and more data-driven, distribution-free and robust machine learning methods. Recently, the SVM approach has been introduced to several financial applications such as credit rating, time series prediction, and insurance claim fraud detection (Fan & Palaniswami, 2000; Gestel et al., 2001; Huang et al., 2004; Kim, 2003; Tay & Cao, 2001; Viaene, Derrig, Baesens, & Dedene, 2002). These studies reported that SVM was comparable to and even outperformed other classifiers including ANN, CBR, MDA, and Logit in terms of generalization performance. Motivated by these previous researches, this study applies SVM to the domain of corporate credit rating prediction, and compare its prediction performance with those of MDA, CBR, and BPNs.

A simple description of the SVM algorithm is provided as follows. Given a training set $D = \{x_i, y_i\}_{i=1}^N$ with input vectors $x_i = (x_i^{(1)}, \ldots, x_i^{(n)})^T \in \mathbb{R}^n$ and target labels $y_i \in \{-1, +1\}$, the support vector machine (SVM) classifier, according to Vapnik's original formulation, satisfies the following conditions:

$$\begin{cases} \mathbf{w}^{\mathrm{T}}\phi(x_i) + b \ge +1, & \text{if } y_i = +1\\ \mathbf{w}^{\mathrm{T}}\phi(x_i) + b \leqslant -1, & \text{if } y_i = -1 \end{cases}$$
(1)

which is equivalent to

$$y_i[\mathbf{w}^{\mathrm{T}}\phi(x_i) + b] \ge 1, \quad i = 1, \dots, N$$
⁽²⁾

where **w** represents the weight vector and *b* the bias. Nonlinear function $\phi(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n_k}$ maps input or measurement space to a high-dimensional, and possibly infinite-dimensional, feature space. Eq. (2) then comes down to the construction of two parallel bounding hyperplanes at opposite sides of a separating hyperplane $\mathbf{w}^T \phi(x) + b = 0$ in the feature space with the margin width between both hyperplanes equal to $\frac{2}{\|\mathbf{w}\|^2}$. In primal weight space, the classifier then takes the decision function form (3):

$$\operatorname{sgn}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(x) + b) \tag{3}$$

Most of classification problems are, however, linearly non-separable. Therefore, it is general to find the weight vector using slack variable (ξ_i) to permit misclassification. One defines the primal optimization problem as

$$\min_{\mathbf{w},b,\xi} \qquad \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{i=1}^{N}\xi_{i} \qquad (4)$$

Subject to
$$\begin{cases} y_i(\mathbf{w}^{\mathrm{T}}\phi(x_i)+b) \ge 1-\xi_i, & i=1,\dots,N\\ \xi_i \ge 0, & i=1,\dots,N \end{cases}$$
(5)

where ξ_i 's are slack variables needed to allow misclassifications in the set of inequalities, and $C \in \mathbb{R}^+$ is a tuning hyperparameter, weighting the importance of classification errors vis-à-vis the margin width. The solution of the primal problem is obtained after constructing the Lagrangian. From the conditions of optimality, one obtains a quadratic programming (QP) problem with Lagrange multipliers α_i 's. A multiplier α_i exists for each training data instance. Data instances corresponding to non-zero α_i 's are called *support vectors*.

On the other hand, the above primal problem can be converted into the following dual problem with objective function (6) and constraints (7). Since the decision variables are support vector of Lagrange multipliers, it is easier to interpret the results of this dual problem than those of the primal one.

دريافت فورى 🛶 متن كامل مقاله

- امکان دانلود نسخه تمام متن مقالات انگلیسی
 امکان دانلود نسخه ترجمه شده مقالات
 پذیرش سفارش ترجمه تخصصی
 امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
 امکان دانلود رایگان ۲ صفحه اول هر مقاله
 امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
 دانلود فوری مقاله پس از پرداخت آنلاین
 پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات
- ISIArticles مرجع مقالات تخصصی ایران