Support Vector Machine을 이용한 지능형 신용평가시스템 개발

김경재*

Development of Intelligent Credit Rating System using Support Vector Machines

Kyoung-jae Kim*

이 논문은 2004년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2004-003-B00069)

ABSTRACT

In this paper, I propose an intelligent credit rating system using a bankruptcy prediction model based on support vector machines (SVMs). SVMs are promising methods because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. This study examines the feasibility of applying SVM in predicting corporate bankruptcies by comparing it with other data mining techniques. In addition, this study presents architecture and prototype of intelligent credit rating systems based on SVM models.

키워드

Support vector machines, Credit rating, Data mining

Ⅰ. Introduction

There have been many studies for the prediction of bankruptcies. The early days of these studies focused on application of statistical methods such as discriminant analysis, logit, probit and logistic regression. In addition, artificial intelligence (AI) techniques such as decision tree, k-nearest neighbor and artificial neural networks (ANNs) have been applied to this area.

Among these techniques, ANN has been popularly applied to this area because it produced superior prediction performance than other techniques. However, ANN has some limitations such as requirement of a big size of data samples, possibility of overfitting and poor explanatory power for the results.

Recently, a support vector machine (SVM), a novel neural network algorithm, was developed by Vapnik [7]. Many traditional neural network models seek to minimize the mis-classification error or deviation from correct solution of the training data, but SVM searches to minimize an upper bound of generalization error. In addition, the solution of SVM may be global optimum while other neural network models may tend to fall into a local optimal solution. Therefore, overfitting of the results is unlikely to occur with SVM.

In this paper, we apply SVM to predicting corporate

* 동국대학교 경영대학 정보관리학과 조교수

접수일자 : 2005. 8. 29

bankruptcies. In addition, this paper examines the feasibility of applying SVM in financial forecasting by comparing it with other data mining techniques. This study also presents architecture and prototype of intelligent credit rating system.

This paper consists of six sections. In Section 2, the basic concept of SVM and their applications in finance are presented. Section 3 describes research design and experiments. In Section 4, empirical results are summarized and discussed. Section 5 presents architecture and prototype for intelligent credit rating systems. Finally, Section 6 discusses the conclusions and limitations of this study.

Ⅱ. Support vector machines and their applications in finance

As mentioned above, SVM is proposed by Vapnik [7]. It uses linear model to implement nonlinear class boundaries through some nonlinear mapping the input vectors x into the high-dimensional feature space. A linear model constructed in the new space can represent a nonlinear decision boundary in the original space. An optimal separating hyperplane is subsequently constructed in the new space. Therefore, 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 the 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.

For the linearly separable case, a hyperplane separating the binary decision classes in the three-attribute case can be represented as the following equation:

$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 \tag{1}$$

where y is the outcome, x_i are the attribute values, and there are four weights w_i to be learned by the learning algorithm. The maximum margin hyperplane can be represented as the following equation in terms of the support vectors:

$$y = b + \sum \alpha_i y_i \mathbb{X}(i) \cdot \mathbb{X} \tag{2}$$

where y_i is the class value of training example $\mathbb{X}(i)$, $\cdot$ represents the dot product. The vector $\mathbb{X}$ represents a test example and the vectors $\mathbb{X}(i)$ are the support vectors. In this equation, b and α_i are parameters that determine the hyperplane. From the implementation point of view, finding the support vectors and determining the parameters b and α_i are equivalent to solving a linearly constrained quadratic programming.

As mentioned above, SVM constructs linear model to implement nonlinear class boundaries through the transforming the inputs into the high-dimensional feature space. For the nonlinearly separable case, a high-dimensional version of Eq. (2) is simply represented as follows:

$$y = b + \sum \alpha_i y_i K(\mathbb{X}(i), \mathbb{X}). \tag{3}$$

The function $K(\mathbb{X}(i), \mathbb{X})$ is defined as the kernel function. There are some different kernels for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space. Choosing among different kernels the model that minimizes the estimate, one chooses the best model. Common examples of the kernel function are the polynomial kernel $K(x,y) = (xy+1)^d$ and the Gaussian radial basis function $K(x,y) = exp(-1/\delta^2 (x-y)^2)$ where d is the degree of the polynomial kernel and δ^2 is the bandwidth of the Gaussian radial basis function kernel.

For the separable case, there is a lower bound 0 on the coefficient α_i in Eq. (3). For the non-separable case, SVM can be generalized by placing an upper bound C on the coefficients α_i in addition to the lower bound [8].

As mentioned above, the backpropagation (BP) network has been widely used in the area of financial forecasting because of its broad applicability to many business problems and preeminent learning ability. However, the BP network has many disadvantages including the need for the determination of the value of controlling parameters and the number of processing elements in the layer, and the danger of overfitting problem. On the other hand, there are no parameters to tune except the upper bound C for the non-separable cases in linear SVM [2]. In addition, overfitting is unlikely to occur with SVM because it may be caused by too much flexibility in the decision boundary, but the maximum margin hyperplane is relatively stable and gives little flexibility [8].

Although SVM has the above advantages, there are few studies for the application of SVM in financial forecasting. Mukherjee et al. [4] showed the applicability of SVM to time-series forecasting. Tay and Cao [6] examined the predictability of financial time-series with SVMs. They showed that SVMs outperformed the BP networks on the criteria of normalized mean square error, mean absolute error, directional symmetry and weighted directional symmetry. Kim [3] applied SVM to predicting future direction of stock price index. In his study, SVM outperformed BP networks and case-based reasoning for the prediction of stock price index. Recently, Shin et al. [5] investigated the efficacy of applying SVM to bankruptcy prediction. The results of their research presented that the accuracy and generalization performance of SVM was better than that of BP network as the training set size got smaller.

Among these prior studies, the most popular application area was not bankruptcy prediction but financial time-series forecasting. Although Shin et al. [5] applied SVM to corporate bankruptcy prediction, but they did not investigate the effect of various kernel functions. In this study, we investigate the effect of different kernel functions on the performances of bankruptcy prediction models.

Ⅲ. Research data and experiments

2.1. Research data

Research data used in this study consists of financial ratios and the status of bankrupt or non-bankrupt for corresponding corporate. The sample of bankrupt companies was 1335 ones in heavy industry which filed for bankruptcy between 1996 and 2000. The non-bankrupt companies were 1335 ones in heavy industry which filed between 1999 and 2000. Thus, the total number of samples is 2670 companies. For independent variables, we first generate 164 financial ratios from the financial statement of each company. Finally, we get 15 financial ratios as independent variables through the two independent sample t-test and the forward selection procedure based on logistic regression.

2.2. Experiments

In this study, the polynomial kernel and the Gaussian radial basis function are used as the kernel function of SVM. Tay and Cao [6] showed that the upper bound C and the kernel parameter δ^2 play an important role in the performance of SVMs. Improper selection of these two parameters can cause the overfitting or the underfitting problems. Since there is few general guidance to determine parameters of SVM, this study varies values of parameters to select optimal values for the best prediction performance. This study uses LIBSVM software system [1] to perform experiments.

For comparison purpose, this study employs discriminant analysis (DA), case-based reasoning (CBR), and backpropagation neural networks (BPN) because they are often used as benchmark models in similar prior studies.

Ⅳ. Experimental results

One of the advantages of linear SVM is that there is no parameter to tune except the constant C. But the upper bound C on the coefficient α_i affects prediction performance for the cases where the training data is not separable by a linear SVM [2]. For the nonlinear SVM, there is an additional parameter, the kernel parameter, to tune. This study uses two kernel functions including the Gaussian radial basis function and the polynomial function.

For the polynomial function, this study varies the degree of the polynomial kernel, d, from 1 to 3. Under this experimental condition, the prediction performances of the polynomial function fell between 53.8642% and 87.6815% in the training data set and between 53.4579% and 85.7944% in the holdout data set.

For the Gaussian radial basis function, this study lists the prediction performance with respect to various kernel parameters and constants. Table 1 shows the prediction performance of SVMs with various parameters.

Table 1. The prediction performance of SVMs (%)

δ^2	C	Training data set	Holdout data set
1	20	90.8665	85.2336
	40	91.8033	85.0467
	60	92.4122	85.0467
	80	92.6932	84.8598
	100	93.0211	85.2336
2	20	88.5246	86.5421
	40	89.2272	86.1682
	60	89.6487	85.9813
	80	90.0703	85.6075
	100	90.8197	85.2336
5	20	86.4637	84.1121
	40	87.4005	86.1682
	60	87.4473	85.7944
	80	87.9157	86.3551
	100	88.0094	86.1682
10	20	85.7611	83.5514
	40	86.1827	83.9252
	60	86.4169	84.1121
	80	86.6042	84.4860
	100	86.8384	85.0467
30	20	85.0117	83.5514
	40	84.9180	83.5514
	60	85.1054	83.5514
	80	85.2927	83.7383
	100	85.3396	83.7383

In Table 1, the best prediction performance of the holdout data is recorded when δ^2 is 2 and C is 20. This result outperforms the best performance of the polynomial function.

In addition, this study compares the best SVM model with the best DA, CBR and BPN models. Table 2 gives the best prediction performances of these models.

Table 2. The best prediction performances of DA, CBR, BPN and SVM (%)

	DA	CBR	BPN	SVM
Training data	83.75	N/A	81.22	88.52
Holdout data	83.55	80.75	81.68	86.54

In Table 2, SVM outperforms other models for the holdout data. The results may indicate the feasibility of SVM in credit rating. The McNemar tests are performed to examine whether SVM significantly outperforms the other three models. Table 3 shows the results of the McNemar test to compare the prediction performance of the holdout data.

Table 3. McNemar values for the holdout data

	CBR	BPN	SVM
DA	2.154	0.880	5.114*
CBR		0.137	11.688**
BPN			8.446**

** significant at the 1% level, * significant at the 5% level

As shown in Table 3, SVM performs better than BPN and CBR at 1% statistical significance level. In addition, SVM outperforms DA with 5% statistical significance level.

Ⅴ. System architecture and prototype

In this section, we present system architecture and prototype for the intelligent credit rating system based on SVM models. Fig. 1 shows general system architecture and components for the intelligent credit rating system. In Fig. 1, the system integrates external database and model-base based on SVM model.

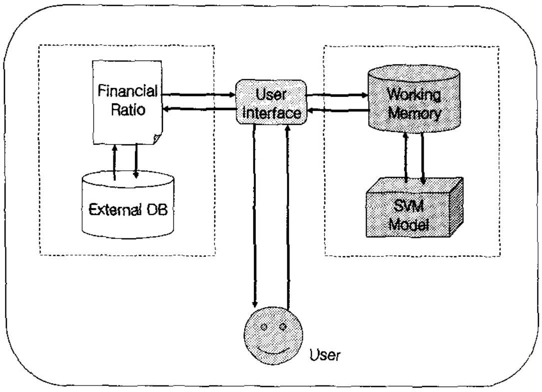

Fig. 1. Architecture and components of the system

This study also presents prototype of the system. Fig. 2 shows system prototype of the intelligent credit rating system using SVM model.

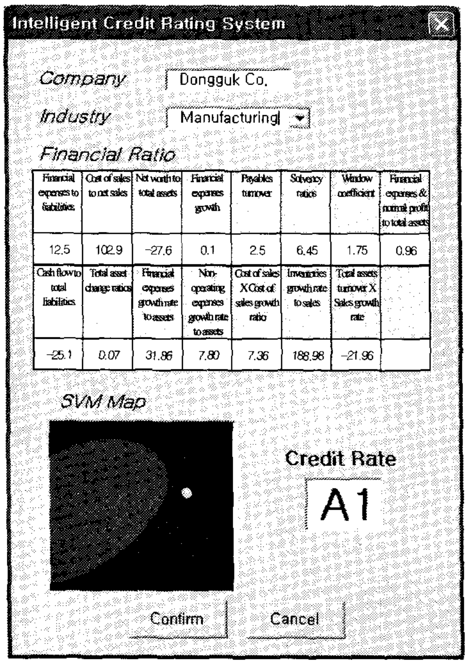

Fig. 2. Prototype of the system

In Fig. 2, there are five components including company name, industry, financial ratio, SVM map, and credit rate. Financial ratio shows major financial ratio of the company, credit rate means final credit rate of the company. In SVM map, green and black color of circles represent hyperplane between two different classes of the holdout examples and green and red bullets means two different classes of examples of the holdout data. In addition, yellow bullet in SVM map represents geographical position of the company.

Ⅵ. Concluding remarks

This study tested the feasibility of SVM to corporate credit rating. In addition, we investigated the effect of the value of the upper bound C and the kernel parameter δ^2 in SVM. The experimental results showed that the prediction performances of SVMs are sensitive to the value of these parameters. In addition, this study compared SVM with DA, BPN and CBR. The experimental results showed that SVM outperformed the other models. The results may be attributable to the fact that SVM implements the structural risk minimization principle and this leads to better generalization than conventional techniques. Finally, this study concludes that SVM provides a promising alternative for credit rating problem. However, the prediction performance may be increased if the optimum parameters of SVM are selected and this remains a very interesting topic for further study.

References

[1] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines", Technical Report, Department of Computer Science and Information Engineering. National Taiwan University, Available at http://www.csie.edue.tw/~chlin/papers/libsvm.pdf, 2001.

[2] H. Drucker, D. Wu, and V.N. Vapnik, "Support vector machines for spam categorization", IEEE Transactions on Neural Networks, Vol. 10, No. 5, pp. 1048-1054, 1999.

[3] K. Kim, "Financial time series forecasting using support vector machines", Neurocomputing, Vol. 55, No. 1-2, pp. 307-319, 2003.
[4] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines", in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, Amelia Island, FL, pp. 511-520, 1997.
[5] K.S. Shin, T.S. Lee, and H. Kim, "An application of support vector machines in bankruptcy prediction model", Expert Systems with Applications, Vol. 28, No. 1, pp. 127-135, 2005.
[6] F.E.H. Tay and L. Cao, "Application of support vector machines in financial time series forecasting", Omega, Vol. 29, pp. 309-317, 2001.
[7] V.N. Vapnik, "Statistical Learning Theory", Wiley, New York, 1998.
[8] I.H. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann Publishers, San Francisco, CA, 2000.

저자소개

Kyoung-jae Kim

Kyoung-jae Kim is an assistant professor of Department of Information Systems at Dongguk University. He received his Ph.D. degree in management information systems from KAIST. He has published in *Applied Intelligence, Expert Systems,* · *Expert Systems with Applications, Intelligent Data Analysis, Intelligent Systems in Accounting, Finance & Management, Neural Computing & Applications, Neurocomputing*, etc. His research interests include data mining, knowledge management, and intelligent agents.