Support Vector Machines Approach to Pattern Detection in Bankruptcy Prediction and Its Contingency

Kyung-shik Shin[1], Kyoung Jun Lee[2], and Hyun-jung Kim[1]

[1] Ewha omans University, College of Business Administration
11-1 Daehyun-Dong, Seodaemun-Gu, Seoul 120-750, Korea
ksshin@ewha.ac.kr, charitas@empal.com
[2] School of Business, Kyung Hee University
Hoegi-Dong, Dongdaemun-Ku, Seoul, Korea
klee@khu.ac.kr

Abstract. This study investigates the effectiveness of support vector machines (SVM) approach in detecting the underlying data pattern for the corporate failure prediction tasks. Back-propagation neural network (BPN) has some limitations in that it needs a modeling art to find an appropriate structure and optimal solution and also large training set enough to search the weights of the network. SVM extracts the optimal solution with the small training set by capturing geometric characteristics of feature space without deriving weights of networks from the training data. In this study, we show the advantage of SVM approach over BPN to the problem of corporate bankruptcy prediction. SVM shows the highest level of accuracies and better generalization performance than BPN especially when the training set size is smaller.

1 Introduction

Early techniques of bankruptcy prediction include statistical techniques such as multiple discriminant analysis, logit and probit. Recently artificial intelligence techniques such as neural networks have been an alternative method for the classification problems and numerous theoretical and experimental studies reported the usefulness of the back-propagation neural network (BPN) in classification studies. However, back-propagation neural network models have several limitations in building the model. First, finding an appropriate neural network model among numerous candidates is an artistic work because there are large numbers of controlling parameters and processing elements in the model. Second, its gradient descent search process for computing the synaptic weights can converge to a local minimum that is only a good fit for the training examples. Third, its empirical risk minimization principle seeking the minimization of the training error does not guarantee good generalization performance. Finally, the size of the training set is also the main issue to be resolved because the sufficiency and efficiency of the training set is one of most influencing factors.

In this paper, we investigate the effectiveness of support vector machines (SVM) approach, as an alternative BPN, to corporate failure prediction tasks. SVM classification exercise finds hyperplanes in the possible space for maximizing the distance from the hyperplane to the data points, which is equivalent to solving a quadratic

N.R. Pal et al. (Eds.): ICONIP 2004, LNCS 3316, pp. 1254–1259, 2004.

optimization problem [9,10]. The solution of strictly convex problems for SVM is unique and global. In SVM, the structural risk minimization (SRM) principle is implemented and known to have high generalization performance. As the numbers of support vectors increase, SVM model is constructed through the trade-off between decreasing the number of training errors and increasing the risk of overfitting the data. SVM captures geometric characteristics of feature space without a need to derive weights of networks from the training data, therefore it can extract the optimal solution only with the small training set size. There are several arguments and observations supporting the high accuracy of SVM in the small training set size as well as the results showing that the accuracy and generalization performance of SVM is better than that of the standard BPN.

Byun and Lee [2] present a comprehensive survey on applications of Support Vector Machines (SVMs) for pattern recognition. Since SVMs show good generalization performance on many real-life data and the approach is properly motivated theoretically, it has been applied to wide range of applications. The support vector machines (SVMs) has been adopted for predicting bankruptcies [7] and compared with NN, MDA and learning vector quantization (LVQ) [5]. SVM obtained the best results, followed by NN, LVQ and MDA. Van Gestel et al. [8] also reports on the experiment with least squares support vector machines, a modified version of SVMs, and shows significantly better results in bankruptcy prediction when contrasted with the classical techniques.

The structure of this paper is organized as follows: The next section provides a brief description of several superior points of the SVM algorithm compared with BPN. The third section describes the research data and experiments. The fourth section summarizes and analyzes empirical results. The final section discusses the conclusions and future research issues.

2 SVM Versus BPN

Compared with the limitations of the BPN, the major advantages of SVM are as follows:

Few Controlling Parameters: Because a large number of controlling parameters in BPN such as the number of hidden layers, the number of hidden nodes, the learning rate, the momentum term, epochs, transfer functions and weights initialization methods are selected empirically, it is a difficult task to obtain an optimal combination of parameters that produces the best prediction performance. On the other hand, SVM has only two free parameters, namely the upper bound and kernel parameter.

Existence of Unique and Global Solution: Because the gradient descent algorithm optimizes the weights of BPN so that the sum of square error is minimized along a steepest slope of the error surface, the result from training can be massively multi modal, leading to non-unique solutions, and be in the danger of getting stuck in a local minima. On the other hand, training SVM is equivalent to solving a linearly constrained quadratic programming therefore SVM guarantees the existence of unique, optimal and global solution.

Good Generalization Performance: SVM implement the structural risk minimization (SRM) principle that is known to have a high generalization characteristic. SRM is the approach to trading off empirical error with the capacity of the set called VC (Vapnik-Chervonenkis) dimension, which seeks to minimize an upper bound of the generalization error rather than minimize the training error. In order to apply SRM, the hierarchy of hypothesis spaces must be defined before the data is observed. In SVM, the data is first used to decide which hierarchy to use and then subsequently to find a best hypothesis from each. Therefore the good generalization performance of SVM is not always attributable to SRM since the result of SVM is obtained from a data dependent SRM [1]. There exists no explicitly established theory that good generalization performance is guaranteed for SVM. However, it seems plausible that performance of SVM is more robust than that of BPM in that the two measures in terms of the margin and number of support vectors give information on the relation between the input and target function according to different criteria, either of which is sufficient to indicate good generalization. On the other hand, BPN trained based on minimizing a squared error criterion at the network output tends to produce a classifier with the only large margin measure. In addition, flexibility from choosing training data is likely to occur with weights of BPN model, but the maximum hyperplane of SVM is relatively stable and gives little flexibility in the decision boundary.

Training by Geometric Characteristics with the Small Data Set Size: SVM learns through capturing geometric picture corresponding to the kernel function, therefore it is constructed from just a subset of the training data, the support vectors. Moreover, no matter how large the training set size is, SVM has infinite VC dimension. That is why SVM is capable of extracting the optimal solution with the small training set size. On the other hand, as for the case of BPN containing a single hidden layer and used as a binary classifier, the number of training examples should be approximately 10 times the number of weights in the network. With the 10 input and hidden nodes, the learning algorithm will need more than 1000 training set size that is sufficient for a good generalization [6]. In most practical applications there can be a huge gap between the actual size of the training set needed and that is available. Utilizing the feature space images by the kernel function SVM is applicable in circumstances that have proved difficult or impossible for BPN where data in the plane is randomly scattered and where the density of the data's distribution is not even well defined [4].

3 Research Data and Experiments

The research data we employ is provided by Korea Credit Guarantee Fund in Korea, and consists of externally non-audited 2,320 medium-size manufacturing firms, which filed for bankruptcy (1,160 cases) and non-bankruptcy (1,160 cases) from 1996 to 1999. We select 1,160 non-bankrupt firms randomly from among all solvent firms, so the choice covers the whole spectrum from healthy to borderline firms in order to avoid any selection bias. The data set is arbitrarily split into two subsets; about 80% of the data is used for a training set and 20% for a validation set. The training data for SVM is totally used to construct the model and for BPN is divided into 60% training set and 20% test set. The validation data is used to test the results with the data that is not utilized to develop the model.

Using this first data set, 7 different datasets are constructed that differ in the number of cases included in the training and test subsamples. The validation set of all datasets consisting of 464 cases is identical, which ensures that the obtained results from validation data are not influenced by the fluctuation of data arrangement.

We apply two stages of the input variable selection process. At the first stage, we select 52 variables among more than 250 financial ratios by independent-samples t-test between each financial ratio as an input variable and bankrupt or non-bankrupt as an output variable. In the second stage, we select 10 variables using a MDA (multivariate discriminant analysis) stepwise method to reduce dimensionality. We select input variables satisfying the univariate test first and then select significant variables by the stepwise method for refinement.

In this study, the radial basis function is used as the kernel function of SVM. Since SVM does not have a general guidance for determining the upper bound C and the kernel parameter δ2, this study varies the parameters to select optimal values for the best prediction performance. The MATLAB support vector machine toolbox version 0.55 beta executes these processes [3]. In order to verify the applicability of SVM, we also design BPN as the benchmark with the following controlling parameters. The structure of BPN is standard three-layer with the same number of input nodes in the hidden layer and the hidden and output nodes use the sigmoid transfer function. For stopping the BPN training, test set that is not a subset of the training set is used, but the optimum network for the data in the test set is still difficult to guarantee generalization performance.

4 Results and Analysis

To investigate the effectiveness of the SVM approach trained by small data set size in the context of the corporate bankruptcy classification problem, the results obtained from various data set sizes are compared with those of BPN. The test results for this study are summarized in Table 1. Each cell of Table 1 contains the accuracy of the classification techniques. The results in Table 1 show that the overall prediction performance of SVM on the validation set is consistently good as the number of training set size decreases. Moreover, the accuracy and the generalization using a small size of data set (5th, 6th, and 7th set) are even better than those using a large size of data set (3rd and 4th set). Especially the performance of 5th and 6th set is similarly excellent compared to that of the 1st set.

The experimental result also shows that the prediction performance of SVM is sensitive to the various kernel parameter δ2. The accuracy on the training set of the most data set decreases as δ2 increases; on the other hand, the accuracy on the validation set shows a tendency to increases with increasing δ2. In addition, however, the prediction performance of the validation set is more stable and insensitive than that of training set. This indicates that a small value for δ2 has an inclination to overfit the training data and an appropriate value for δ2 plays an important role on the generalization performance of SVM. The results of SVM in the range of δ2 from 25 to 75 show the best prediction performances. While the results of BPN are comparable with SVM in large size of data set (1st and 2nd set), in case that a training set size is less

Table 1. Classification accuracies (%) of various data set sizes in which C=100

data set	SVM number of set	SVM accuracy (δ^2) 1	25	50	75	100	BPN number of set	BPN accuracy
(a) 1st set								
	train (1,856)	95,7	75,4	75,2	74,5	74,2	train (1,392)	71,7
	val (464)	60,3	73,9	76,5	76,3	76,3	val (464)	72,2
	total (2,320)						test (464)	74,1
(b) 2nd set (SVM training set of (a)*25%)								
	train (464)	99,1	80,6	79,7	79,3	78,7	train (370)	75,7
	val (464)	59,9	71,8	72,6	73,7	73,7	val (464)	71,6
	total (928)						test (94)	72,3
(c) 3rd set (SVM training set of (b)*50%)								
	train (232)	100,0	73,2	71,8	70,9	71,8	train (184)	57,1
	val(464)	57,1	65,5	66,2	67,5	67,0	val (464)	56,0
	total(696)						test (48)	56,3
(d) 4th set (SVM training set of (c)*50%)								
	train (116)	100,0	84,0	79,2	75,7	70,8	train (92)	55,4
	val (464)	54,5	67,0	60,8	59,5	57,5	val (464)	54,5
	total (580)						test (24)	54,2
(e) 5th set (SVM training set of (d)*50%)								
	train (58)	100,0	91,4	87,9	84,5	81,0	train (46)	47,8
	val (464)	59,9	72,6	75,2	73,3	73,5	val (464)	50,9
	total (522)						test (12)	50,0
(f) 6th set (SVM training set of (e)*50%)								
	train (28)	100,0	100,0	92,9	89,3	89,3	N/A	
	val (464)	58,8	70,0	73,3	70,9	70,3		
	total (492)							
(g) 7th set (SVM training set of (f)*50%)								
	train (14)	100,0	100,0	92,9	92,9	92,9	N/A	
	val (464)	54,3	62,1	64,0	63,8	62,9		
	total (478)							

than 200, it is hard (3rd, 4th, and 5th set) or even impossible (6th and 7th set) to learn the model.

5 Conclusions

From the results of experiments, we can conclude that SVM is the better approach to learn a small size of data patterns as opposed to ordinary BPN, which confirm existing researches [2,5,7,8]. In addition, our research confirms that SVM has the highest level of accuracies and better generalization performance than BPN as the training set size is getting smaller sets. In this study, we show that the proposed classifier of SVM approach outperforms BPN to the problem of corporate bankruptcy prediction. In addition, we investigate and summarize the several superior points of the SVM algorithm compared with BPN.

Our study has the following limitations that need further research. First, in SVM, the choice of the kernel function and the determination of optimal values of the parameters have a critical impact on the performance of the resulting system. We also need to examine the effect of other factors that is fixed in the experiment such as various values of the upper bound C and the kernel function. The second issue for future research relates to the generalization of SVM on the basis of the appropriate level of the training set size and gives a guideline to measure the generalization performance.

References

1. Burges, C., A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2(2):955-974, 1998.
2. Byun, H. and Lee, S., Applications of Support Vector Machines for Pattern Recognition: A Survey, Lecture Notes in Computer Science 2388:213-236, 2002.
3. Cawley, G., Support Vector Machine Toolbox, University of East Anglia, School of Information Systems, http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox/, 2000.
4. Friedman, C., Credit Model Technical White Paper, Standard & Poor's, New York, McGraw-Hill, 2002.
5. Fan, A. and Palaniswami, M., A new approach to corporate loan default prediction from financial statements, in Proc. Computational Finance/Forecasting Financial Markets Conf. CF/FFM-2000, London (CD), UK, May 2000.
6. Haykin, S., Neural Networks: A comprehensive foundation. Macmillan College Publication, New York, 1994.
7. Häardle, W., Moro, R., Schäafer, D., Predicting Corporate Bankruptcy with Support Vector Machines, Working Slide, Humboldt University and the German Institute for Economic Research, available on http://ise.wiwi.hu-berlin.de/~isedoc03/slides.pdf, 2003.
8. Van Gestel T., Baesens B., Suykens J., Espinoza M., Baestaens D.E., Vanthienen J., De-Moor B., Bankruptcy Prediction with Least Squares Support Vector Machine Classifiers, Proceedings of the IEEE International Conference on Computational Intelligence for Financial Engineering (CIFEr2003), Hong Kong, pp. 1-8, March 2003.
9. Vapnik, V., "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.
10. Vapnik, V., Statistical Learning Theory, John Wiley & Sons, New York, 1998