Communications of the IIMA

Volume 7 | Issue 4 Article 8

2007

Using SVM with Financial Statement Analysis for Prediction of Stocks

Shuo Han
Department of Management Science and Engineering University of Science and Technology Beijing

Rung-Ching Chen
Department of Information Management Chaoyang University of Technology

Using SVM with Financial Statement Analysis for Prediction of Stocks

Shuo Han
Department of Management Science and Engineering
University of Science and Technology Beijing,

Rung-Ching Chen
Department of Information Management
Chaoyang University of Technology

ABSTRACT

At present, there are many technical analyses for prediction in stock market. However, the technical indices are fluctuated with the quantity of stock exchanges. The financial indices are more reliable, nonvolatile and valid compared with the technical indices. In this paper, we propose an original and universal method by using SVM with financial statement analysis for prediction of stocks. We applied the SVM to construct the prediction model and select Gaussian radial basis function (RBF) as the kernel function. The experimental results show our method not only improve the accuracy rate, but also meet the different stockholders' expectations.

INTRODUCTION

Support vector machine (SVM) is a useful technique for data classification (Burges, 1998), regression (Smola et al, 1998) and prediction (Müller et al, 1997). Previously there has been a lot of study using artificial neural network (ANN) in these areas, especially in the field of prediction. However, in the stock market, because the data often has enormous noises and complex dimensionality, the ANN method has some limitations (Kim, 2003). Recently, SVM has been successfully used in the field of prediction. SVM can treat higher dimensional data better even with a relative low amount of training set. Further more, it can present a good ability of generalization for complex model (Thissen, 2003).

Some applications by SVM to predict the stock market have been issued, but the degree of accuracy rate and the acceptability of certain prediction are measured by the predictors' deviation from their own experiences or the ineffective data (Huang, 2005). In the field of prediction for stock market, the most important thing is to improve the prediction accuracy rate (Huang et al, 2006. Chen et al, 2006). However, little study has justified the suitability of stock market prediction by SVM.

In this paper, we propose an original and universal method by using SVM with financial statement analysis for prediction of stocks. Commonly there are many technical analyses for prediction in stock market. But these technical indices such as RSI, BIAS, etc. appear to fluctuate with the quantity of stock exchanges. Compared with the technical indices, the financial indices from the financial statement are much more reliable, nonvolatile and valid. The goal of this paper is to improve the accuracy rate of prediction and to meet different kinds of stockholders' expectations.

This paper is organized as follows. In section 2, we will briefly explain the theory of SVM and some concepts from finance and accounting. In section 3, the methodology is given. In section 4, the experiment and the experimental result is shown. In section 5, we will present the conclusion and some suggestions.

THEORY

SVM

The support vector machine (SVM) is a type of learning machine that is based on statistical theory and it is a popular technique for classification. In order to perform binary deviation, the SVM uses a high dimension space to find a hyper plane where the error rate is minimal. The methodology of SVM can be stated briefly as follows:

Given a training set of instance-label pairs (x_i, y_i), i = 1, . . . , l where $x_i \in R^n$ and $y \in \{1,-1\}^l$, the support vector machines require the solution of the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}w^T w + c\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i\left(w^T z_i + b\right) \geq 1-\xi_i, \\ & \xi_i \geq 0, \qquad i = 1,...,l. \end{aligned} \tag{1}$$

The training vector x_i are mapped into a higher dimensional space by the function φ as $z_i = \varphi(x_i)$. C > 0 is the penalty parameter of the error term. Thus the problem is solved as follows (Chang et al, 2007):

$$\begin{aligned} \min_{\alpha} \quad & F(\alpha) = \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \text{subject to} \quad & y^T\alpha = 0, \\ & 0 \leq \alpha_i \leq C, \qquad i = 1,...,l \end{aligned} \tag{2}$$

Where e is the vector of all, C > 0 is the upper bound, Q is an l by l positive semidefinite matrix, $Q_{ij}=y_i y_j K(x_i, x_j)$, and $K(x_i, x_j)= \varphi(x_i)^T\varphi(x_j)$ is the kernel function. Thus the decision function is:

$$\begin{aligned} & \sin\left(w^T\varphi(x)+b\right) = \sin\left(\sum_{i=1}^{l}\alpha_i y_i K(x_i, x)+b\right) \\ & w = \sum_{i=1}^{l}\alpha_i y_i \varphi(x_i) \end{aligned} \tag{3}$$

In this paper, the kernel function of the research is called Gaussian kernel (Keerthi, 2003):

$$K(\tilde{x}, \bar{x}) = \exp\left(\frac{-\left\|\tilde{x}-\bar{x}\right\|^2}{2\sigma^2}\right) \tag{4}$$

LIBSVM is a library for SVM. The goal of LIBSVM is to produce a model which predicts target value of instant datas in the testing set which are given only the attributes and let the users can easily use SVM as a model. Often in LIBSVM the kernel function is radial basis function (RBF), and the γ is called kernel parameter (Hsu, 2003). It is represented as follows:

$$K(x_i, x_j) = \exp(-\gamma\left\|x_i - x_j\right\|^2), \qquad \gamma > 0 \tag{5}$$

Concepts of Finance and Accounting

As mentioned above, there are many technical analyses for stock market prediction. The stockholders often make decisions by applying the technical indices such as RSI, BIAS, MACD, PSY, KDJ, etc. But these technical indices are easily fluctuating with the quantity of stock exchanges. Besides, for some man-made reasons, these technical indices are not very trustful. Consequently there are some limitations to these applications when the stockholders want to evaluate their target stocks or make decisions.

The stock companies will release their financial statements for each financial year to public. In that case, we can get

some useful financial indices from the target financial statement. For instance, current assets, non-current assets, total assets, shareholders equity, share capital, current liabilities, non-current liabilities, main business income, profit from main business, gross profit, earnings before interest and tax, net profit, net cash income per share, earnings per share, book value per share, adjusted book value per share, return on equity, the interests of shareholders ratio, gross margin, net profit growth rate, shareholders rate, net rate of return on capital-growth, main business income growth and main profit growth.

The stockholders can process different evaluation by selecting the related account elements they needed. In this way, the stockholders are able to do the relative financial analyses. For example, by selecting the net profit margin, gross profit margin, net profits on assets, return on assets, net profit ratio, return on equity and growth rate of earnings before tax, we can do the profitability analysis. By selecting the main business income growth, net profit growth rate, total asset growth rate, shareholders rate and main profit growth, we can do the growth rate analysis. By selecting the debt to total assets ratio, debt to total assets ratio, long term debt ratio and fixed ratio, we can do the financial structure analysis. By selecting the current ratio, quick ratio, current cash debt coverage ratio, the interests of shareholders ratio, current liabilities ratio and gross gearing, we can do the solvency analysis. By selecting the inventory turnover, asset turnover, receivables turnover, main business cost ratio, operating expense ratio, administrative expense ratio and finance expense ratio, we can do the operating efficiency analysis.

In the stock market, the operating status is the pivotal factor for the stock value. It could be reflected by the financial statement. Therefore, analyzing and researching the account elements from the financial statement are very important. If we can do the advanced analyses for these items, we will be more aware of the financial and operating status of those stock companies. Certainly it will be helpful for stockholders making decisions to buy the target stocks. Besides, compared with the technical indices, these indices are much more reliable, nonvolatile and valid.

For the common stockholders, they can not evaluate and predict the target stock by analyzing all the financial indices as the professionals. They often concern more on the primary financial indices such as earning per share growth rate, book value per share growth rate, return on equity, net profit growth rate, shareholders rate, net rate of return on capital-growth and the main business income growth. If the common stockholders can integrate the professional advices with the primary indexes analysis, they will be more confident in making decisions of the stock investment.

The goal of this paper is to process a new method by using SVM with financial statements analysis for prediction of stocks to improve the accuracy rate and meet different stockholders' expectations. Normally, some professionals or experts will predict the trend of stocks in future by their own field of knowledge, information and techniques, and release the results to public. We improve the accuracy rate of prediction based on these professional results combined with our results of primary financial indices. In that case, the results will be more reliable. We hope the method would provide a better way of application in the real world, which will be explained in the following sections.

METHODOLOGY

Data Collection

The research data used in this study is selected from the financial statements released by the stock companies of the stock exchange in Shanghai and Shenzhen. By the ending of this study, the numbers of the financial statements released by the stock companies is 251, so the research data are brand-new and dependable. The primary indices we selected are Earnings Per Share (EPS), Book Value Per Share (BVPS) and Net Profit Growth Rate (NPGR). Based on the results released by the professionals and experts, Outstanding Achievement Growth Rate (OAGR), which we have classified two choices, +1 and -1, and denoted as SVM. Consequently, in the experiment, we combined the SVM with three primary indices respectively. Then we can generate eight kinds of prediction modules. The data set and the prediction modules are shown in Table 1 and Table 2.

Table 1: The data set (using SVM to present OAGR and different financial indices).

SVM	OAGR	EPS	BVPS	NPGR
+1	0.296	0.351	4.360	81.792
-1	-0.799	-0.702	-0.381	-12.089
-1	0.083	0.056	3.138	-32.148
…	…	…	…	…
+1	0.360	0.409	3.099	12.241

Table 2: Prediction modules.

Experiment	SVM	OAGR	EPS	BVPS	NPGR
1	√	√			
2	√	√	√		
3	√	√		√	
4	√	√			√
5	√	√	√	√	
6	√	√	√		√
7	√	√		√	√
8	√	√	√	√	√

In this paper, we will compare the eight different kinds of experimental results from the eight modules. The first module (Experiment 1: SVM and OAGR) is the basic module. Other seven experimental results will compared with it. In that case, for different stockholders, they can select different items to compare. Because the first experimental results is only reflected the accuracy rate of the professional or experts, not concluding the stockholders' analyses. We define the first result as the basic result. Different analyzer can get different results by selecting the target items they need.

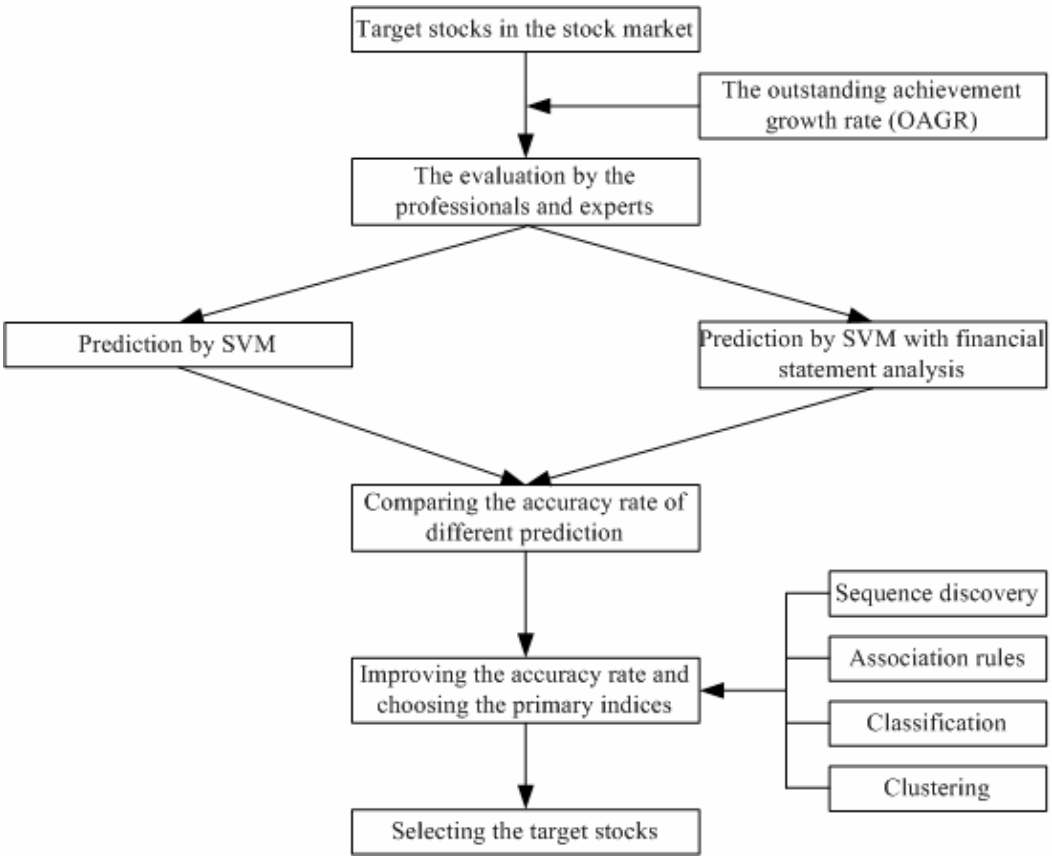
Theoretical Model

Normally the professionals or experts will predict the trend of target stocks based on the Outstanding Achievement Growth Rate (OAGR), and the results will be classified to five levels. From good to bad, they are making profit, favorable of outstanding achievement, advising, making up the deficits, deficit. In this paper, our method is based on the hypothesis that all the stockholders are sensible and rational people. Commonly they will select the stocks belonging to the level of making profit or level of favorable of outstanding achievement for the target stocks. Other

levels will be denied. In this way, we define the former choice as class +1. Similarly, we define the latter choice of denying the other levels of stocks as class -1. Thus, the method will be suitable of the SVM. And then, we will combine the SVM with some primary financial indices we selected to make prediction. We hope that not only will the accuracy rate be improved, but also will the different financial indices selected by the stockholders be presented different results for their needs.

In this paper, we compare the accuracy rate of the eight experiments. The experimental results can help us to make decision whether we choose the target stocks or not. Besides, by analyzing the accuracy rate, we can evaluate the effectiveness of the primary indices we selected. If the indices are not useful to improve the accuracy rate of prediction, we can reevaluate or regulate other financial indices from the financial statement until we find the best combination for the prediction of stocks. In this way, for different stockholders, they can choose their favorite financial indices to satisfy their different predictions. The theoretical model is shown in Figure 1.

Figure 1: Theoretical model.

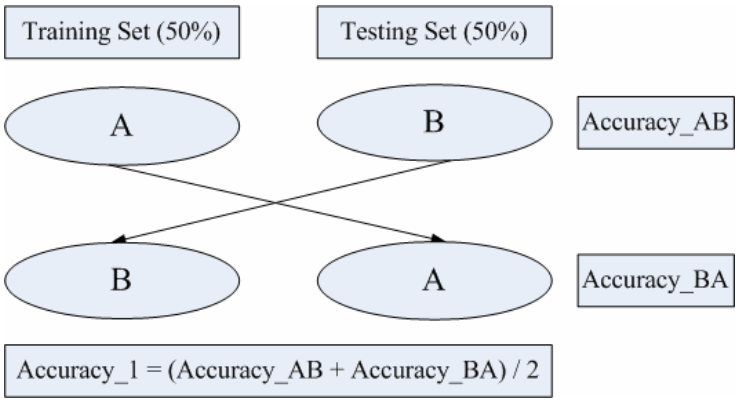


Researching Methodology

As mentioned above, we apply two methods to accomplish the experiment. The first method is implemented as follows. For the 251 samples of stocks, we divide them to two parts. Every part has the 50% of the samples. Besides, the way we divide the samples at random. The way of this method is cross-validation (Pandya, 1995). Firstly, we choose the first part as the training set, and the second part is the testing set. We produce a model by the training test, and then we predict the target value in the testing set. The experimental result is Accuracy_AB. On the contrary, in the second time, we choose the second part of the experimental samples as training set, the first part of the samples

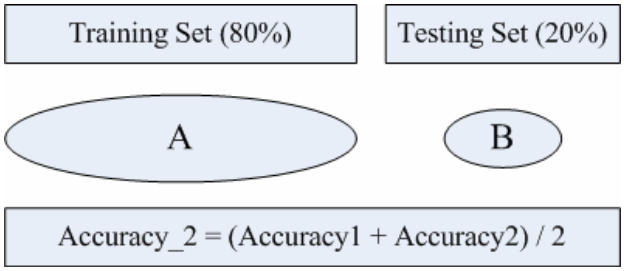
is testing set. Similarly, we can get the experimental results as Accuracy_BA. After that, we use the k-mean method to get the final accuracy rate which is called accuracy_1. This method is shown in Figure 2. The advantage of this method is that we can avoid the deviation of accuracy rate from only doing the experiment once, the experimental results will be more practical and dependable.

Figure 2: Mehod_1.



The different of the second method and the first one is the proportion of the two parts of samples. In the second method, we divide the sample as the eighty-twenty rules. It means, in the first part which is called part A is contented 80% of the samples. The second part which is called part B is contended 20% of the samples. We choose the part A as the training set, and choose the part B as the testing set. Also we divide the samples at random. In order to compare with the former method, we do the experiment by this way twice. We get the experimental results accuracy1 and accuracy2. Then, the accuracy_2 by the k-means is the final experimental result. This method is shown in Figure 3.

Figure 3: Method_2.



EXPERIMENT

In this paper, we applied the SVM to construct the prediction model and select Gaussian radial basis function (RBF) as the kernel function. Using SVM with financial statement analysis as mentioned above, we can calculate all the experimental results shown in Table 3. From these results, we can obviously find that after choosing some financial indices as the parameters, the accuracy rate is better than only predicting by the Outstanding Achievement Growth Rate (OAGR) given by the professionals and experts. Whatever we apply method_1 or method_2, the results of experiment 2, 5, 8, 7 and 6 are better than experiment 1 which is the basic experiment. That is to say, if we choose some financial indices correctly, we will improve the accuracy rate of prediction combined with the prediction results given by the experts such as the ranking of OAGR.

Table 3: Experimental results of different composition.

Experiment	SVM	OAGR	EPS	BVPS	NPGR	Accuracy_1	Accuracy_2
1	√	√				77.69%	81%
2	√	√	√			84.45%	86%
3	√	√		√		76.10%	79%
4	√	√			√	74.91%	85%
5	√	√	√	√		80.47%	86%
6	√	√	√		√	77.70%	86%
7	√	√		√	√	78.51%	85%
8	√	√	√	√	√	78.90%	85%

Table 4: Experimental results sorted by accuracy rate.

Method	**The number of the experiments (%)**							
Method_1	**2**	**5**	**8**	**7**	**6**	**1**	**3**	**4**
(50%-50%)	(84.45)	(80.47)	(78.90)	(78.51)	(77.70)	(77.69)	(76.10)	(74.91)
Method_2	**2**	**5**	**6**	**8**	**7**	**4**	**1**	**3**
(80%-20%)	(86)	(86)	(86)	(85)	(85)	(85)	(81)	(79)

In addition, we will find that in Table 4, there are some similarities after we sorted the results based on the accuracy rate. For example, the experiment 2 and the experiment 5 both have the best results of accuracy rate. That means, we select the Earnings Per Share (EPS) as the parameter, or we select both Earnings Per Share (EPS) and Book Value Per Share (BVPS) as the parameters. We can get the better results for our goal of prediction. In the future tasks of prediction, we only choose the Earnings Per Share (EPS) and Book Value Per Share (BVPS) as the parameters are enough. Meanwhile, we have to acknowledge that the results of experiment 3 or experiment 4 are somewhat worse than experiment 1. That is to say, it is not so much of only adding the Book Value Per Share (BVPS) or only adding the Net Profit Growth Rate (NPGR) as using nothing parameters. In that case, unless we have some special analysis, we had better not use these indices separately. However, after analyzing the results of experiment 5, 6, 7 and 8, we can obviously find that the results by combining those three financial indices in any patterns are better than experiment 1. Consequently, there are some association rules among these financial indices in the financial statement, because the financial indices are restricted relatively.

CONCLUSION

In this paper, we proposed a new method by using SVM with financial statement analysis for prediction of stocks. There has some advantages comparing our method to the common technical analyses for prediction in stock market. The experimental results show that there is a higher accuracy rate of prediction than using SVM to predict only, because the financial indices as the experimental parameters are directly selected from the financial statements which are released in different periods by the stock companies to public, and the experimental results are much more reliable, nonvolatile and valid. In addition, we can discover some association rules among the financial indices in the financial statement, and evaluate some special stocks, because the financial indices are restricted relatively. It may be useful in the practical of the advanced analysis of stock prediction. Based on this method, not only can we improve the accuracy rate of prediction, but also can we create an original and universal method for stock predictions, because the financial indices can be selected differently to meet the different stockholders' expectations. In that case, the stockholders are able to avoid the subjective discrepancies of prediction and make decisions to buy the target stocks.

However, each method has its own advantages and disadvantages. In this paper, we choose the Outstanding Achievement Growth Rate (OAGR) as the basic predictive standard which is evaluated by the professionals or experts, but there are some limitations for the method. Sometimes, we have to consider the data manipulated by people and the useless information from the professional analyzers' own perspectives. Besides, the ranges of the experimental samples are not large enough. In that case, we can not describe the overall status of the stock market so that the accuracy rate is going to be improved in further experiments. For the following research, we will still focus on the stock market. After comparing the experimental results with the practical situation, we will do some advanced studies based on this method.

REFERENCES

Burges, C. J. C., (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121-167.

Chang, C. C., & Lin, C. J. (2007). LIBSVM: a Library for Support Vector Machines. Available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Chen, W. H., & Shih, J. Y., (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, 30, 427-435.

Huang, C. L., & Tsai, C. Y., (2006). Using SOM-SVR with Filter Feature Selection for Prediction of Taiwan Stock Index Future, CSIM IMP 2006 Taiwan.

Huang, W., Nakamori, Y., & Wang, S. Y., (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* 32, 2513-2522.

Hsu, C. W., Chang, C. C., & Lin, C. J., (2003). A Practical Guide to Support Vector Classification. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm.

Keerthi, S. S., & Lin, C. J., (2003). Asymptotic Behaviors of Support Vector machine with Gaussian Kernerl. *Neural Computation,* 15, 1667-1689.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55,307-319. Available at http://www.ComputerscienceWeb.com.

Müller, K. R., Smola, S. A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V., (1997). Predicting time series with support vector machines, in: W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (Eds.), *Proceedings of ICANN'97*, Spring LNCS1327, Berlin, 999-1004.

Pandya, A. S., & Macy, R. B., (1995). *Pattern Recognition with Neural Networks in C++*. IEEE Press.

Smola, A. J., & Schölkopf, B., (1998). A tutorial on support vector regression. *NeuroCLOT Technical Report* NC-TR-98-03, Royal Holloway College, University of London, UK. Available at http:// www.kernel-machines.org.

Thissen, U., Brakel, R. V., Weijer, A. P., Melssen, W. J., & Buydens, L.M.C., (2003). Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems,* 69, 35- 49.