ISSN No. 0976-5697

Volume 7, No. 7, Nov-Dec 2016

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Stock Market Prediction by Non-Linear Combination based on Support Vector Machine Regression Model

K. Ashwin Kumar, T. Niranjan Babu, Nitish Vaishy, K. Lavanya

School of computer science & Engineering, VIT University, Vellore-632014, Tamilnadu, India

Abstract: Stock market predictions comprise challenging applications of modern time series forecasting and are essential to the success of many businesses and financial institutions. In this paper, stock market forecasting is based on Support Vector Machine (SVM) regression. Firstly, using different linear regression model to extract linear characteristics of stock market system. Secondly, using different Neural Network algorithms to extract nonlinear characteristics of stock market system. Finally, the SVM regression is used for the nonlinear combination forecasting model of different stock exchange prices. Empirical results obtained reveal that the prediction by using the nonlinear combination model is generally better than those obtained using other models presented in this study in terms of the same evaluation measurements. Those results show that that the proposed nonlinear modeling technique is a very promising approach to financial time series forecasting.

Keywords: Linear Regression, Neural Network, Support Vector Machine, Forecasting.

1. INTRODUCTION

Effective data analysis and forecasting plays an importantrole in the field of financial investment. World financialmarkets function in a very complex and dynamic mannerwhere high volatility and noisy data are routine. Due to the high degrees ofirregularity, dynamic manner and nonlinearity, it is extremelydifficult to capture the irregularity and nonlinearity hiddenin financial time series by traditional linear models suchas multiple regression, exponential smoothing, autoregressiveintegrated moving average, etc [3], [4].

Recently, more forecasting models have been developed to improve prediction accuracy in stock market prediction such as, Neural Network (NN), Support Vector Machine (SVM) [5], [6]. Every single model has its own unique ability to derive and interpret results under different criterion. Nevertheless, dealing with stock market, it is not adequate to predict with only one particular forecasting model. As a result, employing the combination forecasting method is essential.

2. STATE-OF-ART-STUDY

The main purpose of the combination forecasting methodis to utilize the information provided in various models comprehensivelyand to increase prediction accuracy as much aspossible. In many areas, different combination forecasting hasbecome the critical way to improve the prediction, especiallyin economics, management and statistical research [7], [8].It has been proved that the combination of a variety offorecasting models under certain conditions can develop themodel fitting ability and improve the prediction accuracyeffectively. At present, the research of the combination forecastingmodel focuses on two aspects. One is how to generatethe combination model, the other is how to combine thevarious sub-models and derive the final conclusion. The generationof sub-model gives first place to linear or nonlinearregression, while the conclusion derivation depends uponlinearity combination [9].

In this paper, a novel method is presented for the studyof different stock market series (S&P 500, Dow Jones Industrial Average, NASDAQ, Prime Interest Rate) basedon SVM Regression combination model (SVR–CM) linearregression with nonlinear regression. Firstly, four linear regressionare used to extract linear features of the stock marketsystem. Secondly, four different NN algorithm are used toextract nonlinear features of the stock market system. Finally,SVM Regression is used to combine all output results. Therest of this study is organized as follows. Proposed approach describesthe building process of the combination of forecasting modelbased on SVM in detail.

3. EXISTING APPROACH

NeuroEvolution of Augmenting Topologies (NEAT) is an efficient genetic algorithm which is capable of evolving structures at the same time as weights. It was developed by Kenneth O. Stanley at the University of Texas [1]. It uses historic markings in order to detect homology between genes. It should protect innovation through speciation and minimize structure in order to minimize the dimensions being searched by starting out with a minimum topology namely no hidden nodes. In order to evolve structure a flexible encoding is required. Because NEAT adds structure to the starting topology a encoding must be dynamic and expandable.Mutation in NEAT can change both connection weights and network structures. Each connection weight is perturbed with a fixed probability by adding a floating point number chosen from a uniformdistribution of positive and negative values.

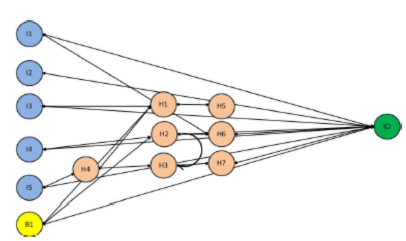

Fig 1: NEAT Champion Topology

Topological complexificationhappens in two ways. Each mutation expands thesize of the genome by adding genes. In the add connectionmutation, a single new connection gene is added connectionis split and the new node placed where the old connectionused to be. The old connections are disabled and the newones are added to the genome. Crossover must be able torecombine networks with different topologies, this process is made difficult by the dynamic and expandable nature of thetopology encoding [2].

4. PROPOSED APPROACH

SVM was a significant result of machine learning researchin recent years, which has been introduced by Cortes andVapnik in 1995 [10]. It was developed on the foundationof small samples statistical learning theory that proposed byVapnik etc., and its algorithm is based on the structural riskminimization principle [11]. Compared with the traditionalneural network, the support vector machine not only simple instructure but also all sorts of technical performances are betterthan neural network obviously, which were testified by lots ofexperiments. Originally, SVM has been presentedto solve pattern recognition problem, However, with theintroduction of Vapniks ε-insensitive loss function, SVMhas been developed to solve nonlinear regression estimationproblems, such as new techniques known as support vectorregression (SVR),which have been shown to exhibit excellentperformance. At present, SVR has been emerging as analternative and powerful technique to solve the nonlinearregression problem. It has achieved great success in bothacademic and industrial platforms due to its many attractivefeatures and promising generalization performance.

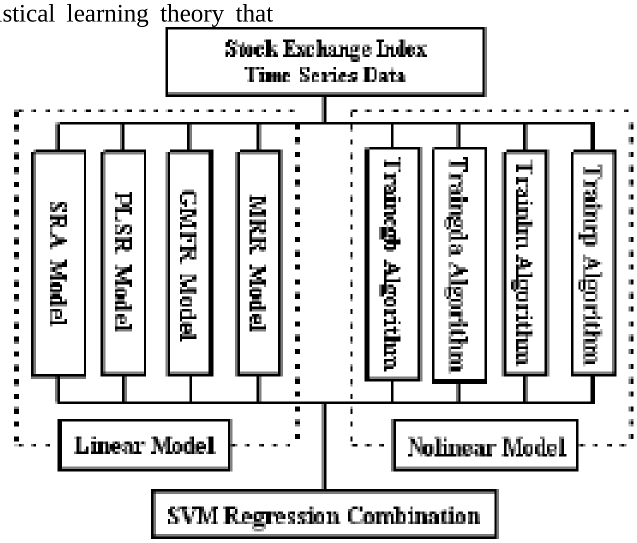

Fig 2A Flow Diagram of the proposed SVR-CM Model

Statistical models are established as the regression modeland regression coefficients are very significant, and each ofthe neural network structure adopted the $n - n - 2$ form.Namely, the number of the input layer nodes is the sameas the hidden layer nodes and the nodes of input layerare determined by the input variables, the output is thestock market's opening price and closing price.The forecasting model of the SVR–CM can be summarizedas follows:

(1) The four regression models are used to extract linear character.

(2) The four from different neural networkalgorithms are used to nonlinear character.

(3) The finalpredictive output is generated by the SVR.

5. PERFORMANCE ANALYSIS

The performance analysis shows that the performance of SVM-CM model is better than the NEAT genetic algorithm in terms of training error and evaluation error.

Table1: RMS Error by NEAT and SVM on different Hidden units

Hidden units	RMS Error by NEAT	RMS Error by SVM
10	0.63	0.34
15	1.24	0.84
30	1.73	0.96
45	2.12	1.12
60	2.48	1.24
100	2.94	1.52
120	3.26	2.15
140	3.32	2.26
150	3.45	2.42
200	3.69	2.78

Table 1 demonstrates the RMS error of different hidden units. While calculating the errors by different algorithms it has been shown that SVM RMS error is less when compared with NEAT RMS error.

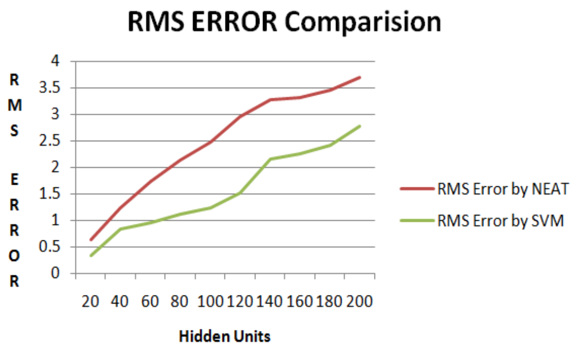

Fig 3: RMS ERROR Comparision on different hidden units

4 outputs correspond to each of indexes on the input (*S&P500*, *DOW*, *NASDAQ Composite*, and *Prime Interest Rate*). The neural network job will be to find hidden patterns in the input data which influences the overall output. After training the network using 40-41-41-4 topology (40 input units, 2 hidden layers with 41 units, 4 outputs), and trying to predict the values, the following results have been obtained:

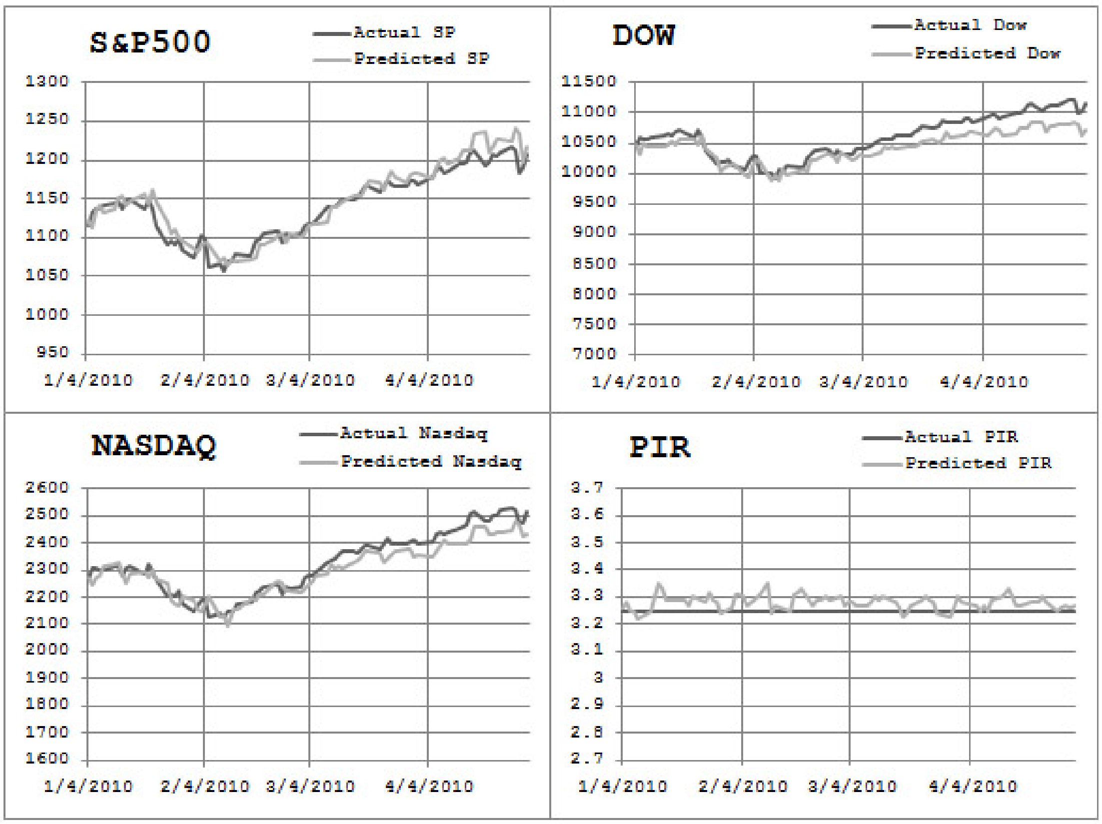

Fig4:Comparision Graph of Actual and Predicted Stock Market Prices

6. CONCLUSION

In this paper, fourkinds of different linear regression models are used to extractthe linear features of the market system, and four kinds of differentneural network algorithms are used to extract the non-linear features of the market system. Two groups of prediction individualsare combined with support vector machine regressiontime Examplesof calculation shows that the method can significantlyimprove the system's predictive ability, prediction accuracy,and with a high prediction accuracy of the rising and fallingtrend of the stock market. Empirical results obtained revealthat the proposed nonlinear combination technique is a verypromising approach to financial time series forecasting.

7. REFERENCES

[1] K. O. Stanley, B. D. Bryant and R. Miikkulainen, Real-timeneuroevolution in the NERO video game, IEEE Transactionson Evolutionary Computation, 2005.
[2] T. Hellstrom and K. Holmstrom, Predicting the Stock Market,Technical Report IMa-TOM-1997-07, Center of MathematicalModeling, Department of Mathematics and Physis, MalardalenUniversity, Vasteras, Sweden, August 1998.
[3] P. Z. Guoqiang, " An investigation of neural network for linear timeseries forecasting," Computers & Operations research, Vol. 28, pp.1183–1202, 2001.
[4] G. P. Zhang, " Time series forecasting using a hybridARIMA and neural network model," Neurocomputing, Vol. 50, pp. 159–175, 2003.
[5] S. Q. Ma, G. Ma, " Stock market risk, income and market efficiency: ARMA–ARCH–M model," World Economy, Vol.5, pp.19–28, 2000.

[6] K. J. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, Vol. 55, pp. 307–319, 2003.
[7] F. Y. Wei , H. X. Cao, "Mathematical models of long-term prediction and its application," Beijing: Meteorological Press, pp.9–12, 1990.
[8] F. M. Tseng, H. C. Yu and G. H. Tzeng, "Combining neural network model with seasonal time series ARIMA model," Technological Forecasting and Social Change, Vol. 693, pp. 71–87, 2002.
[9] Y. K. Ma , X. W. Tang, " Research on the problem of optimizing linear combination forecasting model," Systems Systems Engineering Theory & Practice, Vol.18, pp.110–114,123, 1998.
[10] C. Cortes , V. Vapnik, "Support vector networks," Machine Learning,Vol.20, pp.273–295,1995.
[11] P. L. Bartlett, J. S. Taylor, "Generalization performance on supportvector machines and pattern classifiers," B. Sholkopf, C. Burges, andA. Smola Eds., Advances in Kernel Methods-Support Vector Learning,Cambridge, MA: MIT Press, 1999.