2011 Fourth International Conference on Business Intelligence and Financial Engineering

Price Prediction of Stock Index Futures Based on SVM

Hezhong Dai, Yichi Zhang
Dongling School of Economics and Management
University of Science and Technology Beijing, Beijing, 100083, China
Email:daihezhong@hotmail.com

Dapeng Wang
Department CECIC
Research Center of Asset Management
Beijing, China
Email: jerrywangustb@yahoo.com

Abstract—Though accurately forecasting the price of stock index futures is impossible, it is of great significance if the price's variation trend can be estimated to a certain extent. In this paper, we adopted a Support Vector Machines method to predict the prices of Stock index futures in the next 5 trading days. First, with an information granulation method, the original data of 3 stock index futures were transformed into a series of fuzzy granules. Then the maximum, medium and minimum values of futures' opening price in each single granule are all extracted. After utilizing the SVM model to regress the values in fuzzy granules, we came up with the variation range of futures' price in the next few days. These predicted results are consistent with the actual one, which proves the feasibility of our method.

Keywords-stock index futures; SVM; price forecast; information granulation

I. INTRODUCTION

Price forecast plays an important role in the financial market. On the basis of price information, the investors can adjust their investing strategies to gain more profit and hedge the bidding risks [1]. On April 16, the China Financial Futures Exchange launched the first stock index futures in mainland China. It is a milestone in Chinese financial market and has paved way for the introduction of other derivatives and financial products. Considering that the trade of futures is quite risky, finding the variation rule of futures' price and thus mitigate risks caused by markets' fluctuation become an an important job [2,3,4].As is known, price forecast on the stock index future market is a hard task due to the high pool price volatility. The dynamism of the market is complicated and often considered as random, especially in a short-term. Nevertheless, from an overall and long-term view, the price of stock index future still demonstrates a certain level of forecast capability [5].

Nowadays, many researchers have made contributions in analyzing the price movement and developed many methods applied in price prediction, such as the ARCH models, ARMA models and the Genetic Algorithm. Auto Regressive Conditional Heteroskedasticity (ARCH) models were firstly introduced by Robert Engle in 1982 [6]. They have been widely utilized in modeling financial time series. The time series exhibit time-varying volatility clustering which means the periods of relative clams following periods of swings. Also, the auto regressive moving average (ARMA) models were reported by George Box and Gwilym M. Jenkins [7,8]. They are also called Box-Jenkins models and are typically used to auto-correlated time series data. Additionally, the Genetic Algorithm (GA) is often utilized to reduce the dimension of the input vector and aimed at ensure prediction accuracy by selecting better model parameters [9].

In this paper, we adopted another new neural network model, the SVM, to predict the price variation in the stock index futures' market.

II. SVM & INFORMATION GRANULATIONI.

A. Theory of SVM

Developed by Vladimir Vapnik in 1995, the Support Vector Machine (SVM) is regarded as one of the most important breakthroughs in the field of Machine Learning and can be applied in both classification and regression [10,11,12]. The goal of SVM modeling is to select the optimal hyperplane in a high dimensional space ensuring that the upper bound of the generalization error is minimal. Additionally, though the SVM can only directly deal with linear samples, mapping the original space into a higher-dimensional space can make the analysis of non-linear samples possible [13,14].

Given a set of data points (xi,yi), （i=1,2,3...n, $xi \in R$, $yi \in \{-1,1\}$,） randomly and independently generated from an unknown function, SVM approximates the function in the following form : $g(x)=w\phi(x)+b$. $\phi(x)$ is the feature and nonlinear mapped from the input space x. w and b are both coefficients and can be estimated by minimizing the regularized risk function.

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L(d_i, y_i) + \frac{1}{2}\|\omega\|^2 \quad (1)$$

$$L(d,y) = \begin{cases} |d-y| - \varepsilon & |d-y| \geq \varepsilon, \\ 0 & other, \end{cases} \quad (2)$$

Where both C and ε are prescribed parameters. C is called the regularization constant while ε is referred to as the regularization constant. $L(d,y)$ is the intensive loss function and the term $C\frac{1}{N}\sum_{i=1}^{N} L(d_i, y_i)$ is the empirical error while the $\frac{1}{2}\|\omega\|^2$ indicates the flatness of the function. C

978-0-7695-4527-1/11 $26.00 © 2011 IEEE

DOI 10.1109/BIFE.2011.96

measures the trade-off between the empirical risk and the flatness of the model.

By introducing he positive slack variables ζ and ζ^*, equation 2 can be transformed to the following constrained formation:

$$R(w,\zeta,\zeta^*) = \frac{1}{2} ww^T + C^* (\sum_{i=1}^{N} (\zeta,\zeta^*)) \qquad (3)$$

Subjected to :

$$w\phi(x_i) + b_i - d_i \le \varepsilon + \zeta_i^* \qquad (4)$$

$$d_i - w\phi(x_i) - b_i \le \varepsilon + \zeta_i \qquad (5)$$

$$\zeta_i, \zeta_i^* \ge 0 \qquad (6)$$

Finally, after the Lagrange multipliers are introduced and the optimality constraints exploited, we can come up with the decision function (kernel function) and can be expressed in the following form:

$$f(x) = \sum_{i}^{l} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \qquad (7)$$

α_i^* are Lagrange multipliers. They satisfy the equalities $\alpha_i \times \alpha_i^* = 0, \alpha_i \ge 0, \alpha_i^* \ge 0$

The value of the Kernel is the same with the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$ .that is . Functions that satisfy

Mercer's condition can be regarded as the Kernel function. The Gaussian Kernel function, $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2))$ is an example. As Gaussian Kernels deliver good performances under general smoothness assumptions, the SVMs can be used to estimate the non-linear behavior of the forecasting data set.[15~18]

B. Introduction of information granulation

The concept of information granulation was first introduced by Professor A. Zadech in 1979 [19]. In his paper, the method employed is the fuzzy granulation, a primary method of information granulation.

granule in each window. The key in the above process is that the fuzzy granular time series should optimally fit the original time series. Under many conditions, the fuzzy granule can be expressed as follows [20]:

$$g \underline{\underline{\triangle}} x \text{ is } G \text{ is } \lambda \qquad (8)$$

The fuzzy granulation method includes two steps: first is to divide given time series into a set of windows based on the the condense degree of data, and then is to create a fuzzy

G is a fuzzy subset of U and U refers to a Universe of Discourse. x is the parameters in the set U. x is a variable in the set U. The qualifyer λ denotes a fuzzy probability.

During the process of fuzzy granulation, the type of the fuzzy subset should at first be defined. Meanwhile the fuzzy membership functions can also be determined. Under most conditions, the fuzzy subset can be replaced by the fuzzy granule. The primary types of fuzzy granule include the triangular, Gaussian and parabolic types. The type we employed in this paper is the triangular fuzzy granule.

III. EXPERIMENTS & RESULTS

A. Experimental design

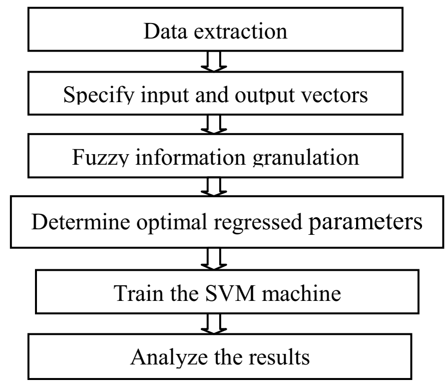


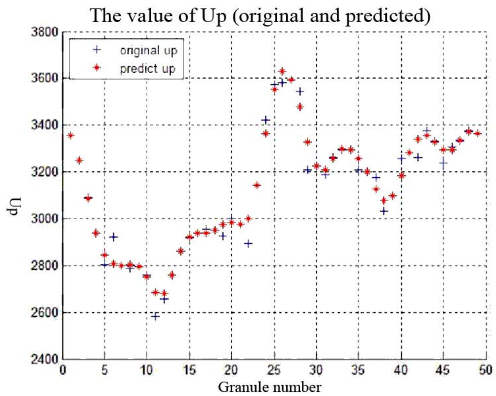
Figure 1. Flow chart of the SVM prediction

Establishing SVM models can recognize the variation rule of the price of the stock index future. Also, it is possible to predict the price in the near future based on the current data. Three stock index future's data in mainland China were examined in the experiment. They are the Shanghai Shenzhen CSI 300 Index (from Apr. 19, 2010 to Dec. 31, 2010), IF00 Contract (from Apr. 19, 2010 to Apr. 22, 2010) and IF1009 contract (from Apr. 19, 2010 to Sept. 10, 2010).

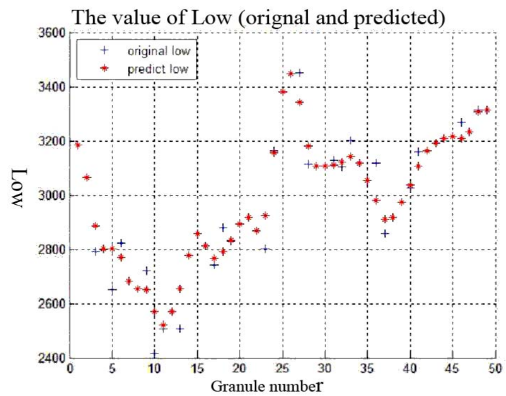
The daily opening price of stock index futures everyday is affected by the daily opening price, daily maximum, daily minimum and daily closing price. So, we used these influential factors as input vectors and the daily opening price as the output vector.

The original data can be transformed into a series of granules with the fuzzy granulation method. We selected the triangular fuzzy granulation model to process the original data. The granulation was accomplished by applying the function: [Low, R, Up]=FIG_D(ts',' triangle', win_num).

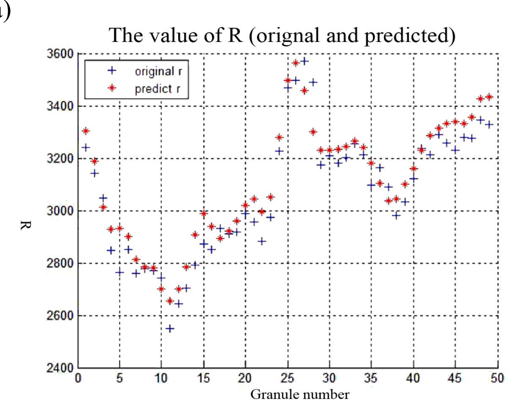
Ts is the time series waiting for granulation. Triangle indicates the triangular type granules are adopted. Win_num reveals the number of granules. The original data were divided into many windows and each window would generate one granule. The Low, R and Up mean the minimum, mean and maximum value of futures' prices in the every granule respectively.



(a)



(b)



(c)

Figure 2. The original and predicted value of IF100: (a) Low, (b)R, (c)Up

After the fuzzy granulation of original data, we used the SVM model to regress the granules' parameters Low, R and Up.

There are three main steps. First, the data should be preprocessed to normalize into a certain range [100, 500]. Second, search and select the optimal parameters by the SVMcgRorRegress function. Third, train the SVM model and complete the prediction.

B．Experimental results and discussions

After the training of SVM model, we came up with the value of granules' parameters: Low, R, Up. The original and predicted values of IF00 is indicted in Figure 2.

Based on the above data, the granules' parameter, Low, R and Up, in the following 5 trading dates could be predicted.

As Low, R and Up mean the minimum, medium and maximum values in each granules, the variation range of stock index futures' opening price can be expressed by [Low, Up and R]. By comparing the predicted and actual variation range in the next 5 days, both the variation trend and the prediction accuracy can be analysed.

TABLE I. VARITATION RANGE OF CSI 300'S OPENNING PRICE

Date	10/12/27	10/12/28	10/12/29	10/12/30	12/12/31
Opening Price	3222.68	3178.87	3176.41	3313.49	3388.28
Variation range	[2961.4，3162.9，3323.8]				
Date	11/1/4	11/1/5	11/1/6	11/1/7	11/1/10
Opening Price	3155.55	3170.18	3177.83	3156.36	3162.08
Predicted Variation range	[3071.3，3166.7，3270.3]				

From above table, it can be seen that the prediction of the CSI 300 future' price in 2011's first 5 days is reasonable. Compared with the former 5 trading days, the opening price in the latter 5 days had a relatively small variation range and downward trend.

TABLE II. VARITATION RANGE OF IF100'S OPENNING PRICE

Date	11/4/18	11/4/19	11/4/20	11/4/21	11/4/22
Opening Price	3368.0	3343.4	3318.8	3328.6	3323
Variation range	[3299，3348，3381.9]				
Date	11/4/25	11/4/26	11/4/27	11/4/28	11/4/29
Opening Price	3302.0	3256.0	3258.2	3238	3191.2
Predicted Variation range	[2929.6，3011.4，3036.2]				

The prediction of IF00's opening price in 5 trading days in 2011 is consistent with the true price. The variation range of opening price in the second trading week is larger than that in the first trading week. The trend went down in general.

TABLE III. VARITATION RANGE OF IF1009'S OPENNING PRICE

Date	10/9/6	10/9/7	10/9/8	10/9/9	10/9/10
Opening Price	2948.0	2990.0	2987.0	3004.0	2955.0
Variation range	[2851.8，2916.4，2934]				
Date	10/9/13	10/9/14	10/9/15	110/9/16	10/9/17
Opening Price	2952	2969.8	2968.8	2920	2867.6
Predicted Variation range	[2929.6，3011.4，3036.2]				

According to the above analysis, the opening price of IF1009 in the second trading week varied in a way different from that of the first week. As was predicted, the variation range of the second week's price should be small and there should be an upward trend. However, it turned out that the price varied in a more radical way and went downward generally.

II. CONCULSION

Predictions are prerequisites before decisions or plans are making. With the foundation and development of stock index futures' market, the prediction of futures' price attracts the attention from researchers and investors. The stock index futures' market is a complicated and non-linear one whose prices present strong inaccuracy. However, by introducing the SVM model, it is possible to forecast the future on a certain level. As the futures' market is sensible to various factors at home and abroad, the accuracy of the prediction will be mitigated by all kinds of information. To reduce such mitigation, we used an information granulation model to process the original data and then built an SVM model to complete the data regression. After the price's variation ranges were calculated, the prediction's accuracy was discussed after comparing the predicted results and the actual results.

ACKNOWLEDGMENT

Many thanks for the China Scholarship Counsil to sponsor this project [2008104351].

REFERENCES

[1] Y. B. Xu and K. Nagasaka, “Demand and Price Forecasting by Artificial Neural Networks (ANNs) in a Deregulated Power Market,” International Journal of Electrical and Power Engineering, vol. 3, pp. 268-275, 2009.

[2] J. M. Poterba and L. H. Summers, “The persistence of volatility and stock market fluctuations,” American Economic Review, vol.76, pp. 1142-1151, 1986.

[3] K. R. French, G. W. Schwert, and R. Stambaugh, “Expected stock returns and volatility,” Journal of Financial Economics, vol. 19, pp. 3-29, 1987.

[4] K. A. Chan, “further analysis of the lead-lag relationship between the cash market and stock index futures market,” Review of Financial Studies, vol. 5, pp. 123-151, 1992

[5] X. Y, Wang and Limin Lin, “Short-term Prediction of Shanghai Composite Index Based on SVM”, Softerware Engineering and Service Sciences (ICSESS), 2010 IEEE International Conference.

[6] R. F. Engle, “Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation,” Econometrica, vol. 50, pp. 987-1008, 1982.

[7] G.Box, G.M. Jenkins, and G. C. Reinsel. Time Series Analysis: Forecasting and Control, third edition. Prentice-Hall, 1994.

[8] J. L .Torres，A. Gareia, and M.De Blas, etal, “Forecast of hourly average wind peed with ARMA models in Navarre,” Solar Energy, vol. 79, pp. 65-77, 2005.

[9] L. Yu, S. Wang, K. K. Lai, “Mining stock market tendency using GA based support vector machines,” Lecture Notes in Computer Science, vol. 3828, pp. 336－345

[10] S. R. Gunn, “Support Vector Machines for Classification and Regression”, ISIS Technological 1998, Dept. of Electronics and Computer Science, University of Southampton.

[11] Corinna Cortes and V. Vapnik, “Support-Vector Networks” Machine Learning, vol.20, pp. 273--297, 1995

[12] SVM - Support Vector Machines, Introduction to Support Vector Machine (SVM) Model http://www.dtreg.com/svm.htm

[13] V. Vapnik, N. Golowich and S. Smola, “A support vector method for function approximation,” Regression Estimation and Signal Processing, in Proceeding of NIPS, pp.281~287, 1996.

[14] F. E. H. Tay, L. J. Cao, “Application of support vector machines in financial time series,” Forecasting omega, vol.2001, pp. 309-317, 2001.

[15] L. J. Cao and Francis E. H. Tay, “Support Vector Machine with adaptive parameters in financial time series forecasting,” IEEE Transactions on Neutral Networks, vol.6, pp. 1506-1518, 2003.

[16] L. J. Cao and F. E.H. Tay, “Financial Forecasting Using Support Vector Machines,” Neural comput & applic, vol.10, pp.184－192, 2001

[17] T. B. Trafalis and H. Ince, “Support vector machine for regression and applications to financial forecasting,” Conference: IJCNN - IJCNN , pp. 348-353, 2000

[18] Zan Huanga,*, Hsinchun Chena and Chia-Jung Hsua, “Credit rating analysis with support vector machines and neural networks: a market comparative study,” Decision Support Systems, vol. 37, pp. 543－558, 2004

[19] L. A. Zadeh, Fuzzy Sets and Fuzzy Information Granulation, Beijing Normal University Press, 2000.

[20] L. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic Fuzzy Sets and Systems”, vol. 90, pp.111-127, 1996.