

Research Article

The Interval Slope Method for Long-Term Forecasting of Stock Price Trends

Chun-xue Nie and Xue-bo Jin

School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

Correspondence should be addressed to Xue-bo Jin; jinxuebo@btbu.edu.cn

Received 3 November 2015; Accepted 3 February 2016

Academic Editor: Doojin Ryu

Copyright © 2016 C.-x. Nie and X.-b. Jin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A stock price is a typical but complex type of time series data. We used the effective prediction of long-term time series data to schedule an investment strategy and obtain higher profit. Due to economic, environmental, and other factors, it is very difficult to obtain a precise long-term stock price prediction. The exponentially segmented pattern (ESP) is introduced here and used to predict the fluctuation of different stock data over five future prediction intervals. The new feature of stock pricing during the subinterval, named the interval slope, can characterize fluctuations in stock price over specific periods. The cumulative distribution function (CDF) of MSE was compared to those of MMSE-BC and SVR. We concluded that the interval slope developed here can capture more complex dynamics of stock price trends. The mean stock price can then be predicted over specific time intervals relatively accurately, in which multiple mean values over time intervals are used to express the time series in the long term. In this way, the prediction of long-term stock price can be more precise and prevent the development of cumulative errors.

1. Introduction

Stock price is a typical and complex type of time series data. The prediction of stock prices has been an active area of research in econometrics, signal processing, pattern recognition, and machine learning for some time. Stock traders and investors are extremely interested in stock market prediction because of the considerable profits that can be reaped by trading stocks. Traditionally, the basic methodology for financial time series has been statistical methods such as autoregressive and moving average model (ARMA), autoregressive integrated moving average model (ARIMA), and generalized autoregressive conditional heteroskedasticity (GARCH), which require the linear variation in the stock prices to remain stationary. In general, the statistical models cannot adapt to changes in the process. Accordingly, traditional statistical methods cannot predict stock performance very well when tracking the complexity of the stock markets [1].

Recently, many machine learning systems have been used to predict stock prices. These include artificial neural net (ANN) [2, 3], Bayes networks [4], genetic programming [5],

support vector regression (SVR) [6], user analysis [7], sentiment analysis [8], and hybrid networks [9–15]. Accordingly, machine learning methods can be used to track the complexity and nonstationary nature of the stock markets in short-term prediction. These methods predict long-term trends only with great difficulty. Existing methods of long-term stock prediction mainly include the following: use of the recursive iteration prediction to obtain the long term prediction trend [12]; however, this method involves accumulative error, and the cumulative error increases with the number of steps in the prediction process. By using a moving window algorithm to delete older data and take in new data, the prediction model can be updated in sequence [6]. The length of the moving window also has considerable influence on the accuracy of the modeling process. This system can only show the mean stock price during the prediction interval and cannot show the details of changes in the stock trend during this interval. In addition, the present methods use the mean value directly as a feature of the stock trend prediction. Regarding the fluctuations in larger time series, the mean values of the interval weaken the fluctuation characteristics of the time series and reduce the long-term accuracy of the forecast.

For these reasons, the prediction of stock price is still a worthwhile issue.

Long-term time series forecasts have other applications, such as the host load prediction. Load prediction is crucial to efficient resource utilization in dynamic cloud computing environments. Di et al. used the exponentially segmented pattern (ESP) [16] to predict the host load in the cloud. They proposed the use of 9 different features to characterize the recent load fluctuation in the evidence subinterval. They were able to predict the mean load over consecutive time intervals.

In this paper, the exponentially segmented pattern (ESP) is used to predict the fluctuation of stock price over consecutive future time intervals. While we give a new feature of stock price in the subinterval, namely, interval slope to characterize the stock price fluctuation over a set period. The interval slope can be used to determine the mean of stock in the subinterval. The support vector regression and the Bayes classifier were used to predict the stock price trend and verify the effectiveness of the interval slope of the stock price in the subinterval.

In this paper, the following contributions are made:

- (i) The exponentially segmented pattern (ESP) is here used to predict fluctuations in different stock data over a long period and can accurately predict not only mean stock price over a future time interval but also the mean stock price over consecutive future time intervals. In this way, the prediction of long-term stock price can be more precise and the generation of cumulative errors can be prevented.
- (ii) The use of new features of stock pricing in the subinterval, namely, interval slope, is here proposed to better characterize the stock price fluctuation over some time period.

The rest of the paper is organized as follows. In Section 2, the long-term stock price prediction model is introduced. In Section 3, experiments and comparisons of different models are made. Conclusions are given in Section 4.

2. Model-Based Prediction of Long-Term Trends in Stock Price

The predictive objective is to predict the fluctuation of opening price over a long period. Multiple precise mean values over time interval are used to express the time series long-term trend.

The proposed stock price trend prediction model involves the following three steps: first, using the ESP principle, the estimated data segment is split into a set of consecutive segments, whose lengths increase exponentially. The interval slope is used to describe the features of each interval. Then, machine learning methods, SVR and MMSE-BC, were used to produce the transformation model of the data which is used to predict the mean stock price for the next interval. Multiple precise mean values over time interval are used to express the long-term trends in the time series. In this way, the prediction of stock price in the long term can be performed precisely without generating cumulative errors.

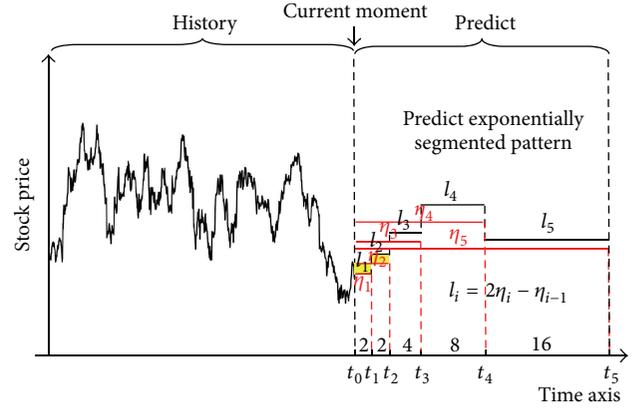


FIGURE 1: Illustration of ESP and the relation of l and η .

2.1. Exponentially Segmented Pattern (ESP) and Transformation of Segments. The objective of the current work is to predict trends in the patterns of fluctuation of stock price over the consecutive future time intervals. The most important step of the proposed stock price trend pattern prediction is that the estimated data segment is split into a set of consecutive segments by ESP principle, whose lengths increase exponentially. An example of ESP is shown in Figure 1. At a current time point t_0 , the estimated data segment is split into a set of consecutive segments whose lengths increased exponentially. The length of each following segment was 2^i , where $i = 1, 2, 3, 4 \dots$. For each segment over the consecutive future time intervals, the mean values were denoted by l_i , where $i = 1, 2, 3, 4 \dots$.

However, the mean stock price over the consecutive time intervals l_i is hard to predict, and the mean stock price over a single future time interval is easy to predict. The length of each following segment is $t_0 + 2^i$, where $i = 1, 2, 3, 4 \dots$. The mean predicted stock price of each time segment is given here as η_i , where $i = 1, 2, 3, 4 \dots$.

A set of the mean stock prices for a single time interval is then available. The aim of the current work was to predict the mean stock price over the consecutive future time intervals (l). In fact, the vector l can be converted from the vector η through the following induction according to a previous work [16].

Suppose that the current moment is t_0 , and the user has already predicted two mean stock prices (shown by η_{i-1} and η_i , the solid red line segment) over two different intervals ($[t_0, t_{i-1}]$ and $[t_0, t_i]$). Then, for the two shaded areas, which are of equal size, the mean stock price in $[t_{i-1}, t_i]$ can be derived. The transformation is given in

$$l_i = \eta_i - \frac{t_{i-1} - t_0}{t_i - t_{i-1}} (\eta_i - \eta_{i-1}). \quad (1)$$

Here, l_i is the predicted mean stock price in the new segment $[t_{i-1}, t_i]$ corresponding to the black line segment in Figure 1.

Taking into account $t_i = 2t_{i-1}$ and $t_0 = 0$, (1) can be simplified further, producing

$$l_i = 2\eta_i - \eta_{i-1}. \quad (2)$$

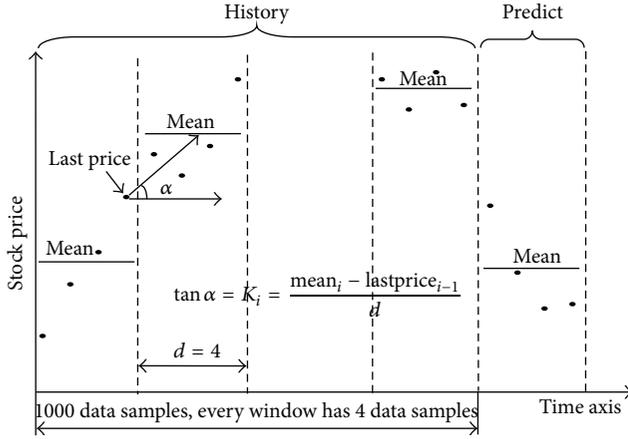


FIGURE 2: Illustration of features for stock price fluctuation ($d_j = 4$ days).

2.2. Features of Fluctuations in Stock Price. The aim of the current work is to predict the mean stock price over a future time interval (η) starting at the current time t_0 . Here, every future time interval is called a subinterval. Based on the features of stock price trend, a new feature of the stock price in the subinterval was proposed here. This feature is here called the interval slope, and it can be used to characterize the fluctuations in stock price fluctuation over a specific time period. The time series is denoted in the subinterval as $\mathbf{e} = (e_1, e_2, \dots, e_i)^T$, where $i = 1, 2, 3, 4, \dots$, and e_i is the sample stock price in the subinterval. For example, the subinterval is 4 days (i.e., $i = 4$), whose features for stock price fluctuation are shown in Figure 2.

Interval Mean Price. meanprice_j is the mean stock price when the conditions of the subinterval equal d_j :

$$\text{meanprice}_j = \frac{1}{d_j} \sum_{i=1}^{d_j} e_i. \quad (3)$$

Interval Last Price. The last price is the most recent price value in the subinterval.

Interval Slope (K). K_j interval slope is the slope of linear equation in the condition of the subinterval equals d_j .

First, the last price and the mean price of the subinterval were computed. Then, the linear equation $y = kx + b$ was used to fit the samples in the subinterval; that is,

$$\text{meanprice}_j = K_j d_j + \text{lastprice}_{j-1}. \quad (4)$$

This produces the following:

$$K_j = \frac{\text{meanprice}_j - \text{lastprice}_{j-1}}{d_j}. \quad (5)$$

Interval slope can be transformed into mean by (5).

For example, the value of d_j can be set as follows: $d_j = 2, 4, 8, 16, 32$. The feature of stock price fluctuation based on $d_j = 4$, history data, and predicted data are shown in Figure 2.

Next, future data features were predicted (the interval mean and the interval slope) through learning history data features. The prediction methods are presented in the next section.

2.3. Long-Term Forecasting Based on Interval Slope. To verify the effectiveness of the developed feature and assess the interval slope of stock price in the subinterval, the support vector regression and the Bayes classifier were used for long-term forecasting. Two machine learning methods, SVR and MMSE-BC, are used to produce the transform model of the data and the mean stock price is used to predict the next interval slope.

(1) Support Vector Regression (SVR). The aim of SVR algorithm was to minimize ε -sensitive errors on the subset of data, here called the support vectors. SVR algorithm uses nonlinear kernel functions in order to project initial data to a higher dimensional space and project linear classifiers from the higher space to the original space. The formulation of SVR is represented as follows:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{l} \sum_{i=1}^l L_\varepsilon \quad (6)$$

$$L_\varepsilon = \begin{cases} |y_i - w \times \psi(x_i - b)| - \varepsilon, & |y_i - w \times \psi(x_i - b)| \geq \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

Here, w is a weight vector which is used to determine the maximum margin hyper plane, the term $\|w\|$ is called a regularized term, and it should be as flat as possible. The second term is the empirical error as measured by Vapnik's ε -insensitive loss function. C is the regularization constant.

The following commonly used kernel functions are included:

Linear: $k(x_1, x_2) = \langle x_1, x_2 \rangle$.

Polynomial: $k(x_1, x_2) = (\langle x_1, x_2 \rangle + R)^d$.

Sigmoid: $k(x_1, x_2) = \tanh(\langle x_1, x_2 \rangle + r)$.

Radial basis function: $k(x_1, x_2) = \exp(-\|x_1 - x_2\|^2 / 2\sigma^2)$.

(2) Bayes Classifier. The MMSE-BC has been considered the best strategy that uses Bayes method with the single feature mean load based on the evaluation type A in a previous work [16]. The MMSE-BC used here was the minimized MSE (MMSE) based Bayes classifier. It is a classic supervised learning classifier used in data mining. The formulation of MMSE-BC is represented in

$$P(\omega_i | \chi_j) = \frac{P(\chi_j | \omega_i) P(\omega_i)}{\sum_{k=1}^m P(\chi_j | \omega_k) P(\omega_k)}, \quad (7)$$

$$\hat{\omega}_i = E(\omega_i | \chi_j) = \sum_{i=1}^m \omega_i P(\chi_j | \omega_i). \quad (8)$$

```

Input: stock dataset, interval, duration of prediction
Output: CDF of the prediction MSE on different dataset and different methods
(1) Split dataset into training dataset and testing dataset
(2) for (newdataset = dataset[:n])/* n is data number, increasing by 40 */do
(3)   for (interval = 2, 4, 8, 16, 32) do
(4)     Determine the feature of the mean and interval slope in every interval
(5)     Predict the mean price  $\eta_i$ , using SVR or MMSE-BC method in training dataset.
        /* Use the mean and interval slope as feature of the stock price trend */
(6)   end for
(7)   Segment transformation based on (2):  $\eta \rightarrow l$ 
(8)   Calculate the MSE of this dataset
(9) end for
(10) Statistic 80 MSE of different dataset and plot the cumulative distribution function (CDF) of MSE

```

ALGORITHM 1: Stock price trend prediction model.

It is important for the Bayes classifier to compute the prior probability distribution $P(\omega_i)$ for the target states based on the samples and compute the joint probability distribution $P(\chi_j | \omega_i)$ for each state ω_i . Then, the posterior probability $P(\omega_i | \chi_j)$ was computed according to Formula (7).

2.4. Trend Prediction Model. The following trend prediction model is proposed here as a way of preventing the generation of cumulative errors. The proposed stock price trend prediction model has the following three steps: first, using the ESP principle, the estimated data segment is split into a set of consecutive segments whose lengths increase exponentially. The interval slope is used to describe the features of each interval. Then, the machine learning methods, SVR and MMSE-BC, were used to produce the transform model of the data and by which the mean stock price was predicted in the prediction of the next interval.

First, the stock opening price data were selected. Second, the time series (stock opening price data) was split into a set of a future time interval segments (η), whose lengths increase exponentially. The length of following subinterval was $t_0 + 2^i$, where $i = 1, 2, 3, 4, \dots$. Third, the mean and interval slope were computed for every subinterval, and the feature dataset was split into training dataset and prediction dataset. Next, MMSE-BC and SVR were trained in order to produce the model parameters. For example, it can compute the prior probability $P(\omega_i)$ and the conditional probability ($P(\chi_j | \omega_i)$ in (8)) and produce a boundary that leads to the largest margin from both sets of points in SVR and predict the mean stock price and the interval slope in prediction interval over a single interval. The interval slope must transform into the mean of the interval based on (5) because the mean values over consecutive future time intervals are used to express the long-term trends in the time series. Then, the mean values over consecutive future time intervals I can be converted from the vector η based on (2). At last, the mean squared error of this dataset can be calculated.

In order to evaluate the performance of MMSE-BC and SVR, the entire dataset prediction mean squared error was computed. For example, the price over the first 1000 trading days was selected for training and the price over the next 32

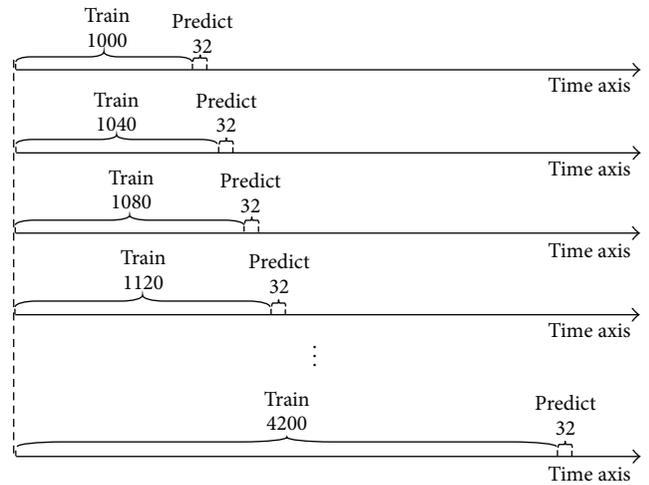


FIGURE 3: Setting the time window.

days was selected for prediction. The entire process follows the procedure of the trend prediction model mentioned. Then, the first 1040 trading days' price can be learned and the next 32 days' price can be predicted. Next, each process is to find the mean squared error of the prediction process. The process with higher prices prediction performance continues to predict the future stock price. The method of setting the time window is shown in Figure 3.

Algorithm 1 gives the pseudocode of the stock price trend prediction model.

3. Experiments and Comparison

This section presents experiments of the trend prediction model on the stock open price forecasting. The trend prediction model was here shown to be able to capture the dynamics of highly nonlinear, nonstationary time series.

3.1. Evaluation Indicator. To evaluate the accuracy of these predictions, the overall mean squared error (MSE) between

TABLE 1: Optimized parameters for the method.

Method	Key parameters	Values/compute method
SVR	Kernel	Sigmoid/radial basis function
	Penalty parameter C of the error term	1000
	Slack variables L_e	0.1

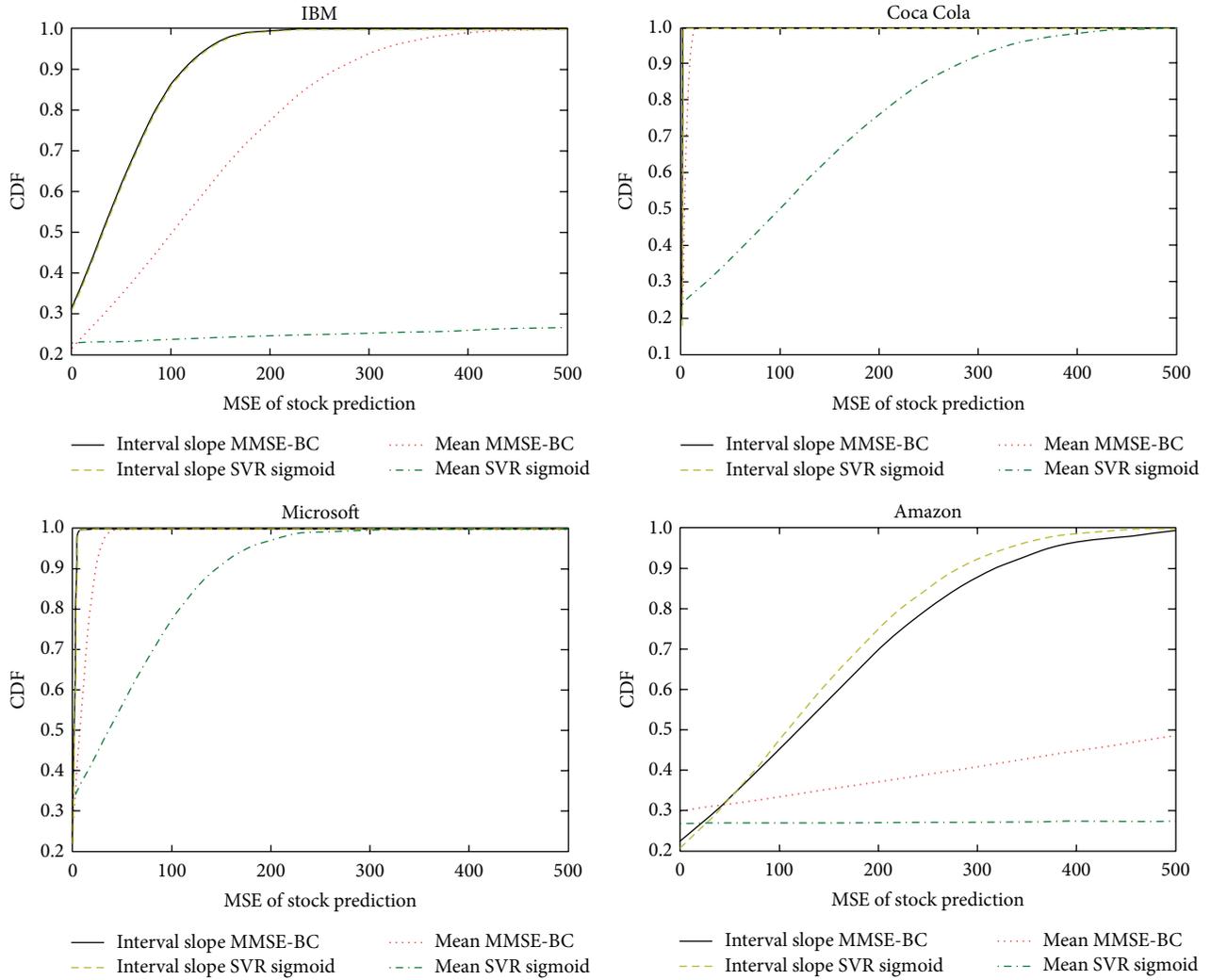


FIGURE 4: CDF of MSE for prediction of trends in stock price.

the predicted stock price values and the true values in the prediction interval can be calculated as follows:

$$\text{mse}(s) = \frac{1}{s} \sum_{i=1}^n s_i (l_i - L_i)^2. \quad (9)$$

Here, $s_i = 2^{i-1}$, $i \geq 1$, $s = \sum_{i=1}^n s_i$, l_i is the predicted mean of testing dataset, L_i is the true mean of testing dataset, and n is the total number of the segments in the prediction interval.

3.2. Method of Training and Evaluation. Eight opening stock price data samples were selected at random for these experiments: IBM, Coca Cola, Microsoft, Amazon, Sony,

Kimberly-Clark, Bank of America, and Walgreens in 1999.1.1–2014.10.30.

SVR and MMSE-BC were here used to predict the trends in opening stock price, and some key parameters are listed in Table 1.

3.3. Experimental Results. The results of MMSE-BC and SVR were compared to the classic mean and the interval slope. Eight stock opening price data samples, IBM, Coca Cola, Microsoft, Amazon, Sony, Kimberly-Clark, Bank of America, and Walgreens in 1999.1.1–2014.10.30, were compared to MSE. Figures 4 and 5 show the cumulative distribution function (CDF) of MSE of different prediction methods, in which SVR's kernel is sigmoid.

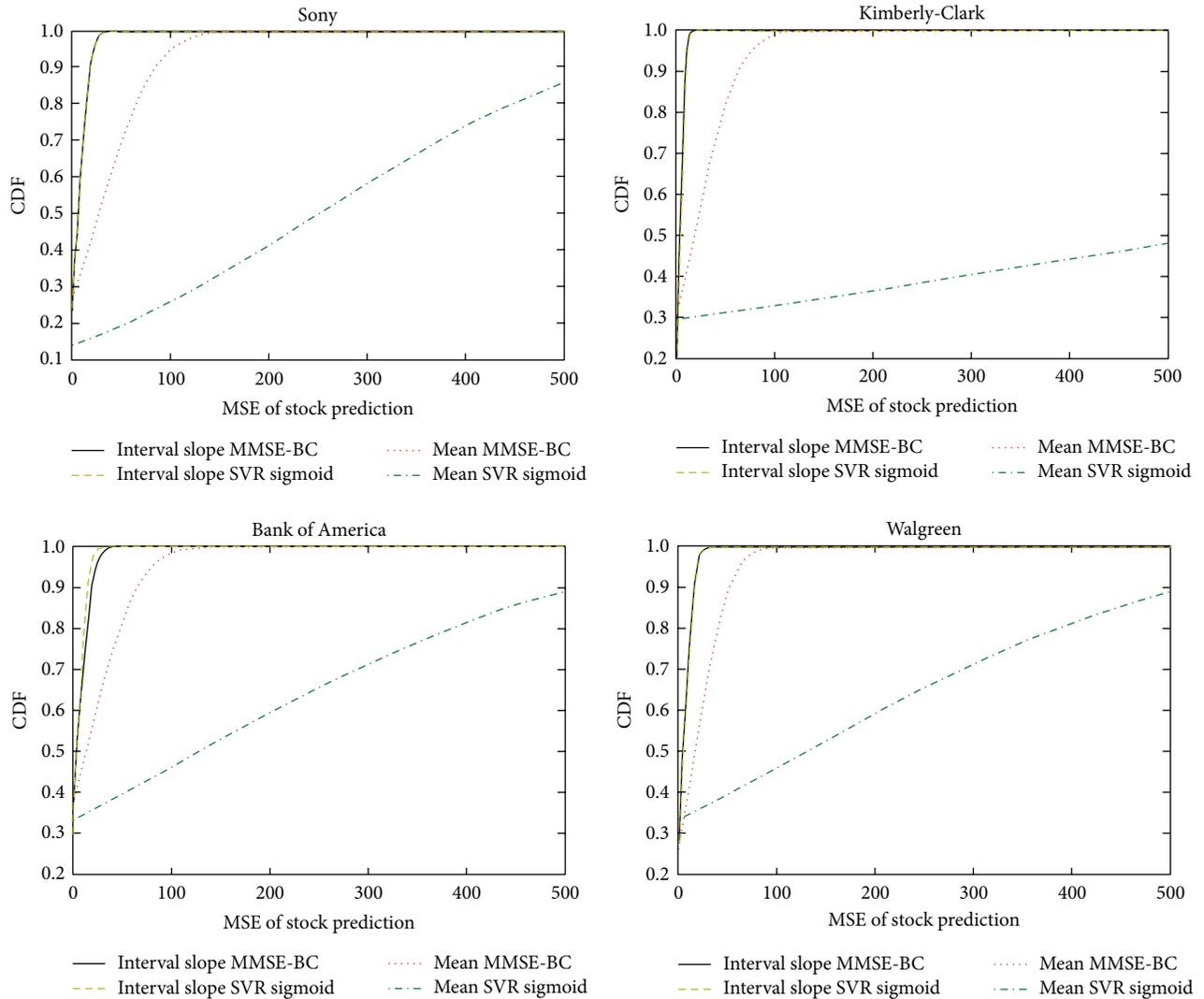


FIGURE 5: CDF of MSE for prediction of trends in stock price.

As shown in Figures 4 and 5, the interval slope curve is above the interval mean curve. This indicates that the interval slope's cumulative probability is greater than that of the interval mean when the value of MSE is below a certain threshold. For example, the IBM interval slope's cumulative probability was larger than the interval mean curve when the MSE value was less than 100. That means that 88% of the MSE values using interval slope were below 100, and only 52% of the MSE values using interval mean were below 100.

It is clear that interval slope's performance was better than that of the mean as indicated by the MMSE-BC and SVR methods. In this way, the interval slope can indicate more complex dynamics, such as change trends. In contrast, the mean can smooth out the dynamic fluctuations in stock price.

As an example of prediction results, Figure 6 shows IBM stock price trend prediction, that is, $t = 3440-3972$, by SVR based on interval slope, in which SVR's kernel is a radial basis function.

Both the mean stock price over a future time interval and the mean stock price over consecutive future time intervals

can be predicted accurately. This shows that the prediction of long-term stock price can be performed precisely without generating cumulative errors. The mean stock price over consecutive future time intervals can express future trends, such as sharp falls, slight falls, concussions, slight increases, sharp increases, falls followed by increases, and increases followed by falls. According to the prediction of the fluctuation of opening price over a long-term period, fund allocation models and trading strategies can be developed in advance.

4. Conclusion and Future Work

In this paper, ESP, which does not generate cumulative errors, was introduced and used to predict fluctuations in the opening prices of different stocks over a long period. The use of a new feature of stock price in the evidence subinterval, interval slope, was proposed to characterize the stock price fluctuation over some time period. It can be concluded that the interval slope can capture complex dynamics, such as trends in the changes in stock price.

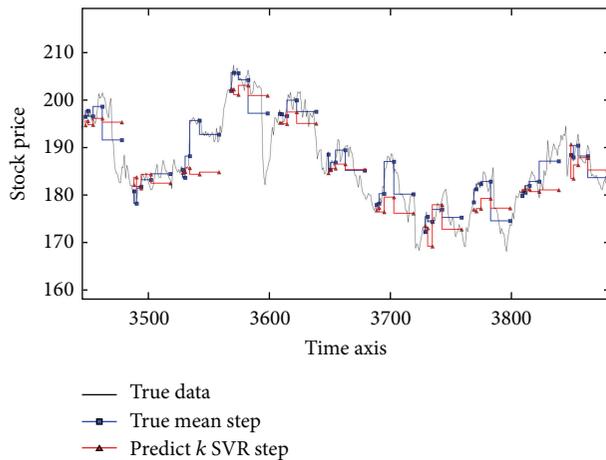


FIGURE 6: IBM stock price trend prediction.

The premise of this method of trend prediction is that future markets will change gradually rather than abruptly. The complexities of changes in stock price can greatly increase the difficulty of prediction. Future work should evaluate different learning methods and even combine different learning methods. Some new methods of evaluation should be used to evaluate the interval slope, the classic mean method, and the rate of return.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by NSFC under Grant no. 61273002, the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions no. CIT&TCD201304025, and the Key Science and Technology Project of Beijing Municipal Education Commission of China no. KZ201510011012.

References

- [1] R. S. Tsay, *Analysis of Financial Time Series*, John Wiley & Sons, Hoboken, NJ, USA, 2005.
- [2] G. Sermpinis, C. Dunis, J. Laws, and C. Stasinakis, "Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage," *Decision Support Systems*, vol. 54, no. 1, pp. 316–329, 2012.
- [3] P.-C. Chang, D.-D. Wang, and C.-L. Zhou, "A novel model by evolving partially connected neural network for stock price trend forecasting," *Expert Systems with Applications*, vol. 39, no. 1, pp. 611–620, 2012.
- [4] E. Kita, M. Harada, and T. Mizuno, "Application of Bayesian Network to stock price prediction," *Artificial Intelligence Research*, vol. 1, no. 2, 2012.
- [5] Z. Er-bo, M. Huan, and H. Zhan-Gang, "Applying genetic programming to analyze moving average and long & mid-term trends of stock prices," *Application Research of Computers*, vol. 27, no. 6, 2010.
- [6] P. Meesad and R. I. Rasel, "Predicting stock market price using support vector regression," in *Proceedings of the 2nd International Conference on Informatics, Electronics and Vision (ICIEV '13)*, pp. 1–6, IEEE, Dhaka, Bangladesh, May 2013.
- [7] P. Dondio, "Stock market prediction without sentiment analysis: using a web-traffic based classifier and user-level analysis," in *Proceedings of the 46th Annual Hawaii International Conference on System Sciences (HICSS '13)*, pp. 3137–3146, IEEE, Wailea, Hawaii, USA, January 2013.
- [8] M. Hagenau, M. Hauser, M. Liebmann, and D. Neumann, "Reading all the news at the same time: predicting mid-term stock price developments based on news momentum," in *Proceedings of the 46th Annual Hawaii International Conference on System Sciences (HICSS '13)*, pp. 1279–1288, Wailea, Hawaii, USA, January 2013.
- [9] D.-Y. Xu, S.-L. Yang, and R.-P. Liu, "A mixture of HMM, GA, and Elman network for load prediction in cloud-oriented data centers," *Journal of Zhejiang University: Science C*, vol. 14, no. 11, pp. 845–858, 2013.
- [10] R. Bisoi and P. K. Dash, "A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter," *Applied Soft Computing Journal*, vol. 19, pp. 41–56, 2014.
- [11] B. Bican and Y. Yaslan, "A hybrid method for time series prediction using EMD and SVR," in *Proceedings of the 6th International Symposium on Communications, Control and Signal Processing (ISCCSP '14)*, pp. 566–569, Athens, Greece, May 2014.
- [12] Z. Huang and M.-L. Shyu, "k-NN based LS-SVM framework for long-term time series prediction," in *Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI '10)*, pp. 69–74, IEEE, Las Vegas, Nev, USA, August 2010.
- [13] J. Fan and Y. Tang, "An EMD-SVR method for non-stationary time series prediction," in *Proceedings of the International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE '13)*, pp. 1765–1770, IEEE, Chengdu, China, July 2013.
- [14] T. Fletcher and J. Shawe-Taylor, "Multiple kernel learning with fisher kernels for high frequency currency prediction," *Computational Economics*, vol. 42, no. 2, pp. 217–240, 2013.
- [15] Q. Yang, C. Peng, Y. Yu et al., "Host load prediction based on PSR and EA-GMDH for cloud computing system," in *Proceedings of the 3rd IEEE International Conference on Cloud and Green Computing (CGC '13)*, pp. 9–15, IEEE, Karlsruhe, Germany, October 2013.
- [16] S. Di, D. Kondo, and W. Cirne, "Google hostload prediction based on Bayesian model with optimized feature combination," *Journal of Parallel and Distributed Computing*, vol. 74, no. 1, pp. 1820–1832, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

