Stock Price Prediction Based on a Network with Gaussian Kernel Functions

Dong Kyu Kim and Rhee Man Kil*

College of Information and Communication Engineering, Sungkyunkwan Univesity
2066, Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do, 440-746, Korea
goldencrosser@gmail.com, rmkil@skku.edu

Abstract. This paper presents a new method of stock price prediction based on the phase space analysis for stock price series. For the prediction model, a network with Gaussian kernel functions is selected and this network is optimized by using a noise variance estimate. As a result, the proposed model provides the high accuracy of predicted values. Through the simulation for the prediction of KOSPI 200 stock price values, the effectiveness of the proposed prediction model has been demonstrated.

Keywords: stock price prediction, regression model, nonparametric estimation, noise variance.

1 Introduction

It has been known that stock price values are generated by a nonlinear and non-stationary dynamic system according to the trend of market which is associated with economic situation. For the problem of time series prediction, the linear regression models such as ARMA and ARIMA models are usually used. However, these models may not provide good estimates of predicted stock price values. If the stock price values are complete random, it is not possible to predict the future values. However, in many cases, the stock price values are not completely random. Rather, in most cases, they show coherent or sometimes chaotic behavior [1]. In this case, stock price values can be described by a nonlinear dynamics. In this respect, the nonlinear and nonparametric regression model is investigated to provide the better estimates of predicted values. As a nonlinear regression model, a network with Gaussian kernel functions is considered since this model is a nonparametric estimation model and good for incremental learning due to the locality of kernel functions. Here, the problems are to determine the input structure for the given prediction of stock price values and also to determine the proper size of regression models for the given stock price series. For this purpose, the input structure is investigated by analyzing the phase space of stock price dynamics and the proper size of regression models is investigated using the estimation of noise variances embedded in stock price series. As a result, the proposed model provides an accurate estimation of predicted values.

---

* Corresponding author.

To demonstrate the effectiveness of the proposed model, the Korea Composite Stock Price Index (KOSPI) 200 stock price series from June 2008 to June 2012 is used for the prediction of stock price values. The proposed model can also be applied to other cases of time series prediction since our model is not just restricted to the problem of stock price prediction.

2   Phase Space Analysis of Stock Price Series

The goal of time series prediction is to predict future values from a series of past values observed at regular time intervals. For the modeling of time series data, we usually rely on the theory called the delay coordinate embedding proposed by Packard et al. [2] and Takens [3]. Their purpose is to find the appropriate dimension of attractors generated by the dynamical system. To describe the dynamical system, let us assume that there exists a D dimensional state vector $\boldsymbol{x}(t)$, $t \in \mathbb{R}$; that is,

$$\boldsymbol{x}(t) = [x(t), x^{(1)}(t), \cdots, x^{(D-1)}(t)]^T \tag{1}$$

where $x^{(i)}(t)$ represents the ith derivative of $x(t)$ with respect to t. This vector can be produced by a dynamical system,

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}, t) \tag{2}$$

where $\boldsymbol{F}$ represents the D dimensional mapping function from the state vector $\boldsymbol{x}$ to the vector $\dot{\boldsymbol{x}}$ at time t. With no information on the state space (or phase space), we can only observe $x(t)$. Moreover, we usually consider the sampled discrete time series data $x(t_k)$; that is,

$$x(t_k) = x(t_0 + k\Delta t) \equiv x(k), \quad k = 0, 1, 2, \cdots \tag{3}$$

where t_0, k, and Δt represent the initial time, sampling step, and sampling interval, respectively. Here, let us construct E-dimensional vectors

$$\boldsymbol{x}_{\tau,E}(k) = [x(k), x(k-\tau), \ldots, x(k-(E-1)\tau)]^T, \quad k = (E-1)\tau, \cdots \tag{4}$$

where τ and E represent the delay time measured as the unit of Δt and embedding dimension, respectively.

For the prediction of future values, we assume that the state vectors in the reconstructed state space are governed by a nonlinear function f; that is,

$$x(t_k + P) = f(\boldsymbol{x}_{\tau,E}(k)) \tag{5}$$

where P represents the prediction step. Then, our goal is to make a prediction model $\hat{f}$ estimating the unknown target function f. However, since the geometric structure of f is determined by the embedding parameters τ and E, the performance of the prediction model is strongly affected by the choice of the embedding parameters. One of the typical methods is to choose the embedding dimension E by estimating the correlation dimension of the system [4], and to

choose the delay time τ at which the first minimum of the mutual information occurs. However, the correlation dimension does not indicate any information on the target function f. In this context, we consider to determine the embedding parameters using the measure related to the smoothness of the target function.

For the analysis of time series data, let us denote the rth nearest neighbor of the vector $\boldsymbol{x}_{\tau,E}(k)$ by $\boldsymbol{x}_{\tau,E}^r(k)$. Since the difficulty of estimation depends on the smoothness of the target function f, we define the gradient of f at each point $\boldsymbol{x}_{\tau,E}(k)$ by

$$\Delta f(\boldsymbol{x}_{\tau,E}^r) = \frac{\left| f(\boldsymbol{x}_{\tau,E}(k)) - f(\boldsymbol{x}_{\tau,E}^r(k)) \right|}{|| \boldsymbol{x}_{\tau,E}(k) - \boldsymbol{x}_{\tau,E}^r(k) ||}, \tag{6}$$

where the norm in the denominator represents the Euclidian distance in $\mathbb{R}^E$. In this paper, $r = 1$ is used; that is, the nearest neighbor of the state vector. Here, the smoothness measure $S(\tau, E)$ of a target function f [5] is defined by using an average of the gradient values for time series $x(t_k)$, $k = 0, 1, \cdots, n - 1$; that is,

$$S(\tau, E) = 1 - \frac{1}{n - (E-1)\tau} \sum_{k=(E-1)\tau}^{n-1} \Delta f(\boldsymbol{x}_{\tau,E}^1(k)). \tag{7}$$

The smoothness measure $S(\tau, E)$ is closely related to the unfolding of the orbits of the vectors in the reconstructed state space. If the embedding dimension E is too small, the state space is yet restricted to the low dimensional space. This restriction makes different orbits of vectors to be folded and placed closely to each other; that is, even the nearest neighbor vectors might have come from different orbits. Since the vectors on the different orbits are likely to have largely different target function values, large gradient values imply that the embedding dimension is too small. On the other hand, if the embedding dimension E is large enough, the nearest neighbor vectors are obtained from the same orbit, which results in the small gradient values. Also, the smoothness measure $S(\tau, E)$ will not be much influenced by changes in E when E is large, because the state space is already unfolded. From these observations, the optimal embedding dimension and delay time can be determined by identifying the points where the smoothness measure $S(\tau, E)$ changes rapidly from the smaller values to the larger values. For the analysis of one step prediction for KOSPI 200 stock prices, the smoothness measure of (7) is calculated for every delay time τ between 1 and 20, and every embedding dimension E between 2 and 10. The calculated smoothness measure of KOSPI 200 stock prices is illustrated in Fig. 1. From the plot of smoothness measure, the smallest embedding dimension is selected first. Here, the proper embedding dimension is selected as $E = 3$. Then, for the given embedding dimension $E = 3$, the proper delay time is selected as $\tau = 4$. As a result, the model of one step prediction is determined as

$$x(t + 1) = f(x(t), x(t - 4), x(t - 8)). \tag{8}$$

From this phase space analysis, it is evident that the information of 3 data within 8 day window provides an important cue to determine the next day prediction of KOSPI 200 stock price values.

<div align="center">(a)   (b)</div>
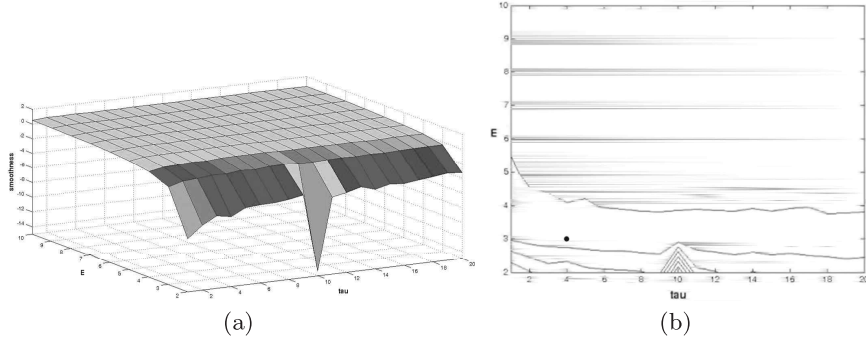
Fig. 1. The plot of smoothness measure for stock price prediction: (a) and (b) represent the 3D plot and contour map of stock price series. The dot in the contour map represents the selected delay time $\tau = 4$ and embedding dimension $E = 3$.

3   Nonlinear Stock Price Prediction Model

As a nonlinear prediction model for stock prices, a network with Gaussian kernel function is selected since this network is able to perform nonlinear and nonparametric estimation and good for incremental learning due to the locality of kernel functions. The suggested prediction model $\hat{f}$ with m kernel functions is described by

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{m} w_i \psi_i(\boldsymbol{x}), \quad \psi_i(\boldsymbol{x}) = e^{-||\boldsymbol{x}-\boldsymbol{\mu}_i||^2/2\sigma_i^2}, \tag{9}$$

where w_i represents the connection weight between the output and the ith kernel function ψ_i in which $\boldsymbol{\mu}_i$ and σ_i represent the mean and standard deviation, respectively. In (9), we need to determine the parameters of m and also kernel related parameters w_i, $\boldsymbol{\mu}_i$, σ_i. Foe kernel related parameters, [6] suggested an efficient estimation method in such a way of minimizing the mean square error (MSE). However, for our problem of prediction model, an optimal number m of kernel functions should be determined to be fitted to the stock price series so that the generalization error is minimized. Here, in the case of time series $x(t_k + P)$, it can be described by

$$x(t_k + P) = f(\boldsymbol{x}_{\tau,E}(k)) + \epsilon, \tag{10}$$

where f represent an embedded function for the generation of time series data and ϵ represents a noise term expressed as a random variable with mean 0 and variance σ^2.

Then, for $x(t_k + P)$, the predicted value $\hat{x}(t_k + P)$ is described by

$$\hat{x}(t_k + P) = \hat{f}(\boldsymbol{x}_{\tau,E}(k)). \tag{11}$$

The expected risk (or MSE) between the true and predicted values is described by

$$E[(x - \hat{x})^2] = E[(f - \hat{f})^2] + Var(\epsilon), \tag{12}$$

where $E[(f - \hat{f})^2]$ represents the regression error and $Var(\epsilon)$ represents the variance of noise.

From the above equation, it is clear that the expected risk is always greater than or equal to the variance of noise. This implies that the empirical error (or training error) should not be less than the variance of noise to avoid over-fitting of the regression model for the given time series. Here, the variance of noise should be estimated. One method of estimating the variance of noise is using the squares of time series differences [7]:

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{k=1}^{n-1} (x(t_k) - x(t_{k-1}))^2, \tag{13}$$

where $\hat{\sigma}^2$ represents an estimator of noise variance.

Assuming that the random noise follows a normal distribution, an estimator of (13) follows a chi-square distribution; that is,

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{14}$$

Then, with a confidence level of $1 - \alpha$, the following probability of (14) is described:

$$P\left(\chi^2_{1-\alpha/2,n-1} \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}\right) = 1 - \alpha. \tag{15}$$

From the above equation, a $100(1 - \alpha)\%$ confidence interval for noise variance σ^2 is determined by

$$\frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n-1}}. \tag{16}$$

In the case of large sample size, this interval is also effective even if the distribution of random noise is unknown because both an estimator of (13) and chi-square random variable approximately follow a normal distribution for large n according to the central limit theorem. Then, from (12) and (16), it is recommended that the training error should not be within the confidence interval for noise variance. From this point of view, the following algorithm of constructing a stock price prediction model is suggested:

Construction of Nonlinear Stock Price Prediction Model

Step 1. From the given stock price series $x(t_k)$, $k = 0, 1, \cdots, n - 1$ and prediction time P, determine the values of smoothness measure of (7). Then, determine the embedding dimension E and delay time τ when the value of smoothness measure is large (usually, greater than -1) at the smaller embedding dimension.

Step 2. Determine the structure of nonlinear prediction model of (5).

Step 3. Estimate the noise variance using (13) and determine a $100(1 - \alpha)\%$ confidence interval for noise variance using (16).

Step 4. Train the nonlinear prediction model using an incremental learning algorithm such as [6] as the number of kernel functions increases and determine the following training error:

$$R_{emp}(\hat{f}) = \frac{1}{n} \sum_{k=0}^{n-1} (x(t_k) - \hat{x}(t_k))^2. \tag{17}$$

Step 5. If the condition

$$R_{emp}(\hat{f}) > \frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n-1}} \tag{18}$$

is met, continue the learning process of Step 4. Otherwise, stop the learning process.

This construction algorithm was applied to KOSPI 200 stock price series using the prediction model of (8). In this experiment, an 80% of stock price series was used as training data and the remaining 20% was used as test data. Here, the noise variance was estimated as $\hat{\sigma}^2 = 0.000127$ and a 99% confidence interval for noise variance σ^2 was determined by

$$0.000115 \leq \sigma^2 \leq 0.000159.$$

Here, to compare with the training and test error with respect to the number of kernel functions, the training error, test error, and confidence interval were plotted as illustrated in Fig. 2.
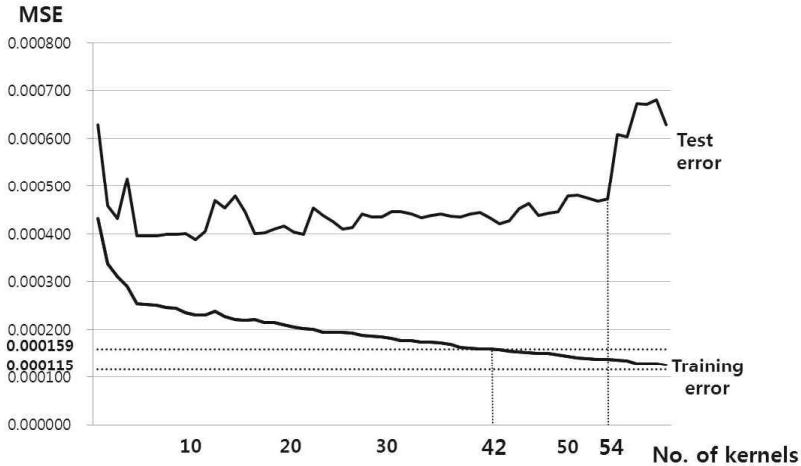
Fig. 2. Selection of the number of kernel functions by comparing the training error with the confidence interval for noise variance

As shown in Fig. 2, the over-fitting of regression model occurs when the number of kernel function is greater than or equal to 54. In this case, the training error is within the confidence interval for noise variance. Therefore, in this

experiment, it is clear that the proper number of kernel functions is less than or equal to 42 which makes the training error just greater than the upper limit of the confidence interval.

4   Simulation

For the simulation for stock price prediction, the data set of Korea Composite Stock Price Index (KOSPI) 200 stock price series from June 2008 to June 2012 was selected and normalized by dividing the price values by 200 so that approximate price range was between 0.5 and 1.5. In this data set, the first 70%, 80%, and 90% of data were used as training data sets, and the remaining data sets were used as test data sets. First, to identify the dynamics of stock price series for the given prediction time $P = 1$, the embedding dimension E and delay time τ were determined using the smoothness measure of (7). As a result, the prediction model of (8) was obtained. Then, the noise variance was estimated using the estimator of (13) and a 99% confidence interval for noise variance was determined using the form of (16). For the given prediction model, the parameters of Gaussian kernel function network (GKFN) of (9) were trained using the non-parametric estimation method of [6] as the number of kernel functions increased. This training was continuously performed until the training error reached the upper limit of the confidence interval for noise variance. Here, to measure the performances of stock price prediction, the following root mean square error (RMSE) and the coefficient of determination R^2 were use for the test data $y(t_i)$, $i = 0, \cdots, l - 1$:

$$RMSE = \left(\frac{1}{l} \sum_{i=0}^{l-1} (y(t_i) - \hat{y}(t_i))^2 \right)^{1/2}, \tag{19}$$

where l represents the number of test data and

$$R^2 = 1 - \left(\sum_{l=0}^{l-1} (y(t_i) - \hat{y}(t_i))^2 / \sum_{l=0}^{l-1} (y(t_i) - \bar{y})^2 \right), \tag{20}$$

where $\bar{y}$ represent the sample mean of $y(t_i)$; that is, $\bar{y} = (1/l) \sum_{l=0}^{l-1} y(t_i)$. Here, the range of R^2 is between 0 and 1, and R^2 indicates the degree of how well the regression model covers the variation of time series data; for example, if $R^2 = 0.9$, the regression model covers 90% of the variation of time series data and the remaining 10% of the variation is not explained. Then, the simulation results for stock price prediction using the proposed GKFN were obtained and described in Table 1. For the comparison, simulation results for stock price prediction using the kernel support vector machine (k-SVM) [8], one of popular methods in nonlinear regression models, were also obtained. In the k-SVM, the same form of the prediction model of (8) was also used.

These simulation results showed that the RMSE was decreased as the ratio of training data was increased and the proposed GKFN outperformed the k-SVM from the view points of the RMSE and also R^2. In general, if R^2 is greater

Table 1. Simulation results for stock price prediction

Ratio of training data	k-SVM		GKFN	
	RMSE	R^2	RMSE	R^2
70%	0.058423	0.60	**0.035879**	**0.87**
80%	0.039960	0.64	**0.020517**	**0.90**
90%	0.036820	0.41	**0.016172**	**0.88**

than or equal to 0.9, we consider that the regression model performs very good fit to the given data. The proposed GKFN was very close to this performance. The main reason was due to the fact that the prediction model was determined by phase space analysis and the optimization of regression model was done by comparing the training error with the noise variance estimate.

5   Conclusion

The stock price prediction involves the analysis of stock price series and also optimization of regression models. In this work, the stock price series is analyzed by the phase space analysis method [5]. As the prediction model, a network with Gaussian kernel functions is selected since this model is good for incremental learning due to the locality of kernel functions. For the optimization of regression model, the noise variance is estimated and used to determine the proper number of kernel functions. As a result, the proposed model provides the high accuracy of predicted values. Through the simulation for the prediction of KOSPI 200 stock price values, the effectiveness of the proposed prediction model has been demonstrated. The proposed model can also be applied to various problems of time series prediction.

References

1. Peters, E.E.: Fractal market analysis: applying chaos theory to investment and economics. John Wiley & Sons, New York (1994)
2. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a Time Series. Phys. Rev. Lett. 45, 712–716 (1980)
3. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.S. (eds.) EAMT-WS 1993. Lecture Notes in Mathematics, vol. 898, pp. 366–381. Springer-Verlag, Berlin (1981)
4. Havstad, J.W., Ehlers, C.L.: Attractor dimension of nonstationary dynamical systems from small data sets. Phys. Rev. A 39(2), 845–853 (1989)
5. Kil, R., Park, S., Kim, S.: Time series analysis based on the smoothness measure of mapping in the phse space of attractors. In: International Joint Conference on Neural Networks, vol. 4, pp. 2584–2589 (1999)
6. Kil, R.: Function approximation based on a network with kernel functions of bounds and locality. ETRI Journal 15, 35–51 (1993)
7. Rice, J.: Bandwidth choice for nonparametric regression. Annals of Statistics 12(4), 1215–1230 (1984)
8. SVMlight ver 6.02, Cornell University, `http://svmlight.joachims.org`