Forecasting Share Price using Wavelet Transform and LS-SVM based on Chaos Theory

Jianguo Zhou, Tao Bai, Aiguang Zhang, and Jiming Tian
School of Business Administration
North China Electric Power University
Baoding, China
dldxzjg@126.com, ncepubaitao@126.com

***Abstract*—In the analysis of predicting share price based on least squares support vector machine (LS-SVM), the instability of the time series could lead to decrease of prediction accuracy. On the other hand, three SVM parameters, c, ε and σ, must be carefully predetermined in establishing an efficient LS-SVM model. In order to solve the problems mentioned above, in this paper, the hybrid of wavelet transform (WT) with LS-SVM model was established. First the chaotic feature of share price is verified with chaos theory. It can be seen that share price possessed chaotic features, providing a basis for performing short-term forecast of share price with the help of chaos theory. Average Mutual Information (AMI) method is used to find the optimal time lag. Then the time series is decomposed by wavelet transform to eliminate the instability. Cao's method is adopted to determine free parameters of support vector machines. Additionally, the proposed model was tested on the prediction of share price of one listed company in China. Especially, In order to validate the rationality of chosen dimension, the other three random dimensions were selected to compare with the calculated dimension. And to prove the effectiveness of the model, PSVM algorithm was used to compare with the result of WT-SVM. Experimental results showed that the proposed model performed the best predictive accuracy and generalization, implying that integrating the wavelet transform with LS-SVM model can serve as a promising alternative for share price prediction.**

***Keywords*—Share Price Forecast, Chaotic Time Series, Wavelet Transform, LS-SVM**

I. INTRODUCTION

Along with the development of stock market, share price forecast has been an important topic among people, especially in stockholders and senior managers. The evolution of share price embodies a variety of information in stock market that comprehensively reflects the law of motion of a variety of market factors. Therefore, an analysis and a mastery of the law of evolution of the share price are of critical importance to improve investment performance. From a managerial perspective, the fall of their share price will give them a signal that some problems appear in their company, and they can take timely strategic actions to avoid distress. For stockholders, the variation of share price allows to arrange rational trading plan to maximize their economic benefit.

Stock market which is a nonlinear dynamic system is influenced by many factors, such as policy, financial condition and market expectation. The joint effect of these factors leads to the extremely complicated random variation of the share price. However, it has recently been found that behind the seemingly random appearance, definite regularity is hidden. Subsequently, researchers have made a quantity of work on the predicting of share price [1].

Originally, Statistical methods have been used for developing share price prediction models. The statistical methods include autoregressive approaches, least square method, and exponential smoothing approaches. To develop an accurate and general applicable prediction approach, data mining and machine learning techniques are employed, e.g. neural network model (NN)[2] and Neuro-fuzzy architectures. However, NN has many shortages, such as local optimization, slow speed of training, and low efficiency. Recently, new algorithms in machine learning support vector machines (SVMs)[3] developed by Boster, Guyon, and Vapnik (1992) provide better solutions to design boundary than that of neural network. Since the new model was proposed, SVM has been successfully applied to many complicated system, including electricity price, traffic flow and power load.

On the other hand, chaos theory points out that the complicated things which are originally believed unpredictable have predictability [4]. This encourages us to apply the chaos theory to the time series forecast. Average Mutual Information (AMI) [5] technique is used to choose the optimal time lag. After eliminating the fluctuant components using wavelet transform [6], the trend and periodicity are illustrated explicitly. Cao's method [7] is used to select the optimal embedding dimension of decomposed series. Therefore, the chaos theory can be used to reveal its intrinsic regularity until more accurate and rational analysis results and prediction models are obtained.

According to the analysis as mentioned above, firstly, the seeming chaotic features of share price are verified with Lyapunov exponents. Secondly, in order to improve the accuracy and stability of share price prediction, wavelet transform [8] is used to decompose the time series and eliminate the fluctuant components. Subsequently, the predicting model of WT-SVM algorithm [9] based on chaos theory is constructed. Lastly, applying the model to a practical

978-1-4244-1674-5/08 /$25.00 ©2008 IEEE

share price forecast and comparing the outcomes with factual share price data and predicting outcomes of PSVM Algorithm, the result proves that the predicting model proposed in this paper can perform successfully in share price predicting.

II. Basic Concep of Models

A. Chaos Theory

Chaos is a nonlinear behavior that exists between the realms of periodic and random [4]. The exact system state can be written:

$$X(t)=\left(x(t),x(t-T),x(t-2T),\cdots,x\left(t-(k-1)T\right)\right) \quad (1)$$

Where t is a scalar index for the data series and T is the interval of observations. For a discrete time series, the next value of the state can be obtained as a function of the current state:

$$x(t+1)=f\left(X(t)\right) \quad (2)$$

Dynamic systems may evolve over time and generate time series with regular appearance, but it is possible that the system evolves to a chaotic attractor. The goal of a time series predictor is to reconstruct the chaotic dynamics of the space state from the measurements of one component of the state vector of a dynamic system, and to predict the evolution of the measured variable.

Takens' embedding theorem proved that if the dynamics occurs in a d dimensional Eucidean space, a reconstruction of the system can be obtained in a $2d+1$ dimensional space built using delay coordinates, which define the delay coordinate vectors. A delay coordinate is simply an observed variable with a time τ .For example, if we have the time series $x_1,x_2,\cdots,x_N$, the i -th delay coordinate vector with embedding d is:

$$y_i(d)=\left[x_i,x_{i+\tau},\cdots,x_{i+(d-1)\tau}\right],i=1,2,\cdots,N-(d-1)\tau \quad (3)$$

Where τ is called time lag. The number of delay coordinates necessary to reconstruct the state of a dynamic system represents the minimum embedding dimension d .

To reconstruct the phase space, good choices for time lag τ and embedding dimension d are needed. We use the first minimum of the average mutual information (AMI) function:

$$I(\tau)=\sum_{n}^{n+\tau}P(x_n,x_{n+\tau})\log_2\left[\frac{P(x_n,x_{n+\tau})}{P(x_n)P(x_{n+\tau})}\right] \quad (4)$$

Where $P(x_n)$ is the probability density of x_n , while the $P(x_n,x_{n+\tau})$ is the probability density of x_n and $x_{n+\tau}$.

The false nearest neighbor (FNN) method is a approach to find the optimal embedding dimension, however, we prefer to Cao's method for the reason as follows:

1) does not contain any subjective parameters except for the time delay for the embeddings

2) does not strongly depend on how many data points are available

3) is computationally efficient

Cao's method can be expressed as follows:

$$\text{Let } EI(d)=\frac{E(d+1)}{E(d)} \quad (5)$$

$$\text{With } E(d)=\frac{1}{N-d\beta\tau}\sum_{t=0}^{N-d\tau-1}\frac{\left\|y_{d+1}(t)-y_{d+1}\right\|}{\left\|y_{d(t)}-y_c^{NN}(t)\right\|} \quad (6)$$

$$\text{And } \left\|y_d(t)-y_{d(t)}^{NN}\right\|=\max\left|x(t+j\tau)-x^{NN}(t+j\tau)\right| \quad (7)$$

Where T is the length of the original date series, d refers to the embedding dimension, NN means the nearest neighbor to other vector. As d gets large enough, will tend to one. The appropriate embedding dimension is given by the value of d where E1(d) stops changing.

B. Wavelet Transform

Suppose the function $\varphi(t)\in L^2(R)$ and its Fourier transform $\psi(\omega)$ satisfies the condition:

$$\int_R\frac{|\psi(\omega)|^2}{\omega}d\omega<\infty \quad (8)$$

Then $\varphi(t)$ can be called mother wavelet. By dilations and translations of mother wavelet, a family of wavelet functions as follows can be obtained:

$$\psi_{a,d}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-d}{a}\right)\quad(a\neq 0,d\in R) \quad (9)$$

Where a denotes the dilation factor and d denotes the translation factor.

Let $a=2^j$ and $d=k2^j$, discrete wavelet transform (DWT) can be realized:

$$\psi_{j,k}(t)=2^{-j/2}\psi\left(2^{-j}t-k\right) \quad (10)$$

Where, k denotes the shift parameter and j denotes the resolution level. The larger the value of j , the lower the frequency.

According to (10), the reconstruction expression of $f(x)$ can be presented as follows:

$$\begin{aligned}f(t)&=\sum_k c_{j,k}\varphi_{j,k}(t)+\sum_k\sum_j d_{j,k}\psi_{j,k}(t)\\&=a_j(t)+\sum_j d_j(t)\end{aligned} \quad (11)$$

Where, a_j and d_j are the approximate and detail parts of original signal, respectively.

C. Least Squares Support Vector Machines

Suppose we have the independent uniformly distributed data $\{x_1,y_1\}\cdots(x_N,y_N)$, where each $x_i\in R^n$ denotes the input space of the sample and has a corresponding target value $y_i\in R$ for $i=1\cdots N$, where N corresponds to the size of the training data. The estimating function takes the form as follows:

$$f(x)=\left(w\cdot\phi(x)\right)+b$$

Where $\phi(x)$ denotes the high dimensional feature space which is nonlinearly mapped from the input space. This leads to the optimization problem for standard SVM:

$$\text{Minimize} \quad \frac{1}{2}w^T w + \gamma\sum_{i=1}^{N}\xi_i \tag{12}$$

$$\text{Subject to} \quad \begin{cases} y_i\left[w^T\phi(x_i)+b\right] \geq 1-\xi_i \\ \xi_i \geq 0, i=1,\cdots,N \end{cases} \tag{13}$$

Where ξ_i denotes a slack variable and γ denotes a positive real constant which determines penalties to estimation errors.

For LS-SVM, (13) has been modified as follows:

$$\text{Minimize} \quad \frac{1}{2}w^T w + \gamma\sum_{i=1}^{N}\xi_i \tag{14}$$

Subject to the equality constrains:

$$y_i\left[w^T\phi(x_i)+b\right] = 1-\xi_i \qquad i=1,\cdots,N \tag{15}$$

By constructing the Lagrange function and according to KKT Conditions, the equation as follows can be obtained:

$$\begin{cases} w=\sum_{i=1}^{N}\alpha_i y_i \phi(x_i) \\ \sum_{i=1}^{N}\alpha_i y_i = 0 \\ \alpha_i = \gamma\xi_i \\ y_i\left[w^T\phi(x_i)+b\right]-1+\xi_i=0 \end{cases} \tag{16}$$

Then we define:

$$\begin{cases} Z=\left[\phi(x_1)^T y_1;\cdots;\phi(x_i)^T y_i\right] \\ Y=\left[y_1;\cdots;y_i\right]^T \\ \vec{1}=\left[1;\cdots;1\right] \\ \xi=\left[\xi_1;\cdots;\xi_i\right] \\ \alpha=\left[\alpha_1;\cdots;\alpha_i\right] \end{cases} \tag{17}$$

After substituting (17) into (16) and eliminating w and γ, we can obtain:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \tag{18}$$

By defining $\Omega = ZZ^T$ and applying Mercer's Condition within the Ω the matrix, each element of the matrix is in the form:

$$\Omega_{i,j} = y_i y_j \phi(x_j) = y_i y_j K(x_i, x_j) \tag{19}$$

Where $K(x_i, x_j)$ is defined as kernel function. The value of a kernel function equals to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is $K(x_i, x_j)=\phi(x_i)\phi(x_j)$. The typical examples of kernel function are polynomial kernel, RBF kernel and sigmoid kernel. The, resulting LS-SVM model for regression can be expressed as follows:

$$f(x)=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)K(x_i,x)+b \tag{20}$$

Mean squared error (MSE) is an index using to evaluate the performance of the forecasting model, must be designed before starting to search optimal values of SVM parameters.

$$(MSE)=\sqrt{\frac{1}{K}\sum_{i=1}^{K}\left(X_i-X_j\right)^2}$$

Where X_i denotes the actual value; X_j denotes the forecasting value; K denotes the number of test sample.

This MSE is used herein as the indicator of model performance to compare the results achieved by the proposed model with those obtained using other models. The MSE denotes the value of fitness function in LS-SVM.

III. Research Design

A. Research Data

In order to verify the effectiveness of proposed model in predicting the share price, the daily closing prices of Xichang Electricity stock from Jul 22, 2003 to Nov 23,2004 (totaling 320 trading days) which is listed in shanghai market in China are selected as the data sets. The typical sample of share price is illustrated in Fig. 1.

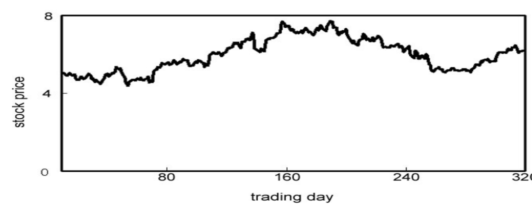

Fig.1 The Share Price From 2003 to 2004

In cases of SVM, each data set is splited into two subsets: a training set of 300 (from Jul 22,2003 to Oct 26,2004) and a test set of the other 20 data, respectively. The test data is used to test the results but not to develop the model. In case of BPN, the data set is splited into three subsets: a training set of 220 (from Jul 22,2003 to Jun 29,2004), a validation set of 80 (from Jun 30,2004 to Oct 26,2004) and a testing set of the other 20 data, respectively, Where the validation set is used to check the results.

B. Analysis Steps

This study is conducted according to the analysis steps as follows:

Step 1. According to historical data of share price, applying Lyapunov exponents to verify the chaotic feature of this time series.

Step 2. Applying wavelet transform to decompose the time series.

Step 3. Defining delay time, extracting embedding dimensions and phase points and reconstructing the phase space using the time series.

Step 4. Implementing the proposed WT-SVM to perform prediction task.

Step 5. We verify the rationality and effectiveness of proposed model and compare the prediction accuracy of WT-SVM with PSVM.

C. The PSVM Modeling

The particle filter algorithm (PFA) is usually used to deal with the time series which is interfered by noise. The principle of applying PFA to eliminate noise is following: introducing the observed value into distribution of samples and using particle swarm optimization algorithms to optimize the sampling process, it makes distribution of samples move to a high posterior probability region. Thus the instability of time series is eliminated. In this paper, we construct a predicting model combining PFA with LS-SVM. The number of particles is determined 50.

IV. EXPERIMENT AND ANALYSIS

Due to share price is inconstant, we must apply discrete process to it for identifying its regularity behind the seemingly random evolving process. In this paper we utilized wavelet transform to decompose the time series.

A. Analysis of Chaotic Feature

Chaos is an irregular, seemingly random behavior. In order to verify the whether the evolving process of the share price is chaotic or not, the Lyapunov exponent can be extracted. If the Lyapunov exponent is greater than 0, the dynamic system is chaotic. The computed Lyapunov exponents of training set were shown in Table I.

TABLE I. LYAPUNOV COMPONENT AND EMBEDDING DIMENSION

Embedding dimension	5	6	7	8	9
Lyapunov component	0.062	0.050	0.044	0.042	0.039

From Table I, one can see the Lyapunov exponents are all greater than 0, showing that the slight changes of the share price will grow exponentially with the passage of time, lead to great changes. Therefore, the evolving process of share price of Xichang electricity is chaotic.

B. Electing Delay Time

The mutual information method is suggested to optimally choose τ , which determines the delay time when the mutual information first reaches minimum.

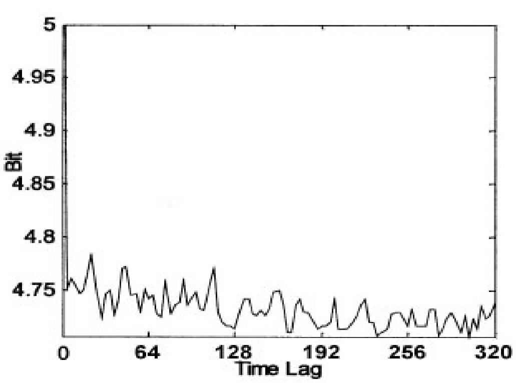

Fig.2 AMI Function for Daily Share Price

For the time series of the share price of Xichang Electricity stock, the mutual information is calculated and shown in Fig. 2. We can see from Fig.2 that the first minimum of the AMI function is three, so the time lag is three. At last, we use the daily share price of training set to predict those of test samples.

C. Decomposed by Wavelet Transform

With further scrutiny, one can find that the definite regularity is found behind the seemingly random. Besides, it also has random parts. This kind of characteristics inspires us to decompose the series by wavelet transform.

The decomposed level is hard to choose. After trying different level, we find: if the level is too low, prediction error will increase; when it increases, prediction result will be improved greatly, but when the level is more than three, there is only tiny improvement. At last, db3 is chosen as the mother wavelet and decomposed level is three. The decomposed series is shown in Fig.3.

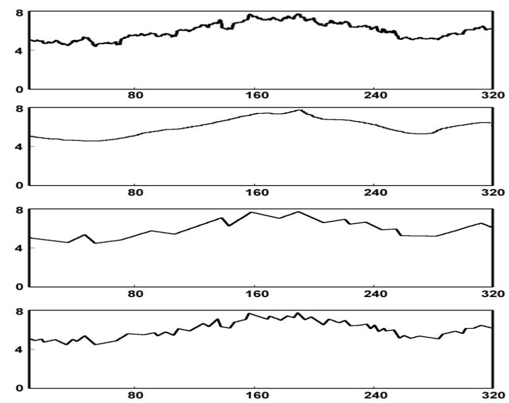

Fig.3 The Original Series (at the top) and Decomposed Series

D. Selecting Embedding Dimension

With the selected time lag, embedding dimension of each level is produced by Cao's method. The value of embedding dimension is shown in Table II. It can be seen from Table II that, with the higher decomposition level, there is the larger embedding dimension. That is because the information in high frequency parts is contaminated by the noise, and thus more data points are needed to predict the value of next day.

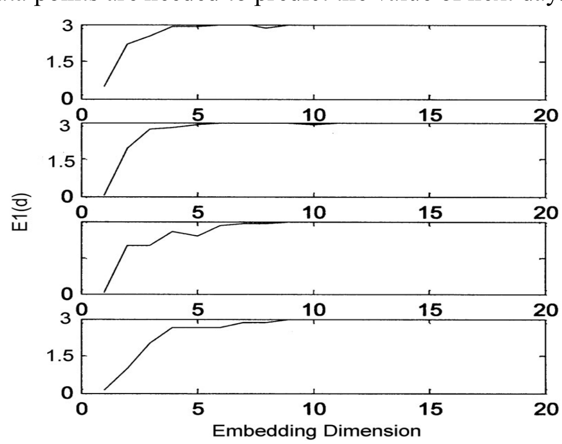

Fig.4 E1(d) graph of each lever produced by Cao's method. From top to bottom, they belong to: approximate parts, detail parts in 3,2,1 respectively.

E. Share Price Predicting based on WT-SVM

In this paper, we chose the RBF as the kernel function. By using the obtained time lag τ and embedding dimension m and according to (3), the input vectors and output vectors for LS-SVM predictor are obtained by Libsvm toolbox. The parameters of LS-SVM predictor for each part are shown in Table II.

TABLE II. PARAMETERS OF EACH LEVER

LS-SVM predictor of each part	C	ε	σ	m
Approximate parts	187	0.031	0.16	4
Detail parts in level 3	151	0.025	0.17	5
Detail parts in level 2	137	0.023	0.13	7
Detail parts in level 1	126	0.019	0.11	8

The reconstruction of the predicted value of each part is used as the final forecasting result as shown in Fig.5. The comparison with the competition winners is also made and the result is shown in Table III.

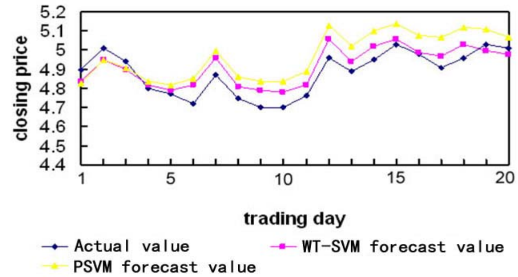

Fig.5 The Final Predicted Result

TABLE III. COMPARISON OF THE PROPOSED MODEL AND OTHER MODELS

Trading day	Actual value	Forecast value(m=8)	Error rate(%)	Forecast value(m=7)	Error rate(%)	Forecast value(m=5)	Error rate(%)	PSVM forecast value(m=8)	Error rate(%)
1	4.90	4.84	1.31	4.83	1.45	4.83	1.39	4.83	1.44
2	5.01	4.95	1.22	4.94	1.31	4.95	1.28	4.95	1.12
3	4.94	4.90	0.87	4.89	0.96	4.89	1.01	4.91	0.57
4	4.80	4.82	-0.39	4.79	0.24	4.79	0.18	4.84	-0.87
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	4.91	4.97	-1.22	5.03	-2.35	5.03	-2.79	5.07	-3.20
18	4.96	5.03	-1.51	5.09	-2.64	5.11	-2.98	5.12	-3.62
19	5.03	5.00	0.64	4.99	0.86	5.04	-0.26	5.11	-1.52
20	5.01	4.98	0.57	5.03	-0.42	5.06	-0.96	5.07	-1.14

It can be seen from Fig.5 and Table III that forecasting errors of share price are acceptable. We defined that the error rate is less than 2% is effective; there are 18 qualified outcomes when embedding dimension was evaluated to 8, and only 14 or 15 qualified outcomes when dimension was defined 7 or 5. The PSVM was also 15 qualified outcomes. On the other hand, the vibration amplitude of error rate using proposed model is lower than other models. These results show that the short-term forecast of share price is feasible and the forecasting performance of WT-SVM outperforms the PSVM model under the same condition.

V. CONCLUSION

In this study, many approaches are employed to improve the accuracy of share price forecast. Utilizing AMI function to select time lag and Cao's method to select embedding dimensions are much more reliable than conventional techniques. Wavelet transform has eliminated the fluctuant components of time series and decreased forecast error. The proposed WT-SVM model has great generalization ability and guarantees global optimality, thus enhancing the forecasting accuracy.

On the other hand, further research is necessary of this study. In the actual prediction, there are many factors having their influence on stock market. In addition, the parameter selection is the critical of predict work, applying advanced algorithm to optimize the parameters that can improve the predict accuracy.

REFERENCES

[1] Van Gestel, T, et al. "Financial time series prediction using least squares support vector machines within the evidence frame". *IEEE Trans.Neural Networks*, vol 12, no.4, pp.809-821, 2001.

[2] R.L. Wilson, R. Sharda, "Bankruptcy prediction using neural networks," *Decision Support Systems* vol.11, pp. 545–557, 1994.

[3] N. Cristiamini, J. Shawe-Taylor, "An Introduction to Support Vector Machines," *Cambridge university press*, Cambridge, England, 2000.

[4] Lü J H,Lu J A. Chaotic time series analysis and application [M]. Wuhan: publishing company of Wuhan University, 2002.

[5] Abarbanel, Henry D.I., "Analysis of Observed Chaotic Data" New York: Springer-Verlag, 1996.

[6] S. Mallet, A Wavlet Tour of Signal Processing Acadamic Press, San Diego, 1999.

[7] L.Y. Cao, "Practical Method for Determining the Minimum Embedding Dimension of a Scalor Time Series", *Physical D*, vol. 11O, pp.43-50, 1997.

[8] Fu-Chiang Tsui, Mingui Sun, Ching-Chung Li and Robert J.Sclabassi, "A comparative study of two biorthogonal wavelet transforms in time series prediction", *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, Oct. 1997, pp1791-1796.

[9] R.L. Milidiu, R.J. Machado, R.P. Rentera, Time-series forecasting through wavelets transformation and a mixture ofexpert models, Neurocomputing 28 (1999) 145–146.

[10] Kimoto, T., Asakawa, K., Yoda, M. and Takeoka, M. "Stock Market Prediction System with Modular neural Networks." Proceeding of the International Joint Conference on Neural Networks, 1990, Vol 1, pp. 1-6.