Short-term Stock Forecasting based on Fractal and Chaos Theory

Yingjie Wang[a], Yuqian Liu and Le Li

School of artificial intelligence, Zhengzhou Railway vocational & technical college, Zhengzhou 450000, China.

[a]wyjlw18@163.com

Abstract

The fractal and chaos theories has been applied for stock forecasting. However, traditional methods exists several disadvantages such as vertical scale factor difficult to calculate, low-precision. Therefore, a method is proposed combined with fractal theory and chaos theory. In this paper, the theory of phase space reconstruction is used to select sample data, the Hurst parameter values is used to calculate vertical scaling factors in Iterative Function Systems (IFS) of each group sample data, and we can achieve the statistical IFS according weight, then the stock forecasting curve was generated by the iterations system. This algorithm was used to forecast stock from three aspects: phase space reconstruction, fractal interpolation and fractal extrapolation. Compared with the existing fractal algorithm, the simulation result shows that the reconstructed sample has better self-similarity and the forecast accuracy of this method is improved, and it can be applied in the field of short-term stock prediction.

Keywords

The Fractal Theories; The Chaos Theories; short-term Stock forecasting.

1. Introduction

Stock prediction is a difficult task due to the nature of the stock data which is noisy, non-stationary, highly uncertain and time varying[1]. Many factors interact in finance including political events, general economic conditions, and traders' expectations. Therefore, improving the technical level of stock forecasting can accurately analysis stock market and predict the stock trend, index or price.

In recent years, many different methods and techniques have been presented. These methods include artificial intelligence theory, fractal theory, chaos theory, support vector machine, artificial neural network and so on[1-5]. Since Mandelbrot created fractal geometry, which as a branch of nonlinear theory have penetrated into data compression, time series prediction, nonlinear system modeling, and many other branches[6,7]. Fractal interpolation proposed by M.F.Barnsley is a method of constructing fractal curve based on iterated function system[8], HY Wang used R/S analysis and Hurst exponent analyze the structure feature of the Shanghai Composite Index and establish fractal interpolation models predict the tendency of change of the index in the short term[9]. Ming-Yue Zhai proposed a method combined with self-similarity theory and fractal interpolation theory to solve load forecasting[10]. Liang Jian-Kai select historical similar days' loads to predict power load based on fractal interpolation theory[11].

The phase space reconstruction theory was applied in time series prediction in recent years. G Guo used the chaotic theory discover chaotic characteristics of stock price time series[12]. CH Peng proposed a railway passenger flow prediction algorithm combined with the phase space reconstruction and similarity principle[13]. A hybrid forecasting model based on fractal prediction model and chaos prediction model linear combination was proposed in [14]. However most analysis and predict stock based on fractal theory or chaos theory is respective, and the result of prediction is not very good.

Therefore, a method for short-term stock forecasting based on fractal and chaos theory is proposed in this paper from three aspects: phase space reconstruction, fractal interpolation and fractal extrapolation. This algorithm uses phase space reconstruction theory to select sample data, and uses the Hurst parameter values to calculate vertical scaling factors in Iterative Function Systems (IFS), then predicts short-term stock by fractal extrapolation. The simulation results show this algorithm improves the prediction accuracy.

2. The Fractal and Chaos Theory

2.1 The Fractal Dimension

A data set has fractal character if its local has the same property such as structure, statistic distribution with the whole [15]. The fractal dimension is a measure of complexity, roughness, and irregularity of a fractal data set, and the box counting algorithm is often used to estimate it. Suppose that the $N(r)$ is the number of boxes required to cover the set, the boxes' radius is r and $N(r)$: r^{-D_0}. Then the box-counting dimension is defined as

$$D_0 = \lim_{r \to 0} \frac{\ln N(r)}{\ln(1/r)} \tag{1}$$

2.2 The Fractal Interpolation

In 1986, Barnsley proposed the fractal interpolation method based on fractal collage principle. The fractal interpolation algorithm can construct an iterated function system(IFS) over the entire range, so it can maintain the most characteristic of the original sample curve and the sample interpolation points can be displayed with rich details.

Fractal interpolation method is based on the theory of iterated function systems. For data sets: $\{(x_i, y_i), i = 0,1,\dots,n\}$, an iterated function system could be constructed, we need use the affine transformation to achieve an iterated function system. Each w_i in IFS is a affine transform function, which is given by the following structure formula [11]:

$$w_i \begin{bmatrix} x \\ y \end{bmatrix} = \begin{pmatrix} a_i & 0 \\ c_i & d_i \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \end{bmatrix} \tag{2}$$

where is a coordinate, a_i, c_i, d_i, e_i, f_i are transformation ratio of w_i, The d_i is a free variable called the vertical scaling factor. When d_i is given, the other parameters can be expressed as

$$\begin{aligned} a_i &= \frac{1}{x_n - x_0}(x_i - x_{i-1}), \\ c_i &= \frac{1}{x_n - x_0}(y_i - y_{i-1}) - \frac{d_i}{x_n - x_0}(y_n - y_0), \\ e_i &= \frac{1}{x_n - x_0}(x_n x_{i-1} - x_0 x_i), \\ f_i &= \frac{1}{x_n - x_0}(x_n y_{i-1} - x_0 y_i) - \frac{d_i}{x_n - x_0}(x_n y_0 - x_0 y_n), \end{aligned} \tag{3}$$

From Eq.(3), the selection of d_i has a great influence on the calculation of the other four parameters of affine transformation. Therefore, an accurate estimate of d_i is particularly important for predicting results.

2.3 The Method of Seeking Attractor

According to fractal theory, G is the attractor of the iterated function system, if $\sum_{i=1}^{n}|d_i| > 1$, and the interpolation points are not collinear. Then fractal dimension of the fractal interpolation function attractor satisfies the equation:

$$\sum_{i=1}^{n} |d_i| a_i^{D-1} = 1 \tag{4}$$

For Eq. (4), we can use the fractal box dimension D to calculate d_i. However, the fractal box dimension D and the Hurst value H has the following relationship

$$D = 2 - H \tag{5}$$

Therefore, use an accurate estimate of value H to calculate d_i is a reasonably simple method, assuming that each vertical scaling factor value d_i equal to the size, then:

$$d_1 = d_2 = \cdots d_n = \frac{1}{\sum_{i=1}^{n} a_i^{1-H}} \tag{6}$$

According to Eq.(6), after estimate the parameters Hurst, we will get the value of d_i. Then calculate the other iterative parameters by Eq.(3), the complete iterated function systems will be constructed.

2.4 The Phase space reconstruction

The basic theory of phase space reconstruction is Takens Embedding Theorem[16], which provides the theoretical foundation for the analysis of time series generated by nonlinear deterministic dynamical systems. Informally, according to the embedding theory, for a scalar time series $\{x_i, i = 1,2, \dots, n\}$, the phase space can be reconstructed according to:

$$y_j = \left\{x_j, x_{j+\tau}, x_{j+2\tau}, \dots, x_{j+(m-1)\tau}\right\} \tag{7}$$

where $j = 1,2, \dots, N\text{-}(m\text{-}1)\tau$, m is the dimension of the vector y_j, called as embedding dimension, and τ is the delay time.

The phase space reconstruction can reduce the redundancy of the original data and not lose the relationship between points, at the same time, it improves the similarity between the samples by reconstructing a single-variable series into a multi-dimensional phase space, and reduces the error of fractal prediction.

3. The design of short-term stock forecasting based on fractal and chaos theory

3.1 Data sample selection based on phase space reconstruction

Short-term forecasting relies on large amounts of historical data, we select and collect the original data regardless of what methods forecasters adopt, and the true and reliable degree of data will have great influence on the extent of these predictions and directly affect the accuracy of prediction. In the paper, we get the original data from trading software, and select data sample by phase space reconstruction.

The phase space reconstruction is a method to extract the dynamic characteristics of the original system in time series by a series of nonlinear approximation algorithms. Suppose the length of time series is N, the embedding dimension is m and the delay time is τ, then we can get $N\text{-}(m\text{-}1)\tau$ groups reconstructed series that length is $m$by phase space reconstruction. According to the embedding theory, reconstructed vector set keeps the main features of the original space, and m is the minimum number of elements in reconstructed vector set, so we make m as radix to obtain the sample(the length of sample data is the integer times of m). By this way, we can get multigroup sample data.

3.2 The Design of short-term stock forecasting based on fractal and chaos theory

The basic steps of algorithm based on fractal and chaos theory used in stock forecasting are as follows.

Step1. Fractal characteristic analysis, use R/S algorithm to analyze the time series.

Step2. Obtain samples, calculate the m and τ of original data set, according to phase space reconstruction, we can get multigroup historical sample data.

Step3. Determine the interpolation points. Take the sampling time order as the abscissa, the value as the ordinate in each group sample, and these points are interpolation points.

Step4. Normalizing the sample data according formula:

$$\begin{aligned} x_i &= \frac{X - X_{\min}}{X_{\max} - X_{\min}} (i = 1, 2, \cdots, n) \\ y_i &= \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} (i = 1, 2, \cdots, n) \end{aligned} \tag{8}$$

where (X, Y) is interpolation points, n is the length of sample data.

Step5. Calculate fractal interpolation parameters by Eq.(6) and Eq(3), and establish IFS of each group sample.

Step6. Obtaining the weighting average fractal interpolation parameters of each group sample, get a statistical meaning of IFS.

Step7. Predicting, according fractal extrapolation, suppose (X_{n+1}, Y_{n+1}) is extrapolation point, the initial value of Y_{n+1} is 0, and the Y_{n+1} is changing with the $\delta(0 < \delta < 1)$, make (X_{n+1}, Y_{n+1}) as new interpolation point establish IFS and draw the curve. The average error between prediction curve and historical curve is:

$$e = \frac{1}{n+1}\left(\sum_{i=0}^{n} |y_i{}^* - y_i|\right) \tag{9}$$

where y_i is the predictive value, $y_i{}^*$ is the historical value, n is the number of forecast points. If the average error is least, the Y_{n+1} is the prediction value, and we can get the stock price by formula (8).

4. Experiments

This section will predict the stock price and prove the accuracy of short-term stock forecasting based on fractal theory and chaos theory. The original data are provided by Great Wisdom 365. The calculation process is as shown in Fig.1.

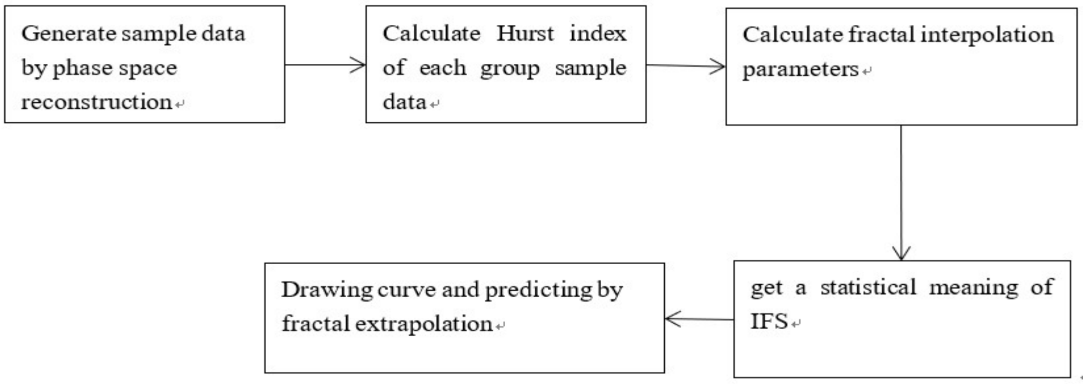

Fig.1 The calculation process

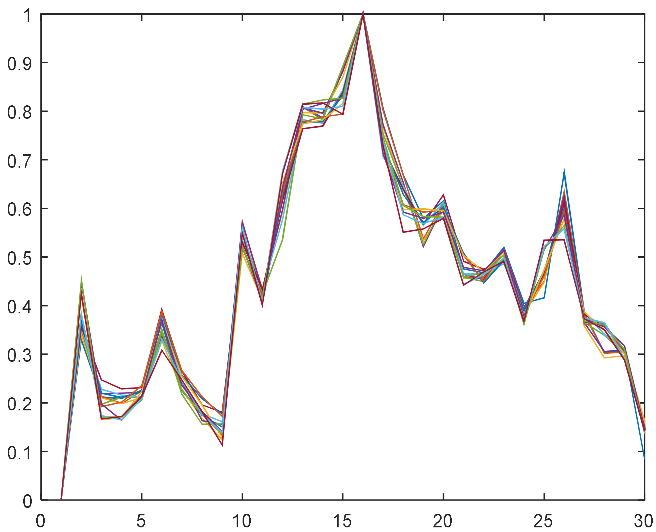

Fig. 2 The reconstructed sample data

4.1 The data sample selection

In this paper, we choose the data from December 26 of 1996 to June 24 of 2005 as the original data, and calculate the delay times τ and the embedding dimension m by using auto correlation algorithm and G-P algorithm. The Fig.2 is the 14 groups sample data after phase space reconstruction, and the length of data is 30(the five times of m). As shows in Fig.2, the reconstructed data has very good self-similarity, and it's reliable sample data for fractal interpolation.

4.2 The IFS of sample data

After phase space reconstruction, we choose the 13 groups sample data, and make the last group as testing sample. As table1 shows that we calculate the the Hurst parameter by R/S algorithm, and calculate the d_i according formula (6), and it shows that the data set has good self-similarity. The Table2 is the IFS code of one group sample by fractal interpolation method.

Table 1. The d_i and Hurst parameter of sample data

Data set	Hurst parameter	d_i
1	0.64556	0.1517
2	0.68016	0.1375
3	0.69141	0.1332
4	0.69417	0.1322
5	0.71945	0.1230
6	0.69512	0.1318
7	0.6748	0.1396
8	0.71371	0.1250
9	0.67843	0.1382
10	0.67127	0.1376
11	0.7231	0.1289
12	0.6867	0.1429
13	0.7482	0.1201
14	0.6935	0.1402

Table 2. The IFS code of one group sample data

IFS	a_i	c_i	d_i	e_i	f_i
1	0.0345	0.0140	0.1201	0.9655	-0.0140
2	0.0345	-0.0095	0.1201	1.9655	0.4337
3	0.0690	0.0010	0.1201	2.9310	0.1656
4	0.0345	0.0027	0.1201	4.9655	0.2110
5	0.1034	-0.0073	0.1201	5.8966	0.3154
6	0.0345	0.0144	0.1201	8.9655	0.0985
7	0.0345	-0.0057	0.1201	9.9655	0.5546
8	0.0345	0.0088	0.1201	10.9655	0.3924
9	0.1034	0.0036	0.1201	11.8966	0.6685
10	0.0345	0.0065	0.1201	14.9655	0.7871
11	0.0690	-0.0161	0.1201	15.9310	1.0161
12	0.0345	-0.0036	0.1201	17.9655	0.5513
13	0.0690	-0.0046	0.1201	18.9310	0.5622
14	0.1034	-0.0031	0.1201	20.8966	0.4453
15	0.0345	0.0051	0.1201	23.9655	0.3629
16	0.0690	-0.0062	0.1201	24.9310	0.5406
17	0.1034	-0.0085	0.1201	26.8966	0.3816

The sample data is reconstructed, and the sampling interval is same, so the weight of each group sample data is same, and then we can get the statistical meaning of IFS. The Figure3 is the fractal interpolation curve of one group sample data.

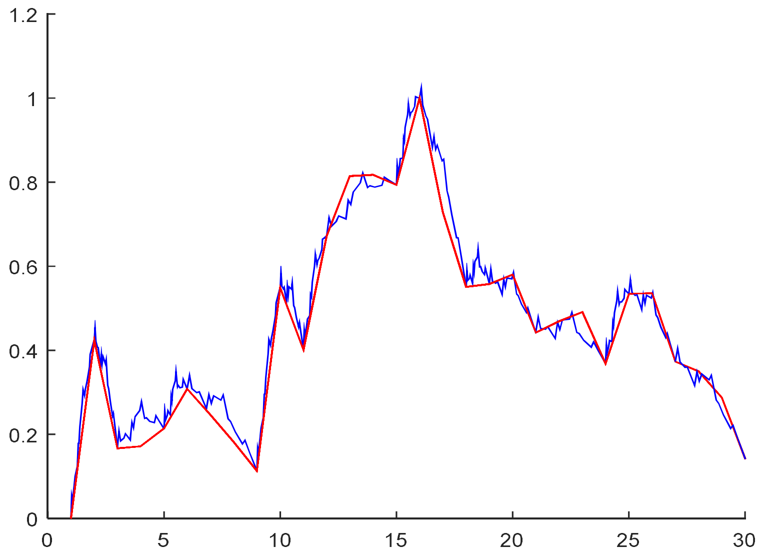

Fig. 3 The fractal interpolation curve of one group sample data

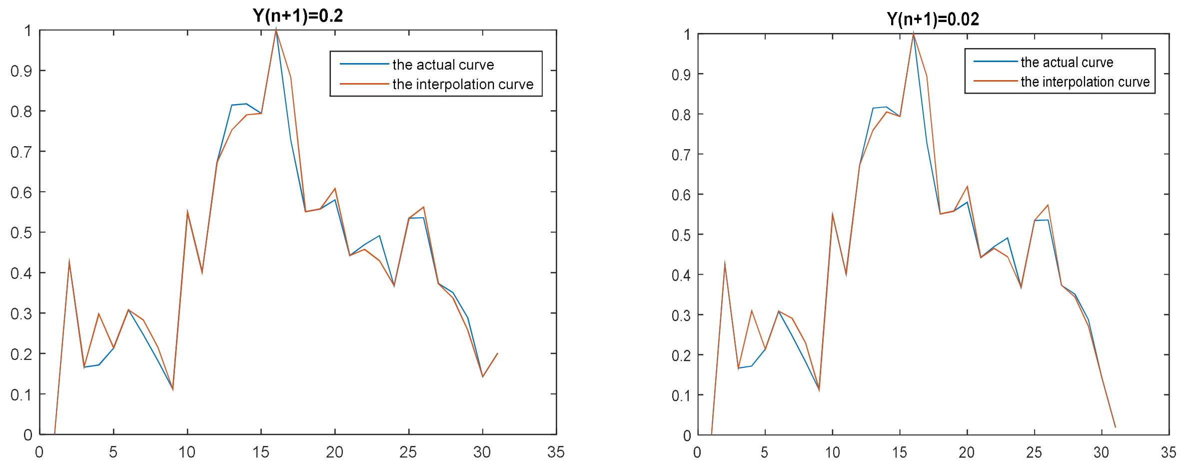

Fig.4 The interpolation value and actual value

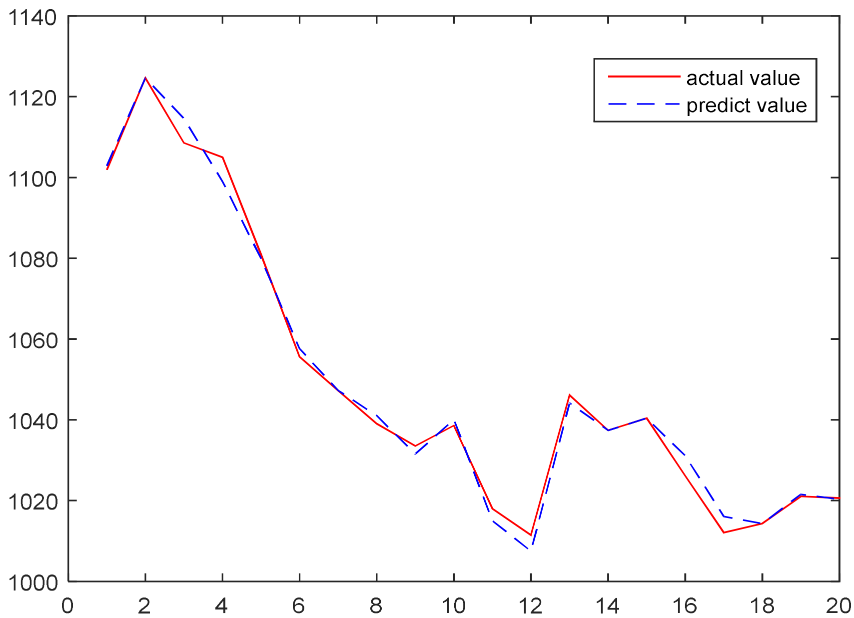

Fig. 5 The actual and predicted values curve

4.3 Forecasting

Suppose the $Y_{n+1} = 0$, make $\delta = 0.02$, and the Y_{n+1} is changing with δ, according to the fractal extrapolation, make (X_{n+1}, Y_{n+1}) as a new interpolation point to establish IFS and draw the curve. The figure 4 is the interpolation value and actual value with different Y_{n+1}.

As shows in figure4, the interpolation curve is different with Y_{n+1} changing from 0 to 1. When Y_{n+1} is 0.2, the error of interpolation curve and actual curve is least by formula(8). According the formula(9), we can get the prediction value is 1176.21, the actual value of June 27 of 2005 is 1124.64, the relative error is 0.45%.

By this way, we predict the 20 days' price after June 26 of 2005. As shows in figure 5, the prediction accuracy rate of this algorithm is high. The result shows that this algorithm can be used for short-term stock forecasting and improve the precision of prediction.

5. Conclusions

Fractal theory has a great advantage in the study of nonlinear systems. In the paper, combining the fractal characteristics with phase space reconstruction, we can have a deep study on algorithm and forecasting steps. Three main conclusions of this approach are as follows.

(1) Through the concept of fractal theory, we introduce the ways that using the Hurst parameter to estimate the vertical scale factor is feasible for forecasting.

(2) The phase space reconstruction is applied to select data sample, and it improve the self-similarity of sample. Comparing with the existing algorithm, the relative error of this algorithm is smaller, it's better for stock forecasting.

(3) Experiment shows that the algorithm in this paper have advantages of high-precision and easy to use. However, the procedure is a little complicated. The algorithm based on phase space reconstruction is more useful to little self-similarity time series, expand the application of fractal prediction.

Acknowledgments

This work was supported by the research on talent training mode reconstruction of artificial intelligence technology service specialty under the background of industrial transformation (2020JG68).

References

[1] G. J. Deboeck. Trading on the edge: neural,genetic, and fuzzy systems for chaotic 1nancial markets, John Wiley & Sons, Inc. New York, pp.45–65, 1994.

[2] A. S. Chena, M. T. Leungb and H. Daoukc. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index, Computers & Operations Research, vol.30, pp. 901–923, 2003.

[3] S. Günay. Fractal Structure of the Stock Markets of Leading Asian Countries, Journal of East Asian Economic Integration, vol. 18, no. 4 pp. 367-394, 2014.

[4] K. Kim. Financial time series forecasting using support vector machines, Neurocomputing, vol.55, pp. 307 – 319, 2003 .

[5] A. Kazem and E. Sharifi. Support vector regression with chaos-based firefly algorithm for stock market price forecasting, Applied Soft Computing, vol. 13, no. 2, pp. 947-958, 2013.

[6] H. Shen, B. Ma, and F. Liu. Study on power system load characteristics based on fractal theory, Zhejiang Electric Power, vol. 2, no. 2, pp. 15–17, 2007.

[7] M. Yamahara, T. Noguchi, M. Okawa, and N. Yamada. The relationship between subjective sleep disturbance and complexity of 24-hour activity utilizing fractal theory in psychiatric inpatients, Sleep and Biological Rhythms, vol. 7, no. 1, pp. 11–16, 2009.

[8] M. F. Barnsley. Fractal functions and interpolation[J], Constructive Approximation, vol.2, no.1, pp. 303-329, 1986.

[9] H. Y .Wang and Q. G. Zhang. Fractal Interpolation Models and R/S Analysis on Stock Index Time Series, Statistics & Information Forum, 2011.

[10] M. Y. Zhai. A new method for short-term load forecasting based on fractal interpretation and wavelet analysis, Electrical Power and Energy Systems, vol.69 , pp. 241–245, 2015 .

[11] L. J. Kai, C. Cattani and S. W. Qing. Power Load Prediction Based on Fractal Theory, Advances in Mathematical Physics, vol.2, pp. 1-6, 2015.

[12] G. Guo. STOCK MARKET MULTI STEP-FORECAST USING CHAOS THEORY, Information & Control, 2000.

[13] C. H. Peng. Prediction of railway passenger flow based on integrated chaotic phase space reconstruction and similarity theory [J], Journal of Wuhan University of Technology, vol. 31, no. 4, pp. 684-687, 2007.

[14] X. F. He , C. Gu and Y. Gu. Application of fractal chaos hybrid forecasting model in dam safety monitoring [J], Journal of Wuhan University, vol. 41, no. 1, pp. 45-49, 2008.

[15] Y. B. Bao, G. Yu, H. L.Sun and D. L.Wang. Performance optimization of fractal dimension based feature selection algorithm, In International conference on advances in web-age information management, Dalian, China pp. 739–744, 2004.

[16] J. Stark, D. S. Broomhead, M. E. Davies and J. Huke. Takens embedding theorems for forced and stochastic systems, Nonlinear Analysis, vol. 30, no. 9, pp. 5303-5314, 1997.