Time series forecasting based on deep extreme learning machine

Guo, Xuqi; Pang, Yusong; Yan, Gaowei; Qiao, Tiezhu

Time Series Forecasting Based on Deep Extreme Learning Machine

Xuqi Guo[1], Yusong Pang[2], Gaowei Yan[3], Tiezhu Qiao[4]

1. College of Information Engineering, Taiyuan University of Technology, Taiyuan, 030600 E-mail: guoxuqi0330@link.tyut.edu.cn

2. Section Transport Engineering and Logistics, Delft University of Technology, Netherlands, 2628CD E-mail: Y.Pang@tudelft.nl

3. College of Information Engineering, Taiyuan University of Technology, Taiyuan, 030600 E-mail: yangaowei@tyut.edu.cn

4. College of Information Engineering, Taiyuan University of Technology, Taiyuan, 030024 E-mail: qtz2007@126.com

Abstract:

Multi-layer Artificial Neural Networks (ANN) has caught widespread attention as a new method for time series forecasting due to the ability of approximating any nonlinear function. In this paper, a new local time series prediction model is established with the nearest neighbor domain theory, in which the hybrid Euclidean distance is used as the similarity measurement between two sets of time series. In order to improve the efficiency, prediction performance, as well as the ability of real-time updating of the model, in this paper, the recombination samples of the model is derived by Deep Extreme Learning Machine (DELM). The experiments show that local prediction model gets accurate results in one-step and multi-step forecasting, and the model has good generalization performance through the test on the five data sets selected from Time Series Database Library (TSDL).

Key Words: time series prediction, Deep Extreme Learning Machine, local model, hybrid Euclidean distance

1   INTRODUCTION

Time series prediction plays an important role in the financial, medical, industrial control and other decision-making areas in daily world. In the past several decades, researchers have done a lot of research on the statistical model for time series forecasting. Such as the Auto Regressive Moving Average (ARMA), and some derivative algorithm on the basis of this [1]. Besides, in order to solve the stochastic time series with nonlinear feature and non-white noise interference problems, the multi-layer perceptron neural network as an important method of machine learning, has become a popular network architecture in time series prediction [2-4], due to its approximation ability to complex nonlinear function and parallel processing capabilities. But the existing artificial neural network still has some defects in time series prediction problems.

There are two different modeling options for constructing forecasting models in time series prediction [5]: one is the global modeling method, namely, to establish a global model independent from the forecasting sequence; and the other is local modeling approach, which constructs a local model for each prediction sequence respectively. The local modeling method, which employ the nearest neighbor domain theory as the core, has higher prediction accuracy for multi-variable function. Therefore, in this article, we use the hybrid Euclidean distance, based on the theory of nearest neighbor, as the similarity measurement to complete the reorganization of all training samples. The hybrid Euclidean distance does not only take into account the similarity of the two groups of time series, but also related to the similarity of time series trend.

Extreme Learning Machine (ELM) is a new fast learning algorithm for single-hidden layer feedforward neural networks, which has been widely used in a variety of classification and regression tasks due to its fast learning speed and high computational efficiency. In order to capture the input data in high correlation characteristics and make it no longer limited by its single hidden layer structure, in this paper the forecasting model of recombination samples is derived by DELM.

Above all, the hybrid Euclidean distance is used as the similarity measurement to complete the reorganization of al-l training samples in this paper, the prediction model of recombination samples is derived by DELM, and a corresponding learning algorithm is given. Finally, we use multiple data sets as an example to verify the effectiveness of the local prediction model and DELM learning algorithm, the experimental results prove that the DELM local modeling method can enhance the efficiency of massive data time series prediction modeling, improve the prediction performance and the ability of real-time updates of the model.

2   TIME SERIES FORECASTING

Given a set of observations from a process:

$$X_0, Y_0, X_1, Y_1, \ldots, X_t, Y_t \qquad (1)$$

Table 1: Relationship between input and output of the prediction model

Input	Output
$S_1 = [X_0, Y_0, X_1, Y_1, \cdots, X_{q-1}, Y_{q-1}, X_q, Y_q,]$	Y_{q+s}
$S_2 = [X_1, Y_1, X_2, Y_2, \cdots, X_q, Y_q, X_{q+1}, Y_{q+1}]$	Y_{q+s+1}
$\cdots$	$\cdots$
$S_z = [X_{t-s-q}, Y_{t-s-q}, X_{t-s-q+1}, Y_{t-s-q+1}, \cdots, X_{t-s-1}, Y_{t-s-1}, X_{t-s}, Y_{t-s}]$	Y_t

Assume that the time series collected at equally spaced time Δt, where Y_t represent the observed values of the output variables at time $t_0 + i\Delta t$, and $X_i \in R^n$ represents the n observed values associated with the output variables at time $t_0 + i\Delta t$. The intention of time series prediction is to estimate the value of the output variables at a certain future time based on the historical observation data, such as at the $t + s$ moment,

$$\hat{Y}_{t+s} = G(X_{t-q}, Y_{t-q}, \ldots, X_{t-1}, Y_{t-1}, X_t, Y_t) \quad (2)$$

In which, q is the time span of the historical data, that is, the forecast delay, and the G represents the forecast model or function. If s = 1, the next moment of the output variable value is predicted, which is called single step prediction; if s ¿ 1, then called this multi-step prediction. When $n = 0$ is known as a single variable prediction, otherwise, called the multivariate prediction.

3   PROBLEM STATEMENT

In order to make the model more universal, assume that $n = 1$ means the observed values include an output variable Y and a relevant variable X, which is similar to the modeling method with two or more relevant variables. The purpose of the modeling is to estimate the output variables at the time $t + s$ by the historical data, and the corresponding relationship between input and output of the model is shown in Table 1.
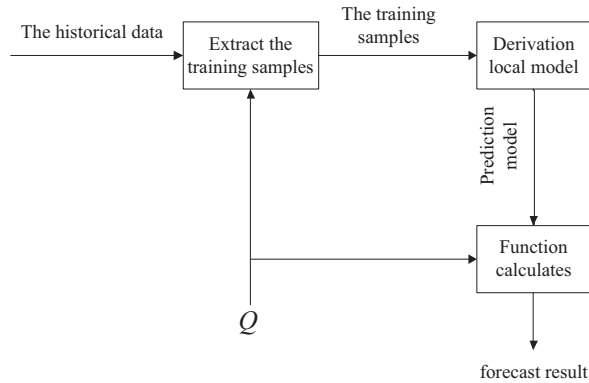

Figure 1: Local prediction model

In which, the sequence $Q = [X_{t-q}, Y_{t-q}, \ldots, X_{t-1}, Y_{t-1}, X_t, Y_t]$ is called the prediction sequence, and the set $S = \{S_1, S_2, \ldots, S_z\}$ is called the nearest neighbor set of Q. In local modeling method we should choose k elements from the z elements of the nearest neighbor set S as $Q's$ nearest neighbor, and regard the k elements as training samples of a local model.

In order to deduce the prediction model of variable Y_{t+s}, we need to employ a machine learning algorithm based on data as local modeling method, as shown in Figure 1. First of all, according to the similarity measurement we extract the samples which are most similar with the predict sequence from the historical data as the recombination samples; Then, according to the recombination samples extracted establish the local prediction model of the time series. In the prediction stage, calculating Y_{t+s} according to the forecasting model was derived above, the algorithm process as shown in Algorithm 1.

Algorithm 1 The proposed prediction algorithm

Input: *A time series* $\mathbf{y} = \{y_1, y_2, \ldots, y_r, y_{r+1}, \ldots, y_t\}$.
Output: *The estimated value* $\hat{y}_{t+s}$.
Step 1. *Preprocess the dataset* $\mathbf{y}$ *and obtain query sequence* $Q = [Y_t, Y_{t-1}, \ldots, Y_{t-r}, Y_{t-r+1}]$.
Step 2. *Calculate the distance between* Q' *and each sample from the dataset.*
Step 3. *Select the first k sequence from s to form neighbor set.*
Step 4. *Train DELM with the neighbor set.*
Step 5. *Calculate* $\hat{y}_{t+s}$ *according to the predictive model derived.*

---

4   TIME SERIES FORECASTING MODEL

In this paper, the local modeling approach is used for time series forecasting, and the forecasting process of the local model is shown in Figure 2. The local prediction model with nearest neighbor domain theory as the core, in which the hybrid Euclidean distance is used as the similarity measurement between two sets of time series and the prediction model of recombination samples is derived by DELM.

4.1   Extract the Training Samples

In order to predict the output variables Y_{t+s} on the $t+s$ moment, we need to extract k samples from all training samples which are the the most similar with the predict sequence Q of Y_{t+s} as the reorganization of training samples, the k recombinant samples are regarded as the input of local model .

In this section we use the hybrid Euclidean distance mentioned in the literature [7] to calculate the similarity between the two groups of time series. Assume that $\mathbf{A}_1$ and $\mathbf{A}_2$ as two sequences of time series. $\mathbf{F}_1$ and $\mathbf{F}_2$ are the differential sequences of the two time series $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively. And $N_E(A_1, A_2)$ is the standardized Euclidean distance between $\mathbf{A}_1$ and $\mathbf{A}_2$, and $N_E(F_1, F_2)$ is the standardized Euclidean distance between $\mathbf{F}_1$ and $\mathbf{F}_2$.

$$\mathrm{N}_H(\mathbf{A}_1, \mathbf{A}_2) = \frac{\mathrm{N}_E(\mathbf{A}_1, \mathbf{A}_2) + \mathrm{N}_E(\mathbf{F}_1, \mathbf{F}_2)}{2} \quad (3)$$

In which $N_H(A_1, A_2)$ is the hybrid Euclidean distance between the two groups of time series $\mathbf{A}_1$ and $\mathbf{A}_2$. The hybrid Euclidean distance not only consider the distance of two groups of time sequence, the addition of $N_E(F_1, F_2)$ makes it possible for the hybrid Euclidean distance to measure the similarity of the trend between the two groups of sequences rising or falling.

4.2 Derivation of DELM Model

In this paper, the DELM is based on the ELM show in [8-10], combines the idea of self coding. The output is coded by minimizing the reconstruction error, so that it could approach the original input infinitely. We can capture the depth features of the original input due to its structure which can obtain the abstract expression of the input.

Figure 3 describes the process of the characteristic expression of multi-layer ELM, regard the extracted training samples $\{\mathbf{S}_i\}_{i=1}^k$ as a input of Multi-ELM network, assuming that the network consists of h hidden-layers, $W = \{W_1, W_2, \cdots, W_{h+1}\}$ represents the weight parameters of network need to learn, and it must be determined while derive the Multi-ELM model.

Each layer of the network can be decoupled as an independent ELM, and to achieve the goal of making output T equal to the input for each ELM. In this way, the low dimensional expression of the input data can be obtained, that is, the output of the hidden layer ELM is used as the input of the next layer. And the low dimensional feature of the input data is mapped back to its original space by minimizing the reconstruction error. Therefore, the mapping matrix is data driven, whose transpose could be regard as the weight parameter of network:

$$\mathbf{W}_k = \hat{\beta}_j^T (j = 1, \ldots, h) \qquad (4)$$

Finally, we can learn the feature representation, which is the outputs of the h_{th} hidden layer, as the hidden layer of a independent ELM, calculate the output weights $\hat{\beta}_{h+1}$ of the
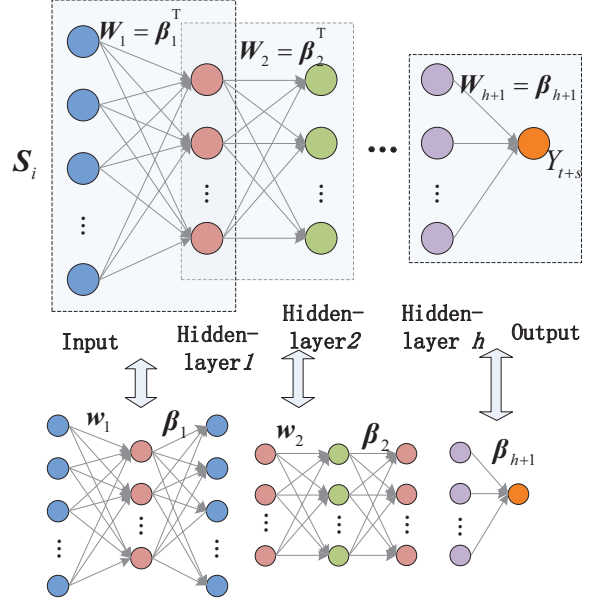

Figure 3: Local model based on DELM

ELM, then regard it as the weight parameter of the network.

$$\mathbf{W}_{h+1} = \hat{\beta}_{h+1} \qquad (5)$$

We can compute the Y_{t+s} according to the Multi-ELM model.

Table 2: The experiment data set division

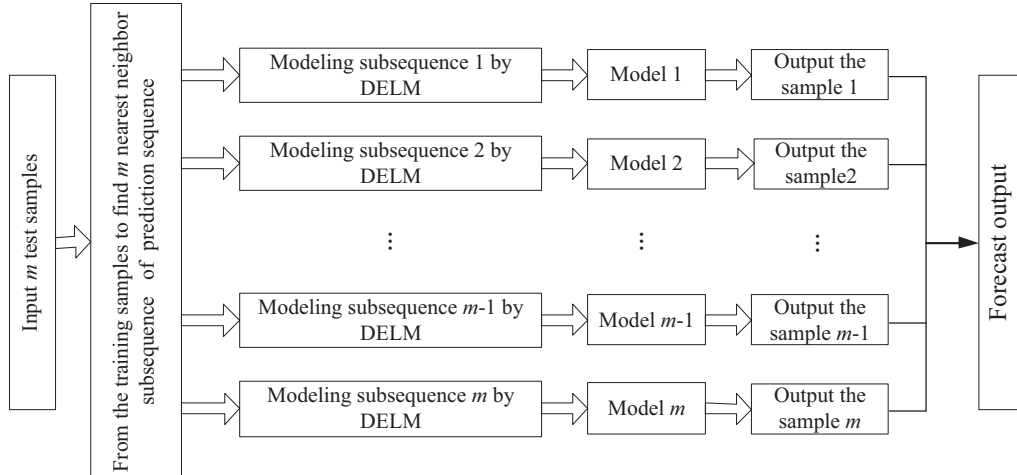| Datasets	Size of training set	Size of testing set
Sunspot	221	67
ESTSP2007	800	75
River Flow	740	76
Poland	1500	101
Darwin	1200	100

Figure 2: The Time series forecasting process based on the DELM

5 ALGORITHM EVALUATION AND APPLICATION

5.1 The Public Dataset

In this paper, we choose five datasets from the Time Series Database Library (TSDL), in which the data were collected by the actual environment. The five datasets can be described as: (1) The sunspot data set contains annual sunspot numbers from 1700 to 1987; (2) The ESTSP2007 data set from the 2007 European Symposium contest about time series prediction; (3) The monthly river flow data set was collected from Georgia Fortaleza ka'ao oostanaula river in 1893 to 1960; (4) Polish electric power data sets from the power load value in Poland in the 1990 s; (5) The Darwin pressure data set is the monthly atmospheric pressure of the Darwin sea level from 1882 to 1998. The specific datasets are divided as shown in Table 2.

5.2 EXPERIMENT RESULTS AND ANALYSES

In order to compare the hybrid Euclidean distance method with the traditional similarity measure method, we choose Dynamic Time Warping (DTW) and Likeness comparing method based on Cloud Model (LICM) to restructure the sample, and the local model is built based on the DELM, to serve as a comparison method based on hybrid Euclidean distance in the DELM model. And while intending to further test the validity of the DELM model, we use multiple regression methods to establish the local prediction model for the samples recombination by Hybrid Euclidean Distance, and the prediction results are compared with the DELM model. In order to quantify the performance of a variety of forecasting methods, the Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE) and the Mean Absolute Error (MAE) were used as a evaluation criteria. In addition, the experiment is given

from two aspects of single step prediction and multi-step prediction, to verify its generalization performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_t} (y_i - \hat{y}_i)^2}{N_t}} \quad (6)$$

$$MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} |y_i - \hat{y}_i| \quad (7)$$

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \quad (8)$$

In which, N_t is the number of test samples, y_i and $\hat{y}_i$ are predicted and true value respectively.

5.2.1 Single-step Prediction

Table 3 shows that in the single-step prediction, the RMSE and MAE of all the comparative method in those five data sets. To verify the effectiveness of the local prediction model DELM. We regard the ELM and Levenberg Marquardt error Back Propagation (LMBP) as comparison method.

As is shown in Table 3, DELM can always get the minimum value whether it is RMSE or MAE in the five datasets. Thus the DELM model can always achieve the most accurate prediction result compared with other methods. As for the three methods DTW, LICM and LMBP on different data sets have the advantage respectively. Through the comparison, RMSE and MAE of ELM is bigger, the prediction accuracy is poor.

While using the local model to forecast the time series, we choose the hybrid Euclidean distance as the similarity measurement criteria here, because the measure takes into account not only the distance but also the trend of the
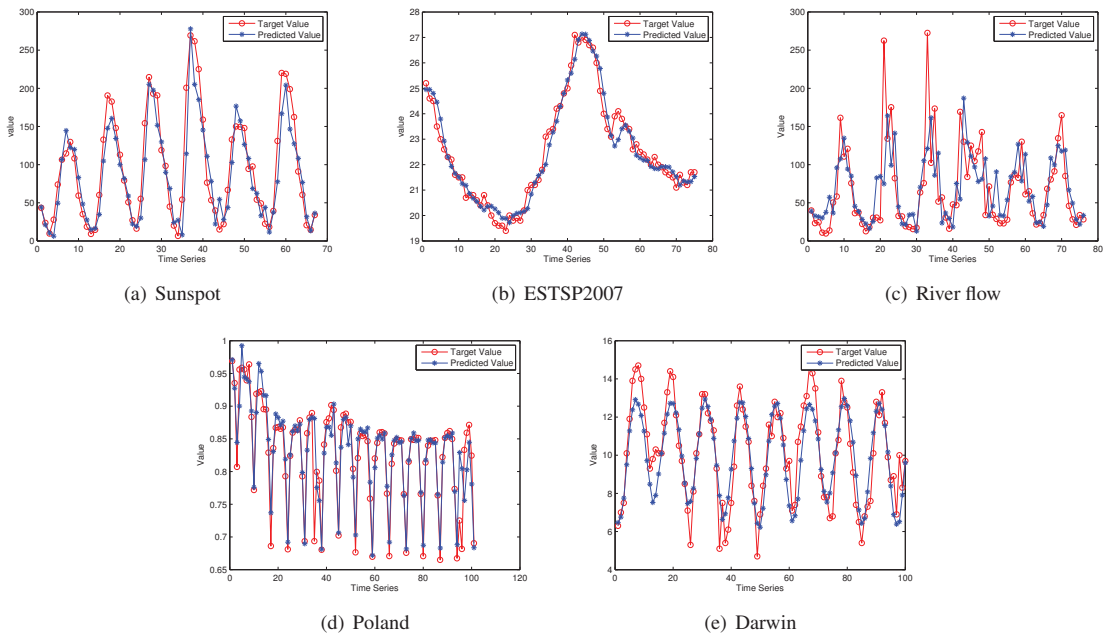

(a) Sunspot (b) ESTSP2007 (c) River flow

(d) Poland (e) Darwin

Figure 4: The prediction results of the five datasets

Table 3: The result of single-step prediction in the five datasets

Datasets	Error	DTW	LICM	Hybrid distance		
		DELM	DELM	ELM	LMBP	DELM
Sunspot	RMSE	43.916	46.158	53.157	36.701	21.948
	MAE	31.940	31.780	43.183	27.404	21.804
ESTSP2007	RMSE	0.5068	0.4483	0.9693	0.4162	0.3764
	MAE	0.0172	0.0158	0.0335	0.0143	0.0128
River flow	RMSE	56.411	51.812	170.04	53.737	41.634
	MAE	40.181	37.273	123.86	39.269	27.727
Poland	RMSE	0.0336	0.0391	0.5145	0.0371	0.0281
	MAE	0.0221	0.0213	0.1904	0.0224	0.0146
Sunspot	RMSE	1.4149	1.3594	2.0062	1.3280	1.1163
	MAE	1.0814	1.0641	1.4886	1.0376	0.9062

sequences involved. Therefore, it can be concluded that, compared to the methods DTW and LICM, the hybrid Euclidean distance method can extract the recombinant training samples closest to the test samples. And DELM can extract deep information of time series through the multi-layer study of the characteristics expression, which break the limitations that the ordinary ELM model is difficult to capture the high correlation characteristics due to its single hidden layer structure. DELM greatly improves the prediction accuracy of forecasting model, and the predicting performance is far better than the LMBP method.

Figure 4 is the comparison of the predicted results and the actual output of the DELM model based on the hybrid Euclidean distance for the five datasets.

5.2.2 Multi-step Prediction

With the purpose of validating the generalization performance of the model, and test the prediction accuracy of the model in the multi-step prediction, in this section we have

a experiment of multi-step prediction based on the single step prediction, with the Sunspot data set as an example.

And in order to detect whether the hybrid Euclidean distance method can extract the most suitable training samples as the reorganization sample in multi-step prediction, this paper make a comparison on the MAE of hybrid Euclidean distance, DTW, and LICM on every prediction point, as is shown in Figure 5 and 6. In selected prediction step length s is 4, 8, and 12 respectively, it is clearly from the figure, the error value of hybrid Euclidean distance method was lower than that of DTW or LICM, therefore, can be concluded that using the hybrid Euclidean distance method can get more accurate prediction results. In the local prediction model contrast of multi-step prediction, this paper references the NN, ANFIS, the LSSVM and ELM as the contrast method. Table 4 is the NRMSE value of various local prediction models for the predicting step length s from 1 to 12. It is clear that, compared to other local models, DELM can get a lower prediction error for the most s values, as is shown in Figure 7.
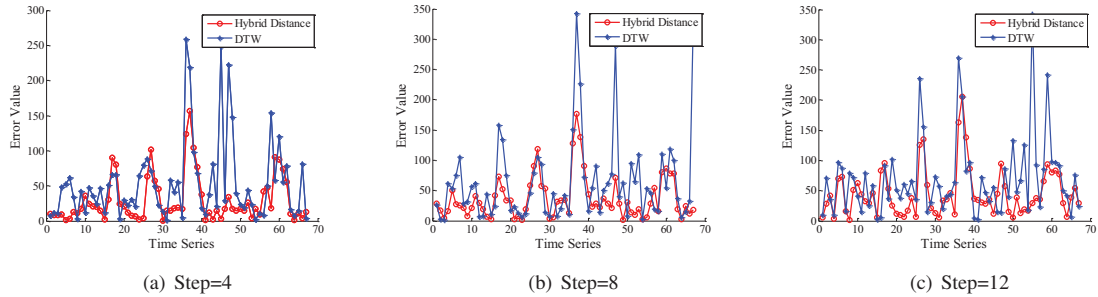

(a) Step=4   (b) Step=8   (c) Step=12

Figure 5: Error comparison of the hybrid Euclidean distance method and DTW in multi-step prediction
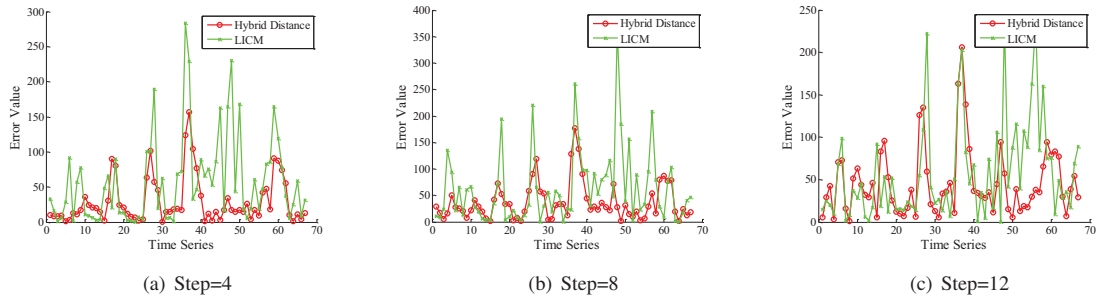

(a) Step=4   (b) Step=8   (c) Step=12

Figure 6: Error comparison of the hybrid Euclidean distance method and LICM in multi-step prediction
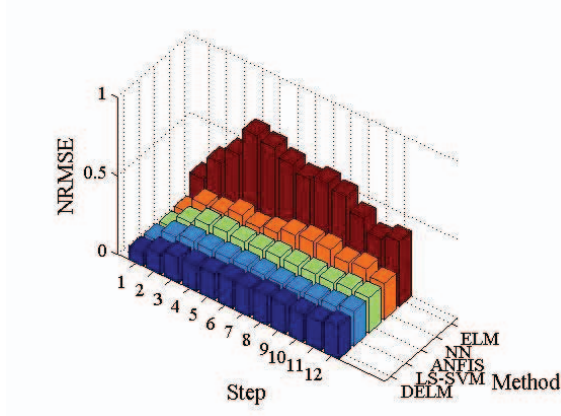
Figure 7: Comparison of various local prediction models in multi step prediction

Table 4: Multi step prediction results of local prediction model

Step	NN	ANFIS	LS-SVM	ELM	DELM
1	0.0856	**0.0738**	0.0768	0.2155	0.0815
2	0.1861	0.1595	0.1526	0.3690	**0.1324**
3	0.1893	0.2003	0.1794	0.4789	**0.1591**
4	0.2312	0.2077	0.1777	0.6534	**0.1612**
5	0.2054	0.1999	0.1800	0.6277	**0.1660**
6	0.2269	0.2023	0.1829	0.5775	**0.1796**
7	0.2660	0.2109	0.1822	0.5350	**0.1802**
8	0.2799	0.2088	**0.1797**	0.5656	0.1879
9	0.2824	0.2109	**0.1817**	0.5362	0.1849
10	0.2521	0.2086	0.1859	0.4495	**0.1778**
11	0.2676	0.2219	0.1962	0.3726	**0.1819**
12	0.2283	0.2220	0.2181	0.4113	**0.2050**

6 CONCLUSION

A new local prediction model based on the nearest neighbor domain theory for time series is established in this paper. In which the hybrid Euclidean distance is used as the similarity measurement, and the forecasting model is derived by DELM. We use the hybrid Euclidean distance to recombine training samples of the prediction, because it consider not only the distance but also the trend of the sequences involved. In order to improve the modeling efficiency, prediction performance, as well as the ability of real-time updating of the model, the training patterns is fed to DELM and the time series prediction model is derived. So as to verify the prediction accuracy of the model, we select five datasets from the TSDL as the experimental datasets, and test the generalization performance of the prediction model derived in single step and multi-step prediction respectively.

According to the experiment, we can conclude that the hybrid Euclidean distance method can extract the training samples which are closest to the prediction samples. And DELM after the multilayer study of the characteristics expression can greatly improve the prediction accuracy of the forecasting model. Because DELM could extract deep information of the time series, breaking the limitations that the ordinary ELM model is difficult to capture the high correlation characteristics due to the single hidden layer structure.

REFERENCES

[1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time series analysis: forecasting and control, Journal of the American Statistical Association, Vol.68, No.342, 343-344, 2016.

[2] E. Alpaydin, Introduction to machine learning(Adaptive Computation and Machine Learning), The MIT press, Vol.5, No.8, 28, 2004.

[3] R. Adhikari, R.K. Agrawal, A homogeneous ensemble of artificial neural networks for time series forecasting, International Journal of Computer Applications, Vol.32, No.8, 1-8, 2013.

[4] L.P. Maguire, B. Roche, T.M. McGinnity, Predicting a chaotic time series using a fuzzy neural network, Information Sciences, Vol.112, No.1, 125-136, 1998.

[5] H. Kantz, T.Schreiber, Nonlinear time series analysis, Cambridge university press, Vol.1, No.4, 491, 2004.

[6] Z. Cai, R.C. Tiwari, Application of a local linear autoregressive model to BOD time series, Environmentrics, Vol.11, No.3, 341-350, 2000.

[7] Z. Huang, M.L. Shyu, Long-term Time Series Prediction using k-NN based LS-SVM Framework with Multi-value Integration, Springer Vienna, 191-209, 2012.

[8] G. Huang, Q. Zhu, C. Siew, Extreme learning machine: Theory and applications, Neurocomputing, Vol.70, No.1-3, 489-501, 2006.

[9] G. Feng, G. Huang, Q. Lin, R. Gay, Error minimized extreme learning machine with growth of hidden nodes and incremental learning, IEEE Transactions on Neural Networks, Vol.20, No.8, 1352-1357, 2009.

[10] G. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Transaction on Systems, Man,and Cybernetics-Part B: Cybernetics, Vol.42, No.2, 513-529, 2012.

[11] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the marquardt algorithm, IEEE Transactions on Neural Networks, Vol.5, No.6, 989-993, 1994.

[12] J.S.R. Jang, C.T. Sun, E. Mizutani, Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence, Prentice-Hall, Inc, 1996.

[13] T. Van Gestel, J. De Brabanter, B.De Moor, Least squares support vector machines, Singapore: World Scientific, 2002.