A Combination Forecasting Model of Extreme Learning Machine Based on Genetic Algorithm Optimization

Zhiheng Yu
Academy of Fine Arts
Northeast Normal University
Changchun, China
E-mail: yuzh496@nenu.edu.cn

Chengli Zhao
School of Management Science and Information Engineering
Jilin University of Finance and Economics
Changchun, China
E-mail: zhaocl069@163.com

***Abstract*—After studying the working principle of feed-forward neural network and analyzing network structure and the learning mechanism of BP neural network and the extreme learning machine (ELM), a prediction model, GA-ELM, is proposed based on genetic algorithm to optimize the learning machine limit. The genetic algorithm is used to select the weights and thresholds of ELM neural network, and the optimal weights and thresholds are used to determine the connection weights between the hidden layer and the output layer. Further, this model is combined with the grey system model to correct the residual of GM, and then GM-GA-ELM combination forecasting model is established. Compared with BP model, GA-BP model and standard ELM model, it is further verified that the predicting accuracy and running time of the proposed model are better.**

Keywords- ELM; GA; GM(1,1); combined forecast

I. Introduction

According to the basic principle of biological neural network, artificial neural networks train and update the weights without considering the internal structure of the mathematical model. They depend on the initial conditions set and a large number of training samples artificially in order to obtain the output mapping the input of the arbitrary nonlinear relationship. The model was originated in 1943 by Pitts and McCulloch, and it is still keen on the research in this field. Kelei[1] proposed a model to introduce a new prototype learning layer into the network. The prototype learning layer used a prototype based metric method to transform instance features into bag features. The network therefore can use label information of bag and learning the whole model in a compact process. Lei [2] focused on the topology and some learning algorithms of infinite deep neural networks and discuss some successful applications in speech recognition and image understanding. In order to reduce the time consumption of running a program, Xiangjuan [3] proposed a method of test data generation for path coverage based on neural networks. The experimental results showed that the method can effectively reduce the time consumption of running a program, therefore improve the efficiency of test data generation.

Among the many neural network models, BP neural network is the most widely used. A novel method of line loss rate calculation in transformer district was presented by Ya [4] and realized by programming, which was combined improved K-Means clustering algorithm with BP neural network model optimized by Levenberg-Marquardt(LM) algorithm. BP neural network and Markov chain were combined by Tong [5] to analyze and forecast a large wastewater treatment plant (WWTP) nitrogen removal efficiency. The results showed that the BP neural network can be used to simulate the WWTP nitrogen removal process, and the accuracy and reliability of the simulation can be improved.

However, the BP network consumes a lot of time in training the process of adjusting various parameters. Different from the traditional feedforward neural network, proposed by Professor Guangbin in Nanyang Technology University, extreme learning machine (ELM) [6] gives randomly initial weights and thresholds to neurons. It calculates the weights between the hidden layer and output layer by inversing the hidden layer and multiplying training output. When the network training process is completed, it no longer goes to the parameter adjustment in the network. Guangbin proves that random node parameters of neural network hidden layer greatly shorten the network training time, but does not affect the convergence ability. Furthermore, neural network constructed by this method can approximate any continuous system. Compared with the traditional feed forward neural network, the training speed of ELM is improved by hundreds or even thousands of times, which stimulates the extensive application of neural network in the predictive control problem. Liang et al [7] proposed the online sequential extreme learning machine based on traditional ELM, which is for the new training data to learn. After the end of the study, the used data will become history data and are discarded for reducing the learning time. The study proposed by Zhena [8] focused on the analysis of the behavior of consumers online travel website and analyzed the influential factors of customer intention as well as the influence processes on the basic of ELM theory. Aiming at the problems of current license plate recognition algorithm, such as slow training speed, low recognition rate, the vehicle license plate recognition algorithm based on Self-adaptive Evolutionary Extreme Learning Machine (SaE-ELM) was studied by Wenwu [9]. Experimental results demonstrate that proposed algorithm has the advantages of high identification rate, fast speed, and good robustness to the complex traffic environment.

978-1-5386-3886-6/17 $31.00 © 2017 IEEE
DOI 10.1109/CIIS.2017.14

In order to improve the learning ability and prediction ability of ELM, the swarm intelligence algorithms have been applied to optimize the weights and thresholds between the input and hidden layers. The genetic algorithm (GA) can be used in the selection process of the weight and the threshold between input layer and hidden layer. It can obtain more effective weights and thresholds than random assignment, so as to improve the effectiveness of the ELM node of hidden layer and enhance the ability to respond to unknown data. Therefore, this paper presents an extreme learning machine based on GA optimization algorithm.

II. GA-ELM MODEL

A. ELM Model

ELM is the improved algorithm based on the single layer feedforward neural network (SLFN). It randomly generated connection weights and thresholds between neurons in hidden layer. In the training process, the values of these parameters need not re-adjust. Only setting the number of hidden layer neurons in advance, we can obtain the optimal solution output.

The typical single hidden layer feedforward neural network is composed of input layer, hidden layer and output layer. Among them, there are n neurons in input layer, corresponding to the input dimension of the sample. There are k neurons in hidden layer and m neurons in output layer. The connection weights and thresholds between input layer and hidden layer are ω and b respectively and the connection weights between hidden layer and output layer are β.

In single layer feedforward neural network model, giving S different training samples$(x_i,t_i)\in R^N\times R^M(i=1,2,\dots,S)$, the input and output matrices of the samples are expressed as X and T. Supposing the activation function of neurons in the hidden layer is φ(.),the mathematical formula model of Network output Y is shown in (1).

$$Y=[y_1,\dots,y_j,\dots,y_S],\ y_j=\begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{Mj} \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{K}\beta_{i1}\phi(\omega_i,b_i,x_j) \\ \sum_{i=1}^{K}\beta_{i2}\phi(\omega_i,b_i,x_j) \\ \vdots \\ \sum_{i=1}^{K}\beta_{iM}\phi(\omega_i,b_i,x_j) \end{bmatrix},\ j=1,2,\dots,S \tag{1}$$

If the output matrix Φ of hidden layer is expressed as a concrete expression, as shown in formula (2), the learning error of the network is represented by O, and ELM can be expressed as (3).

$$\Phi=\begin{bmatrix} \phi(\omega_1,b_1,x_1) & \phi(\omega_2,b_2,x_1) & \cdots & ..,b_K,x_1) \\ \phi(\omega_1,b_1,x_2) & \phi(\omega_2,b_2,x_2) & \cdots & ..,b_K,x_2) \\ \vdots & \vdots & & \vdots \\ \phi(\omega_1,b_1,x_S) & \phi(\omega_2,b_2,x_S) & \cdots & ..,b_K,x_S) \end{bmatrix} \tag{2}$$

$$O=\left\|\Phi\beta-T\right\| \tag{3}$$

According to two theorems proposed by Guangbin [10], when the function φ(.) is infinitely differentiable, ω and b in training can be selected randomly. They are not changed during the training process, and β can be obtained through the least squares solution of equations. The solution is shown in formula (4), in which Φ+ is the generalized inverse of Φ.

$$\hat{\beta}=\Phi^{+}T \tag{4}$$

B. GA-ELM Model

The workflow of GA-ELM based on genetic algorithm optimization is shown in Fig. 1. GA-ELM can be divided into three parts: the determination of ELM network structure, GA optimization and ELM training and prediction [11].

The dimensions N of input layer, the dimensions M of output layer should be determined in the stage of ELM network structure. According to the actual needs of application, the number of nodes in hidden layer of K is determined. At the same time, the activation function of the hidden layer needs to be determined.

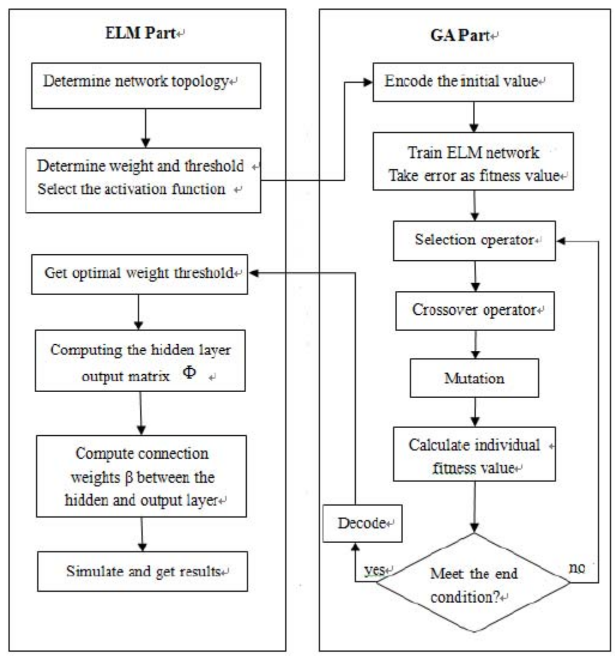

Figure 1. workflow of GA-ELM

According to the network structure of ELM, the length of individual L is determined by GA. Each individual needs to contain all the connection weights and thresholds between the input layer and the hidden layer of the network, so L=K*N+K. Secondly, the initial population and the number of iterations are determined. Population size is usually 20~100. Then, the error of each individual is calculated according to the training sample, which is the basis of individual fitness. The new population and the new fitness value are calculated again and again according to the selection operator, crossover operator and mutation operator to get find the optimal individual. After the end of the optimization process, the best individual will be decoded and submitted to ELM.

In ELM network, the connection weights and thresholds between the input layer and the hidden layer are evaluated by using the optimal individual decoding value submitted by GA. Then the connection weights between the hidden layer and the output layer are calculated. In the simulation prediction, the weights and the thresholds are no longer changed in the network model. According to the input of the test sample, the output of the sample is predicted.

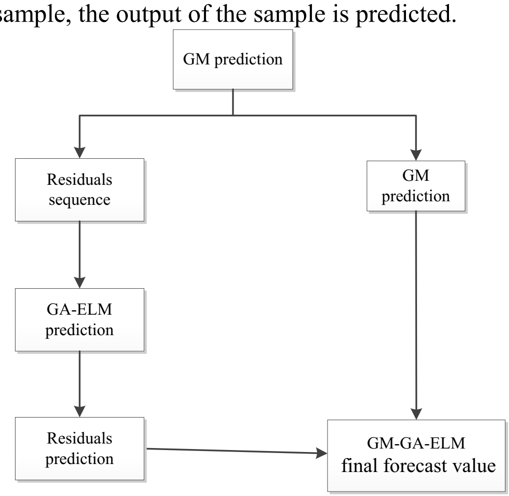

Figure 2. work flow of GM-GA-ELM.

C. GM-GA-ELM

It is called grey system (GM) that part of the information is known and part of the unknown information. Grey prediction is an important part of grey system theory, and the most representative one is GM(1,1) model. GM(1,1) requires less sample data, and the data does not need to meet a certain law. Therefore, it is simple and easy to calculate. But GM(1,1) has inherent bias. In order to improve the prediction accuracy of GM(1,1) and reduce the error, the residual error can be corrected further. In this paper, GA-ELM is used to modify GM(1,1), and GM-GA-ELM combination forecasting model is constructed as Fig. 2.

III. EXPERIMENTS AND ANALYSIS

A. Data Sources and Sample Representation

In experiment, the data is provided by California expressway traffic capacity measurement system (PeMS). They are from April 6 to April 24 in 2015, and the time interval is 1 hour and the total of data is 360. In this experiment, the sample set is generated by using the rolling mode.

B. Comparison of Experimental Results

In experiment, the structure of ELM network is 3-100-1 and the population size of GA is 50.The maximum iteration number is 100,the crossover probability is about 0.2 and the mutation probability is about 0.1. In addition to the relative error, the evaluation index increases the time cost.

BP, GA-BP, standard ELM and GA-ELM are comparative analysis. The experimental results are shown in Table 1, which is the average result of 20 runs of each model. Among them, the index is 1.0e+07.

TABLE I. PREDICTION RESUL OF EACH MODEL

Serial number	Original Data	BP	GA-BP	ELM	GA-ELM
1	0.498	0.465	0.467	0.464	0.438
2	0.364	0.376	0.365	0.381	0.362
3	0.330	0.319	0.326	0.320	0.318
4	0.368	0.422	0.407	0.412	0.386
5	0.652	0.623	0.637	0.645	0.650
6	1.316	1.317	1.328	1.281	1.340
7	1.985	1.985	1.986	1.996	1.970
8	2.447	2.399	2.414	2.393	2.381
9	2.345	2.384	2.410	2.383	2.440
10	2.172	2.129	2.109	2.063	2.166

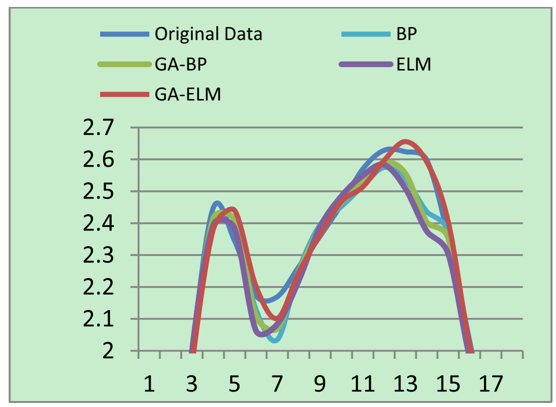

Figure 3. comparison of the original and predicted curves.

The prediction results of various models can better fit the original data. In order to compare the differences of the prediction results of the models, some of the data are shown in Fig. 3. It can be seen that all kinds of prediction models are very close to the original data in general, while GA-ELM can get closer to the original data.

In addition, the performance evaluation of various prediction models is shown in table 2. As can be seen in the prediction accuracy, GA-ELM is the best and ELM is the worst. From the running time, ELM is the best, followed by BP, GA-ELM is worse and GA-BP is the worst. The optimization of GA takes more time, which makes the standard model and GA optimization model run time difference of two orders of magnitude, but in the prediction accuracy has been improved. Further, compared with the optimization of the BP model, GA is more suitable to optimize the ELM model.

TABLE II. PERFORMANCE COMPARISON OF FOUR MODELS

Model	error	Running time
BP	2.873%	0.44831
GA-BP	2.397%	**41.77756**
ELM	**3.111%**	**0.034504**
GA-ELM	**1.985%**	2.685754

The experiment is carried out by using the monthly traffic flow of Yuhang expressway, and the prediction error is shown in table 3. It can be seen that the prediction accuracy corrected by ELM is obviously higher than that of pure GM, and the prediction accuracy corrected by GA-ELM is highest.

TABLE III. PERFORMANCE COMPARISON OF THREE MODELS

Model	error
GM	10.28%
GM-ELM	6.49%
GM-GA-ELM	**5.89%**

IV. CONCLUSIONS

In this paper, after studying the working principle and the workflow of the genetic algorithm, GA-ELM based on genetic algorithm is proposed. The genetic algorithm is used to improve the performance parameters of the learning machine. The validity of the model is verified by comparing with other models. In addition, the model is used to modify the GM model to improve the prediction accuracy of GM.

ACKNOWLEDGMENT

This paper is sponsored by "Twelfth-Five-Year Plan" Science and technology research projects of Jilin Province Department of Education under Grant No.2015412; "Twelfth-Five-Year" research projects of Jilin Province Education Science under Grant No. GH14219 and No. GH150287. We would like to thank the organizations for their supports.

REFERENCES

[1] HE Ke-Lei, SHI Ying-Huan, GAO Yang etc.A Prototype Learning based Multi-Instance Convolutional Neural Network, CHINESE JOURNAL OF COMPUTERS,vol.39(2016) pp.1-12.

[2] Zhang Lei and Zhang Yi. Big Data Analysis by Infinite Deep Neural Networks. Journal of computer research and development.vol.53(1):68-79,2016.

[3] YAO Xiang-Juan, GONG Dun-Wei, LI Bin. Evolutional Test Data Generation for Path Coverage by Integrating Neural Network, Journal of Software,2016,27(4):828-838.

[4] LI Ya, LIU Liping. Calculation of Line Loss Rate in Transformer District Based on Improved K-Means Clustering Algorithm and BP Neural Network, Proceedings of the CSEE, Vol.36 No.17 Sep. 5, 2016 ,4543-4551.

[5] LI Tong ,LI Jun.The prediction of denitrification efficiency of a wastewater treatment plant by using BP neural network and Markov chain method. Acta Scientiae Circumstantiae, Vol. 36,No.2,Feb,2016,576-581.

[6] Huang G B, Zhu Q Y,Siew C K.Extreme learning machine:a new learning scheme of feedforward neural networks[C]. In: IEEE International Joint Conference on Neural Networks(IJCNN 2004). Oakland: IEEE, 2004:985-990.

[7] Liang N Y, Huang G B, Saratchandran P, et al. A fast and accurate online sequential learning algorithm for feedforward networks[J]. IEEE Transactions on Neural Networks, 2006,17(6):1411--1423.

[8] Kanjana.A Study of Consumer behavior Using Online travel website based on ELM model,Chinese Journal of Management.2016.6,vol.13.no.6,889-898.

[9] Wen Wu, Qiao Longhui, He Peng. License plate recognition based on self-adaptive evolutionary extreme learning machine. Application of Electronic Technique, 2017, 43 (1):133- 136.

[10] Huang G B, Zhu Q Y, Siew C K. Extreme Learning Machine: Theory and Applications[J]. Neurocomputing, 2006, 70(1-3):489-501.

[11] Zhiheng Yu. Research on Combinational Forecast Models for the Traffic Flow Based on Intelligent Theory, Northeast Normal University,Phd thesis,2016.