Application of GA-DA-Elman Neural Network Algorithm to Urban Air Quality Evaluation

Haiyan Liu, Xuan Geng

Information Engineering College, Shanghai Maritime University, 201306 Shanghai, China

*Corresponding author e-mail: 1528578459@qq.com

Abstract. In order to improve the accuracy of air quality assessment, this paper presents an Elman neural network model based on dragonfly genetic algorithm fusion optimization. This model used genetic algorithm to optimize the weight and threshold of Elman neural network, introduced dragonfly algorithm in the process of searching for the optimal chromosome set, and used the best dragonfly decoded to replace the worst chromosome, thus forming an improved GA-DA-Elman neural network algorithm. According to the data from the atmospheric environment monitoring of Shanghai, the air quality evaluation models of Elman neural network and GA-Elman neural network were established and compared with GA-DA-Elman neural network. The results show that the GA-DA-Elman air quality evaluation model has the highest accuracy (95%), and the error control stability is optimal.

1. Introduction

The air quality evaluation methods are mainly divided to statistical method and non-statistical method. The traditional air quality evaluation methods are mainly statistical methods including analytic hierarchy process (AHP), fuzzy comprehensive evaluation (FCE) and grey relational analysis method [1-3]; those methods are strongly subjective, and could not make comprehensive analysis by effectively using interconnectedness, hierarchy and index system between indexes. The non-statistical methods mainly contain support vector machine (SVM), RBF and BP artificial neural network [4-8], etc.. Chen Zuyun, et al. [5] adopted SVM to predict and classify air quality and improved the robustness and recognition capability improved to certain degree; the feasibility of the model is verified from the perspective of optimization. Yin Qi et al. [6] combined simulated annealing algorithm (SA) and particle swarm optimization (PSO) to optimize SVM so as to establish air quality evaluation model; they made breakthrough in precision. However, the regional meteorology and pollution source are main factors affecting precision of air quality evaluation, and of strong nonlinear characteristics. The model could not solve the problem favorably. Further research shows artificial neural network as an effective tool to solve nonlinear problems could favorably solve this problem, and has been extensively applied to air quality evaluation field. Liu Jie et al. [7] established air quality evaluation model with BP and RBF neural network; Liu Dujin [8] adopted bee colony algorithm to optimize air quality prediction model based on BP neural network, and perfectly solved nonlinear relationship between air quality data and quality grade.

Elman neural network is a type of typical feedback network, and a optimization model of BP neural network. It has favorable dynamic characteristics and great overall stability; it could not only be applied extensively to nonlinear problem treatment but also could dynamically process complex data. It has been verified in numerous fields [9-12]. Therefore, this paper plans to introduce Elman neural network to air quality evaluation, and establishes a GA-DA-Elman neural network model based on GA and DA.
meantime, Elman neural network learns the training sample so as to further reduce error of subjective evaluation, and make results of evaluation system more objective. It has favorable actual application value.

2. Main algorithms

2.1 Elman neural network

Elman neural network is comprised of input layer, hidden layer, receiving layer and output layer, as shown in Fig.1. The hidden layer receives node input of input layer and node feedback input of the receiving layer; the receiving layer as the feedback of the hidden layer remembers and stores the output value of hidden layer neuron a moment before, and plays the role of delay operator; the output layer carries out linear or nonlinear weighted calculation of data, and outputs operation results.

\[
\begin{align*}
\text{Input layer} & \quad w_1 \quad w_2 \quad \ldots \quad w_r \\
\text{Hidden layer} & \quad x_1(k) \quad x_2(k) \quad \ldots \quad x_n(k) \\
\text{Receiving layer} & \quad x_c(1) \quad x_c(2) \quad \ldots \quad x_c(n) \\
\text{Output layer} & \quad y(k) = \frac{w^3 x(k) + b_2}{\ldots} \\
\end{align*}
\]

In Formula (1), (2) and (3), the input vector \(u = [u_1, u_2, \ldots, u_r]\), output vector \(y = [y_1, y_2, \ldots, y_m]\), output vector of the hidden layer \(x = [x_1, x_2, \ldots, x_n]\), output vector of receiving layer \(x_c = [x_{c1}, x_{c2}, \ldots, x_{cn}]\), \(w^1\), \(w^2\) are the connecting weight matrix from input layer to output layer, and from the hidden layer to the output layer; \(w^3\) is the weight matrix comprising hidden layer and receiving layer. \(f()\) and \(g()\) are respectively the activation function of hidden layer and output layer neurons.

2.2 Genetic optimization algorithm and dragonfly algorithm

The genetic algorithm (GA) encodes the object to be optimized according to the principles of natural choice and genetics, forms coded strings, and finishes iteration selection by screening individuals of the group through fitness function calculation and genetic operation (selection, cross, and mutation) so that the individual fitness is increasing gradually to the set value and new group suitable to the environment is generated thus.

Dragonfly algorithm (DA), proposed by Mirjalili, is a new intelligent optimization algorithm based on hunting behaviors of dragonfly groups. It takes the food as the optimal solution and migration destination by observing foraging and migrating behavior of dragonfly. During foraging, the individual behaviors of dragonfly could be summarized to separation, alignment, cohesion, food attraction and repulsion.

Compared to BP neural network, Elman neural network could adapt to time-variation characteristics, but it is learning slowly and easy to be trapped to local minimum. GA and DA both transmit the optimized weight and threshold value to Elman neural network; however, the problems such as early converge and low precision may be generated. Therefore, this paper integrates GA and DA and optimizes the weight and threshold value of the Elman neural network with the integrated algorithm to seek for the optimal parameters, so as to restrict local minimum and get the Elman neural network more approaching to
authentic rules of data.

3. Realization steps and flow of improved GA-DA-Elman neural network

This paper integrates GA and DA. Based on GA, DA is embedded in the same time to replace the worst chromosome in GA with the decoded optimal dragonfly. Specific realization steps of GA-DA-Elman algorithm can be seen in Fig.1:

![Diagram](image_url)

**Fig. 2** Dragonfly Genetic algorithm optimization neural network flow chart

Step 1: Initialize parameters setting (GA, DA), and establish Elman neural network Parameters setting: population size, iterations, mutation, crossover probability, and initial position; network structure, weight and threshold, etc.;

Step 2: Calculate fitness value of GA and DA;

Step 3: Carry out genetic operation, and seek for the optimal groups;

Step 4: DA is introduced to confirm position of food and predators, and seek for optimal individual and the worst individual of dragonflies; the optimal individual is taken as food and corresponding position will be saved, while the worst individual is taken as the predator;

Step 5: Calculate five behavior degrees of DA, i.e., predator factor, food factor, cohesion, alignment and separation;

Step 6: Update position of dragonflies according to conclusion acquired in aforesaid steps;

Step 7: Judge whether it is far from inferior solution; if yes, apply secondary dragonfly repulsion and seek for the optimal solution; if no, carry out dragonfly decoding, calculate target function value; the objective function value is error rate;

Step 8: Input the weights and thresholds acquired by optimization through mixed algorithm into neural network for training and testing;

Step 9: Replace the optimal dragonfly after decoding with the worst chromosome decoded through GA so as to form new optimal solution;

Step 10: Judge whether the algorithm has finished iteration; if no, apply Step 2.
4. Experimental process and results analysis

4.1 Experiment data processing

The data in this paper is the air quality data (SO₂, NO₂, CO, PM₁₀, PM₂.₅, O₃) and meteorological data (temperature, humidity, wind scale) of Shanghai from January to July 2019 from Shanghai air quality online monitoring and analysis platform (https://www.aqistudy.cn). Referring to the Ambient Air Quality Standard (GB 3905 - 2012) and Technical Regulation on Ambient Air Quality Index (on trial) (HJ 633 - 2012), the air quality grade is divided into 6 grades according to the index value of air quality: Grade I (excellent, value: 0-50); Grade II (good, value: 51-100); Grade III (light pollution, value: 101-150); Grade IV (moderate pollution, value: 151-200); Grade V (severe pollution, value: 201-300); Grade VI (gross pollution, value >300).

4.2 Model building

The atmospheric environment evaluation model is mainly built by three parts: modeling, training and prediction of Elman neural network; The neural topology in this paper shall be 9-12-1. The number of neuron nodes on neural network output layer depends on number of meteorological factors. The evaluation factors in this paper are nine, respectively SO₂, NO₂, CO, PM₁₀, PM₂.₅, O₃, temperature, humidity and wind scale; thus, the number of neurons on the input layer is 9; the nodes of the hidden layer shall be confirmed according to the empirical formula. The empirical formula is as follows:

\[ p = \sqrt{N_{in} + N_{out}} + \alpha \]  

In the formula, \( p \) is the number of nodes on the hidden layer; \( N_{in} \) is the number of nodes on the input layer; \( N_{out} \) is the number of nodes on the output layer; \( \alpha \) is the empirical parameter, which shall be rounded in interval [1:10]. The number of nodes on the hidden layer is calculated according to formula (4), and then Elman debugging is carried out. Results show when the nodes on the middle layer are 12, the predicted error of the network is the smallest. Therefore, for the case in this paper, the number of nodes on the hidden layer shall be 12. The output layer represents environment grade, and the number of node is 1.

For network training, certain data is selected from data set to train network continuously so as to reach iteration. This paper totally uses 680 groups of test training sample data, in which 200 groups are for three grades each, and respectively 50 and 30 groups are for the other two grades. Thus, according to the proportion of 4:5, 150, 150, 150, 40 and 24 groups are selected respectively as network training data; moreover, 100 groups are selected from the remaining 164 groups as test sample data set.

Parameters setting: The parameters of GA are respectively: population size: 200, iterations: 100, crossover probability: 0.1, and mutation probability: 0.8; parameters of DA are respectively: population size: 20, and maximum iterations: 100; the node function of hidden layer is tansig, output node function is purelin, training function is trainlm, the learning efficiency of neural network \( \eta \) is 0.1, minimum error of training objective is 0.00001 and maximum training time is 100.

4.3 Results analysis

The five atmospheric environment classifications respectively correspond to the five grades of air environment quality. This paper adopts two algorithms (GA and GA-DA) to optimize the neural network. The objective function is the error rate (average error between predicted output value of Elman neural network and expected output value). The smaller the average error is, the higher the system precision will be.

Firstly, MATLAB is used to simulate air quality classification with standard Elman neural network and GA-Elman. The model effect can be seen obviously and results can be seen in Fig.3.
According to Fig.3, Elman model has error in Grade I, II and III during air classification process, and error frequency is respectively 3, 2 and 5, in which, the proportion of errors to total samples in the grade is respectively 15%, 5% and 16.7%, and average error rate is 10%. The precision of predication is low. Compared to Elman neural network, GA-Elman has improved classification precision, and eliminated error when air quality is Grade I. The average error rate is 7%.

In the end, MATLAB is used to test Elman neural network improved by GA-DA mixed algorithm. The convergence curve and predicted results of the model can be seen in Fig.4.

For environmental quality classification with aforesaid three atmospheric environment evaluation methods, the environment grade evaluation precision with GA-DA-Elman, GA-Elman and standard Elman algorithms are respectively 95%, 93% and 90%. The precision of GA-DA-Elman algorithm is the highest, and the error stability is favorable.

5. Conclusion

For air quality evaluation, this paper combines DA and GA, and embeds DA to GA to establish a GA-DA-Elman neural network air quality evaluation model. It compares the new model to the standard Elman and GA-Elman Neural network air quality evaluation model, and results show: GA-DA algorithm has better local and overall optimization ability than single GA algorithm, and could more effectively solve the local minimum of Elman neural network; the prediction precision of GA-DA-Elman is higher than that of GA-Elman and standard Elman neural network, and the error control performance is the best.
References

[1] Lv W, Ji S. Atmospheric environmental quality assessment method based on analytic hierarchy process [J]. Discrete & Continuous Dynamical Systems-S, 2019, 12(4&5): 941-955

[2] Olvera-García M Á, Carbajal-Hernández J , Sánchez-Fernández L P, et al. Air quality assessment using a weighted Fuzzy Inference System[J]. Ecological informatics, 2016, 33: 57-74

[3] Dong Shengwei, Su Ting, Huang Lingling, et al. Fuzzy comprehensive evaluation and curve fitting prediction model of air quality [J]. Practice and understanding of mathematics, 2018, 48(06): 102-108

[4] Feng R, Zheng H, Gao H, et al. Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China[J]. Journal of Cleaner Production, 2019, 231:1005-1015.

[5] Chen Zuyun, Jin Bo, Wu Changfu. Application of Support Vector Machine in Environmental Air Quality Assessment[J]. Environmental Science and Technology, 2012,35(S1):395-398

[6] Yin Qi, Hu Hongping, Bai Yanping, et al. Air Quality Assessment of Taiyuan City Based on Improved SAPSO Optimization Support Vector Machine [J]. Practice and understanding of mathematics, 2017, 47(21):215-222

[7] Liu Jie, Yang Peng, Lv Wensheng, et al. BP and RBF artificial neural network modeling and classification evaluation of urban air quality [J]. Safety and environmental engineering, 2014, 21(06): 129-134

[8] Liu Dujin, Pu GuoLin, Wang Guangqiong. Prediction of Ambient Air Quality Based on Bee Colony Optimization Neural Network [J]. Computer and Digital Engineering, 2018, 46(04):639-643

[9] Cheng Guojian, Ma Wei, Liu Ye, et ai. Intelligent calculation of sandstone reservoir physical property parameters based on hybrid dimensionality reduction Elman neural network [J]. Science Technology and Engineering, 2014, 14(03): 24-28

[10] Song Pinggang, Zhou Jun. Rolling Bearing Fault Diagnosis Based on EMD-1(1/2) Dimensional Spectral Entropy [J]. Science Technology and Engineering, 2014, 14(02):149-153

[11] Jia W, Zhao D, Zheng Y, et al. A novel optimized GA–Elman neural network algorithm [J]. Neural Computing and Applications, 2019, 31(2): 449-459

[12] Sun Y, Wang L, Wang X, et al. Reliability Growth Prediction Method Based on GA-Elman Neural Network[C]/2018 12th International Conference on Reliability, Maintainability, and Safety (ICRMS). IEEE, 2018: 135-141

[13] Zhang Xiao, Qian Yuliang, Qiu Zheng, et al. Gas Turbine Fault Diagnosis Based on BP Neural Network Optimized by dragonfly Algorithm[J]. Thermal Power Engineering, 2019, 34(03):26-32