Nature Environment and Pollution Technology
An International Quarterly Scientific Journal
p-ISSN: 0972-6268
e-ISSN: 2395-3454
Vol. 18
No. 5 (Special Issue)
pp. 1567-1572
2019
Original Research Paper
Open Access

Atmospheric Quality Testing Based on Deep Learning

Wei Li*, Yan Li*, Lei Wang* and Haroon Rashid**
*Hebei Institute of Communications, Shijiazhuang, Hebei, 051430, China
**Department of Civil Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Punjab 64200, Pakistan

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 25-08-2019
Accepted: 11-10-2019

Key Words:
Deep learning
Dalian city
Neural network
Atmospheric quality
Genetic algorithm

ABSTRACT

The purpose of this research is to apply the deep learning algorithm to the research of atmospheric quality detection. In this study, firstly, based on genetic algorithm and artificial neural network, the training process of genetic algorithm is optimized and improved, and a new hybrid accelerated genetic algorithm is proposed. Then combined with these algorithms, a universal air quality evaluation method for a variety of air pollutants is proposed. Taking the atmospheric quality inspection report of Dalian city for two months as the research sample, the hybrid accelerated genetic algorithm combined with the artificial neural network algorithm are applied to the BP neural network, which is optimized and improved to predict the atmospheric quality of the next month, and the prediction results are compared with the actual situation. The results show that the new algorithm is excellent in convergence speed and prediction accuracy and has certain value and prospect in the practical application of atmospheric quality prediction in the future.

INTRODUCTION

The quality of atmospheric environment is closely related to people's health and life. With the rapid development of human society, the ecological environment is also deteriorating. The serious deterioration of air quality and the threat of air pollution to human life also directly affect the sustainable development of human society (Flórez et al. 2018). China, as a developing superpower, is also faced with various tests from environmental problems. How to accurately evaluate the air quality and ensure the quality of life has become an important environmental issue concerned by people. The severity of air pollution is assessed by the air pollution hazard index. The air quality evaluation model is an application model that evaluates the air pollution hazard index and determines the level of air pollution hazard by measuring the relative concentration of various pollutants in the atmosphere (Jiang et al. 2017). The perspective of current environmental quality assessment is multi-directional, and up to now, China has not formed a unified series of assessment methods. In recent years, scholars at home and abroad have discussed the comprehensive evaluation and prediction of atmospheric quality detection and put forward some models of atmospheric evaluation, such as the air pollution index evaluation method, air pollution comprehensive index method, air pollution economics evaluation method, air pollution biological evaluation method, fuzzy comprehensive evaluation method, expert evaluation method, analytic hierarchy process, etc. (Sharma & Rani 2017, Mousavi et al. 2018). However, most of the evaluation methods fail to reflect the multi - factor comprehensiveness of the environmental quality structure and the scientific comprehensiveness of the environmental quality standard, so the standard level of environmental quality can't be quantitatively determined. These methods presuppose the model or subjectively specify some parameters. For example, the environmental quality index evaluation method involves too many subjective factors in the evaluation process, losing the accurate reflection of the objectivity of environmental quality structure, and the evaluation conclusion is abstract.

With the establishment of new science and the development of computer technology, a variety of new methods for evaluating atmospheric quality have been proposed, such as neural network method, matter-element extension method, set pair analysis method, genetic optimization method, projection pursuit analysis method, etc. In this study, genetic algorithm and neural network method were combined to optimize the application in atmospheric environmental quality assessment.

PAST RESEARCH

Genetic algorithms (GA) is a method of global optimization to find the optimal solution, and its idea is based on the theory of evolution and heredity. In essence, it is a probabilistic global search algorithm that combines natural selection of survival of the fittest, evolution mechanism of survival of the fittest and random information exchange mechanism among individuals in the same group to solve complex problems and seek the optimal solution (Tepecik & Navruz 2018). GA is implemented by turning the solution of the problem to be solved into chromosomes represented by binary code strings, and then all possible solutions of the problem constitute the chromosome group. GA places them in the environment of problems and evaluates, replicates, crosses and mutates the adaptability of chromosomes according to the survival of the fittest principle. After the generation of the initial population, GA conducts crossover and mutation operations according to the survival of the fittest principle, resulting in a new population as the solution to the problem. Each population is generated by the process of selection, crossover and mutation (Azamathulla et al. 2018). Finally, it converges to an individual that is most suitable for the problem environment, which is also called the optimal individual. And the optimal solution is obtained by decoding the best individual.

The entire working mode and inspiration of Artificial Neural Network (ANN) is derived from the mode of simulating the operation of human thinking (Petrozziello et al. 2017). It is equivalent to a nonlinear dynamic system. Neural networks acquire knowledge or experience mainly through learning, which can be generally divided into two stages: training and prediction (Sarajcev et al. 2018). The so-called training is to form a functional mapping relationship between the condition and the result, that is, to give a target error value of the actual output and the expected output. If the result obtained by the output layer is greater than the preset error target value, the neural network will return the error signal layer by layer along the original transmission route and adjust the weight value of the neuron connection between each level. This process will continue alternately until the error reaches the target value, and then the training process is end. Prediction is to test the reliability of the discriminant function. Some samples not included in the training set are used to form the prediction set, and then the prediction set is put into the trained network. Under the action of the discriminant function obtained in the training stage, a test result can be obtained, and the prediction can be made from the output end of the network. The main research direction of artificial neural network is quite extensive, which reflects the characteristics of the interdisciplinary technology field. At present, the main research work focuses on the biological prototype, the establishment of theoretical model, network model and algorithm research, artificial neural network application system, pattern recognition and image processing, control and optimization, financial prediction, management, and communication adaptive balance.

MATERIALS AND METHODS

Research object: The monitoring data of August 1 to September 27 in Dalian city were selected as samples, and the concentrations of SO_2, NO_2 and PM_{10} were taken as the main monitoring objects. The data samples in this research were provided by Dalian Environmental Testing Centre.

Genetic algorithm: The way in which the algorithm obtains the optimal solution is to set the range of the solution of the parameter to be optimized to a two - dimensional space. The whole space is regarded as composed of an infinite number of spatial solutions, the number of which is determined by the precise range of the parameters to be optimized, and each spatial solution is also called a spatial individual. A certain number of spatial individuals are randomly selected from the two-dimensional solution space, and then this group of individuals is called the initial parent group of individuals. The algorithm flow of GA is as follows:

Initialization- Set the counter of the evolutionary generation number, set the number of the maximum evolutionary generation, randomly generated individuals were used as the initial parent group and denoted as P (0);

Calculate individual fitness- Calculate the individual fitness of each individual in the parent group; selection operation-perform selection operation on the parent group, select the i^{th} individual from the initial parent, so that the more suitable individual with greater fitness is selected. Two groups of 500 individuals are selected, and the selected probability values are based on the fitness P_i of each individual.

Crossover operation- From the two groups of newly obtained individuals, an individual is extracted to pair and reconstitute two groups of 500 individuals with each new generation.

Mutation operation- Take any of a group of progeny individuals previously obtained and flip some part of the binary code of the individual according to the mutation rate of the population, that is, the original is 1 becomes 0, while the original is 0 turns into 1. Then the resulting new 500 child generations are used as the new parent of the next cycle;

Termination condition judgment- The fitness of the new parent generation obtained by calculation will be stopped if the accuracy is satisfied; if not, the cycle will be continued in the previous step, and when t=T, the individual with the greatest fitness obtained during the evolution process is used as the optimal solution output, and the calculation is terminated. The operation flow is shown in Fig. 1.

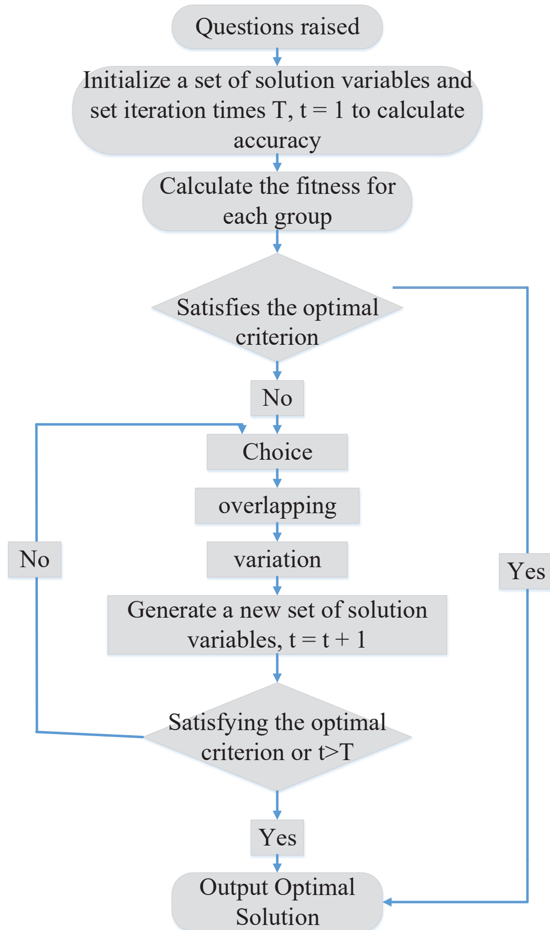

Fig. 1: Genetic algorithm flow chart.

Improved hybrid acceleration genetic algorithm (IHAGA): GA has disadvantages such as poor local search ability and precocity, among which the representative problem is how to balance and coordinate the rapid convergence of algorithm and maintain individual diversity. In this study, the genetic algorithm was improved and the local search algorithm was introduced, which could well solve the global optimization of parameters in the formula model of air quality pollution index. In the selection operation, the selection of elite individual migrators enhances the search ability of the algorithm for continuous global optimization and ensures the superiority of species and the convergence speed of the algorithm. Selecting the worst individual for direct variation ensures the individual diversity of the group, strengthens the global search ability of the algorithm, and well avoid the phenomenon of premature algorithm. The segmentation point crossover operator is adopted, so that individual elements can exchange information more uniformly and more scientifically. The introduction of local search operator greatly improves the local search ability of the whole genetic algorithm. In addition, acceleration operator is adopted to avoid unnecessary iteration and improve the operation efficiency of the algorithm, as shown in Fig. 2.

Neural network algorithm and deep learning: Artificial neural networks use physical devices to simulate some structures and functions of biological neural networks. BP neural network model belongs to multi-layer feed-forward mapping network, and it is a supervised learning algorithm. The network learns and trains the data set through algorithm and determines the connection weight and threshold between each neuron, so as to establish a mapping relationship between the subset of dimensional space and the subset of dimensional space. Each time the learning algorithm adjusts the weights, it includes two processes: signal forward propagation and error response propagation. First, the sample input signal is transmitted from the input layer to the output

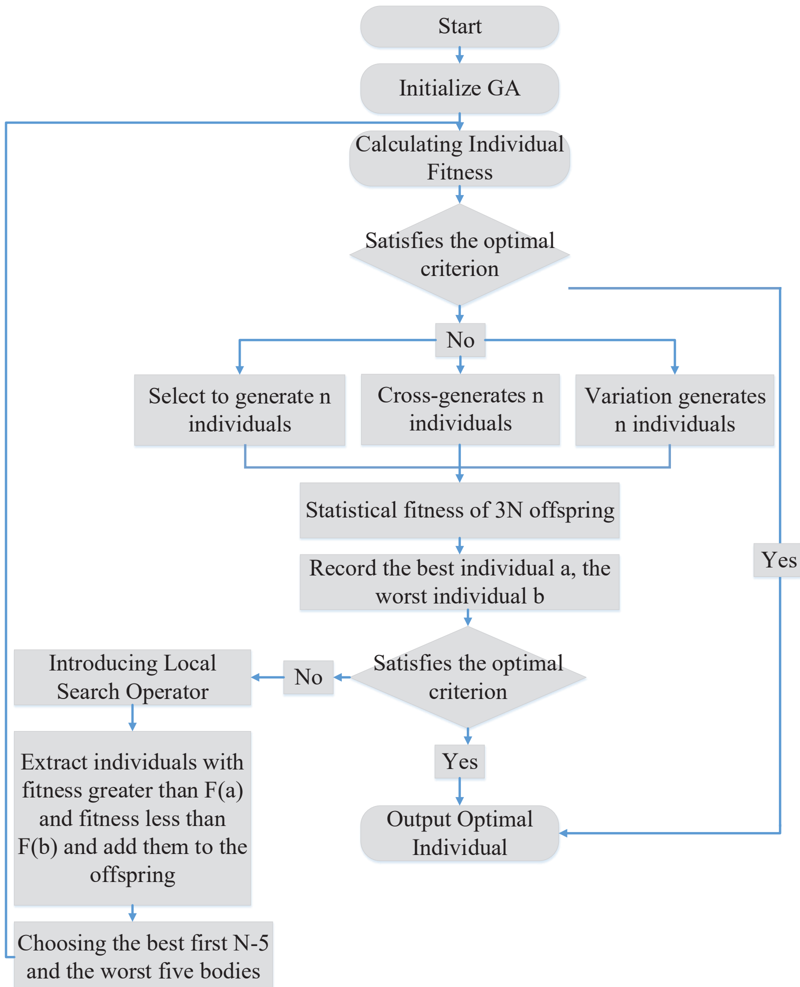

Fig. 2: Improved hybrid accelerated genetic algorithm flow chart.

layer through the neurons in each hidden layer, and the global errors between the output of all sample input signals to the network and the expected sample output are calculated, and then the global error is tested. If the global error does not meet the requirements, it will be propagated back through the error, and the connection weight between neurons in each layer will be modified layer by layer starting from the output layer. After several forward signal propagation and error back propagation processes, the global error between the output value and the expected value of all sample signals obtained by the network finally meets the requirements. In the back - propagation algorithm, gradient method is usually used to correct the weight, so the output function is required to be differentiable, and Sigmoid function is usually used as the output function. In the output layer, it is assumed that y_k is the actual output value and y_{dk} is the ideal output value, the error under this sample is shown in formula 1:

$$E = \frac{1}{2} * \sum k(y_{dk} - y_k)^2 \qquad \ldots(1)$$

Then, the partial derivative is calculated to obtain the weight correction amount. It can be concluded that the learning algorithm of the network belongs to supervised learning, and it is necessary to provide the input samples and the corresponding ideal output at the same time.

Formula for calculating the sub-index of atmospheric pollution damage: With the change of atmospheric pollutant concentration from low to high, the degree of its harm to the environment is constantly changing. Therefore, the growth curve index formula can be used to describe various individual environmental indicators. And the pollution damage rate of the air pollutant to the air quality can be expressed as formula 2:

$$R_i = 1/(1 + a_i e^{-bifi}) \quad \ldots(2)$$

Where, a_i and b_i are undetermined parameters related to pollutant characteristics, which are parameters to be optimized. Another parameter c to be determined, which is independent of the characteristics of pollutants, is introduced, so formula 3 is obtained. Among them, x_i is the actual concentration of the corresponding pollutant. Therefore, formula 3 is a universal formula for calculating air pollution damage index applicable to a variety of air pollutants.

$$R_i = 1/(1 + ae^{-bxi})^c \quad \ldots(3)$$

With the difference of the value range of IHAGA parameter, the values of a, b and c change obviously, and the parameter a tends to the lower bound of the value range of this parameter infinitely. In view of the above characteristics, it indicates that the functional relationship between the air pollution damage rate and the relative concentration of pollutants can be basically determined by two parameters, that is, a and b.

RESULTS AND DISCUSSION

Monitoring results of pollutant concentration: The concentration values of the main pollutants SO_2, NO_2, and PM_{10} in the atmosphere in the 20 -day monitoring data are shown in Table 1.

Predicting outcomes: The neural network is used to predict it. The output nodes of the neural network are set as 3, representing the observed value of pollutant concentration within three days. The output node 4 is the predicted values of the pollutant concentration on the fourth day. Air pollution is forecast every two days in October (Table 2).The main pollutant SO_2 in the atmosphere is taken as the main monitoring object. First, the entire neural network is trained through the training sample. The maximum number of iterations is 1000, and the accuracy is 0.0000099791. The prediction samples of October are introduced into the improved trained neural network and traditional neural network to test their generalization ability. By comparing the results, it is concluded that the average error of the IHAGA-BP neural network is 0.000172, and the average value of the traditional BP neural network is 0.000374. In the same way, with NO_2 as the main monitoring object, the average error of the IHAGA-BP neural network is 0.000194, and that of the traditional BP neural network is 0.000412. In the same way, with PM_{10} is taken as the main monitoring object, the average error of the IHAGA-BP neural network is 0.000186, and the average error of the traditional BP neural network is 0.000507.

Table 1: Concentration values of atmospheric pollutants SO_2, NO_2, and PM_{10} for 20 days.

Date	SO_2 (mg / m^3)	NO_2 (mg / m^3)	PM_{10} (mg / m^3)
August 1	0.019	0.027	0.211
August 4	0.016	0.040	0.107
August 7	0.021	0.037	0.095
August 10	0.002	0.017	0.015
August 13	0.017	0.063	0.102
August 16	0.028	0.052	0.098
August 19	0.017	0.025	0.113
August 22	0.011	0.053	0.059
August 25	0.028	0.075	0.148
August 28	0.018	0.055	0.108
August 31	0.015	0.042	0.093
September 3	0.005	0.021	0.019
September 6	0.020	0.031	0.042
September 9	0.015	0.044	0.121
September 12	0.020	0.050	0.086
September 15	0.018	0.046	0.079
September 18	0.031	0.042	0.103
September 21	0.022	0.052	0.105
September 24	0.008	0.065	0.112
September 27	0.003	0.044	0.015

Table 2: Predicted concentration values of atmospheric pollutants SO_2, NO_2, and PM_{10}.

Date	SO_2 (mg/m^3)	NO_2 (mg/m^3)	PM_{10} (mg/m^3)
October 1	0.009	0.022	0.055
October 4	0.016	32	0.094
October 7	0.021	0.065	0.078
October 10	0.012	0.017	0.054
October 13	0.013	0.053	0.118
October 16	0.014	0.047	0.137
October 19	0.017	0.032	0.077
October 22	0.015	0.037	0.096
October 25	0.012	0.019	0.034
October 28	0.003	0.46	0.074
October 31	0.011	0.033	0.058

It can be concluded from the calculation results of the above examples that the pre-use of IHAGA-BP neural network for the prediction of major atmospheric pollutants in Dalian city is relatively reasonable in accuracy. It has excellent nonlinear approximation ability, small error and good overall fitting effect, and the comparison test proves that the improved neural network is better than the traditional neural network in the accuracy and generalization ability of prediction.

CONCLUSION

In this study, the traditional genetic algorithm in deep learning and the improved hybrid accelerated genetic algorithm were first introduced. Then, the network structure, learning rules, learning parameters, weights and thresholds of the neural network are optimized by the improved genetic algorithm. Finally, a new neural network algorithm is proposed through research, the model structure, specific implementation steps and implementation process of the algorithm are given, and it is verified by predicting the concentration of atmospheric pollutants in Dalian city.

REFERENCES

Azamathulla, H.M., Rathnayake, U. and Shatnawi, A. 2018. Gene expression programming and artificial neural network to estimate atmospheric temperature in Tabuk, Saudi Arabia. Appl. Water Sci., 8(6): 184.

Flórez, C.A.C., Rosário, J.M. and Amaya, D. 2018. Control structure for a car-like robot using artificial neural networks and genetic algorithms. Neural Comput. Appl., :1-14.

Jiang, S., Chin, K.S. and Wang, L. 2017. Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. Expert Syst. Appl., 82(C): 216-230.

Mousavi, M.S., Ashrafi, K. and Motlagh, M.S.P. 2018. Design of a correlated validated CFD and genetic algorithm model for optimized sensors placement for indoor air quality monitoring. J. Heat Mass Transfer, 54(2): 509-521.

Petrozziello, A., Cervone, G. and Franzese, P. 2017. Source reconstruction of atmospheric releases with limited meteorological observations using genetic algorithms. Appl. Artif. Intell., 31(2): 1-15.

Sarajcev, P., Jakus, D. and Vasilj, J. 2018. Application of genetic algorithm in designing high-voltage open-air substation lightning protection system. J. Electrostat., 93: 43-51.

Sharma, A. and Rani, R. 2017. An optimized framework for cancer classification using deep learning and genetic algorithm. J. Med. Imaging & Health Infor., 7(8): 1851-1856.

Tepecik, C. and Navruz, I. 2018. A novel hybrid model for inversion problem of atmospheric refractivity estimation. Int. J. Electron. C, 84: 258-264.