

Research on Air Quality Prediction Based on Neural Networks

Ruihao Wan

University of Southern California, Los Angeles, California, USA

Abstract: In view of the increasingly serious air pollution problem, to alleviate the harmful effects of air pollution on human body and society, this paper studies the prediction of air quality. Due to the nonlinear, regional and dispersive characteristics of pollutant data, the effective utilization rate of data is low and the prediction process is extremely complicated. How to effectively build a prediction model and improve the prediction accuracy of air quality is a hot issue in current research. This paper mainly introduces the current research status of air quality prediction.

Keywords: Air Quality Prediction; Graph Convolutional Neural Network; Intelligent Optimization Algorithm.

1. Introduction

With the continuous advancement of urbanization, automobile exhaust and industrial emissions are increasing, and the air pollution problem is becoming increasingly serious, which has a very serious impact on the sustainable development of the country and public health. From the individual level, air pollution has a very serious harm to the health of the public. Long-term exposure to polluted atmospheric environment will lead to dizziness, skin lesions and damage to normal cardiopulmonary function, etc., and more seriously, it will induce cancer. Children, the elderly and people with respiratory and cardiovascular problems are particularly affected. The study found that nearly 2 million childhood asthma cases were linked to nitrogen dioxide pollution, two-thirds of which occurred in cities [1], and that in 2019 86% of the global urban population had excessive PM_{2.5} exposure levels, which resulted in 1.8 million deaths [2]. The World Health Organization (WHO) released a report on October 11, 2021, stating that air pollution has become one of the "biggest environmental problems threatening human health", with 13 people dying every minute worldwide due to air pollution [3]. At the national level, air pollution will affect industrial and agricultural production, causing huge human, material and economic losses, and have a long-term impact on the country's economic development [4-5], as well as many other social problems caused by economic impact. Common sources of air pollution include natural pollution sources and man-made pollution sources, natural pollution sources are unavoidable, such as volcanic eruption, natural disasters, volcanic eruption will produce a lot of volcanic ash and harmful gases; Natural disasters include wind and dust storms. Man-made pollution sources are mainly produced along with economic development, including industrial and agricultural gas emissions and transportation gas emissions. In the past, the state ignored the damage to the environment for the sake of economic construction, and now the state has put forward a series of policies to strive for harmony between man and nature, so as to achieve the purpose of sustainable development and alleviate the adverse effects of air pollution on human beings. The state has put forward many policies, from the "Twelfth Five-Year Plan for National Environmental Protection" proposed by The State Council in 2011 to the "Guiding Opinions on Building a Modern environmental

governance System" proposed by the General Office of the CPC Central Committee and The General Office of the State Council in 2020, all of which show the importance of the State to air pollution control. At present, most parts of the country are facing severe air pollution problems, how to improve and control air pollution is still the focus of the national environmental protection department.

2. Current Status of Research on Air Quality Prediction

With the rapid development of neural networks, more and more researchers use neural network-based models in the field of air quality prediction. However, the hyperparameters in the prediction model, such as the number of network layers, the number of neurons in each layer, the learning rate and other parameters that affect the performance of the model, are often determined subjectively, making the prediction accuracy of the model less than the optimal state. As one of the most traditional neural networks, Back-propagation (BP) neural networks have disadvantages such as easy to stop training at the local optimum point and easy to affect the model performance by parameter values [6], which makes many scholars use various methods to improve them. In these studies, a lot of achievements have been made, such as using Bayesian normalization to improve the generalization ability of the network and improving the gradient descent method. In addition, some scholars also use heuristic algorithms to improve the prediction model of BP neural network. Zhou et al. [7] optimized BP neural network by combining Genetic Algorithm (GA) and Simulated Annealing (SA) algorithms, and the results show that the GA-SA based BP neural network has strong generalization ability and global search ability, and high accuracy rate. Huang [8] put forward a kind of based on improved particle swarm optimization algorithm (Particle Swarm Optimization, PSO) method of BP neural network to predict the AQI, make the prediction results more accurate. Other prediction models have also been improved. Fan Wenting [9] et al proposed the optimized Support Vector Machine (SVM) prediction model based on the improved Firefly Algorithm (FA). Gao Shuai et al. [10] proposed a method that combined SVM with the traditional Moth-Flame Optimization (MFO) algorithm. In addition, some scholars have proposed optimization algorithms to optimize LSTM.

Zhang et al. [11] optimized LSTM neural network parameters, and then applied them to air quality prediction of different cities, and achieved certain results. Al-janabi et Al. [12] used PSO to optimize LSTM's weight, deviation, number of hidden layers, number of nodes in each hidden layer and activation function parameters to predict the air pollutant concentration for the next two days.

From the above research, it can be found that the improved neural network based on intelligent optimization algorithm has been widely used and can get very good results. Therefore, it is particularly important to select a suitable optimization algorithm for the parameter optimization of neural network. Sawyer found must search algorithm (Beetle Antennae Search Algorithm, BAS) in time complexity and space complexity is lower than most of the swarm intelligence algorithm and higher efficiency. Manifold method et al. [13] compared BAS algorithm with many new and mainstream intelligent optimization algorithms and found that BAS algorithm has strong competitiveness. Compared with PSO algorithm, BAS algorithm has stronger ability to jump out of local extreme values and faster convergence speed. Compared with GA algorithm, BAS algorithm has the advantages of simple operation, less computation and faster running speed. Compared with FA algorithm, BAS algorithm is less affected by parameter sensitivity. Compared with bat algorithm (BA) and artificial bee colony (ABC), BAS algorithm has higher computational efficiency and lower computational complexity. However, BAS algorithm is also prone to fall into local extreme values and the optimization results are unstable. Therefore, this paper improves BAS algorithm to solve the above problems and applies it to optimize the air quality prediction model. Air quality prediction refers to the analysis of air quality time series data and other factors affecting air quality, such as meteorological data, factory waste emission data and vehicle flow data, and the establishment of a prediction model to predict and estimate the change trend of air quality and air quality index in the future period. Through the study of air quality prediction, the public can be provided with the latest air quality status, so that they can make corresponding protection measures according to their actual conditions, and it can also provide intuitive reference basis for the ecological environment department to formulate various environmental improvement plans to guide the prevention and control of air pollution [14]. Deep learning is developed based on artificial neural networks and belongs to a subset of machine learning. The biggest advantage of deep learning technology compared to traditional methods is the learning and feature extraction capabilities of deep learning models themselves. Based on deep learning methods, manual work in the feature extraction stage can be reduced to a certain extent, and the prediction accuracy of the model has been improved compared to traditional methods. This is why this article proposes to study air quality prediction based on deep learning. Deep neural networks have demonstrated their analytical capabilities in sequence data, but there is still limited research on deep learning models for spatiotemporal feature extraction. In order to design better spatiotemporal sequence prediction models based on deep learning technology, several mainstream deep learning models for air quality prediction will be introduced below. Due to the nonlinear, regional and dispersive characteristics of pollutant data, the effective utilization rate of data is low and the prediction process is extremely complicated. How to effectively build a prediction model and improve the

prediction accuracy of air quality is a hot issue in current research. Based on deep learning, this paper builds a spatial-temporal feature extraction air quality prediction model, dynamically analyzes the spatial relationship between monitoring stations, and learns the internal change rule of historical air quality data based on the deep learning model to achieve deeper air quality prediction. As an important means of air pollution prevention and control, air quality prediction plays an indispensable role for the country to take appropriate prevention and control measures. as is well known, substances in the air are constantly moving, so the concentration of pollutants monitored by air quality monitoring stations is also constantly changing and easily influenced by the surrounding environment, which leads to a certain spatial correlation between different stations. When the trend of most data changes between different sites is the same or opposite and the fluctuation amplitude is similar, it is considered that there is spatial correlation between sites.

During the model training process, combining site data with spatial correlation for joint training can learn effective spatial features that exist within them, which is more conducive to improving prediction accuracy. On the contrary, if site data that is not related to the target site is included, it will affect the model's learning ability and reduce learning accuracy. Choosing relevant sites has a significant impact on the prediction accuracy of the model. In previous studies, although researchers considered the impact of selecting relevant sites on later prediction, they only conducted static correlation analysis on different sites using Pearson correlation coefficients based on AQI time series data between sites, without considering that the correlation between sites in space changes dynamically over time. Therefore, using static correlation analysis methods still results in some noise in the training data, which affects the prediction accuracy of the model.

3. Current Research Status of Air Quality Prediction based on Optimized Neural Networks

With the rapid development of neural networks, more and more researchers are using neural network-based models in the field of air quality prediction. However, hyperparameters in prediction models, such as the number of network layers, the number of neurons in each layer, learning rate, and other parameters that affect model performance, are often subjectively determined by humans, resulting in suboptimal prediction accuracy. Backpropagation neural network is one of the most traditional neural networks, which has drawbacks such as easily stopping training at local optima and being easily affected by parameter values, leading many scholars to use various methods to improve it. In these studies, many achievements have been made, such as using Bayesian normalization to improve the network's generalization ability and improving the gradient descent method. In addition, some scholars have also used heuristic algorithms to improve the prediction model of BP neural networks. Zhou et al. combined genetic algorithm and simulated annealing algorithm to optimize the BP neural network, and the results showed that the GA-SA based BP neural network has strong generalization ability and global search ability, with high accuracy. Huang et al. proposed a BP neural network method based on improved particle swarm optimization algorithm to predict AQI, making the prediction results more accurate.

Other prediction models have also been improved, and Fan Wenting et al. proposed optimizing support vector machine prediction models based on an improved firefly algorithm. Gao Shuai et al. proposed a method that combines SVM with traditional moth to flame optimization algorithms. In addition, some scholars have proposed optimization algorithms to optimize LSTM. Zhang et al. optimized LSTM neural network parameters and applied them to air quality prediction in different cities, achieving certain results. Al Janabi et al. used PSO to optimize the weights, biases, number of hidden layers, number of nodes in each hidden layer, and activation function parameters of LSTM, achieving prediction of air pollutant concentrations for the next two days.

From the above research, it can be found that improving neural networks based on intelligent optimization algorithms has been widely applied and can achieve very good results. Therefore, it is particularly important to choose suitable optimization algorithms for optimizing neural network parameters. Through research, it has been found that the beetle whisker search algorithm has lower time and spatial complexity compared to most swarm intelligence algorithms and higher operational efficiency. Liao Liefu et al. compared the BAS algorithm with many new and mainstream intelligent optimization algorithms and found that it has strong competitiveness. Compared with the PSO algorithm, the BAS algorithm has a stronger ability to jump out of local extremes and converges faster; Compared with GA algorithm, BAS algorithm has simple operation, small computational complexity, and faster running speed; Compared with the FA algorithm, the BAS algorithm is less affected by parameter sensitivity; Compared with Bat Algorithm (BA) and Artificial Bee Colony (ABC), BAS algorithm has higher computational efficiency and lower computational complexity. However, the BAS algorithm also has the problem of easily falling into local extremes and unstable optimization results.

4. Closing Remarks

As we all know, substances in the air are constantly moving, so the concentration of pollutants monitored by air quality monitoring stations is constantly changing and easily affected by the surrounding environment, which leads to a certain spatial correlation between different stations. When the change trend of most data of different sites is the same or opposite and the fluctuation amplitude is similar, it is considered that there is spatial correlation between sites. The air quality situation is influenced by various complex factors, such as meteorological factors, geographical location, national policies, etc. After excluding other factors, this article conducts research on predicting the trend of air quality changes based on historical air quality data and meteorological data. Because substances in the air are constantly moving, the concentration of pollutants monitored by monitoring stations is also constantly changing and easily influenced by the surrounding environment. However, the specific impact is unknown, but this relationship can be explored through existing historical data. How to effectively construct prediction models and improve the accuracy of air quality prediction is currently a hot research topic. Considering the spatial correlation between monitoring stations, this paper proposes using unsupervised learning clustering algorithms to classify stations with similar attributes. Deep learning technology, as a complex machine learning algorithm, autonomously learns the intrinsic features of data by establishing multi-layer neural networks. Based on

this, this article proposes an air quality prediction model based on clustering technology and hybrid deep neural network. In order to obtain more accurate prediction results, the intelligent optimization algorithm is improved and the parameters of the prediction model are further optimized. In recent years, domestic and foreign scholars have also proposed other methods for air quality prediction. Han Xiaoguang et al. effectively combined grayscale correlation methods and radial basis function neural networks (RBF) to select indicator factors using grey methods. The RBF neural network predicted the main indices and achieved accurate prediction of Tianjin's air quality. Gao Shuai et al. further validated the effectiveness and accuracy of neural networks in air quality monitoring and prediction by combining BP neural networks with Mind Evolutionary Algorithm (MEA). Yuan et al. used collaborative filtering to distinguish the weights of different air indicators and simulated the situation of air pollution using BP neural network, providing favorable conditions for air pollution prevention and control work. Hu et al. proposed an Elman neural network prediction method based on chaos theory, which effectively learns nonlinear relationships in air quality data. Implementation analysis shows that the Elman chaotic prediction model has good predictive performance and application value. Sayegh et al. conducted comparative experiments using multiple linear regression models (MLRM), quantile regression models (QRM), generalized additive models (GAM), and enhanced regression trees (BRT), using meteorological data and chemical species from Mecca, Saudi Arabia as covariates to predict PM10 concentrations for the next hour. The experiment shows that QRM has a better predictive effect on PM10 hourly concentration. Compared with other models, the superiority of QRM model lies in its ability to simulate the contribution of covariates at different quantiles of model variables.

References

- [1] Anenber S C, Mohegh A, Goldberg D L, et al. Long-Term Trends in Urban NO₂ Concentrations and Associated Paediatric Asthma Incidence: Estimates from Global Datasets[J]. *The Lancet Planetary Health*, 2022, 6 (1):e49-e58.
- [2] Southerland V A, Brauer M, Mohegh A, et al. Global Urban Temporal Trends in Fine Particulate Matter (PM_{2.5}) and Attributable Health Burdens: Estimates from Global Datasets[J]. *The Lancet Planetary Health*, 2022, 6 (2):e139-e146.
- [3] Luo Shi. Who: 13 people die every minute from air pollution [N]. *Global Times*, 2021-10-13(005).
- [4] Dong K, Zeng X. Public willingness to pay for urban smog mitigation and its determinants: A case study of Beijing, China [J]. *Atmospheric Environment*, 2017, 173:355-363.
- [5] Zhu S, Yang L, Wang W, et al. Optimal-combined model for air quality index forecasting: 5 cities in North China [J]. *Environmental Pollution*, 2018, 243:842-850.
- [6] Xu H. Application of BP neural network optimization based on improved Drosophila algorithm in air quality prediction [D]. Nanchang: Nanchang University, 2020.
- [7] Z. Kang, Z. Qu, Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou[C]. 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCI), 2017:155-160.

- [8] Y. Huang, Y. Xiang, R. Zhao, et al., Air Quality Prediction Using Improved PSO-BP Neural Network[J]. IEEE Access, 2020, 8:99346-99353.
- [9] Fan Wenting, Wang Xiao. PM 2.5 Prediction Based on Improved Firefly Optimization Support Vector Machine [J]. Journal of Computer Systems and Applications, 2019, 28(1): 134-139.
- [10] Gao Shuai, Hu Hongping, Li Yang, et al. Air quality index prediction based on MFO-SVM [J]. Journal of North University of China: Natural Science Edition, 2018, 39(4):7-13. (in Chinese)
- [11] Zhang S, Lin M, Zou X, et al. LSTM-based air quality predicted model for large cities in China[J]. Nature Environment, 2020, 19(1): 229-236.
- [12] Al-Janabi, S., Mohammad, M. & Al-Sultan, A. A new method for prediction of air pollution based on intelligent computation [J]. Soft Computer, 2020, 24:661-680.
- [13] Liao Liefu, Yang Hong. A Survey on the Search Algorithm of Longniu Whiskers [J]. Computer Engineering and Applications, 2021, 57(12):54-64.
- [14] Wang T, Jiang F, Deng J, et al. Urban air quality and regional haze weather forecast for Yangtze River Delta region [J]. Atmospheric Environment, 2012, 58:70-83.