CHEMICAL ENGINEERING TRANSACTIONS
VOL. 71, 2018
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Xiantang Zhang, Songrong Qian, Jianmin Xu

ISBN 978-88-95608-68-6; **ISSN** 2283-9216
DOI: 10.3303/CET1871194

Analysis Technology of Environmental Monitoring Data Based on Internet of Things Environment and Improved Neural Network Algorithm

Lianli Huang[a,*], Kui Zhou[b]

[a] Institute of electrical and information engineering, Hubei University of automotive technology, Shiyan 442002, China
[b] Institute of automotive engineers, Hubei University of automotive technology, Shiyan 442002, China
huangll_dy@huat.edu.cn

With the development of industrialization, the problem of environmental pollution has become increasingly serious. Environmental monitoring data, as a measurement index of environmental quality, is increasingly valued by governments and citizens. However, the accumulated real-time monitoring data of the environment at current stage is mostly used to write basic reports, while the hidden laws or values still need to be further explored. This paper proposes two original environmental prediction models and conducts improvement on these two methods separately to predict the quality of the atmospheric environment. The following research results are obtained: the multivariate linear equation is optimized through stepwise linear regression, which can accurately predict the short-term atmospheric environmental quality; the improved BP neural network can predict the mid-term and long-term atmospheric environmental quality through short-term training.

1. Introduction

With the rapid advancement of industrialization and the rapid increase of urban population, the environmental pollution problem is becoming more and more serious (Beck et al., 1961, Jeffrey et al., 2000), including atmospheric pollution, such as acid rain and fog; water environment pollution, such as black odorous water, water bloom and red tide; solid waste pollution, such as construction waste and tailings pond pollution. In recent years, the atmospheric pollution has been particularly prominent. Atmospheric pollution directly leads to a decline in the quality of people's living environment, affecting human health and causing huge economic losses (Lauth et al., 2010, Abramovitch et al., 2015). In order to reduce the occurrence of air pollution, large funds have been invested in the prevention of air pollution at various levels in our country (Bey et al., 2017) and an atmospheric environment monitoring Internet of Things has been established. Through the monitoring of the quality of the atmospheric environment, the atmospheric environment quality is controlled in real time, which guides the industrial production and the construction of related pollution prevention facilities (Rodgers et al., 2015, Lochner et al., 2011). At present, almost all atmospheric environmental monitoring data is used to prepare environmental reports such as daily newspaper and annual reports (Arsie et al., 2010) while the value of data needs to be further explored (Li et al., 2014). For example, through historical monitoring data,the future trend of atmospheric environmental quality can be predicted so as to better guide people's life production activities (Chien et al., 2005; Chu et al., 2018; Chien et al., 2010). Meanwhile, it provides relevant scientific evidence for relevant government decision-making and management departments in the formulation of relevant systems (Chien et al., 2003).

At present, the methods for predicting the quality of atmospheric environment mainly include numerical prediction and statistical prediction. However, the model precision of numerical prediction is lower than that of statistical prediction (Chien et al., 2007) and the scope of application is limited (Cooper and Ekström, 2005, Reifman et al., 2000). Therefore, this paper adopts the statistical prediction with easy data acquisition and high prediction precision, including the BP neural network (Yang et al., 2018) and multiple linear regression equation for the analysis of atmospheric environment prediction.

Please cite this article as: Huang L., Zhou K., 2018, Analysis technology of environmental monitoring data based on internet of things environment and improved neural network algorithm, Chemical Engineering Transactions, 71, 1159-1164 DOI:10.3303/CET1871194

2. Research Basis

2.1 BP Neural Network

BP neural network is a neural network of error inverse propagation, as is shown in Figure 1: after the data is input, it is forwardly propagated from the input layer to the output layer via the hidden layer. According to the set error, the output layer corrects the weight through the hidden layer, which is the error inverse propagation, thereby achieving a stepwise improvement of the precision of the output value of the neural network.

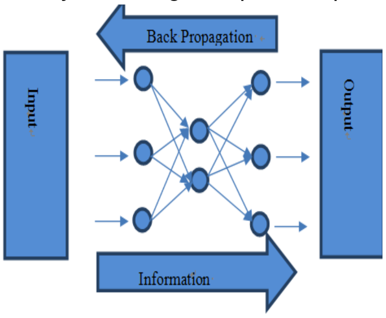

Figure 1: BP Neural Network Topolgy

2.2 Genetic Algorithm

The genetic algorithm is developed from the Darwin's theory of evolution and Mendel's genetics. It uses the imitation of biological gene coding to encode individuals, serving as the initial population. The selection, crossover and mutation are completed according to the principle of survival of the fittest. The new population is formed and the above operation is repeated, thereby realizing the retention of excellent genes and inheriting these genes to the offspring. Therefore, the population can better adapt to the environment and continue to breed and evolve.

3. Data Collection and Processing

The GB3095-2012 standard used in China is used to evaluate the quality of the atmospheric environment. That is, using six pollutant indexes, namely PM2.5, PM10, CO, SO2, O3 and NO2 for the measurement. In addition, meteorological conditions such as temperature, humidity, wind speed, wind direction and air pressure can affect the quality of the atmospheric environment by affecting the diffusion of pollutants in the atmospheric environment. For example, when the humidity is high, the degree of air pollution will increase. Therefore, this paper will collect 6 indexes of atmospheric pollutants and corresponding 5 meteorological indexes.

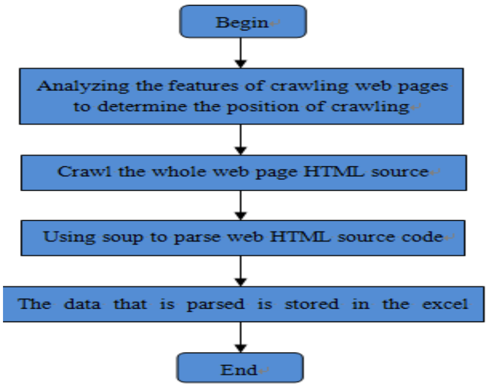

Figure 2: Flow chart of web crawler

This paper downloads the monitoring data of atmospheric environment of the environmental protection bureau in a city by writing the crawler software. The data collection process is shown in Figure 2.

4. Prediction model

The crawler software is used to obtain atmospheric pollutants such as PM2.5, PM10, CO, SO2, O3, NO2, as well as the meteorological data such as temperature, humidity, wind speed, wind direction and pressure. For example, at the monitoring point A, there are 22,000 pieces of PM10 concentration data, of which 20,000 pieces of data are used as training data and the rest 2,000 pieces of data are used as test data.

4.1 Traditional Multivariate Model

The preprocessed data is constructed into a multiple linear regression model, and correlation coefficient test, F test and t test are performed. The predicted result is then compared to the test data.

4.1.1 Traditional Multiple Linear Regression Prediction Model

The dependent variable is PM10 and the independent variable is five types of meteorological data such as temperature and pressure. The modeling method is all input.
(1) The correlation coefficient test results are shown in Table 1 and R2 represents the fitting effect. The larger the value, the better the fitting effect.

Table 1: Test of Correlation Coefficient

Model	R	R^2	R2 adjusted	Standard deviation rate error
1	0.438	0.509	0.507	26.73

(2) The result of the F test is shown in Table 2. As it can be seen from the table, the result of F test is <0.01. Therefore, five meteorological indexes have a significant impact on the concentration of PM10.

Table 2: Test of significance

Model	Squre	Df	Mean squre	F	Significance
Regression	237971.932	5	475897.746	80.792	0.00
Residual	3424801.21	579	5883.049		
Statistics	5804671.02	596			

(3) The result of the t test is shown in Table 3. The non-normalized coefficient is used to list the regression equation. The normalization coefficient is used to reflect the degree of influence of the independent variable on the dependent variable; the partial regression coefficient is used to determine whether the influence of an independent variable on the dependent variable is statistically significant. When it is <0.05, it indicates significant statistical significance; when it is <0.01, the statistical significance is extremely significant.

Table 3: T-test

Modle	Unstandardized Coefficients	Standardized coefficient	T	Significance
constant	398.212		5.461	0.00
pressure	-3.796	-0.248	-5.377	0.00
Temperature	-1.296	-0.0497	-0.981	0.032
Moisture	2.223	0.379	9.201	0.00
Wind speed	-55.059	-0.228	-7.068	0.00
Wind direction	-0.207	-0.179	-5.425	0.00

(4) The linear regression equation of the prediction model is: y=-3.796x1-1.296x2+2.223x3-55.059x4-0.207x5+398.212 and the comparison between the true value and the predicted data is shown in Figure 3.

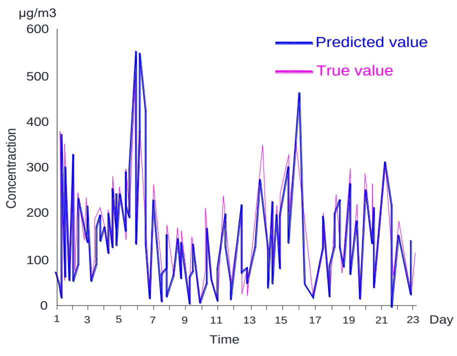

Figure 3: Comparison of predicted value by traditional multiple linear regression model and true value

4.2 Improved Multiple Linear Regression Prediction Model

Table 4: Model abstract

Model	R	R2	R2 adjusted	Standard deviation rate error	Introduce variable
1	0.781	0.739	0.821	20.89	Constant,PM2.5, season
2	0.813	0.769	0.825	20.75	Constant,PM2.5, season, Temperature, O3
3	0.836	0.808	0.825	20.17	Constant,PM2.5, season, wind speed, Temperature, O3
4	0.852	0.817	0.827	19.964	Constant,PM2.5, season, wind speed, Temperature, O3, pressure
5	0.863	0.842	0.829	19.697	Constant,PM2.5, season, wind speed, Temperature, O3, pressure, Moisture
6	0.881	0.842	0.828	19.603	

Considering the physical and chemical reactions between pollutants and the significant impact of seasonal factors on pollutants, other pollutants and seasonal factors are included in the multivariate equation for the optimization. The seasonal variables are: spring 108.1 μg/m3, weight 0.25; summer 97.8 μg/m3, weight 0.2; autumn 112.3 μg/m3, weight 0.25; winter 121.9 μg/m3, weight 0.3.

The significance test is performed by introducing independent variables one by one into the regression model until all significant independent variables are introduced into the regression model. The result of the stepwise regression model is shown in Table 4.

It can be seen from Table 4 that PM2.5 has the most significant impact on PM10, while the impact of atmospheric pollutants SO2, NO2, CO and the meteorological factor, wind direction on PM10 can be negligible. Therefore, after eliminating these indexes, Table 5 can be obtained.

Table 5: Test of Correlation Coefficient

Modle	Unstandardized Coefficients	Standardized coefficient	T	Significance
Constant	-90.087		-4.875	0.00
PM2.5	30.304	0.937	89.851	0.00
Temperature	19.729	0.113	8.	0.032
O3	-0.359	-0.10	-7.61	0.00
Wind speed	8.541	0.038	4.31	0.00
Pressure	0.897	0.059	5.081	0.00
Moisture	5.184	0.029	2.857	0.003
Season	10.280	0.269	20.351	0.005

The regression equation can be obtained: y=30.304x1+19.729x2-0.359x3+8.541x4+0.897x5+5.184x6+10.280x7-90.087

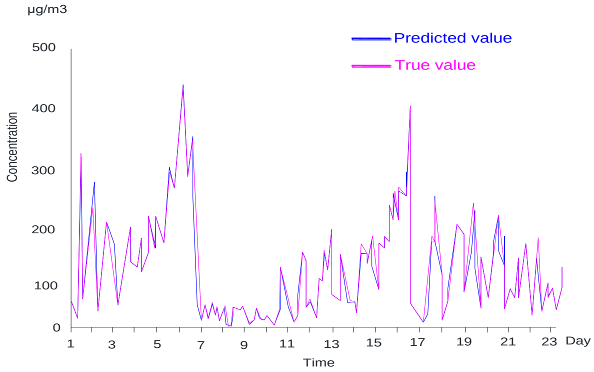

Figure 4: Comparison of predicted value by optimal multiple linear regression model and true value

Using this model, the true value and the predicted value are shown in Figure 4.

It can be seen from the Figure that the fitting degree of the optimized model reaches 0.828, which is significantly higher than that of the original model, indicating that the prediction for PM10 by the stepwise linear regression method is more precious after considering meteorological factors and other pollutants.

In addition, it can also be seen from the Figure that the prediction error in the short term (4 days) is the smallest. At the same time, PM2.5, wind speed, air pressure, humidity and season have an enhancing effect on PM10 concentration. The impact of PM2.5 on PM10 is the greatest; while O3 and temperature has a weakening effect on PM10.

4.3 Traditional BP Neural Network Model

The neural network structure needs to be determined by the number of hidden layers, the number of nodes in the input layer, the number of nodes in the hidden layer, the number of nodes in the output layer, the activation function, training method and training parameters. In this paper, a three-layer neural network with a layer of hidden layer is used and it is set to be 11 and 1 according to the principle that the number of input and output nodes is as small as possible. The number of neurons in the hidden layer is determined by formula $M = \sqrt{n+m} + a$. In the formula, a is a constant between 0 and 10, and m and n are the number of neurons in the input layer and the output layer respectively. The number of hidden layers in this paper is 6. The input function in the hidden layer is $f(x) = \frac{1}{1+e^{-x}}$ and the linear activation function is used on the output layer. The learning rate is 0.01.

The fitting relationship between the predicted value and the true value of the BP neural network is shown in Figure 5:

It can be seen from the above Figure that the traditional BP neural network can better reflect the variation trend of atmospheric pollution in the future and can relatively accurately predict the concentration of atmospheric pollutants. Its goodness of fit is 0.84501.

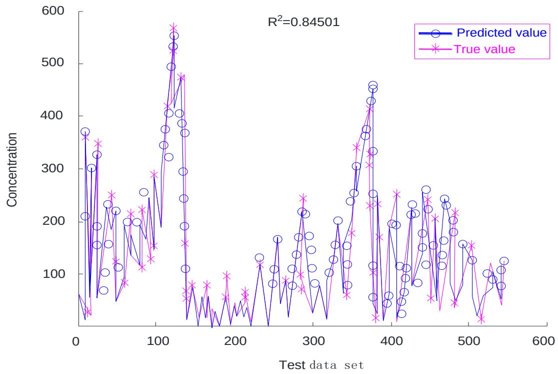

Figure 5: Comparison of predicted value of BP neural network and true value

5. Research Conclusions

With the increase of people's requirement for environmental quality, the prediction for the variation trend of environmental quality using the monitoring data becomes increasingly important. In this paper, the original multiple linear regression model, the original BP neural network and the optimized model are used to predict the atmospheric environmental quality. The following research conclusions are drawn:

(1) The traditional multiple linear regression model can only predict the variation trend of atmospheric environmental quality coarsely; while the other three models can predict the concentration of future atmospheric pollutants accurately.

(2) The stepwise linear regression can be used to predict PM_{10} more accurately after considering meteorological factors and other pollutants. Tthe prediction error for the short term (4 days) is the smallest.

(3) The prediction for the mid-term and long-term atmospheric environmental quality is the best using the optimized BP neural network model.

Acknowledgement

Hubei Provincial Department of Education Scientific Research Project B2016087. Shiyan South-to-North Water Transfer Counterpart Collaborative Project 2017syn043A.

References

Abramovitch A., Pizzagalli D.A., Geller D.A., Reuman L., Wilhelm S., 2015, Cigarette smoking in obsessive-compulsive disorder and unaffected parents of ocd patients, European Psychiatry, 30(1), 137-144, DOI: 10.1016/j.eurpsy.2013.12.003

Arsie I., Marra D., Pianese C., Sorrentino M., 2010, Real-time estimation of engine nox emissions via recurrent neural networks, IFAC Proceedings Volumes, 43(7), 228-233, DOI: 10.3182/20100712-3-DE-2013.00117

Beck A.T., Ward C.H., Mendelson M., Mock J., Erbaugh J., 1961, An inventory for measuring depression, Arch Gen Psychiatry, 4(6), 561-571. DOI: 10.1001/archpsyc.1961.01710120031004

Bey K., Lennertz L., Riesel A., Klawohn J., Kaufmann C., Heinzel S., 2017, Harm avoidance and childhood adversities in patients with obsessive-compulsive disorder and their unaffected first-degree relatives, Acta Psychiatrica Scandinavica, 135(4), DOI: 10.1111/acps.12707

Chien C.F., Chen W.C., Lo F.Y., Lin Y.C., 2005, A case study to evaluate the productivity changes of the thermal power plants of the taiwan power company. IEEE Transactions on Energy Conversion, 22(3), 680-688. DOI: 10.1109/TEC.2006.888026

Chien T.W., Chu H., Hsu W.C., Tseng T.K., Hsu C.H., Chen K.Y., 2003, A feasibility study on the predictive emission monitoring system applied to the hsinta power plant of taiwan power company, Air Repair, 53(8), 1022-1028. DOI: 10.1080/10473289.2003.10466241

Chien T.W., Chu H., Hsu W.C., Tu Y.Y., Tsai H.S., Chen K.Y., 2005, A performance study of pems applied to the hsinta power station of taipower, Atmospheric Environment, 39(2), 223-230, DOI: 10.1016/j.atmosenv.2004.09.062

Chien T.W., Hsueh H.T., Chu H., Hsu W.C., Tu Y.Y., Tsai H.S., 2010, A feasibility study of a predictive emissions monitoring system applied to taipower's nanpu and hsinta power plants, Air Repair, 60(8), 907-913, DOI: 10.3155/1047-3289.60.8.907

Chu X., Zhou Z.X., Xiang X.J., He S.L., Hou X., 2018, Monitoring of long-span bridge deformation based on 3D laser scanning, Instrumentation, Mesure, Métrologie, 17(1), 113-130, DOI: 10.3166/I2M.17.113-130

Cooper D.A., Ekström M., 2005, Applicability of the pems technique for simplified nox monitoring on board ships, Atmospheric Environment, 39(1), 127-137, DOI: 10.1016/j.atmosenv.2004.09.019

Jeffrey B.H., Richard J.D., 2000, Decreased responsiveness to reward in depression, Cognition & Emotion, 14(5), 711-724.

Lauth B., Arnkelsson G.B., Magnússon P., Skarphéðinsson G.Á., Ferrari P., Pétursson H., 2010, Validity of k-sads-pl (schedule for affective disorders and schizophrenia for school-age children--present and lifetime version) depression diagnoses in an adolescent clinical population, Nordic Journal of Psychiatry, 64(6), 409-409.

Li P.K., Pan R., Chen C., 2014, A novel neural network based modeling for control of nox emission in power plant. Applied Mechanics & Materials, 643(643), 385-390.

Lochner C., Serebro P., Van D.M.L., Hemmings S., Kinnear C., Seedat S., 2011, Comorbid obsessive-compulsive personality disorder in obsessive-compulsive disorder (ocd), a marker of severity, Progress in Neuropsychopharmacology & Biological Psychiatry, 35(4), 1087-1092.

Reifman J., Feldman E.E., Wei T.Y., Glickert R.W., 2011, An intelligent emissions controller for fuel lean gas reburn in coal-fired power plants, Air Repair, 50(2), 240-251.

Rodgers S., Ajdacic-Gross V., Kawohl W., Müller M., Rössler W., Hengartner M.P., 2015, Comparing two basic subtypes in ocd across three large community samples: a pure compulsive versus a mixed obsessive–compulsive subtype, European Archives of Psychiatry & Clinical Neuroscience, 265(8), 719-734, DOI:10.1007/s00406-015-0594-0

Yang C., Lv S., Gao F., 2018, Water pollution evaluation in lakes based on factor analysis-fuzzy neural network, Chemical Engineering Transactions, 66, 613-618 DOI: 10.3303/CET1866103