Analysis and Research on Influencing Factors of Haze Weather

Hua Lan* and Yun Pan
School of Computer Science Community University of China, Beijing, China

*1195252827@qq.com

Abstract. In recent years, the causes, prevention and control methods and prediction system of continuous haze weather have become the focus of social research. Taking the meteorological observation data of Hebei province as an example, this paper studies the influencing factors of continuous haze weather. Firstly, the improved TOPSIS algorithm is used to preprocess the data. Then, it combines the data mining methods such as hierarchical analysis and grey correlation analysis to carry out modeling analysis, and it is concluded that the particulate matter in the pollution source has the greatest impact on the haze weather. Moreover, as the aerodynamics factor affects the diffusion and aggregation of the haze pollution source, it is found that it has a linear influence on the formation of haze weather. Finally, through the current meteorological observation data, BP neural network is used to predict haze weather changes. A large number of experimental results show that, on the premise of allowing a certain error rate, the prediction effect of the BP neural network model is relatively accurate, and it also indicates that the formation of haze weather is closely related to the air quality index factor and meteorological index factor.

1. Introduction
In recent years, with the rapid development of heavy industry in our country, haze weather in our country has been frequent, and the total number of days of haze weather has also increased year by year, and the hazards caused by haze weather have become more and more serious. Especially since 2013, the haze weather has led to great economic losses in China and affected the health of residents[1]. In the winter of 2016, several provinces in north China, including Beijing, Shanxi, Hebei and Tianjin, suffered from severe haze weather[2]. For example, due to the haze weather, the air is cloudy and the visibility drops sharply, so the flights in China are forced to stop sailing, even to the point of being forced to land on the spot. Besides, the haze weather has led to cars being unable to drive normally. Vehicles have been forced to stop their journey, and even some places have closed down the expressway. In addition, it is also worth noting that as the regions affected by haze weather are expanding year by year, the harm degree of haze is also increasing. More and more people are affected by haze weather, especially the elderly and children with low immunity[3]. During the periods of frequent haze weather in spring and winter, they would fall ill to various degrees due to the influence of haze weather. In view of this, the whole society's attention to haze weather is constantly increasing, and therefore, the research on the influencing factors of haze weather becomes a hot topic.

Based on the mechanism of the formation of haze weather and prompted haze weather form the influence of the factors can be roughly divided into two groups[4], one is the source of pollution in the composition of haze weather (i.e., sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone and
PM10 and PM2.5), The other (i.e., humidity, temperature, wind velocity, visibility, and atmospheric pressure) is through the diffusion of the composition of the haze weather, which in turn affects the weather conditions formed by the haze weather. Based on this, this article takes Hebei province meteorological monitoring system to collect the relevant data as the research object. Firstly, the AHP (i.e., the analytic hierarchy process) method and the improved TOPSIS algorithm are used to preprocess the abnormal data in the original data. Then, data mining methods such as grey correlation analysis and correlation analysis are used to explore the influencing factors of haze weather from the aspects of pollution sources and meteorological indexes respectively. Finally, by using the historical data of meteorological monitoring system, the BP neural network prediction model is constructed to analyze the future haze weather.

2. The improved TOPSIS algorithm

TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) algorithm was first proposed by two foreign scholars in 1981, namely C.L. Wang and K. Yoon. TOPSIS algorithm is to calculate the positive ideal solution and the negative ideal solution of each indicator by calculating the distance from each solution to the positive and negative ideal solution. Therefore, it is sometimes called the ideal solution [5].

Due to the wide variety and complexity of the influencing factors of haze weather, some abnormal data, such as repeated data, negative data, inverted data (PM2.5>PM10) and zero value data, often appear in the data acquired by the actual meteorological monitoring system. How to carry out data cleaning for these abnormal data is the primary task to study the correlation of haze weather factors. In this paper, the algorithm idea of TOPSIS is used to build an improved TOPSIS model, and the ideal solution of each abnormal data was obtained, and the ideal value was used to replace the abnormal data, so as to achieve the goal of removing the abnormal data.

The steps to build an improved TOPSIS model are shown below [6].

(1) The vector normalization method is used to obtain the normative decision matrix, Set the decision matrix of multi-attribute decision problem as $A = (a_{ij})_{mn}$, Set the normalized decision matrix as

$$b_{ij} = a_{ij} / \sqrt{\sum_{n} a_{ij}^2} \quad (1)$$

Where $i = \{i = 1, 2, ..., m\}; j = \{1, 2, ..., n\}$.

(2) Form the weighted canonical matrix $C = (c_{ij})_{mn}$. Set the weight vector of each attribute given by the decision maker $w = [w_1, w_2, ..., w_n]^T$, and

$$c_{ij} = w_j \cdot b_{ij} \quad (2)$$

Where $i = \{i = 1, 2, ..., m\}; j = \{1, 2, ..., n\}$.

The positive ideal solution and negative ideal solution are determined, respectively, $C^*$ and $C^0$. Set the JTH attribute value of the positive ideal solution $C^*$ to be $c_{ij}^*$, Set the JTH attribute value of the negative ideal solution $C^0$ to be $c_{ij}^0$. Then, the positive and negative of ideal solution are

$$c_{ij}^* = \frac{\max_{i} c_{ij}}{\min_{i} c_{ij}} \quad \text{and} \quad c_{ij}^0 = \frac{\min_{i} c_{ij}}{\max_{i} c_{ij}}$$

Where the $j$ of $\min_{i} c_{ij}$ is the attribute of benefit type, the $j$ of $\max_{i} c_{ij}$ is the attribute of cost type,
Using the vector projection method, the best ideal solution is selected to replace the abnormal data. The normalized original data vector and ideal solution vector are \( \hat{d} = (a_1, a_2, \ldots, a_n)^T \) and \( \beta = (b_1, b_2, \ldots, b_n)^T \) respectively, and vector projection is

\[
\gamma^\beta_j = [\|d\| \times \hat{d}^T \beta] / \left[ \|\hat{d}\| \|\beta\| \right] = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} b_i^2}}
\]

(3) Sort \( \gamma^\beta_j \) and select the ideal solution as the replacement value of the abnormal data.

3. Experimentation

At present, China's ministry of environmental protection regulates the use of AQI (air quality index) to evaluate air quality. In this paper, the correlation degree between sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, PM10 and AQI is studied by using the grey relational analysis model, so as to analyze the influence of air quality influencing factors on the formation of haze weather and predict the haze weather.

This paper takes the air quality monitoring data monitored by century park monitoring station in a city of Hebei province from September 1 to 30, 2014 as the experimental data, and designed the experimental process according to the research objectives as shown in figure 1 below.

![Figure 1. Experimental flow graph.](image)

4. Experimental results and analysis

This experimental scheme is implemented on the platform of Matlab 2016b, the 64-bit operating system of Windows 10. According to the experimental steps, the data preprocessing is carried out first,
and then the grey correlation analysis model is constructed to analyze the air quality index factors. Table 1 shows the correlation ranking of air quality indexes on haze weather.

**Table 1.** Ranking of the influence degree of air quality index on formation of haze weather.

| Air quality index | Correlation degree | Order of influence |
|-------------------|--------------------|--------------------|
| PM2.5             | 0.6734             | 1                  |
| SO₂               | 0.5261             | 2                  |
| PM10              | 0.5127             | 3                  |
| O₃                | 0.5072             | 4                  |
| NO₂               | 0.4999             | 5                  |
| CO                | 0.4974             | 6                  |

According to the data in table 1, it can be concluded that PM2.5 has the greatest impact on haze formation, followed by sulfur dioxide, while the influence of carbon monoxide in air quality index is minimized, carbon monoxide to form fog weather correlation is 0.4974 > 0.3, that is, between the two is strongly related. Therefore sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone and PM10 and PM2.5 has greater impact on forming haze weather.

Figure 2 is the histogram drawn by the correlation coefficient matrix of sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, PM10 and PM2.5 and AQI in a ten-day period in September 2014.

![Figure 2: Correlation degree of air quality index to haze weather.](image)

![Figure 3: Influence of meteorological factors on air quality index factors.](image)

From the above figure 2, it can be seen that compared with ozone, nitrogen dioxide and carbon monoxide, PM10, PM2.5 and sulfur dioxide are highly correlated with AQI, and then it can be inferred that PM10, PM2.5 and sulfur dioxide have more significant effects on the formation of haze weather. According to the relevant knowledge of aerodynamics, humidity, temperature, wind speed, visibility and air pressure in meteorological indicators affect their concentration by influencing the speed of aggregation and diffusion of air quality indicators such as sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, PM10 and PM2.5. Therefore, it is necessary to analyze the influence of meteorological factors on air quality index factors at the same time. Figure 3 is a histogram of the correlation degree between meteorological factors and air quality index factors.
From the above figure 3, compared with other meteorological factors, humidity, temperature and wind speed have the most significant influence on air quality index factors, that is, the change of humidity, temperature and wind speed has the most obvious influence on air quality index factor concentration. Finally, BP neural network model is built to predict the autumn haze weather, the input layer as sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone and PM10 five air quality index factors and the humidity, temperature, wind velocity, visibility and air pressure five meteorological index factor, the output layer of BP model is AQI and PM2.5, number of hidden layer nodes is 10 in the middle. In addition, the input sample is the value of each indicator in September 2014, the test sample is the value of AQI and PM2.5 in September 2014, and the prediction sample is the value of AQI and PM2.5 in September 2015. After several trainings until the ideal effect is achieved (that is, the error between the network monitoring value and the actual monitoring value is up to the requirement), figure 4 shows the correlation between the actual measurement value and the predicted value of the model.

As can be seen from the above figure 4, after the BP neural network model has been trained for many times, the correlation between the actual measured value and the predicted value of the model reaches 0.93. Therefore, it can be said that the trained model can be used to predict the haze weather in autumn. Figure 5 shows the fitting between the predicted network output AQI and the actual AQI by the model.

It can be seen from figure 5 that the established BP neural network model has an excellent prediction effect on the autumn haze weather. Within a certain error range, the predicted value is almost consistent with the actual monitoring value. Therefore, this model can be used to predict the future autumn haze weather in Hebei province.

5. Conclusion
In this paper, based on the correlation study of haze weather influence factors in Hebei Province, firstly, TOPSIS model is constructed, and the abnormal data are replaced by ideal solution. Then, after the collection of data preprocessing, and the influence degree of each influence factor on haze weather is quantitatively analyzed by grey correlation method. Finally, according to the autumn weather data collected, The BP neural network model is established for training and learning, predicting the autumn haze weather conditions. In the research on the correlation analysis of haze weather influencing factors, there are still many areas that need to be continued in the future, such as the following aspects:
(1) from the perspective of haze weather impact factors, in addition to the study of the pollutants that affect haze formation provided by this region, other pollutants provided by adjacent regions should also be considered.

(2) from the perspective of establishing the forecast model of haze weather, it is necessary to build a forecast model capable of predicting haze weather in spring, summer and winter in the future.

References

[1] Gao Ge. Climatic Characteristics and Variation Analysis of Haze Days in China from 1961 to 2005. *Journal of Geography*, 2008, 63(7):761-748.

[2] Guo lijun. A study on the types, structural characteristics and mechanism of persistent smog-haze weather in north China. Nanjing, *nanjing university of information engineering*, 2015.

[3] Tao J, Chan C Y, Lin M, et al. Regression analyses between recent air quality and visibility changes in megacities at four haze regions in China. *Aerosol and Quality Research*, 2012, 12(6):1049-1061.

[4] Bi xueyan, deng xuejiao, li fei, wu wei, et al. Study on the visibility reduction caused by atmospheric haze in the pearl river delta, 2015.

[5] Si Shoukui, Sun Xi Jing. Data modeling algorithm and application. Beijing, *National Defense University press*, 2014.

[6] Zhang Kunlong, Yang Wenguang. Computational intelligence. Beijing, *Coal Industry Press*, 2015.