Fuzzy cluster analysis of air quality in Beijing district

Hongkai Liu
China University of Mining and Technology (Beijing), School of Chemistry and Environmental Engineering, Beijing 100083, China.
1812911979@qq.com

Abstract. The principle of fuzzy clustering analysis is applied in this article, by using the method of transitive closure, the main air pollutants in 17 districts of Beijing from 2014 to 2016 were classified. The results of the analysis reflects the nearly three year’s changes of the main air pollutants in Beijing. This can provide the scientific for atmospheric governance in the Beijing area and digital support.

1. Introduction
In recent years, with the acceleration of urbanization, the control of air pollution has become a top priority in urban development and one of the hot spots in society [1-3]. Beijing is the political, cultural and economic center of China and the construction of Beijing is an important part of the country's development strategy of "common prosperity and economic revitalization".

In the evaluation of atmospheric conditions, the methods used mainly include systematic clustering analysis, grey clustering analysis and fuzzy clustering analysis [4-6]. Clustering problem is according to the actual situation and a certain standard to identify the proximity between the objects, and to put close to each other objects, the fuzzy clustering analysis is based on the study of the attributes of the object itself to construct the fuzzy matrix, on this basis, according to certain membership to determine the relationship between classification [7]. The influence of human activities on the air environment is very complex, there are a lot of things and phenomena are fuzzy quality, in air environment impact assessment on the recognition and measurement of exist a lot of fuzziness, so only choose a reasonable method, can perfect to reflect the situation of air environmental quality. Taking Beijing as an example, through to the 17 to collect the average density of major pollutants in years, guided by the principle of fuzzy mathematics, and the annual average concentration of major pollutants were analyzed, and the scientific classification, for Beijing city governance and environmental comprehensive control to provide advice, for the urban environment and urban management to provide scientific basis for air [8].

In 1965, L. A. Zadeh, an American expert on cybernetics, first proposed the concept of fuzzy sets, which marked the birth of fuzzy mathematics. Fuzzy mathematics is to study and deal with ambiguity phenomenon, a kind of mathematical theory and method. The so-called fuzziness refers to the uncertainty and uncertainty in objective things, the root of which lies in the intermediary transition between objective things. Fuzzy mathematics is an emerging subject which has wide application in pattern recognition, artificial intelligence and so on [9-11].

2. Principle and method of fuzzy clustering analysis
The fuzzy clustering analysis is divided into the following steps
2.1. Data standardization

The domain $U$ represents all the objects of all $n$ samples to be classified, where $x_i = (x_{i1}, x_{i2}, \cdots, x_{im})$, $(i = 1, 2, \cdots, n)$ represents the $m$ attribute eigenvalues of the $i$-th sample object to be classified, so that an original data matrix is obtained, denoted as

$$(x_{ij})_{n \times m} = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{im} \\ x_{i2} & x_{i2} & \cdots & x_{jm} \\ \vdots & \vdots & & \vdots \\ x_{in} & x_{in} & \cdots & x_{mn} \end{bmatrix}$$

In order to compare the data of different dimensions, it is necessary to make a standardized transformation of the original data in general, and the common methods are as follows

1. Standard deviation transformation

$$x'_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k}, \quad (i = 1, 2, \ldots, n; k = 1, 2, \ldots, m)$$

where

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_{ik}, \quad s_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}.$$

2. Range transformation

$$x''_{ik} = \frac{x_{ik} - \min_{1s\leq m} \{x_{ik}\}}{\max_{1s\leq m} \{x_{ik}\} - \min_{1s\leq m} \{x_{ik}\}}, \quad (k = 1, 2, \cdots, m)$$

2.2. Establish fuzzy similarity matrix

For any two objects $x_i$ and $x_j$ in the domain $U$, the similarity coefficient $r_{ij} = R(x_i, x_j)$ is determined according to the degree of similarity between them, and then the $n$ order fuzzy similarity matrix is established. There are several ways to determine the similarity coefficients:

1. Scalar product method

$$r_{ij} = \begin{cases} 1, & i = j \\ \frac{1}{M} \sum_{k=1}^{m} x_{ik} \cdot x_{jk}, & i \neq j \end{cases}$$

in which $M = \max_{i \neq j} (\sum_{k=1}^{m} x_{ik} \cdot x_{jk})$.

2. Included angle cosine method

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2} \sqrt{\sum_{k=1}^{m} x_{jk}^2}}, \quad (i, j = 1, 2, \cdots, n)$$

3. Correlation coefficient method
\[
    r_{ij} = \frac{\sum_{k=1}^{m} |x_{ik} - \bar{x}_i| |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^{m} (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{m} (x_{jk} - \bar{x}_j)^2}}
\]

in which \( \bar{x}_i = \frac{1}{m} \sum_{k=1}^{m} x_{ik}, \bar{x}_j = \frac{1}{m} \sum_{k=1}^{m} x_{jk} \).

4. Maximum and minimum method
\[
    r_{ij} = \frac{\sum_{k=1}^{m} (x_{ik} \land x_{jk})}{\sum_{k=1}^{m} (x_{ik} \lor x_{jk})}, \quad (i, j = 1, 2, \ldots, n)
\]

5. Arithmetic mean minimum method
\[
    r_{ij} = \frac{2 \sum_{k=1}^{m} (x_{ik} \land x_{jk})}{\sum_{k=1}^{m} (x_{ik} + x_{jk})}, \quad (i, j = 1, 2, \ldots, n)
\]

6. Geometric mean minimum method
\[
    r_{ij} = \left( \sum_{k=1}^{m} x_{ik} \land x_{jk} \right)^{1/m}, \quad (i, j = 1, 2, \ldots, n)
\]

7. Distance method
\[
    r_{ij} = 1 - c \sum_{k=1}^{m} |x_{ik} - x_{jk}|, \quad (i, j = 1, 2, \ldots, n)
\]

or
\[
    r_{ij} = 1 - c \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}, \quad (i, j = 1, 2, \ldots, n)
\]

The value of the parameter \( c \) in the formula is properly chosen so that the similarity coefficient \( r_{ij} \in [0,1] \).

2.3. Calculation transitive closure of fuzzy similar matrices
The square of the fuzzy similarity matrix \( R \) is computed by the square method in turn
\[
    R \rightarrow R^2 = R \circ R \rightarrow R^4 = R^2 \circ R^2 \rightarrow \ldots \rightarrow R^{2^t} \rightarrow \ldots
\]

The transitive closure of \( R \) is \( t(R) = R^{2^t_0} \) until the \( R^{2^t_0} = R^{2^t_0} \) is present.

2.4. Clustering
Make a dynamic clustering map based on transitive closure.

3. Research results and discussion
According to the Beijing Municipal Environmental Protection Bureau published on June 2, 2017, the major pollutants in the Beijing Municipal Environmental Status Bulletin include: \( PM_{2.5}, SO_2, NO_2 \) and respirable particulate matter \( PM_{10} \), the specific data in Table 1 [12].
Table 1. Annual average concentration of major pollutants in Beijing in 2016
(Unit: $\mu g/m^3$).

| Region                  | $PM_{2.5} (x_1)$ | $SO_2 (x_2)$ | $NO_2 (x_3)$ | $PM_{10} (x_4)$ |
|-------------------------|------------------|--------------|--------------|-----------------|
| Dongcheng district ($x_1$) | 77.0             | 11.0         | 51.0         | 95.0            |
| Xicheng district ($x_2$)  | 78.0             | 12.0         | 53.0         | 98.0            |
| Chaoyang district ($x_3$) | 75.0             | 12.0         | 51.0         | 91.0            |
| Haidian district ($x_4$)  | 72.0             | 11.0         | 58.0         | 87.0            |
| Fengtai district ($x_5$)  | 79.0             | 11.0         | 53.0         | 99.0            |
| Shijingshan district ($x_6$) | 78.0         | 11.0         | 50.0         | 107.0           |
| Mentougou district ($x_7$) | 68.0             | 10.0         | 42.0         | 91.0            |
| Fangshan district ($x_8$) | 83.0             | 15.0         | 57.0         | 102.0           |
| Tongzhou district ($x_9$) | 80.0             | 15.0         | 55.0         | 98.0            |
| Shunyi district ($x_{10}$) | 71.0             | 10.0         | 43.0         | 82.0            |
| Daxing district ($x_{11}$) | 89.0             | 15.0         | 56.0         | 107.0           |
| Changping district ($x_{12}$) | 61.0           | 8.0          | 46.0         | 85.0            |
| Pinggu district ($x_{13}$) | 70.0             | 11.0         | 30.0         | 85.0            |
| Huairou district ($x_{14}$) | 61.0             | 7.0          | 28.0         | 77.0            |
| Miyun district ($x_{15}$) | 61.0             | 8.0          | 32.0         | 75.0            |
| Yanqing district ($x_{16}$) | 60.0             | 10.0         | 34.0         | 74.0            |
| Beijing Economic and Technological Development Zone ($x_{17}$) | 81.0           | 12.0         | 51.0         | 99.0            |

In the domain $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}\}$, the similarity coefficient is obtained by using the absolute distance method and $c = 0.1$

$$r_{ij} = 1 - 0.1 \sum_{k=1}^{m} |x_{ik} - x_{jk}|$$

(11)

The fuzzy similarity matrix $R = (r_{ij})_{17x17}$ is obtained, and the clustering map is obtained by MATLAB software, as shown in Figure 1.

Figure 1. 2016 Dendrogram.
According to the above dynamic clustering map, the following results can be obtained:

When \( \lambda = 1 \), **U** is divided into 17 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_8 \}, \{ x_9 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \}, \{ x_{17} \} \]

When \( \lambda = 0.9425 \) **U** is divided into 16 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_9 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.9366 \) **U** is divided into 15 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_9 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.9318 \) **U** is divided into 14 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_5 \}, \{ x_7 \}, \{ x_9 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.9229 \) **U** is divided into 13 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_9 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.9138 \) **U** is divided into 12 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.9136 \) **U** is divided into 11 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8846 \) **U** is divided into 10 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.882 \) **U** is divided into 9 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8753 \) **U** is divided into 8 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8704 \) **U** is divided into 7 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.864 \) **U** is divided into 6 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8515 \) **U** is divided into 5 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8089 \) **U** is divided into 4 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.8023 \) **U** is divided into 3 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.7923 \) **U** is divided into 2 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

When \( \lambda = 0.7677 \) **U** is divided into 1 categories
\[ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \}, \{ x_7 \}, \{ x_{10} \}, \{ x_{11} \}, \{ x_{12} \}, \{ x_{13} \}, \{ x_{14} \}, \{ x_{15} \}, \{ x_{16} \} \]

According to the value of the threshold \( \lambda \) from big to small and comparing the difference between the adjacent two values, it is concluded that when \( \lambda = 0.8155 \), the domain **U** is divided into 5 categories, which is the best classification.
Similarly, according to the Environmental Protection Bureau issued by Beijing Environmental Protection Bureau in 2014 and 2015 respectively, the air pollution status maps of 17 districts in Beijing were obtained, as shown in Figure 2 and figure 3.

![Figure 2. 2015 Dendrogram.](image1)

![Figure 3. 2014 Dendrogram.](image2)

4. Conclusions

Through the above analysis, the results of the classification of 17 districts in Beijing can be classified according to the air pollution situation. In 2016, the 17 districts of Beijing were divided into 5 categories, specific classification is as follows: the first category includes the Dongcheng district, Xicheng district, Chaoyang district, Haidian district, Fengtai district, Shijingshan district, Fangshan district, Daxing district, Tongzhou district, Beijing Economic and Technological Development Zone; The second category includes the Mentougou district and the Shunyi district; The third kind is Changping district; The fourth kind is Pinggu district; The fifth category includes Huairou district, Miyun district and Yanqing district, moreover, with the increase of categories, the annual average concentration of major pollutants decreases in turn, which indicates that the air quality is getting better and better.

Similarly, the concrete classification of 2015 is as follows: The first category includes Dongcheng district, Xicheng district, Chaoyang district, Haidian district, Fengtai district and Shijingshan district. The second category includes Fangshan district, Tongzhou district, Daxing district and Beijing Economic and Technological Development Zone; The third category includes Mentougou district, Shunyi district, Changping district, Miyun district and Yanqing district; The fourth kind is Pinggu district; The fifth category is Huairou.

The specific classification of 2014 can also be obtained as follows: The first category includes Dongcheng district, Xicheng district, Chaoyang district, Haidian district, Fengtai district and Shijingshan district and Beijing Economic and Technological Development Zone. The second category includes Tongzhou district and Daxing district; The third category is the Fangshan district; The fourth category includes the Mentougou district and the Shunyi district; The fifth category includes Changping district, Pinggu district, Huairou district, Miyun district and Yanqing district.

Based on the above results, Dongcheng district, Xicheng district, Chaoyang district, Haidian district, Fengtai district and Shijingshan district have been in the same category for three consecutive years, showing that atmospheric environmental conditions in the six districts of Beijing are very similar, and there is little difference. Changes in these four areas in Fangshan district, Tongzhou district, Daxing district, Beijing Economic and Technological Development Zone tends to the 6 urban areas, especially in 2016, the four suburban areas already belong to the same category with them, which shows that with the development of Beijing City, atmospheric environment in southern Beijing
has been affected by a lot. In the suburbs of Beijing, the Mentougou district and Shunyi district appear in the same category for three consecutive years. Huairou district, Miyun district and Yanqing district in the far suburbs appear in the same category in 2014 and 2016. In the suburbs, the Changping district and the Pinggu district in the far suburbs have changed over the three years, indicating that the air quality of the two regions has fluctuated greatly in the past three years.

Therefore, the government departments can rationally arrange the air test site through the classification of each district and the change of each district over time to control the air pollution.

References

[1] Zhao Lili 2011 China's atmospheric pollution and its control measures *Shanxi Architecture*. 37 (25) 194-196

[2] Li Mingming 2011 Analysis of the impact of air pollution on human beings and governance *Summary China's new technology and new products*. (21) 180

[3] Li Pei 2011 China's urban air pollution control integrated management countermeasures *Environment and sustainable development*. (5) 8-14

[4] Lin Changcheng, Wang Hong et al. 2010 Spatial and Temporal Distribution of Air Quality in Xiamen City and Its Relationship with Meteorological Conditions *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*. 39 (1) 79-83

[5] Wang Yan, Liu Peng, etc. 2011 Urban air quality assessment and forecasting model *Enterprise technology development*. 30 (2) 44-45

[6] CHAI Fa-he 2011 Discussion on Management Strategy of Urban Air Quality in China *Environmental and Sustainable Development*

[7] ZHU Xin-yuan, LIU Lai-fu, XU Ru-mei 1982 Application of fuzzy clustering analysis to classify river pollution in the southeastern suburbs of Beijing *Acta Ecologica Sinica*, 2(3) 201-209

[8] Wang Cuiyun, Hu Xueping 2014 Cluster analysis and its application in classification research of economic development level of provinces in China *Journal of Anqing Teachers College (Natural Science Edition)*. 20 (4) 36-41

[9] Ma Lidong 2003 Fuzzy clustering based network cooperative learning research and algorithm implementation *East China Normal University*

[10] Yan Huigang 2011 Study and implementation of intelligent recommendation system based on fuzzy clustering. *Dalian University of Technology*

[11] Yang Xin, Wang Hongbing, Xing Yun, Luo Weizhong 2006 China International Petroleum Investment Fuzzy Mathematics Comprehensive Evaluation Method *Journal of Tsinghua University (Science and Technology)*(06)

[12] Beijing Environmental Protection Bureau. 2016 *Beijing Environmental Status Bulletin*, June 2, 2017. http://www.bjepb.gov.cn/2016zt_jsxl/index.html