Real-time air quality assessment based on the EM-TOPSIS-PROMETHEE integrated assessment method

F Xu$^{1,2}$ and A S Chernikov$^1$

$^1$Bauman Moscow State Technical University (BMSTU), Moscow, Russia
$^2$Harbin Institute of Technology (HIT), Harbin, China

1171520201@stu.hit.edu.cn

Abstract. The goal of this article is to discuss the application of the EM-TOPSIS-PROMETHEE integrated assessment method in the real-time assessment of urban air quality and to compare it with other methods. By using the web crawler based on Selenium + Python we obtained the concentration of five air pollutant indicators in 150 cities. We imported the data into MATLAB, used the Entropy evaluation method (EM) to get the weight of each air pollutant indicator, then used TOPSIS-PROMETHEE to obtain the score of air quality in each city. This method is more comprehensive than AQI, and it can also be combined with IAQI. Finally, we discussed the advantages and disadvantages of this method.

1. Introduction

Air pollution is one of the main factors affecting the environment. The main reasons of air pollution include industrial emissions, forest fires, evaporation of salts from the oceans’ surface, and other natural and anthropogenic factors. Air pollution not only harms human health, but also significantly affects climate (decreased solar radiation, increased precipitation, and acid rain). In China the problem of air pollution is very serious.

The most commonly used simple method to assess air quality is AQI (air quality index). AQI is calculated as the maximum value of IAQI – air quality index of a single pollutant. It is unwise to use AQI for grading air quality, as it considers only one, although the most important pollutant [1].

The goal of this research is to discuss the application of EM-TOPSIS-PROMETHEE in assessing the air quality of cities [2]. The criterion of air quality assessment is the concentration of air pollutants. The lower the concentration of air pollutants, the better air quality is. In this article, factors that determine good quality are the concentrations of five air pollutants: PM10, PM2.5, CO, NO2, and SO2, they can be found on weather sites. Based on the concentration of these five air pollutants in 154 Chinese cities at 14:00 on July 13, 2020 (Beijing Time), we estimated the air quality.

2. Methods

We show the block diagram of EM-TOPSIS-PROMETHEE (Fig. 1).

The Entropy Method (EM) is used to determine the weight of each indicator (concentration of a pollutant). TOPSIS and PROMETHEE are used to determine the score of each city.

We assume that there are $m$ objects that we need to assess (cities) and $n$ indicators (air pollutants): $X = [x_{ij}], i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n$ (1)

where $X$ is the evaluation matrix, $x_{ij}$ – concentration of the $j$-th air pollutant in the $i$-th city.
First, we normalize the data using min-max normalization:

\[
x^*_ij = \begin{cases} 
\frac{x_{ij} - \min(x_{1j}, \ldots, x_{mj})}{\max(x_{1j}, \ldots, x_{mj}) - \min(x_{1j}, \ldots, x_{mj})} & \text{if } x_{1j}, \ldots, x_{mj} \text{ are positive indicators} \\
\frac{\max(x_{1j}, \ldots, x_{mj}) - x_{ij}}{\max(x_{1j}, \ldots, x_{mj}) - \min(x_{1j}, \ldots, x_{mj})} & \text{if } x_{1j}, \ldots, x_{mj} \text{ are negative indicators}
\end{cases}
\]

(2)

For positive \( j \), the more \( x_{ij} \), the better. Obviously, all the concentration of air pollutants are negative indicators. Thus, we got a normalized score matrix.

\[
X^* = [x^*_ij].
\]

(3)

In order to determine the weight of each indicator, we can use EM. Principal Component Analysis (PCA), Analytic hierarchy process (AHP), and EM are the three most commonly used methods for calculating weights. PCA can reduce the number of indicators. For example, if there are 14 indicators, then they can be condensed into 4 (also called factors or principal components) by using PCA. Since we have only five indicators (concentration of five pollutants), we do not need to use this method. AHP requires experts to compare the importance of various pollutants and it is very difficult to operate (mainly because the author is not an expert in this field and he is not familiar with any of them).
EM is based on the concept of entropy in thermodynamics. It describes the average size of information and can determine the weight of each indicator accordingly. The advantage of the model is that it is an objective assessment method. Compared with AHP, EM does not require human experts. The disadvantage of the model is that it evaluates the importance of the indicator only by values of $x_{ij}$, ignoring the physical meaning. In addition, EM cannot reduce the dimension of data (number of indicators) like PCA.

The algorithm of determining the weight of each indicator (air pollutant) $\omega_j$ is as follows:

1. Calculation of probability $p_{ij}$:
   \[ p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n \]  

2. Calculation of the entropy of the indicator $j$:
   \[ e_j = -k \sum_{i=1}^{m} p_{ij} \ln p_{ij} \]  

   where $k = \frac{1}{\ln m} > 0$, $e_j \geq 0$, for $p_{ij} = 0$ we assume $p_{ij}\ln p_{ij} = 0$.

3. Calculation of the redundancy of information entropy $d_j$:
   \[ d_j = 1 - e_j \]  

4. Calculation of the weight of each indicator $\omega_j$:
   \[ \omega_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \quad j = 1, 2, \ldots, n \]  

The concentrations of the main five air pollutants in different cities of China and indicator weights are presented in Table 1.

| Cities   | Weight $\omega_j$ | PM10, $\mu g/m^3$ | PM2.5, $\mu g/m^3$ | CO, mg/m³ | NO₂, $\mu g/m^3$ | SO₂, $\mu g/m^3$ |
|----------|------------------|-------------------|-------------------|------------|------------------|------------------|
| Beijing  | 0.2031           | 0.2533            | 0.1713            | 0.2916     | 0.0806           |
| Shanghai | 46               | 59                | 0.9               | 14         | 2                |
| Tianjin  | 21               | 23                | 0.8               | 33         | 5                |
| Chongqing| 35               | 46                | 1.0               | 18         | 7                |
| Liuzhou  | 16               | 10                | 0.7               | 26         | 5                |
| Haikou   | 32               | 14                | 0.5               | 6          | 4                |
| Sanya    | 16               | 4                 | 0.2               | 2          | 2                |

Then we use TOPSIS to assess the air quality of each city. This method does not have restrictions on the distribution and amount of data, and the calculation of data is simple. The advantage of the model is that it uses the relationship between objects (cities) and positive and negative ideal objects for valuing. The disadvantage of the model is that when there are few objects (cities) to be evaluated, the positive and negative ideal objects are not typical and representative, which will affect the credibility of the evaluation results. However, in the example in this article, there are many objects (over 150 cities).

Let us introduce the concept of the positive and negative ideal objects. The positive ideal objects select the best values of each air pollutant concentration in all cities. By contrast, the negative ideal objects select the worst values of each air pollutant concentration in all cities. For example, if we need to compare only Beijing and Shanghai (rather than 154 cities), and we can find the pollutant...
concentration of these two cities in Table 1, then the pollutant concentration of positive ideal object of these two cities is: PM10 – 21 µg/m³, PM2.5 – 23 µg/m³, CO – 0.8 µg/m³, NO2 – 14 µg/m³, and SO2 – 2 µg/m³.

First, we define the matrix Z:

\[
z_{ij} = \frac{x_{ij}^*}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}
\]

(8)

\[
Z = [z_{ij}]
\]

(9)

\[
Z_j^+ = \max\{z_{1j}^*, z_{2j}^*, \cdots, z_{mj}^*\}, j = 1,2,\cdots, n
\]

(10)

\[
Z_j^- = \min\{z_{1j}^*, z_{2j}^*, \cdots, z_{mj}^*\}, j = 1,2,\cdots, n
\]

(11)

where \(Z_j^+\) represents the value of the indicator \(j\) of a positive ideal object. \(Z_j^-\) represents the value of the indicator \(j\) of a negative ideal object.

To determine the geometric distance \(D_i^+\) between object \(i\) and the positive ideal object:

\[
D_i^+ = \sqrt{\sum_{j=1}^{n} \omega_j (Z_j^+ - z_{ij})^2}, \quad i = 1,2,\cdots, m
\]

(12)

where \(\omega_j\) is the weight of the indicator \(j\).

To determine the weighted geometric distance \(D_i^-\) between the object \(i\) and the negative ideal object:

\[
D_i^- = \sqrt{\sum_{j=1}^{n} \omega_j (Z_j^- - z_{ij})^2}, \quad i = 1,2,\cdots, m
\]

(13)

Score \(C_i\) of the object \(i\) is determined as follows:

\[
C_i = \frac{D_i^-}{D_i^- + D_i^+}
\]

(14)

Obviously, \(0 \leq C_i \leq 1\).

Next we use PROMETHEE II to assess the air quality of each city. The method compares the objects (cities) in pairs by comparing each indicator (concentration of air pollutants), then sums the results up and calculates the positive and negative flows for each object. The disadvantage of the model is that we have to determine the preference function and weight for each indicator. The operation is complex.

The algorithm inputs are shown in Table 2. Therefore, in addition to the score matrix \(X^*\) and the weight \(\omega_k\), it is also necessary to select the preference function \(P_k(x_{ik}, x_{jk})\) for each indicator (air pollutant). \(P_k(x_{ik}, x_{jk})\) represents the score of object \(i\) (the \(i\)-th city) compared to object \(j\) (the \(j\)-th city) when examining the \(k\)-th indicator \(C_k\) (concentration of the \(k\)-th air pollutant). Suppose \(d = x_{ik}^* - x_{jk}^*\). Define the preference function \(P_k(x_{ik}, x_{jk})\) as:

\[
P_k(x_{ik}, x_{jk}) = \phi_k(x_{ik}^* - x_{jk}^*) = \phi_k(d)
\]

(15)

The function \(\phi_k(d)\) must satisfy the restrictions:

1. \(\phi(d)\) – monotone non-decreasing function,
2. \(0 \leq \phi(d) \leq 1\), and when \(d \leq 0, \phi(d) = 0\).
Table 2. Inputs of PROMETHEE II.

| Indicators (air pollutants) | $C_1$ | $C_2$ | … | $C_k$ | … | $C_n$ |
|-----------------------------|-------|-------|---|-------|---|------|
| Weight                      | $w_1$ | $w_2$ | … | $w_k$ | … | $w_n$ |
| Preference functions        | $P_1$ | $P_2$ | … | $P_k$ | … | $P_n$ |
| Object (city) 1             | $x_{11}$ | $x_{12}$ | … | $x_{1k}$ | … | $x_{1n}$ |
| Object 2                    | $x_{21}$ | $x_{22}$ | … | $x_{2k}$ | … | $x_{2n}$ |
| Object i                    | $x_{i1}$ | $x_{i2}$ | … | $x_{ik}$ | … | $x_{in}$ |
| Object j                    | $x_{j1}$ | $x_{j2}$ | … | $x_{jk}$ | … | $x_{jn}$ |
| Object m                    | $x_{m1}$ | $x_{m2}$ | … | $x_{mk}$ | … | $x_{mn}$ |

Here we select the most commonly used function:

$$\varphi_k(d) = \begin{cases} 0 & d \leq 0 \\ 1 & d > 0 \end{cases}, \quad k = 1, 2, \ldots, n \quad (16)$$

In other words, we assume that if $x_{ik}^* \leq x_{jk}^*$, then $P_k(x_{ik}^*, x_{jk}^*) = 0$. If $x_{ik}^* > x_{jk}^*$, then $P_k(x_{ik}^*, x_{jk}^*) = 1$.

After determining the preference function $P_k(x_{ik}^*, x_{jk}^*)$, we can determine the positive and negative flows of each object (city) according to the following algorithm.

$$\Pi(i, j) = \sum_{k=1}^{n} w_k P_k(x_{ik}^*, x_{jk}^*) \quad (17)$$

$$\Phi_i^+ = \frac{1}{m-1} \sum_{j=1}^{m} \Pi(i, j) \quad (18)$$

$$\Phi_i^- = \frac{1}{m-1} \sum_{j=1}^{m} \Pi(j, i) \quad (19)$$

where $\Phi_i^+$ is the positive flow of the $i$-th object, and $\Phi_i^-$ is the negative flow of the $i$-th object. $m$ is the total number of objects ($m = 154$: number of cities in our research) and $n$ is the total number of indicators ($n = 5$: number of air pollutants in our research).

In the end, we combine the results of TOPSIS and PROMETHEE II. First, we use min-max normalizations to normalize $C_i$ and get $c_i$, normalize $\Phi_i^+$ and get $\phi_i$. Then define the final score as:

$$S_i = \zeta \phi_i^+ + \tau c_i \quad (20)$$

where $\zeta$ and $\tau$ depend on user’s preference. In our research, we assume that $\zeta = \tau = 0.5$.

3. Results

Using Python + Selenium, we obtained the concentrations of five air pollutants (PM10, PM2.5, CO, NO₂, and SO₂) at 14:23 on July 13, 2020 in 154 cities. Data can be found on the website China Weather (Table 1) [3]. The calculation results are shown in Table 3.

For a more convenient comparison, we can compare Table 3 with Table 4 showing concentration of the major air pollutants in the cities from Table 3. After min-max normalization, we obtain results presented in Table 5.
Table 3. Results obtained using different methods.

| City     | TOPSIS $C_i$ | PROMETHEE $\Phi_i$ | Score $S_i$ |
|----------|--------------|---------------------|-------------|
| 1 Sanya  | 0.97956      | 0.92382             | 1           |
| 2 Beihai | 0.96738      | 0.89091             | 0.98399     |
| 3 Yangjiang | 0.95185    | 0.78735             | 0.94393     |
| 4 Liaocheng | 0.94598    | 0.76789             | 0.94244     |
| 151 Shenyang | 0.47089    | -0.93662             | 0.20456     |
| 152 Anshan | 0.37448     | -0.95678             | 0.14047     |
| 153 Jiayiguan | 0.29458     | -0.95975             | 0.09092     |
| 154 Benxi | 0.16042      | -0.99438             | 0           |

Table 4. Concentrations of major air pollutants in some cities.

| City     | PM10, $\mu g/m^3$ | PM2.5, $\mu g/m^3$ | CO, mg/$m^3$ | NO$_2$, $\mu g/m^3$ | SO$_2$, $\mu g/m^3$ |
|----------|--------------------|--------------------|--------------|---------------------|---------------------|
| Sanya    | 16                 | 4                  | 0.2          | 2                   | 2                   |
| Beihai   | 11                 | 6                  | 0.3          | 4                   | 4                   |
| Yangjiang | 22             | 4                  | 0.4          | 3                   | 4                   |
| Liaocheng | 11             | 8                  | 0.4          | 5                   | 4                   |
| :        | :                 | :                  | :            | :                   | :                   |
| Shenyang | 88                 | 48                 | 1.0          | 33                  | 14                  |
| Anshan   | 92                 | 59                 | 2.0          | 32                  | 23                  |
| Jiayiguan | 158          | 40                 | 1.6          | 45                  | 65                  |
| Benxi    | 135                | 75                 | 3.2          | 49                  | 24                  |

Table 5. Results after min-max normalization.

| City     | $PM10$ | $PM2.5$ | $CO$ | $NO_2$ | $SO_2$ |
|----------|--------|--------|------|--------|--------|
| Sanya    | 0.95   | 1.00   | 1.00 | 1.00   | 1.00   |
| Beihai   | 0.99   | 0.97   | 0.97 | 0.96   | 0.97   |
| Yangjiang | 0.91  | 1.00   | 0.93 | 0.98   | 0.97   |
| Liaocheng | 0.99 | 0.94   | 0.93 | 0.94   | 0.97   |
| :        | :      | :      | :    | :      | :      |
| Shenyang | 0.47   | 0.38   | 0.73 | 0.34   | 0.81   |
| Anshan   | 0.44   | 0.23   | 0.40 | 0.36   | 0.67   |
| Jiayiguan | 0.00 | 0.49   | 0.53 | 0.09   | 0.00   |
| Benxi    | 0.15   | 0.00   | 0.00 | 0.00   | 0.65   |

The sort result obtained by this method are generally in line with expectations. However, when comparing some of these cities, it is difficult to immediately judge which city's air quality is better or worse. For example, $PM10$ and $SO_2$ of Benxi is better than Jiayiguan. $PM2.5$ and $CO$ of Benxi is worse than Jiayiguan. But the weight of the $SO_2$, calculated by the Entropy Method, is smaller, therefore Benxi's advantage in $SO_2$ is difficult to influence its position in the sort result.
Figure 2 presents a histogram of the distribution of scores. The horizontal axis represents the value of each indicator after min-max normalization. The vertical axis represents the number of cities within a specified range of values for each indicator. Figure 2 shows that after min-max normalization, the value of the $SO_2$ is mainly concentrated in the range from 0.9 to 1. EM assumes that such data contain less information, so the weight of $SO_2$ is less than those of other indicators. The distribution of other indicators is sparser, so their weights are greater. The weights of each indicator are presented in Table 6.

Table 6. Weights of indicators.

| Indicator | PM10  | PM2.5 | CO    | NO2   | SO2   |
|-----------|-------|-------|-------|-------|-------|
| Weight $\omega_j$ | 0.2031 | 0.2533 | 0.1713 | 0.2916 | 0.0806 |

To compare TOPSIS, PROMETHEE, and TOPSIS-PROMETHEE, we normalized the results of these three methods by using min-max normalization, and calculated their standard deviation (SD), average difference between the best score and other scores (DAB), and score difference between the first and second best scores (DFS). Below is the algorithm.

We assume that the scores after min-max normalization is $X'$.

$$X' = [x'_1 x'_2 ... x'_m]^T$$

(21)

where $m$ is the number of cities.

Then we obtain the following expressions for SD and DAB:
\[ SD = \sqrt{\sum_{i=1}^{m} (x'_i - \bar{x})^2} / (m - 1) \]
\[ \bar{x} = \frac{\sum_{i=1}^{m} x'_i}{m} \]
\[ DAB = \frac{\sum_{i=1}^{m} \max\{x'_1, x'_2, \ldots, x'_m\} - x'_i}{m - 1} \]  

(22)

Obviously, the larger SD, FAB, and DFS, the better the method is. From Table 7, we can conclude that PROMETHEE is the best method in this example. TOPSIS-PROMETHEE is the second best. TOPSIS is the worst method. However, when assessing other examples, PROMETHEE may behave wrong. Since TOPSIS-PROMETHEE combines the two methods, it usually has good reliability. In other articles which compared this method, TOPSIS-PROMETHEE is often the best assessment method. Therefore, the use of TOPSIS-PROMETHEE for air quality assessment is still reasonable.

**Table 7.** Statistical parameters of different methods.

| Parameter | TOPSIS | PROMETHEE | TOPSIS-PROMETHEE |
|-----------|--------|-----------|------------------|
| SD        | 0.1475 | 0.2244    | 0.1805           |
| DAB       | 0.2168 | 0.4848    | 0.3508           |
| DFS       | 0.0149 | 0.0172    | 0.0160           |

4. Conclusion

EM-TOPSIS-PROMETHEE is a reliable method for assessing air quality. Widely used method of assessing air quality (AQI) takes into account only one pollutant; therefore, the use of this method in scientific research (especially in sorting) is not very reasonable. The main advantage of EM-TOPSIS-PROMETHEE is complete combination of various methods, considering various air pollutants [2].

The disadvantages of this method are also obvious. Real-time air quality assessment depends on the data scanned by the web scanner. The data quality depends on the data source quality. Moreover, the method is very complex. We need to select the preference functions when using PROMETHEE.

For the further research we can:

- select other preference functions when using PROMETHEE to assess air quality. This article has used the simplest preference function.

- use IAQI as elements of \( X \) matrix. In this article, the concentrations of pollutants are used as elements of the \( X \) matrix, but the relationship between the harm of an air pollutant and its concentration is obviously not just proportional; therefore, it is inappropriate to use the concentrations of pollutants as elements of the \( X \) matrix. We can consider using the IAQI of each pollutant. However, we have not yet found a good data source for collecting IAQI. IAQI also depends on countries. American standard of AQI is deferent from Chinese standard [4].

**References**

[1] Agency E P 1999 *Guideline for Reporting of Daily Air Quality – Air Quality Index (AQI)* (Durham, North Carolina)

[2] Zhang L 2017 *Research on the Performance Evaluation of Renewable Energy Development* Dissertation (Wuhan, China University of Geosciences)

[3] China Weather 2020 http://www.weather.com.cn/air/

[4] Gao Q X, Liu J R, Li W T and Gao W K 2015 *Environ. Sci.* 36 1141