Traffic safety evaluation based on fuzzy clustering analysis: case studies in the city-cluster region of Beijing-Tianjin-Hebei

TingTing Gao*, XiaoFeng Liu, Rui Zhang
School of Auto and Transportation, Tian Jin University of Technology and Education, Tianjin 300222, China

Abstract. Traffic safety is a hot topic. Although traffic safety has improved, overall, for many developing countries, traffic safety fluctuates from year to year. Many studies have explored the issue of traffic safety evaluation from the perspective of the road, but little traffic safety evaluation is from the perspective of the city-cluster region. This paper adopts the fuzzy clustering theory to analyse the traffic safety level in the Beijing-Tianjin-Hebei region. Through calculation and analysis, it can be concluded that the traffic safety level in the city-cluster region of Beijing-Tianjin-Hebei differs among the three. The result shows that the overall traffic safety is improved in city-cluster of region of Beijing-Tianjin-Hebei, they are divided into two categories according to calculations. The traffic safety is relatively stable in Beijing and Hebei, but the traffic safety is more severe in Tianjin. The conclusions can provide a theoretical basis for the integrated traffic safety of Beijing-Tianjin-Hebei.

1. Introduction
With the rapid development of motorization and the increasing number of motor vehicles, traffic safety has become an important problem around the world, not only in China, but also in countries such as India, S. Africa, and Russia, to name but a few. Although traffic safety has improved in recent years, the mortality rate due to traffic accidents is still very high, necessitating the evaluation of traffic safety. In order to evaluate traffic safety efficiently, we must stand on one point of the system. So far, we have not found any research on traffic safety evaluation from the perspective of the city-cluster region. For this reason, this paper adopts the fuzzy clustering theory to analyse the traffic safety level in the Beijing-Tianjin-Hebei region. In 1949, Smeed, of the University of London, established the Smeed model based on the relationship between vehicle ownership, population and deaths from road traffic accidents. The Smeed model was perhaps the first to use this method to solve traffic safety problems. In 1964, Solomon used vehicle speed as the main method of road traffic safety assessment. This method played an important role in later research. Solomon calculated and analysed the traffic accident rates of different road sections and developed the initial vehicle speed - accident rate model [1]. In 2009, Elvik mentioned the Meta model in the “Dynamic Model of Speed and Road Traffic Safety” report. He published a meta-analysis and meta-regression analysis in 2014, and established a speed and traffic safety model based on meta-analysis [2-4]. Herman and Shen proposed the application of benchmarks in road traffic safety and used DEA for modeling analysis [5-6]. Ghazwan [7] established a comprehensive urban road safety evaluation index system in 2003. Based on the above research results, Chung broadened the impact of traffic safety and research results in various aspects of urban transportation in 2008[8].In order to make the urban transportation system more in line with urban development needs, Kononov incorporated urban traffic safety factors into the
mathematical model of urban transportation system planning. Kononov and Allery proposed the concept of urban traffic safety service level in the safety performance model. Based on SPF (Safety Performance Function), the process of urban transportation system service is summarized, and the safety service level framework is proposed [9].

This present study was aimed at traffic safety evaluation in the city-cluster region of Beijing-Tianjin-Hebei by use of fuzzy clustering analysis. First, we analysed the traffic safety data for the city-cluster region of Beijing-Tianjin-Hebei between 2012-2014 and then we used fuzzy clustering analysis according to the indicators of the different cities, and finally we compared the results after clustering.

2. Material and methods

We say that the mathematical method of classifying the things studied according to certain criteria is called cluster analysis. However, in many cases, classification has strong ambiguity, so the classification of such things must use fuzzy mathematics method, so the method using fuzzy mathematics we called fuzzy clustering analysis. This section we will present the fuzzy clustering model.

We set the domain \( U = \{ x_1, x_2, ..., x_n \} \) as the object to be classified, and each object is represented by \( m \) indicators [10]:

\[
x_i = \{ x_{i1}, x_{i2}, ..., x_{im} \} \quad (i = 1, 2, ..., n)
\]

Thus, the original data matrix is obtained as \( X = (x_{ij})_{nm} \).

In solving practical problems, in order to make different dimensions comparable, it is necessary to make some appropriate transformations, such as the transformation of the standard deviation:

\[
x_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad (i = 1, 2, ..., n; k = 1, 2, ..., m)
\]

Where,

\[
\bar{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_{ik}
\]

\[
s_k = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}
\]

Fuzzy similarity matrix

There are two general methods for determining the degree of similarity: the similarity coefficient method and the distance method. The similarity coefficient method includes the angle cosine method and the correlation coefficient method; the distance method is: Hamming distance, Euclidean distance, and Chebyshev distance. Here, each of the two methods is listed separately.

(1) Angle cosine method

\[
r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} \cdot x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2} \cdot \sqrt{\sum_{k=1}^{m} x_{jk}^2}}
\]

(2) Hamming distance

\[
d(x_i, x_j) = \sum_{k=1}^{m} |x_{ik} - x_{jk}|
\]

3. Theory/calculation

3.1. City-cluster region of Beijing-Tianjin-Hebei

The concept of the Beijing-Tianjin-Hebei mega region was developed by the capital economic circle, including the two municipalities directly under the Central Government of Beijing and Tianjin, Baoding, Tangshan, Langfang, Shijiazhuang, Qinhuanndao, Zhangjiakou, Chengde, Zhangzhou, Hengshui and Xingtai in Hebei. In order to realize the integration of Beijing-Tianjin-Hebei city-cluster
transportation, share development risks, and jointly establish the strategic goal of world city traffic safety, relevant theories are urgently needed to analyse the traffic safety level. The map of city-cluster region of Beijing-Tianjin-Hebei is shown as figure 1.

![Figure.1 the map of city-cluster region of Beijing-Tianjin-Hebei](image)

3.2. Research on the road traffic safety level in the city-cluster region of Beijing-Tianjin-Hebei

The traffic accident data for the city-cluster region of Beijing-Tianjin-Hebei can be found in the People's Republic of China road traffic accident statistical annual report from 2012 to 2014. They are sorted out as shown in a Table (see Table 1).

| City | Year | Number of Accidents | Number of Death | Number of Injured | Number of Vehicle | Number of Motorists |
|------|------|---------------------|----------------|------------------|-------------------|-------------------|
| Beijing | 2012 | 3196 | 918 | 3613 | 5145470 | 7547714 |
|      | 2013 | 3063 | 860 | 3359 | 5379733 | 8266010 |
|      | 2014 | 3196 | 851 | 3333 | 5535381 | 9077391 |
| Tianjin | 2012 | 3101 | 848 | 3429 | 2372234 | 3115363 |
|      | 2013 | 4313 | 836 | 4920 | 2765335 | 3379323 |
|      | 2014 | 5322 | 828 | 6171 | 2880460 | 3681850 |
| Hebei | 2012 | 5285 | 2503 | 4738 | 15394385 | 14160189 |
|      | 2013 | 5204 | 2501 | 4770 | 15631659 | 14915933 |
|      | 2014 | 5009 | 2499 | 4533 | 16574228 | 16152794 |

From the above direct data, in the three years from 2012 to 2014, the number of accidents in Beijing and Tianjin was similar, and the number of accidents in Hebei was relatively high. In comparison to the large area of Hebei Province, the area of Beijing and Tianjin is relatively small. By the end of 2014, there is 179200 kilometers of highways in Hebei, 21,849 kilometers of highways in Beijing and 16,110 kilometers of highways in Tianjin, combined with motor vehicle ownership and motor vehicles. The number of drivers can be seen as indicating a greater frequency of road traffic accidents in Tianjin. However, the above relative indicators are not comprehensive enough to evaluate the road traffic safety level of the three cities. There will be great deviations and lack of objectivity. Therefore, it is necessary to establish relative indicators in the evaluation process for the subsequent cluster analysis.

The three relative indicators are calculated in Section 3: the number of accidents per 10,000 cars, the number of deaths per 10,000 people, and the number of injuries per 10,000 people. The final data is shown in a Table (see Table 2).
Table 2. Fuzzy clustering indicators of road traffic safety level in Beijing-Tianjin-Hebei 2012-2014

| City    | year | 10000 car accidents | per10000 population deaths | per10000 population injuries |
|---------|------|----------------------|-----------------------------|------------------------------|
| Beijing | 2012 | 6.2113               | 1.2163                      | 4.7869                       |
|         | 2013 | 5.6936               | 1.0404                      | 4.0636                       |
|         | 2014 | 5.7738               | 0.9375                      | 3.6718                       |
|         | 2012 | 13.0721              | 2.7220                      | 11.0067                      |
| Tianjin | 2013 | 15.5967              | 2.4739                      | 14.5597                      |
|         | 2014 | 18.4762              | 2.2489                      | 16.7606                      |
|         | 2012 | 3.4331               | 1.7676                      | 3.3460                       |
| Hebei   | 2013 | 3.3291               | 1.6767                      | 3.1979                       |
|         | 2014 | 3.0222               | 1.5471                      | 2.8063                       |

4. The process of clustering

**Step 1**: We built a data matrix.

\[
X = \begin{pmatrix}
6.2113 & 1.2163 & 4.7869 \\
5.6936 & 1.0404 & 4.0636 \\
5.7738 & 0.9375 & 3.6718 \\
13.0721 & 2.7220 & 11.0067
\end{pmatrix}
\]

**Step 2**: We standardized the matrix.

It was standardized as follows:

\[
X = \begin{pmatrix}
0.3362 & 0.4468 & 0.2856 \\
0.3082 & 0.3822 & 0.2424 \\
0.3125 & 0.3444 & 0.2191 \\
0.7075 & 1 & 0.6567
\end{pmatrix}
\]

**Step 3**: We established a fuzzy relationship. That is, we established a fuzzy similarity matrix, which was obtained by using the maximum and minimum similarity law coefficients. It is as follows:

\[
R = \begin{pmatrix}
1 & 0.8729 & 0.8198 & 0.4520 & 0.4076 & 0.3781 & 0.6547 & 0.6607 & 0.6535 \\
0.8729 & 1 & 0.9302 & 0.3946 & 0.3558 & 0.3301 & 0.6397 & 0.6456 & 0.6374 \\
0.8198 & 0.9302 & 1 & 0.3705 & 0.3341 & 0.3100 & 0.6180 & 0.6234 & 0.6140 \\
0.4520 & 0.3946 & 0.3705 & 1 & 0.8379 & 0.7301 & 0.4377 & 0.4175 & 0.3804 \\
0.4076 & 0.3558 & 0.3341 & 0.8379 & 1 & 0.8728 & 0.3947 & 0.3765 & 0.3430 \\
0.3781 & 0.3301 & 0.3100 & 0.7301 & 0.8728 & 1 & 0.3662 & 0.3492 & 0.3182 \\
0.6547 & 0.6397 & 0.6180 & 0.4377 & 0.3947 & 0.3662 & 1 & 0.9538 & 0.8692 \\
0.6607 & 0.6456 & 0.6234 & 0.4175 & 0.3765 & 0.3492 & 0.9538 & 1 & 0.9112 \\
0.6535 & 0.6374 & 0.6140 & 0.3804 & 0.3430 & 0.3182 & 0.8692 & 0.9112 & 1
\end{pmatrix}
\]

**Step 4**: cluster
The transitive closure $R^t$ of the fuzzy similar matrix is obtained by the squared method, and then the $\lambda$-cut relationship is obtained, and the $\lambda$ is reduced from 1 to 0. When $\lambda$ changed, a dynamic classification result is given. It is as follows: The matrix in Table 3 was as follows:

$$
R^t = R^4 = \begin{pmatrix}
1 & 0.8729 & 0.8729 & 0.4520 & 0.4520 & 0.4520 & 0.6607 & 0.6607 & 0.6607 \\
0.8729 & 1 & 0.9302 & 0.4520 & 0.4520 & 0.4520 & 0.6607 & 0.6607 & 0.6374 \\
0.8729 & 0.9302 & 1 & 0.4520 & 0.4520 & 0.4520 & 0.6607 & 0.6607 & 0.6140 \\
0.4520 & 0.4520 & 0.4520 & 1 & 0.8379 & 0.8379 & 0.4520 & 0.4520 & 0.4520 \\
0.4520 & 0.4520 & 0.4520 & 0.8379 & 1 & 0.8728 & 0.4520 & 0.4520 & 0.4520 \\
0.4520 & 0.4520 & 0.4520 & 0.8379 & 0.8728 & 1 & 0.4520 & 0.4520 & 0.4520 \\
0.6607 & 0.6607 & 0.6607 & 0.4520 & 0.4520 & 0.4520 & 1 & 0.9538 & 0.9112 \\
0.6607 & 0.6607 & 0.6607 & 0.4520 & 0.4520 & 0.4520 & 0.9538 & 1 & 0.9112 \\
0.6607 & 0.6607 & 0.6607 & 0.4520 & 0.4520 & 0.4520 & 0.9112 & 0.9112 & 1
\end{pmatrix}
$$

When $\lambda=1$, it is divided into nine categories $\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.9538$, it is divided into eight categories $\{x_1,x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.9302$, it is divided into seven categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.9112$, it is divided into six categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.8729$, it is divided into five categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.8728$, it is divided into four categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.8379$, it is divided into three categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.6607$, it is divided into two categories $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

When $\lambda=0.4520$, it is divided into one category $\{x_1,x_2,x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}$.

From the above cluster analysis and Figure 3, it can be seen that we can cluster two categories. The traffic safety levels in Beijing and Hebei are close to one category, and Tianjin is classified as another. The traffic safety level in Beijing in 2013 was similar to that in 2014. The road traffic safety level in 2013 was improved compared to 2012, but the road traffic safety level in 2014 was worse than in 2013. The road traffic safety level in 2012-2013 in Hebei Province is relatively close and tends to be stable. The number of accidents per 10,000 cars in 2014 and the death rate and injury rate per 10,000 people are low. Therefore, the road traffic safety level in Hebei Province is gradually getting better, and the growth rate in 2014 is larger. The road traffic safety conditions in Tianjin in 2013 and 2014 are relatively close, more stable than in 2012. From the number of accidents per 10,000 cars, and deaths and injuries per 10,000 people, the figures for 2013 and 2014 are better than those for 2012. However, they are high, and gradually rising, with a slower rate of increase in 2014; therefore, Tianjin's overall traffic safety is a serious level year after year.
5. Discussion and conclusions
In comparing the road traffic safety in Beijing, Tianjin and Hebei, the number of road traffic accidents in Hebei is the lowest, followed by Beijing and Tianjin. Combined with the other two relative indicators, the road traffic safety level in Tianjin is relatively high. The road traffic safety level in Beijing and Hebei is relatively close, but the level of road traffic safety in Hebei is higher. At the same time, the road traffic safety evaluation in Hebei is getting better year by year, compared with Beijing, which is more stable, and Tianjin, which is getting worse and worse. In addition, in the future work, we will further explore the analysis of the traffic safety of the city-cluster region from the inherent point of view, and strive to achieve regional coordinated security in essence.

Acknowledgment
Research supported in part by Tianjin Science and Technology Plan Project (19YFSLQY00010), the National Natural Science Foundation of China (61503284) and the project of Tianjin University of Technology and education (KJ1728)

References
[1] Nain, R.: ‘Road safety model: some fundamental ideas’, Transportation Research, 1963
[2] Elvik, R.: ‘Area-wide urban traffic calming schemes: a meta-analysis of safety effects’, Accident Analysis and Prevention, 2001, 33(2), pp.327-336
[3] Elvik, Rune.: ‘Dimensions of road safety problems and their measurement’, Accident Analysis and Prevention, 2008, 40, pp.1200-1210
[4] Elvik, Rune.: ‘Why some road safety problems are more difficult to solve than others’, Accident Analysis and Prevention, 2010, 42, pp.1089-1096
[5] Shen, Yongjun, Hermans, Elke, Brijs, Tom.: ‘Road safety risk evaluation and target setting using data envelope analysis and its extensions’, Accident Analysis and Prevention, 2012, 48, pp.430-441
[6] Shen, Yongjun, Hermans, Elke, Bao, Qiong.: ‘Serious injuries: an additional indicator to fatalities for road safety benchmarking’, Traffic Injury Prevention, 2015, 16, pp.246-253
[7] Ghazwan, al-Haji.: ‘Road safety development index (RSDI) [D]’, Sweden: Linköping University, 2003
[8] Lee, J.Y., Chung, J.H., Son, B.: ‘Analysis of accident size for Korean highway using structural equation models’, Accident Analysis and Prevention, 2008, 40, pp.1955-1963
[9] Kononov, J., Allery, B. K.: ‘Explicit consideration of safety in transportation planning and project scooping[C]’, Transportation Research Board 2004 Annual Meeting, 2004, 10, pp.201-223
[10] Zhu, J.Y.: ‘Intelligent system non-classical mathematical method’, M., Huazhong University of Science and Technology Press: Wuhan, CN, 2001, pp.99-133