Application of Data Mining Technology in Analyzing the Impact of Mileage on Unqualified Rate of Vehicle Emission Inspection

Van Tuu Nguyen\textsuperscript{1,2,*} and Jiangwei Chu\textsuperscript{1,*}

\textsuperscript{1}Traffic College, Northeast Forestry University, Harbin, China
\textsuperscript{2}Vietnam National University of Forestry, Hanoi, Vietnam

*Corresponding author e-mail: cjw_62@163.com, nvtuu@nefu.edu.cn

Abstract. The study uses data mining method to analyze the impact of mileage on the unqualified rate of vehicle emission inspection. The result of data mining shows that with the increase of mileage, the unqualified rate of commercial vehicles and non-commercial vehicles increased. Specifically, in the range of 0-200,000 km, the emission inspection unqualified rate of commercial vehicles is twice that of non-commercial. In the range of 200,000-600,000 km, the emission inspection unqualified rate of commercial vehicles is always higher than that of non-commercial vehicles. This result shows that mileage has a significant impact on vehicle emission status, and the unqualified rate of vehicle emission inspection is proportional to mileage mission inspection data.

1. Introduction
In recent years, with the rapid economic growth of many countries in the world, social life has changed day by day, automobiles from being seen as a symbol of the wealth of society, now, cars have become very popular in most cities in the world. Along with that, the rapid development of the automobile industry has made the number of cars in use worldwide increasing rapidly. However, cars are also the source of fuel consumption and automobile emissions, which is one of the leading causes of air pollution in cities with high motor vehicle density such as Beijing, Hong Kong, Delhi, et al. [1-7].

To solve the problem of urban air pollution caused by automobile exhaust pollutants, scholars at home and abroad have carried out various studies. While discussing the harmfulness of automobile exhaust pollutants, You Dalong et al. put forward Macro-countermeasures for controlling automobile exhaust pollutants, such as optimizing automobile design, improving fuel quality, optimizing road conditions, encouraging the development and sale of new energy vehicles, and perfecting automobile exhaust laws and regulations [8]. Governments, automobile manufacturing industries and relevant departments in the world have invested a lot of workforces, material resources, technology and capital to solve the environmental pollution caused by automobile exhaust emissions, but the results are not satisfactory. There are two main reasons: one is the rapid growth of car ownership, another is that with the increase of the service life of the car, its parts are worn to varying degrees, which makes the overall technical state of the car worse gradually, and it needs to be tested and controlled or maintained before it can continue to operate. Vehicle inspection data sets include vehicle technical
indicators, use status information and vehicle exhaust pollutant data, which can be accumulated continuously to form a large amount of inspection data.

In this paper, the decision tree algorithm in data mining technology is used to mine and analyze the emission inspection data of passenger cars, and to explore the impact of the mileage on vehicle emission status and its law, which can provide reference for environmental protection departments to solve the problem of vehicle exhaust pollutants on urban air pollution.

2. Data and methods

2.1. Data
The data collected in this paper are from several motor vehicle inspection centers in a city and include nearly 100,000 in-use vehicle emission inspection data. The emission inspection data set includes the necessary information of the vehicle, such as inspection code (inspection ID), registration date, reference quality, applying character (commercial vehicles, non-commercial vehicles), mileage and other information, as well as the relevant information of test results and inspection conclusions (pass: 0; not pass: 1).

2.2. Methods
This study takes advantage of data mining techniques with decision tree algorithms to build a classification model for exploring the relationship between useful lives and applying character of vehicles to Vehicle exhaust concentration. Based on the data mining results, analyze the relationship between useful life and applying character to the average concentration of emissions.

3. Results and analysis

3.1. Basic statistical results of vehicle emission inspection data
The primary research object of this paper is whether the emission status of the light vehicle is qualified based on the ASM inspection method. Therefore, the collected data are preliminarily classified and select the needed data for the data mining process by classification. The rest of the data was ignored or excluded from consideration.

According to the inspection method, 100,000 in-use vehicle emission inspection data are classified, and the results are shown in Table 1.

| No. | Inspection method               | Number of vehicles | Vehicle ratio (%) |
|-----|--------------------------------|--------------------|-------------------|
| 1   | ASM steady state operation method | 57997              | 58.00             |
| 2   | Free acceleration method        | 23937              | 23.94             |
| 3   | Double idle method              | 14519              | 14.52             |
| 4   | Other methods                   | 3547               | 3.54              |
|     | **Total**                       | **100000**         | **100**           |

From the classification results of the above table, in the 100,000 data has 57,997 (58.00%) belong to the ASM inspection method, 23,937 belong to free acceleration method, 14,519 belong to double idle speed method and 3547 belong to other inspection methods (including visual measurement, loading, and deceleration method, Lingman blackness determination). In this paper, nearly 60,000 data based on ASM inspection method has been used as a data source for data mining in this study.

3.2. The impact of mileage on emission inspection unqualified rate of vehicles
Mileage is an important index to evaluate the intensity of vehicles use, and the frequency of use of the vehicles can be known through mileage, which is also a primary reference basis for buying used vehicles. Two cars registered in the same year, the more mileage indicates the higher the strength of
the car; in contrast, the lower mileage indicates that the vehicle is less intense to use. Another in the process of using the car, with the increase in mileage, the vehicle's technology, such as the degree of engine deterioration, emission of pollutant indicators and other indicators have corresponding changes. In this paper, the data mining method is used to carry out data mining on the unqualified rate of vehicle emission inspection by using the historical emission inspection data for two kinds of vehicles with different properties of commercial vehicles and non-commercial vehicles. The following is the result and discussion of the data mining process.

Through quick browsing and statistical methods, in the test data, nearly 55% of the vehicle test samples are concentrated in the mileage range (0-50,000) km, and about 32% of the vehicle test samples are concentrated in the mileage range (50,000-200,000) km. Therefore, the mileage (L) is subdivided into four intervals for data mining and analysis. Mileage is divided as shown in table 2.

Table 2. Classification of mileage.

| No. | Mileage (L) (104km) | Category |
|-----|---------------------|----------|
| 1   | L≤5                | A        |
| 2   | 5<L≤20             | B        |
| 3   | 20<L≤40            | C        |
| 4   | 40<L≤60            | D        |

3.2.1. The impact of mileage on the emission inspection unqualified rate of commercial vehicles. After classifying the mileage, the corresponding data mining model is established and processed. Through browsing the mining model viewer, all the mining result about the impact of mileage on the unqualified rate of commercial vehicles emission inspection is obtained and shown in table 4.

Table 3. The emission inspection unqualified rate of commercial vehicles based on mileage.

| No. | Mileage (L) (104km) | Test sample | Category | Unqualified number | Unqualified rate (%) |
|-----|---------------------|-------------|----------|--------------------|----------------------|
| 1   | L≤5                | 2333        | A        | 192                | 8.29                 |
| 2   | 5<L≤20             | 1703        | B        | 213                | 12.58                |
| 3   | 20<L≤40            | 1692        | C        | 242                | 14.37                |
| 4   | 40<L≤60            | 2925        | D        | 477                | 16.34                |
| Total|                    | 8653        |          | 1124               | 13.06                |

As can be seen from the data in table 3, the mining model processes a total of 8,653 emission test results data on commercial vehicles. The excavation results showed that 1124 test results were unqualified, accounting for 13.06% of the proportion. From category A to Category D, with the increase of mileage, the unqualified rate of commercial vehicles shows a significant increase, which shows that mileage is closely related to the emission status of the vehicle.

For more clearly describe the influence of mileage on the vehicle emission status. This article charts used some of the data in table 3 to get the chart shown in figure 1.
As can be seen from figure 1, the mileage is proportional to the unqualified rate of commercial vehicles emission inspection. With the increase in mileage, the rate shows a significant upward trend.

### 3.2.2. The impact of mileage on the emission inspection unqualified rate of non-commercial vehicles.

A data mining model based on the mileage and the unqualified rate of non-commercial vehicles has been established, through processing the model to get the mining results as shown in table 4.

| No. | Mileage (L) (10^4 km) | Category | Test sample | Sample ratio (%) | Unqualified number | Unqualified rate (%) |
|-----|----------------------|----------|-------------|------------------|-------------------|---------------------|
| 1   | L ≤ 5                | A        | 25600       | 60.66            | 823               | 3.22                |
| 2   | 5 < L ≤ 20           | B        | 14907       | 35.33            | 1635              | 10.98               |
| 3   | 20 < L ≤ 40          | C        | 1532        | 3.63             | 259               | 16.97               |
| 4   | 40 < L ≤ 60          | D        | 160         | 0.38             | 26                | 16.91               |
|     | Total                |          | 42199       |                  | 2743              | 6.52                |

As can be seen from the data in table 4, the mining model has processed a total of 42,199 emissions test results for non-commercial vehicles, of which 2,743 have failed to qualify, accounting for 6.52% of the proportion. Further browsing of the excavation results also found that in the mileage Category A and Category B, the mileage of 50,000 km and mileage in (50,000-200,000) km of non-commercial vehicles accounted for the majority (95.99%) of the total tested vehicles, the remaining two categories of mileage of the non-commercial vehicle accounted for less than 5% of the proportion. This data mining result indicates that the mileage of the non-commercial vehicle in the table is mainly within 200,000 km. From the data of non-commercial vehicle emission inspection unqualified rate, from Class A to class C mileage stage, with the increase of mileage, non-commercial vehicle emission inspection unqualified rate is on the rise. In the mileage phase of Class D, the non-commercial vehicle detection unqualified rate is slightly lower than the previous stage, but the difference is small, and in this mileage range, the number of samples is minimal (only 0.38%). Therefore, it is not accurate to reflect the real mileage of the non-commercial vehicle detection of the irregular effect of the law.
To more intuitively show the impact of mileage on vehicle emission status. In this paper, some of the data in Table 4 are graphed to get the graph as shown in Figure 2.

![Graph showing the relationship between Mileage and Unqualified Rate of Non-Commercial vehicles Emission Inspection.](image)

**Figure 2.** The relationship between Mileage and Unqualified Rate of Non-Commercial vehicles Emission Inspection.

From Figure 2, we can see the emission inspection unqualified rate of non-commercial vehicles increases with the increase in mileage. In the 200,000-400,000 km range, the unqualified rate of non-commercial vehicles reached a peak of 16.97%. The results again show that the more frequent use and run with more miles, the higher probability there is of a non-commercial vehicle failing the emission inspection.

3.2.3. Comparison between the unqualified rate of commercial and non-commercial vehicles. Now, let's take a look at the difference between the mileage impact on the commercial vehicles and the non-commercial vehicles. There are apparent differences in the mileage at that stage. Using the unqualified rate data in Tables 3 and 4, the unqualified rate comparison chart of operating and non-commercial vehicles based on mileage as shown in Figure 3.

As can be seen from the chart in Figure 3, in the two mileage ranges (0-50,000) km and (50,000-200,000) km, the emission inspection unqualified rate of the commercial vehicle is much higher than that of the non-commercial vehicles. Especially in the mileage range of (0-50,000) km, the unqualified rate of the commercial vehicles is 2.57 times higher than that of the non-commercial vehicle. However, in the two mileage ranges greater than 200,000 km, the unqualified rate of non-commercial vehicles has a significantly higher trend than the unqualified rate of commercial vehicles, because the number of non-commercial vehicles test samples is too small, so it is difficult to reflect the real unqualified rate situation. In general, the emission inspection unqualified rate of the commercial vehicles and the non-commercial vehicles is different, in the 0-200,000 km stage of mileage, the non-commercial vehicle unqualified rate is lower than the commercial vehicle unqualified rate, which indicates that the emission status of non-commercial vehicles is better than that of commercial vehicles.
Figure 3. Comparisons of Unqualified Rate of Commercial vehicles and Non-commercial vehicles Based on Mileage.

4. Conclusion
The emission inspection unqualified rate of the commercial vehicles twice times the unqualified rate of the non-commercial vehicles. Specifically, the unqualified rate of commercial vehicles is 13.06%, while the rate of non-commercial vehicles is only 6.52%. The vast majority of non-commercial vehicles travel within 0~200,000 km, of which 60.66% of non-commercial vehicles are within the mileage range of 0~50,000 km, and 35.33% of non-commercial vehicles are within mileage of 50,000~200,000 km. Less than 5% of non-commercial vehicles are traveling at a mileage greater than 200,000 km. Within the mileage range of 200,000 km, the emission inspection unqualified rate of the commercial vehicles is always higher than that of non-commercial vehicles. In the 200,000~600,000 km of mileage, the emission inspection unqualified rate of non-commercial vehicles is slightly higher than that of commercial vehicles, but because of the proportion of non-commercial vehicles with more than 200,000 km of mileage are tiny, it is difficult to reflect the emission status of vehicles correctly. With the increase of mileage, the emission inspection unqualified rate of commercial vehicles and non-commercial vehicles are on the rise. It shows that mileage has a significant impact on vehicle emission status; the emission inspection unqualified rate is proportional to the mileage.

Acknowledgments
This study was supported by the China Scholarship Council (CSC) fund and the Vietnamese Ministry of Education (No.911 Project). Here, I sincerely thank the Chinese government and the Vietnamese Ministry of Education.

References
[1] Zhu, K., et al., Health financing and integration of urban and rural residents’ basic medical insurance systems in China. International Journal for Equity in Health, 2017. 16(1): p. 194.

[2] Zhang, Y., et al., Development and Application of Urban High Temporal-Spatial Resolution Vehicle Emission Inventory Model and Decision Support System. Environmental Modeling & Assessment, 2017. 22(5): p. 445-458.

[3] Uddin, W., Mobile and Area Sources of Greenhouse Gases and Abatement Strategies, in
Handbook of Climate Change Mitigation and Adaptation, W.-Y. Chen, T. Suzuki, and M. Lackner, Editors. 2017, Springer International Publishing: Cham. p. 1657-1721.

[4] Malik, L. and G. Tiwari, Assessment of interstate freight vehicle characteristics and impact of future emission and fuel economy standards on their emissions in India. Energy Policy, 2017. 108: p. 121-133.

[5] Guevara, M., et al., An emission processing system for air quality modelling in the Mexico City metropolitan area: Evaluation and comparison of the MOBILE6. 2-Mexico and MOVES-Mexico traffic emissions. Science of The Total Environment, 2017. 584: p. 882-900.

[6] Alvim, D.S., et al., Main ozone-forming VOCs in the city of Sao Paulo: observations, modelling and impacts. Air Quality, Atmosphere & Health, 2017. 10(4): p. 421-435.

[7] Huo, H., et al., Vehicular air pollutant emissions in China: evaluation of past control policies and future perspectives. Mitigation and Adaptation Strategies for Global Change, 2015. 20(5): p. 719-733.

[8] You Dalong, Hu Tao, Zheng Fang, et al. Study on Preventing and Controlling Motor Vehicle Exhaust. Environmental Science and Management. 2015, (01): 102-105.