OBD Quality Evaluation System and Control

Hongshuai Xue, Zhili Liu, Xiangyang Yu and Xinmei Wei

Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China.
Email: 15733168529@163.com

Abstract. Based on the characteristics of OBD data and its role in detecting automobile exhaust emissions, the data from January to August was selected from 368 vehicles of 12 companies to establish an OBD data quality evaluation system. Based on the results of quality evaluation, a method of controlling variables was adopted in order to compare the factors that affect data quality and quantitatively analyse the relationship between different factors and data quality results, which can provide technical support for law enforcement supervision and policy development.

1. Introduction

With the increase in the number of resident cars in China, vehicle energy consumption emissions have become the main cause of environmental pollution[1]. OBD data has an important role in effectively monitoring vehicle exhaust emissions[2]. The full name of OBD is On Board Diagnostics, which is an extended detection system for vehicle fault diagnosis[3]. OBD can be used to analyse the oxygen sensor and engine misfire[4] and get working conditions of related components. In addition, it can improve the level of vehicle maintenance technology[5] and use remote OBD monitoring and diagnosis system to establish data interaction between PC server and on-board ECU, using GPRS / GSM as the wireless transmission medium to remotely monitor the vehicle[6]. It can also study the characteristics and rules of travel behavior of cars[7] in order to formulate relevant traffic control methods. OBD data should be collected and transmitted before application.

In order to ensure the integrity and accuracy of vehicle status, OBD monitors systems and components including the engine, catalytic converter and oxygen sensor[8] and collects relevant data online. According to the OBD-II protocol, real time periodic collection of vehicle status information is required and one-time collection method[9] can be used for gathering vehicle identification information and fault information. OBD can determine the vehicle data collection list and collection cycle and generate vehicle status query data frames, then analyse vehicle status real-time data through vehicle status response data frames.

OBD data transmission is divided into wired and wireless transmission. When the local online data terminal selects the PND device based on winCE system, the data transmission method selects the USB wired transmission method[9] and the OBD data collector and vehicle perform asc-II code encoding to complete the wired transmission of vehicle status data; when the data is transmitted to the remote server, the transmission method uses GPRS wireless transmission method, which requires the GSM mobile communication module and the SIM card with GPRS service to be installed in the OBD data collector in order to establish a connection between data and a remote server. With the development of social networking, OBD data can also interact with the Bluetooth module of the Android platform[10]. The transmission format adopts a lightweight data exchange format JSON[11].

OBD data has the characteristics of comprehensive diagnosis, real-time monitoring, fault reminder, and auxiliary management, so it is widely used in the fields of automobile detection[12], emission...
control[13] and intelligent transportation system[14]. The method of line data flow diagnosis[15] is used to diagnose the cause of vehicle failure. This article makes a comparative analysis of the factors that affect the quality of OBD data and quantitatively analyses the relationship between the influencing factors and the rating results, providing a basis for the formulation of energy conservation and emission reduction methods.

2. Basic Data Quality of OBD
Based on the current statistical analysis of OBD data, OBD data often has three types of problems: data missing, abnormal data characteristics and logical abnormalities. Typical problems are shown in table 1.

| month | Negative mileage | The speed mostly shows 65.535, which is suspected to be the system return value | A small range of negative values of NOx downstream sensors | MIL state and OBD diagnostic protocol fields missing |
|-------|------------------|--------------------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------------|
| 1     | 0.027            | 0.156                                                                            | 0.259                                                     | 0.296                                             |
| 2     | 0.028            | 0.470                                                                            | 0.397                                                     | 0.364                                             |
| 3     | 0.028            | 0.328                                                                            | 0.368                                                     | 0.322                                             |
| 4     | 0.038            | 0.415                                                                            | 0.320                                                     | 0.342                                             |
| 5     | 0.029            | 0.000                                                                            | 0.409                                                     | 0.281                                             |
| 6     | 0.024            | 0.000                                                                            | 0.427                                                     | 0.321                                             |
| 7     | 0.022            | 0.106                                                                            | 0.450                                                     | 0.314                                             |
| 8     | 0.037            | 0.271                                                                            | 0.435                                                     | 0.442                                             |

The problem that the accumulated mileage is negative has the similar frequency of occurrence among months and the average proportion of each month is 0.029. The system return value problem is obviously different among months. As for the rest problems, there are few differences among months. This article establishes a data quality assessment system and a rating method from four perspectives of completeness, continuity, accuracy and logic to score the monitored vehicle data. When the score of a car is from 80 to 100, it is a first-grade car; when the score is from 60 to 80, it is a second-grade car; when the score is from 0 to 60, it is a third-grade car. The lowest score of 368 cars is 51, and the highest score is 95. The distribution of scores and ratings is shown in figure 1 and figure 2.

3. Comparative Analysis of OBD Data Based on Sample Data
This section conducts a horizontal comparative analysis of the data of different months, different brands and different indicators, and adopts the method of controlling variables to analyse the factors that affect the quality of the data.
3.1. Impact of Different Months on Data Quality Ratings

Sections should be numbered with a dot following the number and then separated by a single space: Study the relationship between data ratings and months. Statistical results are shown in Table 2.

**Table 2. Data quality ratings in different months**

| month | Vehicle number | First grade | Vehicle number | proportion | Second grade | Vehicle number | proportion | Third grade | Vehicle number | proportion |
|-------|----------------|-------------|----------------|------------|--------------|----------------|------------|-------------|----------------|------------|
| 1     | 328            | 57          | 0.174          | 248        | 0.756        | 23             | 0.070      |
| 2     | 247            | 107         | 0.433          | 120        | 0.486        | 20             | 0.081      |
| 3     | 323            | 60          | 0.186          | 244        | 0.755        | 19             | 0.059      |
| 4     | 366            | 116         | 0.317          | 241        | 0.659        | 9              | 0.025      |
| 5     | 281            | 0           | 0.000          | 266        | 0.947        | 15             | 0.053      |
| 6     | 330            | 0           | 0.000          | 293        | 0.888        | 37             | 0.112      |
| 7     | 322            | 41          | 0.127          | 267        | 0.829        | 14             | 0.044      |
| 8     | 269            | 143         | 0.532          | 124        | 0.461        | 2              | 0.007      |

In different months, the proportion of first-grade vehicles is quite different. The least proportion is 0 in May and June and the highest is 0.532 in August, but the difference in the proportion of grades above second-grade is not obvious among months.

Taking the data of Jing B10164 vehicle as an example, the two columns of the month and the corresponding score are imported into SPSS and the correlation between the month and the score is analysed. And the result is shown in Figure 3.

![Figure 3. Analysis result of correlation](image)

Similarly, the results of Jing B14739 and Jing AEV216 are shown in Figure 4 and Figure 5.
From the above results, it can be seen that the Sig values are all larger than 0.05. So it can be concluded that in the sample data, the quality score of the same vehicle data is not strongly correlated with the month. In short, different months don’t have obvious effects on the data quality score.

3.2. Impact of Different Brands on Data Quality Ratings

Study the relationship between data ratings and brands. The statistical results are shown in Table 3.

**Table 3.** Data quality ratings of different brands

| Vehicle Brands | First grade |   | Second grade |   | Third grade |   |
|----------------|-------------|---|--------------|---|-------------|---|
|                | Vehicle number | proportion | Vehicle number | proportion | Vehicle number | proportion |
| Futian         | 164          | 0.390       | 95             | 0.579       | 5             | 0.031       |
| Ouman          | 58           | 0.293       | 41             | 0.707       | 0             | 0.000       |
| Beifang        | 43           | 0.116       | 37             | 0.861       | 1             | 0.023       |
| Dongfeng       | 23           | 0.087       | 21             | 0.913       | 0             | 0.000       |
| Jianghuai      | 20           | 0.450       | 11             | 0.550       | 0             | 0.000       |
| Jinlv          | 19           | 0.947       | 1              | 0.053       | 0             | 0.000       |
| Qingnian       | 19           | 0.158       | 15             | 0.790       | 1             | 0.053       |
| Jinlong        | 10           | 0.600       | 4              | 0.400       | 0             | 0.000       |
| Jiefang        | 7            | 0.000       | 7              | 1.000       | 0             | 0.000       |
| Kangfei        | 3            | 1.000       | 0              | 0.000       | 0             | 0.000       |
| Total          | 366          | 0.347       | 232            | 0.634       | 7             | 0.019       |

Different vehicle brands have a large difference in the proportion of first-grade vehicles. Considering vehicle brands with more than 10 vehicles, Jinlv is the vehicle brand with the highest proportion of first-grade vehicles and the difference in the proportion of grades above second-grade is not obvious among brands.

3.3. Impact of Different Evaluation Indicators on Data Quality Ratings

Taking the data of 368 vehicles in May as examples, then calculate the integrity rate of the MIL state and the intake air volume index, the accuracy rate of the speed, the net output torque, the friction torque, the SCR upstream and downstream NOx sensor output values, atmospheric pressure, engine fuel flow, engine speed, SCR inlet and outlet temperature, DPF pressure difference, engine coolant temperature and tank level and the continuity rate of time. Then take these indicators as independent variables, take the quality score as the dependent variable and use a forced entry strategy to establish a multiple linear regression model. The significance level is 0.05. The model results are as follows:
It can be seen from figure 6 that the square value of R is close to 1, indicating that the goodness of fit is great. The forced entry strategy excludes three indicators of engine speed, SCR upstream NOx sensor output and SCR inlet temperature, indicating that these three indicators have no significant effects on the quality score.

| Model       | Input Beta | t    | Sig | Partial correlation | Collinearity statistics |
|-------------|------------|------|-----|---------------------|-------------------------|
| engine speed|            |      |     |                     | 0                       |
| SCR upstream NOx |      |      |     |                     | 0                       |
| SCR inlet temperature |      |      |     |                     | 0                       |

Figure 6. R-squared test and excluded variables

It can be seen from figure 7 that the Sig of the other indicators is less than 0.05 and the corresponding coefficients are all larger than 0, indicating that the other indicators have a significant impact on the quality score and have a positive correlation. Since the coefficient of time continuity is the largest, it indicates that the time continuity ratio has the greatest impact on the quality score. Similarly, it can be seen that the accuracy rate of the engine fuel flow has the least impact on the score.

| Model                        | Non standardized coefficient | Standardized coefficient | t     | Sig     |
|------------------------------|------------------------------|--------------------------|-------|---------|
| constant                     | 5.275                        | 0.857                    | 0.392 |
| MIL state                    | 5.918                        | 0.418                    | 0.000 |
| speed                        | 6.396                        | 0.418                    | 0.000 |
| the net output torque        | 4.274                        | 0.615                    | 0.000 |
| the friction torque          | 5.694                        | 0.609                    | 0.000 |
| the SCR downstream NOx       | 5.968                        | 0.106                    | 0.000 |
| the intake air volume        | 1.962                        | 0.103                    | 0.002 |
| time continuity              | 24.703                       | 6.143                    | 0.000 |
| atmospheric pressure         | 2.678                        | 0.207                    | 0.000 |
| engine fuel flow             | 1.702                        | 0.410                    | 0.001 |
| SCR outlet temperature       | 7.165                        | 0.082                    | 0.000 |
| DPF pressure                 | 9.988                        | 0.295                    | 0.000 |
| engine coolant temp          | 13.135                       | 0.596                    | 0.000 |
| tank level                   | 5.263                        | 0.157                    | 0.000 |

Figure 7. Linear regression model result

4. Conclusion

By analysing the problems in OBD data, a data quality evaluation system is proposed, the monitoring data results are evaluated and analysed and factors affecting data quality are studied. The conclusions are as follows:

1. There is a large difference in the proportion of first-grade vehicles in different months, and the difference in the proportion of grades above second-grade is not obvious among months.

2. The factors influencing the data quality score include vehicle brands and indicators. From the proportion of the first-grade vehicles, “Jinlv Brand” is the vehicle brand with the best data quality.

3. Except the three indicators of engine speed, SCR upstream NOx sensor output and SCR inlet temperature, the other indicators have a significant impact on the quality score and the time continuity ratio has the largest impact on the score.

4. The engine model also affects the data quality rating. Among them, there are 37 engine models. The data is too scattered and no statistical analysis is performed.
5. Acknowledgments
I would like to express my gratitude to all those who helped me during the writing of this article. I gratefully acknowledge the help of my professor, Ms. Liu Zhili, who has offered me valuable suggestions in the academic studies. In the preparation of the article, she has spent much time reading through each draft and provided me with inspiring advice. Without her patient instruction, insightful criticism and expert guidance, the completion of this article would not have been possible. I also owe a special debt of gratitude to my classmates at school, Yu Xiangyang and Wei Xinmei, from whose devoted teaching and enlightening lectures I have benefited a lot and academically prepared for the article and data analysis.

In addition, special thanks should be given to Beijing Transportation Institute because the program in this article is funded by "OBD-based research and development and demonstration application of metering and diagnostic technology for emission reduction devices for heavy operating vehicles" (Beijing transportation technology project).

6. References
[1] Liu Yang and Guo Jianbin 2008 Analysis and prospects of ecological environment pollution caused by automobiles J. Environ. Prot. Circ. Econ 09 44–50
[2] Zhang Xiang 2006 Research on OBD of modern automobile J. Urban. Veh 04 41–3
[3] Niu Xiaoxiao and Liu Qiang 2010 Overview of OBD vehicle diagnostic system J. Automot. Parts 04 73–5
[4] Wang Jianhai, Fang Maodong, Gao Jidong, Lu Hongyu and Dai Chunbei 2006 Basic principle and application of OBD for gasoline vehicles J. Automot. Eng 05 491–4
[5] Qiu Yali 2010 Research on the development of automobile fault diagnosis technology J. Transp. Stand 11 223–5
[6] Zhong Yong and Li Minxu 2009 The current status and future of OBD technology J. Annex. Eng. Coll 01 55–8
[7] Guo Jifu, Sun Jianping, Wen Huimin, Zhao Xiang and Lei Fangshu 2017 Analysis of car trip characteristics based on OBD data —Taking Beijing as an example J. Urban. Transp 15 70–7
[8] Feng Lisha, Feng Guosheng, Niu Xiaoyan and Lin Hao 2012 Design of multi-channel data acquisition system for vehicle OBD performance parameters J. Automot. Eng 03 20–2
[9] Li Man 2016 Vehicle Condition Monitoring and Fault Diagnosis Based on OBD System (Beijing: Tsinghua University) pp 14–22
[10] Luo Haibin 2019 Research on the application of OBD technology based on Android platform J. Electron. Compon. Inf. Technol 03 105–7
[11] Liao Hailin, Li Youxin, Ye Shaoxiang and Zhu Zhengjia 2017 OBD vehicle data collection and upload system in Android platform J. Appl. Comput. Syst 26 271–4
[12] Yang Yongchang and Chen Kai 2013 Application of OBD automobile diagnostic system in automobile detection J. Value. Eng 32 172–3
[13] Lin Hongliang 2007 Application of OBD in vehicle emission control J. Agric. Equip. Veh. Eng 11 47–9
[14] Liu Liang, Wang Shiping, Liu Fuhua and Chen Dan 2016 Talking about the application and trend of Internet of Vehicles based on OBD function expansion J. Commun. World 05 236
[15] Song Danping and Huang Miaohua 2008 Line data flow diagnosis method of OBD system and its application J. Beijing. Automob 03 35–8