Recognition of Abnormal Driving Behavior of Highway Vehicles Based on Data Characteristics

Jiahao Feng[1], Jinsong Ye[2(✉)], and Lei Zhou[2]

[1] Beijing Wuzi University, Beijing, China
[2] China Academy of Transportation Sciences, Beijing, China
yejs@motcats.ac.cn

Abstract. Using the data characteristics of traffic big data to intelligently identify abnormal driving vehicles, it is of great significance to fight against toll evasion and improve operation management & efficiency of highway. To improve the efficiency of joint inspection of existing highway vehicle toll evasion behaviors, this paper carries out behavioral analysis and data analysis on the five existing abnormal driving behaviors, and summarizes the data characteristics of five abnormal driving behaviors: breakthrough type, alternate use of OBU, vehicle type history issues, vehicle timeout and overspeed. According to data characteristics of each type, this paper presents an online calculation method of abnormal driving behavior based on data features of threshold table, sliding window. This paper concluded that the characteristics of vehicle data can effectively identify abnormal driving behaviors on highways and improve the efficiency of inspection.

Keywords: highway · traffic big data · abnormal driving behavior · data identification characteristics · online calculation

1 Introduction

By the end of 2015, the highway network toll collection system has been basically realized in China, which means vehicles only need to charge once among provincial highway networks. As we start to see the way highway toll fees system shows the most promise, we must also be aware of where it might stumble. One of the weaknesses is the abnormal driving behaviors, which not only disrupts the normal order of toll collection, but also causes economic losses to highway operating units [1].

Much has been written and debated about such highway toll evasion behavior among academic circles, especially on the classification and characteristics of abnormal high-speed driving behavior. Starting from the psychological characteristics and the purpose of evading tolls by illegally operating vehicle owners, Hu Guoyong expounds the types of highway vehicle reversal to escape tolls, and introduces four types of reversal: two vehicles in the same direction change the license plate and IC card, change the license plate and IC card between two vehicles, reverse the card multiple times, and use no card to avoid fees [2]. Based on actual experience, Tan Yuan introduced 8 common toll evasion

Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 272–284, 2023.
https://doi.org/10.2991/978-94-6463-024-4_29

behaviors in ETC lanes and analyzed the process of common toll evasion behaviors in ETC lanes [6]. Based on actual work experience, Sun Min divided ETC lane evasion into customs clearance type, multi-entry or no exit type, switching type, counterfeiting type and non-conforming type [5]. In response to the loopholes in the ETC system, Wu Lieyang introduced the inconsistency of the vehicle model, the inconsistency of the vehicle card, the mobile OBU, the opposite sign exchange, following the car maliciously, falsely collecting the card, exporting no card, turning around and evading the fee [8].

According to the above research, due to the limitation of the technical means at that time, the identification of vehicles with abnormal driving behaviors is mainly through on-site inspection and manual inspection, and the behavioral characteristics of toll evasion vehicles are classified. However, with the advancement of technology, CPC cards and ETC cards that can record paths have gradually replaced traditional pass cards, and some evasion methods have become a thing of the past. At the same time, in the era of big data, it's not enough to combat evasion only from the behavioral characteristics of vehicles [7].

In this regard, this paper uses highway traffic big data to identify abnormal vehicle driving behaviors based on vehicle data characteristics, and divides vehicle abnormal driving behaviors into five types: breakthrough, OBU alternate use, vehicle-type issues, vehicle timeout, and vehicle overspeed. According to the data identification features of vehicles with abnormal driving behaviors, the traffic data results of vehicles with abnormal driving behaviors are screened out, and the data results intuitively present the license plate number, OBU number and other information of the vehicles evading tolls. This is not only conducive to data inspection of vehicles, but also saves manpower and material resources for on-site inspections, and greatly improves the efficiency of joint inspections.

2 General Design and Data Preprocessing

2.1 General Design

The identification features of the data are unique among different abnormal driving behaviors on highways. In order to accurately identify vehicles with abnormal driving behaviors, this paper summarizes different data characteristics corresponding to each abnormal driving behavior by analyzing the traffic data, and also designs an online calculation method for different data identification characteristics.

The design of the online computing method is divided into three layers, namely the basic data layer, the data analysis layer, and the application layer [3]. The basic data layer stores the traffic data of the highway vehicles, mainly including the entrance table, the gantry table, the exit table and the index fields in the table. The basic data layer also includes the construction of a station threshold table and a sliding window according to the data identification features of different abnormal driving behaviors. In the data analysis layer, the vehicle traffic data is firstly analyzed in a centralized manner, and vehicles that meet the five identification characteristics of abnormal driving behavior data are screened. The vehicles with abnormal driving behavior after screening are classified and integrated according to different data identification characteristics. Later take the province where the vehicle license plate number is located and record

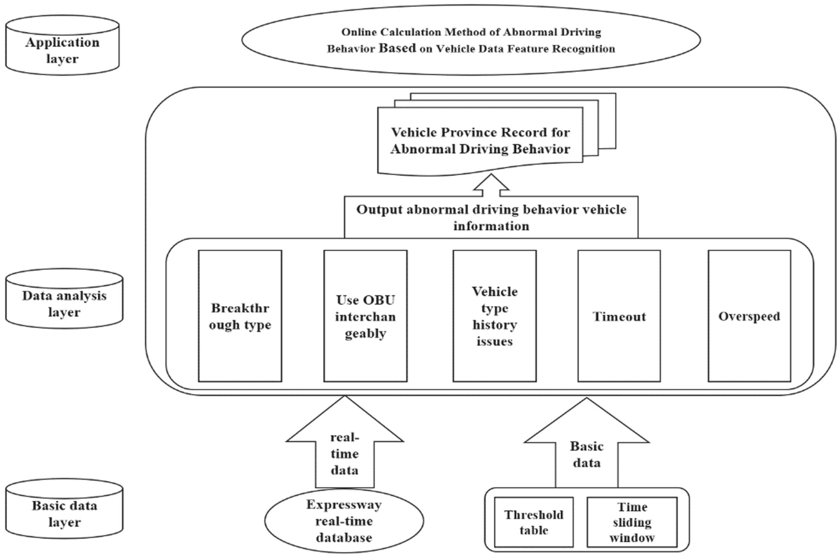

Fig. 1. System Architecture

the vehicles with abnormal driving behavior in the vehicle province record form. The application layer visually presents the license plate numbers of vehicles with abnormal driving behaviors. The system architecture is shown in Fig. 1.

2.2 Data Preprocessing

The huge, complicated and low update frequency of data in the highway real-time database contains many data with empty fields, input errors or repeated input. This part of data needs to be eliminated. For vehicles that are not within the scope of this study, data also needs to be excluded. For the entering wrong data, the license plate number, pass ID, OBU number and model code of the limited vehicle are not empty. The non-OBU vehicle traffic data does not in this paper's research, only the vehicles with OBU and CPC cards as the traffic medium are studied, so the traffic medium type is limited to 1 or 2. Special vehicles on highways such as police cars, ambulances or emergency vehicles can reasonably park on the highway for a long time or drive at excessive speed. Therefore, the vehicle types are limited to passenger cars and trucks, and the travel time and distance are all greater than 0.

For the five abnormal driving behaviors, a station threshold table and sliding window are constructed. The station threshold table mainly defines two interval ranges, namely the minimum speed of highway - maximum speed of highway, and the minimum travel time of highway - maximum travel time of highway. The construction of the threshold table is mainly for overtime and overspeed [4]. The sliding window is mainly based on the time series similarity clustering technology. It starts a new segment from the first point of the time series and continues to grow backward until the fitting error between the segment and the time series exceeds a specified threshold [9].

Then, the following sequence point is used as the beginning of the new segment, and the above process is repeated until the end of the time series. Besides, the sliding window is mainly aimed at breakthrough type and vehicle type issues.

3 Identification Features of Abnotmal Driving Behavior Based on Highway Big Data

3.1 Breakthrough Type

3.1.1 Behavior Analysis

Breakthrough type evasion means that in the ETC lane, the evasion vehicle uses the OBU recognition result of the car in front or behind, and uses the automatic railing machine to lift it, and it runs out of the high speed [10]. In terms of data representation, there are two types of evasion fees: the ID is in the entry table but not in the exit table, that is, there is no entry, and the ID is in the exit table and not in the entry table, that is, there is no entry. This study takes the type of entry and exit as an example to discuss.

There are two main types in abnormal driving behavior with entrance but without exit. One is that the vehicle does not have an OBU. At the entrance, the CPC card is obtained from the manual passage to enter the highway. At the exit, the vehicle uses the rear car with an OBU to use it. The ETC recognition result of the car passes through the ETC lane and exits the highway. The other is that the vehicle has an OBU, enters the highway through the ETC lane, follows the front vehicle with an OBU at the exit, and uses the ETC identification result of the front vehicle to break through the ETC lane to get out of the highway.

3.1.2 Data Analysis

For vehicles with abnormal driving behavior, the information of the vehicle is only in the entry table, and the vehicle information disappears at a certain gantry at the high speed. The specific analysis is as follows.

The pass ID of each vehicle passing through is unique, the data in the entry table is complete, and there is no data in the exit table. The pass ID, entry date and time are used as the criteria to obtain the data characteristics of abnormal driving behavior with entry or exit, that is, the pass ID is in the entry table. Not in the export table, there is no export table data for seven days after the date of import.

Further, there is no exit table data for vehicles with abnormal driving behaviors with entrance but without exit. In order to determine the location where the vehicle information disappears, the gantry number in the vehicle gantry table is used as the criterion to obtain the gantry number that the vehicle last passed in this passage. By determining the gantry number where the vehicle disappears and whether the vehicle traffic data disappears at this gantry number for one or more times, it can provide reference for rapid inspection and improve the inspection efficiency.

3.1.3 Data Identification Characteristics

The data characteristic of abnormal driving behavior with entrance but without exit is that the traffic ID is in the entry table but not in the exit table, and there is no exit

Field	Traffic ID	Actual license plate number
Entry table	Complete and eligible for screening	Complete andeligible for screening
Export table	The vehicle data Traffic ID existing in the entrance table is not in the exit table	
Gantry table	Complete and eligible for screening	The first gantry number in the order of cumulative mileage after the transaction is the same

Fig. 2. Data Characteristics Table of Abnormal Driving Behavior with Entrance without Exit

table information within seven days after the entry date; the first gantry number of the accumulated mileage after the transaction is the same, and the vehicle traffic data is once or disappear at the same gantry number multiple times. The data identification feature table is shown in Fig. 2.

3.2 OBU-Using Interchangeably

3.2.1 Behavior Analysis

Two or more OBUs are handled in violation of the vehicle regulations and used alternately. For example, if a vehicle illegally handles OBU① and OBU②, alternately uses OBU① and OBU② to enter and exit the highway in July, and also alternately uses OBU① and OBU② to enter and exit the highway in August, and the actual license plate number in the OBU is the same, it can be determined that this vehicle has the behavior of handling multiple OBUs in violation of regulations.

3.2.2 Data Analysis

OBU using interchangeably means the vehicle has abnormal driving behavior of alternate use of OBUs. In short, the vehicle alternately uses multiple OBUs in adjacent months. The specific analysis is as follows.

In the vehicle exit table of the first month, the OBU number is an abnormal field, there are two different OBU numbers, and vehicles use two different OBUs alternately in this month; in the vehicle exit table of the second month, OBU The number is still an abnormal field, there are two different OBU numbers, and the OBU numbers are the same as the two OBU numbers of the previous month, and the vehicles are still used alternately within this month. Taking the OBU number and the actual license plate number as the criterion, the data characteristics of the abnormal driving behavior of the OBU alternately used are obtained, that is, the vehicle has used 2 or more OBUs in each month for two consecutive months, showing the form of ABA, and the actual driving behavior. The license plate number information is the same.

Field	Traffic ID	Actual license plate number	OBU number	Type of vehicle	Remaining fields
First month	Complete and correct	A	① ②	a or b	Complete and eligible for screening
Second month	Complete and correct	A	① ②	a or b	Complete and eligible for screening

Fig. 3. OBU Alternate Use of Data Characteristics

3.2.3 Data Identification Characteristics

The data characteristic of the alternate use of OBUs is that the vehicle has used two or more OBUs each month in two consecutive months, and the license plate number information in the two OBUs is the same. That is, the vehicle uses two OBUs, in which the content of the OBU number field is different, the content of the actual license plate number field is the same, and the information of the two OBUs has appeared every month for two consecutive months. The data identification feature table is shown in Fig. 3.

3.3 Vehicle Type Issues

3.3.1 Behavior Analysis

Illegal drivers generally use the following behaviors to tamper with the OBU model information. The OBU issuer writes the model information of the low model into the OBU of the high model due to manual error or partnership evasion, so that the high model can press Low when passing the ETC lane. The model is charged; the illegal driver will disassemble and install the OBU of the low model to the high model for use, so as to achieve the purpose of evasion.

Vehicle type problems are mainly divided into three types: (1) There were two models in the vehicle OBU in the early years, but now there is only one model; (2) There was only one model in the vehicle OBU in the early years, but now there are two types in the OBU. (3) In the early years, two models were recorded in the OBU, which still exists.

3.3.2 Data Analysis

Vehicles with abnormal driving behaviors with model problems mainly refer to the models entered in the OBU before and after a certain time point that do not match the actual models. The specific analysis is as follows.

Taking a certain time point as a constraint node, before the time node, the vehicle traffic data is divided into two situations: A. In the vehicle exit table, the model code is an abnormal field, and more than two models have been entered in the OBU. B. There is no abnormal field in the vehicle traffic data. After this time node, the vehicle data is still divided into two cases: A. In the vehicle export table, the model code is an abnormal field, and more than two models have been entered in the OBU. B. There is no abnormal field in the vehicle traffic data. Therefore, the abnormal driving behavior of the model problem will take the following forms: AA, BA, AB. Using the vehicle license plate number, OBU number, and model code as the discrimination criteria, the data characteristics

Field	Actual license plate number		OBU number		Cartype	
Time frame	Before a time node	After a time node	Befter a time node	After a time node	Befter a time node	After a time node
AB type	same	same	same	same	different types	same type
AA type	same	same	same	same	different types	different types
BA type	same	same	same	same	same type	different types

Fig. 4. Data Characteristics Table of Vehicle Type Issues

of the abnormal driving behavior of the vehicle type issues are obtained. Content has changed.

3.3.3 Data Identification Characteristics

The data characteristic of the vehicle type issues is that when the content of the vehicle license plate number and OBU number fields are the same, the content of the vehicle model code field changes. When a certain time point is used as a constraint node, there are mainly three situations: (1) There is a situation where the same OBU and different models under the license plate number exist before a certain time node of the vehicle, and there is no such situation after a certain time node, that is, the AB type. (2) Before a certain time node of the vehicle, there is a situation where the same OBU is different from the model under the license plate number, and this situation also exists after a certain time node, that is, the AA type. (3) Before a certain time node of the vehicle, there is no situation that the same OBU and the vehicle type under the license plate number are different, but this situation exists after a certain time node, that is, the BA type. The data identification feature table is shown in Fig. 4.

3.4 Vehicle Timeout Behavior

3.4.1 Behavior Analysis

The traditional U-shaped fee evasion will lead to overtime, but it has gradually become history. Therefore, this paper mainly studies the vehicle timeout phenomenon, which mainly refers to the time when the vehicle passes through the road section is lower than the minimum speed.

3.4.2 Data Analysis

For vehicles with timeout behavior, it is necessary to compare the actual passing time of the vehicle with the maximum passing time. The specific analysis is as follows.

If the vehicle traffic data is complete, we can calculate the actual traffic time and the longest traffic time under the current traffic ID according to the exit date and time, entry date and time and total mileage in the exit table. Taking the actual traffic time and the longest traffic time as the criteria, we can get the data characteristics of overtime abnormal driving behavior, that is, the actual traffic time used by vehicles passing through the section is greater than the longest traffic time.

Field	Traffic ID	Actual license plate number	Entry date and time(h_1)	date and time of export(h_2)	Total billable mileage(s)	longest transit time(h_{max})	Actual transit time(h)	Abnormal identification
Export table	Complete and correct	Complete and correct	Complete and correct	Complete and correct	Complete and correct	[s/(40km/h)]+0.4h	h_2-h_1	V>120km/h

Fig. 5. Timeout Data Characteristics

3.4.3 Data Identification Characteristics

The data characteristic of abnormal driving behavior over time is that when the exit driving data information is complete, the actual travel time used by the vehicle to pass through the highway section is greater than the maximum travel time. The maximum travel time is achieved by adding 0.4 h to the time it takes for vehicles to cross the highway at the minimum speed limit of 40 km/h. The calculation formula is [exit date and time (h2) – entrance date and time (h1)] > [(total mileage/40 km/h) + 0.4 h]. The data identification feature table is shown in Fig. 5.

3.5 Vehicle Overspeed Behavior

3.5.1 Behavior Analysis

Speeding is a common abnormal driving behavior on highways, which means that the driver's driving speed exceeds the speed stipulated by laws and regulations while driving.

3.5.2 Data Analysis

Vehicles with abnormal speeding behaviors mainly compare the average speed of vehicles passing through the highway and the upper limit of the speed limit of the highway. The specific analysis is as follows.

The vehicle traffic data is complete. According to the exit date and time in the exit table, the entrance date and time, and the total billed mileage, under this traffic ID, the traffic time and average speed are calculated, and the traffic time and average speed are used as the discrimination criteria., to obtain the data characteristics of the abnormal driving behavior of speeding, that is, the average speed of the vehicle passing through the road section is greater than 120 km/h, which is a speeding vehicle.

3.5.3 Data Identification Characteristics

The data characteristic of abnormal speeding behavior is that when the exit data information is complete, the average speed of vehicles passing through the highway section is greater than the upper limit of the highway speed limit. The calculation formula is, [total mileage charged/(exit date and time - entrance date and time)] > 120 km/h. The data identification feature table is shown in Fig. 6.

Field	Pass ID	Actual license plate number	Entry date and time(h_1)	date and time of export(h_2)	Total billable mileage(s)	Average apeed(v)	Abnormal identification
Export table	whole correct	whole correct	whole correct	whole correct	whole correct	$s/(h_2-h_1)$	V>120km/h

Fig. 6. Overspeed Data Identification Characteristics

	With entrance without exit	Time out	Overspeed
July	1374342	74516	120984
August	4550473	73322	119926
September	1058562	106305	118244
October	1440598	158607	137515

Fig. 7. With Entrance Without Exit, Timeout and Overspeed of Data Situation

4 Results and Analysis

Based on the data identification features of vehicles, this paper uses SQL language to query the vehicle traffic data in the traffic database, and analyzes the data query results of five abnormal driving behaviors.

4.1 Analysis of Data Results with Entrance But Without Exit, Timeout, and Overspeed

The query results of vehicle data with three abnormal driving behaviors in Beijing from July to October 2021. The data are shown in Fig. 7.

Through the comprehensive analysis of the query results with entrance but without exit, timeout and overspeed, we find that the data volume distribution of abnormal driving behavior with entry and without exit is quite different. The data volume in August exceeds the sum of the data volume in July and September, and the reasons for the high data volume in August still need to be further explored. Overtime and overspeed are also common abnormal driving behaviors on highways. Only the Beijing Highway has a high amount of overtime and speeding vehicle traffic data every month. From the distribution point of view, the data distribution of overtime and speeding abnormal driving behaviors is relatively uniform from July to October, and the data volume reaches the highest level in October. The reason for the most overtime and speeding vehicles in October may be related to the intensive travel of people during the National Day Holiday.

Comparing the three types of abnormal driving behaviors, the data volume of abnormal driving behaviors with entry and without exit is much higher than the data volume of overtime and speeding in each month, and even the data volume of one month with entry and no exit exceeds that of abnormal driving with overtime and speeding. The sum of the data volume for four months. The increase and decrease trend of data volume with incoming and outgoing data fluctuates greatly, while the line trend graph of overtime and overspeed tends to be flat. In terms of the horizontal proportion, the total data of the three types of abnormal driving behaviors accounted for 51% in August, most of which were vehicles with abnormal driving behaviors; the total data of the three abnormal driving behaviors in other months accounted for are less than 20%. From the perspective of vertical proportion, the total data of abnormal driving behavior with entry and without exit in four months accounted for 90%, while the total data of overtime and speeding from July to October accounted for about 5%. In view of the high amount of data of abnormal driving behavior, it is necessary to further analyze the data of the vehicle disappearing gantry of this abnormal driving behavior.

The graph of the data results of the vehicle disappearing gantry shows that the data volume in August is still far ahead of the other two months, but with the increase in the number of disappearances, the data volume also decreases significantly. The amount of data that a vehicle disappears once in a certain gantry is much more than the number of times it disappears. However, it cannot be regarded as a fee evasion behavior if it disappears only once. It may be caused by other unexpected situations, such as abnormal data collection of the gantry, and unexpected situations in the vehicle. Wait. Compared with the reason why the vehicle disappears once in a certain gantry, most of the two or more times can be attributed to the high-speed toll evasion. The data analysis diagram is shown in Fig. 8.

4.2 Data Result Analysis of OBU Alternate Use and Model Problems

The query results of OBU's alternate use of vehicle data from May to November 2021 in Beijing, see Fig. 9 for the data.

Model problem Abnormal driving behavior is mainly divided into three types: (1) Before January 1, 2021, there are different models with the same OBU and license plate number, and there is no such situation after January 1, 2021, AB type. (2) Before January 1, 2021, there are different models under the same OBU and license plate number. After January 1, 2021, this situation still exists, AA type. (3) Before January 1, 2021, there are no different models under the same OBU and license plate number. After January 1, 2021, there is such a situation, BA type. See Fig. 10 for the query results of the model question table data.

Through the comprehensive analysis of the query results of OBU alternate use, vehicle model problems and abnormal driving behavior, we find the distribution of the amount of data alternately used by OBU is relatively uniform, and the largest difference in data amount between months is only about 1,000 vehicles. The amount of data alternately used by OBU reaches the highest level from September to October. Compared with other abnormal driving behaviors, the amount of data that OBU alternately uses for abnormal driving behaviors is insignificant. Some unscrupulous car owners do not pay or pay less to escape the highway tolls and rack their brains to crack the OBU, so the relevant technical departments still need to improve the OBU technology.

Regarding the abnormal driving behavior of the model, the AA type has the largest amount of data, accounting for 59%; while the BA type has the least amount of data, accounting for only 8%. This also shows that during the initial release of OBU, there were many mis-issued OBUs. With the improvement of OBU equipment and the increase in efforts to crack down on fee evasion, car owners changed the model code in the OBU

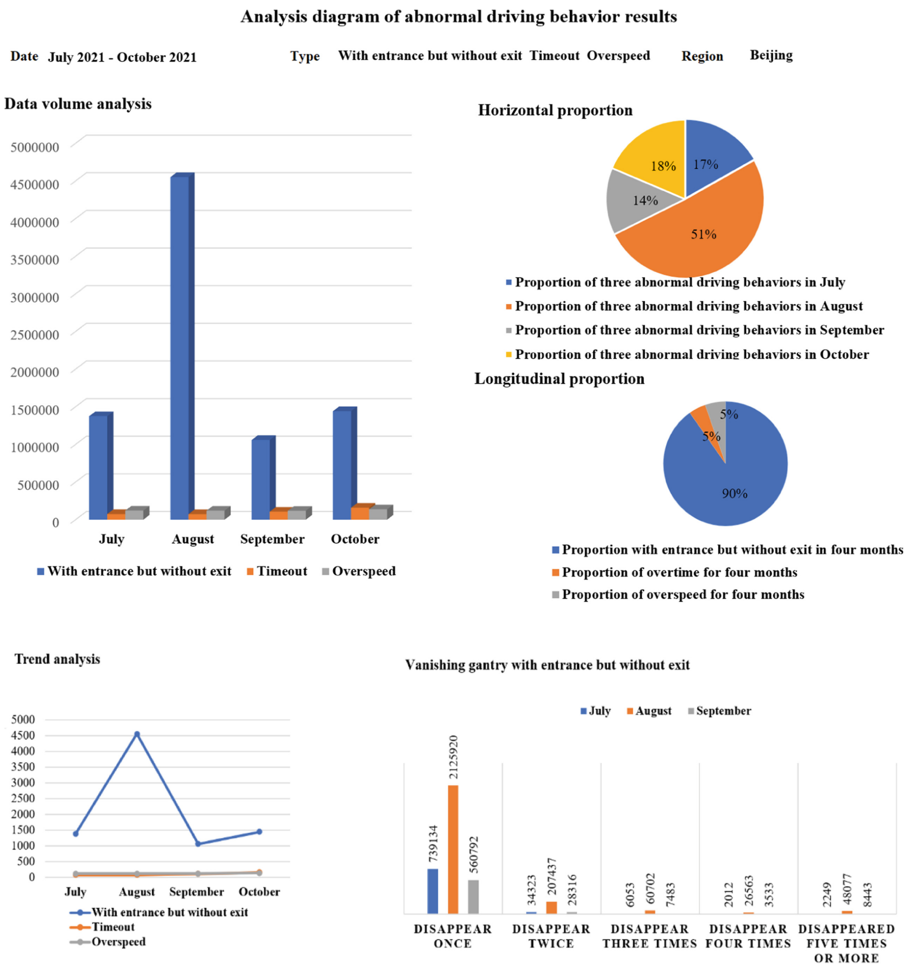

Fig. 8. Analysis Chart of Abnormal Driving Behavior in With Entrance but Without Exit & Timeout & Overspeed

Month	Data (unit: bar)
May to June	5999
June to July	5905
July to August	5433
August to September	5566
September to October	6315
October to November	5392

Fig. 9. OBU Alternate Use Data

Type	Year	Data(unit:bar)
AB type	2020-2021	66828
AA type	2020-2021	118446
BA type	2020-2021	15058

Fig. 10. Vehicle Type Issues Data

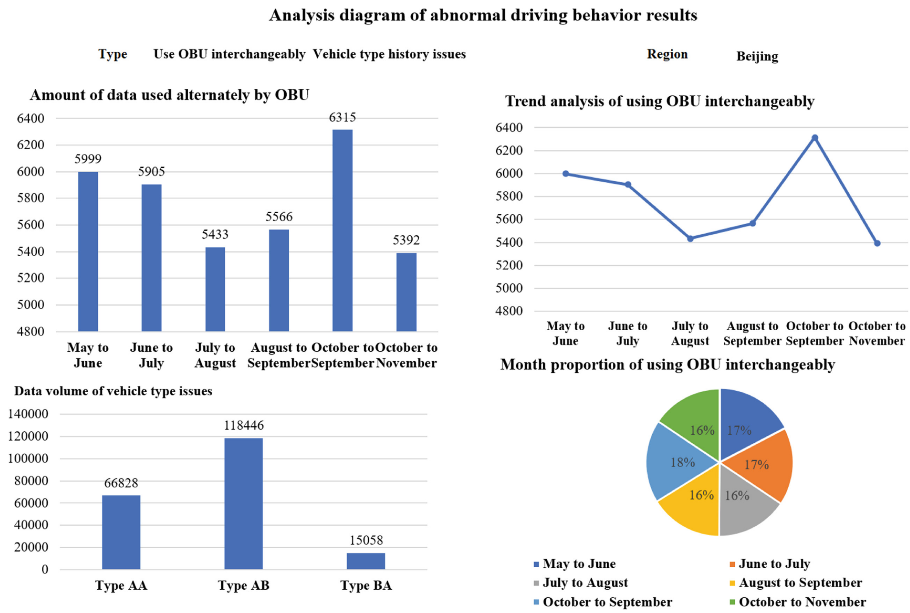

Fig. 11. Analysis Chart of Abnormal Driving Behavior in OBU-using Alternately & Vehicle Type Issues

to the model code consistent with the actual model; from the data Judging from the fact that the AB type has a large amount of data, it shows that there are still a large number of illegal car owners who have tampered with the OBU model code through illegal means in order to escape the high-speed toll. The result graph of the data is shown in Fig. 11.

5 Conclusion

To promote the efficiency of joint inspection of existing highway vehicle toll evasion behaviors, this paper presents the data identification characteristics of five kinds of abnormal driving behaviors, and constructs the online calculation method of abnormal driving behavior based on threshold table, sliding window through the analysis of highway vehicle traffic behavior and its data. This paper further analyzes the query results that meet the data characteristics of five abnormal driving behaviors. In this way, the data volume and trend of each abnormal driving behavior in a certain period of time can be accurately and clearly understood, so as to achieve the purpose of more efficient data audit and joint audit.

Acknowledgement. Fund Project: Open Foundation of Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport (2020B1202).

References

1. Hu B (2020) Research on the control of toll evasion in Sichuan Highway Group. University of Electronic Science and technology
2. Hu G (2011) Analysis of highway vehicles "reversing cards" to evade tolls. West Commun Technol (03):49–51 + 77
3. Li S, Zhou Z, Li Y, Wang Y, Song X, Wang P (2018) Prediction of highway fare evasion based on IGA-IBP algorithm. Comput Eng Des 39(12):3840–3845
4. Ma Q, Xu Y, Ding W, Zou J. Real time detection method for Highway fare evasion vehicle screening. Comput Technol Dev 189:07–189
5. Sun M (2014) Analysis of highway ETC charging method. China Sci Technol Inf (05):201–202
6. Tan Y, Li XY (2014) Analysis and countermeasures of highway ETC fee evasion. Transp Technol Econ 16(03):30–33
7. Wang H (2020) From manual to big data + AI: "one net" audit acceleration. China Transp Inform (11):18–22
8. Wu L (2016) lichuan on the types and inspection methods of highway ETC fee evasion. China Transp Inform (08):31 + 85
9. Xiang H, Yang P, Yi J (2020) Prediction model of highway fare evasion vehicle state based on RF-LR. J Chongqing Norm Univ (Nat Sci Ed) 37(01):75–80
10. Yang X, Peng Y (2020) Discussion on toll management for ETC free flow China Transp Inform (09):40–43