Proposal of classification method of bus operation states using sensor data

Takuya Yonezawa*, Ismail Arai*, Toyokazu Akiyama† and Kazutoshi Fujikawa*
* *Graduate school of Information Science, Nara Institute of Science and Technology, Japan*
†*Faculty of Computer Science and Engineering, Kyoto Sangyo University, Japan*
Email: yonezawa.takuya.yq2@is.naist.jp, ismail@itc.naist.jp, akiyama@cc.kyoto-su.ac.jp, fujikawa@itc.naist.jp

***Abstract*—In bus companies, it is important for the operation manager to grasp operation states of the vehicle from the viewpoint of safety management and improving the operation efficiency. Currently, for allowing the operation manager to grasp operation states of the vehicle, the driver records operation states by manually operating a recorder called 'Digital-tachograph'. However, operating a digital tachograph is a heavy burden to the driver. In this research, in order to solve this problem and to realize efficient operation, we propose a method for automatic classification of operation states using sensor data obtained from the buses. We implemented a classifier using Random Forest with the sensor data. The correct answer rate was 0.9 or more in each condition unless it was irregular operation.**

***Keywords*-Machine learning, Advanced driver assistance, Public transportation**

I. Introduction

In the recent Japanese passenger transportation industry, the following two are issues.

- Degradation of labor environment
- Lack of management of operation managers

About the first one, due to the increase in the number of bus companies, shortage of workers has occurred, and the working environment has deteriorated, such as an increase in the total travel distance per person. About the second one, the reason for this is a decrease in profit due to an increase in the number of bus companies and a lack of knowledge to efficient management of drivers. Each bus company has an employee of a position called 'operation manager'. The operation manager has the role of setting bus routes and instructing safe driving to drivers. However, due to the reduction in staffing due to intensified market competition, the burden of management for operation managers is increasing.

According to the above background, it can be said that in bus companies, it is important for operation managers to accurately and in real time grasp operation states of the vehicles from the viewpoint of safety management and efficient operation.

II. Related technologies and issues

Currently, for allowing an operation manager to grasp operation states of vehicles, drivers manually operate a terminal of a digital tachograph and record operation states. By using the digital tachograph, it became possible to record new data (e.g. GPS, acceleration, temperature etc...) in addition to the data recorded by the conventional analog tachograph (e.g. timestamp, vehicle speed, engine speed)[1]. However, burdens of terminal operations during driving and its human errors are major problems.

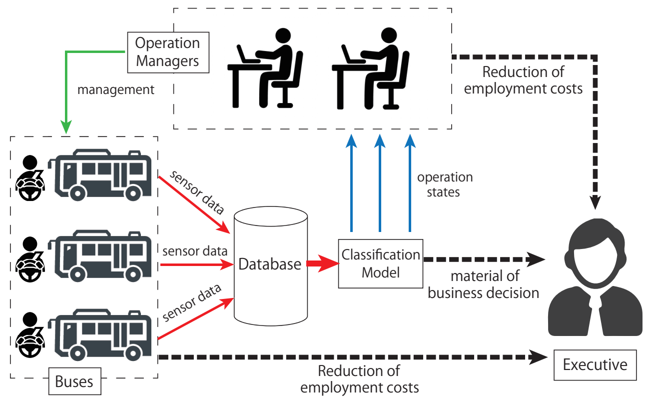

Figure 1. Model of service realized by this research

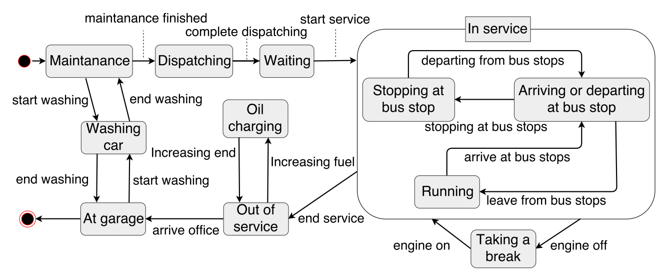

Figure 2. Flow of bus operations

To solve these problems and realize accurate grasp of the operation state, this research aims at automatic classification of operation states using sensor data obtained from buses. An overview of this proposal is shown in Figure 1.

III. Methodology

A. Requirements

Table I shows nine operation states classified in this research. Also, the flow of the bus operations is shown in Figure 2. These operation states are recorded in the transportation industry and are used by operation managers and executives for grasping the working contents of drivers. The 'In service' state in Table I is further classified into

Table I
STATE TO BE CLASSIFIED

State	Definition
Maintenance	inspection work before dispatching
Washing car	washing and cleaning the bus
Dispatching	moving towards the first bus stop
Waiting	Waiting at the first bus stop
In service	bus is in service
Taking a break	taking less than 4 hours break
Out of service	moving to garage after bus finishes service
Oil charging	refueling at gas station
At garage	stopping at garage

Table II
SUBDIVISION OF THE 'IN SERVICE' STATE

State	Subdivision
In service	Stopping at bus stops
	Arriving or departing at bus stops
	Running

three subdivisions as shown in Table II. In particular, with regard to the three states in 'In service' shown in Table II, it is heavy burden because it must manipulate the terminal of the digital tachograph in addition to instrument operation in the bus. However, in order to accurately grasp operation states by the operation manager and the bus company, we try classification in this research.

These matters interviewed the bus company and were defined as requirements. In this research, we try to classify 11 states in Table I and Table II.

B. Obtained sensor data and newly defined features

Various sensor data can be obtained from digital tachograph which is independently developed in collaborative research. However, from the viewpoint of versatility, only data (GPS, Vehicle speed, Engine speed, Total travel distance, Engine pulse) that can be obtained from general digital tachographs are used.

In this research, in order to improve the accuracy of classification, five features are newly defined. These new features can be computed only from sensor data and geographic information on routes. Table III shows the features used to construct the classification model.

- A speed change per second
- Linear distance to a nearest bus stop
- Linear distance to the garage
- Linear distance to the car wash place
- Garage flag (The coordinates are plotted in the garage)

'A speed change per second' and 'Linear distance to a nearest bus stop' are newly defined to improve the classification accuracy of operation states in the vicinity of the bus stop during 'In service'. 'Linear distance to the garage',

Table III
FEATURES USED FOR CONSTRUCTION OF CLASSIFICATION MODEL

Feature	Data resources
GPS (Latitude, Longitude, Altitude)	raw sensor data
Vehicle speed	raw sensor data
Engine speed	raw sensor data
Total travel distance (ODO)	raw sensor data
Engine pulse	raw sensor data
A speed change per second	Vehicle speed
Linear distance to a nearest bus stop	Latitude, Longitude
Linear distance to the garage	Latitude, Longitude
Linear distance to the car wash place	Latitude, Longitude
Garage flag	Latitude, Longitude

'Linear distance to the car wash place' and 'Garage flag' are defined for classifying similar operation states such as 'Maintenance', 'Washing car', 'Oil charging', 'Waiting'.

In this research, in order to classify operation states shown in Table I and Table II, we perform multi-class classification using supervised learning. From the viewpoint of stability and a real time classification of classification model, in this research, Random Forest was adopted as a supervised machine learning method[2].

IV. EXPERIMENTS AND DISCUSSIONS

In this research, we evaluated the same route classification model and multiple route classification model. The sensor data used in this research was obtained from route buses traveling in Kobe City, Hyogo prefecture, Japan every one second. These sensor data are obtained from buses traveling on two routes (S-Route and D-Route). The dataset used for the evaluation is data of six operations in total for three days of two buses traveling on S-route and D-route. The number of bus stops on each line is 22 for S-route and 19 for D-route.

Below, the classification result of D-route in 'Same route classification model', the classification results of S-route and D-route in 'Multiple route classification model' are shown in Table IV.

A. Same route classification model

In the classification performance evaluation on the same route, the model was constructed using the data set of the first day of the D-route as the training data. In addition, the performance evaluation of the classification model was carried out using the D-route data set on the second and third days as test data.

Focusing on the correct answer rates on the 2nd and 3rd days on 'Same route result (D-Route)' in the Table IV, we can see that there is a big difference between 0.819 on the 2nd day and 0.949 on the 3rd day. The reason for this is that D-route's 3rd day flow of operation was irregular. The flow of the operation on the 3rd day of the D-route was

Table IV
ALL CLASIFICATION RESULTS

	Same route result (D-route)				Multiple route result (S-route)				Multiple route result (D-route)			
	2nd day		3rd day		2nd day		3rd day		2nd day		3rd day	
	Records	Rate	Records	Rate	Records	Rate	Records	Rate	Records	Rate	Records	Rate
Maintenance	168	1	123	1	2496	0.400	2223	0.026	168	1	123	0.976
Washing car	1996	0	1788	0	485	0.188	659	1	1996	0	1788	0.071
Dispatching	1475	0.985	3481	0.416	976	0.981	1221	0.947	1475	0.994	3481	0.418
Waiting	123	0	1265	0.041	1535	0.111	1345	0.604	123	0	1265	0.070
In service												
- Stopping bus stops	1847	0.945	2267	0.948	5037	0.945	3991	0.944	1847	0.956	2267	0.956
- Arriving or departing	3627	0.939	3209	0.857	3939	0.906	3884	0.917	3627	0.937	3209	0.863
- Running	27023	0.989	25141	0.916	37249	0.988	38740	0.989	27023	0.988	25141	0.909
Taking a break	19474	0.999	16619	0.929	3409	0.999	3034	0.997	19474	1	16619	0.932
Out of service	1329	0.985	2987	0.609	872	0.994	673	0.987	1329	0.992	2987	0.721
Oil charging	104	1	117	1	67	1	99	1	104	1	117	1
At garage	980	0.845	595	0.420	249	0.968	194	0.974	980	0.337	595	0.237
Total	58146	0.949	57592	0.819	56314	0.922	56063	0.932	58146	0.941	57592	0.824

dispatching, service for a certain period of time, returned to garage, dispatching, service for a certain period of time, and returned to garage. This flow was different from the regular operation flow shown in Figure 2. Also, paying attention to ’Dispatching’, the correct answer rate on the 2nd day is 0.985, while the correct answer rate on the 3rd day is 0.416; the correct answer rate is greatly different. The reason for this is that the features that are a major factor in discrimination of ’Dispatching’ during normal operation are ’Total travel distance’ and ’Engine pulse’. However, on 3rd day, since the transition to the ’Dispatching’ state occurs twice, it is considered that it could not be classified by the classifier using the regular operation data set. It can be seen that the two states of ’Washing car’ and ’Waiting’ show low classification accuracy. ’Washing car’ is done in the garage. ’Maintenance’ is also done in the same garage, and the correct answer rate is 1 on both the 2nd and 3rd days. Records of ’Washing car’ are misclassified as ’Maintenance’ state, which is such a result.

B. *Multiple route classification model*

In the classification performance evaluation of multiple routes, the model was constructed using the data set of the 1st day of S-route and D-route as training data. In addition, the performance data of the classification model was evaluated using test data of S-route and D-route data set on the 2nd and 3rd days.

Focusing on ’Multiple route result (S-Route)’ and ’Multiple route result (D-Route)’ in the Table IV, it shows that the correct answer rate of ’Waiting’ is very low as compared with others as well as the result of the same route classification model. The reason for this is attributable to the fact that ’Waiting’ was misclassified as either ’In service-Stopping bus stops’ or ’Taking a break’ as described in Section IV-A. ’Maintenance’, ’Washing car’ shows that there is unevenness in the correct answer rate, and the overall correct answer rate is not stable. If the correct answer rate of either state is high, the correct answer rate of the other state tends to be low. Five operation states of ’Dispatching’, ’In service’, ’Taking a break’, ’Out of service’, and ’Oil charging’ show high correct answer rates except for the 3rd day of the D-route which was irregular operation. Moreover, the overall correct answer rate during regular operation is about the same as the correct answer rate shown in Table IV, and it is about 0.92 at the lowest. From the above, it can be said that the classification model integrating different routes shows versatility.

V. CONCLUSION

In this paper, from the background of the operation management in the transport industry in recent years, we evaluated operation states classification model constructed from the data obtained from onboard sensors.

As a result of classification in Section IV-A and Section IV-B, if it was not irregular operation, the correct answer rate was 0.92 or more. Although the correct answer rate of some operation states is low, in consideration of the burden that drivers manipulate the digital tachograph and record each time the state changes in ’In service’, proposed operation states classification model is considered useful.

REFERENCES

[1] M. Koike and M. Isogai, “Telematics Service for Commercial Vehicles to Realize Safe Traffic Society” *Fujitsu Scientific and Technical Journal*, vol.51, no.4, pp.43–47, Oct. 2015.

[2] A. Liaw and M. Wiener, “Classification and regression by randomForest.” *R News*, vol.2, no.3, pp.18–22, Dec. 2002