A Method of Driving Characteristics Recognition on Vehicle Operation Sequence

Yu Zhang[1], Bo Shen[1], Yi-Chih Kao[2], Hsin-Hung Chou[3]

[1] School of Electronic and Information Engineering, Beijing Jiaotong University, China
[2] Information Technology Service Center, National Chiao Tung University, Taiwan
[3] Department of Information Management, Chang Jung Christian University, Taiwan
{yu_zhang, bshen}@bjtu.edu.cn, ykao@mail.nctu.edu.tw, chouhh@mail.cjcu.edu.tw

Abstract

Driving behavior has been proved to have a great influence on road safety. Recognizing driving characteristic is an essential part of reducing traffic fatalities and developing intelligent traffic system. In this paper, we propose a method of driving characteristics recognition through mining vehicle operation data such as GPS, velocity and direction collected by the On-Board Diagnostic (OBD) port of vehicles. Based on the feature extracted from the vehicle operation sequence, we employ K-means algorithm to cluster and recognize different driving characteristics after reasonable normalization and dimensionality reduction of the features. Analysis and experimental results indicate that the proposed method has good application significance on mining effective information in vehicle operation data sequence.

Keywords: Driving characteristics recognition, Vehicle operation data, Data mining, Cluster analysis

1 Introduction

The coming of information age and the emergence of machine learning have greatly improved our way of study and working. In the aspect of traffic, we can improve road traffic safety by exploring the potential value of vehicle operation data, instead of expanding road and adding facilities only. Modeling and analyzing driver behavior to promote driver safety have attracted many researchers. And great progress has been made from different aspects including computer vision [1], biological technology [2], sensor information [3] and social psychology.

According to official statistics, China has 217 million automobiles and 342 million car drivers by the end of 2017. With the huge number of automobiles and car drivers, traffic accidents occur more and more frequently. A study by the American Automobile Association (AAA) Foundation for traffic safety in 2009 indicated that as many as 56% of deadly crashes between 2003 and 2007 involve one or more unsafe driving behaviors typically associated with aggressive driving [4]. The aggressive driving includes fatigue driving, speeding, improper car-following, erratic lane changing, etc., which is difficult to be recognized directly. So, how to find these driving behaviors from moving trajectories of vehicles is the key issue to predict and prevent possible dangers. In this paper, we aim at categorizing drivers by extracting characteristics from dynamic monitoring data.

On the other hand, the popularization of the IoT (Internet of things) has produced massive data. Although the mature database technology has provided technical guarantee for the storage and management of these massive data, the world has fallen into the situation of “rich data and poor knowledge [5]” due to the relatively backward technology of information extraction and information processing. Data mining is an intelligent meth-od to extract effective decision-making information and knowledge from massive data. Traditional data mining tasks include Classification Analysis, Clustering Analysis and Association Rules. Our paper mainly aims at Clus-tering Analysis. According to a given similarity measurement standard, the database records without category labels are divided into several disjoint subsets (clusters), so that the internal records of each cluster have high similarity, but low similarity among different clusters. The proper features extraction of the categories to be classified is the most critical step of unsupervised clustering without category labels.

A CAN bus is a robust vehicle bus standard designed to facilitate microcontrollers and devices communi-cating with each other without a host computer, and it carries all the necessary information to describe the state of a car [6]. When it is permitted, people can obtain the data of CAN bus through the On-Board Diagnostic (OBD) port of vehicles.

This paper is organized as follows. Section I introduces the basics of our work, including the

*Corresponding Author: Yu Zhang; E-mail: yu_zhang@bjtu.edu.cn
DOI: 10.3966/160792642019102006029

necessity of driv-ing characteristic analysis and the source of vehicle operation data we used in this paper. Section II introduces the related work of driving characteristic recognition based on vehicle operation sequence. In Section III, we present our model utilize data mining algorithm to deal with vehicle operation data. After evaluating and ana-lyzing the model in Section IV, we draw a conclusion and make discussion about future work in the last section.

2 Related Work

The study shows that the driver's driving tendency does affect the driving behavior. Dula and Ballard [7] created a research instrument to detecting dangerous driving. The examination among 119 college students showed that dangerous driving was positively correlated to traffic citations and accidents. And males are significantly more aggressive, risky, and angry driving than females according to the reporter.

There are two main approaches for analyzing driver behavior to improve road traffic safety. One relies on computer vision. Camera-based system are used extensively for lane, obstacle, and pedestrian detection at first. In [8-9], Bertozzi et al. gave a comprehensive survey of the use of computer vision in intelligent vehicles. As people realized that the monitoring of driver state is as important as the monitoring of surroundings to improve road safety. Some researchers start to analyze driver behavior through computer vision. A Looking-In and Looking-Out driver-support system was proposed in [10]. However, computer vision for driver behavior is often data-enormous and expensive. The other approach is based on vehicle operation sequence collected from various sensors. Combined with statistics, machine learning and other methods to analyze driving behavior. In this paper, we utilize the vehicle operation sequence (Longitude, latitude, velocity and direction) collected by the On-Board Diagnostic (OBD) port of vehicles to cluster driver characteristics.

At present, dangerous driving behavior recognition based on vehicle operation data is more prosperous in some developed countries. Most of these methods divide the driving behaviors in sub-tasks [11]. And the requirement of vehicle operation data is more detailed. For example, driving behaviors can be divided into fatigue driving, over speed, car-following, lane changing and curve driving, etc. Aiming at the problem of fatigue driving detection, Huang and Wang proposed a CNN method combined with image processing technology to identify facial fatigue expression [12]. In the paper of [13], a model-based approach was presented to detect unintended acceleration as well as other vehicle problems. Alireza Talebpour, Hani S. et al. even use A Game Theory approach to model lane-changing behavior in a connected environment. All this task-specific detection methods have been proved to be effective in their respective fields. However, it needs more specific experimental data, and the analysis results can only reflect a small part of driver behavior detection. Therefore, comprehensive driving behavior analysis model is required. In this paper, we classification the drives by their driving tendency which is hidden in vehicle operation data to find dangerous driving behavior.

Methods of driving characteristic recognition generally can be divided into traditional machine learning method and neural network. Kuge et al. [14] proposed a method based on HMM to characterize and detect driving behaviors. Various classifiers (SVM, Bayes Net, C4.5, K-NN, etc.) are applied to recognize different driving behaviors based on a novel physical model in [15]. Clusters are used to identify potentially aggressive driving behaviors in [16]. A new deep learning framework (DeepRSI) conducted real-time road safety prediction with considering the spatio-temporal relationship between vehicle GPS trajectories and external environment factors [17]. Liu et al. [18] proposed a visualization method to recognize different driving behavior through the deep learning of driving behavior data. However, those model-based approach usually require empirical assumptions and parameters setting. Therefore, this paper proposes an unsupervised clustering method based on partial GPS satellite data.

3 Model Description

In this section, we will introduce our cluster model in detail, including the visualization of raw data, the indicators extraction, the data preprocessing methods and theoretical of the model we adopted. This section consists of the following four parts.

3.1 Indicators Extraction on Vehicle Operation Data

For the public privacy and security, enterprise do not usually expose vehicle operation data. Limited by the access authority, we can't obtain all the information about vehicle status (fuel consumption, engine data and load data). Fortunately, it has been confirmed that most of valuable information can be mined from basic data (time, velocity, direction), which is more efficient.

The first step performed for building clustering model is driver's feature extraction. We collected drivers GPS trajectory data with the csv storage format (as show in Table 1). What we consider is to use small features of data to mine a large proportion of information. In other word, variables that carry none or little information or carry repetitive information with other variables are removed. For example, we believe that remove the ODO state is more cost-effective than consider this redundant data as a feature. On the other hand, we calculated the acceleration (α) through time (t)

and velocity (v). The relationship between velocity and acceleration can be represented as (1).

$$\alpha = \frac{\Delta(v)}{\Delta(t)} \quad \textbf{(1)}$$

Table 1. Vehicle operation data categories (one car)

License Plate Number	Direction Angle	Longitude	Latitude	ACC	Time	GPS Velocity	ODO
**	213	115.842196E	27.105588N	on	2018/8/5 2:09:16	31	5537
**	211	115.842151E	27.105520N	on	2018/8/5 2:09:17	30	5537
**	221	115.842103E	27.105460N	on	2018/8/5 2:09:18	29	5537

Intuitively, a good driver should be concentrative and a safe driving behavior should be smooth and unaggressive. The curves of velocity, acceleration are supposed to be smooth. We visualized two driver's segments of velocity and acceleration to have a sense of variation tendency as show in Figure 1. From the figure below, we can intuitively see that the maximum speed of the first driver is higher than the second driver. And we can infer that the second driver's driving stability is better than the first driver by observing their acceleration curve. Therefore, we extracted the available features from the existing data as much as possible to cluster drivers' different characteristics.

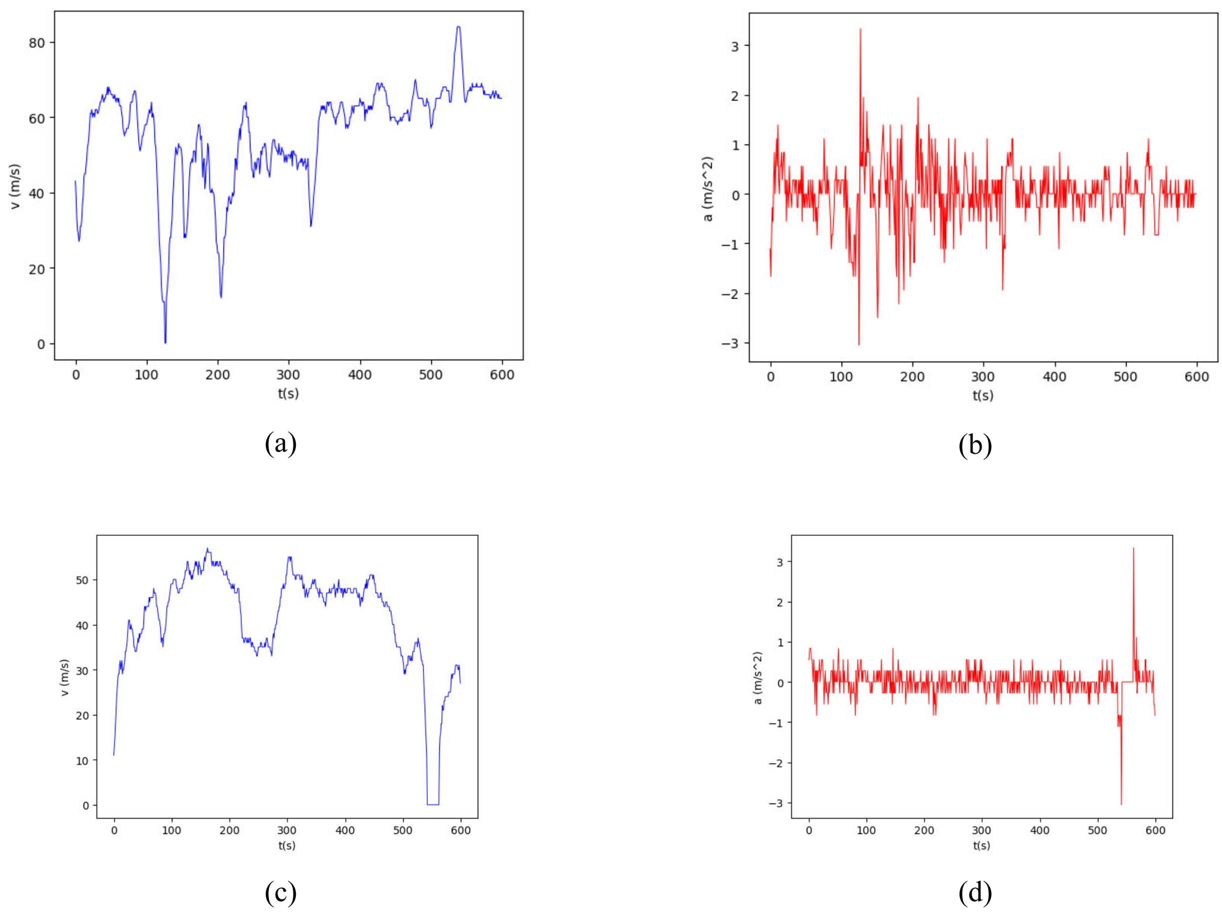


Figure 1. (a) and (b) are the visualization of the first driver's velocity and acceleration about ten minutes. (c) and (d) show the second driver's velocity and acceleration about ten minutes

Up to now, the first-level indicators include velocity, acceleration and positioning time, which can basically represent the position of a car at a certain moment. It means to represent the trajectory of the car. From these first-level indicators, we extract the second-level indicators through some simple rules that can describe driver's driving behavior as show in Table 2.

(1) Fatigue driving: Drivers' reaction will be slow when they are tired. And they cannot take corresponding driving control measures in time for different traffic conditions. So, fatigue driving is very likely to cause traffic accidents. Generally, driving a motor vehicle continuous for more than 4 hours without stopping for a rest period of less than 20

minutes is considered to be fatigue driving. We calculated the number of fatigue driving times of each driver within 24 hours. That is denoted as Fatigue in Table 2

(2) The mean and standard deviation of the velocity. These two indicators describe the driver's speed tendency and driving stability respectively. We deleted the data with velocity of 0 in the calculation process for the accuracy of results. They are denoted as $\overline{v}$ and $\sigma(v)$ respectively in Table 2.

(3) The mean and standard deviation of the velocity in night. Considering that drivers' vision is limited, but vehicles are rare in night, we calculated the mean and standard deviation of night speed respectively. The time range for the night is set at 10PM to 4AM. They are denoted as $\overline{v_n}$ and $\sigma_n(v)$ respectively in Table 2.

(4) Rapid acceleration and rapid deceleration. The rapid change of velocity not only cause uncomfortable experience to passengers in the car, but also is a hidden danger to driving safety. We set the rapid acceleration threshold of 3m/s^2, the rapid deceleration threshold of -3m/s^2 and the duration of 2s, and calculated the percentage of the rapid acceleration and deceleration in driving. The rate of rapid acceleration is denoted as a_{ac}. The rate of rapid deceleration is denoted as a_{de}.

(5) Sharp turn. It's dangerous for vehicles to pass the curve especially at a high speed. So, we count a turn when the change of internal direction angle is greater than 90 degrees for a continuous 5s, and a sharp turn when the average speed exceeds 30 m/s^2.

(6) Idle and Slide off. Prolonged idling and sliding off will damage the car's condition. In downhill, sliding off is also likely to cause brake failure. Therefore, they are dangerous driving behaviors we should consider. We calculated the percentage of idling and sliding time respectively by (2), which are denoted as Idle (%) and Slide (%) in Table 2.

$$\begin{aligned}&\text{Idling: ACC = 'on' \& } v \neq 0\\&\text{Sliding: ACC = 'on' \& } v = 0\end{aligned} \tag{2}$$

Table 2. The second-level indicators (a few cars)

id	Fatigue	$\overline{v}$	$\sigma(v)$	$\overline{v_n}$	$\sigma_n(v)$	a_{ac}	a_{de}	Sharp turn	Idle(%)	Slide off(%)
**	1	39.8336	15.4694	43.2889	16.7586	0.00703	0.00163	3	0	0.16214
**	1	45.4749	13.8518	46.1215	14.1954	0.00909	0.00210	1	0	0.20969
**	1	47.4077	14.4913	45.8099	14.1948	0.00687	0.00223	0	0	0.15408
**	1	49.1476	16.7026	49.6797	16.5001	0.00857	0.00295	4	0	0.24715
**	1	40.9582	15.2496	39.1181	14.5411	0.00669	0.00240	16	0	0.17338
**	……	……	……	……	……	……	……	……	……	……
**	2	57.5442	26.115	59.8965	26.3631	0.00546	0.00157	61	0	0.27899

3.2 Normalization on Indicators

The magnitude of the second level indicators are different. And some of the indicators are expressed in numerical data, while others are expressed in percentage. Therefore, we must normalize the indicators to make them available. Normalization of data can also speed up the search of optimal solution and improve the precision, which is very important in data mining. Considering those variables are continuous data described ten kinds of driving behaviors, we implemented improved 0-1 normalization to normalize the features. Standard 0-1 normalization transform the data into an interval between 0 and 1. However, 0-1 standardization is sensitive to discrete points. If there is a particularly large or small value, the result of 0-1 standardization will not be uniform enough. Therefore, we use the standard 0-1 normalization to transform the mean and standard deviation of the velocity, the mean and standard deviation of the velocity in night these four indicators. And the improved two-threshold normalization is carried out for other indicators. We artificially set two thresholds for some indicators according to actual situation and the visualization of each dimension. If the value of indicator is greater than the maximum threshold, it is denoted as 1. If the value of indicator is less than the minimum threshold, it is denoted as 0. The value of indicator be-tween the two thresholds is processed by standard 0-1 normalization. The thresholds of some indicators are shown in Table 3. And the normalization can be presented as (3).

$$f(x)=\begin{cases}0, & x < threshold_{min}\\ \dfrac{x-threshold_{min}}{threshold_{max}-threshold_{min}}, & threshold_{min} < x < threshold_{max}\\ 1, & x > threshold_{max}\end{cases} \tag{3}$$

Table 3. Threshold of part indicators

Indicators	Minimum threshold	Maximum threshold
Fatigue	0	2
a_{ac}	0.004	0.01
a_{de}	0.0015	0.0025
Sharp turn	0	20
Idle (%)	0.15	0.3
Slide off (%)	0.001	0.1

Thus, each driver is described by a ten-dimensional vector. Each dimension represents a feature of one driver's driving behavior.

3.3 Features Reduction

We calculated a ten-dimensional vector for each driver in the previous step. The influence of these features on the classification of drivers' driving behavior is different and uncertain. And there may be correlations between these variables, which increase the complexity of problem analysis. Therefore, it is necessary to find a reasonable method to minimize the loss of information contained in the original index while reducing the index to be analyzed, in order to achieve the purpose of comprehensive analysis of the ten-dimensional vectors. We performed the Principal Component Analysis (PCA) to reduce the dimension of the vectors.

PCA is the most prominent data dimensionality reduction algorithms. The main idea of PCA is to map the n-dimensional features to the m-dimensional ones ($m<n$). The reconstructed m-dimensional features which are orthogonal are called principal components. The calculation process of PCA includes the following steps:

(1) Calculating the covariance matrix for the normalized data **X**. Each drive has ten features, so the covariance matrix is the square matrix of 10th order.

(2) Calculating the eigenvalues and eigenvectors of the covariance matrix. In the physical, linear transformation of a high-dimensional space can be understood as the transformation of a vector in different levels and all directions. The eigenvectors of a matrix are orthogonal vectors whose linear transformation only change its length not directions. The eigenvalues express the extent of to which the length is stretched.

(3) Sorting the eigenvalues from large to small, and selected the largest m eigenvalues. The corresponding m eigenvectors are column vectors to form the matrix **W**.

(4) Calculating **Y=XW** which is to project the data set **X** to the selected eigenvectors. Matrix **Y** is the data with reduced dimensions that we need.

Generally, the features after dimensionality reduction with the cumulative variance contribution rate greater than 85% can be considered as covering the integrity of original information.

3.4 Vehicle Operation Data Clustering

In the previous step, we ended our preprocessing steps of the vehicle operation data. For each drive, an individual vector is generated to describe driver's driving characteristics. In order to recognize the unknown driving characteristics without making any unnecessary assumptions, we adopted the K-means algorithm to cluster the vector of preprocessed vehicle operation data. K-means algorithm is a typical distance-based clustering algorithm, which calculates the distance among the objects as the evaluation index of similarity and can flexibly determine clusters in an unsupervised manner. The implementation process of K-means is described as follows.

(1) K samples ($\mu_1^{(1)}$, $\mu_2^{(1)}$, ... , $\mu_k^{(1)}$) are randomly selected as the centroid of the initial K clusters.

(2) All the samples ($x_{i(i=1,\ldots,N)}$) are divided into the nearest cluster by calculating the distance between the samples and initial centroid. In this paper, we used Euclidean distance to calculate the distance between the samples and each centroid. The formula of calculating Euclidean distance is shown in (4).

$$d_{ij} = \sqrt{\left\| x_i - \mu_j^t \right\|^2} \quad \textbf{(4)}$$

where x_i is the ith sample, $i \in \{1, 2, 3, \ldots, N\}$, N is the number of samples. μ_j^t is the jth center of cluster in the tth iteration, $j \in \{1, 2, \ldots, K\}$, K is the number of clusters.

(3) Recalculating the centroids of the obtained clusters according to (5). n_j^t is the total number of samples at the jth cluster in the tth iteration. x_{ij} is the sample x_i belong to jth cluster.

$$\mu_j^{(t+1)} = \frac{1}{n_j^t} \sum x_{ij} \quad \textbf{(5)}$$

(4) Repeating the calculation to divided the samples into the nearest cluster until the centroids of those clusters are no longer changing.

So far, our system framework of driving characteristics recognition on vehicle operation data is shown in Figure 2.

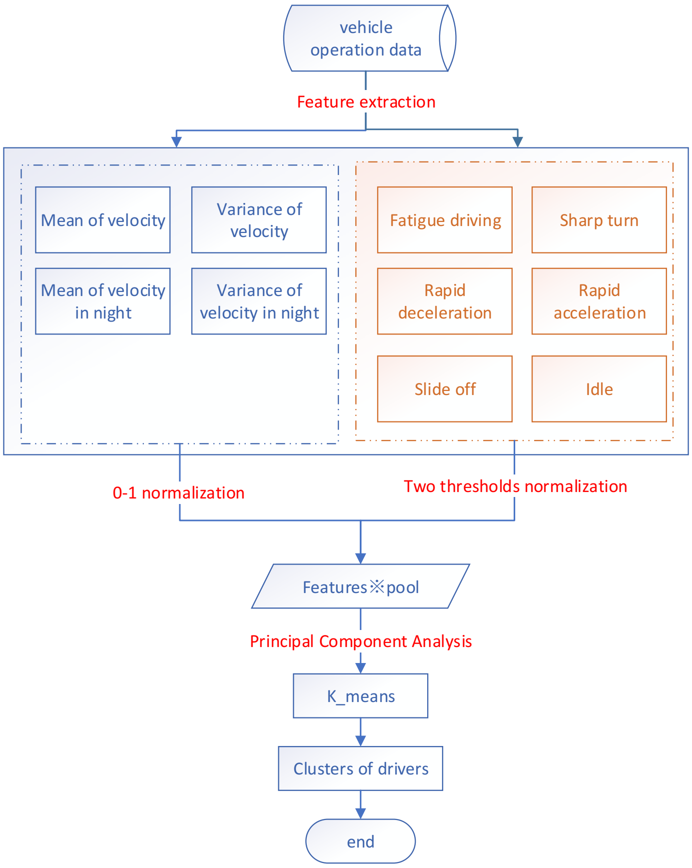


Figure 2. System framework of driving characteristics recognition

4 Experiment and Discussion

In this section, we introduce our experimental platform, experimental process and result analysis. We empirically evaluate the performance of our driving characteristics recognition model.

As a powerful Python library, NumPy has an aptitude for handling arrays. Besides, Python integrates a variety of machine learning packages which make it suitable for us to execute our experiment. Therefore, we implemented the algorithm of recognizing driver characteristic by Python.

Experiments are conducted on the GPS positioning driving sequence data of 20 drivers within 24 hours on August 7, 2018. For each driver, there are more than 20,000 points in the GPS positioning driving sequence. According to the model of feature extraction, we extracted 10 driving behaviors as the features for each driver. Features are fatigue driving, the mean and standard deviation of the velocity, the mean and standard deviation of the velocity in night, rapid acceleration, rapid deceleration, sharp turn, slide off, idle driving, etc. The extracted features are shown in Table 2. Then we normalized the ten-dimensional features according to the 0-1 normalization method or two-thresholds normalization method. After that, PCA is applicated to reduce dimension. We selected the principal components whose cumulative variance contribution rate is greater than 85%. There are three columns in total and their variance contribution rate are [0.41532659, 0.3114285, 0.12542401] respectively. In the last step, we performed K-means algorithm on the processed data for clustering analysis. We set the number of cluster K to 3 and 2, respectively, and did the experiment. The text results and visualization of results are shown in Figure 3 and Figure 4 respectively. Statistical results are shown in Table 4.

From the Table 4 Proportion of each category, we could conjecture cluster 1 and cluster 2 when $K = 3$ are cor-respond to cluster 1 when $K = 2$. In order to understand the clustering results clearly. We analyzed the samples of each category when K=3. Cluster 2 corresponds to the drivers whose driving average speed and variance are low. And all sorts of bad driving behavior are rare. Therefore, Cluster 2 can be regard as ‘gentle driver’. The driving average speed and variance of the drivers are corresponded to Cluster 0 are higher than that of Cluster 2. And the frequency of bad driving behaviors is higher. Therefore, Cluster 0 is classified as ‘normal driver’. As for drives in Cluster 1, the driving average speed and variance, all sorts of bad driving behavior are significantly increased. So, Cluster 1 is classified as ‘dangerous driver’. Our experiment is limited by the small dataset, but the algorithm has a comprehensive evaluation of driving behavior with a strong coverage ability. And the results are interpretable. We believe our experiment can be used for the supervision and warning of drivers by road traffic safety administration or vehicle information system.

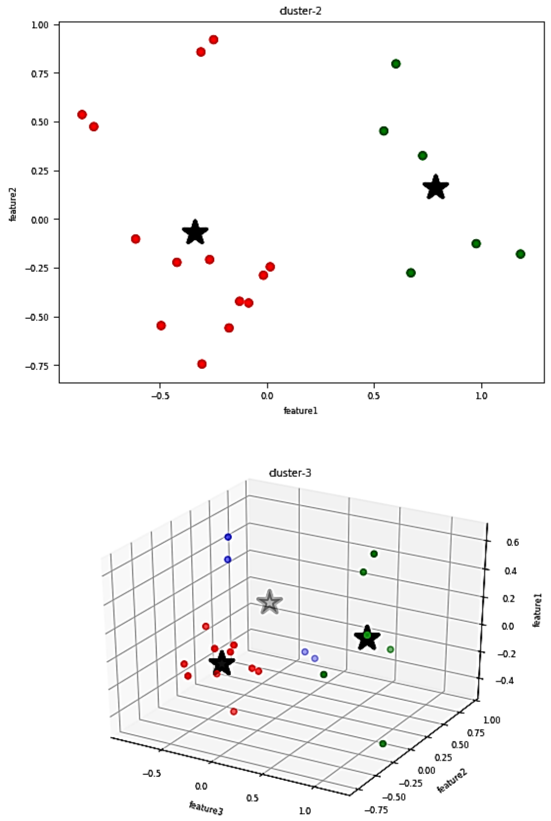


Figure 3. The visualization of clustering results. The first figure is the clustering result only used two principle components when $K = 2$, and the right one is the clustering result used all three principle components when $K = 3$

0	−0.176529	−0.559150	0.126090	0	0
1	−0.493812	−0.547005	−0.039841	0	0
2	−0.303219	−0.744318	−0.025313	0	0
3	−0.420265	−0.222739	−0.211999	0	0
4	−0.613078	−0.102796	0.062679	0	0
5	0.546730	0.451374	0.446268	1	1
6	0.978416	−0.126798	0.258052	1	1
7	0.603383	0.795586	−0.217674	1	1
8	−0.267367	−0.208530	−0.472647	0	0
9	0.673274	−0.276479	−0.018498	1	1
10	0.728121	0.324560	0.638305	1	1
11	0.015685	−0.245653	−0.113746	0	0
12	−0.127167	−0.422342	0.063144	0	0
13	−0.248756	0.919569	−0.465641	2	0
14	−0.808553	0.473361	0.495175	2	0
15	1.185565	−0.180267	−0.468367	1	1
16	−0.015965	−0.288800	−0.082038	0	0
17	−0.085075	−0.431080	0.121661	0	0
18	−0.308025	0.856422	−0.402341	2	0
19	−0.863365	0.535083	0.306732	2	0

Figure 4. The text results of clusters (Column 2, 3 and 4 are the three principal components that describe the driving characteristics of each driver. Column 5 is the clustering result when K=3. Column 6 is the clustering result when K=2.)

Table 4. Proportion of each category

	K=3	K=2
Cluster 0	0.3/ normal	0.3/ normal
Cluster 1	0.2/ dangerous	0.7/ dangerous/ gentle
Cluster 2	0.5/ gentle	—

5 Conclusion

In this paper, we proposed a method of driving characteristics recognition based on vehicle operation data. After the normalization and dimension reduction of the raw data, we applied K-means algorithm to cluster the drivers. Our algorithm provides an idea to discover driving behavior in an unsupervised way. Which is more effective for it requiring no manually labeled data. Our method also has some shortcomings. For instance, the small data set limited the accuracy of our method. Using a large amount of data is crucial for unsupervised learning to discover knowledge. Besides, we did not consider the weather, road environment, traffic conditions, etc. In the future research, we can use the latitude and longitude information mining above factors, as the indicators affecting driver classification. For further research, we will consider applying other clustering methods to analyze a large number of vehicle operation data.

Acknowledgements

This research was funded by the National Key Research and Development Program of China, grant number 2017YFC0840200, and the Fundamental Research Funds for the Central Universities, grant number 2017JBZ107.

References

[1] S. Sivaraman, M. M. Trivedi, Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 4, pp.1773-1795, December, 2013.

[2] L. Yang, R. Ma, H. M. Zhang, W. Guan, S. X. Jiang, Driving Behavior Recognition Using EEG Data from a Simulated Car-following Experiment, *Accident Analysis and Prevention*, Vol. 116, No. C, pp. 30-40, July, 2018.

[3] E. Massaro, C. Ahn, C. Ratti, P. Santi, R. Stahlmann, A. Lamprecht, M. Roehder, M. Huber, The Car as an Ambient Sensing Platform, *Proceedings of the IEEE*, Vol. 105, No. 1, pp. 3-7, January, 2017.

[4] AAA Foundation for Traffic Safety, Aggressive Driving: Research Update, http://www. aaafoundation.org, April 2009.

[5] D. C. Huang, *Data Warehouse and Data Mining Tutorial*, Tsinghua University Press, 2016.

[6] A. Sathyanarayana, S. O. Sadjadi, J. H. L. Hansen, Leveraging Sensor Information from Portable Devices

towards Automatic Driving Maneuver Recognition, *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Anchorage, AK, USA, 2012, pp. 660-665.

[7] C. S. Dula, M. E. Ballard, Development and Evaluation of a Measure of Dangerous, Aggressive, Negative Emotional, and Risky Driving, *Journal of Applied Social Psychology*, Vol. 33, No. 2, pp. 263-282, February, 2003.

[8] M. Bertozzi, A. Broggi, A. Fascioli, Vision-based Intelligent Vehicles: State of the Art and Perspectives, *Robotics and Autonomous Systems*, Vol. 32, No. 1, pp. 1-16, July, 2000.

[9] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, M. Porta, Artificial Vision in Road Vehicles, *Proceedings of the IEEE*, vol. 90, No. 7, pp. 1258-1271, July, 2002.

[10] M. M. Trivedi, T. Gandhi, J. McCall, Looking-In and Looking-Out of a Vehicle: Selected Investigations in Computer Vision Based Enhanced Vehicle Safety, *IEEE International Conference on Vehicular Electronics and Safety*, Xian, China, 2005, pp. 29-34.

[11] W. S. Wang, J. Q. Xi, H. Y. Chen, Modeling and Recognizing Driver Behavior Based on Driving Data: A Survey, *Mathematical Problems in Engineering*, Vol. 2014, No. 1, pp. 1-20, February, 2014.

[12] M. Huang, Y. Wang, Driver Fatigue Expression Recognition Research Based on Convolutional Neural Network, *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, 2018, pp. 6-9.

[13] H. Yu, R. Langari, Detection of Unintended Acceleration in Longitudinal Car Following, *SAE International Journal of Passenger Cars- Electronic and Electrical Systems*, Vol. 8, No. 2, pp. 306-313, April, 2015.

[14] N. Kuge, T. Yamamura, O. Shimoyama, A. Liu, A Driver Behavior Recognition Method Based on a Driver Model Framework, *SAE Technical Paper Series*, Vol. 2000, No. 1, pp. 339-349, March, 2000.

[15] M. Wu, S. Zhang, Y. Dong, A Novel Model-Based Driving Behavior Recognition System Using Motion Sensors, *Sensors*, Vol. 16, No. 10, Article ID 1746, October, 2016.

[16] J. Lee, K. Jang, A Framework for Evaluating Aggressive Driving Behaviors Based on In-vehicle Driving Records, *Transportation Research Part F: Traffic Psychology and Behaviour*, https://trackandknowproject.eu/wp-content/uploads/simple-file-list/Track-and-Know-relevant-literature-and-research-papers/WP4-Big-Data-Analytics/4.1.-Knowledge-Discovery-in-Big-Data/4.1.7.-Driver-profiling/A-framework-for-evaluating-aggressive-driving-behaviors-based-on-in-vehicle-driving-records.pdf.

[17] Z. Peng, S. Gao, Z. C. Li, B. Xiao, Y. Qian, Vehicle Safety Improvement through Deep Learning and Mobile Sensing, *IEEE Network,* Vol. 32, No. 4, pp. 28-33, July, 2018.

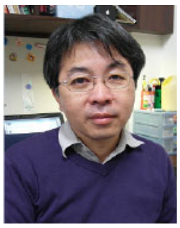
[18] H. Liu, T. Taniguchi, Y. Tanaka, K. Takenaka, T. Bando, Visualization of Driving Behavior Based on Hidden Feature Extraction by Using Deep Learning, *IEEE Transactions on Intelligent Transportation Systems,* Vol. 18, No. 9, pp. 2477-2489, September, 2017.

Biographies

Yu Zhang received the B.S. degree in communication engineering from Beijing Jiaotong University in 2017. Since 2018, she has been studying for her doctor's degree in information and communication engineering at School of Electronic and Information Engineering. Her research interesting includes data mining and NLP techniques such as sentiment analysis.

Bo Shen received the Ph.D. degree in communication and information systems from Beijing Jiaotong University (BJTU), Beijing, China, in 2006. He has been a Professor in BJTU. Pro Shen has published about 40 professional research papers. His research interests include cyber security, Information network, and electronics and communication engineering.

Yi-Chih Kao is currently the director of the Network and System Division of Information Technology Service Center at National Chiao Tung University (NCTU). He received his Ph.D. degree in Industrial Engineering and Management from NCTU. His research interests include cyber security, network performance, software-defined networking, and IT service design.



Hsin-Hung Chou is an assistant professor at the Department of Information Management, Chang Jung Christian University. He received the Ph.D. degree in Computer Science and Information Engineering from National Taiwan University at 2004. His research interests include Computer Game, Machine Learning, Algorithm Design, Cloud Computing, and so on.