A Learning-Based Driving Style Classification Approach for Intelligent Vehicles

Peng Mei[1], Hamid Reza Karimi[2(✉)], Cong Huang[3], Shichun Yang[1(✉)], and Fei Chen[1]

[1] School of Transportation Science and Engineering, Beihang University, Beijing, China
yangshichun@buaa.edu.cn
[2] Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy
hamidreza.karimi@polimi.it
[3] School of Transportation and Civil Engineering, Nantong University, Nantong, China

Abstract. Driving behavior is crucial to the energy consumption analysis of electric vehicles. This paper proposes an unsupervised learning method to classify driving behavior for three typical road conditions. First, three specific road conditions are selected from the open access data, including characteristic information such as speed and acceleration. Besides, the characteristic data is processed, so each distinct value has the same weight. Second, two unsupervised learning clustering algorithms are introduced and compared in typical working conditions. Finally, the clustering results under three working conditions are obtained. Specifically, we can classify driving styles in high-speed conditions into aggressive, standard, and calm; besides, the classification method of K-medoids is more advantageous. In intersection conditions, driving styles are usually divided into standard and calm. Considering the calculation time and other factors, the K-means algorithm shows superior effects compared to the K-medoids algorithm. The driving style can be divided into standard and calm in campus conditions. In this case, K-medoids have a more significant advantage. The research results have implications for the classification of driving styles under different road conditions.

Keywords: Driving style classification · unsupervised learning · intelligent vehicles

1 Introduction

Electric vehicles (EVs) are the most widely used carrier of intelligent network technology in the automotive industry [1]. Still, their shortcomings, such as short driving range and long charging time, have become the bottleneck for developing pure electric vehicles. Under the existing technical background, combining the advantages of network technology, developing estimation algorithms with practical value, and improving the prediction accuracy of the driving range are the primary means to alleviate the "range anxiety". Many scholars and researchers have conducted in-depth research on the "range anxiety" problem caused by the poor accuracy of EV's cruising range estimation. Many factors affect the driving range, among which the main elements are the state estimation of the power battery, the driver's driving style, and future driving conditions.

H. R. Karimi and N. Wang (Eds.): S-Cube 2022, LNICST 487, pp. 179–190, 2023.
https://doi.org/10.1007/978-3-031-34899-0_12

Driving style is an inherent attribute accumulated by drivers when driving on the road for a long time, and different drivers have similarities and differences in their driving styles [2]. Elassad et al. [3] proposed the "driver-vehicle-environment" framework, which can explain the factors affecting driving behavior. First, the driver's age, gender, and personality directly impact driving style. Second, external environmental factors such as road type, traffic conditions, and weather can also indirectly impact driving behavior. Finally, car features such as head-up displays and in-vehicle aids can also affect driving behavior. These effects ultimately fall into three categories: driving events (turning or following), physiological states (fatigue and distraction), and psychological states.

Much experimental research have been carried out on driving styles, and the classification methods can be divided into rule-based, model-based, and machine-learning methods [2]. The principle of the rule-based method is to artificially give a threshold for the factors that affect the determination of driving behavior. When the value of an indicator exceeds the point, the driving style can be determined as a specific category. First, the acceleration information of the experimental vehicle was recorded, and the more frequent the longitudinal or lateral acceleration changes were obtained from the experimental data, the more aggressive the driving behavior style. Then the acceleration probability was used as the driving style. The style classification indicator divides driving styles into three categories: calm, moderate, and aggressive, with an overall accuracy rate of 68.49% [7]. Bejani et al. [8] proposed a driving style evaluation system based on environment perception and designed a rule-based fuzzy controller to classify the driving style, which proved the experiment's reliability. The fuzzy controller becomes the first choice for rule-based classification when there are many input characteristic parameters. Filev et al. [9] used fuzzy logic to evaluate factors such as acceleration change rate, speed change rate, and steering angle as cautious and moderate, more radical and radical. However, the fuzzy control logic mostly depends on the driver's experience, can only cover some of the situations well, and the accuracy rate needs to be improved.

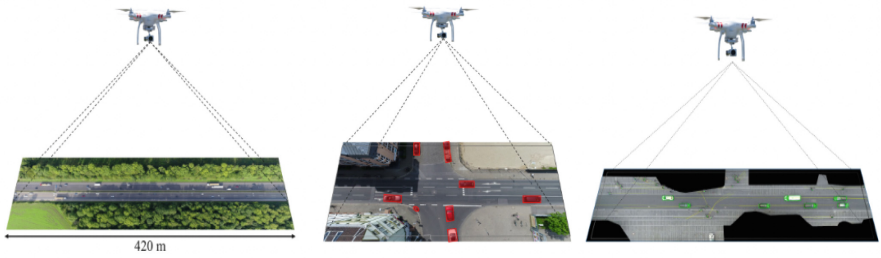
Vaitkus et al. [4] used the supervised learning algorithm of KNN to extract five main features from all the data. In a specific route, the classification accuracy rate was as high as 100%; extracting three main parts, the classification accuracy rate also reached 98%. Since supervised learning methods usually need to label the training data accurately, these labels are difficult to obtain in practical applications, making applying supervised learning difficult, and it takes much time to mark the data manually. To solve the above problems, Wang et al. [5] proposed an SVM method for semi-supervised learning, which divides driving styles into aggressive and standard types according to several labels. Mohammadnazar et al. [6] used the relevant information generated by the Internet of Vehicles technology to measure the instantaneous driving behavior and used unsupervised learning to classify the driving styles of different types of roads.

Meanwhile, to quantify the driving style, information such as speed, longitudinal acceleration, and lateral acceleration are extracted while the car is driving. K-means and K-medoid methods classify drivers into aggressive, regular, and calm types. The study shows that the thresholds of the evaluation indicators for drivers of different road types are also different, and the proportion of vehicles driving in urban areas is higher than that of drivers in high-speed road conditions.

To accurately predict the subsequent driving range, the content of this paper is mainly aimed at the driving style clustering problem of unsupervised learning organized as follows; the second section introduces the data sources and the eigenvalues of the data in different scenarios. In Sect. 3, the unsupervised learning approach, containing k-means and k-medoids, is presented. The two different methods are compared in the different conditions in Sect. 4. In the last section, all work is summarized.

2 Data

The classification results of driving styles are different under different road conditions. A driver who behaves aggressively in high-speed road conditions is likely to act as an average driver in an urban environment full of light traffic intersections. Therefore, this chapter conducts a driving style analysis for three typical driving conditions, including highways, campus conditions, and urban hubs. Many scholars have shared open access data online to study better driving characteristics analysis, such as Next Generation Simulation, the Highway Drone Dataset, and the Interaction dataset. Nevertheless, most datasets are based on highway conditions and are not universal. Therefore, the data set selected in this paper includes data sets such as high-speed, intersection, and intersection, given in Fig. 1 [10, 11].



(1) The bird's eye view of the highway road. (2) The bird's eye view of the intersection. (3) The bird's eye view of the university road.

Fig. 1. The bird's eye view of three typical driving road.

The highD dataset is a new dataset of natural vehicle trajectories recorded on German highways [11]. Traffic flows were recorded at six locations, including more than 110500 cars. The course of each vehicle is extracted, including vehicle type, size, and maneuverability. Because the dataset includes different kinds of vehicles, such as cars and trucks, and the dataset is huge. Therefore, it is necessary to filter the data first and select the car's driving data recorded at the exact high-speed location. Since the length of the expressway is 410 m and records about 13 s, the amount of information on each vehicle is limited, which is conducive to analyzing driving behavior.

The inD dataset is a new dataset of natural vehicle trajectories recorded at German intersections [10]. The data type of the intersection is different from that of the expressway. The data on bicycles and pedestrians are added, and the working conditions are more complicated. The maximum speed limit on the road is 13.8 m/s. We first filter out

the car's data and then process the data to obtain the vehicle's average speed, lateral acceleration, and longitudinal acceleration during the drone shooting process, and then comprehensively analyze the driving style. Similarly, the UniD dataset is a new dataset of natural road user trajectories recorded on the campus of RWTH Aachen University, and we apply the same approach.

3 Methodology

Since the above data has no labels, it is a challenge to classify the trajectory data of cars by traditional methods; we need to use the clustering algorithm to organize the data. As a typical algorithm in unsupervised learning, the clustering algorithm can divide similar sample data into a specific category; the common ones are K-means and K-medoids, and the two algorithms will be introduced.

3.1 K-means

The main idea of the K-means algorithm [12] is that given the K value and K initial cluster center points, each data point is divided into clusters represented by the nearest cluster center point. After all the points are allocated, recalculate the center point of the group according to all the points in the cluster. Then iteratively performs the steps of assigning points and updating the cluster center points until the change of the cluster center points is minimal or the specified number of iterations is reached. The flow of the K-means algorithm is shown in Table 1. The minimize error function is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} d(x, \mu(C_i)) \tag{1}$$

where $C_1, \ldots, C_K$ mean the k clusters, $\mu(C_i)$ is the centroid of cluster C_i, and $d(x, \mu(C_i))$ represents the distance between the observation x and $\mu(C_i)$. The Euclidean Distance d from x to μ is calculated using the equation below:

$$d = \sqrt{\sum_{k=1}^{n} (x_k - \mu_k)^2} \tag{2}$$

3.2 K-medoids

Since the center point of K-means is located at an arbitrary value in the continuous space, it is sensitive to noise. Unlike K-means, K-medoids can only take a particular data point in the dataset as the center point [13]. The flow of the K-medoids algorithm is shown in Table 2. The error function of K-medoids is as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} (x - m(C_i)) \tag{3}$$

The silhouette coefficient is an effective index proposed to solve the validity of the clustering results, and the following formula can express it [14]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & a(i) < b(i) \\ 0 & a(i) < b(i) \\ \frac{b(i)}{a(i)} - 1 & a(i) < b(i) \end{cases} \tag{5}$$

where $a(i)$ represents the average distance from sample i to other samples; $b(i)$ means the average distance from sample i to other clusters C. As the silhouette coefficient approaches 1, the sample clustering becomes more reliable.

In addition, among the selected eigenvalues, the average velocity value is much larger than the acceleration. Therefore, in the cluster analysis, the weight of the speed will account for a higher weight by default. In this case, we need to normalize the extracted eigenvalues with the following formula [15]:

$$x_{scale} = \frac{x - x_{min}}{x_{max} - x_{\min}} \tag{6}$$

where x_{scale} means the normalized value, x denotes the original feature data; x_{min} and x_{max} are the minimum and maximum values of the original feature data, respectively.

Table 1. The flow of the K-means algorithm.

Input: dataset D=$\{x_1, x_2, \ldots, x_n\}$
Step 1: Randomly select k initial center points in D;
Step 2: Divide other data points into clusters with the smallest distance from a center point;
Step 3: Recalculate the center point according to the distance from the data point in each cluster to the center point;
Step 4: Reassign data points to the closest clusters according to the obtained center points;
Step 5: Repeat (3) and (4) until the data points for each cluster no longer change;
Output: Clusters C=$\{C_1, C_2, \ldots, C_k\}$

Table 2. The flow of the K-medoids algorithm.

Input: dataset D=$\{x_1, x_2, ..., x_n\}$
Step 1: Randomly select k initial center points in D; Step 2: Divide other data points into clusters with the smallest distance from a center point; Step 3: Calculate the error function of formula (3); Step 4: Randomly select a non-center point data point from D, and compute the error function each time, assuming that one of the current center points is exchanged for the selected non-center point data; Step 5: Repeat (3) and (4) until the data points for each cluster no longer change;
Output: Clusters C=$\{C_1, C_2, ..., C_k\}$

4 Results and Discussion

4.1 Highway Road

We screened characteristic data for high-speed driving, including average vehicle speed, lateral acceleration, and average lateral acceleration. The original data points are shown in Fig. 2, the average speed ranges from 20 m/s to 50 m/s in the highway condition. We apply K-means and K-medoids approaches to divide the data points into clusters, and the results are given in Fig. 3. To select the optimal number of clusters, we choose the K value with the highest SC value as the clustering; the comparison chart is depicted in Fig. 4, where the SC value of K-means is 0.7584, and the SC value of K-medoids is 0.7616. Following our assessment, the K-medoids method is the optimal choice for

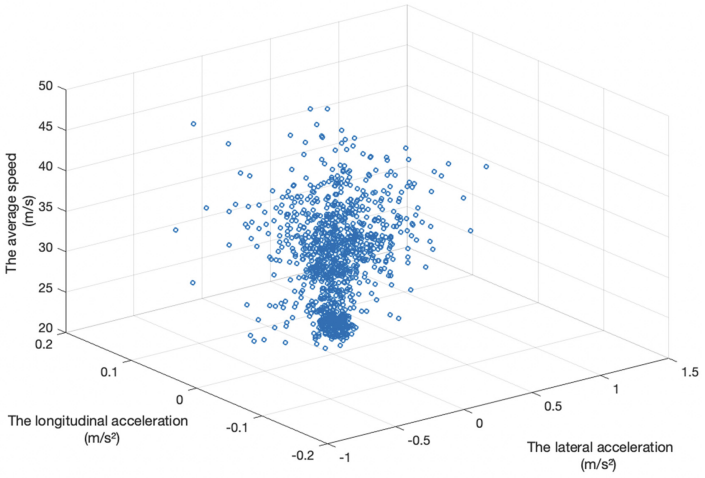


Fig. 2. Filtered feature data points in highway condition.

high-velocity operations. In this case, it makes more sense to divide driving styles into three categories.

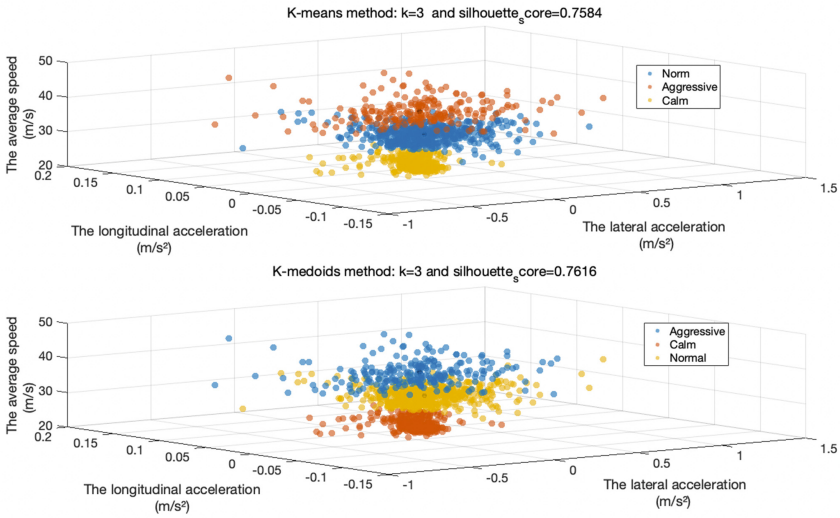


Fig. 3. Cluster analysis comparison in highway condition.

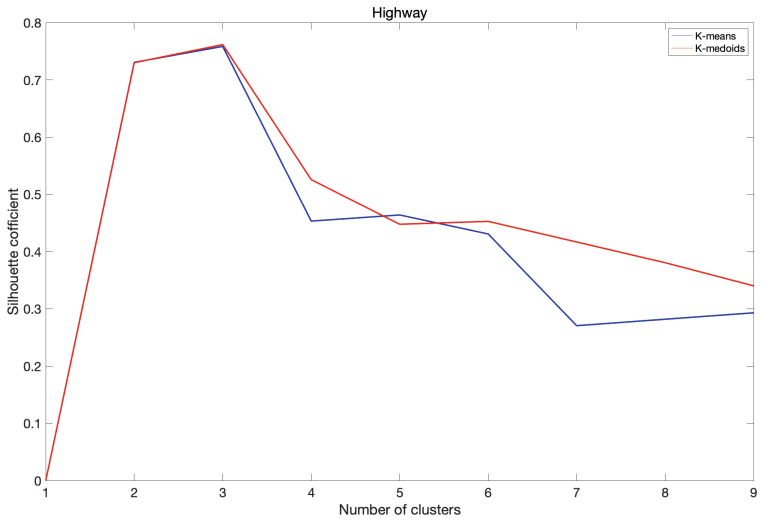


Fig. 4. Comparison of K-means and K-medoids clustering performance using SC in highway condition. (Color figure online)

4.2 Intersection Road

Unlike highways, vehicles at intersections are more cautious; the data points are shown in Fig. 5. Using the same method, we get that when k = 2, the SC value is the highest. The clustering results are given in Fig. 6, the red dots in the graph indicate a standard driving style, and the blue ones show a cautious driving style. The comparison results are depicted in Fig. 7; the red line indicates the relationship between the silhouette coefficient and the number of clusters using the K-medoids method; the blue line indicates the K-means method. It can be seen that when the number of groups is 2, the silhouette factors of the two ways reach the highest value. Moreover, the K-means method is more suitable for driving style cluster analysis in the intersection road.

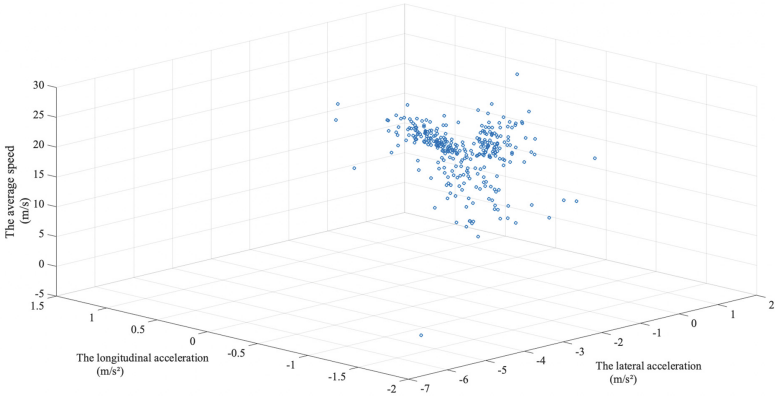


Fig. 5. Filtered feature data points in intersection condition.

4.3 University Road

The campus road conditions have more pedestrians and bicycles and relatively few motor vehicles. The data distribution is depicted in Fig. 8. Similarly, when k = 2, the SC value is the highest, the clustering results are shown in Fig. 9. It can be seen from the figure that the classification effect of the K-means method is different from that of the K-medoids method. Blue dots indicate a calm driving style, and red dots indicate a regular driving style. The comparison results are given in Fig. 10; the silhouette factors of the two ways reach the highest value in the two cluster conditions. Moreover, the K-medoids method is more suitable for driving style cluster analysis on the university road.

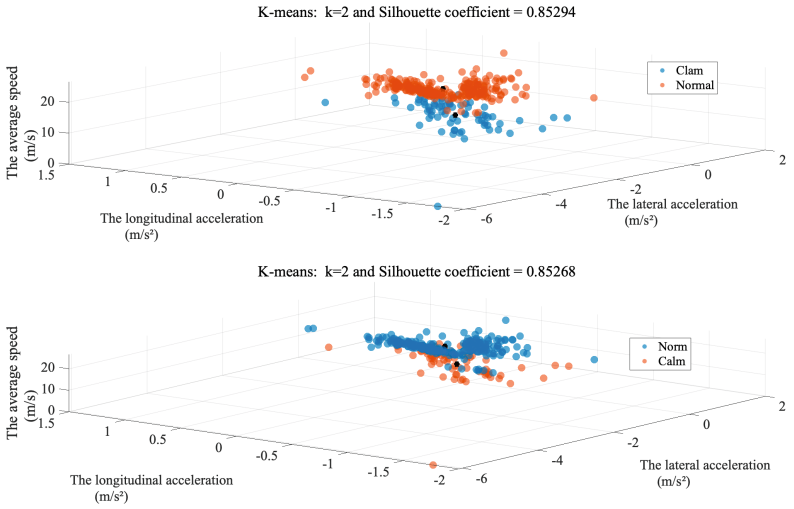


Fig. 6. Cluster analysis comparison in intersection condition. (Color figure online)

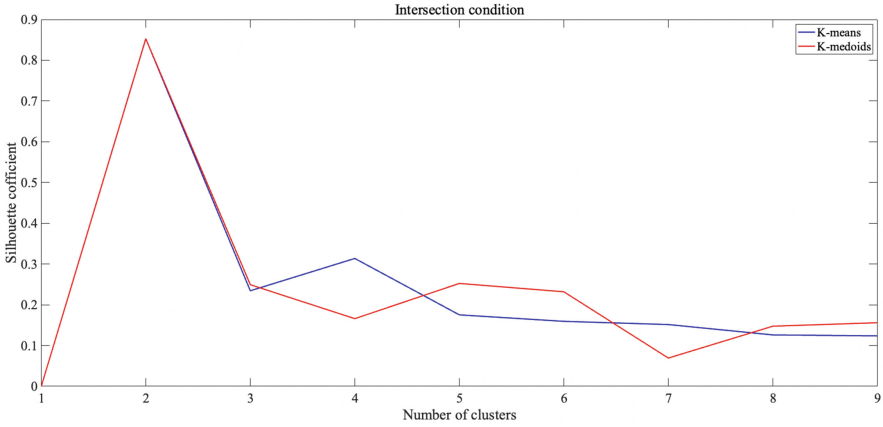


Fig. 7. Comparison of K-means and K-medoids clustering performance using SC in intersection condition. (Color figure online)

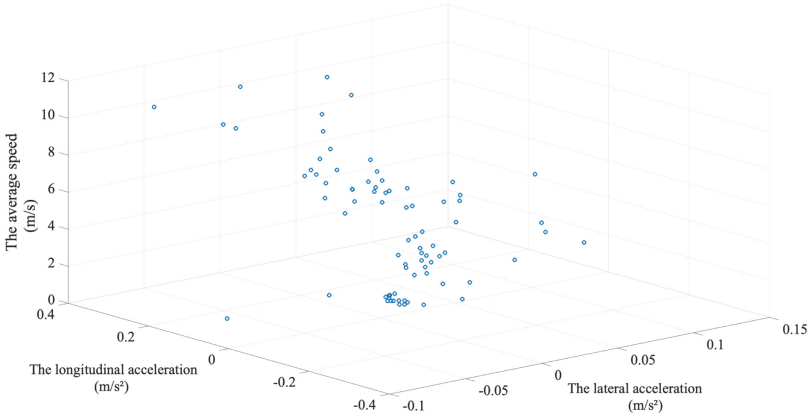


Fig. 8. Filtered feature data points in campus condition.

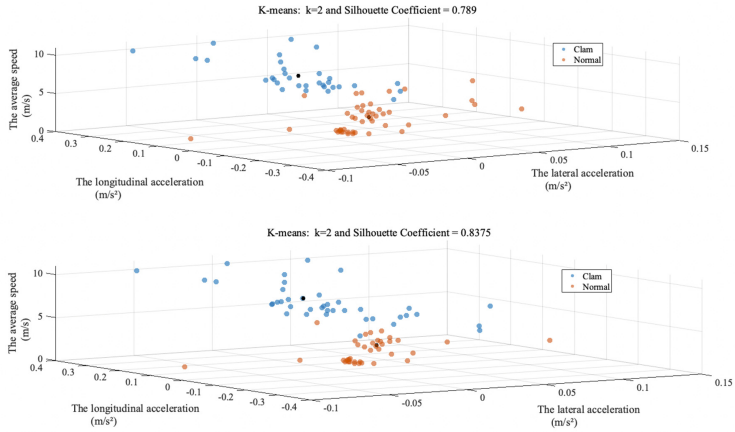


Fig. 9. Cluster analysis comparison in campus condition.

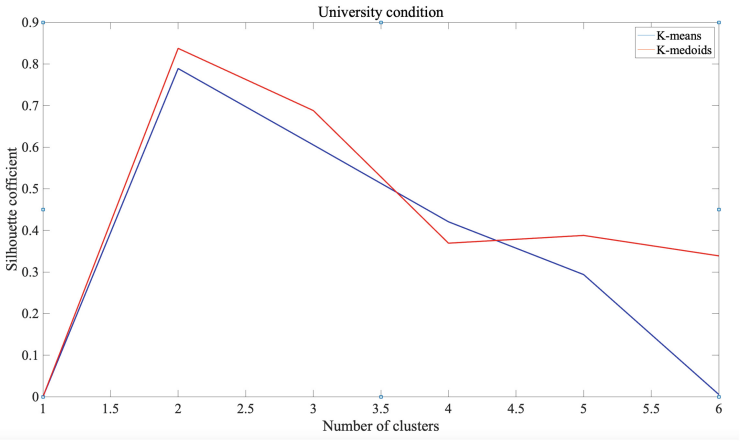


Fig. 10. Comparison of K-means and K-medoids clustering performance using SC in campus condition. (Color figure online)

5 Conclusion

Due to different road conditions, the classification methods of driving styles are also different. Based on the experimental data of three road conditions, this paper uses different unsupervised learning algorithms to cluster and analyze them. We can classify driving styles in high-speed conditions into aggressive, standard, and calm. The classification method of K-medoids is more advantageous; its silhouette factor is 0.7616. In intersection conditions, driving styles are usually divided into standard and calm. The effects of K-means and K-medoids are not much different; both silhouette factors are nearly 0.853. Considering the calculation time and other factors, we recommend the K-means algorithm. In campus conditions, the driving style can be divided into two categories: standard and calm, and in this case. The K-medoids approach has a more significant advantage; its silhouette factor is 0.8375. The research results have implications for the classification of driving styles under different road conditions. In the follow-up research, we will use supervised learning to identify driving behaviors based on the labeled data classified by unsupervised learning.

Acknowledgements. This work is supported by the National Key Research and Development Program of China (2021YFB2501705), partly by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (Grant No. 22KJB510040), and the Basic Science Research Program of NanTong City (Grant No. JC12022028).

References

1. Mei, P., et al.: An adaptive fuzzy sliding-mode control for regenerative braking system of electric vehicles. Int. J. Adapt. Control Signal Process. **36**(2), 391–410 (2022)

2. Martinez, C.M., Heucke, M., Wang, F.Y., et al.: Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey. IEEE Trans. Intell. Transp. Syst. **19**(3), 666–676 (2017)
3. Abou Elassad, Z.E., Mousannif, H., Al Moatassime, H., et al.: The application of machine learning techniques for driving behavior analysis: a conceptual framework and a systematic literature review. Eng. Appl. Artif. Intell. **87**, 103312 (2020)
4. Vaitkus, V., Lengvenis, P., Žylius, G.: Driving style classification using long-term accelerometer information. In: 2014 19th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 641–644. IEEE (2014)
5. Wang, W., Xi, J., Chong, A., et al.: Driving style classification using a semisupervised support vector machine. IEEE Trans. Human-Mach. Syst. **47**(5), 650–660 (2017)
6. Mohammadnazar, A., Arvin, R., Khattak, A.J.: Classifying travelers' driving style using basic safety messages generated by connected vehicles: application of unsupervised machine learning. Transp. Res. Part C: Emerg. Technol. **122**, 102917 (2021)
7. Jardin, P., Moisidis, I., Zetina, S.S., et al.: Rule-based driving style classification using acceleration data profiles. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6. IEEE (2020)
8. Bejani, M.M., Ghatee, M.: A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data. Transp. Res.Part C: Emerg. Technol. **89**, 303–320 (2018)
9. Filev, D., Lu, J., Prakah-Asante, K., et al.: Real-time driving behavior identification based on driver-in-the-loop vehicle dynamics and control. In: 2009 IEEE International Conference on Systems, Man and Cybernetics, pp. 2020–2025. IEEE (2009)
10. Bock, J., Krajewski, R., Moers, T., et al.: The ind dataset: a drone dataset of naturalistic road user trajectories at German intersections. In: 2020 IEEE Intelligent Vehicles Symposium (IV), pp. 1929–1934. IEEE (2020)
11. Krajewski, R., Bock, J., Kloeker, L., et al.: The highd dataset: a drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2118–2125. IEEE (2018)
12. Krishna, K., Murty, M.N.: Genetic K-means algorithm. IEEE Trans. Syst. Man Cyber. Part B (Cyber.) **29**(3), 433–439 (1999)
13. Park, H.S., Jun, C.H.: A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. **36**(2), 3336–3341 (2009)
14. Aranganayagi, S., Thangavel, K.: Clustering categorical data using silhouette coefficient as a relocating measure. Int. Conf. Comput. Intell. Multimedia Appl. (ICCIMA 2007) IEEE **2**, 13–17 (2007)
15. Eck, N.J., Waltman, L.: How to normalize cooccurrence data? An analysis of some well-known similarity measures. J. Am. Soc. Inform. Sci. Technol. **60**(8), 1635–1651 (2009)