Collaborative driving style classification method enabled by majority voting ensemble learning for enhancing classification performance

Yi Guo, Xiaolan Wang*, Yongmao Huang, Liang Xu

School of Electrical and Electronic Information, Xihua University, Chengdu, China

* wangdxhu@163.com

Abstract

The classification of driving styles plays a fundamental role in evaluating drivers’ driving behaviors, which is of great significance to traffic safety. However, it still suffers from various challenges, including the insufficient accuracy of the model, the large amount of training parameters, the instability of classification results, and some others. To evaluate the driving behaviors accurately and efficiently, and to study the differences of driving behaviors among various vehicle drivers, a collaborative driving style classification method, which is enabled by ensemble learning and divided into pre-classification and classification, is proposed in this paper. In the pre-classification process, various clustering algorithms are utilized com-positely to label some typical initial data with specific labels as aggressive, stable and con-servative. Then, in the classification process, other unlabeled data can be classified accurately and efficiently by the majority voting ensemble learning method incorporating three different conventional classifiers. The availability and efficiency of the proposed method are demonstrated through some simulation experiments, in which the proposed col-laborative classification method achieves quite good and stable performance on driving style classification. Particularly, compared with some other similar classification methods, the evaluation indicators of the proposed method, including accuracy, precision, recall and F-measure, are improved by 1.49%, 2.90%, 5.32% and 4.49% respectively, making it the best overall performance. Therefore, the proposed method is much preferred for the autono-mous driving and usage-based insurance.

Introduction

The occurrence of traffic accidents was directly related to the driver’s operation, and bad driving styles lead to more traffic accidents [1]. According to data released by China’s Ministry of Public Security, accident deaths caused by improper driving account for 88.91% of all deaths [2]. Similarly, data released by the National Highway Traffic Safety Administration (NHTSA) shows that subjective driver error is responsible for 94% of all crashes [3]. Therefore, an
An effective driving style classification model is of great significance for driving data analysis and traffic safety. Analysis of driving behaviors and classification of driving styles not only can prevent traffic accidents effectively, but also can be applied to human-centric vehicle control systems [4], intelligent transportation systems [5], and power management for electric vehicles [6]. A driving style classification model, which can classify driving styles more efficiently is proposed in this paper.

The driving style mirrors the driver’s personalized vehicle operation mode, including driving speed, concentration level, vehicle control strategy, and so on, which can well reflect the driver’s individual driving characteristics [7, 8]. Lots of approaches had been studied to recognize driving style of unlabeled data in previous research, which could be roughly categorized into two groups: clustering-based and classification-based. Clustering-based methods mainly include K-means [9, 10], The density-based spatial clustering of applications with noise (DBSCAN) [11], agglomerative hierarchy [12], fuzzy c-means (FCM) [13], and some others. For example, in term of 10 driving characteristics, K-means algorithm was applied to classify drivers into calm type, normal type and aggressive type by applying in [9]. The framework of clustering-based was shown in Fig 1. The research method based on classification was a more in-depth study based on clustering method, the framework of it was shown in Fig 2. In this framework, the results of cluster analysis were used to train and test the classification algorithm. The main classification methods include various neural networks [14–18], decision tree [19], random forest (RF) [20], extreme gradient boosting (XGB) [21], support vector machine (SVM) [22, 23], Bayes classifier [24, 25], AdaBoost [26], and Dempster-Shafer (D-S) evidence theory [27]. Driving style classification could be realized by conventional methods, but there were still many limitations: (1) The clustering-based methods have a problem of re-cluster analysis for newly added data, as new data were generated, it needed to re-analyze the whole data set. (2) Bayes and neural networks belong to the conventional statistical learning classification methods. A large number of training samples were required for them. The larger the number of samples, the closer the test results will be to the real results, which was rare in practical application. (3) Although the decision tree and SVM were suitable for classifying small...
number samples, the results of a single classifier were unstable and the model was easy to fall into overfitting. Generally speaking, the conventional driving style classification models were generally plagued by low accuracy rate of classifications, poor robustness of models, high complexity of algorithms, and single evaluation indicator of results. In solving the above problems, ensemble learning had a good performance, and it has been proved widely in other fields [28, 29]. In the field of vehicle driving safety, an ensemble method based on CNN and GRU was proved that it could increases the detection performance of attack detection and subsequently brings about improved detection performance [30]. To enhance the efficiency and accuracy of the classification models, a collaborative driving style classification method enabled by majority voting ensemble learning was proposed in this paper. It was enabled by ensemble learning and divided into pre-classification and classification. The ensemble learning was adopted in both pre-classification and classification processes. In the pre-classification process, FCM and spectral clustering (SC) were utilized to initially classify the driving behavior data and labeled some typical initial data with specific labels as aggressive, stable and conservative. Then, the labeled data were used to train the classification algorithms. Finally, the majority voting ensemble method incorporating classification and regression tree (CART), SVM, and K-nearest neighbor (KNN) were used to classify other unlabeled data. Compared with other ensemble learning methods, such as RF and AdaBoost, the proposed method’s accuracy, precision, recall and F-measure were improved by 1.49%, 2.90%, 5.32% and 4.49% respectively. What’s more, the proposed method had better generalization performance and efficiency than some other machine learning methods under the same data set.

The remainder of this paper is structured as follows: Section II introduces some theoretical background. The majority voting ensemble method is detailed in Section III, following with the experiment results in Section IV. Section V discusses the results, with the conclusions given finally.

Guidelines for manuscript preparation

Evaluation indicator

In the final evaluation, we comprehensively evaluate the model from two aspects: the rationality of the classification results and the validity of the classifier. Comprehensive evaluation of the classification results' rationality is carried out by introducing some clustering evaluation indicators, including internal and external effectiveness indicators. The internal effectiveness indicator is mainly based on the structural information of the data set, and the cluster partition is evaluated from the aspects of compactness, separability, connectivity and overlap. The external effective indicator means that when the external information of the data set is available, the performance of different clustering algorithms can be evaluated by comparing the matching degree between clustering division and the external criteria. Moreover, two internal validity indicators, namely the Davies-Bouldin index [31] and Calinski-Harabasz index [32] are utilized to evaluate the results here owing to the data used for experiment being unlabeled.

The Davies-Bouldin index uses the distance from the sample point in the cluster to its cluster center to estimate the tightness within the cluster, and the distance between the cluster centers represents the separation between clusters. The smaller the Davies-Bouldin Index is, the better the classification effect is. Davies-Bouldin Index is defined as:

\[
DBI = \frac{1}{K} \sum_{i=1}^{K} R_i
\]
Where $k$ is the number of clusters and $R_i$ is defined as:

$$R_i = \max_{j \neq i} R_{ij}$$  \hspace{1cm} (2)

Where $R_{ij}$ is the similarity measure between cluster $C_i$ and $C_j$, and is defined as:

$$R_{ij} = \frac{S_i + S_j}{|m_i - m_j|}$$  \hspace{1cm} (3)

Where $S_i$ is the standard error between the sample point and $m_i$ in the $C_i$, and is defined as:

$$S_i = \sqrt{\frac{1}{|C_i|} \sum_{x \in C_i} |x - m_i|^2}$$  \hspace{1cm} (4)

Where $|C_i|$ is the number of data in cluster $C_i$ and $m_i$ is the center of $C_i$.

The Calinski-Harabasz index measures the tightness of the cluster by calculating the square sum of the distance between each point in the cluster and the center of the cluster, and measures the separation of the data set by calculating the square sum of the distance between various centers and the data center point. The Calinski-Harabasz is the ratio of the distance between the clusters to the distance within one cluster [33]. The Calinski-Harabasz index is defined as:

$$CH = \frac{T_r(S_B) \times (K - 1)}{T_r(S_w) \times (n - K)}$$  \hspace{1cm} (5)

Where $n$ is the number of clusters, $K$ is the current cluster, and the $T_r(S_B)$ and $T_r(S_w)$ are defined as:

$$T_r(S_B) = \sum_{i=1}^{K} n_i \times d(v_i, v)$$  \hspace{1cm} (6)

$$T_r(S_w) = \sum_{i=1}^{K} \sum_{j=1}^{n} d(x_j, v_i)$$  \hspace{1cm} (7)

$T_r(S_B)$ is the trace of the dispersion matrix between clusters, $T_r(S_w)$ is the trace of the dispersion matrix within the cluster, $n_i$ represents the number of clusters, $K$ represents the current class, and $v$ represents the center of mass of the entire dataset, $v_i$ represents the center of class-$i$. The larger the Calinski-Harabasz index is, the closer the cluster itself is and the more dispersed it is between clusters, which is the better clustering result.

To compare the effectiveness of the classifiers, accuracy, precision, recall and F-measure are used to evaluate the classification model. The indexes are defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$  \hspace{1cm} (8)

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (9)

$$\text{Recall} = \frac{TP}{TP + FN}$$  \hspace{1cm} (10)
\[ F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]  

(11)

Where TP is the number of samples belonging to Class C that were correctly classified into class C. FP is the number of samples that did not belong to Class C that were misclassified into class C. TN is the number of samples that do not belong to category C and are correctly classified into other classes. FN is the number of samples belonging to category C that were misclassified into other classes.

**Driving style classification model based on majority voting ensemble learning**

For strengthening the generalization ability, accuracy, and stability, some improvements have been made to model 2 in Fig 2. The main improvement is that the majority voting ensemble learning is utilized in feature clustering and classification model, which changes the way of pre-classification and classification of the original model using a single algorithm. For pre-classification module, typical samples were labeled as aggressive, stable and conservative through two clustering algorithms. After experimental comparison, FCM and spectral clustering were used to classify and label the data initially. By comparing the clustering results of the two methods, the same results are labeled as typical samples, the rest data are retained as unclassified data. For classification module, excellent generalization ability, stability, and accuracy are reached by classifying the unclassified data in a manner of majority voting ensemble learning. Through theoretical comparison and experimental analysis, CART, SVM, and KNN are selected for the majority voting ensemble learning. The flow of the model is shown in Fig 3.
The main steps of the driving style classification model based on majority voting ensemble learning are as follows.

1. Data pre-processing: The useless attributes, and the noise data of the 450 transport vehicles are removed, and the missing data are filled with some filling regular. Moreover, the trip of each vehicle is divided into micro-trips.

2. Pre-classification module:
   1) Feature extracting: Based on the pre-processed data, the feature parameters of driving behavior are extracted by using the identification method of bad driving behavior;
   2) Feature clustering: Determining the number of clusters, and clustering the feature parameters with the FCM and SC;
   3) Data labeling: The samples with the same clustering result of the two clustering algorithms are labeled with specific labels as aggressive, stable and conservative, and the other unlabeled data are taken as unclassified samples for subsequent reclassification.

3. Classification module:
   1) Training individual classifiers: Using labeled data as training dataset to train CART, SVM, and KNN model;
   2) Classification of unclassified data: The trained CART model, SVM model, and KNN model are used to classify the unclassified data;
   3) Ensemble decision: Combining the three models’ classification results by majority voting, thereby realizing the evaluation of the driving style of each driver.

Data preparation

The dataset of question C of the “2018 Teddy Cup Data Mining Challenge” was adopted for the experiment in this paper. The details of the dataset are shown in Table 1. The dataset includes 13 attributes of 450 transportation vehicles, such as vehicle number, latitude, and longitude coordinates, direction angle, ignition status, mileage, instantaneous speed, acquisition time, and some others. Attributes and their meanings of the data are presented in Table 2. Among them, the value of the four attributes of the left-turning and right-turning signals and the handbrake and footbrake is always 0, which regard as invalid attributes. The total number of data reached 25 million.

The preprocessing procedures are shown in Table 3. Attribute specification, missing value filling, outlier correction and other operations were carried in the data preprocessing. Unavailable attributes should be deleted, partial missing mileage should be filled with average value of before and after value, and abnormal longitude and latitude should be replaced.

Extraction of driving behavior feature parameters

According to the relevant research foundation of vehicle driving behavior recognition and bad driving behavior detection [34], some methods of identifying bad driving behavior are utilized to extract the feature parameters. In extraction of feature parameters, eight bad driving behaviors including fatigue driving times, long idling hot car times, extra-long idle times, rapid lane

| Table 1. The details of the dataset. |
|-------------------------------------|
| Data source                        | 2018 Teddy Cup Data Mining Challenge |
| Whether open source                | Yes                                    |
| Feature dimension                  | 13                                     |
| Number of vehicles                 | 450                                    |
| Amount of data                     | 25000000                               |

https://doi.org/10.1371/journal.pone.0254047.t001
changes times, rapid acceleration times, rapid deceleration times, coasting with engine off times, and overspeed time are extracted. The identification methods of various types of bad driving behavior are as follows:

(1) Fatigued driving behavior: Single fatigue driving event is defined as a single continuous driving event of more than 4 hours during, which a single rest time is less than 20 minutes. If the cumulative driving time exceeds 8 hours in a single day, it is the cumulative fatigue driving event.

Suppose the vehicle has a total of \( n \) trips in a day, and the start time of the \( i \)-th trip is \( T_{i_s} \) and the end time is \( T_{i_e} \), where \( i = (1, 2, \ldots, n) \). Then, according to (12)–(14), the time \( T_i \) of the \( i \)-th trip, the time interval \( \Delta T \) between the \( i \)-th trip and the \( (i+1) \)-th trip, and the total travel time of the single day \( DT \) can be calculated. If \( T_i \) is greater than 4 hours and \( \Delta T \) is less than 20 minutes, it is a single fatigue driving event; If \( DT \) is greater than 8 hours, it is a cumulative fatigue driving event.

\[
T_i = T_{i_e} - T_{i_s} \tag{12}
\]

\[
\Delta T = T_{(i+1)s} - T_{is} \tag{13}
\]

\[
DT = \sum_{i=1}^{n} T_i \tag{14}
\]

Table 3. Data preprocessing procedure.

| Input: The original data \( D \) |
|----------------------------------|
| /* Step1: Data reduction */ |
| 1) Remove useless attributes; |
| /* Step2: Outlier handling */ |
| 2) Abnormal mileage deleted; |
| 3) Zero speed filling; |
| 4) Anomalous latitude and longitude correction; |
| /* Step3: Micro-trip division */ |
| 5) Remove the data of vehicle stop period and divide the journey into sections; |
| Output: Prepared data \( D' \) |

https://doi.org/10.1371/journal.pone.0254047.t003
(2) Rapid speed change: Define that the acceleration of the vehicle is greater than \(3\, \text{m/s}^2\) as a rapid acceleration event, and the acceleration is lesser than \(-3\, \text{m/s}^2\) as a rapid deceleration event.

Suppose the speed of the vehicle is \(v\) the time increment is \(\Delta t\) and calculate the acceleration \(A\) according to (15). If \(A \geq 3\, \text{m/s}^2\), we think it is a rapid acceleration behavior; if \(A \leq -3\, \text{m/s}^2\), we think it is a rapid deceleration behavior.

\[
A = \frac{v(t + \Delta t) - v(t)}{\Delta t}
\]  

(3) Rapid lane change: Suppose the speed of the vehicle is \(v\), the direction angle at the moment \(i\) is \(R_i\), and the time is \(t_i\) then the lane change angular speed \(V\) of the vehicle can be calculated by (16). As \(V \geq 20\, \text{°/s}\), recording \(T_1\) and the direction angle \(R_1\) of that moment. As \(V \leq 5\, \text{°/s}\), record the time \(T_2\) and the direction angle \(R_2\) of that moment. According to formulas (17)–(18), calculate the lane change duration \(T\) and the direction angle deflection \(D\) when restoring the lane change. If \(T \leq 10\, \text{s}\) and \(D \leq 5\, \text{°}\), it is judged as a sudden lane change behavior, and \(T_1, R_1, T_2, R_2\) are reset as zero.

\[
V_i = \frac{R_{i+1} - R_i}{t_{i+1} - t_i}
\]

(16)

\[
D = |R_2 - R_1|
\]

(17)

\[
T = T_2 - T_1
\]

(18)

(4) Bad idle speed: When the vehicle is ignited and the speed is 0, it is an idle speed behavior. In this paper, the idle time of the bad idle warm-up behavior is defined as 2-10min, and the idle behavior over 10min is defined as an extra-long idle behavior.

Suppose the time to start idling as \(t_s\) and the time to end the idling as \(t_e\), then the idle time of the vehicle \(T\) can be calculated according to formula (19). If 2min \(\leq T \leq 10\, \text{min}\), it is judged as a bad idle warm-up behavior, and if \(T > 10\, \text{min}\) it is judged as an extra-long idle behavior.

\[
T = t_e - t_s
\]

(19)

(5) Flameout slide: When the vehicle speed \(v \neq 0\), the status of the engine is 0, and the duration exceeds 2s, it is judged as a flameout slide event.

(6) Overspeed: According to "City Standards for Urban Road Engineering Design" (CJJ37-2012) [35], three speed limit thresholds are defined for vehicle speed, which are 60km/h, 80km/h, and 120km/h, meaning is in turn: the boundary of low speed and high speed, the boundary of highways and city roads, the boundary of ultra-high speed in highway. If the speed of the vehicle exceeds the relevant thresholds, it is determined to be an overspeed behavior.

Calling Baidu API to detect the driving section of the vehicle, the current speed is set as \(V_i\) and the speed threshold of the corresponding section of the driving vehicle is \(V_{max}\). If \(V > V_{max}\), then the vehicle is judged to be overspeeded and the overspeed time \(T\) is recorded.

After extracting each bad driving behavior, the feature parameters are extracted as below:

\[
\text{Rate}_i = \frac{C_i}{M} \times 1000
\]

(20)

\[
\text{OS} = \frac{T_{\text{over}}}{T_{\text{total}}}
\]

(21)
Rate, is the rate of vehicle bad driving behavior, \( C_i \) is the number of single bad driving behaviors, \( i \in \{1, 2, \ldots, 7\} \), corresponding to vehicle fatigue driving times, bad idling warm-up times, extra-long idling times, rapid lane changes times, rapid acceleration times, rapid deceleration times, and flameout slide times, \( M \) is the total mileage of the vehicle. OS is the ratio of vehicle overspeed time, \( T_{\text{over}} \) is the overspeed time, and \( T_{\text{total}} \) is the total time the vehicle is running.

The feature parameters and their parameter units are shown in Table 4.

### Pre-classification of driving behavior based on FCM and SC

In pre-classification, there are some processes for achieving the purpose of dividing vehicle data into labeled dataset and unclassified dataset. Firstly, cluster the vehicle data into \( k \) types by FCM and spectral clustering. In this process, set the category number \( k \) and suppose the clustering result of FCM be expressed as \( C_f(i) \), the cluster center as \( m_j \), and the clustering result of spectral clustering be expressed as \( C_s(i) \), where \( i = 1, 2, \ldots, n \), \( n \) is the number of samples. Then, compare \( C_f(i) \) with \( C_s(i) \), if \( C_f(i) = C_s(i) \), the sample \( i \) is labeled with its type label and divided into labeled dataset \( X_i \). Otherwise, the sample \( i \) is divided into unclassified dataset \( P_i \). The pre-classification procedure is presented in Table 5.

### Driving behavior classification based on majority voting ensemble learning

The driving behavior classification model based on majority voting ensemble learning integrates CART, SVM, and KNN model to classify driving behavior. On the basis of

| Parameter | Unit          |
|-----------|--------------|
| Fatigue driving rate (FDR) | Times / 1000 km |
| Long idling hot car rate (LIHR) | Times / 1000 km |
| Extra-long idle rate (EIR) | Times / 1000 km |
| Rapid lane change rate (RLCR) | Times / 1000 km |
| Rapid acceleration rate (RAR) | Times / 1000 km |
| Rapid deceleration rate (RDR) | Times / 1000 km |
| Coasting with engine off rate (CEOR) | Times / 1000 km |
| Overspeed time ratio (OTR) | - |

https://doi.org/10.1371/journal.pone.0254047.t004
pre-classification, labeled samples are trained and tested on the classification model. 80% of the labeled samples are used as training dataset $X$ and 20% of the labeled samples are used as $\bar{X}$, and then select the three models that perform best in test dataset $P$. Firstly, every individual classifier was trained this task to classify the unclassified dataset $P$. Finally, the individual classifiers are combined by the majority voting strategy and thus obtain the final classification results. The classification procedure is presented in Table 6. There are some classifier parameters that need to be set. The penalty coefficient “C” of SVM is set as “0.8”, the kernel function is set as “linear”, and the decision method “decision_function_shape” is set as “ovr”. The maximum depth of the CART decision tree “max_depth” is set as “6”, the minimum impurity reduction amount “min_impurity_split” is set as “0.1”, the maximum number of leaf nodes “max_leaf_nodes” is set as “28”, and the minimum sample number of leaf nodes “min_samples_leaf” is set as “1”. KNN does not need to set any other parameters. In addition, the computational complexity of the proposed method is $O(MN)$.

### Experiment and analysis

#### Experiment setup and procedures

As sketched in Fig 3, the experimental procedures mainly consist of four steps: (1) Data pre-processing; (2) Extraction of feature parameters; (3) The feature parameters are clustered by FCM and SC, and the data are divided into labeled data and unclassified data; (4) Use labeled data to train and test CART, SVM, and KNN models, then use the three models to classify the unclassified data, finally, combine the results of three models by majority voting.

The experimental environment configuration and platform details of this experiment are shown in Table 7. Python was used as the development language for the entire experiment,
and PyCharm 2018 and Anaconda 3.0 were used as the development environment. The exper-
imental computer’s central processing unit (CPU) is Intel core i5-3230M, 2.6GHz, random
access memory (RAM) is 4GB and the operating system is Windows 10 64-bit.

**Experimental result and analysis**

First of all, in the preprocessing, all empty attributes are deleted, and only valid attributes are
retained. When analyzing the data, it is found that there are abrupt mileage data and missing
speed. For the abrupt mileage data, method of deleting the abrupt term is used for processing,
and for the missing speed, mean filling is used. For abnormal longitude and latitude coordi-
nates, angular velocity and traveling speed are used to predict the longitude and latitude coordi-
nates of the next moment, and the predicted longitude and latitude coordinates are used to
replace the abnormal longitude and latitude coordinates. Take car AA00002 as an example, the
collected vehicle speed, and mileage are shown in Fig 4. Due to the large amount of data, we
only show part of the data in the figure. And in the collected data, the starting mileage of the
vehicle was 8,865 kilometers and the ending mileage was 9,614 kilometers, thus the vehicle
traveled a total of 749 kilometers over 5 days. We can get the average speed of the vehicle by
the vehicle mileage and travel time. Different attributes can be combined and verified with
each other, thus the richer the data were, the more the accuracy of the detect driving events
will be. At the same time, we divide the micro-trips of the vehicle trajectory data, and mark
the data of all vehicles according to the micro-trip number. Take car AA00002 as an example,
between August 4, 2018 and August 7, 2018, the vehicle made a total of 11 micro-trips, with
the start and end times of each trip shown in Fig 5. As can be seen from the figure, the driving
time of this vehicle is mostly concentrated at night, which conforms to the driving law of
freight vehicles.

Secondly, according to the relevant research basis of vehicle driving behavior recognition
and bad driving behavior detection mentioned in “Extraction of driving behavior feature
parameters”, 450 transport vehicles were featured. The extraction results are shown in Table 8.
The value of the overspeed time ratio is the result of magnifying the original value by 1000
times in Table 8. After features extraction, it was found that there are 3 vehicles have an empty
data, so the data of 447 vehicles were finally put into use in the next stage.

Thirdly, in pre-classification experiments, four clustering algorithms, i.e. K-means,
DBSCAN, FCM and SC, are used for clustering. Finally, two algorithms with higher similarity
of clustering results were selected as sub-algorithms of the pre-classification module. Experi-
ment shows that the DBSCAN is a clustering algorithm based on density division, which has a

![Fig 4. Data of AA00002. (a) Display of the first 10,000 speed data of the vehicle; (b) Display of the first 10,000 direction angle data of the vehicle; (c) Display of the vehicle mileage data.](https://doi.org/10.1371/journal.pone.0254047.g004)
good effect on the detection of noise points. However, the cluster radius and density thresholds need to be set in advance, which is not applicable to the dataset in this paper. And compared with FCM and spectral clustering, the clustering results of K-means have more differentiated samples, which can’t label the data as many as possible. Therefore, FCM and spectral clustering are selected as sub-algorithms in this paper. In the experiment, the value of cluster number is set as 3, FCM and SC cluster the feature parameters data respectively. After clustering by FCM and spectral clustering, the clustering results and initial classification results obtained are shown in Table 9. In the results of FCM, the number of samples classified as Type-I is 98, the number of samples classified as Type-II is 23, and the number of samples classified as Type-III is 327. In the results of spectral clustering, the number of samples classified as Type-I, Type-II and Type-III are 105, 24 and 319. It can be seen that the clustering results of the two clustering methods are generally consistent, which also shows the effectiveness of clustering. Comparing the clustering results of the two algorithms, the number of samples classified as Type-I by both methods is 94, the number of samples classified as Type-II by both methods is 23, and the number of samples classified as Type-III by both methods is 316. Therefore, there are 433 samples labeled as training data. And the number of unclassified samples is 14, which will be classified in the subsequent experimental steps. What’s more, from the clustering results, the number of the Type-I and Type-III occupies the vast majority of the entire dataset. This means that the majority of the company’s drivers behave well.

Finally, according to the relevant theoretical basis in "Driving behavior classification based on majority voting ensemble learning", the driving behavior classification model of majority

**Table 8. The results of feature parameters extraction.**

| FDR  | LIHR | EIR  | RLCR | RAR  | RDR  | CEOR | OTP  |
|------|------|------|------|------|------|------|------|
| 0.94 | 8.45 | 0.47 | 16.89 | 11.73 | 10.32 | 0    | 36.64|
| 2.79 | 11.14| 0    | 8.36 | 261.84| 264.62| 0    | 113.16|
| 0    | 5.03 | 0    | 78.73| 50.25 | 46.90 | 0    | 132.32|

https://doi.org/10.1371/journal.pone.0254047.t008
voting ensemble learning is used to classify the unclassified samples. Firstly, each individual classifier is trained and tested by labeled dataset. Then, the results of the three classifiers are combined by majority voting to get a more stable classifier. The 433 labeled samples were divided into two parts. 80% of the samples were used as training dataset to train the classifiers, and 20% of the samples were used as test dataset to test the performance of the classifiers. With 346 random training samples and 87 test samples, the prediction accuracy of proposed model was fluctuated between 98.85% and 100%. The results of the model on the test dataset were shown in Fig 6. As can be seen from Fig 6, 87 test samples were classified, and the maximum number of samples with classification error was one, which demonstrated the excellent performance of the proposed method.

In addition, three models with the best performance of individual classifiers were selected to predict the unclassified samples, the final classification results are shown in Fig 7. The specific data features and classification results are shown in Table 10. Where 10 of the 14 vehicles were classified as stable, 3 as conservative and 1 as aggressive. It can be seen from the table that the rapid acceleration and deceleration behaviors of vehicles classified as aggressive are significantly higher than those of other vehicles, which conforms to the driving characteristics of aggressive driving. Drivers classified as stable and conservative type have a small amount of bad driving behaviors, and less rapid acceleration and rapid deceleration. Moreover, drivers with severe acceleration and deceleration behaviors also generally have rapid lane change behavior, and drivers without classification do not have coasting with engine off behavior.

Subsequently, according to the final classification results, the relevant feature parameters, and the category labels of the training dataset, the driving style of the drivers can be classified.
Fig 7. Classification results of unclassified samples.

![Classification results of unclassified samples](https://doi.org/10.1371/journal.pone.0254047.g007)

Table 10. Classification results of unclassified samples.

| Car ID | FDR  | LIHR | EIR  | RLCR | RAR  | RDR  | CEOR | OTP  | Result | Style   |
|--------|------|------|------|------|------|------|------|------|--------|---------|
| AA00052| 0    | 1.16 | 2.32 | 6.95 | 192.35| 192.35| 0    | 217.16| 0      | STABLE  |
| AA00154| 0    | 19.91| 0    | 9.19 | 199.08| 194.49| 0    | 79.51 | 0      | STABLE  |
| AA00345| 0    | 277.78| 0   | 833.33| 222.22| 222.22| 0    | 0     | 0      | STABLE  |
| AB00465| 1.03 | 1.03 | 2.06 | 20.64| 62.95 | 288.96| 0    | 71.61 | 0      | STABLE  |
| AD00027| 0.92 | 5.54 | 0    | 11.53| 189.11| 184.04| 0    | 56.02 | 0      | STABLE  |
| AD00050| 1.76 | 2.34 | 0    | 2.93 | 146.87| 145.11| 0    | 125.74| 2      | CONSERVATIVE |
| AD00130| 0    | 38.27| 2.55 | 413.27| 260.20| 283.16| 0    | 52.87 | 0      | STABLE  |
| AD00150| 0    | 16.52| 0    | 15.02| 201.95| 197.45| 0    | 94.68 | 0      | STABLE  |
| AD00163| 0    | 9.64 | 0    | 23.66| 198.07| 201.58| 0    | 62.05 | 0      | STABLE  |
| AD00178| 0    | 19.01| 0    | 44.35| 177.40| 181.62| 0    | 83.42 | 0      | STABLE  |
| AD00267| 1.68 | 3.35 | 0    | 1.68 | 136.87| 130.73| 0    | 159.94| 2      | CONSERVATIVE |
| AD00292| 2.33 | 4.08 | 0    | 4.66 | 167.35| 167.93| 0    | 108.57| 2      | CONSERVATIVE |
| AD00344| 2.05 | 0    | 0.68 | 7.51 | 197.95| 184.30| 0    | 135.35| 0      | STABLE  |
| AF00066| 0    | 11.57| 0    | 76.03| 563.64| 591.74| 0    | 0     | 1      | Aggressive |

https://doi.org/10.1371/journal.pone.0254047.t010
as aggressive, stable and conservative type. T-distributed stochastic neighbor embedding (TSNE) was used to reduce the dimension of the data, and the classification results after dimension reduction were shown in Fig 8. It can be obtained from Fig 8 that 447 vehicles are effectively divided into 3 types, and the boundaries between the types are obvious. The yellow triangle samples are conservative driving style, which corresponds to label 2; the red dot samples are stable driving style, which corresponds to label 0; the blue star samples are aggressive driving style, which corresponds to label 1.

Thereafter, according to statistics, the driving style classification results of 447 drivers are shown in Fig 9. There are 105 drivers with stable driving style, accounting for 23.4% of the total dataset, 24 drivers with aggressive driving style, accounting for 5.4% of the total dataset, and 319 drivers with conservative driving style, accounting for 71.2% of the total dataset. It can be seen that most of the drivers of this company are in compliance with the standard driving behavior and their driving style are more stable or conservative. A few drivers have more aggressive driving style. Drivers with aggressive driving style should be specially managed to regulate their driving behaviors and prevent the occurrence of driving accidents.

**Comparison and discussion**

The validity of the classification results and the classification model are evaluated respectively in this work. The validity of the classification results was evaluated by internal cluster evaluation indicators Davies-Bouldin index and Calinski-Harabasz index, and the validity of the
classification model was evaluated by accuracy, recall, precision and F-measure. The internal cluster evaluation indicator can effectively evaluate whether the classification is reasonable or not. The evaluation indicator of classification model can evaluate the model from the accuracy, stability, generalization ability and other aspects comprehensively.

Firstly, calculate the Davies-Bouldin index and Calinski-Harabasz index of the clustering result of K-means used in [9], FCM used in [13], and another sub-clustering algorithm [36] in proposed pre-classification model which named spectral clustering by (1)–(7). Then calculate the two indexes of the data classified by the proposed model. The results retain 4 decimal places, and the calculation results are shown in Table 11.

It can be seen from Table 11 that in comparison with the Davies-Bouldin index and Calinski-Harabasz index of FCM clustering results in [13], the Davies-Bouldin index value of the data classified by the proposed model is lesser than the result of direct use of FCM, and the value of Calinski-Harabasz index is greater than that of direct use of FCM. This indicates that the three categories separated by the proposed model are farther apart and the samples within each category is closer together. Then, comparing the two indexes of spectral clustering and the proposed method, although the Davies-Bouldin index value of the proposed method is slightly larger than the spectral clustering, the Calinski-Harabasz index value is much larger than the it. The comprehensive comparison of the two indexes also shows that the classification results of the proposed model are more in line with the classification principle that the farther the distance between categories is the better, and the closer the distance within categories is the better. Finally, comparing the two indexes of K-means and the proposed model, the proposed method performs much better than it. Therefore, the classification results of the proposed model make the distance between the categories of data further, the distance within the categories closer, and the classification of data more reasonable and effective.

Secondly, compare the proposed majority voting ensemble learning method with the conventional ensemble learning method RF and AdaBoost. The number of individual classifiers of RF and AdaBoost were set to be 3, which consistent with the number of individual classifiers

| Method          | Value of DBI | Value of CH  |
|-----------------|--------------|--------------|
| FCM [13]        | 0.7614       | 438.3526     |
| Spectral clustering | 0.7377     | 414.7100     |
| K-means [9]     | 0.8112       | 316.0603     |
| This work       | 0.7428       | 457.8629     |

Table 11. Comparison of the values of DBI and CH indicator.

https://doi.org/10.1371/journal.pone.0254047.g009
of proposed model, and the proposed model, RF, and AdaBoost were trained and tested for 20 times. The accuracy, recall, precision and F-measure of three models are obtained by (8)–(11).

And the experiment results are shown in Fig 10, which red, blue and pink dots respectively represent the model of the RF, AdaBoost and this work. The gray areas in the figures are the range of indexes values. It can be seen that the proposed model’s accuracy, recall, precision and F-measure are more stable in 20 experiments and the value of evaluation indexes are higher than the other two methods, while training and testing on the labeled data set. The accuracy rate represents the prediction accuracy of the whole dataset, and the higher the accuracy, the closer the model classification results are to the real results. The recall rate represents the probability of actually positive samples that being predicted to be positive, and the precision rate represents the probability of being predicted to be positive that is actually positive sample, and they restrict and influence each other. F-measure is a harmonic average of
precision and recall rates, which can balance the influence of precision rate and recall rate and evaluate a classifier more comprehensively. Therefore, the higher the value of these four indexes are, the better the classifier is. In addition, we can obtain that ensemble learning methods have a good effect in dealing with classification problems. The model accuracy of the three methods were all higher than 90%, and other indexes are also higher than 80%. The average value of the evaluation indexes for 20 experiments and the running time of the proposed method and the conventional ensemble learning method are presented in Table 12. The data in Table 12 is obtained by retaining the four decimal places for the actual value. As can be seen from Table 12, the running time of the proposed model is slightly higher than that of RF and AdaBoost, but the performances of the other four indexes are significantly better than them. This indicates that the proposed majority voting ensemble learning method has better classification ability and robustness than the conventional ensemble learning method. And it is more suitable for solving complex classification problems.

Finally, compare the accuracy, recall, precision and F-measure of the proposed model with the methods used in [16, 22, 23, 26, 37] for the same dataset. The specific methods and the value of indexes of these models are presented in Table 13. As can be seen from Table 13, the neural network model does not perform well in this task, possibly because the feature dataset is not large enough. SVM performs well in this task compared with neural networks. The main reason is that SVM is more sensitive to small sample dataset and can train better classifier with less training dataset. Moreover, the performance of ensemble learning method in this work is generally higher than other machine learning methods, because ensemble learning avoids the limitation of single algorithm in classification decision making and improves stability while ensuring accuracy. In addition, the method used to label has an impact on the final result. The rationality of labels directly affects the validity of subsequent classification. Therefore, the method selection of pre-classification stage is also very important. It can be seen from the table that the same classifier has different performance while uses different clustering methods, so the effectiveness of the method needs to be considered in many aspects when building the model. With a comprehensive comparison of the four indexes, it can be seen that the proposed method performs better than other methods in the Table 13, which indicates the proposed method is able to provide a reliable support for the realization of automatic driving technology, and also provides a reference for usage-based insurance.

Table 12. Comparison of the average value of evaluation indicators of different ensemble learning models.

| Model         | Time (s) | Accuracy | Recall | Precision | F-measure |
|---------------|----------|----------|--------|-----------|-----------|
| Random forest | 0.3022   | 97.87%   | 94.38% | 96.37%    | 94.98%    |
| AdaBoost      | 0.3352   | 98.56%   | 99.01% | 98.59%    | 98.83%    |
| This work     | 0.3581   | 99.37%   | 99.69% | 99.27%    | 99.47%    |

Table 13. Comparison of the value of evaluation indicators of different models for the same dataset.

| Ref. | clustering algorithm | classifier | Accuracy | Recall | Precision | F-measure |
|------|----------------------|------------|----------|--------|-----------|-----------|
| [16] | FCM                  | BPN        | 81.25%   | 60.74% | 55.56%    | 56.75%    |
| [22] | K-means              | SVM        | 98.89%   | 98.61% | 99.49%    | 98.03%    |
| [23] | PCA+FCM              | SVM        | 98.89%   | 98%    | 99.24%    | 98.60%    |
| [26] | K-means              | AdaBoost   | 98.89%   | 99.49% | 99.61%    | 99.04%    |
| [37] | K-means              | BPN        | 75%      | 26.09% | 33.33%    | 29.27%    |
| This work | FCM+SC            | SVM+KNN+CART | 99.20%   | 99.61% | 99.07%    | 99.33%    |

https://doi.org/10.1371/journal.pone.0254047.t012

https://doi.org/10.1371/journal.pone.0254047.t013
Conclusion

Based on the internet of vehicles data of 450 transport vehicles from the competition platform, this paper extracts and quantifies the parameters of driving behavior features, and comprehensively uses various classification methods to deal with the problem. The model consists of pre-classification stage and classification stage, and the ensemble learning is used in both stages. In the pre-classification, an ensembled clustering method based on FCM and spectral clustering is used to cluster the driving behavior feature parameters, thereby the data are divided into labeled dataset and unclassified dataset. With the results of pre-classification, a majority voting ensemble learning classification method based on CART, SVM, and KNN is trained and tested. A variety of individual classifiers are used for learning, and a majority voting strategy is used to combine the three individual classifiers, and thus classify the driving style of vehicles.

A correlation mechanism of internet of vehicles data, driving behavior characteristics and traffic safety is established by the classification of driving styles and mining of drivers’ driving behavior habits, which provides a reference for the transportation management department to monitor and assess drivers. At the same time, this paper improves the generalization ability, stability and accuracy of driving behavior classification model through the combination of various individual classifiers. However, there are some limitations in this paper, for example, the data dimension is not rich enough, which does not include road information and vehicle condition information. More factors should be considered in future researches. Factors such as weather and road conditions can be taken into account, and more diverse data can make the final evaluation more accurate and complete. What’s more, automatic feature screening can also be studied in future.

Author Contributions

Conceptualization: Yi Guo, Xiaolan Wang.

Formal analysis: Liang Xu.

Investigation: Yongmao Huang.

Methodology: Xiaolan Wang.

Writing – original draft: Xiaolan Wang.

Writing – review & editing: Xiaolan Wang.

References

1. Kumar S., Mahima, Srivastava D. K., Kharya P., Sachan N., Kiran K., “Analysis of risk factors contributing to road traffic accidents in a tertiary care hospital. A hospital based cross-sectional study,” Chin. Jour. Traumat., vol. 23, no. 3, 2020, https://doi.org/10.1016/j.cjtee.2020.04.005

2. Road traffic accident statistical annual report, Traffic Management Bureau of the Ministry of Public Security, 2018.

3. Chan M., Singhal A., “Emotion matters: Implications for distracted driving,” Safety Science, 2015, 72:302–308.

4. Wang W., Xi J., Liu C., and Li X., “Human-centered feed-forward control of a vehicle steering system based on a driver’s path-following characteristics,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 6, pp. 1440–1453, Jun. 2017, https://doi.org/10.1109/TITS.2016.2606347

5. Bolovinou A., Amditis A., and Bellotti F., “Driving style recognition for co-operative driving: A survey,” in Proc. 6th Int. Conf. Adapt. Self-Adapt. Syst. Appl., pp. 73–78, May 2014

6. Dib W., Chasse A., Moulin P., Sciarretta A., and Corde G., “Optimal energy management for an electric vehicle in eco-driving applications,” Cont. Eng. Pract., vol. 29, pp. 299–307, 2014, https://doi.org/1016/j.conengprac.2014.01.005 PMID: 10
7. Wang R. and Lukic S. M., “Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles,” in IEEE Vehicle Power Prop. Conf., Chicago, IL, pp. 1–7, Sep. 2011, https://doi.org/10.1109/VPPC.2011.6043061

8. Engelken M., Römer B., Drescher M., Welpe I. M. and Picot A., “Comparing drivers, barriers, and opportunities of business models for renewable energies: A review,” Renew. Sustain. Energy Rev., vol. 60, pp. 795–809, 2016, https://doi.org/10.1016/j.rser.2015.12.163

9. Li J. W., Zhao Z. G., Shen P. H. and Guo Q. Y., “Research on k-means clustering and recognition method of driving style,” Autom. Technol., vol. 12, pp. 8–12, 2018, https://doi.org/10.19620/j.cnki.1000-3703.20180272

10. Chen J., Wu Y., Huang H., Wu B. and Hou G., “Driving-data-driven platform of driving behavior spectrum for vehicle networks,” in IEEE 20th Intl. Conf. High Perform. Computing Conf.; IEEE 16th Intl. Conf. Smart City; IEEE 4th Intl. Conf. Data Sci. Syst. Exeter, United Kingdom, pp. 518–525, Jun. 2018, https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00099

11. Wen H., Yang X. M. and Wu C. Z., “Analysis of driving behavior characteristics of commercial vehicles under big data environment,” Trans. Int. Security., vol. 36, no. 4, pp. 42–50, 2018.

12. Hui F., Peng N., Jing S. C., Zhou Q. and Jia S., “Driving behavior clustering and anomaly detection method based on agglomeration level,” Computer Eng., vol. 44, no. 12, pp. 196–201, 2018, https://doi.org/10.19678/j.issn.1000-3428.0050708

13. Wang H. X., Wang X. Y., Wang Z. X. and Li X. D., “Cluster analysis of risky driving behavior of dangerous goods transportation based on data mining,” Transp. Syst. Eng. Inf., vol. 20, no. 1, pp. 183–189, 2020, https://doi.org/10.16097/j.cnki.1009-6744.2020.01.027

14. Brombacher P., Masino J., Frey M. and Gauterin F., “Driving event detection and driving style classification using artificial neural networks,” in IEEE Intl Conf. Indust. Techn., Toronto, ON, pp. 997–1002, Mar. 2017, https://doi.org/10.1109/CIT.2017.7915497

15. Kagawa T. and Chandrasiri N. P., “Analysis of driving skills based on deep learning using stacked auto-encoders,” in 4th Intl Conf. Electr. Eng., Compu. Sci. Infor., Yogyakarta, pp. 1–4, Sep. 2017, https://doi.org/10.1109/EECSI.2017.8239200

16. Wu Z. H., Wu Z. C., Zhang J., Chen S. and Chen J., “Research on driving behavior evaluation based on fuzzy c-means and neural network,” Computer Syst. Appl., vol. 27, no. 3, pp. 263–267, 2018, https://doi.org/10.15888/j.cnki.csa.006256

17. Li G., Zhu F., Qu X., Cheng B., Li S. and Green P., “Driving style classification based on driving operational pictures,” in IEEE Access, vol. 7, pp. 90180–90189, 2019, https://doi.org/10.1109/ACCESS.2019.2926494

18. Bejani M. M. and Ghatee M., “Convolutional neural network with adaptive regularization to classify driving styles on smartphones,” in IEEE Transac. Intel. Transp. Syst., vol. 21, no. 2, pp. 543–552, Feb. 2020, https://doi.org/10.1109/TITS.2019.2896672

19. Wang J. Q., Zheng Y., Li X. F., Yu C. F., Kodaka K. J. and Li K. Q., “Driving risk assessment using near-crash database through data mining of tree-based model,” Accident Anal. Prev., vol. 84, pp. 54–64, 2015, https://doi.org/10.1016/j.aap.2015.07.007 PMID: 26319604

20. Huang L., Dong Z. H. and Zhang R. M., “Analysis of driving behavior based on random forest,” Wireless Int. Technol., vol. 15, no. 7, pp. 72–76, 2018.

21. Panagopoulos G. and Pavlidis I., “Forecasting markers of habitual driving behaviors associated with crash risk,” in IEEE Transa. on Intell. Transp. Syst., vol. 21, no. 2, pp. 841–851, Feb. 2020, https://doi.org/10.1109/TITS.2019.2910157

22. Wang W., Xi J., Chong A. and Li L., “Driving style classification using a semisupervised support vector machine,” IEEE Trans. Human-Machine Syst., vol. 47, no. 5, pp. 650–660, Oct. 2017, https://doi.org/10.1109/THMS.2017.2736948

23. Liu Y., Wang J., Zhao P., Qin D. and Chen Z., “Research on classification and recognition of driving styles based on feature engineering,” in IEEE Access, vol. 7, pp. 89245–89255, 2019, https://doi.org/10.1109/ACCESS.2019.2926593

24. Wang X. S. and Bian Z., “Recognition and prediction of driving behavior based on bayesian model,” Journal of Communications, vol. 39, pp. 108–117, 2018.

25. Chen C., Zhang G. H., Tarafder R., Ma J. M., Wei H. and Guan H. Z., “A multinomial logit model-bayesian network hybrid approach for driver injury severity analyses in rear-end crashes,” Accident Anal. Prev., vol. 80, pp. 76–88, Jun. 2015, https://doi.org/10.1016/j.aap.2015.03.036 PMID: 25888994

26. Xu T., Zhang X., Zhang Y. K. and Wang J., “Classification of safety tendency of truck drivers based on adaboost algorithm,” Jour. Safety Envir., vol. 19, no. 4, pp. 1273–1281, 2019, https://doi.org/10.13637/j.issn.1009-6094.2019.04.024
27. Wu Z. X., He Y. T., Yu L. J., Fu L. and Chen P., “Research on driving style recognition algorithm based on big data,” Autom. Technol., vol. 10, pp. 10–15, 2018, https://doi.org/10.19620/j.cnki.1000-3703.20181053

28. Zehra W., Javed A. R., Jalil Z., Gadekallu T. R., Kahn H. U., “Cross corpus multi-lingual speech emotion recognition using ensemble learning,” Comp. & Intell. Syst., 2021(1), https://doi.org/10.1007/s40747-020-00250-4

29. Abbasi A., Javed A. R., Chakraborty C., Nebhen J., Zehra W. and Jalil Z., “ElStream: An ensemble learning approach for concept drift detection in dynamic social big data stream learning,” in IEEE Access, vol. 9, pp. 66408–66419, 2021, https://doi.org/10.1109/ACCESS.2021.3076264

30. Rehman A., Rehman S. U., Khan M., Alazab M. and Gadekallu T. R., “CANintellIDS: Detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU,” in IEEE Transac. Net. Sci. Engin., PP (99):1–1, 2021, https://doi.org/10.1109/TNSE.2021.3059881

31. Davies D. L. and Bouldin D. W., “A cluster separation measure,” in IEEE Transac. Pattern Analysis Machine Intell., vol. PAMI-1, no. 2, pp. 224–227, April 1979, https://doi.org/10.1109/TPAMI.1979.4766909 PMID: 21868852

32. Calinski R. B. and Harabasz J., “A dendrite method for cluster analysis,” Commun. in Statistics, vol. 3, 1974, pp. 1–27.

33. Shen Y., Tang Y. and Dou Q., “The discovery of micro-blog users’ interest groups based on SSLOK-means clustering focus on big data improved by calinski-harabasz,” 2nd Internat. Conf. Data Science and Business Analytics, Changsha, pp. 98–104, 2018, https://doi.org/10.1109/ICDSBA.2018.00025

34. Ren H. J., Xu T., and Li X., “Using GPS trajectory data to realize bus driving safety analysis,” Jour. Wuhan Uni., vol. 39, no. 6, pp. 739–744, 2014, https://doi.org/10.13203/j.whugis20140132

35. CJJ37-2012, Code for design of urban road engineering [S]. Beijing: Ministry of Housing and Urban-Rural Development of the People’s Republic of China, 2016.

36. Wang Y. and Wang H. L., “Application of improved spectral clustering algorithm in image segmentation,” Meas. Ctrl. Technol., vol. 37, no. 4, pp. 11–15, 2018.

37. Zheng H. J., Xiong X. and Zhang S., “Driver behavior analysis based on internet of vehicles data mining,” Info. Comm., vol. 8, pp. 52–55, 2019.