Heterogeneity in the Driver Behavior: An Exploratory Study Using Real-Time Driving Data

Jahnavi Yarlagadda and Digvijay S. Pawar

Transportation Systems Engineering, Department of Civil Engineering, Indian Institute of Technology Hyderabad, Kandi, Medak 502285, India

Correspondence should be addressed to Digvijay S. Pawar; dspawar@ce.iith.ac.in

Received 30 July 2021; Revised 22 December 2021; Accepted 18 May 2022; Published 18 June 2022

Copyright © 2022 Jahnavi Yarlagadda and Digvijay S. Pawar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Driver behavior heterogeneity is a significant aspect to understand the individual behavioral variations and develop driver assistance systems. This study characterizes the heterogeneity in driving behavior using real-time driving performance features. In this context, the study investigates the extent of variations in the individual’s driving styles during routine driving. The driving styles are conceptualized using the vehicle kinematic data, that is, speed and accelerations performed during longitudinal control. The data is collected for 42 professional drivers using instrumented vehicle over a defined study stretch. An algorithm is developed for data extraction and total 7548 acceleration and 6156 braking maneuvers and corresponding driving performance features are extracted. The driving maneuver data are analyzed using the unsupervised techniques (PCA and K-means clustering) and three patterns of acceleration and braking are identified, which are further associated with two patterns of speed behavior. The results showed that each driver is found to exhibit different driving patterns in different driving regimes and no driver shows constantly safe or aggressive behavior. The aggression scores are found to be different among drivers, indicating the behavioral heterogeneity. This study results demonstrate that, driver’s level of aggression in different driving regimes is not constant and characterizing the driver by means of abstract driving features is not indicative of the diversified driving behavior. The proposed method identifies the individualized driving behaviors, reflecting the driver’s choice of driving maneuvers. Thus, the insights from the study are highly useful to design driver-specific safety models for driver assistance and driver identification.

1. Introduction

Road traffic accidents are one of the leading causes of death resulting in approximately 1.35 million deaths every year [1]. The factors associated with road crashes are studied over decades and driver behavior is concluded to be the major contributory factor [2–7]. Therefore, understanding the driver behavior is important for many applications like driver assistance or personalized feedback provision for enhancing the driving safety, economy, and comfort. In addition, the implications of driver behavior research are significant inputs for the design of autonomous vehicles.

Driver behavior indicates the manner of executing various driving tasks, which can be perceived as controlling the vehicle in the longitudinal and lateral directions. The habitual way of performing driving maneuvers is considered as driving style, which characterizes the individual driver or a group of drivers [8]. Many researchers have attempted to classify the drivers and the driving styles based on the outcomes of driving tasks in the perspective of driving safety [9–20]. Most of the studies considered the same predefined criteria (thresholds for abstract features of driving data or for safety-critical events) for all the drivers neglecting the variation in driver attributes and driving skills. The differences in the individual driving characteristics among drivers might result in different driving responses toward a set of driving conditions [21]. Thus, classifying the driver behavior based on the predefined thresholds common to all the drivers may not yield to the proper evaluation of driver behavior. Moreover, the driving styles are assumed to be
constant or consistent in the entire trip, which is unrealistic as the external driving conditions tend to change time to time [22, 23]. In other words, the driver classified as aggressive may not continuously exhibit aggressive behavior in all the driving maneuvers performed in a trip. Therefore, this study proposes a framework to explore the behavioral heterogeneity in the longitudinal control by segmenting the driving profiles into driving maneuvers and identifying different driving patterns. The analysis is conducted using unsupervised machine learning techniques to group and interpret the driving patterns, without any fixed thresholds.

In this study, a high-frequency (10 Hz) GPS instrumentation is used to collect the real-time driving data for passenger car drivers. The longitudinal control exhibited by drivers in terms of speed, accelerating, and braking were considered to explore the driving pattern variations in short-term driving decisions. The next section of the paper presents a summary of the literature review on driving style identification. The third section details the data collection techniques and the fourth section presents the adopted methodology. The fifth section details the results and discussion. The last section of the paper presents the conclusions, and future work.

2. Literature Review

Over the last three decades, many researchers have conducted the safety analysis of driving behavior that prominently involved identifying and classifying driving styles. The driving style detection and classification were mainly regarded as a means of differentiating drivers in the perspective of driving safety and to evaluate individuals. The existing literature on driving style classification is synthesized as per the nature of the data used for the analysis and presented in the following subsections.

2.1. Predefined Thresholds of Safety Critical Events. With the advancement in the data collection techniques, the driving styles were conceptualized using different driving performance features depending on the research motive and method of data collection. Majority of the studies classified the driving styles based on the predefined thresholds of safety-related events, such as sudden acceleration, sudden braking, and sharp turning [9, 11–13, 16, 17, 20, 24–29]. Johnson and Trivedi [11] developed smartphone-based application, which categorizes the driving styles as aggressive and nonaggressive based on the nature of detected driving maneuvers. The sudden accelerations and braking, swift swerves, and hard left and right turns were the typical maneuvers considered, for which the reference thresholds were recorded by a single driver and a vehicle. Aljaafreh et al. [12] classified driving styles into four categories using the fuzzy logic inference system. The fuzzy rules for classification were predefined based on the driving performance of three expert drivers in terms of speed and longitudinal/lateral accelerations. Similarly, Feng et al. [24] developed a fuzzy logic driver model to simulate different driving styles. The fuzzy rules were framed based on experts’ knowledge on drivers’ decisions using the parameters such as vehicle speed, headway distance, pedal position, and gear selection. Van Ly et al. [13] explored the possibility of CAN bus data to build driving profiles for characterizing individuals. Authors represented the driving maneuvers using acceleration, braking, and turning events. The k-means clustering and support vector machine were used in the training algorithm and showed an accuracy of 60%. Vaiana et al. [16] classified driving behaviors into safe and aggressive using a g-g diagram developed by referencing a single driver. Eboli et al. [17] constructed a g-g diagram and defined the safety domain for driver behavior classification. The frequency of data points out of the safety domain was considered to classify drivers as safe or unsafe. Mantouka et al. [20] categorized the trips into six distinct groups with the increasing level of safety. The driving performance over the trip was defined using the frequencies of harsh accelerations and harsh brakes per kilometer traveled and the percent of speeding and mobile usage during the trip. Chen et al. [25] developed a supervised hierarchical Bayesian model to understand the latent driving styles using the labeled data. The vehicle motion data, including acceleration, speed, and turning signals were categorized into multiple levels ranging from very low to very high and encoded to input the model. In these studies, the predefined thresholds were needed to detect the driving events, but the thresholds used for defining the driving events were not consistent. Paefgen et al. [26] used ±0.1 g (1 m/s²) for identifying critical acceleration and braking events, whereas Fazeen et al. [27] used the thresholds of ±0.3 g (3 m/s²). Bagdadi [28] used ±0.48 g (4.8 m/s²), and Bergasa et al. [29] considered ±0.4 g (4 m/s²) as thresholds for critical events.

2.2. Abstract Driving Features. Some studies classified the driving styles based on the abstract features of driving performance data observed during the study period [10, 15, 30]. Constantinescu et al. [10] classified driving styles as per the risk-proneness of drivers in the context of road safety. The hierarchical cluster analysis was performed on the abstract features of driving performance data, such as the mean and standard deviation of speed, acceleration, deceleration, percentage of time speeding over speed limit, and the mechanical work over the entire trip. The identified driving styles were compared against the test driver’s performance and categorized into five groups ranging from nonaggressive to aggressive. Similarly, Kalsoom and Halim [30] classified driving styles into slow, normal, and fast categories based on the abstract features such as maximum and average speed, number of brakes, and number of horns during the trip. Given that, the performance aggregated over an entire trip was used for classification, the results of these studies cannot directly indicate the unsafe practices or the individual’s driving faults to diagnose through assistance.

2.3. Driving Profiles. A few studies explored the driving styles using continuous driving profile data, without any prior ground truth [19, 22, 23, 31, 32]. Li et al. [31] considered the transition probabilities between driving
maneuvers as indicators of driving styles and analyzed the transitions among 12 types of maneuvers. All the maneuver episodes were manually identified from the naturalistic data corresponding to 28 drivers. The authors developed random forest classifier based on the optimized set of five transition features and classified drivers into three risk groups, that is, high, moderate, and low-risk representative of the entire trip. Further, the obtained classification was compared against the subjective evaluation scores given by expert team. The authors concluded that transition probabilities were resulting in better estimation accuracy (overall recognition rate of 93%) compared to other traditional methods. Chen et al. [32] developed driving behavior graph for each driving performance feature by considering the driving pattern exhibited over every three seconds. However, the univariate approach in behavior recognition does not account the interdependencies of driving decisions. Higgs and Abbas [23] considered the combination of driving performance features to segment and cluster the car-following periods and explore the driving patterns of individual drivers. Authors analyzed the driving data of 10-truck and 10-car drivers collected as a part of the 100-car naturalistic driving study. The state-action variables corresponding to the segmented car-following periods were used for clustering. A total of 30 clusters were identified representing different driving patterns and the proportion of patterns varied among drivers. Similarly, Chen and Chen [19] analyzed the base-line events (recorded as a part of SHRP-2 study) concerning the driving performance over the event. Three clusters of events were identified representing three different driving styles. Li et al. [32] proposed an unsupervised framework to segment the driving sequences into fragments and cluster the fragments into descriptive driving patterns. The authors utilized two Bayesian algorithms for segmentation, and two extended latent Dirichlet allocation (LDA) models for clustering the driving patterns. Total four driving patterns were identified common to all the driving maneuvers, where each pattern was a combination of both longitudinal and lateral behavioral characteristics. Although, these studies evaluated the driving maneuvers and driving patterns using the driving profile data, the variation in the individual driving characteristics is still unclear and not conclusive. The objectives of studies, instruments used for data collection, the sampling frequency, the driving features, and the kind of data used for analysis across the previous studies is summarized in Table 1.

2.4. Research Gaps. The literature review shows that the number of classes or groups to which the drivers or driving styles are classified was not consistent among the studies. Also, the definition of the class or group was inconsistent across the studies and varied concerning the study methodologies. Based on the in-depth review of the literature on driving style classification, the research gaps observed are highlighted as follows:

1. Most of the studies classified driving styles based on either the abstract performance features or the predefined thresholds. The aggregated features over entire trip does not indicate the nature of short-term driving decisions and the respective pattern variations in the individuals. Whereas in case of predefined thresholds, the thresholds used across the studies were not uniform.

2. In majority of the studies, the drivers and driving styles were characterized by a single classification of safe or aggressive for the entire trip. Very limited research is available which speaks of variations in the individual’s driving styles.

3. None of the previous studies presented the intra-driver and interdriver behavioral heterogeneity in the instantaneous driving decisions in different driving regimes.

2.5. Research Objectives. Given the research gaps, the study is aimed to pursue the following objectives to explore the extent of variability in individual’s driving patterns in short-term driving decisions.

(a) Segmenting the high-frequency (10 Hz) driving profile data into driving maneuvers (acceleration and braking), and extracting the respective driving performance data.

(b) Identifying different driving patterns using unsupervised machine learning techniques, without using predefined ground truth.

(c) Assigning the driving style classification to individuals and quantifying the driver performance using relative aggression score.

(d) Exploring the individual’s behavioral heterogeneity in instantaneous driving decisions.

3. Methodology

The proposed framework for the current study is presented in Figure 1. The framework is divided into five steps: (1) Data acquisition, (2) Maneuver detection and feature extraction, (3) Dimensionality reduction, (4) Maneuver clustering and driving style classification, and (5) Driver behavior heterogeneity. First, the real-time driving profiles of passenger car drivers were collected using high-frequency (10 Hz) GPS instrumentation. Second, the driving profiles were segmented into acceleration and braking maneuvers, and the respective driving performance features were extracted. As the extracted dataset was unlabeled concerning the nature of maneuver, the unsupervised learning techniques were used to explore the underlying patterns. Therefore, in the third step, the principal component analysis was conducted on acceleration and braking datasets, to reduce the feature dimensionality and to improve the clustering efficiency. Fourth, the k-means clustering was used to group the similar patterns of maneuvers. Then, the identified groups were assigned a driving style classification based on the similar characteristics of performance features. Finally, the proportion of each driving pattern was computed for individuals, and assigned a relative aggression score based on the frequency of the aggressive maneuvers exhibited per
kilometer traveled. Then, the behavioral heterogeneity was presented based on the driving style variations observed with-in, and among the individuals.

3.1. Data Acquisition

3.1.1. Study Stretch. Driving behavior depends on several external factors, among which the road geometry and driving environment play a significant role. To keep the driving environment (with respect to road geometry) uniform for all the participants, a predefined study stretch was used for collecting the driving data. The identical test route ensures each participating driver faces similar geometrical elements and other road infrastructure features. The study stretch is of total 23 km length on the four-lane national highway (NH-64) near Hyderabad city. The selected route consists of 12 intersections, 13 mid-block openings, and four gentle curves.

3.1.2. Driver Participation. The recruitment of drivers for the present study was primarily conducted among the fleet management companies surrounding the Hyderabad city. The participation was voluntary, and the interested candidates were asked to fill a short questionnaire consisting of demographic questions of suitability. The knowledge of the route and how frequent the driver travels on a particular route influences the driving outcomes [33].

Considering the route familiarity and minimum driving experience of one year, total 42 drivers were chosen who were frequent travelers on the study route. The age of the
selected drivers ranged from 19 to 45 years with a mean of 29.9 (sd = 6.8) years and an average driving experience of 7.4 (sd = 5.4) years. The participants were introduced to the goals of the study and informed to drive in their natural way as they would generally drive in day-to-day routine. The study was ethically approved by the Institutional Ethics Committee, Indian Institute of Technology Hyderabad, and a declaration of consent was obtained from each participant before the commence of trip. The data were collected for a minimum of two trips for each participant and were compensated by a gift after the study period.

3.1.3. Driving Data Acquisition System. A passenger car was used to collect the driver behavior data and was equipped with a high-frequency GPS instrumentation and four video-cameras (see Figure 2). The vehicle was instrumented before starting the scheduled trip and removed after completion of the trip. The instrument captures the vehicle speed, longitudinal/lateral acceleration, heading, and positional coordinates at a frequency of 10 Hz along with the synchronized video data. The measured absolute position and speed are accurate to ±3 m and 0.1 km/h, respectively. To minimize the effect of weather and traffic conditions, the data was collected in the dry weather conditions, for the trips (ride requests) that were scheduled during off-peak hours. A total of 98 trips data were collected, comprising 65 hours of driving data over 2254 kilometers.

3.2. Maneuvre Detection. The speed and acceleration are the most commonly used kinematic parameters to distinguish the driver behaviors with respect to the level of aggressiveness [11, 12, 17, 26, 34–37]. Considering the knowledge
from previous studies, present study explores the changes in driving styles based on the driving patterns exhibited in acceleration and braking maneuvers. An algorithm was designed to segment the driving profiles into acceleration and braking maneuvers and extract the respective driving performance features. The algorithm works at three levels to extract the significant acceleration and braking maneuvers. The conceptual representation of the segmentation process is depicted in Figure 3. (i) In the first level, the acceleration and braking maneuvers are identified, and the respective driving performance features are extracted. In this level, the instantaneous rate of change in the speed profile was computed and the segments with the positive and negative rate of change were categorized under acceleration and braking segments, respectively. Keeping the sign of rate of change as a reference, the consecutive positive change points/negative change points were grouped (speed bins) to represent a segment. However, this resulted in many number of segments corresponding to minor fluctuations in the high-frequency GPS data.

To eliminate the fluctuations and identify the actual maneuvers, a threshold was applied on the minimum size of the speed bin ranging from 1 to 6 time stamps (0.1 sec to 0.6 sec). To finalize the appropriate minimum bin size, the maneuvers were manually identified for five data files and checked for the identification rate. The minimum bin size of 3 resulted in higher efficiency in the maneuver identification. Whereas, the higher bin size resulted in losing more number of maneuvers and smaller size lead to identification of more inappropriate maneuvers. Thus, all the bins of size less than 3 were eliminated as GPS fluctuations and the resulting sequential positive bins/negative bins were conglomerated representing the acceleration/braking maneuvers. (ii) In the second level, the insignificant maneuvers resulting due to small speed fluctuations (change in speed <5 kmph) and lower speed values (speed <15 kmph) were eliminated. (iii) In the third level, the free decelerations resulted from the release of acceleration pedal were separated from the braking maneuvers. The part of the deceleration segments with a deceleration value less than 0.05 g were deducted. The above
thresholds used for data extraction were framed after the manual video observation for 300 randomly chosen acceleration and braking maneuvers.

3.3. Feature Extraction. The driving performance features of the final acceleration/braking segments were extracted from the respective driving profile data. The driving performance features include the motion data and the respective derived statistical features (see Table 2). For each identified maneuver, the speed and acceleration performance were extracted along with the respective maximum yaw rate. The yaw rate is the rate of change of heading angle, which indicates the vehicular lateral movement while maneuvering. The acceleration/braking maneuvers associated with high yaw rates tend to reflect the aggressive driving styles. Thus, the maximum yaw rate corresponding to each maneuver was

![Graph showing speed over time for acceleration and braking maneuvers.]

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Variable notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum speed (kmph)</td>
<td>$V_{\text{min}}$</td>
</tr>
<tr>
<td>Maximum speed (kmph)</td>
<td>$V_{\text{max}}$</td>
</tr>
<tr>
<td>Change in speed (kmph)</td>
<td>$\Delta V$</td>
</tr>
<tr>
<td>Duration of maneuver (sec)</td>
<td>$\Delta T$</td>
</tr>
<tr>
<td>Mean speed (kmph)</td>
<td>$V_{\text{mean}}$</td>
</tr>
<tr>
<td>Standard deviation of speed</td>
<td>$V_{\text{sd}}$</td>
</tr>
<tr>
<td>Maximum longitudinal acceleration/deceleration (g)</td>
<td>$L_{A_{\text{max}}} / L_{A_{\text{min}}}$</td>
</tr>
<tr>
<td>Mean longitudinal acceleration/deceleration (g)</td>
<td>$L_{A_{\text{mean}}}$</td>
</tr>
<tr>
<td>Standard deviation of longitudinal acceleration/deceleration</td>
<td>$L_{A_{\text{sd}}}$</td>
</tr>
<tr>
<td>Maximum yaw rate (°/sec)</td>
<td>$Y_{r_{\text{max}}}$</td>
</tr>
</tbody>
</table>

Figure 3: Illustration of segmentation and clustering process.

Table 2: Driving performance features of acceleration and braking maneuvers.
computed from the raw data of heading values. The algorithms for segmentation and feature extraction were coded in Python 3.7.

3.4. Dimensionality Reduction Using PCA. The identified acceleration and braking maneuvers are characterized by a set of driving performance features. Each feature is representative of different driving behaviors and interpreting the nature of maneuver by combining all the features is a difficult or more of a subjective task. Thus, a dimensionality reduction technique was used to reduce the number of original variables to a more interpretable combination. The principal component analysis (PCA) is an unsupervised dimensionality reduction technique, which linearly transforms the correlated variables to a new set of uncorrelated principal components (PC). The number of successive PCs explaining 80–90% of total variance are considered to be a good representative of the dataset [10]. To understand the meaning of PCs, the loadings are computed, which shows the correlation between the PCs and original features. Higher loadings represent stronger correlations between features and the respective PCs. Based on the results of the PCA, the feature dimensionality was reduced prior to clustering.

3.5. Clustering and Driving Style Classification. To explore the underlying patterns and group based on the similarity, we performed the $k$-means clustering on the feature data. $K$-means is an unsupervised technique which groups the data based on the intrinsic similarity in the dataset. $K$-means clustering is the simplest and most widely used machine learning algorithm, which finds similar groups by minimizing the Euclidean distance between centroids. As the algorithm works on unlabeled datasets, the target number of centroids/clusters ($k$) should be defined prior to clustering. In this study, we used elbow and silhouette methods to decide the optimal k-value and validate the clusters.

Further, the identified clusters of driving maneuvers were categorized based on the respective cluster characteristics. Then, the driving style classification was assigned to each maneuver as per the cluster number under which it is grouped.

3.6. Driver Performance Score. After assigning the driving styles to each driving maneuver, the proportion of acceleration and braking maneuvers corresponding to different clusters were computed for individuals. The drivers exhibiting higher proportions of aggressive driving maneuvers indicate unsafe driving behavior and thus need to be identified. In this context, the driving performance of each driver was quantified based on the number of aggressive maneuvers exhibited over the observed period. However, the thresholds for safe or unsafe driving behaviors have not been established in the literature. Therefore, in this study, the performance score was computed relative to the maximum number of aggressive maneuvers per kilometer traveled among all the drivers. The drivers exhibiting higher aggression while accelerating, braking, and speeding were considered as a benchmark to assign the relative aggression score to other drivers. The number of aggressive patterns per kilometer traveled is normalized using the reference maximum value, such that each driver takes a score between zero to 100. The lower scores represent lower levels of aggression on a relative scale.

4. Results and Discussion

4.1. Dataset Details. The acceleration and braking maneuvers were identified using the designed algorithm and the final dataset consists of 7548 acceleration maneuvers and 6156 braking maneuvers corresponding to 42 drivers. Each maneuver is characterized by ten driving performance features, which includes minimum speed, maximum speed, mean speed, standard deviation of speed, maximum acceleration/deceleration, average and standard deviation of acceleration/deceleration, change in speed during the maneuver, duration of the maneuver, and maximum yaw rate. Since the units of features are not uniform, the data were prepared for analysis by scaling through the Z-score standardization technique.

4.2. Dimensionality Reduction. The PCA was performed on the scaled data (Z-score standardization) of both acceleration and braking datasets. The correlation circles were developed to understand the correlations between the PCs and the original variables. The PC1 and PC2 loadings against each feature for both datasets are shown in Figures 4(a) and 4(b). The direction of feature lines represents the positive or negative correlations and lengths indicate the strength of the correlation. In other words, the features closer to the circumference of the circle are more important to interpret the respective components [38]. For both acceleration and braking datasets, most of the features are cross-loaded on each PC and are showing higher correlations with PCs. To aid the interpretation, the loadings were rotated using Varimax rotation; the respective rotated component (RC) loadings are shown in Table 3. Four components were chosen as the cumulative explained variance is greater than 85% for both the datasets. The loadings with high magnitude are highlighted in bold (see Table 3), as the higher absolute loadings indicate stronger correlations between features and components.

In case of acceleration dataset (Table 3), RC1 is showing significant and negative correlation with all the longitudinal acceleration features. RC2 shows stronger positive correlations with the mean and standard deviation of speed during the maneuver. RC3 is associated with change in speed and duration of the maneuver and RC4 is highly correlated with the speed features and yaw rate. The similar correlations can be observed for braking dataset, in which RC1 is associated with longitudinal deceleration features and RC2 shows high correlation with speed features and yaw rate. RC3 is showing strong association with change in speed and duration of maneuver and RC4 is significantly correlated with mean and standard deviation of speed in the maneuver.
The PCA resulted in four components, each indicating different driver behavior aspects. However, clustering each component individually (respective features) would result in four sets of clusters for both acceleration and braking datasets, making the interpretation of results a tedious and non-conclusive task. Hence, the rotated components are further grouped based on the interpretation of correlations between RCs and original features. In case of both the datasets, RC2 and RC4 significantly represent the speed choices and speed variability exhibited during the maneuver, whereas RC1 and RC3 are explaining the acceleration/braking behavior and the respective change in speed over the full duration of the maneuver. It is possible to exhibit the aggressive acceleration/braking behavior at lower speeds, and similarly, the high-speed behavior may not be always associated with aggressive accelerations/braking behavior. In other words, the individuals might perform different frequencies of aggressive low-speed maneuvers, aggressive-high speed maneuvers, nonaggressive low-speed maneuvers, and nonaggressive high-speed maneuvers. Each combination speaks a different aspect of driver such as nonaggressive low-speed indicating base-line behaviors, and aggressive-low speed maneuvers indicating rear-end collision tendencies, whereas the nonaggressive high-speed signifies the driving efficiency, and aggressive-high speed indicates the high-risk/faulty behaviors. Thus, to aid the detailed analysis of longitudinal control, the acceleration/braking behavior, and the speed behavior are analyzed independently for each maneuver.

To this end, the original variables are divided into two groups based on the observed correlations between rotated components and the driving performance features. The first group of variables \((L_{A_{\text{max}}} / L_{A_{\text{min}}}, V_{\text{mean}}, V_{\text{sd}}, \Delta V, \Delta T)\) represents the acceleration/braking behavior, and the second group of variables \((V_{\text{max}}, V_{\text{min}}, \Delta V_{\text{mean}}, \Delta T_{\text{mean}}, V_{\text{sd}}, Y_{r_{\text{max}}})\) indicates the speed behavior exhibited while accelerating or braking. Further analysis to identify the underlying patterns of acceleration/braking behavior, and speed behavior is explained in the following section.

### 4.3. Clustering and Driving Style Classification

#### 4.3.1. Level-Wise Clustering

Considering the results of principal component analysis, the k-means clustering analysis was performed at two levels on the acceleration and braking datasets. In the first level of clustering, the driving patterns in terms of the acceleration/braking performance are detected. In the second level of clustering, the speed patterns exhibited during acceleration/braking are identified by clustering the speed variable data. The features used for each level of clustering are given in Table 4.

#### 4.3.2. Level-One Clustering

To characterize the longitudinal control concerning the acceleration and braking behavior, the k-means clustering was performed on level-1 features

---

**Table 3: Varimax rotated loadings of acceleration and braking datasets. Note: RC stands for rotated components.**

<table>
<thead>
<tr>
<th>Features</th>
<th>RC1</th>
<th>RC2</th>
<th>RC3</th>
<th>RC4</th>
<th>Features</th>
<th>RC1</th>
<th>RC2</th>
<th>RC3</th>
<th>RC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_{\text{min}})</td>
<td>0.321</td>
<td>0.116</td>
<td>0.078</td>
<td>0.908</td>
<td>(V_{\text{max}})</td>
<td>–0.007</td>
<td>–0.888</td>
<td>–0.352</td>
<td>–0.179</td>
</tr>
<tr>
<td>(V_{\text{max}})</td>
<td>0.171</td>
<td>0.142</td>
<td>–0.290</td>
<td>0.911</td>
<td>(V_{\text{min}})</td>
<td>–0.320</td>
<td>–0.879</td>
<td>0.198</td>
<td>–0.131</td>
</tr>
<tr>
<td>(\Delta V)</td>
<td>–0.383</td>
<td>0.043</td>
<td>–0.849</td>
<td>–0.110</td>
<td>(\Delta V)</td>
<td>–0.495</td>
<td>–0.112</td>
<td>0.815</td>
<td>0.051</td>
</tr>
<tr>
<td>(\Delta T)</td>
<td>0.260</td>
<td>–0.037</td>
<td>–0.909</td>
<td>0.106</td>
<td>(\Delta T)</td>
<td>–0.111</td>
<td>0.025</td>
<td>–0.955</td>
<td>0.044</td>
</tr>
<tr>
<td>(V_{\text{mean}})</td>
<td>0.010</td>
<td>0.904</td>
<td>0.015</td>
<td>0.121</td>
<td>(V_{\text{mean}})</td>
<td>–0.018</td>
<td>–0.184</td>
<td>–0.028</td>
<td>–0.753</td>
</tr>
<tr>
<td>(V_{\text{sd}})</td>
<td>–0.052</td>
<td>0.913</td>
<td>–0.018</td>
<td>0.045</td>
<td>(V_{\text{sd}})</td>
<td>0.082</td>
<td>0.018</td>
<td>0.032</td>
<td>–0.829</td>
</tr>
<tr>
<td>(L_{A_{\text{max}}})</td>
<td>–0.920</td>
<td>0.008</td>
<td>–0.124</td>
<td>–0.215</td>
<td>(L_{A_{\text{min}}})</td>
<td>–0.922</td>
<td>–0.164</td>
<td>0.179</td>
<td>0.020</td>
</tr>
<tr>
<td>(L_{A_{\text{mean}}})</td>
<td>–0.869</td>
<td>0.056</td>
<td>0.045</td>
<td>–0.319</td>
<td>(L_{A_{\text{mean}}})</td>
<td>–0.934</td>
<td>–0.184</td>
<td>–0.002</td>
<td>0.055</td>
</tr>
<tr>
<td>(L_{A_{\text{sd}}})</td>
<td>–0.915</td>
<td>0.001</td>
<td>0.047</td>
<td>–0.261</td>
<td>(L_{A_{\text{sd}}})</td>
<td>0.957</td>
<td>0.173</td>
<td>–0.011</td>
<td>–0.020</td>
</tr>
<tr>
<td>(Y_{r_{\text{max}}})</td>
<td>–0.386</td>
<td>0.004</td>
<td>–0.214</td>
<td>–0.692</td>
<td>(Y_{r_{\text{max}}})</td>
<td>0.256</td>
<td>0.764</td>
<td>–0.260</td>
<td>–0.026</td>
</tr>
</tbody>
</table>

**Figure 4:** Correlation circles for (a) acceleration maneuvers (b) braking maneuvers.
and the principal components are shown in Table 7. For data of acceleration and braking maneuvers into two clusters, k-means clustering was performed to group the speed separation is shown in Figure 7, indicating well-separated obtained through elbow and silhouette methods. The cluster separation is shown in Table 8. The speed variability appears to be nearly equal in both the clusters with cluster-2 taking relatively higher values (see Figure 8). However, the speed choices are significantly higher in cluster-2 and cluster-1 appears to take low values of speed variables and higher value of yaw rate. Given the interpretation, the cluster-1 is characterized to represent the nominal-speed behavior and cluster-2 indicating the high-speed behavior for both acceleration and braking maneuvers. The cluster centers of all the performance features and the cluster categories are presented in Table 8. The speed clusters of acceleration dataset, and braking dataset are representing the similar speed behavior, which indicates the uniformity in speed choices irrespective of the type of maneuver.

The proportion of maneuvers shared by each acceleration/braking pattern and the respective share of nominal and high-speed clusters are presented in Table 9. The major proportion of maneuvers (56.9% accelerations and 61.6% braking maneuvers) are observed to be classified under smooth driving styles, in which 19.8% accelerations and 18.1% braking maneuvers are also associated with nominal speeding patterns, indicating the safe driving behavior. In case of moderate and aggressive driving maneuvers, 18.2% accelerations and 15% braking maneuvers are associated with high-speed patterns indicating unsafe behaviors. Further, the aggressive maneuvers characterized by aggressive acceleration/braking related with high-speed behavior are observed in 6.2% of accelerations and 8.3% of braking maneuvers, reflecting the actual faulty/critical driving behaviors. Although, the proportions of critical maneuvers are relatively lower, the role of such maneuvers is vital in assessing the driving performance of individuals in day-to-day driving. Since, both the speed and acceleration behavior defines the level of driving safety, further, we analyzed the variations in accelerations, braking, and speeding at individual driver level, which shows the in-detail driving nature of individuals.

4.4. Validation of Driving Patterns. Based on the insights from literature, the ground-truth about driving style of a driver is determined by means of traffic accident data, or subjective evaluation data (self-report/expert scoring), or safety-critical event data [11, 15, 20, 22, 31]. In this study, neither the crash-related information nor the self-reported
driving violation data could be obtained for the participated professional drivers. Hence, the frequency of safety-critical event data is used as a means to describe the driving style of individuals. The safety-critical patterns are obtained through unsupervised techniques and needs to be verified prior proceeding to driving style classification. To this end, the cluster characteristics of critical patterns are compared against the feature thresholds used across studies.

**Table 6: Level-1 Cluster centers of acceleration and braking maneuvers.**

<table>
<thead>
<tr>
<th>Acceleration clusters</th>
<th>ΔV</th>
<th>ΔT</th>
<th>LA(_{\text{max}})</th>
<th>LA(_{\text{mean}})</th>
<th>LA(_{\text{sd}})</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.00</td>
<td>5.51</td>
<td>0.15</td>
<td>0.06</td>
<td>0.036</td>
<td>Smooth</td>
</tr>
<tr>
<td>2</td>
<td>21.08</td>
<td>12.27</td>
<td>0.18</td>
<td>0.07</td>
<td>0.038</td>
<td>Moderate</td>
</tr>
<tr>
<td>3</td>
<td>15.65</td>
<td>4.79</td>
<td>0.26</td>
<td>0.11</td>
<td>0.059</td>
<td>Aggressive</td>
</tr>
<tr>
<td>Braking clusters</td>
<td>ΔV</td>
<td>ΔT</td>
<td>LA(_{\text{min}})</td>
<td>LA(_{\text{mean}})</td>
<td>LA(_{\text{sd}})</td>
<td>Classification</td>
</tr>
<tr>
<td>1</td>
<td>10.35</td>
<td>6.44</td>
<td>0.17</td>
<td>0.06</td>
<td>0.041</td>
<td>Smooth</td>
</tr>
<tr>
<td>2</td>
<td>35.35</td>
<td>13.41</td>
<td>0.29</td>
<td>0.10</td>
<td>0.065</td>
<td>Moderate</td>
</tr>
<tr>
<td>3</td>
<td>18.17</td>
<td>4.93</td>
<td>0.33</td>
<td>0.13</td>
<td>0.085</td>
<td>Aggressive</td>
</tr>
</tbody>
</table>
### Table 7: The PC loadings of speed variables for acceleration and braking datasets.

<table>
<thead>
<tr>
<th>Features</th>
<th>Acceleration maneuvers</th>
<th>Braking maneuvers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PC1</td>
<td>PC2</td>
</tr>
<tr>
<td>$V_{\text{min}}$</td>
<td>0.923</td>
<td>−0.262</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>0.900</td>
<td>−0.208</td>
</tr>
<tr>
<td>$V_{\text{mean}}$</td>
<td>0.466</td>
<td>0.780</td>
</tr>
<tr>
<td>$V_{\text{sd}}$</td>
<td>0.382</td>
<td>0.833</td>
</tr>
<tr>
<td>$Y_{\text{rmax}}$</td>
<td>−0.726</td>
<td>0.348</td>
</tr>
</tbody>
</table>

### Table 8: Speed cluster centers for acceleration and braking maneuvers.

<table>
<thead>
<tr>
<th>Speed clusters</th>
<th>$V_{\text{min}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{mean}}$</th>
<th>$V_{\text{sd}}$</th>
<th>$Y_{\text{rmax}}$</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>acceleration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22.13</td>
<td>36.05</td>
<td>31.38</td>
<td>15.54</td>
<td>13.22</td>
<td>Nominal</td>
</tr>
<tr>
<td>2</td>
<td>49.15</td>
<td>61.13</td>
<td>44.84</td>
<td>16.64</td>
<td>5.45</td>
<td>High speed</td>
</tr>
<tr>
<td>braking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17.83</td>
<td>37.76</td>
<td>31.93</td>
<td>16.37</td>
<td>14.53</td>
<td>Nominal</td>
</tr>
<tr>
<td>2</td>
<td>48.08</td>
<td>61.97</td>
<td>45.59</td>
<td>17.13</td>
<td>5.29</td>
<td>High speed</td>
</tr>
</tbody>
</table>

### Table 9: Proportion of maneuvers shared by each pattern.

<table>
<thead>
<tr>
<th>Acceleration</th>
<th>Smooth</th>
<th>Moderate</th>
<th>Aggressive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal speed</td>
<td>1491 (19.8%)</td>
<td>453 (6%)</td>
<td>1429 (18.9%)</td>
<td>3373 (44.7%)</td>
</tr>
<tr>
<td>High speed</td>
<td>2803 (37.1%)</td>
<td>907 (12%)</td>
<td>465 (6.2%)</td>
<td>4175 (55.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>4294 (56.9%)</td>
<td>1360 (18%)</td>
<td>1894 (25.1%)</td>
<td>7548 (100%)</td>
</tr>
<tr>
<td>Braking</td>
<td>Smooth</td>
<td>Moderate</td>
<td>Aggressive</td>
<td></td>
</tr>
<tr>
<td>Nominal speed</td>
<td>1116 (18.1%)</td>
<td>677 (11%)</td>
<td>759 (12.4%)</td>
<td>2555 (41.5%)</td>
</tr>
<tr>
<td>High speed</td>
<td>2678 (43.5%)</td>
<td>413 (6.7%)</td>
<td>513 (8.3%)</td>
<td>3601 (58.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>3794 (61.6%)</td>
<td>1090 (17.7%)</td>
<td>1272 (20.7%)</td>
<td>6156 (0%)</td>
</tr>
</tbody>
</table>

### Figure 7: Silhouettes of (a) Speed_acceleration clusters (b) Speed_braking clusters.

### Figure 8: Bi-plots of speed clusters identified in (a) acceleration maneuvers (b) braking maneuvers.
The clustering results presented in Table 9 indicates that the aggressive acceleration and aggressive braking maneuvers associated with high-speed behavior are representing the critical/unsafe patterns exhibited by individuals. These patterns are quantitatively represented by means of the cluster centroid plus/minus one standard deviation (SD), and shown in Table 10. The identified aggressive acceleration pattern is in the range of 0.2 to 0.32 g acceleration, and the aggressive braking pattern is varying over the maximum deceleration of 0.25 to 0.41 g, corresponding to the vehicle speed of 50.20 to 73.66 kmph. These aggressive maneuvers are observed to be falling in the range of thresholds used in the previous studies by Johnson and Trivedi [11], Chen et al. [22], Paefgen et al. [26], Fazeen et al. [27], and Bergasa et al. [29]. Thus, the obtained clusters can be effectively used to assess the driving styles of individuals.

4.5. Driver Behavior Heterogeneity. The identified driving patterns are representing the acceleration and braking behavior ranging from smooth to aggressive and the respective speed patterns are identified to be nominal and high-speed behavior. The proportion of maneuvers under each identified pattern for all the 42 drivers are computed and shown in Figure 9. Drivers are not executing driving maneuvers in a single pattern, rather each individual is showing variations in acceleration, braking, and speeding behavior during the study period. In other words, drivers are exhibiting each driving pattern at a certain proportion, which is different from one driver to another, and also from one driving regime to another. For instance, the driver D1 is exhibiting aggressive behavior in 16% of the accelerations, whereas 35% of the braking maneuvers are associated with aggressive behavior. Similarly, each individual driver is showing heterogeneity in the patterns of longitudinal control. No driver was found to constantly exhibit safe or aggressive behavior in the longitudinal control during the entire trip. The differences in driving conditions may result in different driving responses by the same driver, which leads to changing driving styles within the driving period. Thus, the individual’s driving styles are found to be inconsistent and heterogeneous in both acceleration and braking regimes. The study findings are consistent with the existing literature on individual driving characteristics [19, 22, 23]. In the study of Higgs and Abbas [23], the car drivers have shown different proportions of behavior patterns (ranging from 0 to 30 patterns) in the car-following behavior, that differs among individuals. Chen and Chen [19] and Chen et al. [22] also showed that each driver was exhibiting a set of driving patterns that are different from one driver to another.

As shown in Figures 9(a) and 9(b), the smooth acceleration and braking behavior is the dominating driving pattern in most of the drivers, which indicates the base-line or “normal” driving behavior of each individual. The proportion of moderate and aggressive behaviors are changing from one driver to another, which indicates different levels of aggression among individuals. In addition, the speed behavior in accelerations and braking maneuvers (see Figures 9(c) and 9(d)) are notably different among the drivers.

4.5.1. Driving Performance Score. The level of aggression of each acceleration and braking maneuver depends on the combination of acceleration/braking pattern and the respective speed pattern. The aggressive behavior is represented by the aggressive acceleration/braking maneuvers, which are further associated with high-speed behavior. In order to separate the aggressive behavior of each driver, the number of maneuvers of aggressive acceleration, aggressive braking, and the high-speed patterns are computed per kilometer traveled. The speed clusters of acceleration and braking maneuvers are observed to be representing similar characteristics in terms of the driving performance features (see Table 8). Thus, the speed clusters of acceleration and braking maneuvers are aggregated and the total number of high-speed maneuvers per kilometer traveled during acceleration and braking are summed up to propose a single speed score as shown in Figure 10.

The computed relative driver aggression in acceleration, braking, and speed behavior are shown in Figure 10. The driver D24 has exhibited the highest number of aggressive maneuvers per kilometer traveled, in acceleration and braking behavior and the driver D25 has shown the dominance in high-speed behavior. The level of aggression of individuals was found to be different in acceleration and braking regimes, and changing among drivers. For example, the drivers (D6, D18, D27, D29, D42) are showing higher aggression score (>60) in accelerations, whereas the respective individuals’ braking aggression scores are observed to be lower than 40. Similarly, the drivers D1, D11, and D22 are exhibiting higher levels of aggression (>50) in braking maneuvers and moderate-to-lower aggression in acceleration behaviors. The aggression scores are indicative of drivers’ propensity to exhibit harsh driving maneuvers and are particularly relevant in driving assistance.

Each individual is showing a unique combination of driver aggression in acceleration, braking, and speed behavior. The variation in the interdriver aggression might be due to the individual driving characteristics or the external influencing factors such as traffic and road geometry. In the
present study, all the drivers made trips on the same road stretch resulting in similar exposure to the geometric elements. Although, the participants drove vehicle during same time of the day, the observed variations may be due to varying traffic conditions. Thus, the identified patterns are considered to be representing more of driver specific variation or the habitual driving styles of each individual corresponding to the exposed traffic.
5. Conclusion

In the context of assisting the driver to optimize the driving behavior, it is essential to understand the individual's instantaneous driving decisions. Most of the studies in this context were undertaken in the developed countries and no studies analyzed the behavioral heterogeneity of Indian drivers. Moreover, majority of the existing studies conceptualized the driving styles using predefined thresholds of abstract driving features, and the drivers were characterized as safe or aggressive. This study takes the advantage of continuous driving profiles collected using high-frequency GPS instrumentation to explore the extent of driving heterogeneity in different driving regimes. The framework is designed to conceptualize the driving styles exhibited in the longitudinal control by extracting the short-term driving decisions and group the similar behaviors by means of unsupervised techniques.

The methodology is implemented on the driving profiles of 42 professional car drivers, captured in the naturalistic driving conditions over a defined study stretch. An algorithm is developed to extract the acceleration and braking maneuvers and the respective driving performance features as representation of driving decisions in longitudinal control. The similar patterns of decisions are identified using k-means clustering technique and interpreted using the principal component analysis. Total three patterns of acceleration and braking behavior each, characterized by smooth, moderate, and aggressive behaviors, which are associated with nominal and high-speed behavior are identified. The proportion of each pattern observed during the study period of individuals revealed the interesting theories about the extent of driver behavioral variations. The drivers showed varying proportions of patterns in each driving regime, indicating the heterogeneity in driving behaviors. In addition, the driving performance scores were different among individuals in both acceleration and braking maneuvers. No driver was found constantly exhibiting either safe or aggressive driving decisions over the observed driving period.

6. Research Contribution

(a) In the existing literature, most of the studies derived driving styles using predefined thresholds of safety-critical events or the abstract driving performance aggregated over the study period. The thresholds were not consistent across studies and abstract performance did not represent the faulty behaviors in short-term driving decisions. In this context, this study analyzed the driving profile data using unsupervised techniques, without any prior ground
truth. Also, the study proposed a methodology to segment the continuous driving profiles as maneuvers of instantaneous driving decisions in the longitudinal control.

(b) Further, the previous studies assigned a single characterizing driving pattern to an individual. The heterogeneity in driving behaviors pertained to the changes in driver attributes and driving environment was not considered. In the light of these findings, this study explored the individual’s behavioral variations with-in driving period. Rather assigning a single representative classification, the actual at-risk behaviors exhibited by each driver are identified. Moreover, the unsupervised approach presented in this study would theoretically account for the unobserved influencing factors effecting the heterogeneity in individual’s driving patterns.

The findings of the study emphasize the need for continuous monitoring of driver behavior for driver assistance and personalized feedback provision. Thus, the study methodology and study findings are useful for safety risk profiling of drivers, and also risk scoring of roadway segments for hot-spot analysis.

7. Study Limitations and Future Scope

This study accounts few limitations which provides scope for future research.

Experimental design: As the study is short-term instrumented vehicle study, the driving data is collected using instrumented vehicle over a short-period of time. The presence of instrumentation might influence the actual driver behavior even though the drivers were informed that the collected data would be utilized for only research purposes and not for any legal enforcement. The instrumentation used in this study recorded the continuous vehicle kinematic data and the video-data at a frequency of 10 Hz. The advanced sensors such as LiDAR and eye trackers can be used to capture the other prominent driving parameters like headway choices and distractions, that would help to understand the comprehensive nature of various driving patterns.

Influence of amount of the data: Driving data for longer duration and for more number of drivers from different age groups and gender would give detailed insights about the variations in the driving patterns. With the multiple trip data of individuals, the researchers can evaluate the stability in driving behaviors and consistency in different driving patterns.

The methodologies presented in the current study are replicable by future researchers to assess the driving patterns of individuals. However, the study results are subjected to change concerning the local driving habits and the traffic laws corresponding to the study location. The future scope of the study can be extended to understand the driving pattern variations in other driving regimes like car-following or lane-changing scenarios for different vehicle types. The future research efforts can also be focused on certain critical driving patterns to better understand the influencing factors behind such typical behaviors.

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the Science and Engineering Research Board, Department of Science and Technology, Government of India (ECR/2018/002407).

References


