

Research Article

Identifying Driver Behavior in Preturning Maneuvers Using In-Vehicle CANbus Signals

M. Zardosht, S. S. Beauchemin , and M. A. Bauer 

The University of Western Ontario, London, Ontario, Canada N6A 5B7

Correspondence should be addressed to S. S. Beauchemin; sbeauche@uwo.ca

Received 27 June 2018; Revised 21 September 2018; Accepted 30 October 2018; Published 21 November 2018

Academic Editor: Jaeyoung Lee

Copyright © 2018 M. Zardosht et al. 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.

Our objective in this contribution is to categorize driver behavior in terms of preturning maneuvers. We analyze driving behavior in an urban environment prior to turns using data obtained from the CANbus of an instrumented vehicle during a one-hour driving period for 12 different individuals. CANbus data streams such as vehicle speed, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration are collected and analyzed for 5, 10, and 15 seconds of driving prior to each turn. We consider all turns for each driver and extract statistical features from the signals and use cluster analysis to categorize drivers into groups reflecting different driving styles. The results show that using this approach we can effectively cluster drivers into two groups. The results show consistency in the membership within a cluster throughout the different timeframes. We conclude that driver behavior classification from such data streams is possible and we hope in the near future to devise driver descriptors that include additional maneuvers.

1. Introduction

As the number of vehicles and road mileage increases, traffic safety has become one of the main issues for governments and manufacturers. Traffic accidents are one of the main reasons for injuries today. According to a study in 2015 for The National Highway Traffic Safety Administration (NHTSA), in approximately 94% of the accidents examined, the major reason was driver error [1]. In order to improve traffic safety and traffic efficiency, it is essential to understand the characteristics of driver behavior and study its relationship with both vehicle and traffic systems. Understanding driver behavior can help us build models of drivers to improve Advanced Driver Assistance Systems (ADASs), vehicle safety, and privacy and help to detect risky driving styles.

In this contribution, normal driving behavior is analyzed in order to identify different driving styles. Our focus is on the driver's behavior before turns [2]. Several important driving signals have been extracted from an automobile's Controller Area Network (CANbus) data as driving indicators. Vehicle speed, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration are the signals used in this study. Since the activity in preparing for a turn is a challenging and complex driving behavior, and given that it has not

been studied, we decided to focus on this activity. Our goal is to explore whether there are categories of driving behavior that can be identified from driving signals during small periods of time before turning maneuvers and to determine the distinctive characteristics of those categories. Our assumption is that there are different styles of driving when approaching a turning maneuver. We extract statistical features from the driving signals prior to each turn and cluster drivers based on feature vectors.

The rest of this contribution is structured as follows. The next section covers related work on identifying and clustering driving behavior. The data we used for our study is briefly described in Section 3. Section 4 describes our approach to feature selection and clustering; results are presented in Section 5. Finally, Section 6 presents our conclusions and discusses further research.

2. Related Work on Driving Behavior

Since the 1950s, modeling driving behavior has been important to the traffic scientist [3]. While driving is one of the most common tasks individuals perform, it is yet a highly complex task, as it involves multiple essential subtasks and

can be affected by many internal and external factors. Driving consists of a series of complex decisions and actions that a driver performs based on the current traffic and physical environment. It is obvious that different drivers have different driving behavior given a traffic situation. Drivers differ in how they use the pedals, in the way they turn the steering wheel, how they keep their eyes on the road, how much distance they keep when following a car, and so on [4, 5]. In fact, these differences are key factors in identifying individual driving behavior. As a consequence, various driving behaviors need to be understood in order to personalize Intelligent Transportation System (ITS) applications for drivers with different driving styles.

In addition to ITS personalization, identifying different types of driver behavior may also help improve traffic safety by identifying safe or unsafe driving styles [6], aggressive or normal drivers [7, 8], or distracted drivers [9]. Another application of driver behavior analysis is in the area of security and privacy, for example, by looking at driver identification based on driving style [10]. Consequently, an awareness of the differences between driving styles and behavior can augment intelligent Advanced Driver Assistance Systems (ADAS) to improve the performance of each individual driver. This may be accomplished by way of the ability of ADAS to identify and classify drivers based on driving parameters (speed, brake and accelerator pedal pressure, steering wheel angle, and acceleration, for instance) and performing comparisons with an aggregate, canonical driver descriptor representing normal driving. Such an architecture for ADAS could easily establish the problematic elements of driver behavior (as expressed by the driving parameters), as they deviate from the norm, and subsequently provide this information to drivers for behavioral self-correction.

Intelligent vehicle designers need to understand driving behavior in order to make driving assistance systems work properly in dynamic traffic situations. Autonomous vehicle designers need them to make driving driver-free. Traffic engineers need them to improve the safety and reliability of roads and related infrastructure. Driving behavior, as every other human-related task, is complicated and difficult to analyze [11, 12]. Our research is focused on the analysis of driver data, specifically focused on driver behavior based on in-vehicle signals using statistical features.

Current literature emphasizes the need for a comprehensive method to understand normal driving behavior and to distinguish between different driving styles [13, 14]. Drivers have different driving behaviors because there exist various different aspects of driving. They differ in how they use the pedals, how fast or slow they turn a steering wheel, or how often they change speed. Several scholars have considered the role of these driving signals in characterizing driver behavior. Aarts and Van Schagen [15] emphasized the role of vehicle speed in driving behavior in both road and traffic safety. They discussed the relation between vehicle speed and probability of crashes and indicated that high speed not only makes collisions more severe, but also increases the risk of accident occurrence. The relation between vehicle speed and the number of crashes was also investigated in a study conducted by Fildes *et al.* [16].

Miyajima *et al.* [4] conducted a study on modeling driver behavior for the purpose of driver identification. In their study, they show that gas and brake pedal operation signals efficiently model individual driver differences. Gaussian Mixture Models (GMM) are used to model the spectral features of pedal signals. An identification rate of 76.8% was achieved, which suggests that each driver has a distinctive pattern in pedal pressure. The spectral features are frequency-based features obtained by converting the time-based signal into the frequency domain using a discrete Fourier Transform. In another study, Ohta [17] shows that driver behavior differs among drivers by considering the distance they keep when following other vehicles. Drivers were instructed to follow a vehicle at distances that were most comfortable for them, such as most comfortable distance and minimum safe distance. Based on that, a temporal comfort zone for individuals following another vehicle was defined, resulting in time intervals comprised between 1.1 and 1.7 seconds. It was found that any driving behavior in which a driver was closer than 1.1s is critical, while more than 1.7s behind was considered uncomfortable because it was against the social norm. Based on these thresholds for a comfort zone, driver behavior could be classified into three different categories depending on how frequently a driver drove in a particular zone. Alternatively, Chen *et al.* built a driving behavior classification system to model safe and unsafe driving behaviors [6]. In their study, vehicle operation data, including vehicle speed, engine RPM, throttle position, and calculated engine load, were collected via OBD (On-Board Device) interfaces and used AdaBoost algorithms for classification.

Choi *et al.* perform driver analysis using Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) on in-vehicle CANbus signals such as steering wheel angle, brake status, acceleration, and vehicle speed [9]. Three different classification tasks were conducted in this study: action classification, distraction detection, and driver identification. Action classification consisted of categorizing long-term driving behaviors such as turning, lane changing, stopping, and constant driving. In distraction detection the goal was to detect if the driver was distracted by secondary tasks such as working with cellphone and GPS. Driver identification was concerned with driver classification based on driving behavior characteristics. For driver identification they generated HMMs and GMM models based on driving signals collected during neutral and distracted driving periods. They were able to identify drivers based on analysis of signals during distracted and neutral driving about 25% of the time. However, they did not evaluate their model on normal driving behavior to study the differences between drivers in normal driving conditions.

In research performed by Kalsoom and Halim [18], individual driver behaviors were classified based on each driver's statistical driving features such as the ratio of indicators to turns, the number of brake uses, the number of horn uses, the average speed, the maximum speed, and the transmission gear. They applied K-means and hierarchical clustering on their experimental data in order to classify them into slow, normal, and fast driving styles over the entire driving sequence. The data in this study was collected from



FIGURE 1: RoadLAB: the instrumented vehicle depicted with its rooftop stereoscopic vision system and driver gaze tracker (top-left).

a driving simulator and contained 5 minutes of driving data per driver. The clustering was not based on the values of time-series driving signals. Instead, they used the number of operative devices, such as brake, gear, clutch, horn, and turn signal indicators. They also used the average and maximum of gears and average and maximum speed in a 10-second time window. The results of the clustering analyses were inconclusive, with both Hierarchical Clustering Analysis (HCA) and K-means employed with various numbers of clusters. Clustering with HCA resulted in most of the data being grouped into a single class.

In a study by Wu *et al.* [19], GPS data was collected from typical driving of commercial vehicles by professional drivers. After using factor analysis, the attributes related to driving behavioral characteristics were extracted from the GPS data. A total of 8 speed and acceleration related features were extracted for each driver and combined into four aggregated attributes: acceleration-deceleration, speeding-prone, acceleration, and deceleration. Using each of these attributes as indicators, four different cluster analyses of driver behavior were performed with hierarchical clustering. In each analysis, they identified five clusters that indicated how risky the driver behavior was, considering each attribute: minimal, slight, moderate, serious, or severe. The four cluster analyses were performed on 50 drivers. The results were not consistent across all attributes and drivers, but there were some drivers that showed a higher degree of risky driving behavior for all attributes. They characterized the driving style of each driver over the entire driving period.

In both of these studies, the researchers used cluster analysis to categorize drivers across the entire driving sequence. This may be very challenging because over long periods driver behavior may average out to a point at which cluster analysis fails. In light of this, we have chosen to focus on a short driving period immediately before an identifiable driving event, such as turns.

Higgs and Abbas investigated the hypothesis that drivers may have different driving behavior in their daily driving tasks [20]. In order to investigate this assumption, three different truck drivers, considered as low, medium, and high risk, performed 10 different car-following periods. They first divided the car-following periods into different segments prior to performing K-means cluster analysis on them. Each driver displayed a specific distribution of behaviors. However some of these behaviors were common between drivers but at different frequencies. In another study performed by the above authors [5], a two-step algorithm was introduced to segment and cluster car-following behavior based on eight

state-action variables, in order to define driving patterns of drivers. They defined driver behavior as the way a driver responds to the current driving state (the vehicle speed, distance from the following vehicle, and so on) by performing a specific action (steer or push brake pedal). In this study, each car-following period was divided into similar driving state-action signals in the segmentation phase. Specific segments might repeat several times and the clustering phase attempted clustering the repeated segments into groups using K-means as the clustering method.

Driver behavior at intersections has also been studied, mostly from the safety aspect, such as in a review of studies written by Shirazi and Morris in which intersection design and planning are investigated [21]. Alternatively, Aoude *et al.* [22] classify drivers at intersections as compliant and noncompliant with respect to intersection rules using algorithms based on Support Vector Machines (SVM) and HMM. Behavior at intersections with traffic lights has also been characterized in terms of driver reaction on the onset of a yellow phase trigger [23].

3. Data Sources

In order to have a better understanding of normal driving behavior, we need detailed observations on the driver, vehicle, and traffic environment. In 2011, Beauchemin *et al.* [24] equipped a modern vehicle with OBD II CANbus (on-board diagnostics (OBD) is a vehicle's self-diagnostic and reporting capability) channels with stereoscopic video cameras, a GPS system, and a 3D gaze tracker to record the driver's head pose and gaze direction. This vehicle, known as RoadLAB, was used to collect driving data from the vehicle's internal network, the environment, and the driver. This instrumented vehicle, shown in Figure 1, monitors and records the following:

- (i) Environment: 3D front view of traffic environment with two calibrated stereo cameras
- (ii) Vehicle: the internal vehicle functions via a CANbus interface
- (iii) Driver: the driver cephalo-ocular behavior described by head pose and 3D gaze direction

The dataset was recorded with 16 different test drivers, consisting of 7 males and 9 females, with ages ranging between 20 and 47. The drivers drove normally through a predetermined path in an urban area inside the city of London, Ontario. Each driver drove around 28.5 kilometers

TABLE 1: Characteristics of drivers and environment.

Driver	Age	Gender	Weather	Temp.
1	37	Male	Sunny	29C
2	37	Male	Sunny	31C
3	41	Female	Sunny	23C
4	41	Male	Sunny	24C
5	37	Female	Partly Cloudy	27C
6	22	Female	Partly Cloudy	21C
7	31	Female	Sunny	21C
8	21	Male	Sunny	27C
9	21	Female	Partly Cloudy	24C
10	20	Male	Sunny	8C
11	22	Female	Sunny	12C
12	24	Female	Partly Cloudy	18C
13	23	Male	Partly Cloudy	19C
14	47	Female	Sunny	7C
15	44	Female	Partly Cloudy	13C
16	25	Male	Partly Cloudy	14C



FIGURE 2: Map of predetermined route for drivers, located in London, Ontario, Canada. The path length is approximately 28.5 kms and includes urban and suburban driving areas. There are 10 right turns and 7 left turns in the path.

for about 60 minutes over this route, which is illustrated in Figure 2. All drives were performed between 9am and 4pm EST. In total, 3TB of vehicular data was collected over more than 450 kilometers of driving. Table 1 provides general information about each test driver, including weather conditions.

In this study, we exclusively used the data from CANbus (Controller Area Network) to extract measurements about the vehicle such as speed, gas pedal pressure, brake pedal pressure, and steering wheel angle. These signals were originally sampled at 30 Hz in this experiment. We approximate acceleration based on the speed value, such that, for each

one-second interval (30 frames), we calculate $v_1 - v_0$ as the acceleration. (Computing acceleration at every frame (30Hz) provides very small values subject to noise. Measuring acceleration once per second results in numerically stable measurements.)

Figure 3 depicts examples of four driving signals collected during normal urban driving for a period of 300 seconds. The signals correspond to the speed of the vehicle, gas pedal pressure, brake pedal pressure, and steering wheel angle, respectively. The steering wheel angle value is almost zero in most of the frames although the experiment takes place in urban area. As we might expect, the gas pedal pressure

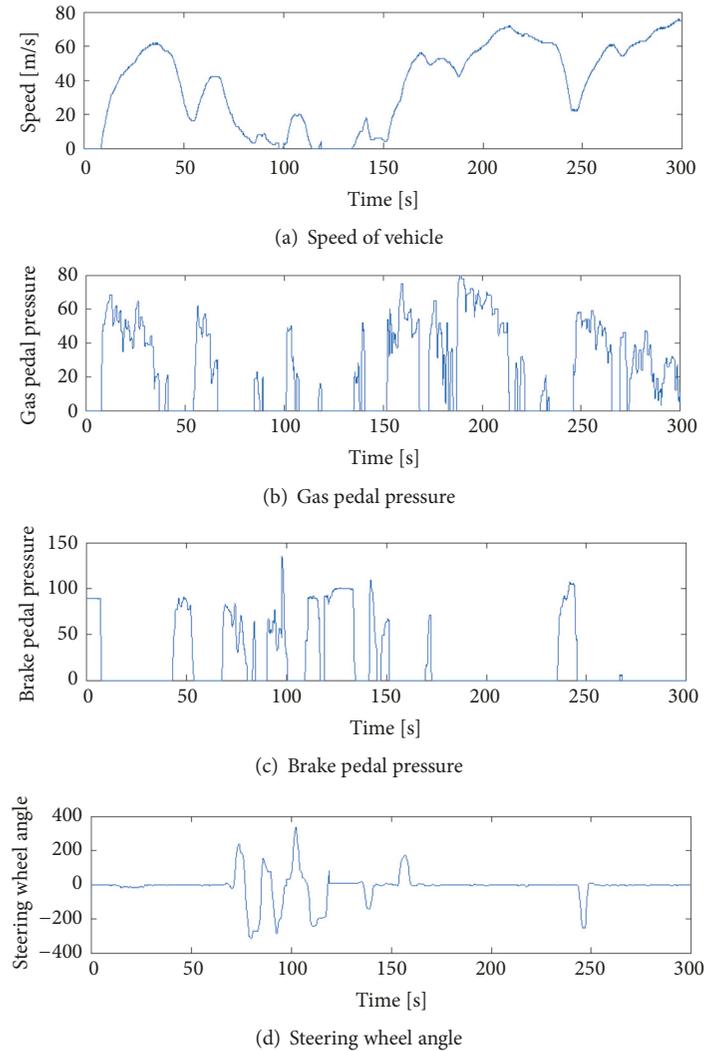


FIGURE 3: Examples of driving behavior signals collected in 300 seconds of normal driving. (a) Speed of vehicle; (b) gas pedal pressure; (c) brake pedal pressure; (d) steering wheel angle.

and brake pedal pressure are complementary to each other (though, of course, they do not have to be as some drivers use both feet). In addition, the direct correlation between gas pedal pressure and vehicle speed is obvious.

Since the main focus of this study is to categorize drivers based on their preturn driving behavior, we needed to extract each of these signals before each turn. Among the 16 driving experiments, we selected 12 subjects (1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 15, and 16) to include in this study. The other four subjects (drivers 5, 6, 13, and 14) were not used due to existence of noise, missing data, or because of changes in the preset driving path. In total, we have data for 10 right turns and 7 left turns for 12 drivers consisting of 6 male and 6 female drivers.

As we are analyzing the driver behavior while driving, we can ignore the time when the vehicle's speed is actually zero. The stopped state usually happens right before turning maneuvers, especially before left turns. Heavy traffic, stop signs, traffic light, yield rights, and so forth are some common reasons for stopping before turning maneuvers. Among

the not-stopped frames before each turn, we collect three different sets with 150 frames, 300 frames, and 450 frames (5 seconds, 10 seconds, and 15 seconds before each turn, respectively).

4. Methodology

For this study, our choice of using CANbus data streams only stems from a need to categorize drivers without using additional modes of sensing, such as 3D stereo vision and range finders, as an attempt to perform driver analysis and classification economically. This choice has implications, among which we find the impossibility to consider traffic conditions at intersections where drivers perform turns. We believe traffic conditions have minimal impact on driving style, partly because drivers were performing under good weather conditions (daytime with no rain or snow) with light to moderate traffic. Other studies of driver behavior used similar strategies to minimize the impact of weather

and traffic conditions on their results [25]. Additionally, we opted to not distinguish between left and right turns as types of maneuvers have minimal effect on observed driving style [26].

Conversely, our study only involves 12 drivers. While each driver provided approximately one hour of driving totaling 12 hours of data, it is clear that the number of participants was insufficient to draw general conclusions. Rather, this study focuses on the potential to cluster drivers based on their behavior (as defined by how they operate the vehicle) using nothing other than CANbus data streams. These signals include speed, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration, collected 5, 10, and 15 seconds before each turning maneuver for different drivers while performing normal driving. Different statistical features are extracted from the signals in order to lower the dimensionality of the data. For each time series sequence, statistical features were calculated to measure different properties of that sequence: mean, standard deviation, kurtosis, and skewness. These statistics were calculated for each of the 5 different signals (speed, gas pedal pressure, brake pedal pressure, steering wheel angle, and acceleration) before each turn. This results in 20 statistical features characterizing the signals for each turn and so represents a single driver's behavior before an individual turn. We compute the average of each of these features over all turns for each driver resulting in a feature vector of 20 values for each driver. We normalize the features by scaling to bring all values into the [0, 1] range. We considered the maximum and minimum values across all drivers and all turns:

$$\bar{X} = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})} \quad (1)$$

where X_{\max} is maximum value and X_{\min} is the minimum value in the set.

While there is a high probability that the chosen statistical features are correlated, we are not aiming at the identification of latent variables from linear combinations of these features (as in factor analysis), since we do not know the nature of the relationships that may exist among these features (linear or nonlinear). We believe that the statistical features extracted from the CANbus data streams are sufficient for an adequate clustering of drivers, since we are not aiming at identifying the causes (factors) of certain types of driving behavior. We justify this assertion by observing the impact of aggressive versus normal driving would have on the features we use for analysis and clustering. Aggressive driving results in higher values for average, standard deviation, and kurtosis for brake pedal and accelerator pressure. In addition, an aggressive driver may accelerate rapidly yet decelerate normally, creating skewness in the statistical distribution of the speed data. We also chose to cluster the data with HCA rather than k -means, so as to avoid providing a value for k (the number of clusters) which is not known *a priori* for our or any existing dataset.

Note that measurements of speed, gas pressure, and brake pressure are all positive (or zero) whereas values for the steering wheel angle and for acceleration can be positive or negative. A negative steering wheel angle indicates a turn

of the wheel to the left and a negative for the acceleration indicates deceleration. Since all feature values are scaled between 0 and 1, this means that *zero* values for the steering wheel position and acceleration are also scaled. Knowing these is useful in interpreting the normalized feature values.

We group these normalized statistical features averaged over all turns in a 5×4 matrix for each driver. We then use HCA on the feature vectors for the 12 drivers. The choice of metrics is key in clustering. For hierarchical clustering, we need two metrics: one to measure the separation between points and another to determine whether to split or merge two clusters. To measure the distance between two points, we settled on Euclidean distance, as it is appropriate for cluster analysis with statistical features. To merge or split two clusters we need to measure the distance between two sets of observations. Here $D(P, Q)$ is the merging cost of combining clusters P and Q . Our goal is to find clusters with minimum variation and Ward's linkage method minimizes the total within-cluster variance [27]. For Ward's method,

$$D(P, Q) = \frac{n_P \times n_Q}{n_P + n_Q} \|\vec{m}_P - \vec{m}_Q\| \quad (2)$$

where \vec{m}_P is the center of cluster P and n_P is the number of instances in it.

We then use hierarchical clustering on the result of all drivers' feature vectors using Ward's linkage criteria and Euclidean distance as the distance metric. The result is a dendrogram clustering drivers based on statistical features extracted from preturns driving behavior. We define a turn as the time at which the turning maneuver begins, as dictated by a steering wheel angle change. We repeated this process for each of the 150, 300, and 450 frame preturn sequences, corresponding to durations of 5, 10, and 15 seconds. We chose these partially overlapping preturn sequences in order to determine the influence these time frames have on the results of the HCA clustering process. The preturn sequences are partially overlapping as to always include the most significant parts of the preturn vehicular dynamics.

5. Results and Discussion

We perform HCA before all turns for 12 drivers using their statistical features. Figure 4 shows the dendrograms for all drivers for 150, 300, and 450 frames, respectively. The vertical axis in these figures represents similarity. For all of our dendrograms in analyzing all preturns using statistical features, we are able to identify two primary clusters of drivers. Considering the dendrograms for 150 frames, one sees that the drivers form two primary clusters: one with drivers 8, 9, 10, 11, and 16 and one with drivers 1, 3, 4, 7, 12, and 15; driver 2 remains in a singleton cluster and so is excluded in this case. Similar patterns are observed in dendrograms for 300 and 450 frames. For 300 frames, the drivers again seem to group into 2 primary clusters: 1, 7, 8, 10, 15, and 16 and the other with drivers 3, 4, 9, 11, and 12. Again, driver 2 remains in a singleton cluster. For 450 frames we obtain drivers 2, 7, 10, 12, and 15 in one cluster and drivers 1, 3, 4, 8, 9, 11, and 16 in the second cluster; in this case all drivers

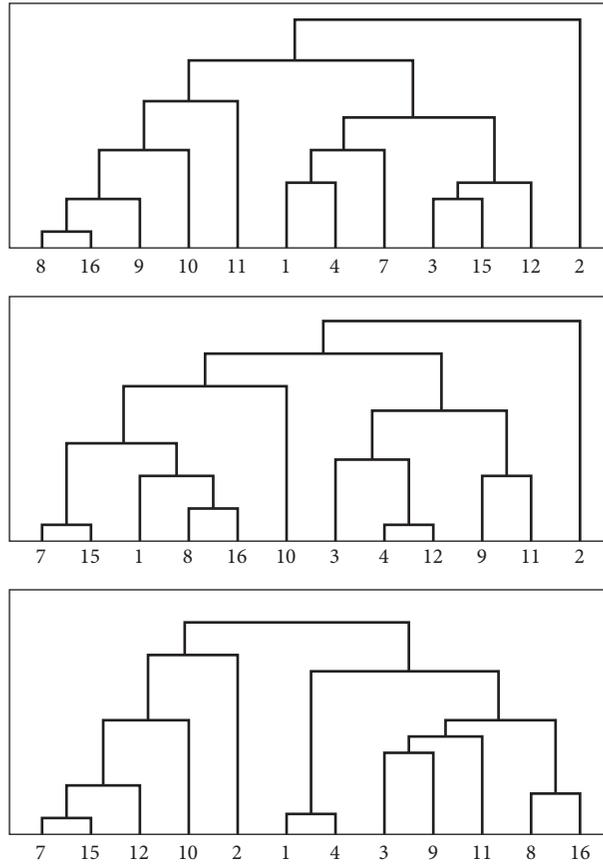


FIGURE 4: Dendrograms from HCA, from top to bottom, for 150, 300, and 450 frames with driver identification numbers as per Table 1.

TABLE 2: Mean values across frames and clusters.

<i>Mean Values Cluster 1</i>			
	<i>150 Frames</i>	<i>300 Frames</i>	<i>450 Frames</i>
<i>Speed</i>	0.4050	0.5445	0.6327
<i>Gas</i>	0.6087	0.7271	0.8942
<i>Brake</i>	0.3294	0.5316	0.5349
<i>SW Angle</i>	0.8471	0.5588	0.5188
<i>Acceleration</i>	0.6988	0.7914	0.6326
<i>Mean Values Cluster 2</i>			
	<i>150 Frames</i>	<i>300 Frames</i>	<i>450 Frames</i>
<i>Speed</i>	0.8333	0.6940	0.5368
<i>Gas</i>	0.2997	0.3181	0.3072
<i>Brake</i>	0.8714	0.8608	0.8215
<i>SW Angle</i>	0.6059	0.6107	0.5486
<i>Acceleration</i>	0.2686	0.3103	0.2989

have been included in one cluster or another. Each driver was characterized by 20 features: 4 statistical measures (mean, standard deviation, kurtosis, and skewness) of each of the normalized vehicle performance metrics, namely, speed, gas pedal pressure, brake pedal pressure, steering wheel position, and acceleration.

Table 2 presents the average values of each CANbus signal for the two clusters that could be identified, while Table 3

summarizes the drivers assigned to each of the two clusters for each different set of frames. It also shows the number of males and females in each cluster.

An examination of the centroid values for *Cluster 1* and *Cluster 2* shows that they significantly differ. To compare the clusters more precisely, we convert each clusters' centroids matrix to a vector of size 20 and then calculate the angle and the correlation (R-value) between these two vectors in

TABLE 3: Clusters of drivers from HCA including all drivers and all frames.

	Cluster 1 (M/F)	Cluster 2 (M/F)	Angle	R-Value
150 Frames	8,9,10,11,16 (3/2)	1,3,4,7,12,15 (2/4)	35	0.32
300 Frames	1,7,8,10,15,16 (4/2)	3,4,9,11,12 (1/4)	33	0.35
450 Frames	2,7,10,12,15 (2/3)	1,3,4,8,9,11,16 (4/3)	35	0.06

TABLE 4: Pairwise cluster comparison.

	HCA: 300 Frames	HCA: 450 Frames
HCA: 150 Frames	0.545	0.322
HCA: 300 Frames		0.606

20-dimensional space. Ideally, we would like the clusters to show little correlation. Table 3 summarizes the angle and R-values between the two clusters in each experiment using 150, 300, and 450 frames.

The R-values and angles show that the clusters are not particularly correlated. Further, there is some consistency in membership of drivers within each cluster across all three timeframes. Even though there is a small number of drivers, the clusters appear distinctive. We also compare the clusters formed across the different timeframes in order to determine whether the clusters formed are similar. For this comparison, we make use of the Sorensen-Dice index, customarily used for comparing the similarity of two samples. The index ranges between zero and one, where 0 indicates no similarity and 1 indicates identical samples. Sorensen's original formula was intended to be applied to the presence or absence of data in the samples and is specified as

$$QS(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

where $|X|$ and $|Y|$ are the numbers of elements in the two samples. We use this index to compare membership in corresponding classes, that is, to compare the membership in *Cluster 1* for 150 frames with membership in *Cluster 1* for 300 frames. We compute a combined score for comparing the clusters for each of the analyses for different frames by averaging the score comparing *Cluster 1* and the score comparing *Cluster 2* of two different analyses. To compare the clusters from 150 frames to 300 frames we compute

$$\frac{QS(\text{Cluster1}_{150}, \text{Cluster1}_{300}) + QS(\text{Cluster2}_{150}, \text{Cluster2}_{300})}{2} \quad (4)$$

Table 4 summarizes the pairwise comparisons. There is good similarity between the clusters formed in the analysis from 300 and 450 frames. It may just be that 150 frames is insufficient data, as 150 frames corresponds to only 5 seconds.

Assuming that we take the two clusters arising from the cluster analyses as indicative of how to group individual drivers, we are interested in determining what this clustering means in terms of driving characteristics exhibited by the drivers. To get some insight, we examined the values of the raw data, such as the averages of the features extracted before being normalized. In comparing the average values of each of

the signals for the drivers in each cluster at different numbers of frames, we see that there are differences between drivers in *Cluster 1* and *Cluster 2*. Although there are only 12 drivers, we compared the groups of drivers (for the same number of frames) in each cluster with a *T*-test. The results show the following:

- (i) For 150 frames, the drivers in *Cluster 1* and *Cluster 2* show significant difference at $p < 0.05$ for both *Speed* and *Gas*; no other measures showed any significance.
- (ii) For 300 frames, the drivers show significant difference in *Speed* and *Brake* at $p < 0.25$, *Gas* at a $p < 0.2$, and *Acceleration* at $p < 0.01$.
- (iii) For 450 frames, the drivers show significant difference in *Gas* and *Acceleration* at $p < 0.01$ and *Brake* at $p < 0.05$.

While the drivers in each cluster for the different numbers of frames are not exactly the same, there is some consistency in which the signals differ. Considering the drivers in the two clusters at 450 frames, we observe that the drivers in *Cluster 2* have higher brake pressure and decelerate (negative acceleration) more than drivers in *Cluster 1*. Those in *Cluster 1* apply more pressure on the gas pedal. Even though the speed of the vehicles in both groups is similar over the 15-second period, drivers in *Cluster 2* are braking harder and decelerating more quickly. We assume that drivers in *Cluster 1* have used their brake earlier and have already slowed down to some extent and so need to maintain some gas pressure. This means that drivers in *Cluster 2* slow down more quickly than those in *Cluster 1*, and as a result they apply the brakes harder and have lower pressure on these gas pedals. This behavior leads us to classify drivers in *Cluster 2* as more aggressive when compared to the drivers of *Cluster 1*, which we classify as moderate drivers. To summarize the behavior of drivers in the two clusters in approaching turns, we have the following:

- (i) *Cluster 1 (moderate drivers)*: moderate speed, some gas pressure approaching turn, gentle braking, and gradual deceleration
- (ii) *Cluster 2 (aggressive drivers)*: higher speed approaching turns, harder braking, and rapid deceleration

Based on our empirical data, we have shown that clustering of drivers using HCA on driver descriptors composed of statistical elements extracted from the OBD-II CANbus data stream is feasible and yields clusters that show significant differences among themselves. Our study focuses only on one type of maneuver, namely, left and right turns. A comprehensive driver classification framework would require that a majority of possible maneuvers be considered in the construction of driver descriptors prior to HCA. This would

provide comprehensive driver descriptors and consequently more meaningful clusterings, with interpretations richer than a simple moderate versus aggressive classification. Reliable driver classification has many uses toward road safety, including driver training, recognition, and driving style analysis. This type of classification forms the basis that is needed for driver behavior correction and eventual enhanced road safety.

6. Conclusion and Directions

We analyzed statistical features of CANbus signals to see if there were distinct clusters of driver behavior before turns. The approach made use of four statistical features, the mean, standard deviation, kurtosis, and skewness, that were extracted from the signals. Using these statistical features as a representative of each driver's behavior, HCA was used to cluster the drivers. The results show that there exist at least two distinct groups of drivers with different return behavior: one cluster includes moderate drivers, while the other cluster contains more aggressive ones. We also examined the distinctiveness of the clusters as well as the consistency of member within the clusters across different timeframes. The analysis shows that the consistency between clusters' members was satisfactory throughout different timeframes.

In this contribution, we introduced the notion of *driver descriptors* that are not unlike those successfully used in image analysis (an example of such descriptors is known as SIFT (Scale-Invariant Feature Transform) and was introduced by D. Lowe in the context of feature detection for computer vision algorithms [28]). Just as image SIFT features can be classified and matched, so can driver descriptors, due to their analogous composition. While this study shows that a simple classification can be obtained from a 20-dimensional driver descriptor and one type of driver-initiated maneuver (right and left turns), one of our immediate goals is to enhance driver classification by defining better descriptors and including the sum of possible driver maneuvers. Most likely the sum of techniques developed around SIFT descriptors can be adapted to driver descriptors to achieve general driver identification, classification, and recognition. Additionally, the analysis of driver descriptors will lead to the identification of inadequate driving and the context in which it occurs.

This research was performed within the context of creating more effective Advanced Driving Assistance Systems (ADAS), where driver behavior could be analyzed in real-time and used by an ADAS to provide feedback to drivers.

Data Availability

The data used in this research is provided by the RoadLAB research programme conducted at the University of Western Ontario, London, ON, Canada.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Partial funding for this research was provided by the Natural Sciences and Engineering Council of Canada (NSERC).

References

- [1] National Highway Traffic Safety Administration, "Traffic safety facts research note," Technical Report DOT HS 812 318, U.S. Department of Transportation, 2016.
- [2] M. Zardosht, *Identifying individual driver behaviour using in-vehicle can-bus signals of pre-turning maneuvers [Master's thesis]*, The University of Western, Ontario, Canada, 2016.
- [3] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic dynamics: studies in car following," *Operations Research*, vol. 6, pp. 165–184, 1958.
- [4] C. Miyajima, Y. Nishiwaki, K. Ozawa et al., "Driver modeling based on driving behavior and its evaluation in driver identification," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 427–437, 2007.
- [5] B. Higgs and M. Abbas, "Segmentation and clustering of car-following behavior: Recognition of driving patterns," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 81–90, 2015.
- [6] S-H. Chen, J-S. Pan, and K. Lu, "Driving behaviour analysis based on vehicle obd information and adaboost algorithms," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, pp. 102–106, 2015.
- [7] J. Carmona, F. García, D. Martín, A. de la Escalera, and J. M. Armingol, "Data fusion for driver behaviour analysis," *Sensors*, vol. 15, no. 10, pp. 25968–25991, 2015.
- [8] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proceedings of the 14th IEEE International Intelligent Transportation Systems Conference (ITSC '11)*, pp. 1609–1615, Washington, DC, USA, October 2011.
- [9] S. Choi, J. Kim, D. Kwak, P. Angkititrakul, and J. Hansen, "Analysis and classification of driver behaviour using in-vehicle can-bus information," in *Proceedings of the Biennial Workshop on DSP for In-Vehicle and Mobile Systems*, pp. 17–19, 2007.
- [10] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '12)*, pp. 234–239, Madrid, Spain, June 2012.
- [11] J. A. Michon, "A critical view of driver behavior models: What do we know, what should we do?" in *Human Behavior and Traffic Safety*, L. Evans and R. C. Schwing, Eds., Springer, 1985.
- [12] T. Toledo, "Driving behaviour: models and challenges," *Transport Reviews*, vol. 27, no. 1, pp. 65–84, 2007.
- [13] R. Fuller, "Towards a general theory of driver behaviour," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 461–472, 2005.
- [14] G. A. M. Meiring and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, pp. 30653–30682, 2015.
- [15] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [16] B. N. Fildes, G. Rumbold, and A. Leening, "Speed behaviour and drivers attitude to speeding," Tech. Rep. 16, Monash University, 1991.
- [17] H. Ohta, "Individual differences in driving distance headway," *Vision in Vehicles*, vol. 4, no. 4, pp. 91–100, 1993.

- [18] R. Kalsoom and Z. Halim, "Clustering the driving features based on data streams," in *Proceedings of the 2013 16th International Multi Topic Conference, INMIC 2013*, pp. 89–94, Pakistan, December 2013.
- [19] C. Wu, C. Sun, D. Chu, Z. Huang, J. Ma, and H. Li, "Clustering of several typical behavioral characteristics of commercial vehicle drivers based on GPS data mining: Case study of highways in China," *Transportation Research Record*, vol. 2581, pp. 154–163, 2016.
- [20] B. Higgs and M. Abbas, "A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles," in *Proceedings of the 2013 16th International IEEE Conference on Intelligent Transportation Systems: Intelligent Transportation Systems for All Modes, ITSC 2013*, pp. 857–862, Netherlands, October 2013.
- [21] M. S. Shirazi and B. T. Morris, "Looking at Intersections: A Survey of Intersection Monitoring, Behavior and Safety Analysis of Recent Studies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 4–24, 2017.
- [22] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 724–736, 2012.
- [23] H. Rakha, I. El-Shawarby, and J. R. Setti, "Characterizing driver behavior on signalized intersection approaches at the onset of a yellow-phase trigger," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 4, pp. 630–640, 2007.
- [24] S. S. Beauchemin, M. A. Bauer, T. Kowsari, and J. Cho, "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 2, pp. 391–401, 2012.
- [25] L. Eboli, G. Mazzulla, and G. Pungillo, "How drivers' characteristics can affect driving style," *Transportation Research Procedia*, vol. 27, pp. 945–952, 2017.
- [26] T. Lajunen, J. Karola, and H. Summala, "Speed and acceleration as measures of driving style in young male drivers," *Perceptual and Motor Skills*, vol. 85, no. 1, pp. 3–16, 1997.
- [27] J. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, pp. 236–244, 1963.
- [28] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, September 1999.



Hindawi

Submit your manuscripts at
www.hindawi.com

