Transportation Research Record

Understanding Speeding Behavior from Naturalistic Driving Data: Applying Classification Based Association Mining

--Manuscript Draft--

Full Title:	Understanding Speeding Behavior from Naturalistic Driving Data: Applying Classification Based Association Mining
Abstract:	Speeding is considered as one of the most significant contributing factors to severe traffic crashes. Understanding the associations between trip/driving/roadways features and speeding behavior is crucial for both researchers and practitioners. This research utilized GPS-based naturalistic driving data collected by the Safety Pilot Model Deployment (SPMD) program and roadway features from a road inventory dataset - Highway Performance Monitoring System (HPMS), provided by the Department of Transportation of the United States, to investigate the hidden rules that associated trip/driving/roadway features with speeding behavior. A classification-based association (CBA) algorithm was adopted to explore the hidden rules from two perspectives of speeding: speeding duration and speeding pattern. Results indicate that longer trips, driving during peak hours, tailgating other vehicles, being delayed in traffic before a speeding event, driving on roadways with a median or driving on higher functional class roadway are highly associated with relatively longer speeding events. In terms of speeding pattern, driving during peak hours, and tailgating other vehicles are two items showing repeatedly in the high speeding pattern cases. Moreover, higher functional class of roadway, presence or absence of median, and speed loss before a speeding event show high association with speeding events. The findings can help practitioners understand the composite effect of these factors more comprehensively and provide corresponding countermeasures to mitigate negative consequences wherever possible. These can also help in calibrating driver behavior parameters for transportation related simulation tools.
Manuscript Classifications:	Data and Information Technology; Artificial Intelligence and Advanced Computing Applications ABJ70; Machine Learning (Artificial Intelligence); Safety and Human Factors; Design, Operations, and Traffic Management; Vehicle User Characteristics AND10; Driver Behavior
Manuscript Number:	20-03307
Article Type:	Presentation and Publication
Order of Authors:	Xiaoqiang Kong
	Subasish Das
	Kartikeya Jha
	yunlong Zhang

Understanding Speeding Behavior from Naturalistic Driving Data: Applying Classification Based Association Mining

Xiaoqiang "Jack" Kong
(Corresponding Author)
Graduate Research Assistant, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 2217
Bryan, TX 77807
Email: x-kong@tti.tamu.edu

Subasish Das, Ph.D.
Associate Transportation Researcher, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 4414
Bryan, TX 77807
Email: s-das@tti.tamu.edu

Kartikeya Jha
Research Associate, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 2238
Bryan, TX 77807
Email: jha_k@tamu.edu

Yunlong Zhang, Ph.D.
Professor, Texas A&M University
3136 TAMU, College Station, TX 77843-3136
Email: yzhang@civil.tamu.edu

Word Count: 6,477 words + 4 tables (250 words per table) = 7,477 words

Submitted [8/1/2019]

ABSTRACT
Speeding is considered as one of the most significant contributing factors to severe traffic crashes. Understanding the associations between trip/driving/roadways features and speeding behavior is crucial for both researchers and practitioners. This research utilized GPS-based naturalistic driving data collected by the Safety Pilot Model Deployment (SPMD) program and roadway features from a road inventory dataset - Highway Performance Monitoring System (HPMS), provided by the Department of Transportation of the United States, to investigate the hidden rules that associated trip/driving/roadway features with speeding behavior. A classification-based association (CBA) algorithm was adopted to explore the hidden rules from two perspectives of speeding: speeding duration and speeding pattern. Results indicate that longer trips, driving during peak hours, tailgating other vehicles, being delayed in traffic before a speeding event, driving on roadways with a median or driving on higher functional class roadway are highly associated with relatively longer speeding events. In terms of speeding pattern, driving during peak hours, and tailgating other vehicles are two items showing repeatedly in the high speeding pattern cases. Moreover, higher functional class of roadway, presence or absence of median, and speed loss before a speeding event show high association with speeding events. The findings can help practitioners understand the composite effect of these factors more comprehensively and provide corresponding countermeasures to mitigate negative consequences wherever possible. These can also help in calibrating driver behavior parameters for transportation related simulation tools.

Keywords: Trip features, driving characteristics, geometric features, speeding duration, speeding pattern

INTRODUCTION

The year-on-year increasing figure on road traffic fatalities is fast becoming a serious global safety concern. The National Highway Traffic Safety Administration reported that 52,274 drivers were involved in 34,247 fatal crashes resulting in 37,133 fatalities in the year 2017 *(1)*. "Seventeen percent of the drivers involved were speeding at the time of the crashes, and twenty-six percent of those killed were in a crash involving at least one speeding driver." These statistics highlight a pressing need for a deeper examination of potential contributing factors that lead to such sequence of events and underline the importance of understanding driver's speeding behavior. This research utilized GPS based trajectory data and road inventory data to explore the associations between trip/driving/roadway features and speeding behavior from two perspectives: speeding duration (how long the speeding event lasts) and speeding pattern (how much does this speeding event exceed the speed limit). It examined features that might encourage drivers to speed for relatively longer or higher by employing a classification and association rule algorithm.

There are different ways of data collection and different methods are applied for analysis in understanding the relationship between roadway environment and speed choice. While some studies collected data from field measurements of spot speeds and then analyzed it with data for road characteristics, others used simulation to collect and capture a wider range of responses from drivers in a constructed environment. The trajectory dataset provided by Safety Pilot Model Deployment (SPMD) program, which is a part of the Connected Vehicle Safety Pilot Program, contains detailed information about trips and driving characteristics. Data for each trip consists of points with spatial information at a centisecond level. The geospatial information (longitude and latitude) of each point is used to conflate its corresponding roadway features. With information on speed limit from road inventory data - HPMS, speeding behavior can be identified based on the duration of speeding and speeding pattern.

The authors anticipated that the duration of speeding and the speeding pattern of a speeding event are associated with the speed loss before that event. Intuitively, drivers are likely to compensate for the travel time loss during congestion by speeding afterward. This research investigated the associations rules of speeding behavior from three perspectives: a) trip features like total travel time; b) driving features like tailgating other vehicles, speed loss before a speeding event and driving during the peak period; c) roadway features like functional class, access control, shoulder width, lane width, median width, pavement type, speed limit, median type, and Annual Average Daily Traffic (AADT). The speeding behavior was investigated from two perspectives: speeding duration and speeding pattern. Speeding duration has been classified as long (speeding lasting more than 4 minutes) and short (speeding lasting 2-4 minutes). Additionally, the speeding pattern is classified as a high speeding pattern (speeding more than 5mph over speed limit) and moderate speeding pattern (speeding less than 5 mph).

This research applied the classification-based association (CBA) algorithm. To authors' knowledge, this is the first of its kind application of the CBA algorithm to this area of research on speeding behavior. An advantage of this algorithm is that it can automatically handle both numerical and categorical variables to mine meaningful association rules. Another advantage is that this non-parametric method avoids making any parametric assumptions, which are normally subjective and sensitive to correctness of underlying hypothesis.

This study explored the possible associations between speeding behavior and trip/driving/roadway features. The contributions of this research are 1) a comprehensive exploration of the features that might be associated with speeding behavior; 2) by using a classification-based association rule algorithm, the rules mined from the dataset are comparable. The comparison could provide in-depth insights into the associations between speeding and trip/driving/roadway features; 3) this research is the first in differentiating the speeding behavior from two perspectives: speeding duration and speeding pattern.

EARLIER WORK AND RESEARCH CONTEXT

In recent years, several studies have investigated the effects of roadway features on the speeding behavior. Kanellaidis et al. *(2)* examined the effects of three demographic characteristics – age, familiarity to the environment, and self-assessment – on driver's risk perception of the roadway environment. The results indicated that the number of curves and the element of curvature of the section significantly influenced driving behavior. Jørgensen et al. *(3)* found that both visual measures (e.g., speed limit signs) and physical measures (speed bumps, roundabouts, etc.) had a perceptible effect on speed reduction; however, the effect of physical measures was higher. Moreover, the interval spacing between these speed-calming measures also had a significant impact on speeding behavior. The density of built-up area around the roadway also showed an effect on driver behavior. A related study conducted in Denmark found that drivers interpreted physical road characteristics such as lane and roadway width, alignment and longitudinal profile, etc., more than roadside characteristics *(4)*.

Andersen et al. *(5)* utilized GPS-based big data collected on homogenous road sections using floating car method and modelled with multivariate linear regression to investigate speed and driving characteristics with roadway characteristics on rural two-lane roads in Denmark. The study found significant effects of road and shoulder width, section length and extent of road markings on driver behavior. Road markings along both the centerline and the shoulders gave a sense of added safety to drivers and encouraged higher speed choice. Presence of woodland (vegetation or shrubbery) and intersections also influenced speed choice along with other contributing factors such as gender, time of day and age of vehicle. Similarly, Lobo et al. *(6)* observed that the presence of upstream and downstream features which cause constrained visibility, bend and density of intersections, etc., resulted in reduced speeds.

Several studies have also highlighted the tendency in drivers to disregard posted speed limits to varying extents depending on the road environment. While Fildes et al. *(7)* found in a study conducted in Australia that a major portion of drivers did not consider driving 30 kmph above the posted limit to be risky in both rural and urban settings, Eksler et al. *(8)* found in Scandinavia that more than half of the driving population violate the speed limit by more than 10 kmph on rural roads. Varhelyi *(9)* concluded that in general, drivers tend to underestimate the risks of speeding, and base their speed choice primarily on their own assessment of the road and traffic environment.

Bassani et al. *(10)* analyzed urban arterial and collector road sections instead of rural roads, and found that the presence of both right and left shoulders resulted in higher speed dispersion, as did the presence of dedicated bus and taxi lane adjacent to the travelled way. On the other hand, the presence of sidewalk and pedestrian crossing significantly reduced speed dispersion. The study concluded that speed reduction is influenced more by transversal geometric characteristics than longitudinal characteristics. Lane position and number of travelled ways (lanes) were found to have the most significant effect on mean speed.

The literature review reveals that there is a need for an in-depth study focusing on the determination of speeding related contributing factors and associated patterns using rich data on observed naturalistic driver behavior. This study applied association rules to determine the patterns of associated factors from a complex dataset to understand speeding behavior and relevant risk factors.

METHODOLOGY

Recent advances in data collection, filtering and storage techniques have produced an abundance of quality data for transportation and traffic safety research. However, the ever-increasing size and complexity of data have gradually induced challenges in employing conventional research methods. This has given rise to the need for utilizing emerging and innovative techniques that can help understand underlying patterns in available data. Datasets consisting of large number of variables and features lend themselves particularly well to analysis techniques involving machine learning, data mining and statistical learning approaches. These approaches help identify patterns in input data without the need to confirm predetermined hypotheses relating examined features. Conventional parametric models work well if the model specification for underlying assumptions is efficient in capturing relationships between dependent and independent variables (*11*). Conversely, violation of any assumption could lead to flawed or inadequate estimations. Specific nonparametric data mining techniques have received increased attention from researchers in traffic safety because of no predefined assumptions (*12*). In several ways, these contemporary approaches provide opportunities to identify complex relationships in data as well as improved algorithmic processing capabilities.

Classification Based Association (CBA) Mining

Data mining involves machine learning, statistical knowledge, modeling concepts and database management. Association rules mining, a descriptive analytics technique, discovers significant rules showing variable category conditions that occur frequently together in a dataset. It involves the use of machine learning models to analyze data for patterns, or co-occurrence, in a database. It identifies frequent if-then (antecedent-consequent) associations, which are called association rules. As a non-parametric method, it avoids making any parametric assumptions as most parametric methods do. Association rules are created by searching data for frequent if-then patterns and using the criteria support and confidence to identify the most important relationships. Support indicates how frequently the items appear in the data. Confidence is an indicator of the number of times the if-then statements are found true. In other words, Support represents how much historical data supports the identified rule and Confidence represents how confident we are that the rule holds. A third metric, called lift, which is the ratio of Confidence to Support, can be used to compare observed confidence with expected confidence *(13)*.

Association rules are derived from *itemsets* which comprise of two or more items. The Apriori algorithm considers any subset of a frequent itemset to also be a frequent itemset. No superset of any infrequent itemset is generated or tested which allows pruning of many item combinations. *Apriori* is a level-wise, breadth-first algorithm which counts transactions. This algorithm can be used to mine frequent itemsets, maximal frequent itemsets and closed frequent itemsets. This algorithm helps researchers mine out frequently occurring itemsets, subsequences, arrangements and interesting associations between various items.

A set of definitions are provided here before demonstrating the method with an example. Let $I= i_1, i_2, \ldots i_m$ be a set of items (e.g., a set of crash categories for a particular crash record) and $C= c_1, c_2, \ldots, c_n$ be a set of database crash information (transaction) where each crash record c_i contains a subset of items chosen from I. An itemset with k items is called as a k-itemset.

An association rule can be demonstrated as A → B, where A and B are disjoint itemsets. Here, A is known as the antecedent and B is the consequent. The strength of the association rule can be measured using the values of measures like support, confidence and lift. For the purpose of our analysis, support is defined as the percentage of casualties in the dataset that contains the itemset. Confidence is the ratio of the number of all crashes in C to the number of crashes that include all items in I. Lift is the ratio of confidence over expected confidence. The equations of support are listed in equation 1 through 3.

$$S(A) = \frac{\sigma(A)}{N} \tag{1}$$

$$S(B) = \frac{\sigma(B)}{N} \tag{2}$$

$$S(A \rightarrow B) = \frac{\sigma(A \cap B)}{N} \tag{3}$$

where,

$\sigma(A)$= Number of incidents with antecedent A

$\sigma(B)$= Number of incidents with consequent B

$\sigma(A \cap B)$= Number of incidents with both A antecedent and B consequent

N= Total number of incidents

S(A)= Support of antecedent

S(B)= Support of consequent

S(A → B)= Support of the association rule (A → B)

The definitions for confidence and lift are provided in equations 4 and 5 respectively. Confidence measures the reliability of the inference of a generated rule. Higher confidence for A → B indicates that presence of B is highly likely in transactions (or observations) containing A. The lift of the rule A → B associates the frequency of co-occurrence of the antecedent and the consequent to the expected frequency of co-occurrence.

$$C(A \rightarrow B) = \frac{S(A \rightarrow B)}{S(A)} \tag{4}$$

$$L(A \rightarrow B) = \frac{S(A \rightarrow B)}{S(A).S(B)} \tag{5}$$

where,
C(A → B)= Confidence of the association rule (A → B)
L(A→ B)= Lift of the association rule (A → B)

A lift value > 1 indicates significant interdependence between the antecedent and the consequent, while a value < 1 indicates low interdependence, and a value of 1 designates independence. A rule with a single antecedent and a single consequent is defined as a 2-product rule; similarly, a rule with two antecedents and single consequent or one antecedent and two consequents is defined as a 3-product rule. A critical inference of the association rules is that the generated rules need not be interpreted as causation, rather as association.

While association rule mining tries to find all rules in the database that satisfy the desired minimum support and minimum confidence constraints, Classification Based Association (CBA) rule mining aims to discover a small set of rules in the database to form an accurate classifier based on class association rules (*14*). Therefore, there is a pre-set target (called class) for the classification rule mining process. Instead of using an itemset as in association rules mining, a *ruleitem* is used, which consists of a *condset* (a set of items) and a class. A rule pruning technique is adopted to prune off non-predictive and overfitting rules. This is an extension to the approach used in association rule mining.

DATA DESCRIPTION

Data Sources

This study uses two data sources for this analysis: 1) SPMD data, and 2) HPMS data of Michigan for the year 2016. The following subsections provide a brief description of these two datasets and the final data used in this study.

Safety Pilot Model Deployment (SPMD) Data

SPMD is a part of the Connected Vehicle Safety Pilot Program *(15)*. SPMD study researchers collected several types of driving data, including data on vehicle trajectory, data about surrounding vehicles, and network data. This study focused on the vehicle trajectory data and the surrounding vehicle data collected by data acquisition system 1 (DAS1). University of Michigan Transportation Research Institute (UMTRI) was responsible for DAS1. It uses Mobileye sensor for data collection *(16)*. Trips with less than ten minutes trip duration were excluded from the dataset. Data for 92 vehicles was available after filtering for trip duration. This data was collected between October 2012 and April 2013.

Highway Performance Monitoring System (HPMS)

HPMS contains roadway inventory data for all the states in U.S. In this study, the authors used the roadway inventory data for Michigan, 2016. The data included roadway and traffic features such as number of lanes, roadway functional classification, and average annual daily traffic (AADT), etc. Table 1 lists the roadway features included in this study. Roadway features such as type of shoulder, AADT and lane width had missing values for some rows, were thus not included in this study.

Data Integration

The researchers reduced the trajectory data before merging it with the roadway inventory data. SPMD data contains data for every 100 centiseconds. The researchers only used data at the second level for merging. This was done to speed up merging process without losing accuracy. Open source GIS tool QGIS was used for merging the two datasets. After conflation of two datasets, every geographical point in the SPMD dataset will have its corresponding information from HPMS, such as speed limit. The authors defined the speeding event as a speeding behavior, driving over the speed limit, lasting more than 2 minutes. Therefore, each trip can be divided into sections like "speeding section" and "not speeding section." At each "not speeding section," the loss of speed is defined as the amount of speed difference between the average speed along this "not speeding section," and the speed limit times the total travel time at this section. In other words, this speed loss is an index of the magnitude of speed lost during this "not speeding section" before a speeding event starts. For each speeding event, three variables, "vstime", "vsspeed" and "pre_spdloss" are generated. "Vstime" refers to the duration of the speeding event. It is classified into two categories: moderate (2-4 minutes) and long (longer than 4 minutes). "Vsspeed" represents the speeding pattern of this speeding event. It has been classified as moderate speeding pattern and high speeding pattern, based on whether the average speed of the speeding event is larger than speed limit +5 mph. Each speeding event follows a normal driving event. For each speeding event, "pre_spdloss" indicates the speed loss during its previous normal driving event. The speed loss of each normal driving event equals average speed difference with speed limit times travel time of this normal driving event before speeding begins. This variable reflects the total amount of speed loss before a speeding event, and generally, this value is negative.

The final dataset contains 1056 speeding events with longer than 2 minutes continuous speeding behavior. This comprises of 416 trips. Table 1 presents all variables used for the association rule mining process.

TABLE 1 Overview of Selected Variables

Variable	Description	Data source	Levels	Frequency	Percentage (%)
vstime	Speeding time for one speeding event	*SPMD*	1: Moderate(2-4mins)	616	58.33
			2: Long(Longer than 4mins)	440	41.67
vsspeed	Average amount of speed over the speed limit during one speeding event	*SPMD*	1: moderate speeding (> speed limit & < speed limit +5)	553	52.37
			2: high speeding (>speed limit +5)	503	47.63
pre_spdloss	Total speed lose before the speeding event occurred	*SPMD & HPMS*	1: high speed loss (< -0.1)	296	28.03
			2: moderate_speedloss (-0.1 to 0.01)	270	25.57
			3: minor speed loss (-0.01 to 0)	385	36.46
			4: no speed loss (>=0)	105	9.94
tailgate	If this vehicle is tailgating another vehicle while speeding	*SPMD*	1: yes	278	26.33
			2: no	778	73.67
f_class	Functional class system	*HPMS*	1: Interstate	394	37.31
			2: Principal Arterial	596	56.44
			4: Minor Arterial	55	5.21
			5: Major Collector	11	1.04
access_con	If there is access control	*HPMS*	1: Yes	956	90.53
			2: No	100	9.47
shoulder_w	Width of the shoulder	*HPMS*	Numerical (0-14ft)	1056	100.00
lane_w	Width of the lane	*HPMS*	Numerical (10-13ft)	1056	100.00
median_w	Width of the median	*HPMS*	Numerical (0-99ft)	1056	100.00
pavement	Surface type	*HPMS*	2=Asphalt	73	6.91
			3= Plain Concrete	138	13.07
			4= Reinforced Concrete	52	4.92
			6=Asphalt-Concrete (AC)	41	3.88
			7=AC overlay	727	68.84
			9=Unbounded Concrete	25	2.37
speed_limi	Speed limit	*HPMS*	Numerical (25-70 mph)	1056	100.00
peak	If the event occurred during peak hour	*SPMD*	0: Non-peak hour	738	69.89
			1: Peak hour	318	30.11
median_t	Median type	*HPMS*	1= None	53	5.02
			2= Unprotected (more than 4 feet)	576	54.55
			3= Curbed	46	4.36
			4= unspecified barrier	2	0.19
			6= semi-rigid barrier	135	12.78
			7= rigid barrier	244	23.11
aadt	AADT of the speeding road segment	*HPMS*	Numerical (1183 - 193400)	1056	100.00
traveltime	Total travel time	*SPMD*	Numerical (10.03 to 156 min)	1056	100.00

RESULTS AND FINDINGS

In order to extract useful association rules, the key parameters like support and confidence were set as 0.01 and 0.8, respectively after performing some trial and error runs. TABLE 2 presents a summary of results from two separate analyses: speeding duration and speeding pattern. The total number of rules obtained from speeding duration events is 60, and there are 35 rules generated from speeding pattern analysis.

TABLE 2 Summary Statistics of Association Rules

RHS	Number of rules	Support			Confidence			Lift		
		Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
Long Speeding Duration (vstime=long)	19	0.013	0.028	0.009	0.861	0.923	0.8	2.066	2.215	1.92
Moderate speeding Duration (vstime=moderate)	41	0.027	0.076	0.009	0.933	1.000	0.8	1.599	1.714	1.371
High Speeding Pattern (vsspeed = High)	18	0.011	0.018	0.008	0.826	1.000	0.727	1.734	2.099	1.527
Moderate Speeding Pattern (vsspeed= Moderate)	17	0.014	0.032	0.008	0.801	0.917	0.727	1.529	1.75	1.389

Rules for Speeding Duration

There are two categories on the right hand side of the algorithm: *vstime = Moderate* and *vstime = Long*. The consequent *vstime = Moderate* indicates that the speeding event lasted for 2-4 minutes. The consequent *vstime = Long* indicates that the speeding event lasted more than 4 minutes. The first chunk of association rules in TABLE 3, from rule 1 to rule 19, shows the mined association rules for long speeding duration events and rule 20 to rule 60 are the mined rules of moderate speeding duration events. The lift value for all association rules is greater than 1 and the confidence value is greater than 0.8. All rules are sorted by decreasing values of lift. A common limitation of association rules mining is the lack of control in identifying important rules from the perspective of interpretation. The following short discussions elaborate on meaningful association rules mined through data analysis. The explanation focuses on only these interesting, informative rules from each class, discussing them in detail, and skipping redundant information after going through all the generated rules listed in TABLE 3.

Rules for long duration of speeding

The first rule with the highest value of lift is *f_class=interstate, Peak=1, traveltime=[>=56.1]*. The assertion is that Interstate highways are associated with long duration of speeding during the peak hours if the travel time is relatively long (more than 56 minutes in this research).

Rule 4 is *pre_spdloss=high,tailgate=yes, f_class=principal arterial, traveltime=[>=56.1]*. This rule suggests that principal arterials are associated with long duration of speeding while tailgating other vehicles after a high speed loss.

Rules for moderate duration of speeding

Rule 28 is *pre_spdloss=high,lane_w=[>= 11.5],median_t=none*. This rule asserts that roads with no median are associated with moderate duration of speeding after a high speed loss.

TABLE 3 Association Rules for Speeding Duration

	LHS	RHS	S	C	L
	Long Duration of Speeding				
1	f_class=interstate, Peak=1,traveltime=[>=56.1]	vstime=Long	0.011	0.923	2.215
2	tailgate=yes,pavement=aspha_conre_overlay,Peak=1,traveltime=[>=56.1]	vstime=Long	0.011	0.923	2.215
3	f_class=interstate,pavement=aspha_concre,median_t=rigid barrier	vstime=Long	0.010	0.917	2.200
4	pre_spdloss=high,tailgate=yes,f_class=principal arterial,traveltime=[>=56.1]	vstime=Long	0.010	0.917	2.200
5	pre_spdloss=high,tailgate=yes,pavement=aspha_conre_overlay,traveltime=[>=56.1]	vstime=Long	0.010	0.917	2.200
6	pre_spdloss=high,f_class=principal arterial,access_con=yes,peak=0,traveltime=[>=56.1]	vstime=Long	0.010	0.917	2.200
7	tailgate=yes,f_class=interstate,pavement=aspha_conre_overlay,Peak=1,median_t=unprotected	vstime=Long	0.009	0.909	2.182
8	f_class=interstate,pavement=aspha_conre_overlay,Peak=1,median_t=unprotected	vstime=Long	0.012	0.867	2.080
9	pavement=aspha_concre,speed_limi=[50, Inf],median_t=rigid barrier	vstime=Long	0.011	0.857	2.057
10	pre_spdloss=minor,median_t=rigid barrier,traveltime=[>=56.1]	vstime=Long	0.011	0.857	2.057
11	tailgate=yes,Peak=1,traveltime=[>=56.1]	vstime=Long	0.015	0.842	2.021
12	tailgate=yes,pavement=aspha_conre_overlay,median_t=unprotected,aadt=[2.36e+04, Inf],traveltime=[>=56.1]	vstime=Long	0.028	0.833	2.000
13	pre_spdloss=high,tailgate=yes,access_con=yes,traveltime=[>=56.1]	vstime=Long	0.017	0.818	1.964
14	tailgate=yes,pavement=aspha_concre,Peak=1	vstime=Long	0.009	0.818	1.964
15	f_class=interstate,pavement=plain concrete,median_t=rigid barrier	vstime=Long	0.009	0.818	1.964
16	pre_spdloss=moderate,tailgate=yes,peak=0,median_t=unprotected,traveltime=[>=56.1]	vstime=Long	0.012	0.813	1.950
17	pre_spdloss=high,pavement=aspha_conre_overlay,median_t=unprotected,traveltime=[>=56.1]	vstime=Long	0.016	0.810	1.943
18	pavement=plain concrete,median_t=rigid barrier	vstime=Long	0.011	0.800	1.920
19	pavement=aspha_concre,speed_limi=[50, Inf],Peak=1	vstime=Long	0.011	0.800	1.920
	Moderate Duration of Speeding				
20	pavement=asphalt,speed_limi=[<= 50],peak=0	vstime=Moderate	0.046	1.000	1.714
21	lane_w=[<= 11.5],speed_limi=[<= 50],aadt=[<= 23600]	vstime=Moderate	0.045	1.000	1.714
22	f_class=minor arterial,peak=0	vstime=Moderate	0.039	1.000	1.714
23	pre_spdloss=high,pavement=asphalt,aadt=[<= 23600]	vstime=Moderate	0.032	1.000	1.714
24	tailgate=no,peak=0,aadt=[<= 23600]	vstime=Moderate	0.018	1.000	1.714
25	lane_w=[>= 11.5],speed_limi=[<= 50],peak=0,median_t=none	vstime=Moderate	0.018	1.000	1.714
26	median_t=unprotected,traveltime=[-Inf,11.3]	vstime=Moderate	0.016	1.000	1.714

	LHS	RHS	S	C	L
27	pre_spdloss=minor,speed_limi=[<= 50],peak=0	vstime=Moderate	0.016	1.000	1.714
28	pre_spdloss=high,lane_w=[>= 11.5],median_t=none	vstime=Moderate	0.014	1.000	1.714
29	tailgate=no,pavement=asphalt,peak=0	vstime=Moderate	0.013	1.000	1.714
30	pre_spdloss=minor,pavement=asphalt	vstime=Moderate	0.011	1.000	1.714
31	speed_limi=[<= 50],peak=0,aadt=[<= 23600]	vstime=Moderate	0.060	0.984	1.688
32	access_con=no,lane_w=[<= 11.5]	vstime=Moderate	0.047	0.980	1.681
33	lane_w=[<= 11.5],aadt=[<= 23600]	vstime=Moderate	0.045	0.980	1.679
34	pavement=asphalt,peak=0	vstime=Moderate	0.051	0.964	1.653
35	tailgate=no,speed_limi=[<= 50]	vstime=Moderate	0.023	0.960	1.646
37	median_t=curbed	vstime=Moderate	0.042	0.957	1.640
38	tailgate=yes,peak=0,traveltime=[-Inf,11.3]	vstime=Moderate	0.020	0.955	1.636
39	lane_w=[>= 11.5],peak=0,median_t=none	vstime=Moderate	0.020	0.955	1.636
40	pre_spdloss=minor,speed_limi=[<= 50]	vstime=Moderate	0.019	0.952	1.633

Note: S=Support, C=Confidence, L=Lift

Rules for Speeding Patterns

There are two categories in the consequent: *vsspeed = High* and *vsspeed = Moderate.* The consequent vsspeed=High indicates that the drivers' average speed during this speed event was 5 mph more than the speed limit. The consequent *vsspeed=Moderate* represents that the average speed during speeding event was more than the speed limit but less than speed limit + 5mph. The research team used open source R software 'arulesCBA' package to perform this analysis *(17)*.

Rules for High Speeding Patterns

The value of lift for all association rules listed in TABLE 4 is greater than 1.3. The table is sorted by RHS and decreasing value of lift, same as Table 3. The following examples and discussion for four interesting rules are provided to better exemplify the understanding of mined rules and provide insights.

Rules for high speeding pattern events in TABLE 4:

- Rule 3 is *tailgate=yes, access_con=no, pavement=plain concrete.* The assertion here is that roads with no access control can be highly associated with high speeding pattern if drivers are tailgating other vehicles.
- Rule 4 is *pre_spdloss=high, pavement=plain concrete,peak=0.* This rule asserts that drivers exhibit high speeding patterns after they experience a high speed loss during non-peak hours.

Rules for Moderate Speeding:

- Rule 20 is *pre_spdloss=minor,f_class=interstate,pavement=aspha_conre_overlay,peak=0,median_t=semirigid barrier.* The assertion is that Interstate highways are associated with moderate speeding patterns after the drivers experience a minor speed loss.

TABLE 4 Association Rules for Speeding Patterns

	LHS	RHS	S	C	L
	High Speeding Patterns				
1	f_class=principal arterial,median_t=curbed	vsspeed=high	0.013	1.000	2.099
2	pavement=plain concrete,median_t=curbed	vsspeed=high	0.012	1.000	2.099
3	tailgate=yes,access_con=no,pavement=plain concrete	vsspeed=high	0.011	1.000	2.099
4	pre_spdloss=high,pavement=plain concrete,peak=0	vsspeed=high	0.012	0.929	1.949
5	f_class=major collector	vsspeed=high	0.009	0.909	1.909
6	access_con=yes,pavement=reinforcement concrete,median_t=rigid barrier	vsspeed=high	0.014	0.833	1.750
7	Peak=1,median_t=curbed	vsspeed=high	0.009	0.818	1.718
8	tailgate=no,pavement=reinforcement concrete,median_t=rigid barrier	vsspeed=high	0.009	0.818	1.718
9	pavement=reinforcement concrete,median_t=rigid barrier	vsspeed=high	0.016	0.810	1.700
10	pre_spdloss=high,median_t=curbed	vsspeed=high	0.018	0.792	1.662
11	tailgate=no,access_con=no,pavement=asphalt	vsspeed=high	0.014	0.789	1.657
12	pre_spdloss=minor,f_class=interstate,pavement=reinforcement concrete,peak=0	vsspeed=high	0.009	0.769	1.615
13	pre_spdloss=moderate,tailgate=yes,f_class=principal arterial,pavement=plain concrete	vsspeed=high	0.009	0.769	1.615
14	pre_spdloss=minor,tailgate=yes,f_class=interstate,Peak=1,median_t=rigid barrier	vsspeed=high	0.009	0.769	1.615
15	f_class=interstate,pavement=reinforcement concrete,Peak=1	vsspeed=high	0.009	0.750	1.575
16	pre_spdloss=no,tailgate=yes,f_class=principal arterial,Peak= 1	vsspeed=high	0.009	0.750	1.575
17	pre_spdloss=moderate,tailgate=yes,pavement=plain concrete	vsspeed=high	0.010	0.733	1.540
18	f_class=interstate,pavement=aspha_concre,Peak= 1	vsspeed=high	0.008	0.727	1.527
	Moderate Speeding Patterns				
19	pre_spdloss=minor,tailgate=yes,f_class=interstate,pavement=aspha_conre_overlay,peak=0, median_t=semirigid barrier	vsspeed=moderate	0.010	0.917	1.750
20	pre_spdloss=minor,f_class=interstate,pavement=aspha_conre_overlay,peak=0, median_t=semirigid barrier	vsspeed=moderate	0.012	0.867	1.655
21	pre_spdloss=minor,tailgate=yes,f_class=interstate,pavement=aspha_conre_overlay, median_t=semirigid barrier	vsspeed=moderate	0.015	0.842	1.608
22	pre_spdloss=minor,f_class=interstate,pavement=aspha_conre_overlay,median_t=semirigid barrier	vsspeed=moderate	0.019	0.833	1.591
23	pre_spdloss=minor,pavement=aspha_conre_overlay,Peak=1,median_t=semirigid barrier	vsspeed=moderate	0.014	0.833	1.591
24	pre_spdloss=minor,tailgate=yes,pavement=aspha_conre_overlay,peak=0, median_t=semirigid barrier	vsspeed=moderate	0.014	0.833	1.591

	LHS	RHS	S	C	L
25	tailgate=no,f_class=interstate,peak=0,median_t=semirigid barrier	vsspeed=moderate	0.009	0.818	1.562
26	pre_spdloss=high,pavement=plain concrete,Peak=1,median_t=unprotected	vsspeed=moderate	0.009	0.818	1.562
27	pre_spdloss=minor,f_class=interstate,peak=0,median_t=semirigid barrier	vsspeed=moderate	0.015	0.800	1.528
28	pre_spdloss=high,tailgate=no,f_class=principal arterial,pavement=aspha_conre_overlay,peak=0,median_t=unprotected	vsspeed=moderate	0.010	0.786	1.500
29	pre_spdloss=high,tailgate=no,f_class=principal arterial,pavement=aspha_conre_overlay,median_t=unprotected	vsspeed=moderate	0.013	0.778	1.485
30	pre_spdloss=minor,pavement=aspha_conre_overlay,median_t=semirigid barrier	vsspeed=moderate	0.032	0.773	1.476
31	pre_spdloss=minor,tailgate=yes,pavement=plain concrete,Peak=1,median_t=unprotected	vsspeed=moderate	0.014	0.750	1.432
32	pre_spdloss=high,tailgate=no,f_class=principal arterial,peak=0,median_t=unprotected	vsspeed=moderate	0.011	0.750	1.432
33	tailgate=yes,f_class=interstate,pavement=aspha_concre,peak=0	vsspeed=moderate	0.009	0.750	1.432
34	f_class=interstate,pavement=aspha_conre_overlay,peak=0,median_t=semirigid barrier	vsspeed=moderate	0.024	0.735	1.404
35	pre_spdloss=no,tailgate=yes,f_class=interstate,pavement=aspha_conre_overlay,peak=0, median_t=rigid barrier	vsspeed=moderate	0.008	0.727	1.389

Note: S=Support, C=Confidence, L=Lift

It would be tedious and infeasible to examine each association rule mined by CBA algorithm. To address this the authors developed Figure 1 using "arulesViz" package of R software *(18)*. This graph uses grouped balloon plots to visualize association rules based on different classes of groups. The size of the balloon represents the aggregate value of support and the color represents the aggregated lift values. Figure 1(a) represents the rules generated for speeding duration. Figure 1(b) represents the rules generated for speeding patterns. The rules are sorted by decreasing value of lift. For example, the first row in Figure 1(a), which has the highest lift value, consists of 2 rules with similar component: Traveltime=[>=56.1], peak=1, and other 3 items. The second row in Figure 1(b) consists of 2 rules, which have similar components: access_con=no, median_t=curbed, and 3 other items. In this way, the audience can easily identify the important rules with high value of lift and support, and the number of rules that have similar components.

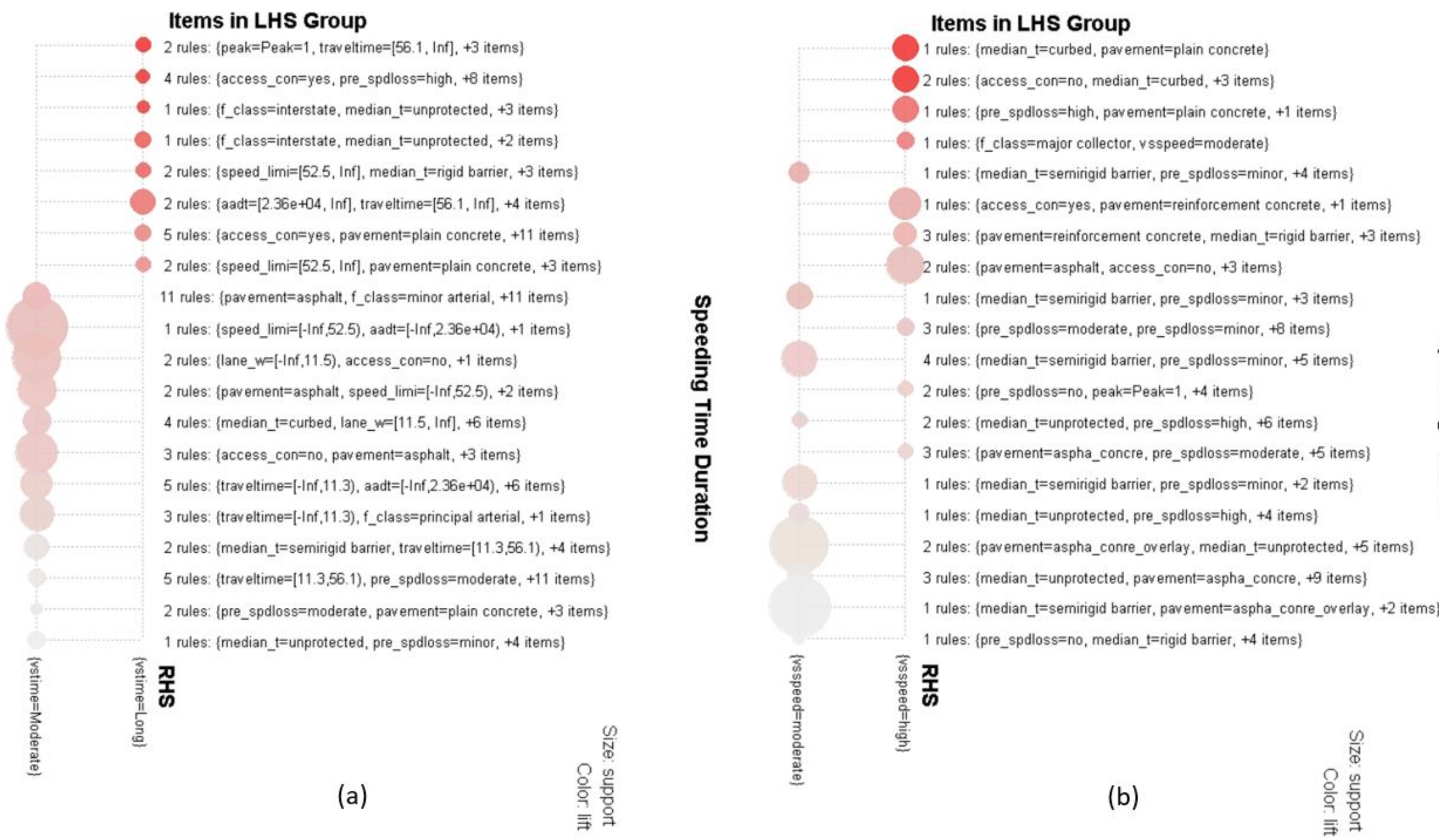

Figure 1 Grouped balloon plots of mined association rules

FINDINGS

Long Duration of Speeding

It is perceived that trip characteristics such as relatively longer travel time, driving features like driving during peak hours, high speed loss before a speeding event, tailgating another vehicle, and road geometric features such as higher functional class, presence of median, etc., are strongly associated with relatively long-lasting speeding events.

For instance, rule 1 in TABLE 3, *f_class=interstate,Peak=1,traveltime=[>=56.1]*, is the rule with the highest value of lift. It asserts that roadway features like functional class of Interstate highway and trip features such as driving during peak hours and long travel time are highly associated with long duration of speeding behavior. Corresponding indices are support = 1.1%, confidence = 92.3%, and lift = 2.215. Those indices represent the following:

(a) 1.1% of long duration speeding events are associated with these three items: driving on an Interstate highway, total trip travel time more than 56 minutes, and driving during peak period
(b) in this dataset, out of all speeding events containing this combination, 92.3% exhibited speeding behavior for longer than 4 minutes
(c) the percentage of all long duration speeding events containing this combination was 2.2 times the percentage of long duration speeding events in the overall dataset.

Out of all rules associated with long duration speeding, the trip characteristic of longer travel time was highly associated with long lasting speeding in 11 (rule1-2, rule 4-6, rule 10 – 13 and rule 16 -17) out of 18 rules. Driving during peak hours is found in 7 (rule 1, rule 2, rule 7, rule 8, rule 11, rule 14 and rule 19) out of 18 rules. High speed loss before speeding events is contained in 5 (rule 4, rule 5, rule 6, rule 13 and rule 17) out of 18 rules. Tailgating behavior features in 9 (rule 2, rule 4, rule 5, rule 7, rule11, rule 12, rule 13, rule 14 and rule 16) out of 18 rules. Roadway features like higher functional class (Interstate highway and principal arterial) and presence of median (unprotected median and rigid barrier median) are contained in several rules as well. It is worth pointing out that the observation that the magnitude of speed loss before a speeding event has an impact on the duration of speeding event falls in line with our preliminary assumption. The *pre_spdloss=high* item appears in 5 rules out of 18. Moreover, there are 5 rules (rule 4, rule 5, rule 6, rule 13 and rule 17) which contain the combination of two items *pre_spdloss=high, Traveltime= [>=56.1]*, which are highly associated the long duration speeding. This indicates that when the trips are longer, drivers are more likely to speed for a longer duration after remaining stuck in traffic for a relatively long time, to compensate for the lost travel time.

Moderate Duration of Speeding

It was observed that moderate duration of speeding is strongly associated with trip features like driving during non-peak hours, and road characteristics such as relatively lower speed limit (less than 50mph), lower functional class, relatively lower AADT (less than 23,600) and no median.

Driving during non-peak hours is highly associated with moderate duration of speeding. It presents itself in 11 (rule 20, rule 22, rule 24-25, rule 27, rule 29, rule 31, rule 34, rule 37-38 and rule 40) of the first 21 rules with highest values of lift. Roadway features like no median, lower speed limit and low AADT are found to be dominant components in the rules for moderate duration of speeding.

Comparison of Two Set of Rules Generated for Speeding Duration Classification

(a) Road segments with higher functional class have significant impact on long duration speeding events. On higher road functional roadways, the speeding behavior tends to last longer. In TABLE 3, a majority of rules for moderate duration of speeding has relatively lower functional class, and the functional class feature was observed in only a few rules. Values for variables such as speed limit and AADT presented in these rules suggested that the road segment has a lower functional class. For example, for rule 21 in TABLE 3, *lane_w=[<= 11.5),speed_limi=[<= 50),aadt=[<= 23600]*, with lane width less than 11.5 ft, speed limit less than 50 mph and AADT less than 23,600, this road segment is less likely to be an Interstate highway or a principal arterial.
(b) The rules indicated that the speeding events last longer during peak hours compared to non-peak hours.
(c) These rules also show that trips with longer travel times are more likely to have long speeding events. The explanation could be that drivers tend to speed for longer durations when their overall travel times are longer, or it is also possible that longer travel times provide drivers more opportunities to speed for longer.
(d) A greater speed loss during a congested period could be a trigger for a long duration speeding event. The impact of speed loss before a speeding event doesn't seem significant for moderate speeding events.
(e) Presence or absence of median is significantly associated with speeding duration. Speeding events tended to not last long on the road segments without median. However, there are 6 different median types, if the median exists, from HPMS dataset. The specific median type does not affect the speeding duration significantly.

(f) Tailgating is highly associated with long duration speeding events but does not significantly affect moderate duration speeding. The drivers might feel more comfortable to speed for longer while tailgating another vehicle.

High Speeding Pattern

Based on the frequency of the features presented in the rules of high speeding pattern, TABLE 4 shows that several trip/driving features and several road characteristics are highly associated with high speeding pattern. Trip/driving features include driving during peak hours, tailgating other vehicles, experiencing great speed loss before a speeding event. Road characteristics include curbed/rigid barrier median and higher functional class of roadways like Interstate highway and principal arterials.

Several rules generated from high speeding pattern dataset contain curbed median or rigid barrier median. With the barrier present between traffic from opposite directions, this might provide more sense of safety that could possibly encourage them to speed more severely. In addition, higher functional class of the roadway such as Interstate highway or principal arterial was found in several rules (rule 1, rule 11-15 and rule 17-18). It is interesting to note that there are several rules with common combination of trip/driving features and roadway features. For example, in the overall dataset, 92.9% of all speeding events contain the combination of features captured by rule 1, *pre_spdloss=high,pavement=plain concrete,peak=0* (support = 1.12%, confidence = 92.9%, and lift = 1.949).

Moderate Speeding Pattern

In this set of rules for moderate speeding pattern events, trip/driving features like minor speed loss before a speeding event, driving during non-peak hours and tailgating, and roadway features such as higher functional class of the road segments and positive barrier median are highly associated with moderate speeding pattern events. It is noteworthy that a majority of rules (rule 19-24 and rule 26-32) generated for moderate speeding pattern events contain a major or minor speed loss before the speeding event occurred. In other words, moderate speeding pattern often occurred after a speed loss.

Comparison of Two Set of Rules Generated for Speeding Pattern Classification

(a) Peak period travel is highly associated with speeding events. High speeding pattern events are more likely to occur during peak period. High speeding pattern could occur when the driver experienced a great speed loss (see rule 4 in TABLE 4).
(b) Tailgating is highly associated with high speeding pattern events.
(c) A higher road functional class such as Interstate highway or principal arterial is a significant factor for both high speeding pattern and low speeding pattern. This factor presented itself in rule 1, rule 5, rules 12-16, rules 18-21, rule 25, rules 27-29 and rules 32-35.
(d) The speed loss before a speeding event is a significant contributing factor to both moderate speeding pattern and high speeding pattern.
(e) Presence or absence of median is a road feature that is significantly associated with speeding events, however, for those road segments where a median exists, the specific type of median doesn't have a considerable impact on the speeding pattern.

CONCLUSIONS

This research investigated the potential associations between speeding behavior and trip/driving/roadway features by exploring a naturalistic driving dataset using classification and association rule mining algorithm. This paper explored the associations from two perspectives: duration of speeding behavior and speeding patterns. In this study, the trajectory data from 92 drivers were conflated with HPMS dataset, which contains information on roadway features. In order to filter meaningful speeding events, the authors defined a speeding event as a continuous speeding behavior (driving speed higher than speed limit) lasting more than 2 minutes (consciously speeding). With this definition of a speeding event, 1056 speeding events were filtered out for the data mining process.

The duration of speeding behavior was classified as long duration and moderate duration. The results provide meaningful insights when comparing the mined rules between the two classes of speeding duration. From trip characteristic perspective, it was found that having trips with relatively longer travel times (greater than 56 minutes in this study), driving during peak period, and being delayed in traffic resulting in a great speed loss before a speeding event are factors that are highly associated with long duration speeding events. From driving characteristic perspective, long duration speeding events were more likely to occur while tailgating other vehicles. From road feature perspective, the higher the roadway functional class of the segment, the more likely a long duration speeding behavior. In general, higher roadway functional class roads like Interstate highways and principal arterials have higher speed limit, wider lanes, asphalt/concrete surface and higher AADT. It is worth noting from the association rules presented earlier in the paper that those features often present themselves together. Presence or absence of median is another road feature that exhibited a significant association with long duration speeding events. The presence of median could impart drivers a feeling of increased safety, however, the type of median, if a median is present, does not make significant difference.

The speeding pattern was classified as either high speeding pattern or moderate speeding pattern. In line with the findings above, the results were analyzed from three different perspectives. From trip characteristic perspective, it was inferred that high speeding pattern events are more likely to occur during peak hours. It is possible that drivers feel rushed to get to work or home during peak hours which could force them to increase their risk threshold. From driving feature perspective, tailgating another vehicle is strongly associated with high speeding pattern. This echoes with the finding for long duration speeding behavior. Tailgating could get the driver feeling safer and make them less alert while driving, which would eventually lead them to speeding more severely and for longer. Moreover, the speed loss – being stuck in the traffic – was a frequent item in the mined association rules. It indicates that the speed loss generally triggers the speeding behavior. From road feature perspective, presence or absence of median had an influence on both speeding patterns in the same general manner.

Analysis speeding behavior at a trajectory level by using this naturalistic driving dataset could help researchers and practitioners to understand driver behavior more comprehensively. For example, the findings could help practitioners understand the associations between speeding and trip/driving/roadway features and provide corresponding countermeasures to mitigate negative consequences wherever possible. The findings can also help researchers in calibrating driver behavior parameters for transportation related simulation tools. To adapt this vast driving trajectory data to our research purpose, the authors have made several assumptions such as considering only speeding for longer than two minutes as a conscious speeding event, defining thresholds for speeding loss, classifying speeding durations into long and moderate categories, and classifying speeding patterns. Therefore, future study is necessary for validating our assumptions, which can provide more solid evidence for identifying or classifying different driving behaviors.

ACKNOWLEDGEMENT

The authors appreciate the assistance provided by Apoorba Bibeka who helped clean and conflate large data files like SPMD and HPMS.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm the contribution to the paper as follows: Data collection, Analysis and interpretation of results: Xiaoqiang "Jack" Kong; method conception: Dr. Subasish Das; draft manuscript preparation: Xiaoqiang "Jack" Kong, Kartikeya Jha and Yunlong Zhang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. National Highway Traffic Safety Administration, US Department of Transportation. Traffic Safety Facts 2017, Washington, D.C.
2. Kanellaidis, G., A. Zervas, and V. Karagioules. Drivers' Risk Perception of Road Design Elements. *Transportation Human Factors*, Vol. 2, No. 1, 2000, pp. 39–48.

3. Jørgensen, M., N. Agerholm, H. Lahrmann, and B. Araghi. Driving Speed on Thoroughfares in Minor Towns in Denmark. 26th ICTCT Workshop, 2013.
4. Havarikommissionen for Vejtrafikulykker. Grove Hastighedsovertrædelser. Havarikommissionen for Vejtrafikulykker, København, 2011.
5. Andersen, C. S., K. H. Reinau, and N. Agerholm. The Relationship between Road Characteristics and Speed Collected from Floating Car Data. Presented at the 17th International Conference Road Safety on Five Continents (RS5C 2016), Rio de Janeiro, Brazil, May 2016.
6. Lobo, A., C. Rodrigues, and A. Couto. Free-Flow Speed Model Based on Portuguese Roadway Design Features for Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2348, 2013, pp. 12-18.
7. Fildes, B.N., G. Rumbold, and A. Leening. Speed Behaviour and Drivers' Attitude to Speeding. *Monash University Accident Research Centre,* Report 16, 1991, p. 186.
8. Eksler, V., M. Popolizio, and R. Allsop. How Far from Zero?: Benchmarking of Road Safety Performance in the Nordic Countries. *European Transport Safety Council ETSC*, 2009.
9. Varhelyi, A. Dynamic Speed Adaptation Based on Information Technology - a Theoretical Background. Bulletin / University of Lund, Lund Institute of Technology, Department of Traffic Planning and Engineering, Vol. 142, 1997.
10. Bassani, M., D. Dalmazzo, G. Marinelli, and C. Cirillo. The Effects of Road Geometrics and Traffic Regulations on Driver-Preferred Speeds in Northern Italy: An Exploratory Analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 25, 2014, pp. 10–26. https://doi.org/10.1016/j.trf.2014.04.019.
11. Das, S., A. Dutta, M. Jalayer, A. Bibeka, and L. Wu. Factors Influencing the Patterns of Wrong-Way Driving Crashes on Freeway Exit Ramps and Median Crossovers: Exploration Using 'Eclat' Association Rules to Promote Safety. *International Journal of Transportation Science and Technology*, Vol. 7, No. 2, 2018, pp. 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001.
12. Das, S., A. Dutta, R. Avelar, K. Dixon, X. Sun, and M. Jalayer. Supervised Association Rules Mining on Pedestrian Crashes in Urban Areas: Identifying Patterns for Appropriate Countermeasures. *International Journal of Urban Sciences*, 2018, pp. 1–19.
13. Das, S., X. Kong, and I. Tsapakis. Hit and Run Crash Analysis using Association Rules Mining. *Journal of Transportation Safety & Security*, 2019, pp. 1-20.
14. Weng, J., J. Z. Zhu, X. Yan, and Z. Liu. Investigation of Work Zone Crash Casualty Patterns Using Association Rules. *Accident Analysis & Prevention*, Vol. 92, 2016, pp. 43–52.
15. D., Henclewood, B., Yelchuru, M., Abramovich, and S., Rajiwade. Safety Pilot Model Deployment - Sample Data Environment Data Handbook. US Department of Transportation, December 2015.
16. Mobileye, "Our Technology," *Mobileye*. Available: https://www.mobileye.com/our-technology/. [Accessed: 18-Jul-2018].
17. arulesCBA: Classification Based on Association Rules. Accessed via https://cran.r-project.org/web/packages/arulesCBA/index.html
18. arulesViz: Visualizing Association Rules and Frequent Itemsets. Accessed via https://cran.r-project.org/web/packages/arulesViz/index.html