# Identifying Safety-Critical Events in Data from Naturalistic Driving Studies

Jamal Raiyn† and Adam Rayan‡

† *Computer Science Department,* Al Qasemi Academic College, Baqa Al Ghariah, Israel
raiyn@qsm.ac.il
‡ *Computer Science Department,* University of Haifa, Israel

***Abstract*** **- Within the next decade, the world will witness a radical change in the field of transportation as technology engineers work to integrate. Autonomous vehicles (AVs) into transportation networks with conventional vehicles. However, this project faces many challenges, and chief among them are safety-critical events. Safety-critical events are difficult to control, but they can be managed through the analysis of naturalistic driving data. Naturalistic driving studies (NDSs) collect data continuously from real traffic activity, so as not to miss any safety-critical event. In NDSs of AVs, the vehicles are equipped with cameras, radar, and other sensors and Internet of Things (IoT) devices to capture as much driving data as possible. This paper describes how this data can be used to identify chain of events that leads to an accident.**

***Keywords - autonomous vehicle, NDS, driving behavior, heterogeneity.***

## I. INTRODUCTION

One of the aims of transportation engineering over the next decade is to integrate autonomous vehicles (AVs) into traffic networks with human drivers in smart cities. AVs [1] equipped with smart devices that use artificial intelligence will be needed to achieve this dream, and there are a number of challenges to overcome, chief among them safety-critical events [8]. These can be detected through the collection and analysis of traffic data; however, predicting them remains a major challenge as regards three types of road traffic; traditional road traffic, AVs traffic, and mixed road traffic, where AVs and human-driven vehicles will interact with each other as societies transition to full AV use.

Due to the different sources of heterogeneity in driver behavior, it is difficult to construct an accurate model based on data collected by smartphone application. The next decade, we are likely to witness three types and phases of vehicular behavior on highways (see Figure 1).

### A. Human Driver Behavior

A number of studies have shown that driving behavior has a significant impact on safety critical events. Human driving behavior is complex because it may be affected by external factors that can be neither detected nor directly measured. To date, studies have classified driving behavior in terms of driving maneuvers (e.g. following, hard braking, lane changing, etc.)

### B. Human - AV Behavior

It is difficult to control safety-critical events caused by human driver behavior, and the number of accidents world – wide accident has increased enormously. Ideally, AVs will replace human drivers and, in so doing, eliminate human error and prevent safety-critical events. During the transition, there will be a mixed phase when human drivers and AVs share the road, and one strategy is to establish communication among human drivers based on physical layer by adding a new interface to human driver communications.

### C. AV - AV Performance

The Introduction of cognitive AVs that can learn, update, and make a productive decisions based on extracted features promises to improve traffic safety on urban roads. To carry out this complex task, AVs will have to be designed with certain human mental characteristics.

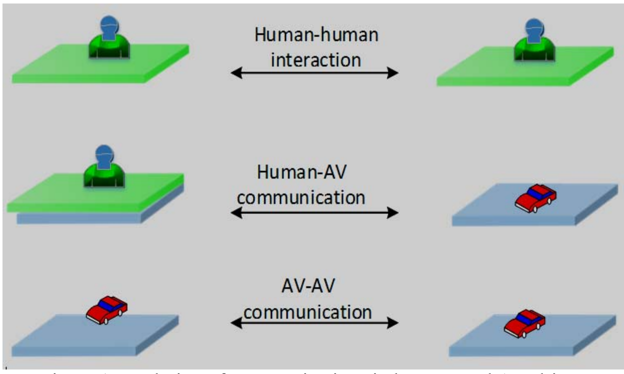


Figure 1. Evolution of communications in human- and AV-driven transportation

The rest of this paper is organized as follows: Section II gives an overview of related research; section III describes the methodologies; and sections IV and V discuss the performance evaluation and conclusions restrictively.

## II. LITERATURE REVIEW

Researchers believe that AVs can help to deter safety-critical events because many studies [2-5] show that human factors are the main cause of vehicle accidents.

### *A. Safety- Critical Events*

Various strategies have been proposed to identify safety-critical events, such as the detection of traditional thresholds. A traditional threshold is a predefined threshold of driving behavior, which, once crossed, that can lead to a safety-related event, some examples are sudden acceleration, sudden braking, and sharp turning. Typical maneuvers that have been quantified with reference thresholds for a single driver and a vehicle are sudden accelerations and braking, swift swervings, and hard right and left turns. Speed and acceleration, measured with data obtained from smartphone services are the most commonly used kinematic parameters to assess driver behaviors [8]. In addition to traditional thresholds, there are also classification strategies that can be used to identify safety-critical events.

### *B. Feature Extraction*

Driving performance features include motion data based on smartphone services and their statistical features, as illustrated Table 1. The driving performance features of the final acceleration and braking segments were extracted from their respective driving profile data. For each identified maneuver, the speed and acceleration performance were extracted, along with their respective acceleration and braking behavior [7].

TABLE 1. DRIVING PERFOMANCE FEATURE OF ACCELERATION

| Feature Description | Variable notation |
|---|---|
| Minimum speed (km/hr) | $V_{min}$ |
| Maximum speed (km/hr) | $V_{max}$ |
| Change in speed (km/hr) | $\Delta V$ |
| Mean speed (km/hr) | $V_{mean}$ |
| Standard deviation od speed | $V_{st}$ |
| Root mean square error | $V_{RMSE}$ |

### *C. Naturalistic Driving Studies (NDS)*

In a NDS, the personal vehicles of subject drivers are used to collect and store detailed data as the drivers operate under normal driving conditions.

Compared to smartphone based applications, NDSs offer a new strategy to complement existing methods for understanding driver and vehicle behavior in normal, impaired and safety-critical situations.

NDS datasets have proven to be extremely useful for the analysis of safety-critical events such as crashes and near-crashes. In short, NDSs make it possible to gather fundamental data on how people drive.

However, finding safety-critical events in NDS data using traditional methods in current use are based on may be difficult and time consuming. The traditional methods in current use are based on kinematic triggers, for example, searching for deceleration below predefined threshold that signifies harsh braking. Searching and filtering are performed by manually reviewing video data to assess whether the events flagged the triggers are actually safety-critical. This procedure is based on subjective decisions, is time-consuming, and is often hard and boring for the analysts.

### *D. Car-Following/Platooning*

Car-following models describe the processes by which drivers follow each other in the traffic stream. The goal of car-following (CF) studies is to recognize the heterogeneity of driving behaviors under different CF conditions, in order to interpret and mimic drivers' behaviors when programming AVs [10-11]. In order to investigate drivers' heterogeneous CF behaviors and the effects of modeling oversimplification, we analyzed observed behaviors using the exponential moving average (EMA).

## III. METHODOLOGY

### *A. Data Collection*

The dataset used in this research consisted of time-series and naturalistic driving data taken from the L3Pilot database [12]. A European research project L3Pilot, which tests the viability of automated driving as a safe and efficient means of transportation on public roads, has developed a common data format (CDF) for both data collection and processing and has implemented a consolidated database for processed data collection. The data consist of performance indicators for four driving scenarios, free driving, following a lead vehicle, driving in traffic jams and changing lanes.

### *B. Feature Extraction*

Feature extraction based on naturalistic driving data is important for the analysis pf driver behavior related to safety-critical events. Human driving behavior can be identified however, it is difficult to control because, although human drivers are affected by external factors that can be estimated and predicted, there are internal factors affect human cognition which cannot be distinguished or controlled. However, for AVs, both internal and external factors are predictable.

DOI 10.5013/IJSSST.a.24.01.02  ISSN: 1473-804x online, 1473-8031 print

### *C. Data Processing*

In last decade, there have been more studies on data mining algorithms such as, classification algorithms, the support vector machine (SVM), convolutional neural networks (CNNs) and deep learning algorithm (DL). These approaches have been widely used in recent traffic research for traffic flow prediction [6], incident detection [3], and accident frequency prediction [9]. A CNN is a type of feed-forward neural network that uses convolutional calculations and has a deep structure. CNNs are often used for feature extraction and identification of traffic events [10]. The algorithm learns effectively from samples of corresponding features, avoiding the complicated feature extraction process; another advantage is that it can process 1D, 2D and 3D. datasets.

Disadvantages of the CNN algorithm are that it requires large samples, and the encapsulation of the feature extraction process complicates performance improvement.

Because naturalistic driving studies focus mainly on time- series data, such as data recording activity on roads and urban highways, they use the exponential smoothing approach forecasting and analysis. This approach is used to identify of safety-critical events reflected in the naturalistic –driving data. This paper introduces a smoothed moving average scheme to characterize the travel flow on freeway [3]. However, the three main types of moving averages are the simple moving average (SMA), the weighted moving average (WMA), and the Exponential Moving Average (EMA), full detail in our previous paper [13]. The simple moving average is calculated by taking the arithmetic mean of a certain number of previous data points in a time series. Each data point is given equal weight in the calculation. The weighted moving average, on the other hand, assigns different weights to different data points in the time series, based on their relative importance. This means that recent data points are given more weight than older data points, making the weighted moving average more responsive to recent movements in the time series. The exponential moving average is similar to the weighted moving average, but it assigns exponentially decreasing weights to previous data points in the time series, with the most recent data point being given the highest weight. This makes the exponential moving average even more responsive to recent movements in the time series than the weighted moving average; it can be expressed as a percentage as follows:

$$tt^F(t) = \alpha * tt(t) + (1-\alpha) * tt(t-1) \qquad (1)$$

where $tt^F(t)$ is the travel time forecast at time $t$. ∝ is the smoothing parameter and $0 \leq \alpha \leq 1$. Alpha the responsive of forecast to sudden jumps and drops. It is the percentage weight given to the prior period, while the remainder is distributed to other historic periods. The advantage of all exponential smoothing methods is they assign weights to past cases such that that recent cases are given more weight than the previous ones. The EMA forecast model which is based on historical information yields impressive forecast performance compared to actual observations as illustrated in 2a, b, and c. In general, the moving average offers good results.

## IV. RESULTS AND DISCUSSION

In raw form, smartphone–based application data are not often informative and converting the data to information usually requires selection and reduction. The use of NDS data to evaluate the effectiveness of in-vehicle counter-measures can be facilitated by a model for rear-ending crashes that identifies important kinematic and behavioral variables.

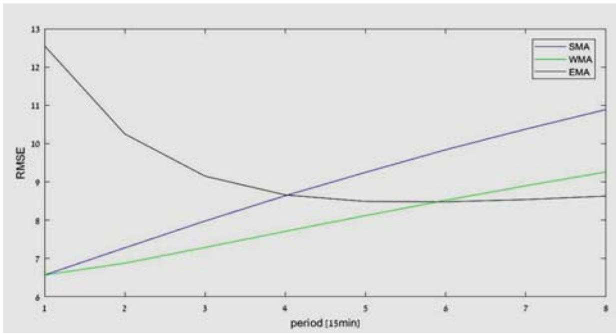

Figure 2a: RSME period-to-period

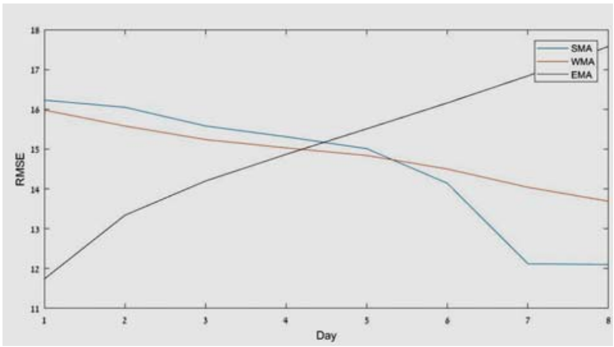

Figure 2b. RSME day-to-day

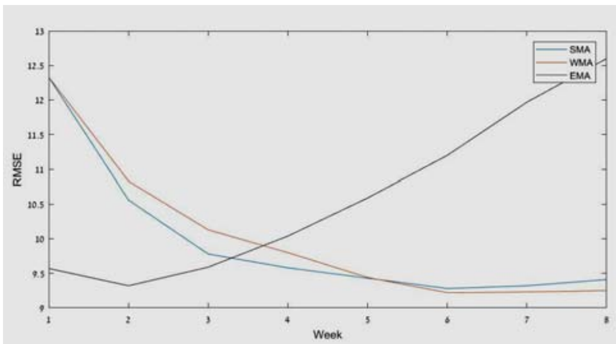

Figure 2c. RSME week-to-week

### *A. Driving Behavior in Highway*

Highway driving behavior was analyzed based on statistical measurement errors (summarized in Table 2). The EMA algorithm showed higher bias proportion than WMA and the SMA showed the lowest bias proportion, which means, the EMA predicted abnormal conditions efficiently as illustrated Figure 3. Use of statistical term proportion of variance ensured that the variance was proportionate among several factors that affected the traffic data (see Figure 4).

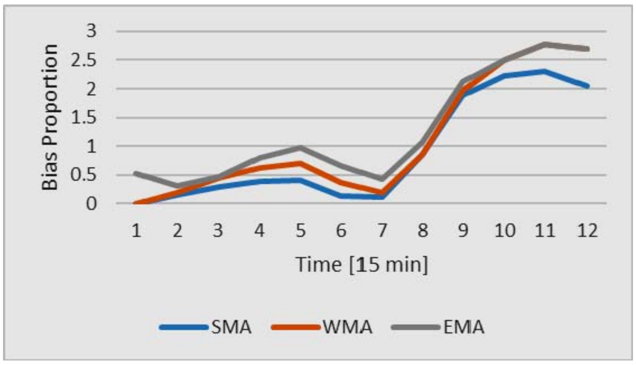


Figure 3. Bias proportions for the three types of MA

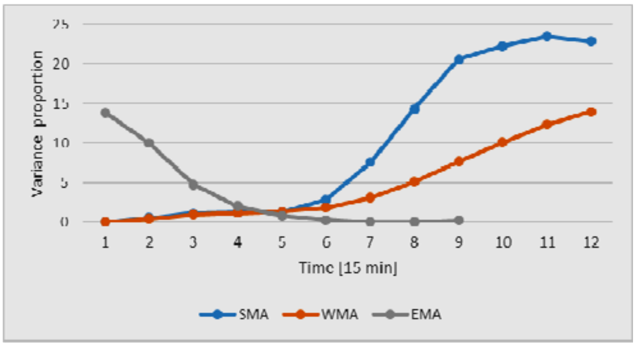


Figure 4. Variance proportion for the three types of MA

TABLE II. STATISTICAL ERROR MEASUREMENT

| Notation | Description | Equation |
|---|---|---|
| $V_{MSE}$ | Mean Squared Error | $MSE = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}{n}$ |
| $V_{RMSE}$ | Root mean square error | $RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}{n}}$ |
| $U^S$ | Variance Proportion | $U^S = \frac{\sigma_x - \sigma_{\bar{x}}}{MSE}$ |
| $V_U{}^M$ | Bias proportion | $U^M = \frac{\left(\bar{x} - \hat{x}\right)^2}{MSE}$ |

### *B. Driving Behavior in Lane Changing*

This subsection discusses the lane- changing with the aim of preventing failed lane- changing attempts, from which come most of the impacts on traffic efficiency and safety. Driving typically change lanes when faced with a slower moving lead vehicle. A lane- changing crash happens when a driver attempt to change lanes and strikes or is struck by another vehicle in the adjacent lane. Lane changing by the ego-vehicle is based on its lateral position with respect to the position of the lane markings [12]. Lane changing starts with the point at which the car starts moving in the direction of the lane marking before crossing it. It ends when the vehicle stops moving away from the lane marking after crossing it. A maximum window interval of 10 seconds before and after crossing the marking is set to limit the start- and endpoints. Figure 5 illustrates the standard deviation for longitudinal acceleration and for speed. Figure 6 illustrates the standard deviation for lateral acceleration and for position in the lane.

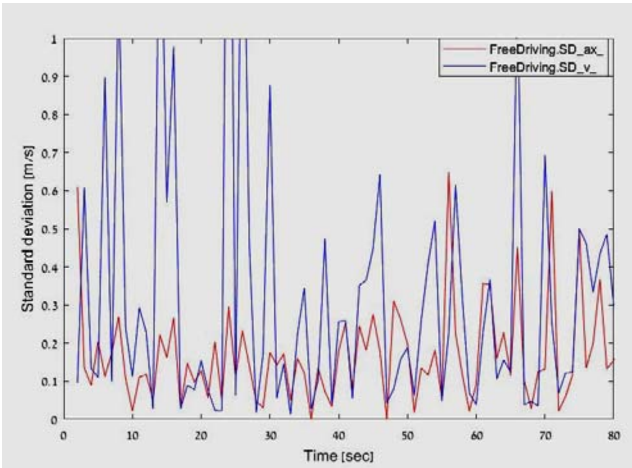


Figure 5. Standard deviation for SD_ax and SD_v

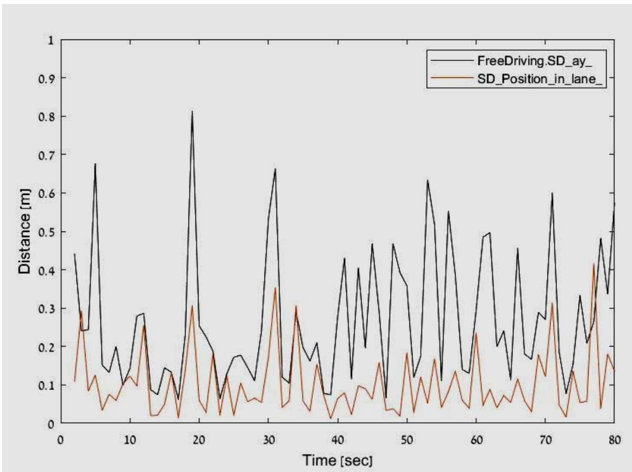


Figure 6. Relation between SD position and SD_ay

## V. CONCLUSION

This study involved a driver behavior analysis based on data colleced by smartphone applications. An exponential moving average scheme was used to analyze road time-series data based on periods of 15 minutes, one day and one week. Based on extracted data features, the EMA method can determine the cause of traffic congestion on freeway. However, it has proved difficult to establish a causal relation between the features extracted from smart-phone application data and the critical- safety events. In contrast to data from smart- phone applications, naturalistic driving serves as a unique source of data from quantifying the features of driver behavior in brake-to-stop events, such as headways distances, reaction times, and braking rates, and analysis of these features can be used to reduce traffic congestion resulting from accidents.

## REFERENCES

[1] F. Arena and G. Pau. “An Overview of Vehicular Communications ,” future internet,11,27, 2019. pp.112.

[2] M. Ahmad, Q. Chen, Z. Khan, M. Ahmad and F. Khurshid. “Infrastructure□based vehicular congestion detection scheme for V2I,” Int. J. Commun. Syst. 2019. pp.1-14

[3] J. Raiyn, “Classification of Road Traffic Anomaly Based on Travel Data Analysis,” International Review of Civil Engineering, 12, 6, 2021.

[4] D. Jia and D. Ngoduy. “ Enhanced cooperative car-following traffic model with the combination of V2V and V2I communication,” Transportation Research Part B, 90, 2016. pp. 172–191.

[5] T. Wang, J. Zhao and P. Li, “An extended car-following model at unsignalized intersections under V2V communication environment,” PLoS ONE 13(2): e0192787. pp. 1-13.

[6] J. Raiyn,”Road traffic congestion management based on search allocation approach”, Transport and Telecommunication. 18. (1), 2017. pp. 25-33.

[7] J. Vos, H. Farah and M. Hagenzieker, “Speed behaviour upon approaching freeway curves”, Accident Analysis and Prevention 159 (2021) 106276.pp.1-13. https://doi.org/10.1016/j.aap.2021.106276.

[8] Z. Wen, H. Zhang and R. Zhang, “Safety-Critical Event Identification on Mountain Roads for Traffic Safety and Environmental Protection Using Support Vector Machine with Information Entropy”, Sustainability **2021**, 13, 4426.pp. 1-15. https://doi.org/10.3390/su13084426.

[9] S. Soni, N. Reddy, A. Tsapi, B. van Arem and H. Farah, “Behavioral adaptations of human drivers interacting with automated vehicles”, Transportation Research Part F: Psychology and Behaviour 86 (2022) 48–64.

[10] L. Pariota, F. Galante, G. N. Bifulco,” Heterogeneity of Driving Behaviors in Different Car-Following Conditions”, Periodica Polytechnica Transportation Engineering, 44(2), pp. 105-114, 2016. DOI: 10.3311/PPtr.8609.

[11] J. Yarlagadda and D. S. Pawar, “Heterogeneity in the Driver Behavior: An Exploratory Study Using Real-Time Driving Data”, Hindawi, Journal of Advanced Transportation, Volume 2022, Article ID 4509071, 17 pages https://doi.org/10.1155/2022/4509071

[12] L3pilot automation driving, https://l3pilot.eu/.

[13] J. Raiyn, “A COGNITIVE RADIO SCHEME FOR DYNAMIC RESOURCE ALLOCATION BASED ON QOE, International Journal of Wireless & Mobile Networks (IJWMN) Vol. 8, No. 1, February 2016. DOI: 10.5121/ijwmn.2016.8101.

.