FINAL REPORT

Predicting changes in driving safety performance on an individualized level under naturalistic driving conditions

Date of report: February 2019

Lora Cauvuto, PhD, Associate Professor, University at Buffalo
Fadel Megahed, PhD, Assistant Professor, Miami University

Prepared by:
Lora Cauvuto, PhD
University at Buffalo
324 Bell Hall
Buffalo, NY 14260

Prepared for:
Transportation Informatics Tier I University Transportation Center
204 Ketter Hall
University at Buffalo
Buffalo, NY 14260
### Abstract

The goal of this project is to examine how driver safety performance varies by location, time of day, hours on duty, and driver workload and to model changes in performance. The following tasks will be completed: 1) model input parameters for characterizing workload; 2) quantify changes in driving performance based on mirror checks and system alerts and evaluate these changes with respect to gold standard guidelines; and 3) investigate data-driven modeling approaches for performance prediction.
Insert your own project cover page here
Acknowledgements

Robert Leonard, PhD
Tessa Chen, PhD
Nicholas Jerdack
Michael Hult
Alex Kwon
Joseph Lim

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation’s University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.
PROBLEM

Transportation incidents remain a pressing public safety issue in the United States and throughout the world, despite significant advancements in vehicle safety technologies. In the U.S., an estimated 17,775 people died in motor vehicle crashes in the first half of 2016, a 10.4% increase from 2015 (U.S. Department of Transportation 2016). Note that many of the most severe crashes involve a mix of commercial and personally-owned vehicles. While the root cause of these crashes often involves several factors, it is estimated that driver-related factors are the leading cause for 75-90% of fatal/injury-inducing crashes (Lal and Craig 2001, Medina, Lee et al. 2004, Craye, Rashwan et al. 2016). The National Highway Traffic Safety Administration (NHTSA) estimates that about 20% of all crashes are fatigue-related (U.S. Department of Transportation 2016). Even more alarming, ~60% of fatal truck crashes are due to the driver falling asleep while driving (Craye, Rashwan et al. 2016). Statistically speaking, it has been shown that drowsy driving increases crash risk by 600% when compared to normal driving (Klauer, Dingus et al. 2006).

Truck drivers operate in a stressful and unique environment. First, truck operations directly affect public interests (Mejza and Corsi 1999). In 2013, there were 411,000 truck-related crashes that resulted in 3,424 fatalities and an estimated 111,000 people injured (U.S. Department of Transportation 2016). Second, truck drivers experience little supervision or contact with fellow employees (Zohar, Huang et al. 2015). Third, a truck driver’s operating environment is very complex. Drivers encounter different routes/paths, weather, traffic conditions, and locations each time they take a trip. In addition, they can drive for long continuous hours. We currently do not understand how these factors interact, and what are their combined effect on driver safety performance (DSP), and consequently, crash risk. In the proposed project, we will investigate how real-time conditions interact to affect DSP changes. From that understanding, practitioners and drivers can make more informed decisions regarding scheduling and driver safety interventions, to reduce the likelihood of a crash.

Our goal is to examine how trucking companies can effectively utilize the massive data collected by wearable driver-worn sensors to: 1) understand how DSP varies by location, time, time of day, and/or driver workload, 2) reduce the dimension of the data to allow for real-time decision-making, 3) identify and eliminate redundant variables within these datasets, 4) model the rate of change in DSP, either parametrically or non-parametrically, and 5) use the insights from the models to predict hazardous behaviors of drivers. The proposed project provides a data analytics perspective on transportation safety. We attempt to see patterns using descriptive analytics approaches, then we will apply a variety of statistical, modeling, and machine learning approaches to characterize historical data for the purpose of making predictions about future performance.

CURRENT STATE OF KNOWLEDGE

Wang et al. (Wang, Yang et al. 2006) classified the current DSP detection systems into three major categories: 1) direct methods based on a driver’s state, focusing on eye/eyelid movements and physiological state changes, 2) indirect methods that focus on driver performance, with an emphasis on vehicle behavior including position and headway, and 3) methods that combine the driver’s state and performance. Much of the literature, however, focuses on directly detecting root causes of any change in performance [see e.g., (Jo, Lee et al. 2014)].

Variables such as the times of lane departures, standard deviation of lateral position (SDLP), and maximum lane deviation were found to highly be correlated with eye closures (Skipper 1985). The mean square of lane deviation, mean square of high-pass lateral position, and SDLP showed good potential as drowsiness indicators (Hartley 1995). Thus, driver performance measures can be sufficient in identifying fatigue, sleepiness, distraction, etc. Studies on driver performance have mainly employed lane tracking alone or in combination with tracking the distance between the vehicle of interest and the car/truck in front (Wang, Yang et al. 2006, Liu, Hosking et al. 2009). While lane departure systems are now deployed in commercial and passenger vehicles, it is unclear how this data can be effectively used to detect changes in a driver’s safety performance. Liu et al. (Liu, Hosking et al. 2009) systematically reviewed 17
behavioral experiments where indirect methods were used to detect increase in drowsiness. In their results, they stated “the studies were too dissimilar to produce any common measure that could be meaningfully pooled. Furthermore, most studies did not report numerical estimates, or did not produce graphs with error bars.” From a trucking perspective, the studies in (Liu, Hosking et al. 2009) are not relevant since they were limited to two hours of driving and primarily involved driving simulators. Studies that compared the effects of drowsiness/fatigue found some discrepancies between simulated and real-world driving conditions (Philip, Sagaspe et al. 2005). The discrepancies were attributed to the greater level of stimulation in the real-world when compared to simulated driving (Philip, Sagaspe et al. 2005).

Based on the this discussion and the detailed surveys in (Wang, Yang et al. 2006, Liu, Hosking et al. 2009), there are several observations to be made. First, there has not been much focus on developing statistical models for detecting a change in trucking performance (especially for massive datasets). It is not clear if: a) the current methods are applicable to a network of vehicles and b) the methods should be individualized by driver since driving styles can affect the baselines for SDLP, time between minor/major lane departure. Second, driver performance studies often ignore data on driving conditions. For example, all hard brakes in these studies are considered adverse events rather than classifying them based on other factors. Third, few datasets have included driver characteristics, driving behaviors, and driver distraction in a naturalistic environment due to limitations in the technology to capture this data previously. The development and implementation of wearable technology to monitor driver parameters now allows for collection of naturalistic driver behavior and the resulting driver performance; however, this data has not yet been incorporated to predict driver safety performance.

**SCOPE OF WORK**

The overall objective of this project is to model changes in driving safety performance based on individualized driver characteristics and the cognitive load of the driving task. To achieve this objective, the following tasks are proposed:

**Task 1:** Model input parameters for characterizing workload: tasks performed, cognitive load, miles driven, road location, hours on duty, time of day, driving characteristics, weather conditions

**Task 2:** Quantify changes in driving performance based on characteristics of mirror checks and system alerts and evaluate these changes with respect to gold standard guidelines

**Task 3:** Investigate data-driven modeling approaches for DSP prediction, including structural analysis and machine learning approaches. Detection once a problem has occurred, such as a critical event like a crash, is often too late. Prediction of the likelihood of a fatigued state, or a change in safety performance, can allow for interventions to be implemented. Once an alert to a behavior change is made, it is also critical to determine the driver responsiveness to the alert to quantify the effectiveness.

In the completion of these tasks, the measurable output (ground truth) will be an observed change in performance, defined here as an “unsafe act”. It is hypothesized that the proposed modeling approaches will be able to predict a change in driving safety performance, with at least a 90% accuracy, based on a change in parameters preceding the unsafe act.

**APPROACH AND METHODOLOGY**

**DATA SOURCE**

This work made use of data collected from the Co-Pilot SE™ from Maven Machines (see Figure 1). We partnered with Maven Machines who provided access to their dataset from a fleet of truck drivers who wore the device. The headset device captured head movements in the left, right, and down directions at a rate of 50 Hz and transmitted to an application running on the driver’s cell phone. Two datasets were received from Maven Machines in October 2017. The data was captured from two versions of the hardware. Along with the data, Maven provided a data dictionary describing the data variables included. Each row of the data is an “event”, considered when a driver looks either at the left or right mirror or
Specific variables in the dataset include date and time, employee id, GPS location (latitude and longitude), event type (left, right, or down), maximum angle of head turn, duration of event, and speed of driving at the time of event. The system also provides a status indicator that notes whether the driver was on the phone and whether the phone screen was on. Data was collected from the period August 14, 2016 through October 26, 2017. More than 10 million events from ~200 drivers were included prior to data cleaning.

A third dataset was received from Maven Machines in fall 2018 for brake events. Brakes were captured based on the acceleration recorded by the headset. The severity of the brake event was determined based on the maximum acceleration. Industry standard defines a hard brake as an acceleration of 10 mph/s. This dataset was recorded with the second version of the hardware and included travel from April through November 2017. Variables in this dataset included date and time of event, employee id, GPS location (latitude and longitude), acceleration during the brake, maximum acceleration during the brake, speed at the start of the brake, speed at the end of the brake, duration and distance covered during the braking event, time since last mirror check prior to the braking event, and distance since last mirror check prior to the braking event. Both mirror check and hard brake events serve as driver safety performance outcomes. The analysis takes advantage of the other variables as explanatory factors for the performance outcomes.

**DATA CLEANING**

The first step of analysis was to merge the datasets and clean any bad data from the files. The files were edited to create consistent variable naming and formatting. Bad data were identified based on the following criteria:

- System status indicators based on binary bit data from events when the system was not being worn or the truck was not being driven (see Table 1)
- Driving speed > 90 mph: deemed to be infeasible for real data
- Sensor position with maximum angle > 180 deg or < -180 deg: infeasible if worn
- GPS latitude or longitude outside of the United States
- Maximum acceleration for a brake event < 3.5 mph/s: considered to be “moderate” braking level
- Invalid data point based on data dictionary (e.g., when speed = -1)

This cleaning brought the dataset from > 12 million events to 9.8 million events where the sensor was connected and worn for a total of 198 drivers.
The second step of data cleaning was to separate the data into separate trips for longitudinal analysis. Separation of one trip from the next was based on a heuristic of the time between events. It was assumed that a gap of more than 30 minutes between mirror check events would only occur if the driver stopped the truck for a break/delivery/other reason, and then resumed driving. Therefore, the time between mirror checks was calculated and the threshold was applied. As the objective was to look at changes in driving safety performance over the duration of a trip, a second trip heuristic was applied to only include trips that lasted longer than 20 minutes. Overall there were 11,673 trips that lasted longer than 20 minutes, with an average trip duration equal to 2.15 hr. The figure below shows a trip example from the dataset and the accompanying proportions of mirror checks that were to the left side, right side, and down.

**DATA ANALYSIS**

**Input Variables**

Following initial data cleaning, exploratory data analysis was conducted to isolate the relevant input parameters. Common data visualization approaches were employed. Variables identified for analysis included: distribution of events that are left, right, or down (mirror checks vs. potential distraction to look at other in-cab items); number of event statuses that indicate the driver was on a phone call or their phone screen was on (a sign of distraction); duration of trips; time of day when the trip occurs; average speed while driving with different events; duration of glances away from center; miles driven; continuous miles driven between breaks; and road types (highway vs. local road). Statistical analysis, including summary statistics and correlations, was conducted to capture central tendency, variability, and distribution of the parameters.

**Outcome Variables**

Two main outcome variables were considered: 1) the time between mirror checks, here defined as the time between consecutive events, and 2) the occurrence of a medium or hard brake, here defined as a maximum acceleration > 3.5 mph/s for a brake event. Only left-left, left-right, and right-left events were considered, since a down event would not indicate a mirror check. The Federal Motor Carrier Safety Administration (FMCSA) recommends that drivers check their mirrors every 5-8 seconds, with a balance of left and right mirror checks. This would result in 7.5-12 mirror checks per minute. For the purpose of this analysis we also considered a threshold at 6 per minute (10 seconds). Since the brake events were only available for a subset of the data, predictors of and changes in mirror check rate were of primary interest. Following this, analysis focused on whether changes in mirror check rate, along with the input variables, were predictors of a brake event. There were 2737 brake events across ~40 drivers.
**Modeling**

With the goal of predicting driver safety performance, logistic regression was used to determine whether the input variables were predictive of an unsafe mirror check rate. For this analysis, the output variable of average mirror check rate was dichotomized to either safe (> 6 per minute) or unsafe (< 6 per minute). The predictors considered for this model were the trip duration, median speed, proportion of events where the driver was looking down, and the balance of mirror checks between the left and right sides. The second level of analysis then considered changes in the mirror check rate over the duration of the trip and the influence on the occurrence of a brake event. This second level of analysis was only performed on the subset of mirror check data that included the drivers present in the brake event data.

**FINDINGS: DOCUMENTATION OF DATA GATHERED, ANALYSES PERFORMED, RESULTS ACHIEVED**

As a part of our critical review of the literature, we have learned (through a novel, data-driven bibliometric approach) that there is a divide between the crash risk prediction models and how they are incorporated in decision-making through prescriptive optimization techniques. Since this was such a significant finding, we have expanded on this analysis and submitted this work to *Transportation Research - Part C: Emerging Technologies*. We invite the interested reader to examine our analysis further at: [https://caimiao0714.github.io/TrafficSafetyReviewRmarkdown/](https://caimiao0714.github.io/TrafficSafetyReviewRmarkdown/)

From the cleaned dataset, we extracted the input parameters for workload, including trip duration, time of day, median speed, percent of time looking down, and percent of events when the driver was on the phone. 61.7% of the events happened during the day and 38.3% at night, with most data collected Monday-Friday. Drivers looked to the left mirror 58.2% of events, to the right 35.5% and down for 6.3%. This distribution of events remained consistent based on the time of day and across drivers. Approximately 45% of longer separation between mirror checks (>8 sec) occurred when the driver was on the phone, and ~10% were when the driver’s phone screen was on.
We then looked more closely at the most recent subset of 974 trips that were collected with the newer hardware during the period that also included the brake data. For this subset, the average trip duration was 192 minutes (3.2 hours). This subset was further reduced to consider trips where the median speed was greater than 30 mph (steady driving) and that had at least 200 events. It was assumed that periods of consistent driving were needed to accurately analyze mirror check rates and brake events. This resulted in 668 trips with an average duration of 206 minutes (3.4 hr). Characteristics of the input variables are shown in the table below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip Duration (min)</td>
<td>206</td>
<td>131.8</td>
</tr>
<tr>
<td>Median Speed</td>
<td>58.0</td>
<td>8.8</td>
</tr>
<tr>
<td>Number of Events</td>
<td>1294</td>
<td>903</td>
</tr>
<tr>
<td>% Down Events</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Ratio of Left to Right Events</td>
<td>2.01</td>
<td>4.84</td>
</tr>
</tbody>
</table>

The response variable of average mirror check rate was used for modeling changes in driving performance. Independent of trip duration, the median number of seconds between mirror checks was 9.73. As shown in the graph below for the full dataset of trips, the average mirror check separation time remained consistent independent of trip duration. The standard recommendation for commercial drivers is to check their mirrors every 5-10 seconds, thus the average separation was within the recommendation. The mirror check rate was considered both as a continuous and dichotomous variable for modeling. For the dichotomous form, we considered the target of 6 mirror checks per minute and classified trips with an average event separation time < 10 seconds as safe and those with an average > 10 seconds as unsafe. This was the average for the full trip, without considering changes over time.
Logistic regression was used to model the effects of median speed, proportion of down events, left-right mirror check balance, and trip duration on whether the driver had a safe or unsafe mirror check rate. These variables were selected as the inputs of interest based on hypotheses from the literature in terms of the influence of distraction (down events), failure to check for other vehicles (left-right event balance), and fatigue (trip duration). The results of the classification for the dichotomous variable of safe driving behavior are shown in the table below. As seen, the overall prediction performance was 66%. Left-right event balance and trip duration were significant predictors ($p < 0.001$ and $p = 0.001$, respectively), whereas median speed and percent of down events were not significant ($p = 0.24$ and $p = 0.75$).

**Classification Table**

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Safe</td>
</tr>
<tr>
<td>Step 1</td>
<td>Safe</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
</tr>
</tbody>
</table>

a. The cut value is .500

Further investigation into the significant predictors showed that a higher ratio of left to right events (as shown in the figure below), meaning a greater imbalance in mirror checks, was associated with the longer separation between events. The average for the “unsafe” rate was 2.2 (standard deviation = 2.4) and the average for the “safe” rate was 1.5 (standard deviation = 0.35). For the other significant factor, trip duration, the average trip duration was shorter for those trips that had a “safe” mirror check rate. The average for the “unsafe” rate was 225 (128) min compared to 185 (104) min. In interpreting these results, it should be noted that inaccuracies in event detection, either due to the motion that the driver used or the sensor in the headset, may have resulted in a failure to capture all of the head movements toward the mirrors.
REFERENCES


