Accident Analysis and Prevention 96 (2016) 162–168

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform

Robert Kluger[a,*], Brian L. Smith[b], Hyungjun Park[b], Daniel J. Dailey[c]

[a] University of Virginia, Department of Civil and Environmental Engineering, 351 McCormick Road, Charlottesville, VA, 22903, United States
[b] University of Virginia, United States
[c] University of Washington, United States

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form 25 May 2016
Accepted 4 August 2016

Keywords:
Crashes
Near-crashes
Safety-critical event
Naturalistic driving study
Discrete fourier transform

ABSTRACT

Recent technological advances have made it both feasible and practical to identify unsafe driving behaviors using second-by-second trajectory data. Presented in this paper is a unique approach to detecting safety-critical events using vehicles' longitudinal accelerations. A Discrete Fourier Transform is used in combination with K-means clustering to flag patterns in the vehicles' accelerations in time-series that are likely to be crashes or near-crashes. The algorithm was able to detect roughly 78% of cras-javascript:void(0)hes and near-crashes (71 out of 91 validated events in the Naturalistic Driving Study data used), while generating about 1 false positive every 2.7 h. In addition to presenting the promising results, an implementation strategy is discussed and further research topics that can improve this method are suggested in the paper.

1. Introduction

High resolution, kinematic vehicle data (second-by-second speed, acceleration, yaw, etc.) is becoming more available than ever in the transportation community. With this influx of data, there are a considerable number of potential benefits to a wide range of safety applications, including monitoring driver performance, identifying unsafe locations on the road (hot spots), or even providing real-time emergency response. However, before these benefits can be realized, there is a need to be able to identify unsafe driving activities, like crashes and near-crashes, amongst a vast amount of regular driving.

The goal this paper to is to develop ways to identify safety-critical events (SCEs), defined in this context as crashes, near-crashes, and other unsafe driving behaviors using kinematic data from single vehicles. Creating an algorithm that can detect SCEs using only the trajectories of single vehicles could have a variety of applications including:

- Allowing infrastructure providers to identify SCEs in connected vehicle environments and evaluate road network safety
- Allowing taxis and shared ride service providers to monitor their drivers and ensure they provide safe rides to customers
- Allowing insurance companies to monitor their customers' driving tendencies and better evaluate risk
- Allowing agencies to monitor fleet vehicles (e.g. buses, snow plows, etc.) for both driver performance and tort liability claims
- Allowing transportation management agencies to monitor traffic and provide emergency response when necessary
- Providing real-time alerts to emergency response services in connected vehicle environments
- Identifying events in large-scale naturalistic driving studies

Each of these applications is slightly different and will likely require varying inputs to a method or algorithm when identifying SCEs but there is a clear benefit to a variety of stakeholders by having the ability to identify them.

Many established methods for identifying unsafe driving, whether it be SCEs, or a specific subset of SCEs, rely on information to be available describing how one or more vehicles are interacting. One example of such information is Time-to-Collision (TTC), which is an estimate of how much time a vehicle has on its current trajectory before it would collide with a lead vehicle. This typically requires access to radar data, which can be expensive to equip on large fleets of vehicles. As a result the analysis was restricted to kinematic data that is native to connected vehicle standards (SAE

* Corresponding author.
E-mail address: rmk7su@virginia.edu (R. Kluger).

http://dx.doi.org/10.1016/j.aap.2016.08.006
0001-4575/

International, 2009) and can be collected from smart phones or aftermarket devices.

For this study, crash and near-crash data was acquired from the SHRP2 Naturalistic Driving Study (NDS) (Virgina Tech Transportation Institute, 2013). The methodology outlined performs subsequence-matching techniques on longitudinal accelerations observed in vehicles during a set of crashes and near-crashes. A Discrete Fourier Transform (DFT) is used to transform subsequences of the observed time-series and a K-means clustering algorithm is then used to classify those subsequences as events or baseline driving.

2. Research goals

The primary goal of this study is to develop a methodology for identifying safety critical events when given a high-resolution time series of kinematic vehicle data, specifically longitudinal acceleration. With recent advancements in vehicle and roadside technology, learning how to identify unsafe driving behavior using high-resolution data streams has become a practical endeavor that can provide benefits in a variety of applications.

Time series data was acquired from the SHRP2 NDS during crash and near-crash events. The goal of the algorithm developed was to identify time series where a crash or near-crash occurred without flagging time series that did not contain any SCEs. Before proceeding, it is necessary to provide definitions relevant to this study.

- *Crash: "Any contact with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated. Includes other vehicles, roadside barriers, objects on or off the roadway, pedestrians, cyclists, or animals."*
- *Near-Crash: "Any circumstance that requires a rapid evasive maneuver by the participant vehicle or any other vehicle, pedestrian, cyclist, or animal, to avoid a crash. A rapid evasive maneuver is defined as steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle capabilities."*
- *Baseline: Any time series without a crash or near-crash.*
- *Safety-Critical Event (will be used synonymously with the term "Event"): Any crash or near-crash event.*

The crash, near-crash and driving definitions were those used by VTTI for their naturalistic driving studies (Guo et al., 2010), since that is the source of the data. The authors defined a safety-critical event as any crash or near-crash, though a case can certainly be made to include other situations and will also be discussed further at a later point.

The proposed algorithm takes the following steps:

- Break time-series into small subsequences or "windows" to examine specific sections in time
- Perform Discrete Fourier Transform to identify the strength of different frequencies present in each window
- Execute K-means clustering to group each window by the strength of different frequencies.

Relevant literature is examined, including additional context for the research motivation as well as some information on previous approaches to this problem. Then a description of the methods used and why they were applied is provided. While the range of applications is diverse, the specific inputs the presented methodology addresses is light vehicle crashes and near-crashes. The discussion section addresses how this algorithm may change based on specifics of each application.

3. Literature review

Discussed in this section will be background information on two key topics relevant to this paper. The first will outline a few studies that collect high-resolution kinematic data on a large-scale. Second, there will be a review of literature that uses this type of data to classify events, or other patterns and behaviors, with a description of the methods being used.

Two studies that have successfully recorded kinematic data during crashes are the 100-Car Naturalistic Driving Study (Dingus et al., 2006) and the larger follow up, SHRP2 Naturalistic Driving Study (Virgina Tech Transportation Institute, 2013). In both of these studies, subjects were recruited to equip their vehicles with cameras, radar, and a Data Acquisition System (DAS) designed by Virginia Tech Transportation Institute (VTTI). High-resolution trip-level data was then generated for subjects over the life of each study. Other studies such as the Safety Pilot Model Deployment in Ann Arbor, Michigan (Harding et al., 2014), the NGSim study in California (Halkias and Colyar, 2006), and Integrated Vehicle-Based System Safety Field Operation Test in Ann Arbor, Michigan (Sayer et al., 2008) Also collected similar data on different scales and in different contexts. Since all of these studies are naturalistic, subjects could, and on occasion did, get into crashes and near-crashes.

In particular, the SHRP2 NDS is unique due to the scale of the study in terms of both network coverage and number of participating subjects, the presence of a system for documenting events, and the presence of a suite of cameras equipped to vehicles for establishing ground truth. While some of the other listed studies also had some of those qualities, they were unable to accomplish all of those at the level of the SHRP2 NDS.

In the SHRP2 NDS, VTTI and the field teams at each site were responsible for identifying when and where their subjects got into crashes. Their approach was to use a collection of criteria to flag potential events in the trip data collected. Those flags include a longitudinal acceleration threshold, a lateral acceleration threshold, some time-to-collision (TTC) thresholds, a yaw rate trigger, and an event button that subjects could press to signal a collision. Individual thresholds alone (e.g. 0.6 *g*'s of longitudinal acceleration) tended to have low recall (true positives/total events) and many of them also had poor precision (true positives/test positive) (Dingus et al., 2006). The SHRP2 Study has adjusted the criteria to flag events by removing most of the radar-based triggers, adding a time-component to the deceleration, adjusting the acceleration thresholds to 0.75 *g*'s of lateral acceleration and 0.65 *g*'s of longitudinal acceleration, and adding some vehicle-safety system activation triggers. The individual triggers often had recall in the single digits, with the best individual flag had around 20% recall. While VTTI was successful in locating crashes despite the low individual identification rates of individual flags, they were able to include some data elements that were not native to the BSM and they have video to verify if an event did or did not occur for trips that were flagged.

Vehicle trajectories from the SHRP2 NDS and similar studies have been analyzed to classify certain occurrences on the road in terms of kinematic data elements. Engström and Victor developed and patented at method using neural networks to classify driving patterns and demonstrated the method on vehicle trajectories in different roadway setting (Engstrom and Victor, 2005). McDonald et. al used a computationally efficient SAX-VOX method to transform time series data into character strings and perform natural language processing to identify commonly observed action and patterns (Mcdonald et al., 2013).

In terms of specifically detecting safety-critical events, Wu and Jovanis proposed a novel algorithm to classify crash types using the maximum differences over time in both lateral and longitudinal accelerations during crashes and near-crashes. They also outlined