Accident Analysis and Prevention 113 (2018) 97–105

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Definition of run-off-road crash clusters—For safety benefit estimation and driver assistance development

Daniel Nilsson[a,b,*], Magdalena Lindman[b], Trent Victor[a,b], Marco Dozza[a]

[a] CHALMERS—University of Technology, Department of Mechanics and Maritime Sciences, Division of Vehicle Safety, Sweden
[b] Volvo Car Safety Centre—Volvo Car Corporation, Sweden

ARTICLE INFO

Keywords:
Run-off-road crash
Road departure
Lane keeping
Crash data
Test scenarios

ABSTRACT

Single-vehicle run-off-road crashes are a major traffic safety concern, as they are associated with a high proportion of fatal outcomes. In addressing run-off-road crashes, the development and evaluation of advanced driver assistance systems requires test scenarios that are representative of the variability found in real-world crashes. We apply hierarchical agglomerative cluster analysis to define similarities in a set of crash data variables, these clusters can then be used as the basis in test scenario development. Out of 13 clusters, nine test scenarios are derived, corresponding to crashes characterised by: drivers drifting off the road in daytime and night-time, high speed departures, high-angle departures on narrow roads, highways, snowy roads, loss-of-control on wet roadways, sharp curves, and high speeds on roads with severe road surface conditions. In addition, each cluster was analysed with respect to crash variables related to the crash cause and reason for the unintended lane departure. The study shows that cluster analysis of representative data provides a statistically based method to identify relevant properties for run-off-road test scenarios. This was done to support development of vehicle-based run-off-road countermeasures and driver behaviour models used in virtual testing. Future studies should use driver behaviour from naturalistic driving data to further define how test-scenarios and behavioural causation mechanisms should be included.

1. Introduction

Significant effort has recently been directed towards development of advanced driver assistance systems (ADAS) that address single-vehicle run-off-road crashes. Run-off-road crashes are an important subset of crashes worldwide and crashes resulting from lane departure constitute a high proportion of severe or fatal crashes, as seen in studies from Germany and in the US (Dissanayake and Roy, 2014; Kuehn et al., 2009; Kusano and Gabler, 2014; Liu and Subramanian, 2009). Considerable effort has been spent on developing infrastructure- (e.g. rumble strips) and vehicle-based (e.g. lane departure warnings) countermeasures (Navarro et al., 2011).

When introducing new vehicle-based safety systems to the market, one key issue is to understand their safety benefit - how well they do what they are designed for. For example, in one prospective (before on-market) study based on US crash data, it was estimated that lane departure warning (LDW) systems could reduce the number of crashes by 26.1% (Scanlon et al., 2015). Another study estimated the potential number of single-vehicle crashes in the US that could be relevant for LDW to be 6% of all single-vehicle crashes, 10% of nonfatal injury crashes, and 31% of fatal crashes (Jermakian, 2011). While retrospective real-world data analysis on the effectiveness of (on-market) LDW is scarcely available, a recent study of passenger car injury crashes in Sweden suggests that the real-world effectiveness of LDW was estimated to 30% (with a lower limit of 6%) on all roads, and up to 53% (with a lower limit of 11%) on dry or wet roads with a speed limit between 70 and 120 km/h (Sternlund et al., 2016). Recent results from a study of insurance claims in Sweden between 2012 and 2015 suggests a 30% reduction from LDW in road departure crashes without loss of control, where the repair cost exceeded 75,000 SEK (Isaksson-Hellman and Lindman, in press).

Reliance on retrospective analysis of real-world crashes for safety performance evaluations creates a delay in prioritization of future ADAS development efforts. For example, the LDW system investigated in the above retrospective analyses was introduced to the market in some 2008 models of Volvo cars (Volvo Car Company, 2007), and almost 10 years later the traffic safety effects can be estimated. Hence the need for prospective, virtual testing, which offers a more timely surrogate to real-world crash data. Existing real-world crashes, where a specific ADAS was not present, can be simulated in a virtual

* Corresponding author at: SAFER, Box 8077, 402 78 Göteborg, Sweden.
E-mail address: daniel.nilsson.3@volvocars.com (D. Nilsson).

https://doi.org/10.1016/j.aap.2018.01.011
Received 8 August 2017; Received in revised form 30 November 2017; Accepted 9 January 2018
0001-4575/

environment, along with a virtual implementation of the ADAS (see Lindman et al. 2010 for an example). These simulations can then be used to perform safety benefit estimations and also used during development of ADAS to improve system effectiveness in the simulated outcomes. Notable examples include application to rear-end crashes (Bärgman et al., 2017), pedestrian (Lindman et al., 2010), and to intersection scenarios (Sander, 2017; Scanlon et al., 2016). However, definition of relevant, representative and robust test scenarios for simulation and test track evaluation requires in-depth knowledge concerning key causation mechanisms and the variability of crash contributing factors in run-off-road crashes.

In most previous efforts to identify the basis for test scenarios using crash data, descriptive statistics have been calculated for a number of defined *conflict situations.*[1] Then, test scenarios have been based on the distribution of various pre-crash parameters (kinematics, road configuration, environment, etc.) in each conflict situation. This approach of developing test scenarios by calculating crash data statistics per conflict scenario has been used in various projects. In the EU project ASSESS, injury severity and fatality were used as weighting factors to rank conflict situations based on the GIDAS, OTS, and STRADA national crash databases (McCarthy et al., 2010). While single-vehicle crashes ranked highest, they were eventually excluded from the final definition of test scenarios which consisted of crashes in longitudinal traffic, with turning vehicles or crossing paths, and with pedestrians (Bartels et al., 2010; Wisch et al., 2010). The CATS project also developed test scenarios for autonomous emergency braking (AEB) systems for cyclists by using descriptive statistics for each of the different conflict situations selected (Uittenbogaard et al., 2016a, 2016b), and the APSECSS project developed test scenarios for forward looking integrated pedestrian safety systems using a similar approach (Wisch et al., 2013). For run-off-road crashes, police-reported US crash data from the National Automotive Sampling System, General Estimates System (NASS-GES) has been used to identify run-off-road test situations using a similar approach (Najm et al., 2002; Najm and Smith, 2007). Using a combination of conflict situation and cause for lane departure, in-depth crash data from insurance companies (UDB) was used to define situations describing crashes caused by unintentional lane departures (Kuehn et al., 2015).

However, large variability in traffic crash data can often cause relationships to remain hidden in rudimentary analysis approaches. It has been suggested that dividing crashes by conflict situation may introduce unwanted heterogeneity, leading to incorrect conclusions in the analysis of the crashes (Depaire et al., 2008). Data-driven cluster analysis provides an alternative approach in the treatment of crash data that may circumvent these issues, as crashes are grouped by data *similarity*[2] rather than by a priori defining a conflict situation and then providing descriptive statistics. Prior to this study, various methods of cluster analysis has been applied to police-reported crash data in order to segment crashes according to temporal (Kumar and Toshniwal, 2016) or spatial (Anderson, 2009; Bíl et al., 2013; Kim and Yamashita, 2007) distribution, or in order to identify patterns in crash data to improve understanding of factors and variations in a crash population. In the latter category of studies, cluster analysis has been applied to crashes with Swedish drivers in training (Berg et al., 2004), multi-vehicle crash data and roadway data from Poland (Nowakowska, 2012), national crash data on pedestrians in Iran (Kashani and Besharati, 2016), in Athens, Greece (Theofilatos and Efthymiou, 2012) and in Switzerland (Sasidharan et al., 2015), fatal and serious crashes involving young New Zealand drivers (Weiss et al., 2016), cyclist-motorist crashes in Denmark (Kaplan and Prato, 2013), highway crashes in Spain (De Oña et al., 2013), and crashes with casualties in Belgium (Depaire et al., 2008). Cluster analysis have previously been used to develop test scenarios, for example, in one study cluster analysis of in-depth crash data (STATS19-OTS) was used to define typical car to pedestrian situations in order to establish test protocols for autonomous emergency braking (Lenard et al., 2014). To our knowledge, no previous study has applied cluster analysis specifically to run-off-road crashes.

Selection of variables strongly influences the outcome of cluster analysis, and can be done in a number of ways (Milligan and Cooper, 1987; Punj and Stewart, 1983). One approach is to select a smaller set of variables that are considered of high interest for the analysis, such as variables that are important in the design of a certain system or technical solution (Conquest et al., 1993; Lenard et al., 2014), or variables contributing to accident and injury severity (De Oña et al., 2013, 2011). Crash data segmentation has also been done through latent class clustering (Depaire et al., 2008; Kaplan and Prato, 2013; Weiss et al., 2016), also known as model-based clustering, where a larger set of variables is used under the assumption that they are generated by a mixture of underlying probability distributions (Vermunt and Magidson, 2002). Latent class clustering has also been compared to *k*-modes for crash segmentation, where the two methods were found to yield similar results (Kumar et al., 2017). Another approach for cluster variable selection used in crash data analysis is iterative inclusion and exclusion of variables based on calculated significance, in order to find an optimal variable set (De Luca et al., 2012, 2011).

The primary aim of this study is to define clusters of run-off-road crashes for use in development of current and future advanced driver assistance systems and for timely safety-benefit estimation based on virtual testing. Therefore, hierarchical agglomerative cluster analysis was used to divide run-off-road crashes from the German In-Depth Accident Study (GIDAS) crash database, the resulting cluster solution provides a basis for run-off-road test scenario definition.

Lastly, as driver state may strongly influence the effectiveness of a countermeasure, observed driver behaviour in crashes is relevant for the development of ADAS (e.g. to describe crash causation mechanisms to be addressed by the system and for development of driver behaviour models for use in virtual testing). To exemplify, if a critical situation occurs as a consequence of driver inattention, a warning may redirect driver attention, so that the driver may resolve the situation. Conversely, if the driver is intoxicated or suffers from a heart attack, warnings can be expected to have less impact on the outcome. Models of driver behaviour should therefore represent real-world distributions of driver state parameters, for successful simulation-based estimation of the effectiveness of a countermeasure. However, driver-related variables in retrospective in-depth crash databases such as GIDAS are often missing or involve uncertainty due to subjective assessment based on data collected at the crash site and from interviews (Larsen, 2004; Ljung Aust et al., 2010; Otte et al., 2009), as opposed to being more objectively coded from pre-crash video as in naturalistic databases. Still, important clues for the development of valid driver behaviour models may be identified using retrospective crash data. Thus, an additional aim for this paper was to analyse the clusters with respect to the distribution of pre-crash factors that may provide valuable information about driver state—and, in essence, the driver's role—to guide future development of driver behaviour models with relevance for run-off-road mitigation simulations.

2. Data and methods

2.1. Data selection

The present study is based on the GIDAS (German In-Depth Accident Study) database, consisting of in-depth traffic crash data collected at sites in and around Dresden and Hannover, Germany (Otte et al., 2003). About 2000 crashes with at least one injured are collected each year, and data collected between 2008 and 2014 were used. This set contains a total of

[1] Conflict situation here refers to the initial situation (e.g. two vehicles travelling in opposing lanes), and not the crash configuration (e.g. whether the two vehicles collide front-front, front-side etc.). Examples of alternative nomenclature used in the literature are: pre-crash scenario, crash type, and accident type.

[2] Similarity, in cluster analysis, refers to a simplified measure that describes the between-object (here: between-crash) difference, and is based on a larger set of variables.