Using data from a smartphone app to analyse distraction and drowsiness of drivers

Sara Ferreira, Faculdade de Engenharia, Universidade Porto, Portugal, sara@fe.up.pt
Zafeiris Kokkinogenis, University of Porto, Portugal, zafeiris.kokkinogenis@gmail.com
Antonio Couto, University of Porto, Portugal, fcouto@fe.up.pt

Keywords: Drowsiness, Distraction, Smartphone, Safety, Human factors, Driver monitoring system

BACKGROUND

Driver distraction and drowsiness have received increased attention during recent years due to the rapid increase in vehicle automation technology that impact on the driver behaviour. Driver distraction and drowsiness contribute to approximately 5–25% of all crashes (European Commission 2015) and around 20% of all fatal and severe crashes (Connor et al. 2002, Kecklund et al. 2011), respectively. These types of driver behaviour are associated to a degraded driving performance as well as to a significant detrimental of the cognitive performance (e.g. reaction time) and therefore, to a negative impact on road safety (Sweeney et al. 1995, Atchley et al. 2017, Fitzharris et al. 2017). Traditionally, safety research is supported on traffic crashes and/or traffic conflicts data. However, reports from traffic crashes and conflicts do not always capture the full dynamics and the conditions under which an event has occurred, particularly if the driver behaviour is the focus. Consequently, other research methods have been applied namely naturalistic (Hanowski et al. 2005; Soccolich et al. 2013) and driving simulator studies (Oviedo-Trespalacios et al. 2017; Papantoniou et al. 2015). However, these studies are both complex and costly to conduct and as a result are frequently of small size.

AIM

In this context, the present study introduces a novel approach by exploring data of driver-monitoring systems (DMS) vendor supported on a smartphone camera-based application freely available for the general drivers’ population. The DMS solution emits an alert when detects distraction or drowsiness of the driver, storing a set of information of that event (timestamp, a GPS position, the instant speed of the vehicle and the type of the alert). Moreover, information about the journey (e.g. date and time of the start and end of the journey), the driver age and sex were also collected. Deidentified data were extracted from the DMS vendor’ database. It should be noted that the research team did not supervise the DMS implementation process, neither data collection and management. Retrospective data obtained under these circumstances, we named ‘opportunistic’ data.

METHOD

In order to full explore the retrospective data, two distinct analyses were performed: 1) clustering analysis to identify drivers’ profiles and 2) a generalized linear model (GLM) to identify risk factors associated to alerts. A data treatment and rearrangement were performed and at the final, a dataset was obtained including 489 drivers to which correspond 1008 observations when aggregated the alerts by driving record. These 1088 driving records correspond to the observations modeled by the GLM, in particularly a binomial negative model. To this analysis, variables such as the number of alerts per journey (dependent variable), journey time, journey breaks, breaking duration time, age and sex of the driver were used. On the other hand, several measurements were created to be used as inputs for the clustering analysis which was based on the Hierarchical Clustering Approach (HCA) and the K-means (KM) method.

RESULTS

The application of the hierarchical approach led to the plausible number of significant clusters between 3 and 10. Additionally, the Silhouette validity index suggests three clusters (S.C. = 0.292). A
separation among the three clusters with respect to the number of inattention events is evident. One cluster is composed of all the drivers in the dataset that did not raise any alert (39% of the drivers), meanwhile another cluster is composed by 23% of drivers with the greatest number of alerts and the highest values respect to the exposition variables (time and distance related variables).

Using the number of alerts separated by distraction and drowsiness as dependent variable of the two models, distraction and drowsiness models, we found that journey time has a similar effect on both distraction and drowsiness alters, indicating that increasing 10% of the time of the journey, the distraction and drowsiness alerts increase 6% and 5%, respectively, ceteris paribus. Additionally, increasing the number of stops during the journey, the distraction and drowsiness alerts decrease however, the elasticity of this impact is not constant, being the alerts frequency sensitive to the magnitude of the number of breaks. Results show that age and sex variables have different effects on alerts of distraction and drowsiness.

CONCLUSIONS
This study explored ‘opportunistic’ data to investigate driver behavior and profile. The findings show that this type of data has potential to be explored to road safety studies. The data and its analysis should be seen as complementary to other studies using controlled methods of data collection (e.g. naturalistic and driving simulator studies), at least at this early stage of data exploitation gathered from emerged technology.

REFERENCES