Using naturalistic driving data to model crash risk at a city level

Marta Faria, Universidade de Lisboa, Portugal, ana.marta.faria@tecnico.ulisboa.pt
Patricia Baptista, Universidade de Lisboa, Portugal, patricia.baptista@tecnico.ulisboa.pt
Tiago Farias, Universidade de Lisboa, Portugal, tiago.farias@tecnico.ulisboa.pt
Sara Ferreira, University of Porto, Portugal, sara@fe.up.pt

Keywords: naturalistic driving data, driving behavior, crash risk, binary logistic regression

1. Introduction

Research suggests that driving behavior is a contributory factor in over 90 percent of crashes. Consequently, identifying the driving behavior that represents greater risk of involvement in a crash brings significant benefits to reduce road accidents. A number of methods have been applied including focusing in demographic profiles, self-reported behavior and risk preferences and personality and risk perceptions. However, results from these methods generally are based on small number of observations and may be under or over-predicted. In contrast, data collected during real world driving (naturalistic driving data) provides information on driving behavior across time and space, generating a great amount of observations for each driver. Considering this, the aim of this research work is to explore the association between driving behavior and crash occurrence, using naturalistic driving data to model the crash risk at a city level.

2. Methods and data

To accomplish the objectives of this work, firstly, a sample of 46 drivers was monitored for a period of at least six months in the Metropolitan Area of Lisbon using an on-board data logger, the i2D device. This device allows acquiring 1 Hz data on vehicle dynamics (speed, acceleration, etc.), engine data and also location information. With this data an extensive database was built, including data on the road type (defined by hierarchical street levels, from level 2 – minor arterials – to level 4 – local streets). To integrate the extensive database with crash data the authors followed a method focused on the city rather than on the driver. Therefore, driving data from all drivers were aggregated per street allowing to characterize driving behavior for each street at a city level. This database, with average driving parameters per street, was then merged with the crash database which is also street-based. Afterwards, a binary logistic regression was performed to model the crash risk as a function of several driving behavior parameters. The factors that were included in the model ranged from percentage of time spent in different speed bands, average acceleration and deceleration, percentage of time drivers accelerate or decelerate in different acceleration values, vehicle specific power (VSP) distribution, percentage of time spent in aggressive driving, among others.

3. Preliminary results and conclusions

Preliminary results reveal the existence of an association between driving behavior at a street level basis and crash risk. It was found that deceleration related parameters, such as percentage of time spent in VSP mode 1 (which occurs either for higher negative slopes or for more extreme deceleration events) and average negative VSP (comprehending all situations with negative slopes and deceleration events), are influential in predicting the crash risk. Furthermore, as expectable, the likelihood to have a crash in level 2 and 3 streets is higher than on more local ones (level 4). However, oppositely to authors’ beliefs, even if non-significant, an increase in aggressive driving (defined by acceleration thresholds as a function of speed) was found to decrease the odds of outcome being crash. Higher percentages of time spent in aggressive driving are typically found for level 2 and 3 streets (comparing with level 4 streets); therefore, an increase in aggressiveness was...
expected to increase the odds of crash outcome. Further analyses must be performed to assess this result.

Concluding, naturalistic driving data provide detailed data allowing performing in-depth analyses that were not possible until these large databases become available. This data will give a better knowledge on driving behavior allowing a more efficient selection of the measures to focus on future training assets in order to improve driving behavior and, consequently, reduce crash risk.