24th International Congress on Modelling and Simulation, Sydney, NSW, Australia, 5 to 10 December 2021
mssanz.org.au/modsim2021

Analysis of Victorian road accident environmental data

Upuli Karunaratne, Richard Watson and Peter Ryan

Ryan Watson Consulting Pty Ltd, Vermont VIC 3133, Australia
Email: richard@ryanwatsonconsulting.com.au

Abstract: Copious research has been undertaken to better understand the causes of road accidents so as to take action to mitigate their frequency and severity. Considerable data is available on road accident events and published as open data on the internet. Such data include variables relating to the accident, such as accident type: vehicle hitting stationary object, collision between vehicles, vehicle hitting pedestrian, etc; accident severity: fatal, serious injury, minor injury, property damage, etc. As well, there are many environmental variables associated with accidents such as the road type, neighbourhood infrastructure, demographics of participants, etc. These datasets are also frequently published on the internet.

Data mining is a process for investigating large datasets such as the above, often termed big data. It discovers and extracts predictable and consistent patterns as well as methodical connections among variables, and after that validates the results (Mor et al. 2021). One type of data mining is association rule mining, which aims to find statistical relationships between data, such as correlations between road accident counts per year and environmental variables for a specific region. Many studies have applied data mining to road accident analysis in various parts of the world (eg Dhanya 2021 in Bangalore, India). Some such studies have been undertaken in Australia, but there is a need for further work using the latest available data and analysis methods.

Road accident data for the state of Victoria has been published by VicRoads since 1987, and much geospatial and demographic environmental data is available from open-source platforms such as the Australian Urban Research Infrastructure Network (AURIN), Victoria's Spatial Datamart, and the City of Melbourne's open data portal. Previously the authors have used predictive machine learning to visualize and analyse Victorian road accident event data (Watson and Ryan 2020). The present study aims to use discovery techniques to establish how environmental variables, particularly geospatial, are related to accident count rates in particular areas. According to Mor et al. (2021), the process of data mining must be preceded by learning about the application domain, identifying data sources and preprocessing the data. Data mining is then used to extract patterns and identify those of interest. This presentation will review the progress achieved so far and the most appropriate analysis methods suggested by the data.

REFERENCES

Mor, N., Sood, H., Goyal, T., 2021. A Critical Review on Use of Data Mining Technique for Prediction of Road Accidents. In Singh, V. et al. (eds.), Computational Methods and Data Engineering, 141-149, Springer Nature, Singapore. https://doi.org/10.1007/978-981-15-7907-3_11141

Dhanya, V., Preethi, K., Nandini, R., Bhuvaneshwari, R., 2021. A Road Accident Prediction Model Using Data Mining Techniques. IEEE 5th International Conference on Computing Methodologies and Communication, Erode, India, April 2021. https://doi.org/10.1109/ICCMC51019.2021.9418336

Watson, R., Ryan, P., 2020. Visualization and Prediction of Road Accident Data using Python Machine Learning. Ryan Watson Consulting Pty Ltd, unpublished discussion paper, February 2020.

Keywords: *Road accident data, data mining, environmental variables, predictive machine learning*