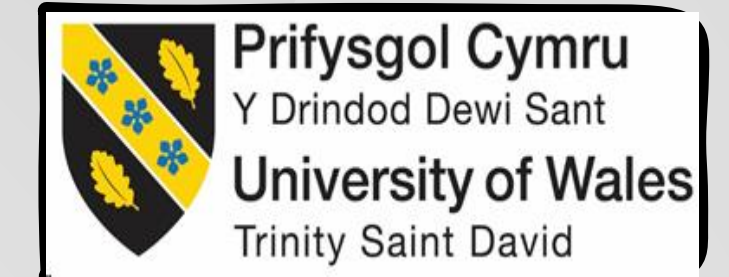


DATA MINING IN THE ROAD SAFETY

YAN ZHANG, *University of Wales Trinity St. David (UNITED KINGDOM), 1804633@Student.uwtsd.ac.uk*



Abstract

With the rapid development of the automobile industry, the number of cars is increasing, but this leads to more and more traffic accidents. Therefore, mining traffic accident data not only can help people understand the factors affecting traffic accidents, but also can help people to take corresponding protection measures to ensure the safety of people's property and life.

To this end, this research will mine road traffic accident data and try to find some factors that may affect traffic accidents.

Dataset Description

This dataset is a part of the UK road traffic accident data and was first published by the department for transport on 12th October 2018[1].

The dataset mainly records some details of road traffic accidents in the UK in 2017 and is normally used to provide the public, researchers or other stakeholders with up-to-date information on road traffic accidents.

The original dataset consists of a data table containing 129,982 records and 32 attributes in CSV format.

Data processing

- Delete some unimportant attributes
- Delete empty values
- Converts encoded numbers into text
- Add Date period column and Time period column
- Divide dataset into training set and test set
- Select important attributes based on Weka's built-in attribute selector

Data analyses

Data visualization analysis method, geographical analysis method, J48 classification method, DBCSCAN clustering method, APRIORI association method and Time series method will be used for further analysis and mining. And the result of data visualization is shown in the figure 1.



Figure 1 Data visualization

Results

- May to July, September to November, Monday to Friday, 8 to 9 in the morning and 5 to 6 in the afternoon is more likely to lead to traffic accidents
- Large cities, economic areas are prone to traffic accidents
- Junction control is Give way or uncontrolled, the speed limit is 30 MPH, the 2nd_road class is unclassified, light condition is daylight, and road type is single carriageway is more likely to lead to traffic accidents

Further work

Further work will combine other datasets to analysis vehicles and casualties involved in traffic accidents.

Relevant departments can add more detailed traffic accident information in their future work (such as whether there is traffic jam), it will help to discover more useful information, and reducing the probability of traffic accidents.

Reference

- [1] D. f. Transport, "Reported road casualties in Great Britain: 2017 annual report," 2017.