ANALYSIS OF ROAD TRAFFIC ACCIDENTS USING DATA MINING.

K. L. P. Liyanaarachchi (klppliyanarachchi@gmail.com), E. Y. A. Charles. 
Department of Computer Science, University of Jaffna.

Description of the research

Traffic accidents are one of the major impact for the development of the societies around the world. Data mining algorithms can be used to analyze accident datasets to predict rules which can help to reduce road traffic accidents. The main objective of this research is to identify more accurate and useful patterns in road traffic accident data using data mining techniques. It is expected that these patterns can be utilized to take measures such that the number of accidents or the severity of the accidents can be reduced.

This research conducted a descriptive statistical analysis to understand the dataset’s attributes and then used 348 decision tree classification method to construct a decision tree. To obtain useful results appropriate features were identified and in addition a measure was taken to reduce the effect of class imbalance. Decision tree constructed in this research work has shown an accuracy of 78.32%.

Results and Analysis

348 Decision tree classifier was used to construct the decision tree. To obtain a useful tree, appropriate features were selected using the correlation values of each attribute.

Class Imbalance problem:

The four classes were merged into two classes namely “Injury” and “Damaged only” to find the solution for the class imbalance problem. The following result was obtained after applying the feature selection on merged dataset:

Conclusion

This research study mainly considered in finding more accurate and useful patterns on the Road Traffic accident data by using 348 decision tree classifier. The classifier showed good results on the dataset and found some useful patterns.

Example rules:

- “Validity of Driving” = 1 (Valid license for the vehicle) AND “Element Type” = 9 (Private bus) AND “Number of vehicles” = 2 AND “Light Condition” = 3 (Night, wet street lighting).
  THEN “Highest Severity” = 1 (Injury).
  * correctly classified number of instances = 1256

- “Validity of Driving” = 1 (Valid license for the vehicle) AND “Element Type” = 7 (Motorcycle) AND “Driver/Rider At Fault” = 3 (Fault in the vehicle).
  THEN “Highest Severity” = 1 (Injury).
  * correctly classified number of instances = 6087

- “Validity of Driving” = 2 (Without valid license for the vehicle) AND “Element Type” = 1 (Dual purpose vehicle) AND “Light Condition” = 5 (Night, good street lighting) AND “Vehicle Ownership” = 3 (Government vehicle).
  THEN “Highest Severity” = 2 (Damage Only).
  * correctly classified number of instances = 1770

- “Validity of Driving” = 2 (Without valid license for the vehicle) AND “Element Type” = 9 (Private bus) AND “Location Type” = 2 (2-lane high way) AND “Light Condition” = 3 (Night, wet street lighting).
  THEN “Highest Severity” = 1 (Injury).
  * correctly classified number of instances = 13586

This research work can be extended to analyze the dataset further based on the obtained results. The rules obtained can be used by traffic department of Police, to prevent accidents or to reduce the severity of injuries and also to develop automated traffic control applications.