

Analysis of Road Traffic Accidents Using Data Mining

K.L.P.P. Liyanaarachchi (klppliyanaarachchi@gmail.com)¹, E.Y.A. Charles¹

¹ Department of Computer Science, University of Jaffna, Sri Lanka.

Abstract

Accident happens unexpectedly and unintentionally, typically resulting in damage or injury or in fatalities. Data mining is the extraction of implicit, previously unknown, and potentially useful information from data collected for various purposes. The main objective of this research is to identify more accurate and useful patterns that would exist in the road traffic accident data using data mining techniques. It is believed that these patterns can be utilized to take measures to reduce the number of accidents or the severity of the accidents.

As part of this research work, details of accidents occurred in Colombo district in the year 2015 were collected from the Traffic Headquarters, Colombo, Sri Lanka. A data set with 9487 accident incidents each detailed with 55 features was created from the collected data. This data consists of four types of accidents, namely, Fatal (154), Grievous (877), Non-Grievous (2028) and Vehicle damage only (6428). There are a quite a few published studies on traffic accident analysis using data mining methods. In most of these studies, J48 classifier has produced higher accuracy than other methods. So far no such study has been reported on accidents occurred in Sri Lankan roads.

A correlation analysis was performed on the data set and as a result 10 attributes were removed. In this study, the J48 decision tree classifier was used in two ways. In the first one all four types of accidents were considered. The decision tree built with 70% of the data was able to achieve an average accuracy of 71.4687%. In the second analysis, three types Fatal, Grievous and non-grievous types were combined into one class and named as Injured. This approach was taken to reduce the effect of the vehicle damage only class, which is around 68% of the total data. The decision tree built with this merged class was able to achieve an accuracy of 78.7288% using a tenfold cross validation. The decision tree was converted into 20 rules, which can predict the type of accident based on the identified attribute values. The results were found to be helpful to identify the factors influencing traffic accidents and can be further analyzed to find more subtle reasons or situations that are causing accidents.

Keywords: *Data mining, Decision tree, Road traffic accident*