Abstract

This paper emphasizes the importance of Data Mining classification algorithms in predicting the vehicle collision patterns occurred in training accident data set. This paper is aimed at
deriving classification rules which can be used for the prediction of manner of collision. The
classification algorithms viz. C4.5, C-RT, CS-MC4, Decision List, ID3, Naïve Bayes and
RndTree have been applied in predicting vehicle collision patterns. The road accident training
data set obtained from the Fatality Analysis Reporting System (FARS) which is available in the
University of Alabama’s Critical Analysis Reporting Environment (CARE) system. The
experimental results indicate that RndTree classification algorithm achieved better accuracy
than other algorithms in classifying the manner of collision which increases fatality rate in road
accidents. Also the feature selection algorithms including CFS, FCBF, Feature Ranking, MIFS
and MODTree have been explored to improve the classifier accuracy. The result shows that the
Feature Ranking method significantly improved the accuracy of the classifiers.

References

- Andreas G.K., Janecek, Wilfried N. Gansterer, Michael A. Demel Michael, Gerhard F.
  Ecker, “On the Relationship Between Feature Selection and Classification Accuracy”, 2008,
- Chang L. and H. Wang, "Analysis of traffic injury severity: An application of
  non-parametric classification tree techniques Accident analysis and prevention", 2006, Accident
- Han, J. and Kamber, M., “Data Mining: Concepts and Techniques”, Academic Press,
- Handan Ankarali Camdeviren, Ayse Canan Yazici, Zeki Akkus, Resul Bugdayci, Mehmet
  Ali Sungur, “A Comparison of logistic regression model and classification tree: An application to
- I-Cheng Yeh, Che-hui Lien, “The comparisons of data mining techniques for the
  predictive accuracy of probability of default of credit card clients”, Expert Systems with
- Isabelle Guyon, Andr´e Elisseeff, “An Introduction to variable and Feature Selection”,
- Lei Yu, Huan Liu, “Feature Selection for high-Dimensional Data: A Fast
  Correlation-Based Filter Solution”, Proceedings of the Twentieth International Conference on
- Mark A. Hall, “Correlation Based Feature Selection for Machine Learning”, Ph.D. Thesis,
  Department of Computer Science, Waikato University, Hamilton, NZ, 1999.
- Nojun Kwak and Chong-Ho Choi , “Input Feature Selection for Classification Problems”,
- Weimin Chen , Chaoqun Ma, Lin Ma , "Mining the customer credit using hybrid support
- Yong Soo Kim, “Comparison of the decision tree, artificial neural network, and linear
  regression methods based on the number and types of independent variables and sample size”,
- Tanagra Data Mining tutorials, http://data-mining-tutorials.blogspot.com
- World Health Organization, Global status report on road safety: time for action, Geneva,
  2009.
- Feature Selection Algorithm, http://featureselection.asu.edu
- J. Rose Quinlan, “Programs for machine learning”.
- Random Tree Algorithm, http://www.answers.com

**Index Terms**

Computer Science  
Data Mining

**Keywords**

Classification Algorithms  
Feature Selection Algorithms  
Manner of Collision  
Fatal  
Severity  
Collision Patterns  
Prediction