IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

Predicting Accident Zone in Thanjavur City Using Data Mining Techniques

S. Sterlin[1*], K. Lakshmi[2]

[1]M.Sc Computer Science, Idhaya College for Women, Kumbakonam, Tamilnadu, India
[2]Department of Computer Science, Idhaya College for Women, Kumbakonam, Tamilnadu, India

Corresponding Author: lakshmi47@gmail.com

Available online at: www.ijcseonline.org

***Abstract*—**The objective of this project is to predict the accident zone in kumbakonam city. This will be very useful to transport department as well as administration department of kumbakonam city and take decision easily about the accident zone. This project will give the references about the accidents places in kumbakonam city. It is very useful to reduce the accidents in the kumbakonam city. This project contains two major modules namely Administrator and User. Both administrator and user may be a police. Administrator must be a higher authority. The privilege of the administrator is to create, update and delete the users' account. Administrator can view and delete the accident data's which was updated by the user. Major operation by an administrator is to maintain accident data's and user data's. The privilege of the user is to add and update accident data's. Those accident data's can be used to generate the report about the accidents happened in kumbakonam city. User can view their profile and also view the accident data.

***Keywords*—**Accident Zone, Prediction, Road Traffic, Data Mining, Analysis.

I. INTRODUCTION

Road traffic accidents (RTAs) are a major public health concern, resulting in an estimated 1.2 million deaths and 50 million injuries worldwide each year. In the developing world, RTAs are among the leading cause of death and injury; Ethiopia in particular experiences the highest rate of such accidents. Thus, methods to reduce accident severity are of great interest to traffic agencies and the public at large. In this work, we applied data mining technologies to link recorded road characteristics to accident severity in Ethiopia, and developed a set of rules that could be used by the Ethiopian Traffic Agency to improve safety.

In recent years, with the growth of the volume and travel speed of road traffic, the number of traffic accidents, especially severe crashes, has been increasing rapidly on a yearly basis. The issue of traffic safety has raised great concerns across the globe, and it has become one of the key issues challenging the sustainable development of modern traffic and transportation. Therefore, it is crucial for engineers to be able to extract useful information from existing data to analyze the causes of traffic accidents, so that traffic administrations can be more accurately informed and better policies can be introduced.

Traffic conditions are a complex system due to many stochastic factors, and traffic accident data has long been known to be very difficult to process. Many attempts have been made in recent years through applying different methodologies and algorithms. Association rules has captured wide attentions and careful studies because of its adoptability and the nature of being easily understood, the focus of study of which is how to increase the accuracy and efficiency of the calculation. Among the researches to date, Geurts et al. used association rules to identify accident circumstances that frequently occur together at high frequency accident locations; Tesema et al. developed an adaptive regression trees to build a decision support system to handle road traffic accident analysis; Marukatat has made noticeable attempts at identifying the degree of importance of Information Entropy for road traffic accident analyses. Dong et al., Lee et al., Hassan and Tazaki, Zhang et al. and other researchers have achieved multileveled data mining of traffic accidents by means of a comprehensive application of data mining techniques.

The researches above all achieved the mining of accident data on a certain level; however, the overall calculating processes are largely too complicated and cannot be applied to all types of data, especially the multiattribute ones. On the other hand, the PSO algorithm has been applied in many fields. Shi and Eberhart studied the parameters' optimization,

based on particle swarm optimizer. Wang et al. propose an association rules algorithm based on particle swarm optimization algorithm to mining the transaction data in the stock market. Moreover, others improved and applied PSO algorithm to their purpose. So far, there have been a lot of researches targeting at different types of data, and due to the "capricious" nature of real-world data, coupled with the innate shortcomings of the algorithm, the association rules still falls short of people's expectations in being less complicated, less time and space-consuming, and more efficient.

Association rules is a data mining method for investigating the associative property of different events, which can be used in traffic accident data mining to mine the importance of attributes, that is, the associative relationship of events with certain types of accident. Its basic idea is to treat each characteristic as an item. Accident site, number of death, and so on can all be called an item. The higher the association, the more likely one event is directly linked to the cause of a certain type of accident. To decide how related two items are, we need to identify how many times some characteristics appear at the same time in a large number of similar events. If items show up at the same time frequently, indicating that there is a statistic pattern behind it, we can start to believe that the items are relevant.

Traffic accidents are events with strong randomness, while entropy is the mathematical method for analyzing an event's uncertainty. From the perspective of statistical mathematics, entropy is a measure of system randomness. Information Entropy is a value for characterizing the statistical characteristics of random variables. It is a measure of the average uncertainty of random variables, an objective description of statistical characteristics of the population.

II. METHODOLOGY

This proposed system is to consider processes complete chains of operations that are needed to produce certain products or services to make organizations perform better. Approaches are available that address the issue of how to actually design a processor, since in many contexts the processes are already in place, how to redesign one. This proposed ranker is very handy for the citizens to select the perfect products by viewing our ranking system. This application provides the feasible environment for the parents with improved options that they can easily enumerate into it. Proposed systems are feasible and updateable for new ones according to the product performance. Proper Styles and CSS applied to attract the viewers. Process based edit options are given for the admin of the system. This project has highly swift loading capacity and easy to use.

III. RESULTS AND DISCUSSION

The results of the experiments in which Naïve Bayes classifier achieved overall higher accuracy and robustness than Poisson regression support the effectiveness of applying Naïve Bayes classifier to the incremental estimation of project failure risk. Bayesian approaches including Naïve Bayes classifier are "incremental" estimators that can improve the accuracy by accumulating predictors as the project progresses. Furthermore, they make possible to treat missing data in a robust manner. On the other hand, regression approaches including the Poisson regression model are "global" estimators that often form monolithic models. They have difficulty in identifying a consistent set of predictors and handling missing data.

There are several directions for future research: First, it is necessary to compare the proposed method with more techniques by using different data sets. Second, in order to improve the performance, other strategies for selecting predictors should be explored, including further investigation of expected lift ratio. Finally, the identification and utilization of control factors that cooperate with prediction models will be essential to achieve early risk control of software projects.

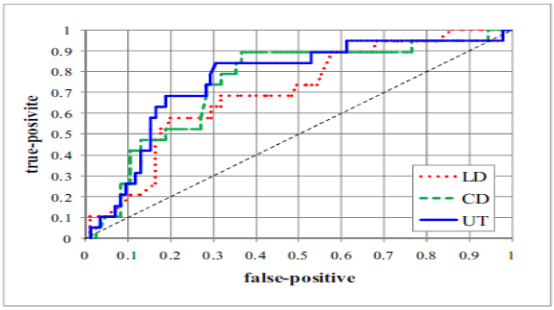

Fig 1: The ROC curves of LD, CD, and UT (Naïve Bayes)

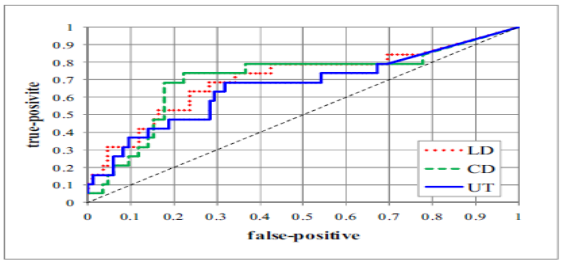

Fig 2: The ROC curves of LD, CD, and UT (Poisson regression)

IV. CONCLUSION AND FUTURE SCOPE

A thorough literature review revealed a gap in published studies on the relationship between road characteristics and RTA severity in Ethiopia. In this paper, we collected and

cleaned traffic accident data, attempted to construct novel attributes, and tested a number of predictive models. The outputs of the models were presented for analysis to domain experts for feedback. The RTA is eager to continue the study to identify areas of interest that should be given resources for traffic safety. Finally, knowledge was presented in the form of rules using the PART algorithm of WEKA. In contrast with the previously published work of the authors, which focused on driver characteristics, here we focused on the contribution that various road-related factors have on the accident severity. The results of this study could be used by the respective stakeholders to promote road safety. While the methods are simple, the results of this work could have tremendous impact on the well-being of Ethiopian civilians. The next step in the modeling will be to combine road-related factors with driver information for better predictions, and to find interactions between the different attributes. This also plans to develop a decision support tool for the Ethiopian Traffic Office.

REFERENCES

[1] Abdalla, I. M., R. Robert, et al. (1997). "An investigation intothe relationships between area social characteristics and road accident casualties."Accidents Analysis andPreventions 5: 583-593.

[2] Beshah, T. (2005). Application of data mining technology to support RTA severity analysis at Addis Ababa traffic office.Addis Ababa, Addis Ababa University.

[3] Beshah, T., A. Abraham, et al. (2005). "Rule Mining and Classification Of Road Traffic Accidents Using Adaptive Regression Trees." Journal Of Simulation 6(10-11).

[4] Chang, L. and W. Chen (2005). "Data mining of tree-based models to analyze freeway accident frequency." Journal of Safety Research 36: 365-375.

[5] Chang, L. and H. Wang (2006). "Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention " Accident analysis and prevention 38(5): 1019-1027.

[6] Chong, M., A. A., et al. (2005). "Traffic Accident Analysis Using Machine learning Paradigms." Informatica 29(1).Getnet, M. (2009).

[7] Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents: the case of Addis Ababa city. Addis Ababa Addis Ababa University.

[8] Mahdi PakdamanNaeini et.al Stock Market Value Prediction Using Neural Networks2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)

[9] Cheng-Tao Chu, Sang KyunKim,Yi-An Lin, Map-reduce for machine learning on multi core CS. Department, Standford University 353 Serra Mall, Standford University, Stanford CA 94305-9025

[10] Jimmy Lin and Michael Schatz Design Patterns for Efficient Graph Algorithms Map Reduce University of Maryland, College Park,MLG_10 Proceeding of the Eighth Workshop on Mining and Learning with Graph Pages 7885

[11] The general inefficiency of batch training of gradient descent learning, D. Randall , Volume 16, Issue 10, December 2003, Pages 1429– 1451,ACM Digital Library, Elsevier Science Ltd. Oxford, UK, UK

[12] MapReduce Implementation of the Genetic-Based ANN [9] Classifier for Diagnosing Students with Learning DisabilitiesTung-Kuang Wu1, et.al. 2013.

[13] Filtering: A Method for Solving Graph Problems in Map reduces, Silvio Lattanzi et.al Google Inc. 2011 ACM 978-1-4503-0743 Web References

[14] Yahoo hadoopTutorial ,

[15] https://developer .yahoo.com/hadoop/tutorial/

[16]GSOC proposal to implement neural network, https://issues.apache.org/jira/browse/MAHOUT-364.

