International Journal of Innovative Computing 7(1)15-18

UTM UNIVERSITI TEKNOLOGI MALAYSIA RESEARCH UNIVERSITY

International Journal of Innovative Computing

Journal Homepage: http://se.fsksm.utm.my/ijic/index.php/ijic

Road Traffic Crash Severity Classification Using Support Vector Machine

Isah Sani Birnin Gwari
Department of Computer Science
Faculty of Computing,
81310 UTM Skudai, Johor Bahru
abbantea@gmail.com

Nor Haizan Mohamed Radzi, Noorfa Haszlinna Mustaffa
Department of Computer Science
Faculty of Computing,
81310 UTM Skudai, Johor Bahru
haizan@utm.my, noorfa@utm.my

***Abstract*— Road traffic crash (RTC) is considered among the leading cause of death in many countries in the world and gives negative impact to the social and economic progress. In Nigeria, 13,583 RTC cases were reported in the year 2013 and this figure rising rapidly. Prediction on injuries severity and analysis on accident contributory factors is vital in order to improve either the road condition or the road safety regulation in attempt to reduce fatalities due to RTC. In this paper, a support vector machine model is developed to predict the road crash severity injuries using human, environment and vehicle contributory factors.**

Keywords — Classification, road crash, support vector machine, severity injuries

I. INTRODUCTION

Death and injuries due to the road traffic crash (RTC) has increased dramatically and posed threat to the community worldwide. The loss of a family member due to accident not only effected the emotion but also the income of the family. It is estimated around 3% of gross domestic income lost to the road traffic worldwide. In the low and middle income countries the loss of GDP is more higher which up to 5%. According to the World Health Organization (WHO) there is about 1.25 million people die each year as a result of road traffic crash and nearly 50 million people were injured and remained disabled for life due to RTC [2]. Majority of the fatalities on the road crash was reported occurred in the low and middle-income countries. Nigeria, as one of the developing countries, has the number of people killed in RTC 47 time higher than Britain [13]. The statistic of the RTC casualty in Nigeria was risen up rapidly between the years 2010-2013. Furthermore in 2013, Nigeria Federal Road Safety Corporation has recorded 13,583 cases RTC with 6,544 people were killed. This figure gives a strong alarm that an immediate action needs to be carried out. Furthermore, work on the RTC data set in Nigeria is lag behind compare to the developed countries.

Study related to the RTC has been carried out worldwide over decades ago in order to eliminate and reduce the fatalities of the road accident. Most of the works focus on classification and prediction of the RTC severity and the identification of the contributory factors of RTC. The identification of significant contributory factors of the RTC associated with the injury severity is a crucial task particularly to the local Road Safety Agency so that the effective preventive measure can be prioritized. Thus, the information obtained in the study can be used to enforce rules and regulation for reducing and prevention of the crash severity injuries. Generally in the RTC study, the contributory factors are fall into three categories: human, environment and vehicle. However, the identification and generalization of the causes of RTC globally is very difficult since the contributory factors of RTC are influenced by the local infrastructure, climate change and the road user attitudes and awareness towards road safety [3]. In fact, in Nigeria the cause of RTC is also mainly due to these three categories contributory factors. But, the attribute of each of these factors may different from other countries due to the different in the road infrastructure and road safety awareness among the people. Thus, this work attempts to investigate the contributory factors of the RTC severity in Nigeria. In this study, the road traffic crash data obtained from the Federal Nigeria Road Safety Corporation were analyzed to identify the most significant factors of crash severe injuries and predicts crash severe injury using Support Vector Machine learning algorithm.

This paper is organized as follows; section two highlights the previous works related to RTC classification analysis. Sections 3 discussed the method of proposed study and data set used in the research.

II. RELATED WORK

Investigation on why accident happens has been the central focused of the researchers over years ago. Many of them used data mining techniques to examine the most influence factor associate with the crash severity. Generally, the level of RTC severity is divided into no injury, injury and fatal. In [3] bayesian network, decision tree and Neural network classifiers are developed to investigate the relation of the road-related and vehicle factors with the crash severity. They have discovered the light conditions, vehicle maneuver and road type are the three most frequent factors influence the RTC. Association of the RTC severity injuries with the road related factors is also studied in [4]. In [6] the combination of the road related and other environment factors is used and fuzzy granular decision tree classifier is developed to predict RTC severity. However, the works conducted in [5] and [8] used all the three contributory factors which combined human, environment and vehicle factors to predict the RTC severity. Variable importance measure (VIM) and regression tree (CART) are applied to classify and identify the important factors affecting the three levels of crash severity injuries and revealed that over speeding, improper overtaking and not using the seat belt as the causes of crash severity[8]. Evolutionary fuzzy classifier is developed in [5] to study the Ethiopian RTC data between the years 1998 to 2000. In this work, the attributes are ranked and only 13 most important features out of 45 attributes are selected from the combination of three road traffic crash contributory factors.

III. METHOD

A. *Support Vector Machine*

Support Vector Machine (SVM) is a nonparametric machine learning algorithm proposed in 1990s by VN Vapnik. This technique firstly maps the input points into a high-dimensional feature space and then find a separating hyper-plane that maximizes the margin between two classes. It is based on optimization theory in order to minimize the classification error that may occur during training stage. The kernel functions are used in training stage for the classifier to select the support vectors with the surface of the function [9]. The data are classifies in SVM using support vectors that outline the hyper – plane in the feature space [10]. SVM is used to solved problems related to classification, prediction and learning.

The support vector machine learning algorithm use kernel function to decide the complexity of the classification function set. Giving training data i = 1, 2…, N where X is the vector of input pattern and Y is the corresponding target output. The SVM decision function learning algorithm is based on

$F(X) = \sum_{i=1}^{n} Y_I \alpha_i K(X_i, X) + b$. type using math equation

Where $K(X_i, X)$ is the kernel function mapping the input vectors into a space, m is the number of input and b is the bias term.

B. *Data Preparation*

The historical crash data set used in this study is obtained from the Federal Road Safety Corporation of Nigeria Birnin Yero Unit Command (an agency that regulates safety, rescue and road crash activities). The data set contains 712 crash cases that cover the period from 2013 to 2015. Initially there are 36 causes of accident reported in the RTC cases that comes from human, environment and vehicle factors. However, 71 crash cases and 11 attributes are removed because of incomplete information, wrongly spell field, missing field and target class cannot be identified in the crash. Thus, a total of 641 crash data and 25 attribute of accident causes were remained for the experiment. The attributes comprise of 11 from human, six from environment and seven from vehicle contributory factors. The RTC severity is divided into 3 classes: Fatal, Serious and Minor. The attributes description is shown in Table 1, 2, and 3 below.

TABLE 1. HUMAN CONTRIBUTORY FACTORS

Attribute Name	**Description**
Response Time	The exact time the rescue teams attend to the crash victim.
Over Speeding	The crash was occur due to the over speeding of the driver
Driving under Alcohol/Drug	The driver was under influences of alcohol/drug while driving which lead to the road crash.
Dangerous Driving	The road crash occurs as a result of dangerous driving of the vehicle by the driver.
Lost of Control	The driver lost control of the vehicle on the road, road crash occurs by hitting another vehicle or the vehicle move out of road.
Overloading	The crash was occurs as a result of excess overloading of the vehicle with either passenger or load by the driver
Route Violation	The road crash occurs as a result of the driver violate route on the highway.
Dangerous Overtaking	The crash occurs during dangerous overtaking by the driver
Sleeping on steering	The driver is on sleeping when the crash occurs.
Use of phone while driving	The road crash occurs as a result of the driving is using phone while driving that is his concentration is not on the road.
Wrongful overtaking	The crash happens when the driver wrongly overtakes in a corner, sharp bed without seeing his front.

TABLE 2. ENVIRONMENT CONTIRBUTORY FACTORS

Attribute Name	Description
Crash Date	The date that the crash occur
Crash Time	An approximate time of the crash
Location	The approximate location of the crash in the highway
Bad Road	The crash occur as a result of bad road or a black spot, pothole, sharp bed etc.
Road Obstruction	The road crash happens as a result of obstructions on the highway
Poor weather	The crash occurs as a result of poor weather in the area like during rainy, haze, etc.

TABLE 3. VEHICLE CONTIRBUTORY FACTORS

Attribute Name	Description
Vehicle Type	The type of vehicle involved in the crash e.g. car, bus, lorry
Vehicle Category	Categories of the vehicle involved in the crash e.g. Honda, Toyota, etc
Vehicle Make	Car maker e.g. Honda, Toyota, etc
Brake failure	Crash occurs as the result of failure of the brake
Mechanical Deficiency	The crash happen as the result of the vehicle has mechanical deficiency
Tire Burst	Crash happen as a result of flat tire

The input attributes are represented into numerical format and normalized ranging from 0 to 1 using for Min- Max formula.

$$X_{i,\,0\text{ to }1} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$$

The data set is divided into two set, one for training and another set for testing data using 80/20 rule. The training data set contain 80% and the testing data contain the remaining 20% of the data set. The 10 – fold cross validation techniques is used to validate the road traffic crash classifier in the SVM. In this cross-validation, the data set was split into 10 parts; 9 were used for training and the remaining 1 was used for testing. This process is repeated 10 times with a different 10th of the dataset used to test the remaining 9 parts during every run of the 10-fold cross validation.

IV. EXPERIMENT AND RESULT

SVM classification algorithm mainly used four basic kernel functions which include linear, polynomial, radial basis and sigmoid. Each of this kernel function has parameters, C and gamma. The Radial Basis Function (RBF) was used in this study because it can take care of cases when the relation between class labels and attributes in nonlinear with the two parameters cost C and gamma ɣ [11]. The optimal and best values of C and ɣ are found using grid search algorithm in WEKA tools with 10 – fold cross validation techniques.

After finding the best pairs parameters for the LIBSVM classifies a series of experiment were performed using the training data set to generate a best classification performance the result is evaluate with the testing data set.

The general prediction performance accuracy is 73% for training data set and 74% for testing data set. In the training dataset the model correctly classified is 36%, 84% and 61% for minor, serious and fatal respectively while in testing dataset 92% correctly classified as serious and 47% fatal are correctly classified and only 1.7% correctly classified as minor class.

TABLE 4. CLASSIFICATION RESULT

Categories	Training Data		Testing Data	
	Correct classified	Incorrect classified	Correct classified	Incorrect classified
Minor	16	28	2	9
Serious	279	52	77	6
Fatal	85	53	16	18
	380	133	93	35

The significant of SVM classifier performance in this paper was demonstrated using the commonly used measures in the literatures [3, 5, 12] of accuracy, precision, recall and F-measure. The general accuracy of the model classifier is 82%, serious class in the model has the highest values of precision, recall, f-measure; 77%, 86% and 81% respectively, while minor class accuracy outperform the other classes with 92%.

TABLE 5: SUMMARY OF PERFORMANCE MEASURE

	Precision	Recall	F – Measure	Accuracy
Fatal	65%	56%	60%	80%
Serious	77%	86%	81%	75%
Minor	56%	33%	41%	92%

IV. DISCUSSION

The main significance of this research is the development of a model for classification of the road traffic crash severity injuries in Nigeria. This model would provides the insights of the road traffic crash causes. This work has combined three main contributory factors identified in the data set, including human, environment and vehicle in the model development. Combination of these factors however, has resulted more attributes or features, as in this case is 23 features. In order to obtain better understanding on the causes of the accident, feature selection is suggested in the future. Feature selection will identify the most significant causes of the accident. By knowing the most significant causes of accident, the local authority and road traffic agencies could able to review and improve the current road traffic regulation or to impose a new one if necessary. Furthermore, as found in many literatures, selection of the significant features possible to improve the performance of classifier as there is a possibility of the existence of irrelevant and redundant features in the data set.

REFERENCES

[1] M.A.Sheriff, M., "Traffic Education and Safety in Nigeria, " Nitours Journal, Vol. II, 2009.

[2] W.H. Organization, Global status report on road safety 2013. WHO, 2013. 2015, Geneva.

[3] Y. Castro, and Y.J. Kim, "Data mining on road safety: factor assessment on vehicle accidents using classification models," International Journal of Crashworthiness, 21(2): p. 104- 111, 2016.

[4] T. Beshah, and S. Hill. "Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia," in AAAI Spring Symposium: Artificial Intelligence for Development. 2010.

[5] T.Beshah, et al. "Learning the classification of traffic accident types, in Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on IEEE. 2012.

[6] H.K. Moghaddam, and X. Wang. "Vehicle Accident Severity Rules Mining Using Fuzzy Granular Decision Tree," in International Conference on Rough Sets and Current Trends in Computing. 2014. Springer.

[7] V. Rovšek, M. Batista, and B. Bogunović, " Identifying the key risk factors of traffic accident injury severity on Slovenian roads using a non-parametric classification tree," Transport (2014) p. 1-10, 2014

[8] A. Tavakoli Kashani, A. Shariat-Mohaymany, and A. Ranjbari," A data mining approach to identify key factors of traffic injury severity," PROMET-Traffic&Transportation, 23(1), p.11-17, 2011.

[9] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: a review" Artificial Intelligence Review, 34(4): p. 369-387, 2010

[10] S.A. Mulay, P. Devale, and G. Garje, "Intrusion detection system using support vector machine and decision tree" International Journal of Computer Applications, 3(3): p. 40-43, 2010.

[11] C.W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification" , 2011

[12] T. Beshah, et al., "Mining Pattern from Road Accident Data: Role of Road User's Behaviour and Implications for improving road safety", International journal of tomography and simulation, 22(1): p. 73-86, 2013

[13] A. O. Atubi, and P.C. Onokala, "Contemporary Analysis of Variability in Road Traffic Accidents in Lagos State, Nigeria,"African Geographical Review, 28(1), 11-41. 2009