IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Support Vector Machine with Principle Component Analysis for Road Traffic Crash Severity Classification

To cite this article: N.H.M. Radzi *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012068

Support Vector Machine with Principle Component Analysis for Road Traffic Crash Severity Classification

N.H.M. Radzi[2], I.S.B. Gwari[2,] N.H. Mustaffa[3], R. Sallehuddin[4]

[1,3,4]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia.
[2]Department of Mathematical Sciences, Kaduna State University, Kaduna. Nigeria.

Email : haizan@utm.my

Abstract- Road traffic crash (RTC) is one among the leading causes of death in the world, including Nigeria. It also turns many victims completely disabled and generally affected the socio–economic development in the society. In this paper, we proposed to predict the road crash severity injuries in Nigeria by identifying the most significant contributory factors using Principal Component Analysis with Support Vector Machine (SVM) used for classification algorithm. Road crash data from year 2013-2015 obtained from Federal Road Safety Corps Nigeria is used in this study. The result shows that and increased to 87% compared to 82% without feature selection.

1. Introduction

Road Traffic Crash (RTC) is ranked as the ninth cause of death in the world [2]. RTC also gives significant impact to the socio-economic to the victim or the family as it incurs additional overhead cost for medical expenses and care taker. RTC is also contributing up to 3% lost in gross domestic product (GDP) in the country [6]. In Nigeria, there were 13,583 cases of RTC in 2013 with 6,544 people killed recorded by the Federal Road Safety of Nigeria [13]. Therefore, identification of significant contributory factors of RTC and the causes of each crash severe injury (fatal, serious and minor) is necessary to provide bases for rules and regulation enforcement [1].

The range of factors that contribute to RTC severity is very wide and different from one country to another. This is due to the different behaviors of the road safety awareness and also road infrastructure. Basically, the causes of RTC can be divided into three categories: human, environment and vehicle[3]. Further, each of these categories consists of several other factors. Therefore, identification and generalization of causes of RTC globally is very difficult [3]. In this study, we attempt to use principle component analysis (PCA)to identity the most significant factors and support vector machine (SVM) to classify the RTC severity in Nigeria.

2. Related Works

A considerable amount of studies on RTC severity classification has been conducted by many of researchers including in [3-8, 10]. Most of researchers used single contributory factor such as either features in human, environmental or vehicle. But, there are some researchers combine the features in environment and vehicle categories to classify the road crash and only a few [1,5, 8] combined the features in all three categories. Combination of all categories, however, may affect the accuracy of the classifier, as it will increase the dimension of the input. Therefore, researchers used various method to assess significant features contribute to RTC. In work [4], PART algorithm is used to identify the crash severity injuries significant features. Fuzzy granular decision tree was used in model proposed in [6]to classify and predict road crash into property damage, injury and fatal classes. The model in [6] used 7features from the environment category which are obtained from 1837 crash cases collected from 13 regions in California. Work in [7] classify the RTC into two classes: injury and fatal using crash data from Slovenia between year 2004 to 2009 with 9 features and used variables importance measure (VIM)to predict the most contributory factors to RTC. Work in [8] also used (VIM) in classification and regression tree (CART). Besides this, another method called Principle component analysis (PCA) is also promising for feature selection. Fengxi[11] used PCA to select significant features in face recognition. PCA also used in [12] and increase the accuracy in the classification algorithm. Therefore, this work attempt to use PCA to select most significant features contribute to RTC severity.

3. Methods

3.1. Principal Component Analysis

PCA learning algorithm will compute the covariance matrix and find all the eigenvalues and eigenvectors. The eigenvectors will be selected corresponded with the first m large eigenvalues as eigenvectors $= \{V_1 \cdots V_m\}$. Calculate the feature extraction result using the j^{th} feature component, $C_j = \sum_{p=1}^{m} \left|V_{pj}\right|$, where $v = \{1,2 \cdots, N\}$, $p = \{1,2, \cdots m\}$, $|V_{pj}|$ is the absolute value of V_{pj}. The feature selection result will be the d_{1th}, d_{2th}, …. d_{nth} of the feature component[11]. The used of PCA feature selection will select the most important variables from all the large variables in the dataset.

3.2 Support Vector Machine

Support Vector Machine (SVM) is a non parametric machine learning algorithm proposed in 1990s by VN Vapnik. The input points are firstly maps into a high-dimensional feature space and then find a separating hyper-plane that maximizes the margin between two classes. The kernel functions are used in the training stage to select the support vectors with the surface of the function [9]. Given training data set (x_i,y_i), where x_i is a real valued n-dimension input vector and y_i is the corresponding target output. The SVM decision function learning algorithm is based on the following formula,

$$F(X) = \sum_{i=1}^{n} Y_I \alpha_i K(X_i, X) + b, \tag{1}$$

where ,

$K(X_i, X)$ - kernel function
n - number of input
b - bias

4. Data Description

This study used 641 crash data recorded from year 2013 to 2015. It was obtained from Federal Road Safety Corps of Nigeria Birnin Yero Unit Command (an agency that regulate safety, rescue and road crash activities) [1]. The data contains11 attributes from human, 6 from environment and 7 from vehicle categories. The target class defined in this study are:fatal, serious and minor[1]. Table 1 shows the description of each attribute. The data is divided into training and testing where 80% used for training and the remaining 20% for testing. The experiment is conducted using 10 – fold cross validation techniques In eachfold, the data set was split into 10 parts; 9 were used for training and 1 part for testing.

5. Experiments and Result

The PCA features selection algorithm in WEKA tools was used to extract the most relevant features in the classification of road traffic crash from the 24 attributes as describe in table 1. PCA firstly calculate the covariance matrix from the original data to find the eigenvalues and eigenvectors of each attributes and ordered them according to the eigenvalues. The eigenvector with highest eigenvalue will be the principal component of the new data set. PCA features selection generated 22 new attributes as in table 2 and ranked according to the eigenvalues.

These new22 attribute are then used as input features for SVM classifier. The RBF kernel functionis selected for training function in SVM with optimal values of gamma, ɣ and cost, C parameters obtained from experiment conducted in the [1]. We then run the experiment with various numbers of features. Table 3 shows the classifier accuracy with new attributes obtained from PCA with different number of features. The accuracy of the classifier is at the highest, which is 80%, whenever 15 features is chosen as the attributes.

Table 1: RTC contributory factors attributes

Attribute Name	Contributory Factors	Description
Crash Date	Environmental	The date that the crash occur
Crash Time	Environmental	An approximate time of the crash
Response Time	Human	The exact time the rescue teams attend to the crash victim.
Location	Environmental	The approximate location of the crash in the highway
Vehicle Type	Vehicle	The types of vehicle involved e.g. car, bus, lorry, etc.
Vehicle Category	Vehicle	categories of the vehicles e.g. Private, Commercial etc.
Vehicle Make	Vehicle	The vehicle make e.g. Honda, Toyota, etc.
Over Speeding	Human	The driver is over speeding
Under Alcohol/Drug	Human	The driver was under influences of alcohol/drug while driving
Dangerous Driving	Human	Dangerous driving of the vehicle by the driver.
Lost of Control	Human	The driver lost control, road crash occurs by hitting another vehicle or the vehicle move out of road.
Bad Road	Environmental	Bad road condition or a black spot, pothole, sharp bed etc.
Road Obstruction	Environmental	Obstructions on the highway
Poor weather	Environmental	Poor weather condition in the area like during rainy, haze, etc.
Sleeping on steering	Human	The driver is sleeping while driving.
Use phone	Human	Driver using phone while driving
Wrongful overtaking	Human	Overtakes in a corner, sharp bed without seeing his front.
Brake failure	Vehicle	Failure of the brake in the vehicle.
Mechanical deficiency	Vehicle	Vehicle has mechanical deficiency.
Overloading	Human	Overloading with either passenger or load of the vehicle
Sign light violation	Vehicle	Light sign in the vehicle is not proper working.
Route Violation	Human	Driver violate route on the highway.
Dangerous Overtaking	Human	Dangerous overtaking by the driver
Tire burst	Vehicle	Flat tires in the vehicle.

Table 2. New attributes

Sn	Eigenvalue	Proportion	Cumulative	New Attributes
1	1.57728	0.06572	0.06572	-0.578VEHICLE MAKE...
2	1.39531	0.05814	0.12386	0.509RESPONSE TIME ...
3	1.3049	0.05437	0.17823	-0.471SPEED VIOLATION...
4	1.22021	0.05084	0.22907	-0.527VEHICLE CATEGORY...
5	1.19383	0.04974	0.27881	0.42DANGEROUS DRIVING...
6	1.15863	0.04828	0.32709	-0.393TYRE BURST...
7	1.1287	0.04703	0.37412	-0.502LOST OF CONTROL...
8	1.08209	0.04509	0.41921	0.638ROUTE VIOLATION...
9	1.07585	0.04483	0.46403	-0.63WRONFUL OVERTAKING...
10	1.06432	0.04435	0.50838	-0.679MECHANICALLY DEFICIENT VEHICLE...
11	1.04029	0.04335	0.55172	0.486USE OF PHONE ON DRIVING...
12	1.03163	0.04298	0.59471	-0.409SIGN LIGHT VIOLATION...
13	1.02432	0.04268	0.63739	0.602 VEHICLE TYPE...
14	1.01856	0.04244	0.67983	-0.452 CRASH TIME...
15	1.01595	0.04233	0.72216	-0.595BAD ROAD....
16	1.01286	0.0422	0.76436	0.522POOR WEATHER...
17	0.95949	0.03998	0.80434	-0.527DRUNG ALCOHOL DRIVING...
18	0.92037	0.03835	0.84269	0.46 BRAKE FAILURE...
19	0.89747	0.03739	0.88008	-0.458VEHICLE CATEGORY...
20	0.86295	0.03596	0.91604	-0.57CRASH DATE...
21	0.7927	0.03303	0.94907	-0.59 DRUNG ALCOHOL DRIVING...
22	0.70243	0.02927	0.97834	-0.633LOCATION...

Table 3. New Attributes Accuracy

Number of Features	Accuracy
22	68%
20	69%
18	69%
16	71%
15	80%
14	69%
13	69%

We then used these 15 features for our PCA-SVM model and the result obtained as in Table 4. In the training phase the PCA-SVM has correctly classified 85% serious, 72% fatal and 36% minor while in testing dataset 96% serious, 85% fatal and 54% minor are correctly classified.The general performance prediction classification accuracy of the proposed build model is 77% for training set and 89% for testing.

Table 4. Classification Prediction Result

CLASS	Training Data		Testing Data	
	Correct Classified	Incorrectly Classified	Correctly Classified	Incorrectly Classified
MINOR	16	28	6	5
SERIOUS	284	47	80	3
FATAL	100	38	29	5
	400	113	115	13

The significant of SVM classifier performance in this paper was demonstrated using the commonly used measures in the literatures [1, 3, 5, 12] of accuracy, precision, recall and F-measure and compare with [1]. 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 experiment performance

	LIBSVM				PCA-LIBSVM			
	Precision	Recall	F – Measure	Accuracy	Precision	Recall	F – Measure	Accuracy
MINOR	65%	56%	60%	80%	78%	75%	76%	88%
SERIOUS	77%	86%	81%	75%	83%	88%	85%	81%
FATAL	56%	33%	41%	92%	55%	40%	46%	92%

Table 5 illustrates the comparative result of the proposed PCA model with the standard LIBSVM model with the 24 input attributes of road crash contributory factors it shows clearly that the proposed PCA –SVM with 15 most significant factors of road crash performed better in precision, recall and *f* measure for both fatal, serious and minor crash class. As early mention the accuracy rate of the proposed model presented are highly recommended for classification and prediction of road traffic crash severe injuries using principal component analysis PCA feature selection method to identify the significant contributory factors.

6. Conclusion

This research develops a PCA-SVM model that would provide an insights and helping road safety agency in Nigeria to predict and classify the RTC severity injuries. It also identified the most road contributory factors to help the policy makers in constructing a good roads with proper road sign and adequate regulation of highway rules to the road users for better traffic safety control policies.

After a thoroughly review of many literature in road traffic crash classification analysis a research gap was identified of using three main contributory factors of road crash data for severe injuries classification and prediction in Nigeria. In this study we collected a raw data from Federal Road Safety Corps Nigeria to clean, preprocess and construct a predictive model for crash severe injuries identification.

References

[1] Radzi N H M *et al* 2017 *International Journal of Innovative Computing, 7*(1) 15-18
[2] Organization, W.H. 2015 Global Status Report on Road Safety 2015 retrived from https://www.who.int/violence_injury_prevention/road_safety_status/2015/en/.
[3] Castro Y *et al* 2016 International Journal of Crash worthiness **21**(2) 104-111.
[4] Beshah T *et al* 2010 *AAAI Spring Symposium 2010: Artificial Intelligence for Development*
[5] Beshah T *et al* 2012 International Conference on Intelligent Networking and Collaborative Systems (INCoS)
[6] Moghaddam H K *et al* 2014 The 9th International Conference on Rough Sets and Current Trends in Computing Springer 280-287
[7] Rovšek V M 2017 Transport **32**(3) 272-281.
[8] Tavakoli K *et al* 2011 PROMET-Traffic&Transportation **23**(1) 11-17.
[9] Hsu C W *et al*, *Tech Rep* Department of Computer Science, National Taiwan University
[10] Beshah T *et al* 2013 *Int Journal of Tomography and Simulation* **22**(1) 73-86.
[11] Song F *et al* 2010 International Conference on System Science, Engineering Design and Manufacturing Information (ICSEM)
[12] Xu Y *et al* 2010 *Pattern recognition* **43**(3) 1106-1115
[13] Sheriff M 2009 Nitours Journal **2**.