Use of Data Mining Technique for Systematic Road Safety Audit of Non-urban Highways

Bincy B.J, Anitha Jacob

Abstract— In India the number of road crashes is raising at frightening rate. There is one death in every four minutes due to road crashes in India. Hence it is necessary to improve the road safety by conducting a detailed Road Safety Audit (RSA) in order to identify road safety issues and to make necessary improvements. Budgetary constraints limit many developing countries from performing the audit on regular basis. This will eventually delay any rehabilitation or repair process making the road conditions the worst and risky. This paper proposes a systematic approach to do the road safety audit on a highway and to do effective and efficient data mining, for deriving knowledge driven decisions in the classification of highway sections. The approach will help to perform safety evaluation of sections and to identify the crash potential locations. Further output of the work is the development of a mathematical model for classification of highway sections based on road safety audit.

Index Terms— Data mining, Weka, Road Safety Audit, Road safety model

I. INTRODUCTION

Road Safety Audit (RSA) is method of evaluating the safety performance of a road by an independent team or trained specialties. Qualitative estimation on potential road safety issues, identification of opportunities for improvements and ensure safety for all road users are the basic objectives of RSA. Government incorporate RSAs into the initial stage of the project such as construction of new roads and intersections, and also encourages RSAs on existing roads and intersections. Thereby all the new and reconstructed roads can be made safe as possible. Since RSA is done based on a clearly defined procedure it can be used at any progressing stages of project. The principles of RSA can be applied throughout the highway project development in order to ensure a growing awareness about road safety principles. Traffic control devices provide safe and secure journey for the road users. These are devices used to inform, guide and control the traffic. Maintenance of traffic control devices is one of the most important aspects of highway management systems. Scientific and well-timed installation of the traffic control devices increases safety as well as significantly decreasing accident rates. For a safer driving environment, timely maintenance of the traffic signs is very important and incorporating these activities makes a viable economic sense. The specialized units of highway authorities that is the Road Safety Authority (RSA) frequently check the safety requirements of the traffic control devices.

Road Safety Audit can minimize the risk and severity of road accidents by the road project and also can minimize the need of remedial work after construction. Road safety inspection based on (IRC SP – 88, 2010) [9] is conducted on an existing road, from Kunnamkulam to Peramangalam, since it is identified as one of the accident black spot.

II. 11. LITERATURE REVIEW

This section, describe a short survey about Road Safety Audit, and classification of road by means of Weka software and analysis

A. Road Safety Audit (RSA)

Road safety audit is done to ensure the operational safety performance of a road. Hence, it has the potential for improving safety when it is applied to a road or traffic design before the project is implemented. Through RSA the identification of potential safety hazards on new road projects at the appropriate stage can be done and so that it can minimize the adverse effects at minimum cost. It can be conducted on any design proposal, which involves changes to the ways road users will interact, either with each other or with their physical environment. Purpose of the audit is to identify hazardous features on existing road so that it can be eliminated or otherwise treated before they become an accident prone location. Mehar and Agarwal [2] presented a systematic approach to improve the road safety by analyzing the crash records thus by identifying the most hazardous locations of the study area. They also describes the difficulties faced in the safety audits and analysis of accidents. They developed a hierarchical frame work for the improvement of road safety. Based on the crash rate safety hazardous locations were identified and ranking is done. They also suggest suitable remedial measures to improve the hazardous locations. Saffarzadeh and Farshad [3] studied the maintenance of highway traffic control device and the problems faced, they found that in many developing countries, due to budget limits and lack of regular maintenance activities and many deficiencies related to signs, guardrails road markings increased. They also tried to propose an appropriate management system for the maintenance of traffic control devices, along with the development of computer software for control devices, which can identifying the necessary time for maintenance.

Manuscript received June 23 2018.

Bincy B.J is with Department of Civil Engineering, Jyothi Engineering College, Cheruthuruthy, Thrissur, Kerala 679531, India (e-mail: bincybj93@gmail.com).

Dr. Anitha Jacob is with Department of Civil Engineering, Jyothi Engineering College, Cheruthuruthy, Thrissur, Kerala 679531, India (e-mail: anithajacob@jecc.ac.i

With those mathematical models it can increasingly lead to the efficiency of operational systems for the maintenance of traffic control devices with in the limitation of the allocated budgets. These models are essentially flexible like any other mathematical model, regarding the alternatives. Farzaneh Moradkhani, Somayya Ebrahimkhani [4] incorporated data mining technique to identify the factors behind road accidents. They collected crash data, human behavior and geometric details of the study area. In order to conduct data mining the variables they collected are converted to numerical forms. They have done the analysis using Weka software. Based on this suitable traffic safety polices can be implemented on corresponding accident and situations. Arun S Bagi, Dheeraj N Kumar [5] conducted studies in (SH 87) relating to various aspects of road safety to identify various factors causing crashes. The crash data is collected from the F.I.R and the trend of crashes on each Kilometer wise is analyzed and also conducted pedestrian safety analysis. A detailed study of accident data, traffic and roadway geometric is done, and the statistical relationship between these parameters were studied. From this analysis the major causative factors leading to crashes can be find out. It will help to identify some of the causative factors responsible for accidents and based on that relative importance can be given to each parameters and suitable remedial measures can be provide accordingly. Griselda Lopez and Jaun de Ona [6] conducted analysis on the deficiency of traffic control devices in relation with crashes in rural highways. They conducted studies related to geometric and environmental road characteristics and found that the road which is improperly or poorly signaled can lead to incorrect placement or maneuvers of vehicles and ambiguous situations that can increase the risk of crashes. They analyze the relationship between road crashes in two-lane rural highways and certain deficiencies in signaling. They also considered increase in driver work load that may cause error in perception, or judgment stages may lead to driver to loss of control. They had done the analysis using CART method, to develop decision trees based on that road sections were classified.

Majid Khalilikhah, Kevin Heaslip [8] studied the effects of damage on sign visibility. Traffic signs are provided to convey information to the drivers. So it is necessary to ensure visibility in night time or low light conditions, traffic signs must be in compliance with the minimum retro reflectivity standards outlined by the road safety manual .They studied about the aging of sign and also other contributing factors affect sign retro reflective performance and are determined the effects of various damage forms on sign retro reflectivity, through statistical methods, including regression models, chi-square test, t-test, and odds ratio were employed to analyze traffic sign data.

III. Study Area

According to the Black spot analysis, prioritized based on MORTH, Thrissur district consist of 21 Black Spots. Peramangalam is one of the identified Black Spot in them. Recent accident data shows that occurrence of accident in Kunnamkulam is very high

Hence study area is selected from Kunnamkulam to Peramangalam. The selected study area is of 15 Km road stretch.

IV. METHODOLOGY

The initial stage of data mining process is to select the suitable parameter from the available data source and to define the data appropriately. Data sources can vary based on the purpose of analysis. The data was collected from two sources

A. Road inventory data

The selected study area is a 15 Km road stretch. The study area is divided into 200m sections for the convenience of study.

B. Traffic collision data

Past five year (2013-2017) accident data of Kunnamkulam and Peramangalam police station is collected from the FIR, and it includes information such as Location, Time Driver, Passenger, Pedestrian data related to (age, sex, alcohol influence)and also severity of crashes (Fatal crashes, Grievous injury crashes and Minor injury crashes)

The data collected are in both quantitative and qualitative type. Quantitative data is measurable form while qualitative data can only be classified also known as categorical data.

V. DATA PRE-PROCESSING

Data pre-processing involves converting the collected data into an understandable form. The collected data is sometimes insufficient, unpredictable and is likely to contain human errors. Data pre-processing is the suitable for resolving such errors. Data pre-processing prepares the unprocessed data for further processing.

A. Data Field Selection

Data gathered from various sources were combined, classified and examined. In which the data which are not relevant for the data mining technique was ignored.

B. Data classification

Data classification process can improve the quality of the data collected by correcting the errors detected. Generally data classification reduces errors and improves the data quality. Data classification can correct the errors such as the entries which are clearly invalid occurred due to human error and error on the problem reporting system can be identified and those which are correctable were corrected.

C. Data Transformation

Data transformation is the process of transforming the raw data into suitable forms which are able for data mining. The data set used in the study contained integer values for the entire attributes and are given in Table 1.The table shows the classification values for the corresponding attributes.

VI. DATA MINING

Data mining, is an emerging technique for data analysis which is useful for collecting, storing data. With the increased volume of database it is very difficult to understand the data without any powerful tool. It can also be used to identify the unknown features, from that suitable decision can be made from a large data. Through data mining the collected data can be organized, evaluated and can identify patterns for predicting the future behavior.

TABLE 1. CLASSIFICATION TABLE

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Classifiers	No: of correctly classified instances	Accuracy in %	Error rate
Classification tree(J48)	62	93.93	0.06
Navie Bayes	64	96.96	0.26
SMO	66	100	0.27
Multi-layer perceptron	66	100	0.02
Simple CART	62	93.93	0.16

VII. WEKA ANALYSIS

Weka (Waikato Environment for Knowledge Analysis) is software developed by the University of Waikato, New Zealand suitable for machine learning. Weka is freely available software and is licensed by GNU General Public License. It consists of algorithms which are useful for solving real-world data mining problems.

A. Experimentation

To predict the classes of road in the study area various classification models were developed using Weka software. In Weka, classification as well as regression can be carried out from that the relationship between various parameters can be found out. WEKA software consists of various classification tools. J48 decision tree algorithm, Simple CART, Navie Bayes, SMO, Multi-layer perceptron were used to perform the classification analysis.

J48. It is the implementation of ID3 algorithm developed by Weka project team and it is used to generate univariate decision trees. [8]

Naive Bayes. It is a classifiers which classifies based on applying Bayes' theorem with strong (naive) independence assumptions between the features.[8]

SMO (Sequentially minimal optimization). It implements sequential minimal algorithm for training a support vector classifier. It replaces all missing values and transforms all nominal attribute into binary one. [9]

Multi-layer perceptron. It develops a network system, and it can be used for difficult to complex problems. They are very good for approximation.[9]

Simple CART. This decision tree is normally applicable in data mining to produce a frame work.[3]

B. Results

For the analysis there are 66 sections of road. All the five classifiers performed for correctly classifying the instances. Accuracy measure represents how far the set of data are being classified correctly. The error rates and accuracies of each classifier are analyzed and are shown in Table 2.Among them Multi layer perceptron classifier sounds better with 0.0268 error rate and 100% accuracy even though

TABLE 2: ACCURACY MEASUREMENT

Attribute	Classification values
Width of road	1:Proper 2:Not sufficient
Shoulder width	1:Proper 2:Not sufficient
Terrain	1:Flat 2:Rolling
Veh/3 min	1: Vehicle count with in limit 2:Exceeds limit
Passing Zone width	1:Sufficient 2:Not sufficient
Warning Sign	1:Properly provided 2:Provided -not clear 3:Not provided
Regulatory Sign	1:Properly provided 2:Provided -not clear 3:Not provided
Informatory Sign	1:Properly provided 2:Provided -not clear 3:Not provided
Pavement Markings	1:Properly provided 2:Partiallyremoved 3:Completly removed
Position of sign board	1:proper sight distance 2:Less sight distance
Post Mounted Delineators	1:Properly provided 2:Very less in number 3:Not provided
Crash Severity	0:No crash 1:Minor injury 2:Major injury 3:Fatal

SMO classifier also give 100% accuracy but the error rate is slightly greater than Multi-layer perceptron. Hence, out of all classifiers Multi layer perceptron suits best for predicting classes of roads.

Accuracy of classification process is can further improved by means of the receiver operating characteristics (ROC) curve. ROC curve is a plot between the fraction of (TPR) true positive rate versus the fraction of false positives FPR (false positive rate).

Here in all cases, the AUCs values obtained were approximating to unity, with the Multi-layer perceptron model showing an AUCs equals to 1. From these results it shows that the Multilayer perceptron suits best for

predicting classes of roads. After the prediction of type of road the top correlated independent parameters which affects the classification of the road has to be find out.

C. *Correlation*

Correlation analysis is done with two third of the data in order to identify top correlated attributes. Correlation is done based on the assumption that data must be ordinal and the score of one variable must be monotonically related to other variable. The relevant features can be obtained by conducting correlation analysis i.e., the (variables, predictors) they are the independent and dependent variables for the model construction. The top 5 attribute obtained are:

- Regulatory sign
- Pavement markings
- Position of sign board
- Post mounted delineators
- Crash

VIII. DATA MODELLING

Through modeling it can formulate models that explain the general patterns in the data. Data Modeling is done to construct mathematical equations by combining variables that reliably predicts a desired outcome. Modeling is done by using SPSS software.

A. *Model calibration*

Model calibration is the reverse process to regression. Two third of the data is taken for model calibration. From this the relationship between a dependent variable and an independent variables can be found out.

B. *Coefficient of Determination (R^2)*

It represents the variation of dependent variable explained by the independent variable included in the model. A value of R^2 is between 0 and 1. The coefficient of determination, R^2, is used to analyze how differences in one variable can be explained by a difference in a second variable.

Table 3 shows the summary of R^2 values. Higher the value of R^2, the model will be better. The obtained value of R^2 is 0.88 it means 88 % of the points should fall within the regression.

TABLE 3: SUMMARY OF R SQUARE

Model	R	R^2	Adjusted R^2	Std. Error of estimate
1	0.94	0.88	0.86	0.26

C. *Regression analysis*

Linear regression analysis is done for modeling a relationship between a dependent variable with one or more independent variables. In linear regression, the relationships are modeled using linear predictor functions whose values are unknown model parameters are estimated from the data.

D. *Model validity*

Remaining one third data is checked for validity. Then the average root mean square error is calculated. Model with minimum root mean square is selected. Regression validation is done to identify whether the numerical results satisfying the obtained relationships from regression analysis, and are acceptable as descriptions of the data. In this case, dependent variable is the class of road and independent variables are Regulatory sign, pavement markings, position of sign board, post mounted delineators, accident. The coefficients of these dependant variable obtained by conducting regression analysis is shown in table 4

TABLE 4: REGRESSION COEFFICIENTS

Model	Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	T	Sig.
Constant	-0.39	0.160		-2.4	0.02
Regulatory sign	0.15	0.049	0.197	3.24	0.02
Pavement Markings	0.35	0.067	0.401	5.30	0.00
Position of sign board	0.40	0.199	0.242	3.38	0.02
Post mounted delineators	0.13	0.055	0.151	2.43	0.02
Crashes	0.257	0.049	0.365	5.50	0.00

By substituting the corresponding values of these road sections into the road safety model the rank of the road will obtain as:

Class of road = -0.394+ 0.159*regulatory sign+0.354*pavement markings+0.401*position of sign board+0.135*post mounted delineators+0.257*accident (1)

Table 5 shows the ranking of road based on the RSA model. Based on these value ranking of each section of road in the study stretch are found out.

TABLE 5: RANKING OF HIGHWAY SECTIONS

Classification	Rank
Good	1
Fair	2
Poor	3

A. *Application of RSM*

Consider a road section with no regulatory sign, completely removed pavement markings, sign board kept with less sight distance, delineators less in number, occurrence of major accident Table.6 shows the classification values corresponding to the conditions of road as obtained by road safety audit. By substituting the corresponding values of these road sections into the road safety model the rank of the road will obtain i.e.

Class of road = -0.394+ 0.159*3+0.354*3+0.401*2+0 .135*2+0.257*2

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$= 2.736 \approx 3 \qquad (2)$$

Hence road is classified as Poor

TABLE 6. ROAD CLASSIFICATION

Condition of road	Classification values
No regulatory sign	3
Completely removed pavement markings	3
Sign board kept with less sight distance	2
Occurrence of major accident	2
Delineators less in number	2

IX. CONCLUSION

The road safety audit will help to identify the accident potential locations and perform safety evaluation of highway sections. In this study, a model for the classification of different sections could be developed. It helps to identify the condition of each road section, hence can make out the section which needs to be improved. Out of the total 66 sections of Kunnamkulam - Peramangalam highway, 33 sections are in good condition with adequate traffic control devices, 25 road sections are in fair condition and 8 sections are in poor conditions. These sections can be improved by providing proper treatments as identified from the audit.

ACKNOWLEDGEMENT

Authors heartily thank Dr. Elangovan T., Executive Director, Kerala Road Safety Authority and Former director of NATPAC for his guidance and suggestions during this project work. The help rendered by Prof. Divya Menon of Department of Computer Science Engineering of Jyothi Engineering College, Cheruthuruthy is gratefully acknowledged.

REFERENCES

1. Pal. S.K. and Mithra S.K. “Multi Layer Perceptorn, Fuzzy sets and classification” in *IEEE Trans on neural network*, Vol. 3, Issue.5, pp.683-692, 1993.
2. Mehar M. and Agarwal P. “A methodology for ranking road safety hazardous locations using analytical hierarchy process" in *Social and behavioral science*, Vol.104, Issue 8,pp. 1030-1037, 2010.
3. Saffrzadeh M. and Farshad S. “Repair and maintenance for highway traffic control devices" in *Journal for science and technology*, Vol. 29,Issue 11,pp.156-164, 2011.
4. Farzaneh M. and Somayya E. “Analysis of the traffic injury severity on two –lane two way rural roads based on classification tree model" in *Safety Science*,Vol.49,Issue.10, pp.1314-1320 2011.
5. Arun B. and Dheeraj K. “Road Safety and Accident Prevention in India: A Review”, in *International Journal of Advanced Engineering Technology*, vol.5, Issue 2,pp.64-68, 2014.
6. Griselda L.and Juan de. O., “Influence of deficiencies in traffic control devices in crashes on two-lane rural" in *Accident analysis and prevention*, Vol.96, Issue.2, pp.130-139, 2016.
7. Ramya V. “Analysis and prediction of Banglore traffic south road accidents” in International Journal on recent and innovation trends in computing and communication 2016, Vol.4, Issue.7, pp. 69-75.
8. Majid K. and Heaslip K. “The effects of damage on sign visibility: An assist in traffic sign replacement" in *Journal for traffic and transportation engineering*, Vol.3, Issue.6, pp. 571-581, 2016.
9. Manual on Road Safety Audit, Indian Roads Congress SP-088, 2010.

Bincy B J is born and brought up at Trivandrum District of Kerala, India. She was born on 24th November 1993. After schoolings, she took Bachelor of Technology in Civil Engineering from Musaliar College of Engineering College, Trivandrum, in the year of 2016. She is pursuing Masters degree in Transportation Engineering in Jyothi Engineering College, Cheruthuruthy, Thrissur, which will be completed in 2019. Her field of interest is in traffic engineering.

Dr. Anitha Jacob is born and brought up at Thrissur District of Kerala State, India. She was born on 21st May 1980. After her schooling at Thrissur, she took Bachelor of Technology in Civil Engineering from Govt. Engineering College, Thrissur during 2001. She took Master’s degree in Traffic and Transportation Planning from National Institute of Technology, Calicut in 2006. Later she had done research in Highway safety and took her Ph.D. in Geometric Design Consistency and Safety of Two-Lane Rural Highway during 2013 from National Institute of Technology, Calicut. She is presently guiding two Ph.D. scholars in the area of Highway safety.She had been working at Jyothi Engineering College, Cheruthuruthy, Thrissur, Kerala since 2002 in various capacities of Lecture, Assistant Professor, and Associate Professor till 2018. Currently she is working as Lecturer in Government Polytechnic College, Chelakkara, Thrissur, Kerala. She was the principal investigator for the 12.7 Lakh project on “Driver Centric Safety Evaluation of Two-Lane Rural Highways” funded by Kerala State Council for Science, Technology and Environment, Trivandrum. Following are a few of her publications

1. Anitha Jacob and Anjaneyulu M V L R, “Operating Speed of Different Classes of Vehicles at Horizontal Curves on Two-Lane Rural Highways” – Journal of Transportation Engineering, Vol. 139 (3) , March 2013, 287-294.
2. Anitha Jacob and Anjaneyulu M V L R, “Development of Crash Prediction Models For Two-lane Rural Highways Using Regression Analysis” – Highway Research Journal, Indian Roads Congress, Vol 6 No. 1, Jan-June 2013.
3. Anitha Jacob, Dhanya R and Anjaneyulu MV L R" Geometric design consistency of multiple horizontal curves on two-lane rural highways", Proceeding of the International Conference on Transportation Researchers Group of India (CTRG-2013), 12-15 December, 2013, Agra.
4. Anitha Jacob, Jinesh K. J., Jisha Akkara & Jose P. Therattil , “Effect of rural highway geometry on driver workload: A step towards safety”, Procedia – Emerging trends in Engineering, Science and Technology for Society, Energy and Environment, Taylor and Francis Group, London, ISBN 978-0-8153-5760-5.

Dr. Jacob was awarded the best paper on road research by Indian Roads Congress in 2015. She is a life member in Indian Roads Congress, ISTE and Institution of Engineers.