Graphical Prediction of Road Accidents using Data Analysis

Swetha P.C,Vaishalli.A,Devika.M,Sneha Balaji

Swetha P.C B.tech ,CSE, SRM Institute of Science And Technology,Tamilnadu,India.
Vaishalli.A B.tech ,CSE, SRM Institute of Science And Technology,Tamilnadu,India.
Devika.M B.tech ,CSE, SRM Institute of Science And Technology,Tamilnadu,India.
Sneha Balaji B.tech ,CSE, SRM Institute of Science And Technology,Tamilnadu,India.

ABSTRACT

Road and traffic accidents are an important concern around the world. Road accidents not only affects the public health with different level of injury but also results in property damage. Data analysis has the capability to identify the different reasons behind road accidents i.e. traffic characteristics, weather characteristics, road characteristics and etc. A variety of research on road accident data analysis has already proves its importance. Some studies focused on identifying factors associated with accident severity while others focused on identifying the associated factors behind accident occurrence. This research analysis uses both the factors associated. Big Data is frequently used method for analyzing road accident data in present research. Trend analysis is another important research area in road accident domain. Trend analysis can assist in identifying the increasing or decreasing accidents rate in different reasons. Large data has been into all aspects of traffic management deeply. It has a significant role especially in traffic accidents. In *traffic trips, people are most concerned about traffic safety. So it has great significance to analyze the big data of traffic accidents. In this study, we have proposed a method to analyze road accident using datasets from the UK government. The motive of this study is to provide an efficient way to analyze and predict the accident zones and the severity. The result shows that the proposed method is capable of efficiently reducing the occurrence of accidents using the graphical predictions*.

Keyword : - *Accident analysis, prediction, severity, casualty* **.**

1. INTRODUCTION

Road and traffic accidents are one of the major cause of fatality and incapacitation across the world. Road contingency can be considered as an event in which a conveyance collides with other conveyance, person or other objects. A road contingency not only provides property damage but it may lead to partial or full incapacitation and sometimes can be fatal for human being. Incrementing number of road accidents is not a good sign for the conveyance safety. The only solution requires the analysis of traffic contingency data to identify different causes of road accidents and taking preventive measures. A variety of research has been done on road contingency data from different countries. UK Government studies used different techniques to analyze road contingency data utilizing statistical techniques and provide fruitful outcomes. Precise and comprehensive contingency records are the substructure for contingency analysis. The efficacious utilization of contingency records depends on three factors, namely, the precision of the data, record retention and data analysis. The desideratum for a high standard of incident reporting is a major prerequisite for the utilization of contingency records to develop road safety measures. If the pristine incident report itself is impecunious, then the analysis and utilization of the results will be impecunious. Erroneous and incomplete contingency data make the results fuzzy, misdirecting, and not very fruitful. Road accidents are unsure and anomalous events and their analysis needs to be cognizant of the factors that affect them. Road accidents are defined by a set of attributes that are often different. The main arduousness of contingency data analysis is its heterogeneity. Consequently, heterogeneity must be quantified through the analysis of the data, or else a number of relationships among the data may stay hidden.

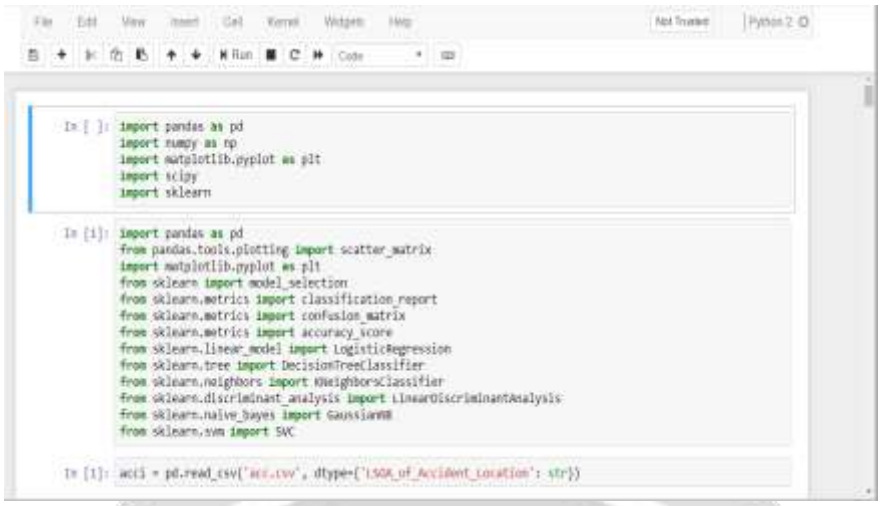

Chart-1: Modules in the coding

2. RELATED WORKS

This topic includes various literature surveys about the prediction and the data analysis of road accidents and data mining research of road accidents. The survey leads about various details like applications and technologies. They gave the detailed view of analysis and the technologies that are useful to the prediction of accidents. They also gave the future enhancement of automated technologies.

2.1 Data Mining Application

E. Bayam, J. Liebowitz, W. Agresti, "Older drivers and accidents: A meta analysis and data mining application on traffic accident data", Expert Systems with Applications, vol. 29, no. 3, pp. 598-629, 2005" - Teenage driving and associated accidents have been thoroughly studied. With the graying of our population in the United States, a focus on senior drivers and related accidents is needed. Unfortunately, there is not one comprehensive study that reviews the major existing studies conducted on senior drivers and accidents. In examining the literature, it also appears that data mining has rarely been applied in studying relationships between senior driver characteristics and accidents. This paper addresses these two needs by providing a meta-analysis of the existing literature on senior drivers and showing how data mining techniques could be used in this application.

2.2 Safety Research Models

L. Chang, W. Chen, "Data mining of tree-based models to analyze freeway accident frequency", "Journal of Safety Research", vol. 36, no. 4, pp. 365-375, 2005 - Statistical models, such as Poisson or negative binomial regression models, have been employed to analyze vehicle accident frequency for many years. However, these models have their own model assumptions and pre-defined underlying relationship between dependent and independent variables. If these assumptions are violated, the model could lead to erroneous estimation of accident likelihood. Classification and Regression Tree (CART), one of the most widely applied data mining techniques, has been commonly employed in business administration, industry, and engineering. CART does not require any pre-defined underlying relationship between target (dependent) variable and predictors (independent variables) and has been shown to be a powerful tool, particularly for dealing with prediction and classification problems.

Method : This study collected the 2001–2002 accident data of National Freeway 1 in Taiwan. A CART model and a negative binomial regression model were developed to establish the empirical relationship between traffic accidents and highway geometric variables, traffic characteristics, and environmental factors.

Results : The CART findings indicated that the average daily traffic volume and precipitation variables were the key determinants for freeway accident frequencies. By comparing the prediction performance between the CART and

the negative binomial regression models, this study demonstrates that CART is a good alternative method for analyzing freeway accident frequencies. Impact on industry : By comparing the prediction performance between the CART and the negative binomial regression models, this study demonstrates that CART is a good alternative method for analyzing freeway accident frequencies.

2.3 Data Mining Techniques

S. Krishnaveni, M Hemalantha, "A Perspective Analysis of Traffic Accident using Data Mining Techniques", International Journal of Computer Applications, vol. 23, no. 7, pp. 40-48, June 2011 - Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligent decisions based on data. Nowadays traffic accidents are the major causes of death and injuries in this world. Roadway patterns are useful in the development of traffic safety control policy. This paper deals with the some of classification models to predict the severity of injury that occurred during traffic accidents. I have compared Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier and Random Forest Tree classifier for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms than other four algorithms.

2.4 Music Signal Analysis

M. Nakano, J. Le Roux, H. Kameoka, T. Nakamura, N. Ono, S. Sagayama, "Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden Markov model", *Proc. of WASPAA*, pp. 325-328, 2011 - This paper presents a Bayesian nonparametric latent source discovery method for music signal analysis. In audio signal analysis, an important goal is to decompose music signals into individual notes, with applications such as music transcription, source separation or note-level manipulation. Recently, the use of latent variable decompositions, especially nonnegative matrix factorization (NMF), has been a very active area of research. These methods are facing two, mutually dependent, problems: first, instrument sounds often exhibit time-varying spectra, and grasping this time-varying nature is an important factor to characterize the diversity of each instrument; moreover, in many cases we do not know in advance the number of sources and which instruments are played. Conventional decompositions generally fail to cope with these issues as they suffer from the difficulties of automatically determining the number of sources and automatically grouping spectra into single events. We address both these problems by developing a Bayesian nonparametric fusion of NMF and hidden Markov model (HMM). Our model decomposes music spectrograms in an automatically estimated number of components, each of which consisting in an HMM whose number of states is also automatically estimated from the data.

3. METHODOLOGY

Data Preprocessing : The data used in this paper is derived from the sample data of traffic accidents in UK government . The collection time of data is of ten years, all of them is about the traffic accident. The total amount of data is more than 2000; the data contains four attributes, namely accident number, accident type, accident location and accident time. Analyzing the above data, among them the accident number is the character type, the type of error is text type, the accident location is text type and the time of the accident is time type. There is a missing value in the data, and the text data is not easy to handle. In literature we can find algorithms that are designed for ne or another VRP(Vehicle routing protocol), where the algorithms are designed to deal with a specific subject or specific constraints. Although mentioned VRP variants mimic some real world situations, these situations do not reflect the whole problem. The mentioned VRPs are criticized for being too focused on specific models that involve non-realistic assumptions. Real-world VRP with various constraints generalizes traditional VRP and is usually called a rich vehicle routing problem (RVRP). Solving RVRP has been a challenging today's task. Metaheuristic is another approach for solving a complex problem that may be too difficult or time-consuming for other techniques. One of the metaheuristics that are investigated for solving VRP is a genetic algorithm (GA). Genetic algorithms are based on ideas of evolution theory. The main principle here is that only the fittest entities survive. Genetic algorithms work with individuals, sometimes also called chromosomes, each representing a possible solution to a given problem. GA typically works with the initial population of solutions; together with each new generation GA creates a new potential offsprings, based on the selected individuals from the previous generation using a set of stochastic transition operators (crossover and

mutation). The iterative process of generations and evaluation of individuals continues until a sufficient stopping criterion is met.

The standard genetic algorithm has limitations in the constrained environment. Due to a stochastic characteristic, genetic algorithms can continue very long until the acceptable solution has been found for a constrained problem. For a constrained problem, the feasible search space is smaller than the whole search space and genetic algorithm operators generate solutions in the whole search space including the infeasible space. The common approaches for constraint handling in genetic algorithms involve additional repair and improvement methods that are designed for a specific constraint to keep the generated solutions in the feasible search space. The repair of one constraint can involve the violation of another constraint. Such approaches can produce an inadequate result when they are applied to different problems and are hardly extendable with new constraints. Specialized algorithms usually are hardly applicable to RVRP.

```
In [12]: vehi.describe()
```

Out[12]:

	Vehicle_Reference	Vehicle_Type	Towing_and_Articulation	Vehicle_Manoeuvre	Vehicle_Location-Restricted_Lane	Junction_Location	Ski
count	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06	
mean	1.592529e+00	9.720490e+00	3.412250e-02	1.278995e+01	1.281139e-01	2.355335e+00	
std	7.494053e-01	7.873360e+00	3.187744e-01	6.151990e+00	9.721220e-01	3.075178e+00	
min	1.000000e+00	-1.000000e+00	-1.000000e+00	-1.000000e+00	-1.000000e+00	-1.000000e+00	
25%	1.000000e+00	9.000000e+00	0.000000e+00	7.000000e+00	0.000000e+00	0.000000e+00	
50%	1.000000e+00	9.000000e+00	0.000000e+00	1.700000e+01	0.000000e+00	1.000000e+00	
75%	2.000000e+00	9.000000e+00	0.000000e+00	1.800000e+01	0.000000e+00	5.000000e+00	
max	2.800000e+01	9.000000e+01	5.000000e+00	1.800000e+01	9.000000e+00	8.000000e+00	

8 rows × 21 columns

Chart-2 : Methods in Coding Data

3.1 Implementation

The Dataset was collected from the UK government for the predicting the road accident analysis. Many related .csv files have been obtained so that the result is obtained and a conclusion can be derived. Large data has been into all aspects of traffic management deeply. It has a significant role especially in traffic accidents. In traffic trips, people are most concerned about traffic safety. So it has great significance to analyze the big data of traffic accidents. Traffic accidents have great importance on whether it is for people to travel recommendations, or the transport sector takes a reasonable and practical way to ease the traffic. In this paper, the data is analyzed and processed, the invalid data is deleted, the dirty data is removed and the data of the traffic accident are analyzed, modeled, forecasted and verified. After the data is cleaned, the correctness of each model is verified and the data is displayed on the map. Find the law between the potential of data, through the adaptive method to choose a high degree of correctness model. Forecast the severity of traffic accident which occurred at a particular time and place to improve the effectiveness of the data.

This system analysis the traffic accidents are occurring every day. These are briefly recorded for the purpose of administration formalities and compensation or insurance purpose. Most of the normal injury cases are not reported or settled at home.Traffic Police write down the description of the incidents at the site without specific form checklist. Moreover, the writing formats vary person to person. Hand written forms are stored and summarized records are reported to the higher authorities of the police. Thus the system improves many factors including graphical representation and a detailed analysis of road accidents.

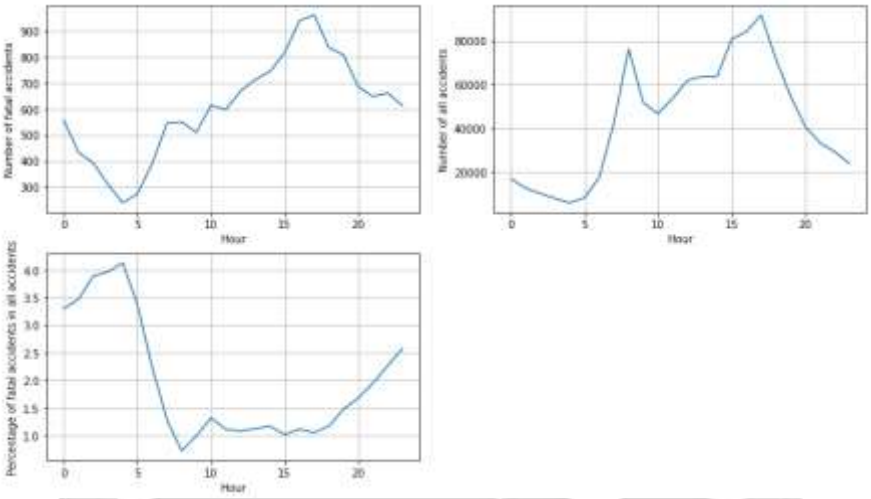

Chart 3: Output for Time And Area

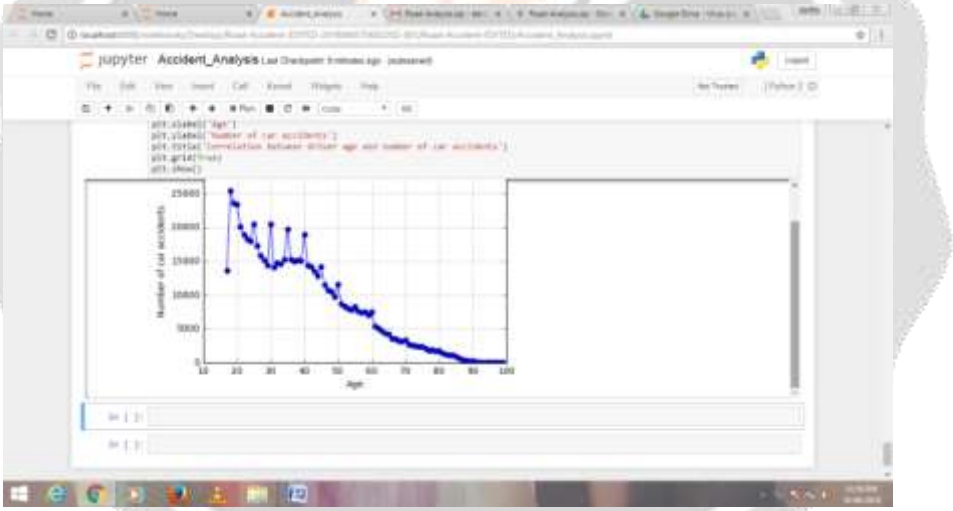

Chart 4 : Output for Time Period

4. CONCLUSIONS

This paper is derived from the sample data of traffic accidents in UK government . The collection time of data is of ten years, all of them is about the traffic accident. The total amount of data is around 2000; the data contains various attributes, namely accident number, accident type, accident location and accident time, etc. Analyzing the above data, the causes and the trend of the accidents occurred for over a period of 10 years (2005-2015) have been predicted by plotting respective graphs based on the attributes.

We also need to mention that all the data is about fatal accident, so no matter how long would it take for to arrive, there would always be fatal. Also, there is no variable recording at what time the death happened, and a lot of records are missing value at time, so very limit information could be inferred from the time relevant attributes.

The percentage of fatal accident happened on different weather is in comparison with percentage of people and fatal involved. Most fatal accidents happened at clear/cloud weather. This is understandable because clear/cloud is the most usual case of weather condition.

In future work, this dataset can be enhanced by using an automated method for collecting data.

5. ACKNOWLEDGEMENT

It is indeed a pleasure to mention about Mrs. A.Shiny, M.E (CSE), Assistant Professor, project guide who have always been patient enough to make the complexities of the project and relentlessly supported me throughout the project.

6. REFERENCES

1. Priyanka A Nandurge Nagaraj V.Dharwadkar, "Analyzing Road Accident Data using Machine Learning Paradigms " **,** International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017).
2. Sachin Kumar and Durga Toshniwal." Analysing road accident data using association rule mining". In Proceedings of International Conference on Computing, Communication and Security, pages 1–6, 2015.
3. Maria Riveiro, Mikael Lebram, and Marcus Elmer , "Anomaly Detection for Road Traffic: A Visual Analytics Framework"**,** 1524-9050 © 2017 IEEE.
4. U.S. Department of Transportation. FARS analytic reference guide, 1975 to 2009 /811352,2010.
5. Sachin Kumar, Durga Toshniwal, "A data mining approach to characterize road accident locations" J. Mod. Transport. (2016) 24(1):62–72.
6. Liling Li, Sharad Shrestha, Gongzhu Hu "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques **" ,** 978-1-5090-5756- 6/17/$31.00 ©2017 IEEE SERA 2017, June 7-9, 2017, London, UK.
7. Chen CHEN, " Analysis and Forecast of Traffic Accident Big Data", Tunnel Traffic Engineering Design Office, Yunnan Province Transportation Planning and Design Institute, Kunming, China 522253113.
8. Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. "Applying association rules mining algorithms for traffic accidents in dubai". International Journal of Soft Computing and Engineering, 2015.