PREDICTIVE ANALYTICS OF ROAD ACCIDENTS IN TRAFFIC VIOLATION USING MACHINE LEARNING APPROACH

1th B.**Mohan Prabu**
Dept.of Computer Science and Engineering
Panimalar Institute of Technology
(Anna university)
Chennai,India

2nd **Neeraj Chandran**
Dept.of Computer Science and Engineering
Panimalar Institute of Technology
(Anna university)
Chennai,India

3rd **Syed Waseem.K**
Dept.of Computer Science and Engineering
Panimalar Institute of Technology
(Anna university)
Chennai,India

---***---

Abstract-*There are many inventories in automobile industries to design and build safety measures for automobiles, but traffic accidents are unavoidable. There is a huge number of accidents prevailing in all urban and rural areas. Patterns involved with different circumstances can be detected by developing an accurate prediction models which will be capable of automatic separation of various accidental scenarios. This cluster will be useful to prevent accidents and develop safety measures. It is believed to acquire maximum possibilities of accident reduction using low budget resources by using some scientific measures. The proposed machine techniques and algorithms to predict the cause of accidents in the areas. In this proposed system, take three attributes such as accidents, casualty and vehicle. Attributes are classified using machine learning Random forest classifier. Linear regression is used to predict the rate of accident casualties and rate of severity of the casualties are and finally the prediction will be reported via graph.*

1.INTRODUCTION

There is a huge impact on the society due to traffic accidents where there is a great cost of fatalities and injuries. In recent years, there is a increase in the researches attention to determine the significantly affect the severity of the drivers injuries which is caused due to the road accidents. Accurate and comprehensive accident records are the basis of accident analysis. The effective use of accident records depends on some factors, like the accuracy of the data, record retention, and data analysis. There are many approaches applied to this scenario to study this problem. A recent study illustrated that the residential and shopping sites are more hazardous than village areas.as might have been predicted, the frequencies of the casualties were higher near the zones of residence possibly because of the higher exposure. A study revealed that the casualty rates among the residential areas are classified as relatively deprived and significantly higher than those from relatively affluent areas.

2.EXISTING SYSTEM

Many research studies focused solely on identifying the fundamental factors that cause road crashes. From these studies, it was noticed that human factors have the most significant impact on accident risk. The basic factors influence on road safety directly related to the driver are i.e., driving behaviour, driver's perception of traffic risks and driving experience. Drivers involve frequently in attitudes that cause road safety issues. Many of these attitudes are dynamic, conscious rule violations, while others are the result of errors due to less driving experience, momentary mistakes, inattention or failure to perform function, the latter often related to age. These behaviours often contribute to traffic collisions. Besides of risky driver behaviour the bad driving practices and poor knowledge along with disrespect for road and safety regulations are the obvious problems.

2.1 Disadvantage

1. The study investigated that the task of driving can be easy or difficult depending on the momentary task demand of driving and the driver's skill to control his/her vehicle correctly.

2. To examine the relationship between driving behaviour and the number of traffic accidents in each country, regression analyses was performed by using forward stepwise procedure. Attackers wishing to take a more active approach to interception may launch one of the attacks.

3.PROPOSED SYSTEM AND METHODOLOGY

Models are created using accident data records which can help to understand the characteristics of many features like drivers behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid accidents. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. The model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it.

Here the road accident study is done by analyzing some data by giving some queries which is relevant to the study. The queries like what is the most dangerous time to drive, what fractions of accidents occur in rural, urban and other areas. What is the trend in the number of accidents that occur each year, do accidents in high speed limit areas have more casualties and so on …? These data can be accessed using Microsoft excel sheet and the required answer can be obtained. This analysis aims to highlight the data of the most importance in a road traffic accident and allow predictions to be made. The results from this methodology can be seen in the next section of the report.

3.1 Advantage

1. There is evidence that driving impairments are related to time of day, loss of sleep or sleep inertia, prolonged work and minor illness.
2. It is important, therefore, to assess the frequency with which people drive when they are potentially fatigued because of this range of risk factors.

4.ALGORITHM

4.1 Random Forest

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

4.1.1 Working

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
- For classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

5.LIST OF MODULES

5.1 Exploratory Data Analysis

During this step we performed some descriptive analysis and determined the target variable. We also explored how many classes were in the target and a selection of other possibly problematic (high cardinality) variables. I also visualized the target variable in a histogram which is a good technique for understanding the distribution of the data to assist in parameter tuning.

5.2 Data Cleaning

We dropped those high cardinality variables during this step as a precursor to the pre-processing step.

5.3 Pre-processing & Transformation

We removed the target variable from the entire data set and transformed the categorical variable into a model matrix with one-hot encoding. This is sometimes the requirements for certain algorithms to process the data in a sparse matrix format. Other statistical software such as R, automates this step when generating models. I imputed the missing values in the data to 0. I scaled the continuous variables using min-max normalization which transforms values onto a scale from 0 to 1 to prevent variables on different scales heavily impacting the coefficients.

5.4 Performance Comparison

It is well known that one of the very important factors causing different injury level is the actual speed that the vehicle was going when the accident happened. Unfortunately, our dataset doesn't provide enough information on the actual speed since speed for 67.68% of the data records was unknown. If the speed was available, it is extremely likely that it could have helped to improve the performance of models studied in this paper.

6.IMPLEMENTATION

6.1 Levels

Unit testing, White-box testing is done during unit testing to ensure that the code is working as intended, before any integration happens with previously tested code. White-box testing during unit testing catches any defects early on and aids in any defects that happen later on after the code is integrated with the rest of the application and therefore prevents any type of errors later on.

Integration testing, White-box testing at this level are written to test the interactions of each interface with each other. The Unit level testing made sure that each code was tested and working accordingly in an isolated environment and integration examines the correctness of the behaviour in an open environment through the use of white-box testing for any interactions of interfaces that are known to the programmer.

Regression testing, White-box testing during regression testing is the use of recycled white-box test cases at the unit and integration testing levels.

White-box testing's basic procedures involve the understanding of the source code that you are testing at a deep level to be able to test them. The programmer must have a deep understanding of the application to know what kinds of test cases to create so that every visible path is exercised for testing. Once the source code is understood then the source code can be analysed for test cases to be created. These are the three basic steps that white-box testing takes in order to create test cases:

Input, involves different types of requirements, functional specifications, detailed designing of documents, proper source code, security specifications. This is the preparation stage of white-box testing to layout all of the basic information.

Processing Unit involves performing risk analysis to guide whole testing process, proper test plan, execute test cases and communicate results. This is the phase of building test cases to make sure they thoroughly test the application the given results are recorded accordingly.

Output, prepare final report that encompasses all of the above preparations and results.

The course of accident can be divided into three stages: pre-collision, collision, and post-collision. Each stage gives us a significant amount of information but also involves several difficulties as discussed below.

Pre-Collision, the pre-collision case is the most vital information to explain an accident scenario. Also, this information may become a good evidence for crime scene investigation. The pre-collision situation is a clear violation of traffic rules by any/both the vehicles, which include violation of traffic lane, violation of signals at intersections, violation of speed limit at congested roads, abrupt motion on the road, etc. Finally, we can say that pre-collision stage is an unusual activity and thus can be easily detected by applying anomaly detection methods based on the various parameters, such as speed, trajectories, position, etc.

Collision, the collisions are essential to accident detection, but it is very complicated to detect and cannot be directly detectable by any general purpose computer vision technique. One way to detect a collision is to identify the joints of the trajectories of the vehicles over spatiotemporal dimensions. However, the major challenge is the discrimination between collision and occlusion. For this we use the trajectories over space-time interest points [39] and improved dense trajectories.

Post-collision, as stated above that the collision and occlusions are hard to classify and may lead to false alarms. These false alarms further can be refined by considering the post-collision scene. The two most common post-collision scenes include: 1) Fallen objects at the collision point: As we stated that the intersection of the trajectories of two vehicles might be a collision or an occlusion. However, after the intersection, if both the trajectories are continued, and no abrupt or zig-zag motion resulted. Then, the intersection is merely an occlusion, not a collision. However, if some abrupt motion or discontinued trajectories have occurred, then the possibility of a collision is high. Measure the time for which the object remains static. 2) Crowd attention towards the collision point: The last and final stage of the accident is the crowded road or pedestrians running towards the collision point.

7.CONCLUSION

The Random forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The algorithm still suffers from the imbalanced dataset problem and requires more pre-processing to give better results at the results is great but it could have been better if more pre-processing have been done on the data.

8.REFERENCES

[1] V. Kostakos, T. Ojala, and T. Juntunen, “Traffic in the smart city: Exploring city-wide sensing for traffic control center aug- mentation,” IEEE Internet Comput., vol. 17, no. 6, pp. 22–29, Nov. 2013.

[2] I. Celino and S. Kotoulas, “Smart cities [Guest editors' introduction],” IEEE Internet Comput., vol. 17, no. 6, pp. 8–11, Nov./Dec. 2013, doi: 10.1109/MIC.2013.117.

[3] European Initiative on Smart Cities, 2010–2020. Accessed: May 15, 2018. [Online]. Available: https://setis.ec.europa.eu/set-plan-implementation/technology-roadmaps/european-initiative-smart-cities

[4] S. Djahel, R. Doolan, G.-M. Muntean, and J. Murphy, “A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches,” IEEE Commun. Surveys Tuts., vol. 17, no. 1, pp. 125–151, Mar. 2015.

[5] A. Dopfer and C.-C. Wang, “What can we learn from accident videos?” in Proc. Int. Autom. Control Conf. (CACS), Nantou, Taiwan, Dec. 2013, pp. 68–73.

[6] C. Regazzoni, A. Cavallaro, Y. Wu, J. Konrad, and A. Hampapur, “Video analytics for surveillance: Theory and practice,” IEEE Signal Process. Mag., vol. 27, no. 5, pp. 16–17, Sep. 2010.

[7] G. Yuan, X. Zhang, Q. Yao, and K. Wang, “Hierarchical and modular surveillance systems in ITS,” IEEE Intell. Syst., vol. 26, no. 5, pp. 10–15, Sep./Oct. 2011.

[8] S. Xia, J. Xiong, Y. Liu, and G. Li, “Vision-based traffic accident detection using matrix approximation,” in Proc. Asian Control Conf. (ASCC), Kota Kinabalu, Malaysia, May/Jun. 2015, pp. 1–5.

[9] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, and Y. Liu, “Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment,” IET Intell.

Transp. Syst., vol. 10, no. 6, pp. 428–437, Aug. 2016.

[10] K. Yun, H. Jeong, K. M. Yi, S. W. Kim, and J. Y. Choi, "Motion interaction field for accident detection in traffic surveillance video," in Proc. 22nd Int. Conf. Pattern Recognit. (ICPR), Stockholm, Sweden, Aug. 2014, pp. 3062–3067.