IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review On Traffic Collision Analysis and Prediction

Abhijeet M. Tote

Student

Department Of Computer Engineering,

Matoshri College Of Engineering & Research Centre, Eklhare, Nashik, India

Abstract—The increase in road fatalities comes as bad news. This cannot be stopped but can be controlled. The Accidents may cause due to driver's health, Driver's feelings, vehicle speed, climate condition, traffic conditions, road conditions, etc. Analysis and prediction on Traffic collision has gained importance in these days. The big dataset is generated every year. The analysis and prediction helps to define prevention measures. This paper defines the study of existing analysis and perdition systems for road and other media. Based on the study of existing work, a new system is for road accidents analysis and prediction is proposed.

Keywords— Road accidents, association rules ,big data ,hadoop, clustering, EMM, feature extraction

I. INTRODUCTION

The increase in road fatalities comes as bad news. This has a negative social and economic impact. This cannot be stopped but can be controlled. World Health Organization (WHO), announces the statistical study every year. 60 million people get injured and 1.60 million people died consistently due to road accidents. A center for Disease Control and Prevention (CDCP) announces the economical impact of road accidents. The accidents cause loss of 100 billion every year The Accidents may caused due to driver's health, Driver's feelings, vehicle speed, climate condition, traffic conditions, road conditions, etc. Analysis and prediction on Traffic collision has gained importance in these days. The big dataset is generated every year. Enormous actions have been taken to enhance Road safety. Conventional strategies can't be utilized in such frameworks as the information produced is in huge volumes.

There is need of such system that analyzes this data in automatic faction. The machine learning algorithms are required for processing. By Analyzing the generated data the common accident types can be clubbed together. It will helpful for statistical analysis. The cause of accident can be predicted by machine learning algorithm. The possibility of accident can be predicted. Such predictions help to add more preventive measures. The requirement in the domain of road accidents motivates to develop a new system. The fusion of existing machine learning algorithms and big data analysis techniques can generate a better analytical report and prediction set.

The accidents may cause due to various parameters such as: climate condition, road conditions, drivers mental status, driver experience, etc. Analysis of accident cases based on such parameters helps to predict the future accident style at certain location or at certain cities. This helps to create preventive measure accordingly.

Machine learning algorithm helps to analyze data using one or more parameters. The clustering algorithm clubs together the similar type of accidents. The association mining rule technique finds the patterns in a dataset. The pattern mining technique helps to extracts the dependency of two or more type of attributes that may cause and accidents. The prediction can be done on the resultant effect of an accident in the form of number of injuries and death count.

For example: Driver age=50, whether condition = fog, driver drink = Not checked, Road surface=Dry may cause accident with injury.

The proposed system works on analysis and prediction of road accidents information data using machine learning algorithms and its efficient execution. For efficient execution process, feature extraction technique and Hadoop processing is used. For analysis same type of accidents are clustered together using EMM algorithm, Association Rule Mining (IARM) algorithm. The generated association rules are then provided to the Congestion control using Machine Framework (CCMF) and Traffic Congestion Analyzer using Map Reduce (TCAMP) algorithms to generate predictions.

Following section includes the study of existing work in the domain of accident analysis and prediction for road and for other media like railways, airline, etc. Based on the existing system analysis, a problem formulation is proposed in section III. Based on the problem identified, a new system architecture is proposed in section IV followed by the conclusion.

II. RELATED WORK

Sarkar Set. al.[2] proposes a prediction system. This system predicts the possible incident in steel plant based on the related stored data. It predicts the occurrence of injury cases and their probable causes. It uses text mining technique with 3 different classification algorithms: Support Vector Machine SVM, Random Forest RF and Maximum entropy Max Ent. These classifiers generate better results in binary and multi-class prediction model. An Ensemble approach is proposed for multi-class prediction model. The Ensemble approach improves accuracy.

Flight crash investigation system is proposed by Sharma S, et.,al.[3]. This system focuses on the flight crash investigation and analysis using data mining techniques. It finds the ground/abroad fatality rate. The clustering algorithm K-Means is used along with the cosine similarity measure. The clustering results group the similar crash information in one group. Similarity finds the relation among different texts of crashes.

A railroad accident investigation reports generation system is proposed by Williams T, et.,al.[4]. This technique generate a statistical analysis report based on Similarity found among different texts of railroad Accidents. In this technique, the text form published articles is analyzed using data mining techniques. The dataset of published articles contains railway accidents. Using LDA technique topics are extracted from text. The K means clustering is applied on extracted text. The clustering results show that there is recurring themes in many major accidents. After grouping the accidents data the main causes of accidents are extracted and relation among multiple accidents is identified.

Sarkar S,et.,al,[5] Works on prediction of occurrence of accidents and its outcome such as injury, near miss, fatal death, property damage, etc. . And finds the inter relationship of factors causing accidents. The two machine learning algorithms : support vector machine (SVM) and artificial neural network (ANN) are used. The parameter passed to this algorithm is optimized using genetic algorithm and particle swarm optimization. After these algorithms association rules are extracted with the help of decision tree algorithm with PSO and SVM. This technique extracts the root cause behind the injury.

Verma A[6] proposes a analysis tool for steel plant incident data. It explores the hidden factors and patterns form the description. It also identifies the anomaly in incidence reporting. The data is in case report text format. It uses singular value decomposition (SVD) and expectation-maximization (EM) algorithm. This paper only focuses on grouping of similar type of accident data.

Williams T. et.al[7] proposes a system to analyze road accidents based on text mining techniques such as: probabilistic topic modeling and k-means clustering. The system works on major accidents that are occurred in the same fashion. Parameters those are analysed in the system are: track defects, grade crossing accidents, wheel defects, and switching accidents.

Ghazizadeh et.al[8] proposes a technique to analyze national highway complaint dataset. The data is analyzed at 2 levels: fatal incident and injury. latent semantic analysis (LSA) and hierarchical clustering technique is used to cluster complaints.

F. Abdat[9] proposes a system that analyzes recurrent accidents caused due to Movement Disturbance. This is called as Occupational Accident with Movement Disturbance (OAMD) scenarios. The dataset is in the form of descriptive text from. A Bayesian Network (BN)-based model is used to extract informative text. Then Most Probable Explanation (MPE) is extracted by creating clusters based on the similarity measurements.

H. R. Marucci, M. R. Lehto, H. L. Corns, [10] proposed system which is having injury details recorded from huge U.S. insurer as an input. Earlier fuzzy bayes technique is used with some improvements. This improvement contains adding of predictors in form of sequenced words and before calculation common subsets are removed before calculations. This pushed prediction strength to next level. Accuracy is achieved in several categories which were to be found difficult to code on past. In this paper two-tired approached is adopted. In this two-tired approach firstly narratives are categorize in broader level like " falls ". Afterwards just before classification it is categorize to the more refined level like "falls from height". 79% of sensitivity is achieved at broader level categorization and 66% for more refined categorization.

T. Rivas, M. Paz, J. E. Martn, J. M. Matas, J. F. Garca, J. Taboada [11] proposed system over traditional surveys conducted for workplace risks. These authors claims that existing survey are unable to predict accidents. Hence proposed system tried model incidence and accidents in two companies working in mining and construction sector. It defines predictive models by analyzing most important causes of accidents. Data-mining techniques including Bayesian networks , decision rules, classification trees, support vector machines. After incident / accident interviews are also conducted and analyzed.

Y. M. Goh, D. Chua [12] proposed system using accidental data from Singapore construction industry. Neural network analysis was carried out on OSHMS (i.e. Occupational Safety and Health Management System). The study is carried out to prove that how OSHMS elements and safety performance are related with each other and how neural network methodology is useful to analyze that. Three most important elements in case study are emergency preparedness, incident investigation and analysis, and group meetings. This study concluded that neural network techniques are useful for analyzing OSHMS input data to find meaningful imminent about improvements about safety performance.

Babu S.N., et.al.[1] works on road accidents prediction theory based on various parameters like driver-age, experience, vehicle type, whether condition, road conditions, etc. The system uses Congestion control using Machine Framework (CCMF) and Traffic Congestion Analyzer using Map Reduce(TCAMP) algorithms for prediction on road accidents. Initially the system forms clusters based on the clusters the association rules are extracted. Using association rules predictions are performed. Lots of parameters are used for mining accidental data. The attribute reduction will be useful for accuracy improvement.

Paper	Description	Analysis
Sarkar S ,et. al.[2]	It proposes a prediction system which predicts the possible incident in steel plant based on the related stored data. Text mining technique is used with 3 different classification algorithms: Support Vector Machine SVM, Random Forest RF and Maximum entropy Max Ent.	These classifiers generate better results in binary and multi-class prediction model. Multi-class prediction model is proposed. Ensemble approach improves accuracy.
SharmaS,et.,al.[3].	This system proposed Flight crash investigation system which focuses on the flight crash investigation and analysis. It uses data mining techniques. It finds the ground/abroad fatality rate. Clustering algorithm K-Means is used along with the cosine similarity measure.	The clustering helps to group similar kind of crashes. Its helps to jump on certain conclusions.
WilliamsT,et.,al.[4].	Railroad accident investigation and analysis is proposed in this system. In this technique, the text form published articles is analyzed using data mining techniques. Using LDA technique topics are extracted from text. The K means clustering is applied on extracted text.	LDA technique is used to extract the accidental causes (topic extractions) and clustering results show that there is recurring themes in many major accidents and relation among multiple accidents is identified
Sarkar S,et.,al,[5]	It works on prediction of occurrence of accidents and its outcome such as injury, near miss, fatal death, property damage, etc. . Support vector machine (SVM) and artificial neural network (ANN) algorithms are used. Also The parameter passed to this algorithm is optimized using genetic algorithm and particle swarm optimization.	Association rules are extracted which needs decision tree algorithm with PSO and SVM. New technique is discussed to extracts the root cause behind the injury.
Verma A[6]	It proposes a analysis tool for steel plant incident data. It explores the hidden factors and patterns form the description. It uses singular value decomposition (SVD) and expectation-maximization (EM) algorithm.	The drawback of this paper is that it only focuses on grouping of similar type of accident data.
Williams T. et.al[7]	It proposes a system to analyze road accidents based on text mining techniques such as: probabilistic topic modeling and k-means clustering. Parameters those are analyzed in the system are: track defects, grade crossing accidents, wheel defects, and switching accidents.	The system works on major accidents that are occurred in the same fashion. Clustering helps to find similar kind of accidental causes and find relations between the clusters
Ghazizadeh et.al[8]	It proposes a technique to analyze national highway complaint dataset. The data is analyzed at 2 levels: fatal incident and injury. latent semantic analysis (LSA) and hierarchical clustering technique is used to cluster complaints.	Due to hierarchical clustering technique interrelations amongst incidentals reasons are easily traceable.

Paper	Description	Analysis
F. Abdat[9]	It proposes a system that analyzes recurrent accidents caused due to Movement Disturbance. This is called as Occupational Accident with Movement Disturbance (OAMD) scenarios. The dataset is in the form of descriptive text from. A Bayesian Network (BN)-based model is used to extract informative text. Then Most Probable Explanation (MPE) is extracted by creating clusters based on the similarity measurements.	Use of Bayesian Network (BN) is done to extract information text which is innovative use of this network technique.
H. R. Marucci, M. R. Lehto, H. L. Corns, [10]	It proposed system which is having injury details recorded from huge U.S. insurer as an input. Earlier fuzzy bayes technique is used with some improvements. In this paper two-tired approached is adopted. In this two-tired approach firstly narratives are categorize in broader level and afterwards just before classification it is categorize to the more refined level.	By using two tired approach system sensitivity level is achieved upto 79% at broader level categorization and 66% for more refined categorization.
T. Rivas, M. Paz, J. E. Martn, J. M. Matas, J. F. Garca, J. Taboada [11]	It proposed system over traditional surveys conducted for workplace risks. Proposed system tried model incidence and accidents in two companies working in mining and construction sector. Data-mining techniques including Bayesian networks , decision rules, classification trees, support vector machines. After incident / accident interviews are also conducted and analyzed.	This system is succeeds in finding relations between the accidental causes and may predicts future risks
Y. M. Goh, D. Chua [12]	It proposed system using accidental data from Singapore construction industry. Neural network analysis was carried out on OSHMS (i.e. Occupational Safety and Health Management System). Neural network methodology is used to analyze that relation between elements and safety performance. Preparedness, incident investigation and analysis, and group meetings are considered.	This study concluded that neural network techniques are useful for analyzing OSHMS input data to find meaningful imminent about improvements about safety performance.
Babu S.N., et.al.[1]	It works on road accidents prediction theory based on various parameters like driver-age, experience, vehicle type, whether condition, road conditions, etc. The system uses Congestion control using Machine Framework (CCMF) and Traffic Congestion Analyzer using Map Reduce(TCAMP) algorithms for prediction on road accidents. Initially the system forms clusters based on the clusters the association rules are extracted. Using association rules predictions are performed. Lots of parameters are used for mining accidental data. The attribute reduction will be useful for accuracy improvement.	This paper is selected for implementation as it is having Map reduced technique along with clustering approach which is useful to find association rules. These association rules are useful for accidental predictions are performed. For accuracy this paper performs special efforts with the help of attribute reduction technique.

III. ANALYSIS AND PROBLEM FORMULATION

Lot of work has been done on road accident analysis using text mining technique. Most of the system generates a statistical analysis by clustering algorithm. The rule extraction helps to analyse the patterns and helps to predict the future occurrence of accidents. There is need to generate analytical report from the big accidental database efficiently and generate a road accident predictions by considering more than one factor at a time that causes accidents.

The problem statements can be defined in three folds:

1. Generate an analytical report from accidental database.
2. Generate predictions on accidents using association rule mining.
3. To improve system efficiency for processing big data.

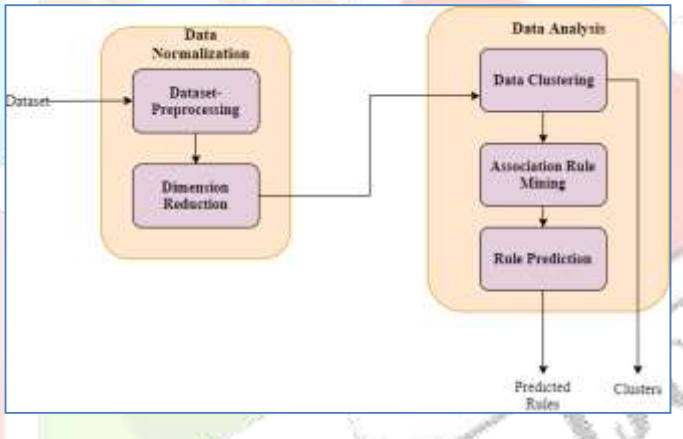
IV. PROPOSED METHODOLOGY

A. Architecture

Following fig.1 describes the architecture of the system. The accident information dataset is input to the system. The Prediction rules and clusters are the output of system. System mainly categorized in 2 sections: Data normalization and data analysis. Data normalization is treated as a preprocessing step whereas in data analysis phase actual data mining techniques are applied such as data clustering, association rule mining and rule prediction algorithms.

In pre-processing step important dimensions of dataset is selected in dimension reduction step. The related data is grouped together using clustering technique. For clustering Expectation-Maximization algorithm is used. Form the generated groups the association rules are extracted using Improved Association Rule Mining (IARM) algorithm. Using Congestion control using Machine Framework(CCMF) and Traffic Congestion Analyzer using Map Reduce(TCAMP) algorithms rules can be predicted.

FIGURE 1 : SYSTEM ARCHITECTURE



B. System Working:

The data processing is mainly described in following 4 sections:

1. Data Preprocessing:

In data preprocessing the noisy data is removed. The dataset contains raw information of all type of accidents. The unwanted information from dataset is removed.

The dataset may contain missing values. The missing values are filled using binning and linear regression technique.[7]

2. Dimension reduction:

A Wrapper technique: CFS subset evaluator is used to reduce the dimension count. The reduced dimension count helps to improve system efficiency. This finds the optimal feature set based on the classifier performance.

3. Data Clustering

The whole dataset of road accidents is initially divided in number of clusters based vehicle type and then the cluster is divide in subgroups using parameter like : time, drivers experience, climate etc. The Enhanced Expectation-Maximization algorithm is used for data clustering.[9] this clustering algorithm focuses on grouping of similar members based on probability distribution.

4. Association Rule Mining

From the clustered data association among data is extracted using Improved Association Rule Mining (IARM) algorithm[3]. This technique extracts the strong association using support value. The rule are extracted based on vehicle class and road accident parameters.

5. Rule Prediction

Congestion control using Machine Framework(CCMF) and Traffic Congestion Analyzer using Map Reduce(TCAMP) algorithms are used for prediction.

The generated clusters of EM algorithm based on vehicle type are given to the input to CCMF algorithm. This algorithm finds the clusters of relevant parameters. The cluster is in the form of tree. Using the generated clusters and association rules the TCAMP algorithm predicts

the possibility of road accidents. This is a hadoop based map reduce program. The map reduce technique improves the efficiency of execution.

V. CONCLUSIONS

This paper includes the study of various analysis and prediction systems in domain of accidents. The accident analysis includes the road, railway, airline and manufacturing plant accidents. In the study of analysis of accident system, the causes of accidents and the implications are studied. Based on the analysis of a prediction can be estimated. The prediction helps to define prevention measures. Based on the analysis of existing system, a new system is proposed for analysis and prediction for road accidents. The proposed system analyzes the accident dataset and finds the clusters form the data. The road accident may caused due to various parameters like climate, road condition, driver status ,etc. The system extracts association rules from the data using IARM algorithm. Based on the rules and accident type road accident predictions are done using CCMF and TCAMP algorithm. To improve system efficiency feature selection and map reduce strategy is used.

VI. REFERENCES

[1] S.N., Tamilselvi J., "Generating road accident prediction set with road accident data analysis using enhanced expectation-maximization clustering algorithm and improved association rule mining", Journal Europeen des Systemes Automatises, Vol. 52, No. 1, pp. 57-63, April 2019. https://doi.org/10.18280/jesa.520108

[2] Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. http://dx.doi.org/10.14299/ijser.2013.01

[3] Sharma S, Sabitha AS. (2016). Flight crash investigation using data mining techniques. In Information Processing (IICIP), 2016 1st India International Conference on. IEEE, pp. 1-7. http://dx.doi.org/10.14299/ijser.2013.01

[4] Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009-V001T06A009. http://dx.doi.org/10.14299/ijser.2013.01.

[5] Sarkar S, Vinay S, Raj R, Maiti J, Mitra P. (2018). Application of optimized machine learning techniques for prediction of occupational accidents. Computers & Operations Research (Elsevier), pp. 343-348. http://dx.doi.org/10.1145/3075564.3078884

[6] Verma A, Maiti J. (2018). Text-document clustering based cause and effect analysis methodology for steel plant incident data. International Journal of Injury Control and Safety Promotion, 1-11. http://dx.doi.org/10.1080/17457300.2018.1456468

[7] Williams T, Betak J, Findley B., "Text mining analysis of railroad accident investigation reports", In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009-V001T06A009. http://dx.doi.org/10.14299/ijser.2013.01.

[8] Ghazizadeh M, McDonald AD, Lee JD., "Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database", Human Factors 56(6): 1189-1203. http://dx.doi.org/10.1504/IJFCM.2017.089439

[9] Abdat F, Leclercq S, Cuny X, Tissot C.,"Extracting recurrent scenarios from narrative texts using a bayesian network: Application to serious occupational accidents with movement disturbance", Accident Analysis & Prevention 70: 155-166. http://dx.doi.org/10.1016/j.aap.2014.04.004

[10] H. R. Marucci, M. R. Lehto, H. L. Corns, Computer classification of injury narratives using a Fuzzy Bayes approach: improving the model,In Human Interface and the Management of Information. Methods,Techniques and Tools in Information Design (pp. 500 - 506). SpringerBerlin Heidelberg, 2007

[11] T. Rivas, M. Paz, J. E. Martn, J. M. Matas, J. F. Garca, J. Taboada,Explaining and predicting workplace accidents using data-mining tech-niques, Reliability Engineering & System Safety, 96(7), pp. 739 - 747,2011

[12] Y. M. Goh, D. Chua, Neural network analysis of construction safety man-agement systems: a case study in Singapore, Construction Managementand Economics, 31(5), pp. 460 - 470, 2013