ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017
8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)
Organized by
Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

# A Survey on Traffic Management Using Bigdata Analytic Tool

Prachi Sahu[1], Roja M H[1], Shwetha K S[2], Dr. Jitendranath Mungara[3]

B. E Student, Dept. of ISE, New Horizon College of Engineering, Bengaluru, Karnataka, India[1]

Assistant Professor, Dept. of ISE, New Horizon College of Engineering, Bengaluru, Karnataka, India[2]

Head of the Department, Dept. of ISE, New Horizon College of Engineering, Bengaluru, Karnataka, India[3]

**ABSTRACT:** Traffic safety is one of the main priorities of any governments. Traffic management stands as important issue to be considered for better transportation. The after effect of traffic accident has affected people one or the other way, such as the traffic being slowed down due to an accident or collision in the same road, which may lead in to a congestion. In past ten years, the worlds transformation in economic and industrial growth has let us to give attention towards traffic management. Traffic data is very huge.Big data ecosystem has the ability to store, manipulate, analyze and mine large traffic accident dataset and can drive knowledge creation that can help decision makers to reduce the number of accidents. The different methods for analyzing the traffic data is used such as using data mining, big data, comparison of clustering technique for traffic accident detection, machine learning paradigms. We need to look at these holistically when we need to design solution for tomorrow's world.

**KEYWORDS**: Big Data, data mining, clustering, machine learning.

## I. INTRODUCTION

Traffic management is one of the major concern to be solved today. What adds to the traffic pressure is the Maximum number of unplanned road network, migration growth in IT industries and increasing number of accidents. With the rising population, is the corresponding increase in number of vehicles leading to more number of accidents. Almost every day, people face traffic accident in one way or the other leading congestion in one or more lanes. Traffic data is located at the core of ITS. Leveraging the big traffic data provides a very good platform to develop Intelligent Transport System (ITS). Although, the traffic data is very huge and rich, there is a lack of useful information. Big data ecosystem has the ability to store, manipulate, analyze and mine large accident datasets and can help us in gaining insight to enhance roadway safety and crashes. Traffic safety aims at reducing the risk of killing seriously injuring of passengers using transportation system and reducing risk of damaging vehicles and transportation infrastructure. Maintaining safety requires a regular monitor and check on different transport elements such as drivers, vehicles, roads, traffic signals and deployed safety processes.

## II. RELATED WORK

TRAFFIC ACCIDENT ANALYSIS USING MACHINE LEARNING PARADIGM[1]: Traffic accidents are analyzed based on the patterns they are involved. Building a model considering such behavior and roadway accident patterns will be useful to develop road safety measures. Here, measure of formulating rules is done by considering objective surveys of the cause of accident and severity of injuries. Different machine learning paradigms is applied in modelling the severity of injury that occurred during traffic accident. The datasets include labels of person number, vehicle number, year, month, gender, alcohol usage, roadway surface condition, speed limit etc.

The five classes like number of injury, possible injury, non-incapacitating injury, incapacitating and fatal injury describes as attributes for injury severity. The different machine learning based models will help us in accurately classifying the severity of injuries. Such classification gives us the information about patters of accidents. The records

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017
8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)
Organized by
Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

of dataset forms input or output pairs where each record has associated output.Artificial Neural Networks using hybrid learning: It consist of input layers of neurons which contains one or more hidden layers and output layers of computational neuron. Backpropagation (BP) learning method with error function is used to update weights and so called as weight training but BP learning neural network is bad option because it gets trapped in local minimum which is by random initialization of weights. To reduce the problem with usage of BP better choice is using combination of BP and SCG.Decision tree: Classification problems can be solved by decision trees. The CART (classification and regression trees) model contains univariate binary decisions in hierarchy system. The best variable is chosen for splitting data into two groups at the root node and splitting happens recursively applied to each branch.

Gini index gini(T) of split data is defined by
Gini split(T) = N1 / N gini(T1) + N2 /N gini(T2)

Hybrid decision tree: hybrid method is basically integrating all learning or decision making models. So that each model depicts one or the other better set of features. In hierarchical hybrid intelligent system each level provides different and better new information. Only when all the layers work correct, overall functionality works better. The traffic datasets are fed to decision tree and records were assigned to one of the terminals nodes. Training data with node information were supplied for training the ANN finally with different number of hidden neurons of models were selected with highest classification accuracy for output class. . As we aim to analyze dataset and predict results in more accurate way machine learning paradigms also fails to provide us information without a prerecorded patterns already being fed to system.

ANALYSIS OF TRAFFIC DATA USING DATA MINING [1] is an another method, Six classification models are used to predict the severity of injury occurred during the traffic accident and then comparing these models to predict the accuracy. i)Naïve Bayes classifier: It is a simple probabilistic classifier with strong independent assumption. Probability can be calculated by using:

$$P(C/F1.....Fn) = \frac{P(C) \cdot P(F1\ldots\ldots Fn/C)}{P(F1\ldots.Fn)}$$

ii)J48 Decision tree classifier: C4.5 builds decision tree from a set of training data using the concept of information entropy. Checks for the base cases. Select the attributes, for each attribute find the normalized information gain from splitting on attribute “a”. Consider “a” be the best attribute with the highest normalized information gain. Create a decision node that splits on “a”. By splitting on “a” the sublist nodes are added as children node. iii)AdaBoost M1 classifier: It is a meta algorithm. It boosts the existing week classifier. Consider instances which are weighted equally, Apply learning algorithm, Algorithm focuses on incorrectly classified hard instances, The instances become harder sometimes softer,A series of diverse experts are generated based on the reweighted data.iv)PART (Partial Decision Tree) classifier: It is rule based algorithm. It generates a set of rules which are related to divide and conquer strategy, remove all the instance from the training collection that are covered by this rule and proceed recursively until no instance remains. PART builds its partial decision tree for the current set of instance and chooses the leaf with the largest coverage as a new rule.v)Random Forest Tree classifier: It works as a target collection of de-correlated decision trees. Choose T number of trees and N number of variables. m<<M where M is the number of input variables. From sampled Sn create a bootstrapof size n with t.choose a tree from the bootstrap and grow. select m variables from each node of growing tree at random and use them to find the best split. Grow the tree to the maximum extent and there is no pruning. Applying genetic algorithm for feature selection for accident datasets, vehicle datasets, causality dataset which helps to determine cause and severity of accident. Random forest algorithm yields best output by selecting all the attributes for classification. Analysis done using data mining techniques would not provide us with perfect results whereas it might lead to misuse of data, inaccurate data values, and security issues.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017
8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)
Organized by
Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

COMPARISION OF CLUSTERING TECHNIQUES FOR TRAFFIC ACCIDENT DETECTION [4] is a Data mining method used to analyze the previous accident records, detects the abnormal behavior in traffic flow. A traffic accident is stimulated and the effectiveness of different clustering techniques is examined for detecting accident. VANET: Idea of utilizing wireless communication in vehicles,Simulation of Urban Mobility(SUMO): This has been used to stimulate traffic that is time discrete and space continuous and produce position and velocity information, Clustering Method: It is used to group the like components together forming intracluster components and the dissimilarity between intercluster components, Exception maximization: This executes exception and maximization steps. The E step includes the latent variables and calculates exception of likelihood. The M step maximizes the expected likelihood value calculated in the E step and hence calculates parameters for maximum likelihood, Hierarchical clustering: In every point in the dataset are considered to be a cluster, close to each other are combined to form a new cluster,K-means clustering: this divides the dataset into k-clusters. Cluster center are calculated using cluster members. Calculating centers and distances continues until cluster center do not change much between successive calculations.

The concept of vehicle infrastructure integration that enable vehicle and infrastructure units to exchange data with each other. The vehicle and roadside infrastructure equipped with wireless communication interface will enable vehicle to send updates on continually changing data such as acceralation/deceleration, speed etc.

RSUs are assumed to be placed at every road segment along the highway. All the vehicle information are collected at one specific database. The collected information is analyzed to check for the anomaly situation and if anomaly situation are increasing, it is identified as an accident.Hierarchical clustering has better performance than exception maximization and k-means clustering at distinguishing accident affected vehicles from other vehicles.

TRAFFIC ACCIDENT ANALYSER USING BIG DATA [4]It is a very advanced method for analyzing the traffic data, it is an important application tool for using big data for storing, integrating and analyzing the traffic accidents using mahout data mining as a part of big data ecosystem. It consists of several functions to analyze and visualize the major traffic accident information. Traffic data is analyzed using a traffic data analyzer which is a software written in C and with the use of SSH. This analysis is done through MySQL aggregation function. It will extract the data required from the operational database and pass that data to Hadoop, the extracted data is analyzed usingmahout. When the user runs the application the application will connect to the server using SSH and a session is established with cent OS. The application provide six different analysis functions. The result is converted into map view, tabular view and chart view. The user has choice to analyses the traffic data for a specific administrated division or as a whole. The functions are;i)Common cause for accident function**,** “Get common cause for accident” will execute the sql query, which will apply the aggregate function COUNT to get the common cause for accidents and sort the results in descending order based on the number of accidents. ii)”Common time for accidents functions” ,“Get Common Time for Accidents” this function will execute the sql query which will apply the aggregate function COUNT, to get the common time for accidents and sort the results in descending order based on the time of accidents. iii)”Frequent Vehicle type participate in accident” function,This function applies the data mining techniques FP growth, using mahout to extract the frequent patterns of vehicle types that appear in accidents.iv)”Districts Accident Functions”This function executes the sql query which uses the aggregation function COUNT and SUM to count the no of accidents for each borough, it will sum up the total injuries and death for that borough.v)”Get Highest Street Death “Function , This function executes the sql query that uses the aggregation function SUM. The result contains street names and the total death which occurred in each street . vi)”Common Streets for Accident “Functions , this function executes a sql query which uses the aggregation function COUNT to get the results which contains the street name and the number of accidents that happen in each street.Developed application stores the massive traffic data on Hadoop with the parallel computing framework for processing and mining based on map reduce technique

## III. **CONCLUSION**

We consider the traffic accident datasets to analyze it in best way to bring out causes of traffic and to avoid it. Since traffic datasets are being large, big data techniques are used like data mining, analyzer applications, clustering

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 5, June 2017
8th One Day National Conference on Innovation and Research in Information Technology (IRIT- 2017)
Organized by
Departments of ISE, CSE & MCA, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

technique comparisons and machine learning algorithms. Analysis done using data mining techniques would not provide us with perfect results whereas it might lead to misuse of data, inaccurate data values, and security issues .Again a traffic analyzer application could also suggest with inaccurate results which will b unjust for the people of concern in the accident scene. As we aim to analyze dataset and predict results in more accurate way machine learning paradigms also fails to provide us information without a prerecorded patterns already being fed to system .Thus further we aim to work in providing accurate analysis and predictions with better set of features that will be affecting traffic issue using secured PYTHON platform for coding along with best classifier algorithms.

## REFERENCES

1. M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident data mining using machine learning paradigms," in Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, ISBN, vol. 1047219710, 2004, pp. 415–420
2. I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005
3. S. Krishnaveni and M. Hemalatha, "A perspective analysis of traffic accident using data mining techniques," International Journal of Computer Applications, vol. 23, no. 7, pp.
40–48, 2011.
4. M. OpenCourseWare, R for machine learning, 2015. [Online]. Available:
http://ocw.mit.edu/courses/sloan-school-of-management/ 15-097-prediction-machine-learning-and-statistics-spring-2012/ lecture-notes/a
5. H. Ibrahim and B. H. Far, "Data-oriented intelligent transportation systems," in
Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on. IEEE, 2014, pp. 322–329.
6. Nejdet DOGRU, Abdulhamit SUBASI, "Comparision of clustering techniques for accident detection", Turkish Journal of Computer Sciences, 2015.