Traffic Prediction for Intelligent Transportation System using Machine Learning

Gaurav Meena
Department of Computer Science
Central University of Rajasthan
Ajmer, Rajasthan, India
E-mail: gaurav.meena@curaj.ac.in

Deepanjali Sharma
Department of Computer Science
Central University of Rajasthan
Ajmer, Rajasthan, India
E-mail:deepz9733@gmail.com

Mehul Mahrishi
Department of Information Technology
Swami Keshvanand Institute of Technology
Jaipur, Rajasthan, India
E-mail: mehul@skit.ac.in
IEEE Member No. 96246958

Abstract—This paper aims to develop a tool for predicting accurate and timely traffic flow Information. Traffic Environment involves everything that can affect the traffic flowing on the road, whether it's traffic signals, accidents, rallies, even repairing of roads that can cause a jam. If we have prior information which is very near approximate about all the above and many more daily life situations which can affect traffic then, a driver or rider can make an informed decision. Also, it helps in the future of autonomous vehicles. In the current decades, traffic data have been generating exponentially, and we have moved towards the big data concepts for transportation. Available prediction methods for traffic flow use some traffic prediction models and are still unsatisfactory to handle real-world applications. This fact inspired us to work on the traffic flow forecast problem build on the traffic data and models.It is cumbersome to forecast the traffic flow accurately because the data available for the transportation system is insanely huge. In this work, we planned to use machine learning, genetic, soft computing, and deep learning algorithms to analyse the big-data for the transportation system with much-reduced complexity. Also, Image Processing algorithms are involved in traffic sign recognition, which eventually helps for the right training of autonomous vehicles.

Index Terms—Traffic Environment, Deep Learning, Machine Learning, Genetic Algorithms, Soft Computing, Big Data, Image Processing

I. INTRODUCTION

Various Business sectors and government agencies and individual travellers require precise and appropriately traffic flow information. It helps the riders and drivers to make better travel judgement to alleviate traffic congestion, improve traffic operation efficiency, and reduce carbon emissions. The development and deployment of Intelligent Transportation System (ITSs) provide better accuracy for Traffic flow prediction. It is deal with as a crucial element for the success of advanced traffic management systems, advanced public transportation systems, and traveller information systems. [1]. The dependency of traffic flow is dependent on real-time traffic and historical data collected from various sensor sources, including inductive loops, radars, cameras, mobile Global Positioning System, crowd sourcing, social media. Traffic data is exploding due to the vast use of traditional sensors and new technologies, and we have entered the era of a large volume of data transportation. Transportation control and management are now becoming more data-driven. [2], [3].However, there are already lots of traffic flow prediction systems and models; most of them use shallow traffic models and are still somewhat failing due to the enormous dataset dimension.

Recently, deep learning concepts attract many persons involving academicians and industrialist due to their ability to deal with classification problems, understanding of natural language, dimensionality reduction, detection of objects, motion modelling. DL uses multi-layer concepts of neural networks to mining the inherent properties in data from the lowest level to the highest level [4]. They can identify massive volumes of structure in the data, which eventually helps us to visualize and make meaningful inferences from the data.

Most of the ITS departments and researches in this area are also concerned about developing an autonomous vehicle, which can make transportation systems much economical and reduce the risk of lives. Also, saving time is the integrative benefit of this idea. In current decades the lots of attention have made towards the safe automatic driving. It is necessary that the information will be provided in time through driver assistance system (DAS), autonomous vehicles (AV)and Traffic Sign Recognition (TSR) [5].
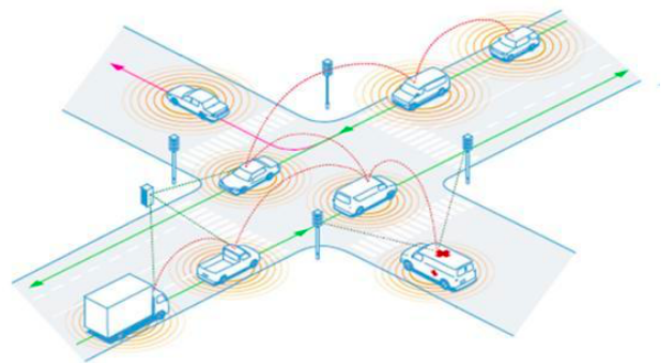

Fig. 1. Intelligent Transportation System

Although already many algorithms have been developed for

predicting the traffic flow information. But these algorithms are not accurate since Traffic Flow involves data having a vast dimension, so it is not very easy to predict accurate traffic flow information with less complexity.

We intend to use Genetic, Deep Learning , Image Processing, Machine Learning and also Soft Computing algorithms for prediction of traffic flow since a lot of journals and research paper suggests that they work well when it comes to Big-Data.

II. BACKGROUND

Intelligent Transportation system (ITS) is adopted in world congress held in Paris, 1994. the ITS has used the application of computer, electronics, and communication technology to provide traveller information to increase the safety and efficiency of the road transportation systems. The main advantage of ITS is to provide a smooth and safe movement of road transportation. It's also helpful in the perspective of environment-friendliness to reduce carbon emission. It provides many opportunities for automotive or automobile industries to enhance the safety and security of their travellers [6].

Irrespective of vehicles increases on roads, the traffic also increases. And the available road network capacity is not feasible to handle this heavy load. There are two possible approaches to resolve this issue. The first one is to make new roads and new highway lanes for the smooth functioning of vehicles. It requires extra lands and also the extensive infrastructure to maintain it, and due to this, the cost of expenditure also high. Sometimes many problems came into the network like in the urban area. This land facility is not available for the expansion of the roads and lanes. The second approach uses some control strategies to use the existing road network efficiently. By using these control strategies, the expenditure also reduces, and it is cost-effective models for the government or the traffic managers. In this control, strategies identify the potential congestions on the roads, and it directed to the passengers to take some alternative routes to their destinations. [7]

Deep learning is a part of machine learning algorithms, and it is a compelling tool to handle a large amount of data. DL provides a method to add intelligencies in the wireless network with complex radio data and large- scale topology. In DL, use concepts of a neural network, by using this feature, it is beneficial to find network dynamics (such as spectrum availability, congestion points, hotspots, traffic bottleneck. [8]

The travel time is the essential aspect in ITS and the exact travel time forecasting also is very challenging to the development of ITS. Support Vector Machine (SVM) is one of the most effective classifiers among those which are sort of linear. It is advantageous to prevent overfitting of data. SVM is great for relatively small data sets with fewer outliers. Another algorithm (Random Forest, Deep Neural Network, etc.) require more data but always came up with very robust models. SVM support linear and nonlinear regression that we can refer to as support vector regression, instead of trying to fit the most significant possible roads between two classes while limiting margin violation. Support Vector Regression (SVR) tries to fit as many instances as possible on the road while limiting margin violations. [9]

III. SPECIFICATIONS

In order to accomplish the objectives of paper, Android Studio, Java, Garmin, PHP, XML, Python and sklearn library are used.

Design

The application was created with minimum buttons, so that user can easily use the applications without any confusion. Also the interface is kept light so that the application loads fastly and doesn't create any problem to user. The geo coordinates of the application are displayed in a Toast View.

- Name of Application:- My application
- Compatibility:- Android 6.01 and above
- Memory:- 20 kb

The screen shot of the application can be easily seen in Figure 2, where you can easily observe that the interface of the application is kept minimal.
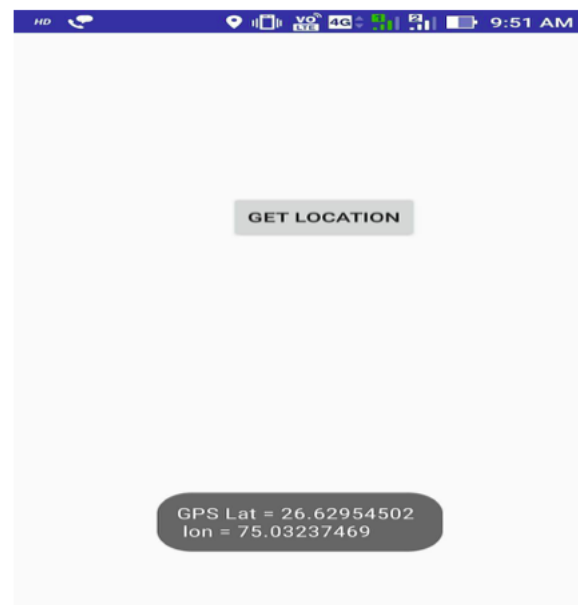

Fig. 2. Screen shot of the application

IV. IMPLEMENTATION AND RESULTS

We have applied and tested different machine algorithms for achieving higher efficiency and accurate results. To identify classification and regression we have used a Decision Tree Algorithm (DT).The goal of this method is to predict the value of the target variables.Decision tree learning represents a function that takes as input a vector of attributes value and return a "Decision " a single output value. Ii falls under the category of supervised learning algorithm. It can be used to solve both regression and classification problem. DT identify its results by performing a set of tests on the training dataset [10].

146

Outliers detection is another critical step for an accurate result, and for this, we have used Support vector machines (SVMs), which is a set of supervised learning methods that can also be used for classification and regression. The SVM is beneficial for high dimensional spaces, and it also helps in the condition where a number of samples are less than the number of dimensions. [11].

The random forest algorithm is a robust machine learning algorithm. It is defined as bootstrap aggregation. The random forest algorithm is based on forecasting models, and it is mostly used to classify the data. The bootstrap algorithm is used to generate multiple models from a single training data sets. A bootstrap algorithm has also used a sample to estimate statistical quantities. [12].

We have proposed the algorithm for predicting the traffic congestion which can be seen below:

Algorithm 1 For identifying the congested situation

---
1. Collect the traffic data in every 5 min with features:
  A. Location (Measured with GPS)
  B. Direction
  C. Speed
  D. Start-End Junction
2. Group every 5 min interval with their corresponding data.
3. Calculate the distance between each vehicle with all another vehicles within specified junction.
if the distance is less than the specific threshold between two vehicles **then**
   those vehicles are considered to be the neighbourhood vehicles
else
   Not considered as neighbour vehicles.
end if

Algorithm 2 For classifying the congested situation

---
1. This will eventually give us the matrix A.
2. Now assign 1 to $A[i, j]$
if $A[i, j] < threshold$ **then**
   $A[i, j]$ = 1
else
   $A[i, j]$ = 0
end if
3. Count $A[i, j]$=1 and label i, j as neighbourhood vehicles
4. Repeat above steps in every 5 min for 45 min
5. Plot the graph between neighbourhood vehicles and time interval.
if the neighbourhood vehicles shows an increasing graph **then**
   the traffic congestion is identified
else
   No traffic
end if

Steps Involved in implementation:-

1) Created the application which can provide us the GPS coordinates.
2) Perform the proposed algorithm
3) Evaluate the matrix for the dataset
4) Divide the the dataset into training and testing.
5) Analyse different machine learning algorithms.
6) Predict the 45 min interval parameters through machine learning algorithm
7) Conclude about the traffic congestion

Following the above steps we can implement this algorithm and can obtain the model which gives the higher accuracy of the machine learning model than the existing ones.

It is easy to coach the deep network by applying the BP methodology with the gradient-based improvement technique. Unfortunately, it's notable that deep networks trained during this method have dangerous performance. So we have not incorporated the deep learning models in my work. Also the dataset developed doesn't have many features so it won't be a sensible decision to use the deep learning and genetic algorithms. Following the proposed algorithm we have solved lot of problems like Big-data issues, also the huge dimensions of dataset is reduced which avoids the over fitting of the model.

Results

Table I shows the results of performance of the models obtained through different machine learning algorithms that are discussed in this paper. In this table we defined the various attributes like Accuracy, Precision, Recall and Time Taken.

TABLE I
EVALUATION MATRIX FOR DIFFERENT MACHINE LEARNING ALGORITHMS

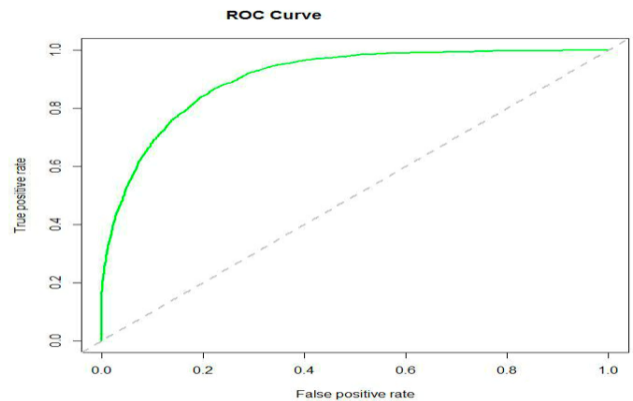| Algorithm	Accuracy	Precision	Recall	Time
Decision Tree	88%	88.56%	82%	108.4sec
SVM	88%	87.88%	80%	94.1sec
Random Forest	91%	88.88%	82%	110.1sec

Fig. 3. ROC curve for Decision Tree Algorithm

Figure 3 represents the ROC curve for the Decision Tree algorithm. In a ROC curve actuality positive rate (Sensitivity)

147

is aforethought in perform of the false positive rate (100-Specificity) for various cut-off points of a parameter. Each purpose on the mythical monster curve represents a sensitivity/specificity try resembling a selected call threshold. The area underneath the mythical monster curve (AUC) could be a live of however well a parameter will distinguish between 2 teams

V. CONCLUSION AND FUTURE WORK

Although deep learning and genetic algorithm is an important problem in data analysis, it has not been dealt with extensively by the ML community. The proposed algorithm gives higher accuracy than the existing algorithms also, It improves the complexity issues throughout the dataset. Also we have planned to integrate the web server and the application. Also the things algorithms will be further improved to much more higher accuracy.

REFERENCES

[1] Fei-Yue Wang et al. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2010.

[2] Yongchang Ma, Mashrur Chowdhury, Mansoureh Jeihani, and Ryan Fries. Accelerated incident detection across transportation networks using vehicle kinetics and support vector machine in cooperation with infrastructure agents. *IET intelligent transport systems*, 4(4):328–337, 2010.

[3] Rutger Claes, Tom Holvoet, and Danny Weyns. A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):364–373, 2011.

[4] Mehul Mahrishi and Sudha Morwal. Index point detection and semantic indexing of videos - a comparative review. *Advances in Intelligent Systems and Computing, Springer*, 2020.

[5] Joseph D Crabtree and Nikiforos Stamatiadis. Dedicated short-range communications technology for freeway incident detection: Performance assessment based on traffic simulation data. *Transportation Research Record*, 2000(1):59–69, 2007.

[6] H Qi, RL Cheu, and DH Lee. Freeway incident detection using kinematic data from probe vehicles. In *9th World Congress on Intelligent Transport SystemsITS America, ITS Japan, ERTICO (Intelligent Transport Systems and Services-Europe)*, 2002.

[7] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu. Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2):68–75, 2017.

[8] C. Zhang, P. Patras, and H. Haddadi. Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys Tutorials*, 21(3):2224–2287, thirdquarter 2019.

[9] Chun-Hsin Wu, Jan-Ming Ho, and D. T. Lee. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, 5(4):276–281, Dec 2004.

[10] Yan-Yan Song and LU Ying. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130, 2015.

[11] Yiming He, Mashrur Chowdhury, Yongchang Ma, and Pierluigi Pisu. Merging mobility and energy vision with hybrid electric vehicles and vehicle infrastructure integration. *Energy Policy*, 41:599–609, 2012.

[12] Jason Brownlee. Bagging and random forest ensemble algorithms for machine learning. *Machine Learning Algorithms*, pages 4–22, 2016.