

Dense Vehicle region Management through smart Traffic Routing using ML Approach

Kislay Kishore, Adhithiya N, Dr.S. Ramamoorthy
SRM Institute of Science and Technology, Chennai, India
kr5621@srmist.edu.in
ag1090@srmist.edu.in
ramamoos@srmist.edu.in

ABSTRACT

Road traffic congestions continue to remain a major problem in most of the cities around the world resulting in massive delays, increased fuel wastage and monetary losses. Traffic congestions can occur due to poorly planned road networks, poor traffic management in critical hotspots of the city, bad weather conditions or due to an increase in the volume of vehicles. The traffic data in smart cities can easily be obtained with the help of sensors. A machine learning model would help us in predicting the traffic congestion in advance by providing the appropriate data to the model. The machine learning algorithm used to predict the traffic congestion is Random Forest Algorithm due to its high accuracy and robustness.

Keywords

Traffic Congestion, Machine Learning, Random Forest Algorithm, Big Data

Introduction

Roadways are widely used for commercial as well as financial purposes. Road networks provide a cheaper and easier means of providing transportation and hence road traffic congestion is a major issue that occurs in various road networks. Traffic congestion leads to stress, excess fuel consumption leading to increase in pollution in the environment etc. The factors that are taken into consideration for a roadway to be congested are decrease in the average speed of vehicles, increase in the travel time or increase in the volume of the vehicles. Free flow of traffic can be achieved with proper planning of the road networks or enhancing the traffic management in zones where traffic congestion can occur.

The main objective of this report is to design an algorithm using machine learning to identify traffic congestion beforehand and hence allow for proper management of the traffic congestion resulting in reduced duration of traffic congestion. The prediction of the congestion should be accurate. The live traffic data can be obtained with the help of sensors, cameras, social media etc. Smart cities have abundance of live traffic data due to the presence of sensors on roads which can be used for finding trends in the data obtained from these sensors.

Literature Review

In the previously proposed models, the authors have used genetic, deep learning, soft computing and machine learning to analyze the big data for the transportation system and predict the congestion accurately for the big data. A decision tree algorithm is used to predict the traffic congestion and it will provide a single output value. Traffic data is collected for every 5 minutes and features such as location, direction, speeds are taken into consideration which help in determining the distance between vehicles. A threshold distance is set between vehicles and these values are compared to identify whether the vehicles are neighborhood vehicles. The time interval set is 45 minutes and a graph is plotted between the neighborhood vehicles and the time interval which will help in determining congestion. Due to lack of features in the dataset, the author has not used deep learning and genetic algorithms in the proposed model.

The previous models used logistic regression for predicting congestion. The machine learning based algorithm provides an easier prediction of congestion for a particular road network. The suggested model's accuracy was compared to various other algorithms for the same dataset, to find the algorithm which has the better accuracy and

has a better performance. The algorithm used had constant complexity hence making it usable in devices with less computation capability and with less resources.

One of the previous models uses the vehicle location to predict the traffic flow congestion at selected road network. Deep learning approach is used for the prediction of congestion. With multiple approaches taken to calculate the average speed of the vehicles for a particular road network, this data is provided to the deep learning model for predicting the traffic congestion for a particular road network. Due to multiple approaches taken in the model, the performance of the model is not. The model used predicts the traffic congestion for a road network with multiple intersections. The Hidden Markov Model is used to predict the traffic congestion at the intersection based the flow of vehicles for a particular period. The flow of vehicle is calculated based on the vehicles entering and exiting the intersection and the count is compared to the previous period for prediction of congestion. The proposed model had better error tolerance and performance compared to SVM. The models performance increased with the increase in the features and size of the training data. The predicted result was similar to the forecasted result.

One of the models proposed a system for real-time traffic analysis and prediction. Analysis of the Twitter data includes extraction of tweets from Twitter and then classifying the tweets based on location. The tweets are classified into different categories based on the hashtags used on the tweets by the user. The model performs a sentiment analysis on the list of relevant tweets and the traffic is predicted using the majority vote rule. The model uses random forest algorithm for prediction due it's better performance with big data and avoiding overfitting of the model. Alternative routes are also suggested for a given source and destination.

From the literature survey we were able to infer that the accuracy of the proposed algorithm is high and the deep learning approach provides easier analysis of the big data and this approach has a wider scope due to the increase in the autonomous vehicles and driver assistance systems. The proposed algorithm provides better solution to the big data issues compared to other algorithms and reduces the complexity of the dataset. The practical application of the deep learning and genetic algorithms based on current datasets is difficult. The deep learning approach provides a more accurate traffic prediction given large and complex data.

Proposed Work

Architecture

In the suggested model, the original dataset containing the large traffic data is split into training set and test set. Later the data is sent for pre-processing and data cleaning. After pre-processing the dataset is given to the system for analysis.

In the system the large dataset is divided into small numbers of data sets. Later the system uses Random Forest algorithm on the data sets to analyze the data and predict the output. Then a conclusion a decided using the output obtained from the system.

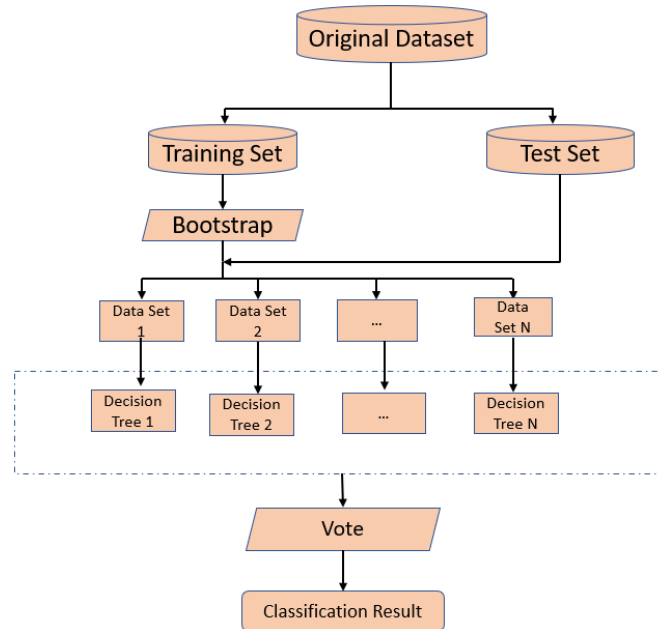


Figure 1. Architecture Diagram

Methodology

The pre-processed data is retrieved from the original dataset and is divided into training and test datasets. After training and testing, the data is divided into N numbers of dataset randomly with the help of random forest algorithm. Each dataset is processed individually using decision tree algorithm. After all the results are obtained, the best result is voted and classified and hence, an output is obtained.

Algorithm

The proposed model uses the Random Forest algorithm to predict congestion for a given road network. Random forest algorithm generates 'n' number of decision trees and the results of these decision tree's are combined to provide a more precise overall result. Due to the abundance of traffic data present in smart cities with the help of sensors, the decision tree algorithm will face with the overfitting issue i.e., the decision tree generated from the model will be generated to its complete depth. But Random Forest algorithm overcomes this issue and provides with a low bias, low variance model. The accuracy of the Random Forest algorithm will not be affected with the change in the dataset, hence making the model flexible. Random forest algorithm provides row sampling and feature sampling which are provided as inputs to the decision trees and helps in determining the key features in the given dataset.

Implementation

The proposed model uses the Random Forest algorithm to predict congestion for a given road network. Python and Jupyter Notebook are used as the platform for performing the required analysis and model prediction. Various python libraries are used to perform the various functions, which are part of Jupyter notebook. Pandas library is used for the data analysis and manipulation of the data. Matplotlib is used in the data exploration part of the project. Sklearn is used for splitting the dataset into train and test data as well as is used to train the model with the suggested algorithm.

The implementation of the proposed system was done by first pre-processing the data. We first try to identify any inconsistent or null values in any of the features present in the data. The pre-processing stage of the data lead to the discovery of some features having inconsistent data, which were identified and replaced with the appropriate values. The pre-processed data was used in the analysis part of the project. The various graphs plotted were used to identify

the peak congestion periods for a single day. With the help of sklearn library, the data is divided between training and testing data in the appropriate size. The training data was used for training the Random Forest model with the number of decision trees being set to default and the model accuracy was also identified after the model was trained with the testing data. The accuracy of the suggested model is also calculated for comparison with that of other suggested models.

Results

The suggested model's accuracy for the given data and compared with another model is depicted below:

Table 1. Accuracy Results

Algorithm	Accuracy
Decision Tree	83.5%
Random Forest	88%

The confusion matrix for the suggested model is depicted below:

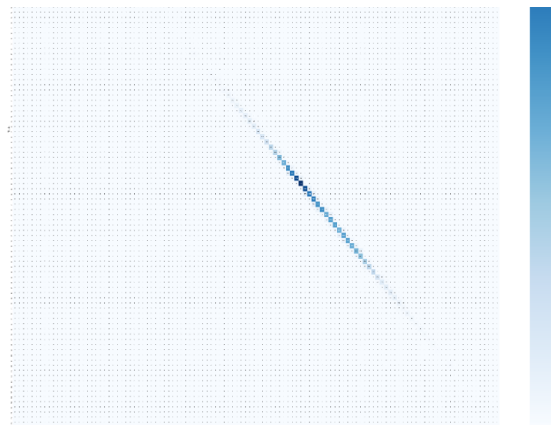


Figure 2. Confusion Matrix

The graph plotted for the confusion matrix depicts:

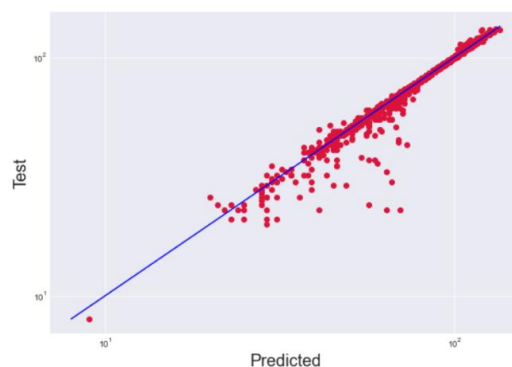


Figure 3. Scatter Plot Graph

Discussions

The test results shows that the accuracy of the decision tree is 83.5% whereas the accuracy of the RF model is 88%. Due to the abundance of traffic data in smart cities, the suggested problem falls under the category of a big data problem. Decision tree algorithm tend to face overfitting because the tree is generated to its complete depth. This issue is overcome by the use of Random Forest algorithm as it has low bias, low variance. The algorithm also has a better accuracy than the decision tree algorithm. The change in the dataset will not affect the accuracy of the model. Due to the generation of multiple decision trees by Random Forest algorithm, there is an increase in the time taken for predicting the congestion. The deep learning approach overcomes this issue but based on current data, this approach is not viable.

Conclusion

The increase in the number of vehicles in the cities has caused a steep rise in the congestions which is a cause of multiple issues. With the help of the current technologies, it is possible to predict such congestions with better accuracy and tackle this issue. The model suggested in this paper can predict the congestion by taking into consideration the average speed and the average flow of vehicles. The future work for the current model is to predict for multiple roads and compare every road and suggest the alternate routes to the end user. With the help of smart city data, the model can further be expanded into providing better solutions to the issues occurring due to traffic congestions.

Limitations and Future Studies

We are planning for constantly training the data set for accurate predictions. The Applications will look forward for improvements by collaborating with Road Traffic Management Authorities for obtaining the live data of the traffic. We are also planning to incorporate more parameters in this system such as suggesting the alternate route to the user or alter them about the possible traffic congestion in the future. These will be more useful for general public.

Acknowledgement

We would like to express my special thanks of gratitude to our Guide, Dr. S. Ramamoorthy and our Professor & Head Dr. B. Amutha who gave us the golden opportunity to do this wonderful project on the topic Dense Vehicle region Management through smart Traffic Routing using ML Approach, which also helped us in doing a lot of Research and we came to know about so many new things. We are really thankful to them.

Secondly, we would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame.

References

- [1] Kalaga, Ramachandra Rao (2013), "Measuring Urban Traffic Congestion
doi: [http://dx.doi.org/10.7708/ijtte.2012.2\(4\).01](http://dx.doi.org/10.7708/ijtte.2012.2(4).01)
- [2] K.R. Priyarakshitha (2018), "Traffic Congestion Study with a Reality Approach- A Review"
ISSN: 1314-3395
- [3] B.Karthika, N.UmaMaheswari, R.Venkatesh, "A Research of Traffic Prediction using Deep Learning Techniques"
ISSN: 2278-3075
- [4] Noureen Zafar, Irfan Ul Haq, "Traffic Congestion based on Estimated Time of Arrival"
<https://doi.org/10.1371/journal.pone.0238200>

- [5] "Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment", Amr Elfar, Alireza Talebpour, Hani S. Mahmassani
<https://doi.org/10.1177%2F0361198118795010>
- [6] "Traffic Prediction Based on Ensemble Machine Learning Strategies with Bagging and LightGBM", Huiwei Xia; Xin Wei; Yun Gao; Haibing Lv
10.1109/ICCW.2019.8757058
- [7] Gaurav Mena, Deepanjali Sharma (2020)," Traffic Prediction for Intelligent Transportation system Using Machine Learning" doi:10.1109/icetce48199.2020.9091758
- [8] Aditya Krishna K.V.S, Abhishek K, Allam Swaraj, Shantala Devil Patil, Gopala Krishna Shyam (2019)," Smart Traffic analysis using Machine Learning"
ISSN: 2249-8958
- [9] Suguna Devi, T.Neetha (2017),"Machine learning based traffic congestion prediction in a IoT based smart city"
e-ISSN: 2395-005 p-ISSN: 2395-0072
- [10] Shuming Sun, Juan Chen, Jian Sun (2019)," Traffic congestion prediction based on GPS trajectory data"
doi:<https://doi.org/10.1177/1550147719847400>
- [11] ZHAO Shu-xu, WU Hong-wei, LIU Chang-rong (2019), "Traffic flow prediction based on optimized Hidden Markov Model"
doi: 10.1088/1742-6596/1168/5/052001
- [12] S Minu, R.L Lija (2018), "A case study on analysis of Traffic flow in Chennai city"
ISSN: 2455-7137
- [13] Sougata Maji (2017),"Traffic congestion and Possible solutions A case study of Asansol
ISSN: 2321-9467
- [14] Kovuru Sridevi, T.Ganesen, B V S Samrat, S Srihari (2019),"Traffic analysis using random forest algorithm considering social media platforms"
ISSN: 2277-3