Indian Journal of Science and Technology, *Vol 9(17), DOI: 10.17485/ijst/2016/v9i17/87334, May 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Cost Effective Road Traffic Prediction Model using Apache Spark

M. Prathilothamai, A. M. Sree Lakshmi and Dilna Viswanthan

Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amrita University, Coimbatore - 641112, Tamil Nadu, India.

Abstract

Objectives: We proposed a cost effective model to predict the traffic to inform the public about the current traffic condition to all persons who are entering the same lane. **Analysis:** In real time application like traffic monitoring, it needs to process huge volume of data in huge size. We analyzed the traffic prediction using the current technologies Apache Hadoop and Apache Spark framework. Spark is processing the 10 Terabytes of data in half-a-second. The main uniqueness from our approach is that we can predict the road traffic using Spark within half-a-second. **Findings:** Road traffic is predicted using Ultrasonic and PIR sensor within a half second. The proposed system uses the vehicle count and speed to predict the traffic condition. Existing system using hadoop will predict the traffic in few seconds. Whereas in the proposed system performance gets increased using Spark. Therefore, the results are more helpful in finding the road traffic condition. **Improvement:** The proposed system predicts it in a half a second by using Spark whereas the existing system predicted the road traffic by consuming more time.

Keywords: Big Data, Hadoop, Road Traffic Prediction, Spark, Real-Time Applications

1. Introduction

Improvements in road traffic congestion have a great devastating effect in every human life. Hence road traffic prediction helps to reduce the transportation delays. Several metrics have been proposed in the literature for predicting road traffic congestions. In[1], author designed an online web based traffic management system for light and heavy weight vehicles that shows the density of the traffic. The system uses edge detection methodology to detect the edges of the objects in the videos. Further a feature extraction technique is used to identify the objects in certain bad conditions such as night vision, shadows and someother occlusions that occur between time gap of the vehicle passage. In[2], paper focused mainly on an intelligent traffic control system by implementing it with embedded system. It is done by counting methods to detect the objects and thereby simultaneously counting the emergency vehicles for controlling the signals at the traffic based on the priority. In[3] proposed a technique for estimating traffic congestion levels from the velocity of traffic data collected using mobile sensors. Traffic data is collected using GPS device and a webcam. Further a decision tree technique is applied on to the mobile sensors from which the vehicle movement patterns are extracted. A sliding window technique is used in order to retrieve the continuous average velocities of the moving pattern, which then classifies the traffic congestion level. In[4], proposed a Global Positioning System (GPS) based traffic predicting technique in which the public is made aware of the traffic situations by sending an alert message in their mobile phones which shows the level of the traffic congestion to take alternative routes. The major drawback of the system is the necessity of GPS in all vehicles which is not possible all ways.

In[5], the algorithm used for classification was decision tree. The data was collected using GPS device and webcam. The velocity of the vehicle determines the congestion level. In[6], proposed a system for predicting congestion using semantic web technology, it considered all types of data such as heterogeneous stream data, real

**Author for correspondence*

and historical data together with live data for predicting the congestion status. Here the system predicts the congestion correctly but the data processing is very complex and time consuming. Further in[7], proposed a solution for controlling traffic congestion using sensors which is used within the infrastructure of the road, as well as at the same time data is collected from vehicles too. In[8], tried a technique for background subtraction and modeling by introducing feature selection algorithm. Real Boost algorithm is used select the appropriate features for every pixel. In order to delete the objects from foreground scene the algorithm acts as a very good tool. In the system proper results were derived for feature selection from the images .In this paper, in[9], explains about how to calculate the probability that is estimated over a certain interval of time. In[10], explained about the real time algorithm for finding about the foreground-background segmentation. In[11], explained about the foreground and background differentiation for analysing the video. In[12], explains about the framework which can be used for vehicle detection tracking systems. This approach is mainly used for vehicle counting applications. In[13], explained about how to count the moving objects from the video by means of video processing. In[14], explained about the vehicle tracking and the classification. The optical flow and the use of camera oriented parameters are used for the processing. In[15], explained about the semantic web technology used in predicting the traffic congestion. In[16], proposed a technique to detect the smoke inside a room using video data. In order to improve the performance of smoke detection and to reduce false detection of smoke, it proposed two techniques one is using Gaussian mixture model for detecting the color pixels of smoke and dynamic features to detect the area of the smoke and to extract it. In[18], proposed a technique for detecting the count of the traffic flow and identification of particular vehicle. For the system to behave very faster and accurately it uses background subtraction method thereby it tracks the vehicle for a particular period of time. In[17], used Adaboost method of classification in order to detect the vehicle on hardware embedded system. The processing of images are done using different resolution structure. In[19], proposed a system for the road traffic congestion by deriving a ontology from the sensor data collected. The datas are also preprocessed. The system mainly consumes a lot of time in road traffic prediction.

In our proposed system, integrating and interrelating observations from multiple data sources such as sensor data, we could infer that the traffic prediction is an initial essence in controlling the traffic in many of the congested cities. Many ways are adopted for traffic prediction; the main aim in prediction is to inform the people about the traffic condition. Thus in many of the data stream applications, prediction is assessed by relating the past and observed values.

In most data streaming applications, prediction is done by means of interrelating the current and past data or by selecting some rules for predicting the future without any accurate data. These techniques are designed for high processing of data and correlations of unprocessed data from sensors like magnetic sensor, ultrasonic sensor. Hence we show that the integration of sensor with Apache spark can be used for interpreting the semantics of data which improves the overall structure of traffic prediction system. To this end, we exploit apache spark and its applications in real time data processing as a way to annotate and interpret the data. We extend the latter work by 1. Presenting the prediction system by means of parallel processing of sensor as well as video data 2. Focusing on prediction by using ontology 3. Alerting the public about the traffic condition. Thus, the revealing part focuses towards the evidential approach that will be able to solve the above mentioned risks. Therefore in the paper, Timely Prediction of Road Traffic Congestion by processing of sensor data in Apache Spark is the main approach.

2. Existing System

In the existing system, Ontology is the main approach used and it predicts the road traffic congestion status. Mainly for processing the data, the unprocessed data are collected using Ultrasonic sensor and Passive Infrared sensor and data preprocessing is done. Based on the collected sensor data, an ontology based rules are derived. After the processing the collected sensor data, the data like vehicle count, inbound traffic and also outbound traffic are obtained. The main drawback in this existing system is that it takes a lot of time in processing the data as well as it doesn't predict output accurately and also prediction takes few seconds.

3. Proposed System

In our proposed system, the severity of the road traffic is predicted by using Apache Spark for increasing the accuracy in prediction. Initially, we have collected data using Ultrasonic as well as Passive Infrared sensors. After processing the collected sensor data, the dataset provides the Speed of the vehicle, actual count of the vehicle, duration in which vehicle has passed by the sensor area. Then those collected sensor data are converted into a CSV (Comma Separated Values) file. The CSV file is then processed in Apache Spark inorder to predict the traffic congestion. The prediction of the road traffic is classified as High congestion, Medium congestion and Low congestion. As well as we have integrated it with Apache Zeppelin using a console application in C#, and the traffic data are retrieved in parallel using a Zeppelin notebook configured for Apache Spark. In our proposed system, we have focused on the idea of traffic prediction based on mainly two factors such as Speed of vehicle and number of vehicle. If the vehicles are passing with a low speed and if the get gradually reduced it can be inferred that it has a high traffic congestion. In another case, if the vehicles are passing at speed than the minimum speed and the speed doesn't get reduced it can be inferred that it has a low traffic congestion congestion. Thus our prediction system gives an accurate result to the public.

3.1 Proposed System Architecture

In Figure 1 the sensor data is collected and they are converted to CSV file to retrieve the count of the vehicles passed and the speed of the vehicles within a particular duration. After the retrieval of the parameters, the file is loaded into Apache Spark and are used to predict the traffic congestion status as high congestion, medium congestion and low congestion.

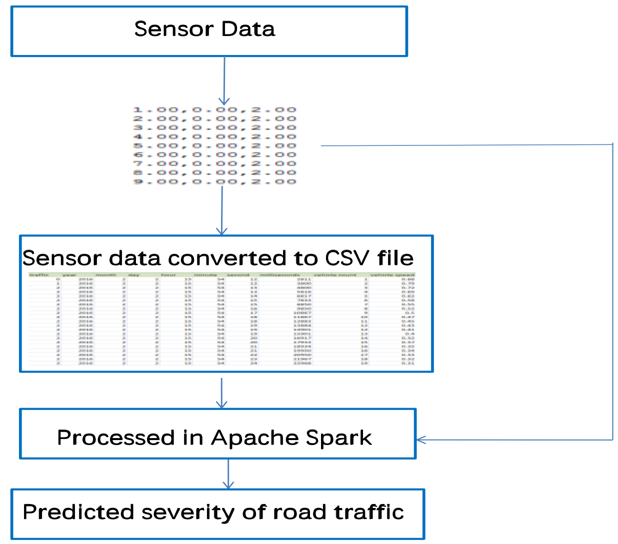

Figure 1. Architecture of proposed system.

3.2 Sensor Data Collection

Initially the sensor data is collected by using the Arduino board connected with two sensors, namely, the Passive Infrared sensor and the Ultrasonic sensor which are used in collecting the data. The PIR sensors are mainly used to detect the human interruption in the area mounded by the sensor. Hence the use of PIR sensor draws a huge devastating effect on predicting the real time traffic condition. The main proof in using the PIR sensor is that, they are mainly used in detecting whether the person has entered the room or not, which proves it has a high sensitivity in detecting the persons. And also the PIR sensors are known as the movement detectors.

Those sensors measure the amount of infrared light that is radiating from the objects moving in the area. For detecting the persons, a threshold will be set earlier and if any of the observed measurements exceeds then it will be considered as a movement. But the PIR sensors don't tell us either the number of persons entered the surveillance area or the distance of how close are they to the sensor. Here comes the role of Ultrasonic sensor. The ultrasonic sensors are used in our project to detect the vehicles. The ultrasonic sensors works with the principle of pulse waveforms. Those pulsed energy transmitted at two known end points makes an incident angle which allows us to calculate the vehicle speed. The speed is calculated as:

$$\text{Speed} = \sum_{i=1}^{n} (X_i - Y_i)$$

Where X_i = Beam separation distance, and Y_i = Time taken to traverse the beams. The summation of all the beams that is projected from the sensor and reflected back to sensor are given by the above formula. That is, the speed is calculated by observing the time at which the vehicle crosses each of the emitted beams. Also, since the beams takes a certain distance to travel, it can be used in calculating the speed more easily. Hence, the people like obstacles are identified using PIR sensors whereas the vehicle count is detected using the ultrasonic sensors.

3.3 Convert Sensor Data into CSV file

For collecting the real time road traffic data we have used the PIR and ultrasonic sensor, but for processing those data they are connected using the arduino board to the system. Further the collected sensor data is successfully loaded. Even after collecting the data, we need to visualize those data taken from the sensors. In Figure 2, the technique used is the conversion of the data collected by sensor into a CSV (Comma Separated Values) file format on the computer which can have gloriously specific column names such as year, month, day, hour, minute etc. This has a stable advantage in building a built-in table for processing. Therefore for processing task, firstly arduino has to shakehands with the processing technique. For processing the sensor data Processing 2.0 software is used. Both arduino and processing software needs to run together for successfully collecting data using the sensors and for converting them into CSV file. After processing the sensor data into CSV file, we further used, Apache Spark, for traffic prediction.

3.4 Sensor Data Processing in Apache Spark

In Apache spark we processed the CSV file for predicting the road traffic congestion. Initially we loaded the CSV file. The CSV file contains data such as the number of vehicles. For retrieving the results from the CSV file Sparksql is used. The main component of Apache Spark is the Spark SQL, where queries are processed on the sensor data (CSV file) to detect the Number of vehicles, human count, and traffic status. In order to predict the traffic by classifying it as low, medium and high traffic we used classification algorithm in Apache spark. Classification is an algorithm which is used to classify the dataset into different classes. Further, Decision tree is used, to predict output from those classes. Decision tree acts as a predictive model in decision analysis which can be used to explicitly and visually represent decision making. So the Decision tree fits well into our system where traffic is needed to be predicted into different classifications as high, medium and low traffic. For processing the Decision tree first the spark MLlib library is imported into spark which a machine is learning library provided by Apache spark. Further the CSV file is loaded where it is processed by using the Decision tree. Therefore, we will be able to retrieve the decision tree based on the CSV datas available. In our system the decision tree is constructed by analyzing the features such as time and vehicle count, i.e., if within a particular time limit if the vehicle count is less then it is predicted to be as low traffic, and further traffic is predicted as high and medium based on the vehicle count values.

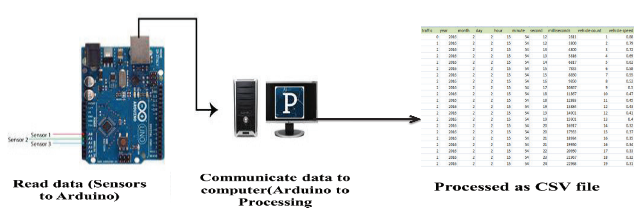

Figure 2. Converting sensor data into CSV file.

4. Result

In our proposed system, we have collected real time data in traffic peek time and also off-peak time. We have uploaded those data in our system and traffic is predicted which is shown in Figure 3 and Figure 4.

We have collected the real time traffic data in peek time for the duration 1 minute and processed it in our system. Our system predicts the heavy traffic in less than a half second. In Figure 3,

X-axis represents number of vehicles and Y-axis represents vehicle speed in 1 minute duration and 10 meter distance. It shows that speed of vehicles get decreased and also number of vehicles also get increased, with that we can make a conclusion that results in high traffic.

We have collected the real time traffic data in off-peak for the duration 1 minute and processed it in our system.

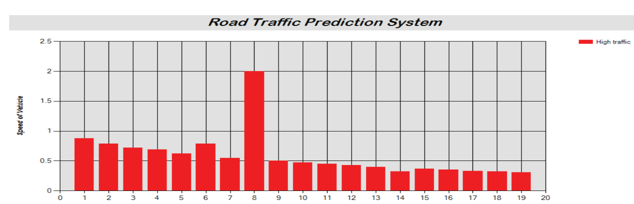

Figure 3. High traffic prediction.

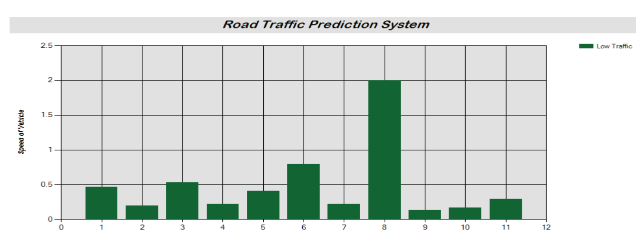

Figure 4. Low traffic prediction.

Our system predicts the traffic status as low in less than a half second. In Figure 4, X-axis represents number of vehicles and Y-axis represents vehicle speed in 1 minute duration and 10 meter distance. It shows that speed of vehicles are not gradually decreased with that we can make a conclusion that results in low traffic.

5. Conclusion and Future Works

The steady increase in the road accidents which are caused mainly due to unawareness of the traffic condition has created a high demand for traffic prediction systems. In our system, high traffic is predicted accurately with vehicle speed and count, but for low traffic prediction vehicle speed and count is not enough. Due to random increase and decrease of vehicle speed with the driver's performance. With that we can't make prediction that the traffic is low. In future, we will take the video of traffic. In our future work by means of parallel processing of sensor as well as video data in spark to give accurate prediction of traffic status as high, medium and low. In that we will face the challenges in video processing in spark and to predict traffic status within a half-second.

6. References

1. Goudar R, Singh A, Kumar A. Online traffic density estimation and vehicle classification management system. Indian Journal of Science and Technology. 2014 Apr; 7(4).
2. Parthasarathi V, Surya M, Akshay B, Siva KM, Vasudevan SK. Smart control of traffic signal system using image processing. Indian Journal of Science and Technology. 2015 Jul; 8(16).
3. Thianniwet T, Phosaard SC, Pattara-Atikom W. Classification of road traffic congestion levels from GPS data using a decision tree algorithm and sliding windows. Proceedings of the World Congress on Engineering; London, UK. 2009 Jul 1–3.
4. Priyadarshana YHPP. GPS assisted traffic alerting and road congestion reduction mechanism. Proceedings of Technical Sessions; 2013.
5. Choocharukul K. Congestion measures in Thailand: State of the practice. Proceedings of the10th National Convention on Civil Engineering; 2005 May. p. TRP111–8.
6. Lecue, F. Predicting severity of road traffic congestion using semantic web technologies. The Semantic Web: Trends and Challenges. Springer International Publishing; 2014. p. 611–27.
7. Neha. Implementation and comparison of the improved traffic congestion control scenario. International Journal of Advanced Research in Computer Science and Software Engineering. 2014 Apr; 4(4). ISSN: 2277 128X.
8. Parag T, Elgammal A, Mittal A. A framework for feature selection for background subtraction. IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2006.
9. Schapiro RE, Singer Y. Improved boosting algorithms using confidence-rated predictions. Machine Learning. 1999; 37(3):297–336.
10. Kim K. Real-time foreground–background segmentation using codebook model. Realtime Imaging. 2005; 11(3):172–85.
11. Banarase-Deshmukh RG, Kumar P, Bhosle P. Video object tracking based on automatic background segmentation: A survey. International Journal for Technological Research in Engineering. 2014 Aug; 1(12).
12. Kamiya K. A framework of vision-based detection-tracking surveillance systems for counting vehicles. Proceedings of the International Conference on Soft Computing; 2012.
13. Raghtate G, Tiwari AK. Moving object counting in video signals. International Journal of Engineering Research and General Science. 2014 Apr-May; 2(3):2091–730.
14. Ambardekar AA. Efficient vehicle tracking and classification for an automated traffic surveillance system. Proceedings of the International Conference on Soft Computing, Pro Quest; 2007.
15. Lecue F. Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City. Web Semantics: Science, Services and Agents on the World Wide Web. 2014; 27:26–33.
16. Wang Y, Weiyi Lv. Indoor video-based smoke detection. Study Program for a Degree of Bachelor of Science in Computer Science; 2011.
17. Alefs B. Embedded vehicle detection by boosting. IEEE Intelligent Transportation Systems Conference, ITSC'06; 2006.
18. Venugopal KR, Patnaik LM. Moving vehicle identification using background registration technique for traffic surveillance. Proceedings of the International Multi Conference of Engineers and Computer Scientists; 2008.
19. Prathilothamai M, Marilakshmi, Majeed N, Viswanathan V. Timely prediction of road traffic congestion using ontology. Proceedings of the International Conference on Soft Computing; 2015 Dec 8. p. 331–44.