

SEMANTIC BASED MOVING WEIGHTED AVERAGE MODEL FOR ROAD TRAFFIC PREDICTION

M. Prathilothamai

School of Computer Science and Engineering,
Vellore Institute of Technology, Chennai, India.
prathilothamai.m2014@vit.ac.in

V.Viswanathan

School of Computer Science and Engineering,
Vellore Institute of Technology, Chennai, India.
viswanathan.v@vit.ac.in

Abstract

In this emerging world, most of the peoples are wasting their precious time in travelling. The drastic increase in the number of vehicles is due to an increase in population. In this contribution, we propose an ontology-based moving weighted average model to forecast traffic congestion. Real-time traffic data was collected from the camera and with the help of an Arduino board using Ultrasonic and PIR sensors. The proposed system uses a moving weighted average model to calculate the threshold distance and speed from the recent history of traffic data like vehicle count, the distance between the front and back vehicles, and vehicle speed. The system interprets the semantic relationship of the collected data with the threshold data to predict traffic congestion. Due to the scalability of the data, we conduct experiments in Apache Spark. Thus, it will increase the speed of traffic prediction. The system will insist on the traffic status as a location-based message to assist the people in planning their commute.

Keywords: Traffic prediction, Naïve based classification, Ontology, Traffic alert, Apache spark, Arduino board.

1. Introduction

People reach everything at their fingertips in the emerging world. Drawing them to a congested environment is a significant issue due to traffic congestion. Traffic congestion makes people rush to reach their destination on time. Such a rush can take even a life. It may lead to accidents. In such situations, we need to address the pressure created by traffic on people. It depicts the need for an accurate traffic prediction system (TPS). Suppose TPS provides a proper route to avoid traffic congestion, and it helps people reach safely to their destination on time. Route guidance and traffic congestion information of TPS is helpful for goods transportation, tourism etc.

Consider various systems like detecting Vehicle movement and traffic congestion prediction, part of route guidance's critical application. Traffic congestion can be reduced by improving the road infrastructure, providing alternate route guidance near traffic congestion areas, and immediately initiating action. Machine learning approaches are primarily used in data streaming applications to predict future trends by relating historical data with real-time data. Thus, the proposed system uses the classification approach to classify traffic congestion based on recent history and real-time data.

The system uses the semantic interpretation of data collected from the camera and PIR and ultrasonic sensor data in the Apache Spark environment to improve traffic prediction. Apache spark is chosen for the environmental set-up to avoid the disk latency issue in Hadoop and best suits the real-time application. In this system, the First level prediction model is built based on video and sensor data in the parallel processing environment. The second level semantic interpretation of sensor and video data is considered to improvise the prediction model's performance. At last, traffic congestion status is broadcasted as a location-based message to assist people in planning their commute.

2. Related Work

[Rayan Goudar, et al. (2014)] designed a web-based application to assist people in travel by showing the traffic density in the route. The system uses video data for measuring traffic density. They apply various combinations of edge detection techniques and feature extraction methods to remove the shadows in the night-time images to identify the vehicles. [V.Parthasarathi, et al.(2015)] develop an embedded system for traffic control depends on the emergency vehicle count. The system controls the traffic signal based on emergency vehicle existence, and it treats emergency vehicles like an ambulance as a higher priority one. Thus, it provides a better route for the

ambulance to reach its destination at the right time. However, it fails to address environmental factors such as weather, fog, etc.

[Priyadarshana (2013)] designed a system for traffic prediction which uses information from a GPS device mounted on the vehicle. Traffic congestion prediction is broadcasted to the public as an alert which is helpful for them to decide the route. However, all vehicles must have GPS devices to make use of this system. To predict traffic hybrid approach is proposed by [Shen and Wynter (2012)]. It uses GPS and sensors to collect the vehicle details like the number of the vehicle, speed of the vehicles, etc. It uses the local link as GPS and global link as sensors subregions to predict traffic congestion. The system did not consider the time, day and weather details. [Freddy Lecue, et al.(2014a)] proposed a semantic-based traffic forecast system to predict the severity of the traffic. The system builds the model by historical data; testing is done on live data. It has a high processing time. [Neha (2014)] introduce a system to predict traffic congestion and control traffic. The system uses sensor data gathered from vehicles to report traffic light servers. It processes the collected data to predict the traffic status, which helps to identify an alternate route. Vehicles get traffic status information from wireless traffic light servers, which help to find an alternate route.

[Freddy Lecue, et al (2014b)] designed a traffic density prediction model incorporating semantic of traffic data. The system uses the following data: road infrastructure, weather information, city event schedule, vehicle details from roads sensors, and historical data to forecast traffic density. However, it is not suitable for urban areas. [Nicholas G.Polson, et al. (2017)] proposed a deep learning-based traffic flow prediction model. It shows traffic flow non-linearity due to accidents and road damage and considers the following events: a football game in Chicago and a snowstorm event. It is suitable for short-term traffic predictions. [Jie Xu, et al. (2015)] created an online framework to predict traffic by considering live traffic and recent history. It best suits traffic prediction in unusual situations like accidents, sudden rainfall, etc. [Byon, et al. (2019)], designed a traffic prediction system that uses GPS, sensor, camera and traffic monitoring sites data. It designs a filter to ignore flaws in the sensor data and applies sampling techniques in different frequencies for traffic prediction. [Xianglong Luo, et al. (2019)], proposed a traffic prediction model for holidays. It builds based on historical data using Discrete Fourier Transform (DFT) and Support Vector Machine (SVM). The system lacks considering environmental factors and weather information which has a significant impact on traffic flow prediction.

[JSharmila, et al. (2019)] proposed the hybrid approach based SVM technique for the travel time prediction technique. This author measures the SVM parameter through the particle filter, and identified parameters are used for prediction. In this method, the SVM technique combines the kernel functions, loss functions and regularization parameters. Similarly, the static and dynamic data evaluation is done through the SVM technique. [Dashan Geo (2001)] used the SVM technique to detect moving vehicles in an automatic traffic monitoring system. Due to effective generalization and efficient classification accuracy, the SVM technique is preferred for the vehicle monitoring system. In this technique, histogram-based data extraction is applied to predict the data in the moving vehicles. SVM identifies the high intensity moving region. However, the integrated technique provides better accuracy compared to the traditional SVM technique. [Azzedine Boukerche (2020)] proposed a neural network model. In that input data experiments in different network structures for making the final prediction. The Convolutional Neural Network is the popular technique based on this neural network. The neural network-based classification belongs to the non-parametric models. When handling a large volume of traffic data, this model automatically adjusts the parameters without changing the base model. This neural network model needs a large volume of data for the training process. If the data size is not sufficient, this technique produces poor prediction results.

Similarly, It does not suit a dynamic change in road and traffic conditions. The system accuracy and rate of calculation depend on the number of layers. When the number of layers is high, the calculation time also takes more computational time to complete the task. Similarly, the accuracy rate depends on the number of neurons in each layer. If a lesser number of neurons is used in a particular layer, the accuracy rate is lesser. The convolutional neural network (CNN) model is suitable for handling spatial data in big road traffic data. The CNN technique is used for extracting the spatial features from the traffic data. The major limitation of the CNN model is that the spatial data can only extract from geographical information.

Traffic density prediction is a research area where various AI techniques are used to build the model. Short-term traffic prediction is made by considering various traffic parameters. Most traffic prediction is made based on historical data, but few are predicted based on real-time traffic data. [Mahmuda Akhtar, et.al (2021)] analysed various existing models based on machine learning methodologies and discussed the strength and weaknesses of the models. In recent research in traffic prediction systems [K. Lee, et al. (2021)] shows massive data production. Nowadays, deep learning methodologies are applied for traffic prediction. Deep neural networks (DNN) are suitable for finding good patterns and features with spatiotemporal relationships. This paper analyzed DNN techniques on Short-term traffic (STTP) and listed the challenges and research problems on STTP.

The above study addresses the following: traffic congestion prediction, vehicle detection, traffic flow prediction and vehicle tracking system, etc. and also it uses different techniques for data collection, which is of high cost. People get traffic congestion information from the corresponding web or mobile application. The

proposed system uses sensors which is less in cost and cameras for the data collection, and it broadcast predicted traffic condition as a location-based message to the public.

3. Proposed System

The proposed system proposed a road traffic prediction model. The system collects the traffic data from the ultrasonic and PIR sensor and camera and processes it in Apache Spark. In this work, the semantic of the sensor and camera location is interpreted with the geospace ontology to broadcast the traffic severity as the location-based message to the people in the same lane. Protege tool is used to create traffic ontology to identify the sensor and camera location. The system predicts traffic in the location based on the weighted average model. Differentiate impact on the traffic with different traffic classes. Weights are assigned to each of the classes.

3.1. System architecture

System architecture shows the parallel input processing in Apache Spark, as shown in figure 1. Environmental factors may affect the sensors and cameras, and thereby system may receive inconsistent data, which leads to incorrect prediction. On a few occasions, like rainy seasons, foggy cameras can't capture vehicle details, but sensors provide details. Semantic rule inferences are applied to the sensor and camera data to improve the sensor and camera data. People plan their commute by receiving location-based SMS as the alert message.

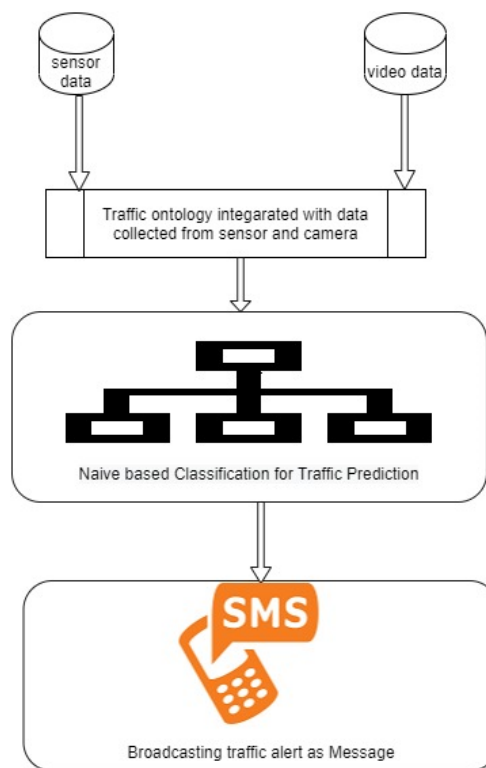


Fig.1 Architecture of the proposed system

3.1.1. Map-Reduce framework configuration

In the proposed system, Apache Spark distributes the input data across map nodes, and it shares the Traffic ontology among map nodes and reduce nodes in the Apache Spark environment. The system interprets the semantic rule inferences on sensors and video data using shared ontology. It also calculates the threshold of vehicle speed, count and distance using the weighted average model on historical data. Naïve Bayes classification builds a classification model on each map node based on threshold values of the vehicle data and semantic interpretation of the real-time data. Reduce nodes merge the models generated by all map nodes to create a final model.

3.2 Data collection and processing

Data is collected from the Coimbatore district in both peak and peak-off duration with the help of sensors and cameras. It collects the data during weekdays in all four lanes of ettimadai to the marapalam stretch. Processing of sensor data is shown in 3.2.1.

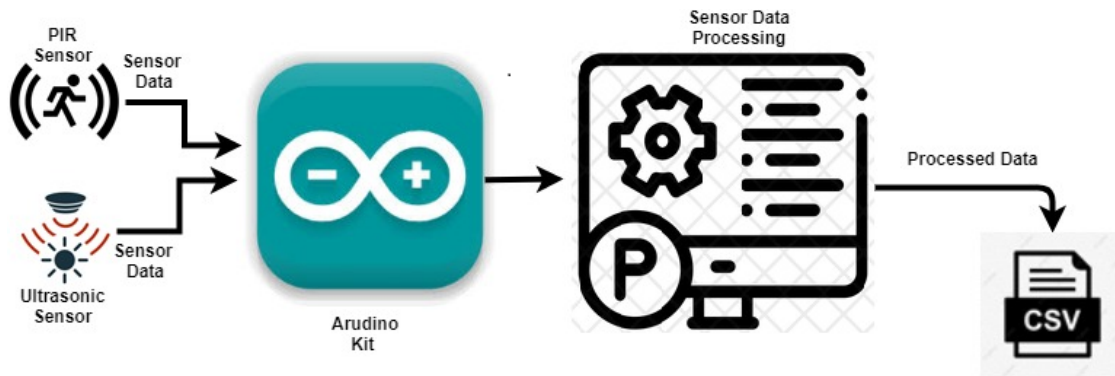


Fig.2. Data collection using Arduino board and Sensor

3.2.1. Sensor data processing

Ultrasonic and PIR sensors are connected to the Arduino board to collect the data, as shown in Figure2. The processing software converts into a CSV file. PIR sensor detects the motion of any object, which is used for detecting the vehicle movement; ultrasonic sensor plays a vital role to count the number of vehicles passed in that duration. PIR sensor detects objects, but it does not count the objects. An ultrasonic sensor works based on the sound propagation principle. Ultrasonic sensors use in the application where we need to count the objects, speed of the movement. It works by emitting ultrasonic waves to the objects and reflecting as an echo. Both PIR and ultrasonic sensors are used to detect the vehicle and count the number of vehicles. The speed of the vehicle is calculated as shown e.q.1

$$Speed = \sum_{i=1}^n (A_i - B_i). \quad (1)$$

where A_i refers transmission waves, and B_i refers receiving waves. It aggregates the difference of all waves emitting from the sensor and reflects back to the sensor is to find the average distance of the vehicle. Time is calculated by finding the time difference between emitted waves. Thus, the speed of the vehicle is measured by e.q.2

$$Vehicle\ speed = \Delta\ distance * \Delta\ time \quad (2)$$

Speed is calculated based on vehicle passage time and wave distance in motion. Hence PIR detects objects, and Ultrasonic detects moving objects such as metal, glass etc. Thereby count difference shows the number of pedestrians which in turn vehicle count also finalized. Ultrasonic sensors usage in traffic has plenty of advantages as listed below: Ultrasonic waves are acoustic waves, so they reflect on materials regardless of color or surface; It detects the vehicle in a foggy and rainy environment; It works well in dark environments such as in the nighttime also; It is less in cost and consumes less power. In the proposed system, Both PIR and Ultrasonic sensors data process in parallel, as shown in Fig.3. It converts the processed data into comma-separated files (CSV). It has the following details: Vehicle speed, vehicle count, and vehicle passage time on the sensor area.

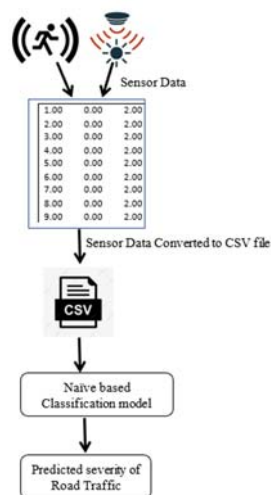


Fig.3. Data collection using Arduino board and Sensor

3.2.2. Camera data processing

Real-time traffic video processing steps captured by the camera is shown in Fig.4. Real-time traffic data is time-series data. However, it has a pattern with infrequent white noise due to uncertain events like accidents and road damage due to environmental factors like rain. Hence classification based on probability like the Naïve Bayes classification is suitable.

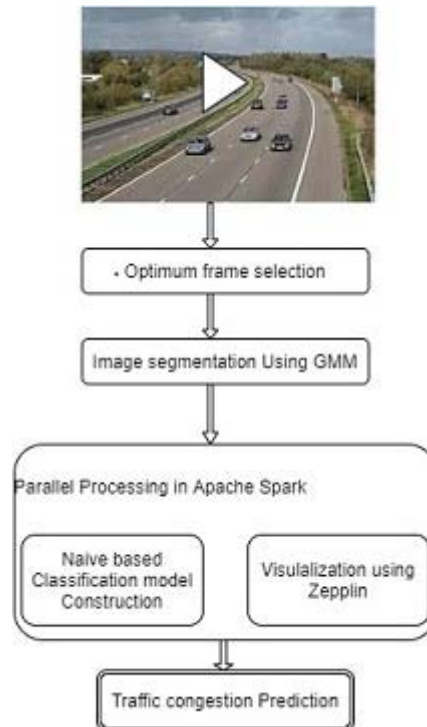


Fig.4. Road traffic prediction using video data

Video data is segregated as the sequence of frames. The system uses an optimum frame selection algorithm to choose specific frames rather than all. Gaussian mixture methodology is used for foreground subtraction. In this work, various techniques were used for vehicle detection and counting. Among those techniques, the Gaussian mixture model (GMM) is better in processing time and performance. Different variations on GMM experiments in that system selects the variant which has taken minimal time for processing. In further, it uses the morphological filter to remove noise in the foreground. i.e., to remove pedestrian, roadside parked vehicle. The system uses blob analysis to detect the vehicles, and thereby it counts the vehicles. After the vehicle detection, it calculates the Euclidian distance between pixels of one vehicle and the other to find the distance between the vehicles. In Apache Spark, both processed sensor and video data is fed into it. Both data are processed in parallel and constructed the classification model, and visualization in map using Zeppelin app. The system predicts the Traffic density with the help of a classification model and ontology.

3.3. Incorporating semantics of data

Even though considering both video and sensor data for traffic density prediction performance is not up to the mark. Traffic ontology is created on top of both sensor and video data to improve its performance, as shown in Fig.5. In addition, System refers to the geospace ontology to provide location-based SMS in the proper location. Uncertain events in the road lead to the wrong prediction because it is not considered in the threshold calculation by the weighted average model, which is based on historical data. Though System creates the ontology on top of the sensor and video data and traffic ontology has rule inferences for uncertain events, semantic interpretation improves traffic density prediction.



Fig.5. ontology for traffic

Semantic Rule inferences are based on the threshold value of vehicle speed, count and distance between vehicles. Though traffic data is time-series data, the moving weighted average model (MWAM) is used to find the threshold values. Since traffic impacts on different traffic classes are varied, different weights are assigned in MWAM to find the threshold. For instance, high traffic has an impact rather than medium and low traffic—so different weights assigned for low, moderate and high traffic as 1,2 and 3, respectively. Threshold calculation is based on the following e.q.3

$$\text{Threshold_value}_t = \frac{\sum_{i=0}^{S-1} x_{t-s+i} \times w_i}{\sum_{i=0}^{S-1} w_i} \quad (3)$$

Where S is a time interval to calculate moving weighted average, i.e., every 5 mins threshold value will be calculated, and W_i represents recent history traffic class weightage, i.e. it may be 1 or 2 or 3. X_t is the observed data at time t. Sample semantic inferences are listed below:

- Traffic.Highway.Laneno.Time.average speed of vehicles \leq Threshold vehicle speed && Traffic.Highway.Laneno.Time. number of vehicle $<$ threshold count && Traffic.Highway.Laneno.Time.distance $<$ threshold distance \rightarrow Traffic. Highway.Laneno.Time.Status.Add (heavy traffic congestion)
- Traffic.Highway.Laneno.Time. average speed of vehicles \leq Threshold vehicle speed \parallel Traffic.Highway.Laneno.Time.Total Count of vehicles = threshold count && Traffic.Highway.Laneno.Time.distance = threshold distance \rightarrow Traffic. Highway.Laneno.Time.Status.Add (moderate traffic congestion)
- Traffic.Highway.Laneno.Time. average speed of vehicles $>$ Threshold vehicle speed && Traffic.Highway.Laneno.Time.Total Count of vehicles \geq threshold Count \rightarrow Traffic.Highway.Laneno.Time.Status.Add (low traffic congestion)

3.4. Traffic alert MMS

Nowadays, time is precious; people are not ready to spend valuable time in traffic, but it is unavoidable. The proposed system provides alerts about the traffic as a location-based SMS for the registered users. The system consumes less data for registration, but there is no need for an internet facility for getting an alert. Blue Mix is embedded into Apache-spark to create an MMS as an alert as shown in Fig.6. It broadcast the MMS to the registered users on the lane during uncertain events on the spot. The registered users make use of the internet to find the alternate route to save time.

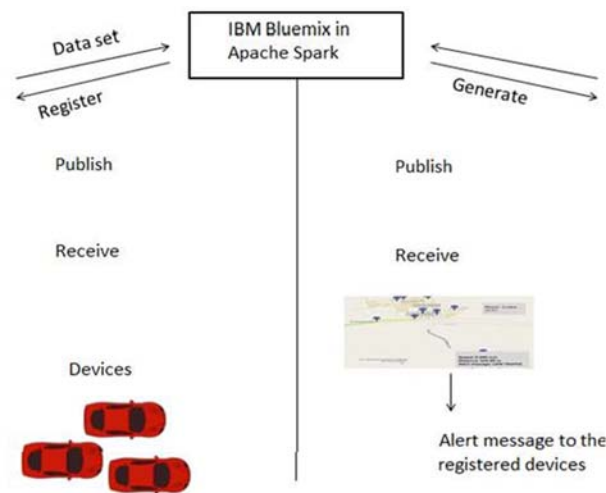


Fig.6. Traffic alert message

4. Results and Discussion

The traffic predictions are made in the following cases: with sensor data collection; with video data collection; combination of both sensor and video data by incorporating semantics of both. The system performance is measured and compared in all the above cases in terms of precision, recall and accuracy. Traffic alert message is also shown in this section.

4.1. Traffic prediction based on sensor data

In Fig.7, X-axis represents number of vehicles and Y-axis represents vehicle speed in kilometers. It shows that speed of vehicles get decreased when number of vehicles get increased accordance with threshold vehicle speed and threshold vehicle count which is calculated based on weighted moving average formula i.e e.q.1. with that we can make a conclusion that results in high traffic.

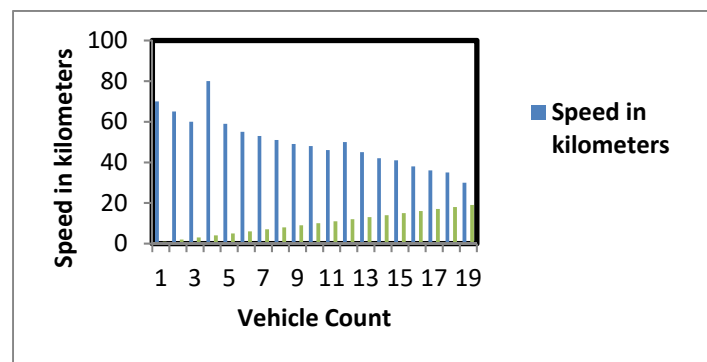


Fig.7. High Traffic prediction of sensor data processing

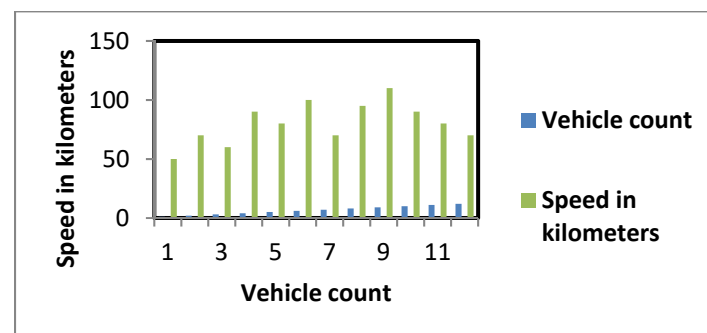


Fig.8. Low Traffic prediction of sensor data processing

Fig.8. shows that speed of vehicles is increased drastically while increase vehicle count accordance with threshold vehicle speed and threshold vehicle count which is calculated based on weighted moving average formula i.e e.q.1 that we can make a conclusion this result in low traffic.

4.2. Traffic prediction based on video data

Fig.9. and Fig.10. Shows the traffic prediction based on video data. In fig.9., X axis represents the vehicle count and Y axis represents the distance between vehicles in meters. From the fig.9. it can be inferred that the distance between vehicles is getting decreased while increasing vehicle count with threshold vehicle count which is calculated based on weighted moving average formula i.e. e.q.1 that shows that it results in High traffic.

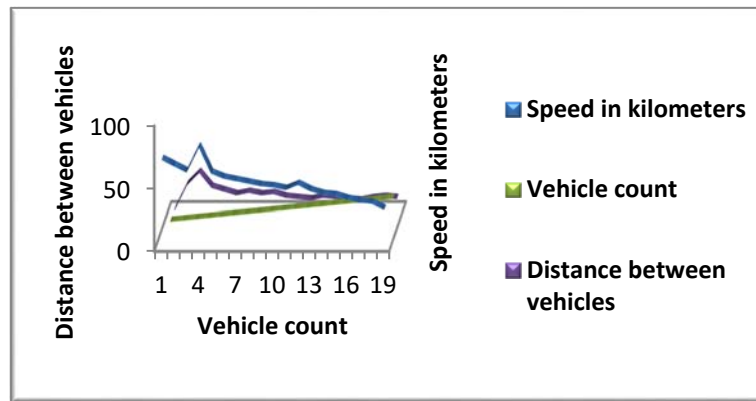


Fig.11. High traffic prediction based on semantic interpretation

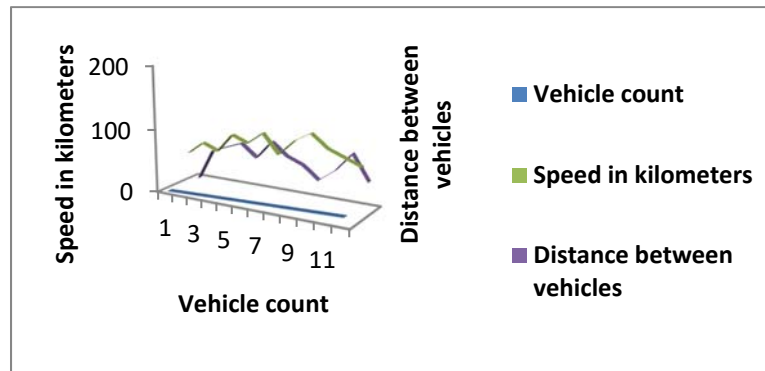


Fig.12. Low traffic prediction based on semantic interpretation

Fig.12 shows that while number of vehicles gets increased, speed as well as distance between the vehicles lies above the threshold value which is found based on weighted moving average formula. It shows that it is low traffic prediction.

4.3 Prediction accuracy, precision and recall

System considers data which is observed on 90 days in both peak time and peak off time. Sensor data and video data are acquired parallel and processed in Apache spark for identifying traffic condition which results in 83.8% accuracy. In order to find the goodness of the classifier the following metrics are used: precision and recall. Precision implies that it is a fraction of relevant instances of that class among predicted result. Recall implies that it is a fraction of relevant instances of that class among the test data. Fig.13. shows that moderate traffic precision is very low and it depicts that classification is very poor in moderate class. It provides irrelevancy in most of the time due to imbalance in the class distribution that is the reason system accuracy is less. To improve the accuracy, system integrated semantics of sensor and video data. Figure.14 show that precision of high traffic is good in addition with other two classes which means that relevancy of the classification get improved. Semantic interpretation of data improves the system accuracy as 90.3%.

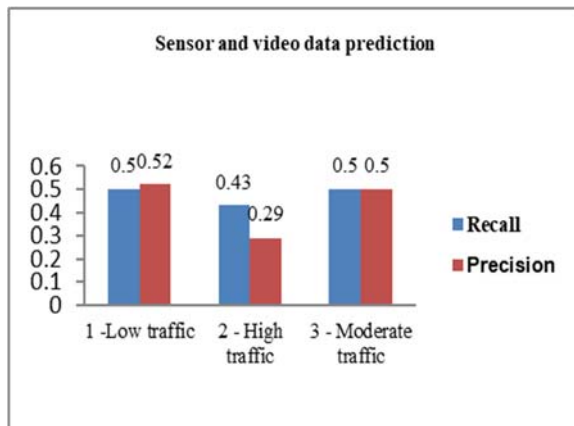


Fig.13. Precision, Recall of prediction based on both sensor & video data

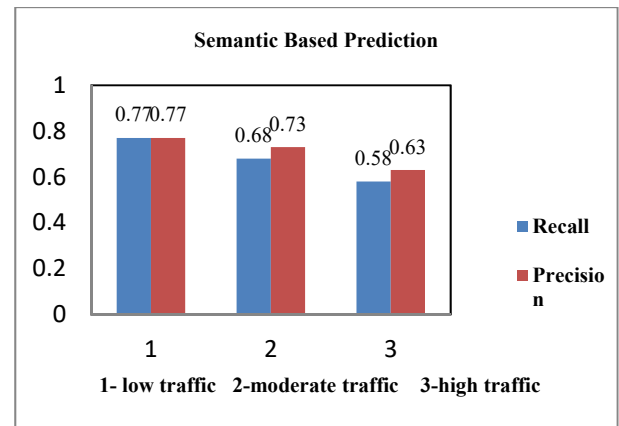


Fig.14. Precision, Recall of prediction based on semantic interpretation of both data

4.4 Location based alert message

System uses geospace ontology to broadcast the location based message by identifying latitude and longitude of sensor and camera. Fig.15 shows the sample Location based alert message on High traffic and Fig.16 shows the sample Location based alert message on low traffic.

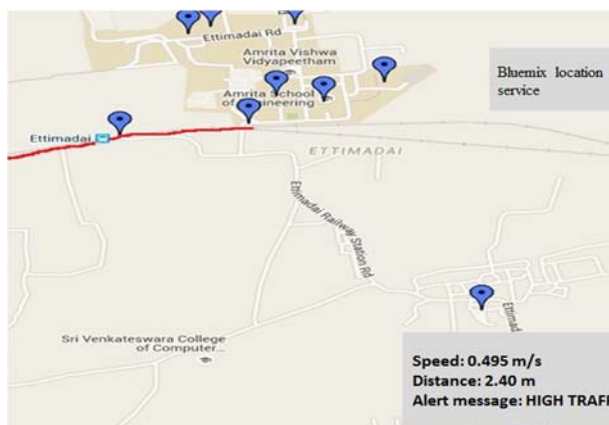


Fig.15. Location based alert message on High traffic



Fig.16. Location based alert message on Low traffic

5. Conclusion and Future Enhancement

The traffic alert is received ahead; people can commute travel plan accordingly. Traffic density prediction needs to be dynamic, and implementation must be a cheaper one. Thus, proposed system uses sensors like PIR, Ultrasonic sensors and cameras to collect real-time data. Both sensor and camera data are processed parallelly in Apache spark and construct the naive based classification model using the moving weighted average model. Traffic ontology is embedded to improve the performance of the traffic prediction system. Geospace ontology is integrated to broadcast traffic alert MMS as a location-based message. The proposed system achieves 90% accuracy. For further improvement, the dataset is analysed and finds the class imbalance in the training dataset. In future, class imbalance in the traffic dataset will be addressed to improve its accuracy. The system fails to consider the planned events like weekend, festival time traffic. The future work concentrates on weekend and festival time traffic data. Thereby public will receive accurate information about the traffic density to plan their commute.

6. References

- [1] Azzedine Boukerche, Jiahao Wang (2020), "Machine Learning based traffic prediction models for intelligent transportation systems", Computer networks.
- [2] Byon, Y.J., Shalaby, A., Abdulhai, B. (2019), Chung-Suk Cho, Hwasoo Yeo and Samah El-Tantawy, "Traffic Condition Monitoring with SCAAT Kalman Filter- based Data Fusion in Toronto, Canada ", KSCE Journal of Civil Engineering vol. 23(2), pp.810-820.
- [3] Dasha Gao, Jie Zhou, and Leping Xin (2001), "SVM-based Detection of Moving Vehicles for automatic traffic monitoring", 2001 IEEE Intelligent Transportation systems conference proceedings, PP 25-29.
- [4] J. Xu, D. Deng, U. Demiryurek, C. Shahabi and M. v. d. Schaar(2015), "Mining the Situation: Spatiotemporal Traffic Prediction with Big Data," in IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 4, pp. 702-715.

- [5] K. Lee, M. Eo, E. Jung, Y. Yoon and W. Rhee (2021), "Short-Term Traffic Prediction With Deep Neural Networks: A Survey," in IEEE Access, vol. 9, pp. 54739-5475.
- [6] Lécué, Freddy (2014), "Predicting severity of road traffic congestion using semantic web technologies.", The Semantic Web: Trends and Challenges. Springer International Publishing, pp.611-627.
- [7] Lécué, Freddy (2014), "Predicting severity of road traffic congestion using semantic web technologies.", The Semantic Web: Trends and Challenges. Springer International Publishing, pp.611-627.
- [8] Mahmuda Akhtar, Sara Moridpour (2021), "A Review of Traffic Congestion Prediction Using Artificial Intelligence", Journal of Advanced Transportation, vol. 2021, Article ID 8878011, 18 pages.
- [9] Neha (2014), "Implementation and Comparison of the Improved Traffic Congestion Control Scenario", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4.
- [10] Nicholas G. Polson, Vadim O. Sokolov (2017), "Deep learning for short-term traffic flow prediction", Elsevier: Transportation Research Part C: Emerging Technologies, Vol.79, pp.1-17, 2017.
- [11] Priyadarshana, Y. H. P. P (2013), "GPS Assisted Traffic Alerting and Road Congestion Reduction Mechanism.", Proceedings of Technical Sessions, Vol. 29.
- [12] Rayan Goudar, Abhijeet Singh, Abhijeet Kumar (2014) "OnlineTraffic Density Estimation and Vehicle Classification Management System", Indian Journal of Science and Technology, Vol. 7, Issue 4.
- [13] R. B Sharmila (2019), Nagendra R.Velaga, Akhilesh Kumar, "SVM based hybrid approach for corridor level travel time estimation", IET Intelligent transport systems, PP 1-11.
- [14] V. Parthasarathi, M. Surya, B. Akshay, K. Murali Siva, Shriram K. Vasudevan (2015), "Smart Control of Traffic Signal System using Image Processing", Indian Journal of Science and Technology, Vol .8.
- [15] W. Shen and L. Wynter (2012), "Real-time road traffic fusion and prediction with GPS and fixed-sensor data," 2012 15th International Conference on Information Fusion, Singapore, pp. 1468-1475.
- [16] Xianglong Luo, Danyang Li, and Shengrui Zhang (2019), "Traffic Flow Prediction during the Holidays Based on DFT and SVR," Journal of Sensors, vol. 2019, pp.1-10.

Authors Profile



M.Prathilothamai, working as Assistant Professor in the Department of Computer Science and Engineering at Amrita University, Coimbatore, India. She received B.Tech degree from Bharathiar University, Coimbatore, India in 2004, and ME computer science and engineering from Anna University, Chennai, India in 2009. She is doing PhD in Vellore Institute of Technology, Chennai, India. Her areas of research include Big Data and Semantic Web. She is specialized in automata theory and operating systems. She has published papers in Scopus indexed journals. She enrolled as a Co-Principal Investigator of CSI minor research grant.



VISWANATHAN VADIVEL received the Ph.D.degree from Anna University, Chennai, India, by contributing his ideas to the field of semantic web technologies and social media marketing. He has teaching experience of over 20 years in the field of computer science. He is currently a Professor with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai. He has authored articles in semantic web technologies for renowned publications. His research interests include data mining, semantic web, and social network analysis.