Averaging Based Predictive Modelling for Traffic Congestion in IoT

Nandkumar Kulkarni
Department of Information Technology, Smt. Kashibai Navale College of Engineering, Pune, India
npkulkarni.pune@gmail.com

Dr. Dnyaneshwar Mantri
Department of E&TC, Sinhgad Institute of Technology, Lonavala, India
dsmantri@gmail.com

Dr. Pranav Pawar
Department of Computer Engineering, MITADT, Loni, Pune, India.
pranav21684@gmail.com

Prof. Neeli Rashmi Prasad
International Technological University (ITU), San Jose, USA
neeli.prasad@ieee.org

***Abstract*—**The Internet of things (IoT) is the system of physical devices, vehicles, and other items embedded with electronics, software, sensors, actuators, and connectivity which empower these objects to accumulate and interchange data. IoT allows objects to be recognized or controlled distantly without human involvement. This result in enhanced efficiency, precision and economic advantage. Traffic blocking is bursting as foremost challenge in every established as well as emerging countries and it needs immediate attention. The amalgamation of machine learning and IoT, Vehicular Adhoc Network (VANET) makes the traffic management more intelligent. Many researchers have proposed numerous answers for covering detecting, estimating and avoiding traffic congestion in a handful of established nations. These solutions are not suitable from Indian perspective because of mixed traffic conditions, population. This paper proposes a novel traffic congestion prediction technique based on averaging under heterogeneous conditions. The proposed system uses real time and historic traffic data for informing accurately on road congestion preceding the journey.

Keywords—IoT; Traffic, Predictive; Averaging; Congestion; Modeling

I. INTRODUCTION

Increasing urbanization in developing countries like India has also led to many problems like pollution, population burst, burst in traffic and so on. The many fold increase in traffic congestion is one of such major problem. Municipal provinces in utmost of the emerging nations are facing foremost challenges in traffic regulatory in present epochs, furthermore India is no exception. It has seen express evolution into economy, bringing about in vehicle proprietorship levels budding at a considerably rapid proportion. For example, the no of certified vehicles in India's six major metropolises went up by 7.75 times during 1981 to 2001, while the population increased only by 1.89 times. Thus, the growth of motor vehicles was almost four times faster than the growth of population. The World-Bank stated that the financial damages incurred on account of obstructive and inferior roads only run as tall as $6 billion for every year in India. Though at hand numerous solution alternatives like road-framework expansion, Transport Scheme Controlling measures and overcrowding estimation, expert use-cases like in Intelligent-Transportation-System (ITS) ascertained to be proficient mode to lessen congestion in technologically advanced nations like U.S.A. Two of the foremost pillars of ITS are the Advanced Traveler Information System (ATIS) and the Advanced Traffic Management System (ATMS).

ATIS deals with consumers' instantaneous traveler info, empowering them to take better as well as up-to-date tourism judgments those will bring about well-organized scattering of voyagers to diverse routes utilizing varied means. ATMS become aware and conveys transportation conditions to governing station via a communication system. It customizes supreme transportation regulatory mechanism by aggregating the accessible transportation info. ATIS as well as ATMS have need of clear-cut evaluations of instantaneous traffic, also forecasting traffic so as to maintain smooth traffic flow. The modus operandi intended for traffic assessment as well as prediction can be assembled into data-driven and model-based methods.

Traffic congestion is popping its head as one of the most major problems in developed as well as developing countries and hence an effective model for the prediction of traffic congestion is the need of the hour. The current traffic prediction is based on real time monitoring of the vehicles and is not efficient and accurate as expected. So in order to deal with this problem this paper proposes easy and efficient traffic prediction algorithm which will be able to monitor and control the traffic based on certain predictions. The predictions will be based on historic as well as real time data and thus enabling us to control and manage the traffic efficient.

II. LITERATURE SURVEY

The following section consists of the studies carried out by various authors followed by an analysis of the current existing modules, the techniques employed by them and their comparative strengths and weaknesses. In [1], the author focuses on solving the traffic management problems only in urban areas. The technique used by the author to predict congestion is based on routes information and he divides the factors affecting traffic into two categories: physical volume of dense traffic meant for present-day and exterior happenings. The major focus is on the latter category. The key trends that contribute to the traffic congestion problems in India are reviewed in [2]. The backdrops and short comings of the existing policies and programs have been reviewed and a set of recommendations has been proposed to tackle these challenges. Only the weather data and factors affecting traffic management are taken into consideration in [3] and the system used is Hadoop and R library. To project and investigate traffic blocking, a set of experimental anticipated climate values have been utilized.

In [4], the authors have suggested the need for a coherent urban transportation policy to avoid the ad hoc interventions

in order to eliminate chaos and confusion. The authors have a more realistic, practical and all-inclusive approach to tackle the problem of urban traffic congestion. The monitoring and modeling system developed in Netherlands for prediction of traffic and guidance in managing the same are described and analyzed in [5]. The fuzzy-AR technique has been scrutinized and been processed by a clustering algorithm in order to predict and manage congestion in high speed networks in [6]. A cost effective, real time application to notify the travelers of the current traffic conditions on a given particular patch of road has been proposed in [7]. The performance of the prediction algorithm has been improved by using Apache Spark and Hadoop framework. In [8], the authors have tried to redefine the global parameters used for traffic predictions i.e. density and speed, in order to predict traffic congestion more accurately under heterogeneous conditions of traffic. In [9], a technique to predict traffic especially under Indian conditions has been developed. The authors have sharply scrutinized the existing systems in [10] and have also performed a vague comparative analysis between these systems. The table I is an overview of the above survey.

III. EXISTING SYSTEMS

A lot of researchers and traffic engineers have studied and investigated the different existing traffic prediction technologies. But none of them have done a comparative analysis of all the technologies that can be found in the literature. Following survey aims to study and compare all the models that have been developed to predict traffic. The studies have been scrutinized to compare the prediction -method, -horizon, -scale, -environment (urban or expressway), -computational effort and -accuracy.

In the study different variables are predicted, the most important being flow (vehicles/hour, also known as intensity), density (vehicles/km) or occupancy (percent of time a detector is occupied by vehicles), minimum travel time on a link or on an entire route and mean speed (km/h) of all vehicles that have passed during a certain time interval. Two modus operandi utilized in statistics to discover degree of accurateness of the anticipated data is RMSE (Root-Mean-Square-Error) and MAPE (Mean-Absolute-Percentage-Error). RMSE quantifies inaccuracy amongst datasets by contrasting a projected-, experimental- (recognized) value. MAPE too, is a measure of the prediction accuracy of a forecasting in statistics. The advantage of MAPE is that it usually describes accuracy as a percentage whereas RMSE outputs the result in the same unit as that of the input data.

This survey lists out all the existing methods that have been described in various literatures. They can broadly be categorized as: naive, parametric and non-parametric (Ref. Fig. 1).

A. *Naive Methods*

The phrase 'naïve' is relatively instinctive, on the other hand can be inferred as 'short of any model hypotheses'. These are one of the most widely applied practical methods owing to their low computational effort and comparatively easy execution. However, compared to any of the parametric or non-parametric methods, the prediction accuracy of naïve methods is very low.

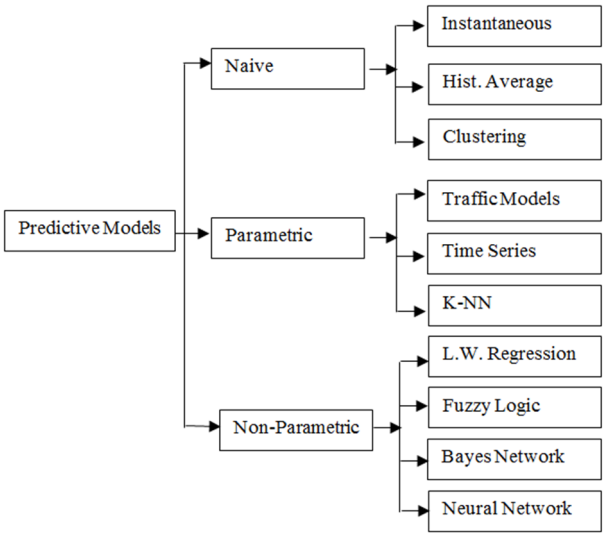

Fig. 1. Taxonomy of Prediction Models

B. *Parametric Models*

The term 'parametric' is used to specify that only the parameters or the variables of the technique are found using provided input data; the organization or the construction of the model is predetermined. These structures are best suited for the implementation of traffic simulation models, where the knowledge of traffic processes can be extracted easily. Also, this technique provides a framework for modeling unseen or unpredictable events such as accidents. This is very helpful for Digital Terrain Model (DTM) purposes. Also, when compared with non-parametric models, the amount of input data required here is very less. With respect to accuracy as well as computational efforts, some parametric methods outperform most of the others.

C. Non Parametric

Initially, the common assumption was that traffic flow follows a chaotic pattern and non-parametric methods are founded on this very theory of chaos. The transitions in the patterns in non-parametric methods are non-linear and deterministic. The phrase non-parametric does not hint at these approaches have any parameters or variables at all. Instead it implies that no predefined parameters are used. The variables are flexible and are decided at run time. The structure or behavior of the model as well as model variables are extracted and defined from the data. Hence the size of input data needed is much larger than the other two methods. The reason these methods stand out from others is that processes found in traffic that are difficult, non-linear as well as dynamic can also be modeled. No knowledge on the underlying processes is required. However, as this model is constructed from data, concealed events such as accidents and other

- Locally Weighted Regression

As the name suggests, this model uses the LR models. The prediction residual of each data point is then weighted proportionally to its proximity to the current measurement. Very good results are reported, in prediction accuracy as well as computation time.

TABLE I. LITURATURE SURVEY OVERVIEW

Reference	Work Done	Advantages	Disadvantages	Key Metrics/ Algorithms
[1]	Traffic congestion in urban areas minimized by eliminating external events.	Provides satisfactory accuracy and computational time.	Does not take into consideration the internal events (density, flow, etc.)	Radial Basis Function
[2]	Analyzed the major drawbacks in India affecting traffic congestion.	A practical solution that will really provide major benefits.	Very difficult to implement.	--
[3]	Impetus given to the weather factors affecting traffic management.	Accurate prediction of traffic parameters during weather changes.	Only weather data taken into consideration, huge computational time.	Multiple Linear Regression
[4]	A coherent policy to tackle urban traffic congestion.	A very practical and realistic solution.	Relies heavily on real-time data, ignores historic data.	--
[5]	A complete analysis of the system proposed.	Detailed workings of the system explained in layman's terms.	Focuses only on one system. No solutions to any problems provided.	Velocity and occupancy
[6]	The fuzzy-AR technique has been closely scrutinized and complete taxonomy of the system is explained.	The performance of the technique is enhanced and practical solutions are provided.	This technique cannot be implemented as a stand - alone model.	Fuzzy Auto-Regressive
[7]	Real time traffic prediction techniques have been processed using Apache Spark framework.	The performance has been considerably enhanced using the Apache and Hadoop framework.	Not a prediction model in true sense. Detects and informs users only in real-time.	Apache Spark, Spark SQL
[8]	Existing technologies have been modified to work under heterogeneous conditions.	Has shown promising results for traffic environments such as in India.	Very complicated and a lot of computational time needed.	Area Occupancy
[9]	A model has been developed to predict traffic under heterogeneous traffic conditions	Takes into consideration external events like pedestrian encroachment	Does not solve spatial dependency and produces over fitting	Fuzzy Logic
[10]	Scrutinizes the effectively of the exiting prediction systems under Indian conditions.	Provides a better insight into the drawbacks of the existing systems.	Does not suggest any solution to the existing shortcomings.	Artificial Neural Networks, Kalman filtering.

- Fuzzy Logic

With fuzzy logic a rule base (a set of IF-THEN rules) is created, manually or automatically. A current situation corresponds to one or more rules. Based on the then and sometimes the degree of correspondence, a prediction is made.

- Bayesian networks

In Bayesian networks, also known as causal models, the data from adjacent links are considered informative to the current link under investigation. A Bayesian network is simply a directed graphical model for representing conditional independencies between a set of random variables. Comparisons with other methods are not made. However, this method is applied in real life by Inrix, a Microsoft spin-off company.

- Neural Networks

Neural networks are one of the most extensively applied models because of their robust self-learning and self-adaptive potentials. We can also map events that are non-linear and dynamic using NNs. NNs are also characterized by the abilities of fault-tolerance and parallel processing. In order to enhance the prediction accuracy and to minimize the processing time, many extensions have been applied to the basic NN model. These extensions can be subdivided by the type of variation: (1) a different training procedure, (2) different internal structures or mathematics, (3) pre-processing input data and (4) include spatial and/or temporal patterns explicitly into the models.

IV. SYSTEM ARCHITECTURE

A. Existing System

The existing system architecture (Ref. Fig. 2) does not adhere to one particular prediction method or class. It is more of a general algorithmic framework which can be executed using different machine learning models. The disadvantage of the existing system is that it exhibits good performance only for those conditions which are very close to those supplied in the training phase. Also the prediction is done just minutes in advance to those vehicles that are within the range of the node deployed at the junction.

- Drawbacks of Existing System
  - Does not adhere to one particular prediction method or class.
  - It is a general algorithmic framework.
  - Produces Over-fitting.
  - Limited to a node or junction.

B. Proposed System

The proposed system (Ref. Fig. 3) uses averaging method to perform the prediction. The vehicles are also classified as heavy vehicles and small vehicles depending upon their distance from the ultrasonic sensor. This gives a better control mechanism. Also, the data is saved every time and

this data is always referred to train the system during each prediction. Due to this, the prediction error decreases with each prediction made. Also, spatial and temporal dependencies shall be solved using the averaging method. Novel Idea

The system is targeted towards making the experience of road travel more comfortable by creating a system that will predict the traffic conditions before-hand thus making navigation a bit easier, an idea which has not yet been effectively implemented in India. The success of this system will not only make it a landmark in Indian Road Transport but will also make planning of trips more controlled and easier.

- Difference between proposed system and Google Maps

Google maps is a real time traffic application. It uses the user's GPS information to detect traffic. It requires the users to be connected to the GPS and also needs permissions from the users. The speed of the vehicles is used as an indicator for calculating the traffic congestion. It provides only real time traffic information, no prediction is done.

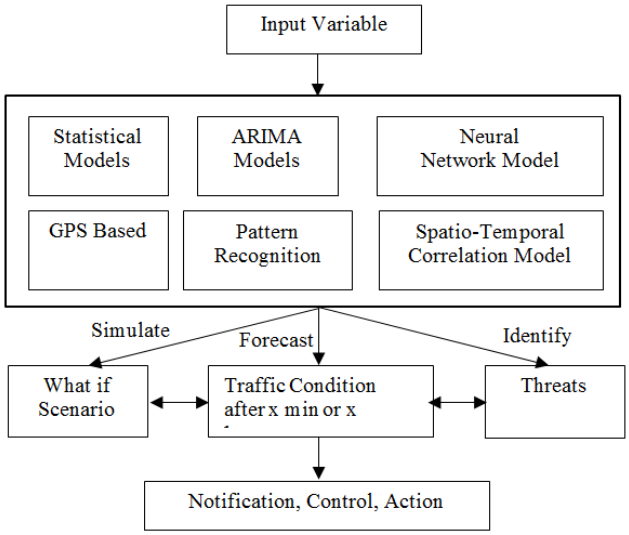

Fig. 2. Existing System Architecture Diagram

C. System Modules

- Data capture module : It consists of
  - Sensor data is collected using the PIC f184520 µC connected with:-
  - Ultrasonic sensors - Ultrasonic sensors are used to calculate the number and speed of vehicles.
  - 16*2 LCD
  - ESP8266 board
- Data processing module: comprises of
  - Transformation of sensor data
  - For processing the collected data, the sensors are connected using the PIC f184520µC to the system.
  - After collecting the data, we need to visualize those data taken from the sensors.
  - This can be done by converting the data into a CSV file.
- Prediction module
  - Averaging Technique used for prediction

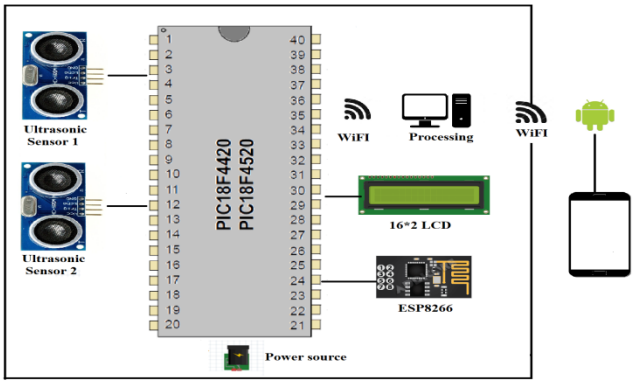

Fig. 3. Proposed System Architecture Diagram

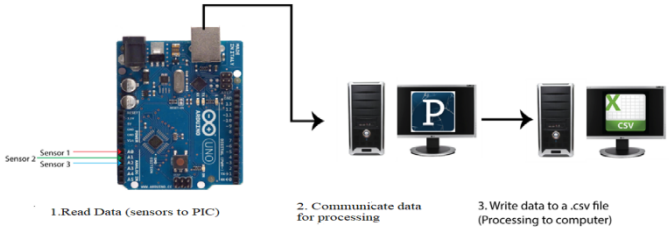

Fig. 4. System Modules

D. Flow and Methodology

Data is collected using two ultrasonic HC-SR04 sensors, each dedicated to one lane. A combination of real-time data and historic data is used in order to make the prediction. Averaging model is used as the predictive model. The working of the system is shown in Fig. 5.

- Server side

The server side is used to collect, process and store the traffic data collected through the sensors (Ref. Fig. 6). The sensor data is stored as per the day and time of the week. As each day's data is added, averaging is carried out on all the data again in order to predict traffic congestion. The accuracy increases with increase in the data. The data is taken lane wise as well as size of the vehicle is taken into consideration for the purpose of prediction.

- Client Side (mobile app)

The user can access the traffic congestion data using the android app. In order to solve the temporal dependency, the prediction is carried out within a specific time period and this is done independently for each day of the week. The user needs to send a request containing the day and the time on which the prediction has to be done (Ref. Fig. 7).

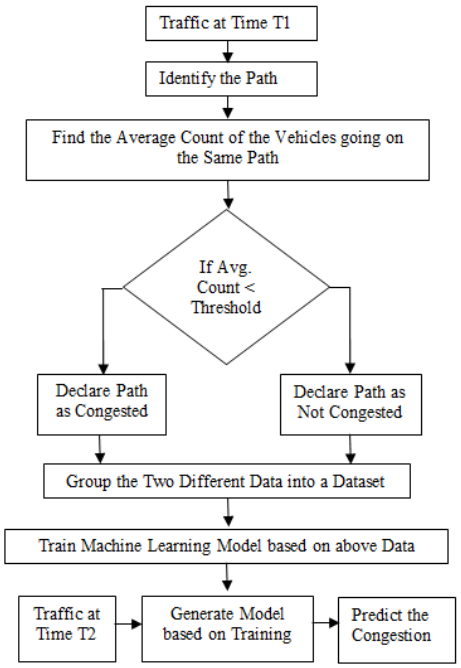

Fig. 5. Working of the System

E. Screenshots of working of System

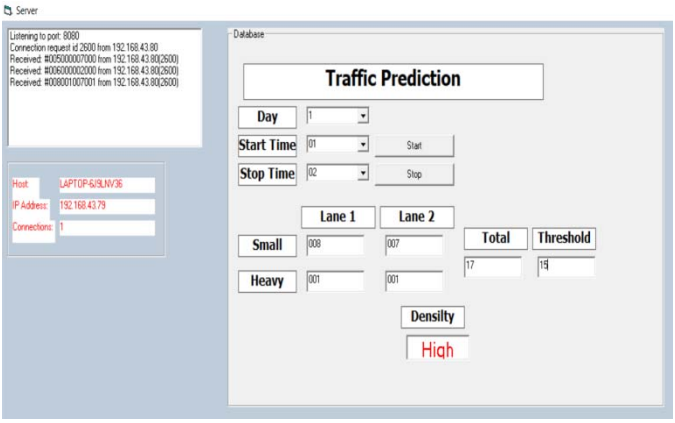

Fig. 6. Server Side Screen Shots

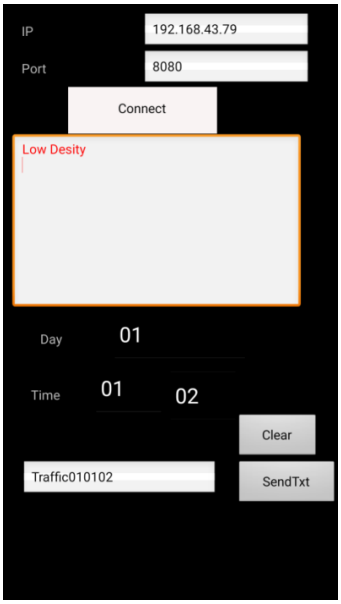

Fig. 7. Client Side Screen Shots

V. CONCLUSION

The proposed system predicts the traffic congestion for both urban and expressway conditions with reasonable accuracy depending on the historic as well as real time data. The proposed system combines the machine learning technique with IoT to reduce the prediction error and it exhibits high performance in order to make the travel plan comfortable. This is a closer step towards the concept of smart cities.

VI. FUTURE WORK

At present, for traffic-flow extrapolation a negligible number of the practices are employed worldwide in both urban and expressway situations. In this paper, the authors have instigated traffic-flow by means of averaging technique built on Naïve predictive model. In future, an innovative hybrid traffic-flow predictive model built on Neuro-fuzzy technique will be instigated for precise prediction. The performance of Neuro-fuzzy technique will be contrasted with averaging technique. Tangible world circumstances such as managing lively traffic-flow, selection-, scheduling- of route for transporters will be taken into account for prediction.

VII. REFERENCES

[1] D. Pescaru, "Urban Traffic congestion prediction based on routes information", IEEE Applied Computer Intelligence and Informatics, the 8th International Symposium On Applied Computer Intelligence and Informatics, pp. 121 – 126, 2013.

[2] "Urban Transport in India: Challenges and Recommendations", Indian Institute for Human Settlements online.

[3] J. Lee, B. Hong, K. Lee and Y. Jang, "A prediction model of traffic congestion based on weather data", IEEE Data Science and Data Intensive Systems, IEEE International Conference on Data Science and Data Intensive Systems, pp 81-88, 2015.

[4] S. Padam and S. K. Singh, "Urbanization and Urban Traffic in India: A Sketch for a policy", Harvard online.

[5] M. Ben-Akiva, G. Cantarella, E. Cascetta, J. Ruiter, J. Whittaker, and E. Kroes, "Real time prediction of traffic congestion," IEEE Vehicle Navigation & Information Systems, the 3rd International Conference On Vehicle Navigation 8 Information Systems, pp. 557 – 562, 1992.

[6] B. Chen, S. Peng and K. Wang, "Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks: A Fuzzy AR Approach," IEEE Transactions on Fuzzy Systems, Vol. 8, No. 5, pp. 491 – 508, 2000.

[7] M. Prathilothamai, A. M. Sree Lakshmi and D. Viswanthan, "Cost Effective Road Traffic Prediction Model Using Apache Spark", Indian Journal of Science and Technology, Vol. 9, 2016.

[8] V. Thamizh Arasan, and G. Dhivya, "Measuring Heterogeneous Traffic Density", World Academy of Science, 2008.

[9] A. K. Das, Krishna Saw and B. K Katti, "Traffic Congestion Modelling Under Mixed Traffic Conditions Through Fuzzy Logic Approach: An Indian Case Study Of Arterial Road", Recent Advances in Civil Engineering for Global Sustainability, Global Research and Development Journal for Engineering, 2016.

[10] A. Padiath, L. Vanajakshi, H. C. Subramanian and H Manda,"Prediction of traffic density for congestion analysis under Indian traffic conditions", IEEE Intelligent Transportation Systems, the 12th International Conference on Intelligent Transportation Systems, pp. 1-6, 2009.