International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Traffic Congestion Prediction System using K-Nearest Neighbour Algorithm

Rishab Menon R[1], Shreyas M S[2], Rajashree P[3], Rohith Thammaiah[4]

[1,2,3 &4] *Department of Computer Science and Engineering, MVJ College of Engineering, Bangalore, Karnataka, India*

Abstract - *Due to urbanization, there is a rapid growth in the number of vehicles on the road resulting in traffic congestion. Traffic congestion is a condition of a segment in the road network where the traffic demand is greater than the available road capacity. With increasing traffic, people in cities face many hurdles that affect their day to day activities. Thus, there is a necessity for a system that can accurately predict the traffic congestion rate at any road, at a future day and time. In this paper, we have explained a traffic congestion prediction system using data mining by implementing the K-Nearest-Neighbour (KNN) algorithm, so that based on the historic traffic data, the model predicts the traffic congestion values for the required locations on a particular day, at a specified time interval. Here, we make an attempt to model the traffic congestions on a particular road based on its spatial and temporal data. In order to make this system scalable to handle big traffic datasets, we have also implemented using the Hadoop framework. This system will be useful for the traffic department officials to plan the control over traffic density on roads.*

Key Words: **Traffic Congestion, Spatial, Temporal, KNN, Data Mining, Big Data, Hadoop, Urban Computing, Planning**

1. INTRODUCTION

There has been a steady increase in both rural and urban traffic in recent years resulting in congestion, accidents and pollution. In fact, traffic congestion is widely regarded as one of the greatest problems faced by the world today. Traffic congestions usually happen during peak hours or periodic events, which includes public celebrations, mid and large-scale business promotions, protests, parades, and other traffic incidents due to accidents etc. When a congestion occurs in one part of the traffic network, they are definitely going to affect the traffic flows of the surrounding roads. This includes all the traffic leading to the already congested roads. To resolve traffic related problems scientifically and reasonably has become a society wide consensus. Building transportation infrastructures can relieve the traffic pressure up to a certain level and for a limited period of time only. One of the important way to increase transport efficiency, reduce traffic congestion and improve traffic safety situation, is to implement traffic guidance and control, effectively use the road resource and give full play to vehicle function. In developing countries, where resources are limited, and due to less attention paid to transportation sector; traffic congestion problem is becoming a major challenge for administrators and planners. As far as Indian condition are concerned, Indian cities are facing traffic problems characterized by mixed traffic flow conditions, levels of congestion, noise and air pollution, traffic fatalities and injuries. Hence, it is essential to have efficient methods to discover the frequent patterns of congestions, and their propagations in the traffic networks.

Traffic congestion prediction and management of traffic based on that prediction is proposed as a solution to manage the fast-growing city traffic. Prediction of traffic is challenging because of non-linearity and the larger variance and quick transition between free flow, breakdown, recovery and congestion. Many works have been attempted in literature to effectively model the traffic and construct prediction models.

In this paper, we model the traffic based on its spatial neighbour data and the temporal data at that road segment. Historic data of volume of traffic at road segment is used to train the model and once the model is trained, it is used to predict the traffic at later point of time. We implement and compare our system against temporal model systems like ARIMA and prove that our prediction method has less error compared to ARIMA prediction model. Since the traffic data is ever growing, it is necessary to make the system scalable i.e. the system should be able to handle big data efficiently. Thus, have also designed the system using Apache Hadoop framework, to make sure the system is scalable to handle big traffic datasets. Furthermore, the presented paper can contribute to the transportation research in the community of urban computing.

2. PROPOSED SYSTEM

In this paper, a prediction model is proposed that implements the K-Nearest Neighbour(KNN) algorithm, which makes use of the spatial, as well as the temporal correlations found in the traffic dataset, in order to predict the traffic congestions at a required location, on a particular day and time-interval. In this mechanism, the traffic at a road segment is dependent on the traffic at its neighbouring roads from where the traffic can flow into the road under consideration. So, the neighbour's traffic data must also be considered to model the traffic at this road, in addition to its temporal data. In other words, the spatial correlations are considered along with the temporal correlations with the help of KNN algorithm to predict the traffic congestions. The proposed solution is about two times more accurate than the existing solutions.

2.1 System Architecture

The system architecture of any system describes about the various components used in the system. In this paper, we have described the system architecture from a GUI point of view, as well as from a functional point of view. This system can be used by users as well as the administrators responsible for maintaining traffic. This system will also have a database on its own. Also, it will use the Google Maps API to mark the congested roads for a particular location. The various components of the system from a GUI point of view is shown below:

Fig - 1: System Architecture (GUI Level)

The various components of the system from a functional point of view is shown below:

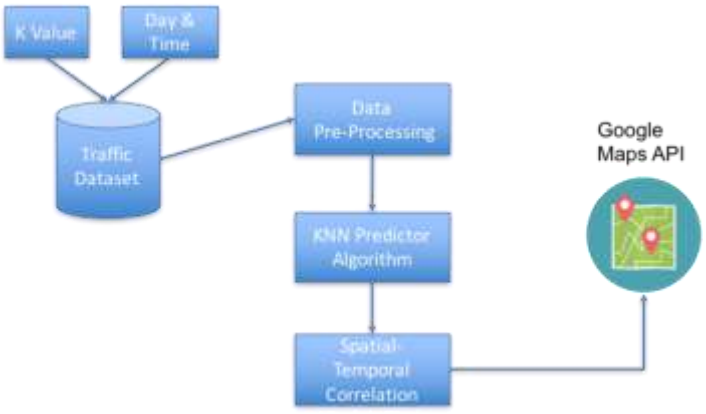

Fig - 2: System Architecture (Functional Level)

2.2 System Design

The datasets are given as input for training the model. The required fields are extracted from the datasets. Temporal modelling constructs the temporal model using moving average model to forecast the traffic based on temporal characteristics. Spatial K-Neighbour finds different locations in the dataset, and for each of these locations, the K-Neighbours are found. K value is configurable within the system. The spatial and temporal correlation module builds a model based on the spatial and temporal properties. Equal weightage is given to temporal, spatial and weighted average model, which is used for correlation. The forecasting module stores the temporal and spatial correlation equations. This answers any traffic congestion queries for any location, at any point of time, in future.

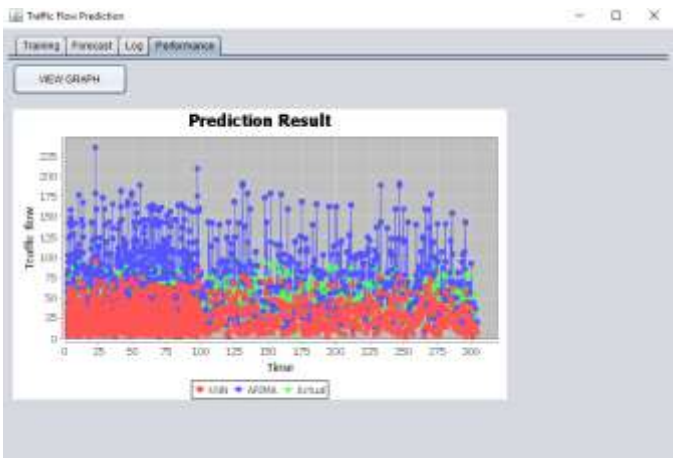

Chart-1 Performance Graph of KNN Model

The above performance graph shows the increase in accuracy when the K-Nearest Neighbour algorithm is used. The forecasted values using both K-Nearest Neighbour model and ARIMA model against the actual real values are compared in the graph.

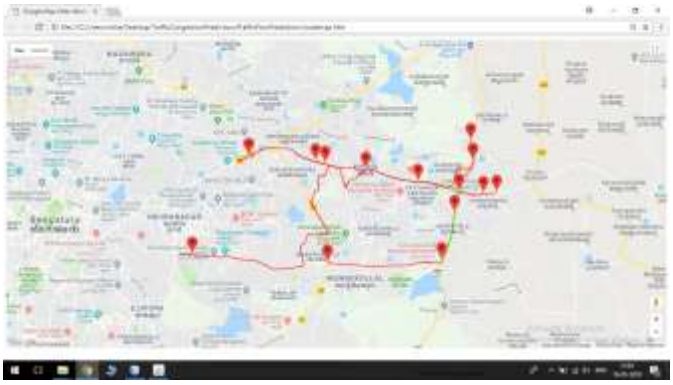

Fig - 3: Predicted Traffic Congestion of Roads

The predicted traffic congestion values are displayed using the Google Maps API. The roads marked in red indicate predicted busy or high traffic congestion. The roads marked in green indicate predicted free or less traffic congestion.

3. CONCLUSION

A scalable, more accurate and easy to use design and implementation of Traffic Congestion Prediction System using Data Mining is proposed. In order to predict the traffic congestion values accurately, we need to consider both temporal and spatial correlations. In the existing systems, only temporal correlations were considered. The proposed prediction model used in this project takes into consideration both the spatial and temporal correlations.
By implementing K Nearest Neighbour (KNN) Algorithm, the project accurately predicts the traffic congestion values for the required day and time interval. The algorithm considers the neighbour nodes based on distance and time alone. The predicted values are projected on to the Google Maps API. The road segments marked in red indicate high traffic congestion, and the road segments marked in green indicate low traffic congestion. A performance graph is charted to show the accuracy of the predicted values to the actual values.

REFERENCES

[1] Huiping Cao, Nikos Mamoulis, and David W Cheung. Mining frequent spatio-temporal sequential patterns. In Fifth IEEE International Conference on Data Mining (ICDM), pages 8–pp. IEEE, 2005.

[2] Enrique Castillo, Jos´e Mar´ıa Men´endez, and Santos S´anchez- Cambronero. Traffic estimation and optimal counting location without path enumeration using bayesian networks. Computer- Aided Civil and Infrastructure Engineering, 23(3):189–207, 2008.

[3] Eric Hsueh-Chan Lu, Wang-Cheng Lee, and Vincent S Tseng. Mining fastest path from trajectories with multiple destinations in road networks. Knowledge and information systems, 29(1):25–53, 2011.

[4] Xiaofeng Wang, Gang Li, Guang Jiang, and Zhongzhi Shi. Semantic trajectory-based event detection and event pattern mining. Knowledge and information systems, pages 1–25, 2011.

[5] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In 20th International Conference on Very Large Data Bases, pages 478–499. Morgan Kaufmann, Los Altos, CA, 1994.

[6] Jiancheng Long, Ziyou Gao, Xiaomei Zhao, Aiping Lian, and Penina Orenstein. Urban traffic jam simulation based on the cell transmission model. Networks and Spatial Economics, 11(1):43–64, 2011

BIOGRAPHIES

Rishab Menon R received the B.E degree in Computer Science & Engineering from MVJ College of Engineering, Bangalore, India in 2018. His primary research interests lie in data science, analytics & machine learning.

Shreyas M S received the B.E degree in Computer Science & Engineering from MVJ College of Engineering, Bangalore, India in 2018. His primary research interests lie in machine learning & artificial intelligence.

Rajashree P received the B.E degree in Computer Science & Engineering from MVJ College of Engineering, Bangalore, India in 2018. Her primary research interests lie in data science, analytics, RPA and automation.

Rohith Thammaiah received the B.E degree in Computer Science & Engineering from MVJ College of Engineering, Bangalore, India in 2018. His primary research interests lie in application development & UX architecture.