Vehicle Traffic Data Management System

Rakshith Gowda VR
CSE
1PI14CS079
PESIT
Bangalore,India
rrakshi10@gmail.com

Tanush Vinay
CSE
1PI14CS119
PESIT
Bangalore,India
tanush.vinay@gmail.com

***Abstract*— In this project we have tried to analyze traffic between two points in a city. We have tried to infer from basic understanding between attribute dependencies and in our data, visualized, tried different methods to understand our data and seek a trend in traffic. Our analysis would be useful in organizations which need this analysis for traffic management etc. We were able to visualize trends in traffic data and its flow in this project.**

I. Introduction

In this project we make use of a city pulse traffic data set. The data set consists of real time prerecorded sensitive traffic data at different points in a city. Using this data we would like to predict if that point has a traffic jam and also if a point has probability if having a traffic jam at a particular point of time. By correlating major attributes and setting up probability factors and analytics. Based on the traffic flows at a signalized intersection, the timing of the signal can be optimized. Variable message signs at bus stops or along the highway can be used to display valuable information to travelers regarding travel times. As for improving the safety of transportation system, data related to the location and timing of occurrence of crashes in the system can be used to develop important plans to improve the transportation infrastructure. In case of emergencies alternate routes for emergency response vehicles and traffic flow can be identified. By doing this project we could see major points to be traffic managed in a city.

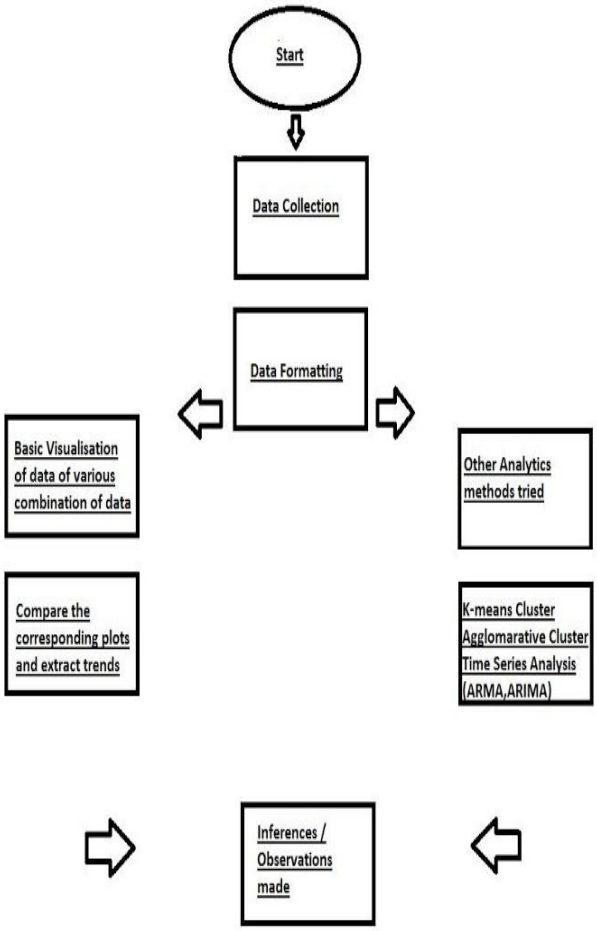
II. Summary Of literature Survey

In our literature survey we saw various ways in which people viewed data and comprehend it as traffic. We saw how people were able to see the attributes in their data and were able to correlate them and finally come to a conclusion based on maximizing and minimizing their attributes. In some cases we saw people seeing the vehicle count, the time it was measured for, the speed maybe, the time of the day the measurement was done etc. we also saw the various mathematical approaches taken to visualize and draw conclusions. We've learnt from our literature survey and tried to implement it in our project effectively with more of our own understanding.

III. Problem Statement

The problem statement of this project would be that we take a dataset of traffic data for a particular city and try to say if that's a high traffic point or not at a particular time. The data set is a traffic data set for a city in Denmark. It contains multiple files, each file for a particular pair of points in a city between which traffic data was measured. It's a very well structured data for three months and with a data measurement interval of five minutes, so the data was huge. The constraints include that we have only about four relevant attributes to work with, the volume of the data and we didn't have too much knowledge about the place and weather so we used basic info we found on the net because it was needed.

IV. Proposed System-Block Diagram



V. Detailed Explanation Of Block Diagram

We initially started with searching for the data online and we found a website which had data about the traffic of a city collected for various location for an interval of 5 months.

The consisted of various attributes such as id of the location, number of vehicles counted, the time slot (every 5 mins), date, averaged measured time, average speed and various other attributes.

After collection of data we had to simplify one of the attribute and split it into two attributes. A python codes separates all the data of a particular location and formats the data as needed and a csv files is created. From this csv file the data of a particular location and date is extracted and the granularity of the time is increased to 10 minute.

This data is then loaded into R and various plots is draw and the plots are compared to make inferences out of visualization is made and concluded.

Under other methods, We tried to first find correlations among various attributes which did not yield fruitful results. So, there was no linear relation at all so linear methods like linear regression failed, even the multivalued regression didn't make sense and logical regression didn't apply for our data.

We then tried clustering the data based on vehicle count using kmeans clustering where min, max and mean of vehicle count as centroids and plotting graphs for these clustered data, we still failed to get a trend, the same case was with dendogram.

We finally tried time series but even after splitting the time series, we couldn't see any demarcation of trend. Hence we resolved to plot data assuming that traffic directly proportional to vehicle count and inversely related to average measured time and average speed and drew conclusions based on trends against daily weekly and monthly time.

VI. Experiments And Results

We were convinced that our attributes had no correlation by the experiments and hence our result was sure that we could not linear models and the plots and code outputs attached say so. We also attached plots of k-means and dendogram which we used to visualize data but they didn't make any sense when it came to inferring about data. Since our data was time sensitive we plotted a time series graph for vehicle count vs time and we were not able to get a good split to actually infer anything. We finally chose to do analysis via visualization for our data.

We first plotted a graph for vehicle count over the three months. We got inference about how the data behaved, we could see that vehicle count during weekdays and weekends were almost the same we started seeing a trend. We Then plotted weekly data for a particular pair of points and then indeed confirmed that our above understanding make sense. We then tried to plot vehicle count in

combination with average measured time and average speed and it showed the same trend so we just stuck to using vehicle count. We then went ahead to plot traffic on a 24 hour basis for each day, weekdays and weekends. We inferred for on particular file that during weekends we saw that there was more traffic during the afternoons and minimal during the morning and evening. During the weekends we saw heavy traffic during the mornings and gradual drop in traffic as the day progresses. We then referred our metadata file and then expected the result to be this way because we assumed that people on weekdays went to school and work in the morning and in the evenings people returned home at different times on this traffic point. On weekends people woke up late and didn't return home early evening so traffic was mostly during the afternoons. We saw this model fitting and tried to see similar trends across other four files. Indeed the other four showed this similar behavior over weekends and weekdays and some with some strange values. One such file had zero traffic everyday for a particular time period, probably because the road wasn't open in those hours of the day. We felt that we could have done more if we had the information about the place and probably the weather pattern as well.

VII. Conclusion

In conclusion I would like to say that we had multiple traffic point and they had their special trends across weekends and weekdays. We also saw abnormal values for Particular time intervals like sudden heavy traffic or suddenly no traffic for long periods of time . We needed more information like weather, events in the city,the road conditions etc to predict the traffic points more comprehensively. But the one valuable conclusion we drew is traffic behavior wasn't constantly the same but showed a slight trend, so traffic jams is more of a real time issue which needs to solved when it happens because it has so many factors apart from past recorded traffic data.

VIII. References

1. Road congestion and incident duration

Martin W. Adler, Jos van Ommeren, Piet Rietveld VU University Amsterdam

De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands Tinbergen Institute, Amsterdam Gustav

Mahlerplein 117, 1082 MS Amsterdam, The Netherlands 15 July 2013

2. Road Traffic Congestion in the Developing World

Vipin Jain Polytechnic Institute of NYU vjain02@students.poly.edu

Ashlesh Sharma New York University ashlesh@cs.nyu.edu

Lakshminarayanan Subramanian New York University lakshmi@cs.nyu.edu

3. ROAD TRAFFIC PREDICTION USING BAYESIAN NETWORKS

Poo Kuan Hoong, Ian K. T. Tan, Ong Kok Chien,Choo-Yee Ting,

Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. {khpoo, ian, ong.kok.chien08, cyting}@mmu.edu.my

4. Prediction of traffic conditions in the urban area of Thessaloniki

K. Lakakis, K. Kyriakou, P. Savvaidis
Department of Civil Engineering, Aristotle University of Thessaloniki, 54006, Thessaloniki, Greece

5. Traffic Data and Analysis Manual

Texas Department of transportation, September 2001.

6. Traffic Archive Data Management

D Sharan and Gitakrishan Ramadurai

Transportation Engineering Division.
Dept of civil engineering,
IIT Madras, October 2011.

IX. Contribution Of Each Member

Since we were a two member team, there was no such individual contribution but was done together. But to show some demarcation, Rakshith was responsible for cleaning the data and basic visualization. Tanush was responsible for trying out major data classification and techniques and drawing conclusions if it was the right way to visualize the data.