



Traffic Flow Calculation using Big Data

S.Kalaiarasi , G. Leela , K. Nikesk , Ch. Prasad

Abstract: Traffic is one of the primary issues in world. It makes numerous medical issues people on foot and bikers. It is additionally one of the practical setting of a nation. U.S.A. alone squandered almost \$160 billion of fuel in year 2014 alone. Mumbai remains at no.1 position in the rundown of most exceedingly awful traffic stream while Delhi taking no.4 position. In this task we use BIGDATA for guaranteeing that the explorers doesn't get struck in the rush hour gridlock. BIGDATA can enable clients to settle on better travel choices, lighten traffic blockage, diminish carbon outflows, and improve traffic activity proficiency. Our goal of traffic stream forecast is to give a superior traffic stream data. Traffic stream forecast has picked up its significance because of fast development in urban areas and increment in rush hour gridlock blockage.

Traffic stream forecast intensely relies upon authentic and ongoing traffic information gathered from different sensor sources, including inductive circles, radars, cameras, portable Global Positioning System, publicly supporting, internet based life, and so on.

In this paper, we propose a profound learning-based traffic stream forecast technique.

Keywords: Deep learning, real time information, traffic stream prediction.

I. INTRODUCTION

The precise expectation of the traffic is one of the hardest undertaking in any Urban Neighborhood. It is given most extreme significance since it has been a serious issue in any city everywhere throughout the world[1][2]. It was causing a great deal of air and clamor contamination[3], irritating the financial condition of a country by causing a cost in millions. In this way, presently it is the ideal opportunity for well and bettersupport of it. Traffic Flow forecast must be satisfactory and precise for clients and voyagers take better educated choices[4], mitigate traffic blockage, diminish carbon discharges and chairmen for better arranging of streets and preventive measures. The goal of traffic stream expectation is to give such traffic stream data. It has acquire consideration and significance because of quick development in urban populace.

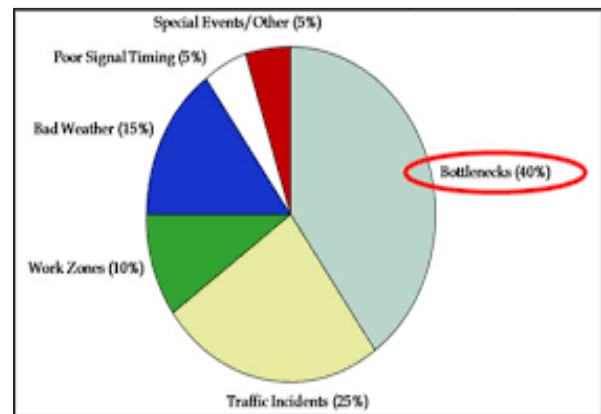


Fig. 1 Pie graph

Traffic stream forecast depends on ongoing and noteworthy information gathered from the voyagers, traffic cameras, GPS and so on. With this colossal pool of information we need to foresee the traffic utilizing Big Data. In spite of the fact that there are numerous frameworks and models that are as of now been utilized in rush hour gridlock expectation none of them are productive and best. The vast majority of them utilizing shallow traffic models which are insufficient.

II. LITERATURE REVIEW

Traffic has always been a bigger problem in almost every part of the earth. Traffic flow prediction is considered a important key concept and has been taken care of. But with rapid increase in urban culture and rise in number of vehicles on road it wasn't sufficient. So we came up with an idea of Intelligent Transportation System[5]. Over the decades many models of traffic flow predictions are created and some of them are very good but not completely efficient.

We have come up with a model of traffic flow prediction using knn algorithm and Feedforward Neural network[6] of multilayer perceptron. In it the data runs in a single direction but not in a loop. Information is passed on from various reliable sources like GPS trackers , toll gates , traffic cameras , police department etc. All this information is then processed in a hidden layer and the output is created as the time required to travel a distance and best route possible with least traffic.

III. METHODOLOGY

In this model we use k-nearest traveller (KNN) estimation nearby Feed Forward Neural Network of Multi-Layer Perceptron. This class of frameworks contains various layers of computational units, regularly interconnected in a feed-forward way.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

G. Leela Krishna Reddy, Pursuing B.Tech, Computer Science Department, SRM Institute of Science and Technology Tamil Nadu, India

K.Nikesk Sai Veda, Pursuing B.Tech, Computer Science Department, SRM Institute of Science and Technology Tamil Nadu, India

Ch. Vara Prasad, Pursuing B.Tech, Computer Science Department, SRM Institute of Science and Technology Tamil Nadu, India

Mrs. S. Kalaiarasi, Assistant Professor, Department of Computer Science SRMIST, Ramapuram.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Each neuron in one layer has guided relationship with the neurons of the following layer.

In various applications the units of these frameworks apply a sigmoid limit as an activation work..

The general estimation speculation for neural frameworks communicates that every relentless limit that maps between times of certifiable numbers to some yield break of veritable numbers can be approximated emotionally eagerly by a multi-layer perceptron with just one covered layer. This result holds for a wide extent of activation limits, for instance for the sigmoidal limits..

Multi-layer frameworks use a collection of learning systems, the most standard being back-inciting. Here, the yield regards are appeared differently in relation to the correct answer with figure the estimation of some predefined goof work. By various methods, the mix-up is then supported back through the framework. Using this information, the figuring changes the heaps of each affiliation to diminish the estimation of the bungle work marginally. Resulting to reiterating this technique for an enough tremendous number of getting ready cycles, the framework will commonly meet to some state where the slip-up of the checks is close to nothing. For this circumstance, one would express that the framework has taken in a particular target limit. To change stacks properly, one applies a general method for non-straight streamlining that is called point dive. For this, the framework calculates the subordinate of the error work in regards to the framework loads, and changes the heaps with the ultimate objective that the bumble reduces (along these lines going downhill outwardly of the goof work). Thus, back-expansion must be applied on frameworks with differentiable activation limits.

All things considered, the issue of demonstrating a framework to perform well, even on tests that were not used as getting ready tests, is an extremely unpretentious issue that requires additional methodologies. This is especially huge for circumstances where in actuality, compelled amounts of planning tests ar-e available. The risk is that the framework overfits the arrangement data and fails to get the certifiable quantifiable technique delivering the data. Computational learning theory is stressed over getting ready classifiers on a compelled proportion of data. Concerning neural frameworks an essential heuristic, called early ending, much of the time ensures that the framework will summarize well to models not in the readiness set.

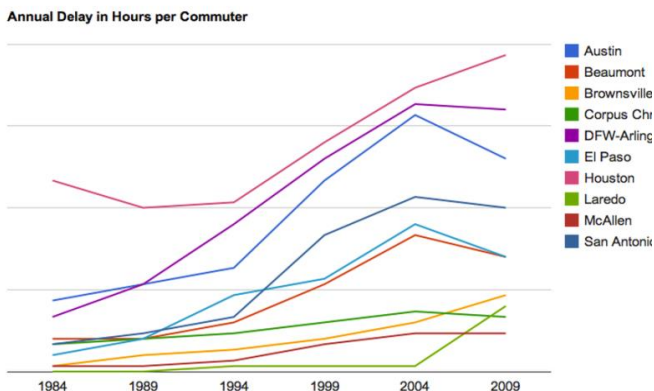


Fig . 2 Graph depicting delay in different cities

By and large, the issue of showing a system to perform well, even on tests that were not utilized as preparing tests, is a very unobtrusive issue that requires extra methods. This is particularly significant for situations where without a doubt, constrained quantities of preparing tests are accessible. The peril is that the system overfits the preparation information and neglects to catch the genuine measurable procedure creating the data. Computational learning hypothesis is worried about preparing classifiers on a constrained measure of information. With regards to neural systems a straightforward heuristic, called early halting, frequently guarantees that the system will sum up well to models not in the preparation set.

Comparative regular issues of back-proliferation calculation are speed of assembly and plausibility of winding up in a nearby least of mistake work. Today there are useful strategies that make back-proliferation in multi-layer perceptrons the instrument of decision for some, AI errands.

One likewise can utilize arrangement of free neural systems directed by couple of delegate, a comparable conduct that happens in cerebrum. These neurons can perform independently and handle huge assignments, and results can be consolidated at last.

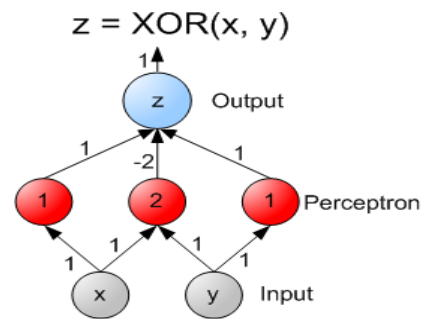


Fig.3.Multilayer Perceptron

A. *K Nearest Neighbour:*

KNN can be utilized for both grouping and relapse prescient issues. Be that as it may, it is all the more broadly

utilized in arrangement issues in the business.

Pseudo Code of KNN

We can actualize a KNN model by following the underneath

steps:

1. Load the information
2. Initialise the estimation of k
3. For getting the anticipated class, emphasize from 1 to add up to number of preparing information focuses

Calculate the separation between test information and each

line of preparing information. Here we will utilize Euclidean

separation as our separation metric since it's the most prevalent technique. Different measurements that can be utilized are Chebyshev,

cosine, and so forth.

- Sort the determined separations in climbing request in view of separation esteems .Get top k lines from the arranged cluster
- Get the most regular class of these columns .Return the anticipated class. KNN calculation is one of the easiest characterization calculation. Indeed, even with such

straightforwardness, it can give exceptionally aggressive outcomes.

KNN calculation can likewise be utilized for relapse issues.

Calculation

B. Algorithm

Distance functions

Euclidean $\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$

Manhattan $\sum_{i=1}^k |x_i - y_i|$

Minkowski $\left(\sum_{i=1}^k (|x_i - y_i|^q) \right)^{1/q}$

Proposed Method: Apart from other customary strategies we

utilized KNN calculation which chooses the most related neighboring stations with test station. The forecast information

from KNN calculation is nourished to Feed forward neural

system where it is weighed alongside information from different sources and the last expectations are expressed.

The flowchart for the proposed model is as per the following:

Step1.Calculate Euclidean separation between neighboring M-1 stations with test station

Step2.Select most related K stations with test station

Step3.Predict Traffic Flow utilizing FNN arrange

Step4. Backtrack the expectation to check for mistakes

Step5. Get the anticipated traffic stream

IV. EXPERIMENT

A. Dataset Description:

We conduct our experiments on a real highway traffic dataset. It contains nearly 3 months traffic flow information (from July. 19, 2016 to Oct. 17, 2016), and collects 1,527,136 road driving records with 77,012 vehicles. This dataset was formally released by Knowledge Discovery and Data Mining Tools Competition (KDD CUP 2017)[7], which can be downloaded from Tian Chi Platform

B. Index of Performance:

For evaluating the efficiency of proposed model, three performance indexes are required.

Mean absolute Error(MAE), Mean Relative Error(MRE), RMS error (RMSE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - \hat{f}_i|$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|f_i - \hat{f}_i|}{f_i}$$

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (|f_i - \hat{f}_i|)^2 \right]^{\frac{1}{2}}$$

Where f_i is observed traffic flow and \hat{f}_i is predicted traffic flow.

C. Efficiency of algorithm

We have tested our model with traffic flow of 15-min , 30- min, 60-min interval. All these information is provided through input layer. It is then processed in hidden layer and output is provided .

V. IMPLEMENTATION

A.Getting data:

Data is collected from GPS trackers in 1779 ravellers mobiles,cars,traffic signals and cameras.

B. Processing data:

The data thus acquired is then divided based on time. Each hour data is divided from one point to another point and every day data is stored accordingly. Now this data is fed to FNN network as one input,the other inputs being weather condition,condition of road etc. All these inputs are taken and are weighed in a hidden layer i.e. suppose if there is rain estimated on that particular day or there is construction being done on road the inputs are given importance or else if it a clear weather and a clean road the inputs need not to be considered. Hence by evaluating the inputs the output is generated.



Fig. 4.Google Maps

VI. RESULT

From the output we can tell that our proposed model is efficient for traffic flow prediction of typical roads with heavy, medium traffic roads. But it does not work efficiently in light traffic areas. Since the inputs from light traffic area are very small and even a small change can be a big error. Hence proposed method is effective for practice in high and medium traffic areas.

VII. CONCLUSION

In this paper we have discussed about Traffic flow prediction using Big Data in combination with knn algorithm and Feed forward neural network. KNN

is used to select mostly related neighboring stations and these input along with others is fed to the Feed forward neural network which processes the data and represents the output. In FNN we use backpropagation which compares the output and helps in reducing the errors. Thus this model is much more efficient than the previous models. FNN models use the weights and by adjusting them they reduce the error.

REFERENCES

1. J.Zhangetal.,“Datadrivenintelligenttransportationsystems:Asurvey,”IEEETrans.Intell.Transp.Syst.,vol.12,no.4,pp.1624–1639,Dec.2011.
2. C. L. Philip Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” Inf. Sci., vol. 275, pp. 314–347, Aug. 2014.
3. Claire Granier, Luis Kornbluh, Stacy Walters, Guy P. Brassuer, “Global impact of road traffic on atmospheric chemical composition & on ozone climate forcing”, Journal of Geophysical Research, vol. 111, May 2016.
4. N. Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, “DynaCAS: Computational experiments and decision support for ITS,” IEEE Intell. Syst., vol. 23, no. 6, pp. 19–23, Nov./Dec. 2008.
5. Kaan Ozbay, Pushkin Kachroo, “Incident Management in Intelligent Management Systems”, UNLV, 1999.
6. Michael May, Dirk Hecker, Christine Korner, Simon Scheider, Daniel Schulz, “A vector-geometry based Spatial Knn-algorithm for traffic frequency predictions”, IEEE International Conference on Data Mining Workshops, 2008.
7. Zibin Zheng, Yatao Yang, Jiahao Liu, Hong-Ning Dai, and Yan Zhang, “Deep and Embedded Learning Approach for Traffic Flow Prediction in Urban Informatics”, IEEE Trans. Intell. Transp. Syst., 2019-10-12.

AUTHORS PROFILE



G. Leela Krishna Reddy currently doing his IIIrd Year B.Tech from SRM Institute of Science and Technology from Computer Science Department.



K. Nikesh Sai Veda currently doing his IIIrd Year B.Tech from SRM Institute of Science and Technology from Computer Science Department



Ch. Vara Prasad currently doing his IIIrd Year B.Tech from SRM Institute of Science and Technology from Computer Science Department.



Mrs. S. Kalaiarasi is pursuing her Ph.D in CSE from Saveetha University and did M.E from Sathyabama University. She is Currently working as a Assistant Professor in Department of Computer Science at SRMIST, Ramapuram.