ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

Implementation of K-NN on Traffic Flow Prediction Using Hadoop

Vaibhav Gholap[1], Sujit Nale[2], Pradeep Patil[3], Renuka Patil [4]

BE. Students, Dept. of Computer Engineering, JSPM'S ICOER, Wagholi, Pune, Maharashtra, India[1,2,3,4]

ABSTRACT:Big data in the major technique to implement the data processing in handling large amount of data's. There are number of techniques that was updating in the big data in order to manage the data which was not handled by the normal human. Big data is used in many large scale industries which are containing millions ofrecord. This is because of the increase in the population and the digitalize community.In big-data-driven traffic flow prediction systems, the robustness of prediction performancedepends on accuracy and timeliness. This paper presents a new Map-Reduce-based nearest neighbour (NN)approach for traffic flow prediction using correlation analysis (TFPC) on a Hadoop platform. In particular,we develop a real-time prediction system including two key modules, i.e., offline distributed training(ODT) and online parallel prediction (OPP). Moreover, we build a parallel k-nearest neighbour optimization classifier, which incorporates correlation information among traffic flows into the classification process.Finally, we propose a novel prediction calculation method, combining the current data observed in OPPand the classification results obtained from large-scale historical data in ODT, to generate traffic flowprediction in real time. The empirical study on real-world traffic flow big data using the leave-oneoutcross validation method shows that TFPC significantly outperforms four state-of-the-art predictionapproaches, he Gaussian Mixture Model (GMM) is used to find out the patterns that how the vehicles are moving around the cities. And also to find out the time interval when the traffic will be high or low.

KEYWORDS: Big data analytics, traffic flow prediction, correlation analysis, parallel classifier, HadoopMap-Reduce.

I. INTRODUCTION

The classification of big data is becoming an essential task in a wide variety of fields such as biomedicine, social media, marketing, etc. The recent advances in data gathering in many of these fields have resulted in an inexorable increment of the data that we have to manage. The volume, diversity and complexity that bring big data may hinder the analysis and knowledge extraction processes. Under this scenario, standard data mining models need to be re-designed or adapted to deal with this data. The k-Nearest Neighbour algorithm (k-NN) is considered done of the ten most influential data mining algorithms. It belongs to the lazy learning family of methods that do not need of an explicit training phase. This method requires that all of the data instances are stored and unseen cases classified by finding the class labels of the k closest instances to them. To determine how close two instances are, several distances or similarity measures can be computed. This operation has to be performed for all the input examples against the whole training dataset. Thus, the response time may become compromised when applying it in the big data context. The traffic data in transportation have beenexploding rapidly with the characteristics of heterogeneity, autonomous sources, and complex and evolving associations (HACE). The big data generated by the IntelligentTransportation Systems (ITS) are worth further exploring tobring all their full potential for more proactive traffic management. The ability to accurately predict the evolutionof traffic in an online and real-time manner that plays acrucial role in traffic management and control applicationsis particularly important. Data-driven intelligent transportation has drawn significant attention in recent years: provided a special session on big data servicesand computational intelligence for industrial systems,including ITS applications. Furthermore, a special issuehighlighted the most recent research progress in big data The traffic is the main problem that occurs in the major cities. It will affect many of the Peoples day to day life and may cause major problem to heavy vehicles to reach the destination on time. Many Problems that will made the traffic flow difficult such as Weather, Road works, Accidents, vehicles that was repaired and make the traffic jam on

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

the road. So, the future the world is going to face traffic problem which is the major problem for the normal peoples. Many of the Projects are also available to control these kind of the traffic problems and still the research is going for these control traffic on Intelligent Traffic System (ITS). And our aim is to predict the traffic for frequent interval of time and also to give the suggestion to the people to reach the destination quickly. In other countries they have already some uses these type of prediction technique to help the peoples to reach the destination and also to reduce the traffic that occur in the cities. Some project the traffic signals are recorder with the cameras and sensors to know the Number of vehicles.

II. RELATED WORK

In the last several decades, considerable research studies were reported on the applications of different empirical and theoretical techniques to traffic flow prediction. These works can be roughly categorized as parametric methods, nonparametric methods, and hybrid integration methods [1]. Recently, there have been various traffic flow prediction systems, models and algorithm susing statistics-based approaches and computational. Intelligence(CI)-based approaches[3].

Lv et al. proposed a novel deep-learning-based traffic flow prediction method, using auto encoders as building blocks to represent traffic flow features for forecasting. Li et al. presented a method which picks the most relevant data from the ``Big Data'' to build concise yet accurate traffic flow prediction model [4]. Tchrakian Lv et al.developed an algorithm or the implementation of short-term prediction of traffic with real-time updating based on spectral analysis. Min and Wynter put forward an approach to providing predictions of speed and volume over 5-min interval for upto 1 h in advance. However, most of them employed the stand-alone learning models with the sequential algorithm son a single machine and were still somewhat unsatisfactory in processing the increasingly explosive traffic big data in realtime, such as massive taxi trajectory data used in this work [6][7].

A. Problem statement:

The traffic data in transportation have been exploding rapidly with the characteristicsof heterogeneity, autonomous sources, and complex and evolving associations (HACE). The big data generated by the Intelligent Transportation Systems(ITS) are worth further exploring to bring all their full potential for more proactivetraffic management. The ability to accurately predict the evolution of traffic in anonline and real-time manner that plays a crucial role in traffic management andcontrol applications is particularly important.

B. Goals and objectives:

To predict real time traffic flow prediction using current big traffic data.
To save memory consumption.
To reduce the computational costs of big calculations.
To display analysis of traffic flow data in form of speedup, scale-up andsize-up.

III. PROPOSED SYSTEM

This paper aims to develop a robust approach for the accurate and real-time Traffic Flow Prediction using Correlation Analysis (TFPC). The major contributions of our work are summarized as follows:

A new prediction system (RPS) on a Hadoop platform is built to improve the capacity of big traffic data processing for forecasting traffic flow in real time, which h contains two key modules of offline distributed training (ODT) and online parallel prediction (OPP).

A robust nearest neighbour classifier (ParKNNO) ona Map-Reduce framework is presented to enhance the accuracy, efficiency and scalability of traffic flow prediction,by discovering correlation information amongtraffic flows and incorporating it into the classification process.

A novel prediction calculation method is put forward to generate the real-time traffic flow prediction, withthe current data observed in OPP and the classificationresults obtained from large-scale historical data in ODT.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

The prediction performance of TFPC is investigated with real-world traffic flow data collected from 12,000 taxis of Beijing during 15 days. The empirical study indicates that the proposed approach is superior toother comparable prediction methods in terms of accuracy, speedup, scaleup, and sizeup.

We aim to develop a new framework (RPS) to handlereal-time prediction applications. Also, a robust nonparametric classifier (ParKNNO) is put forward to improve the classification performance by effectively incorporating correlation of traffic flows, and a novel prediction calculation method is proposed to accurately generate the prediction in real time.

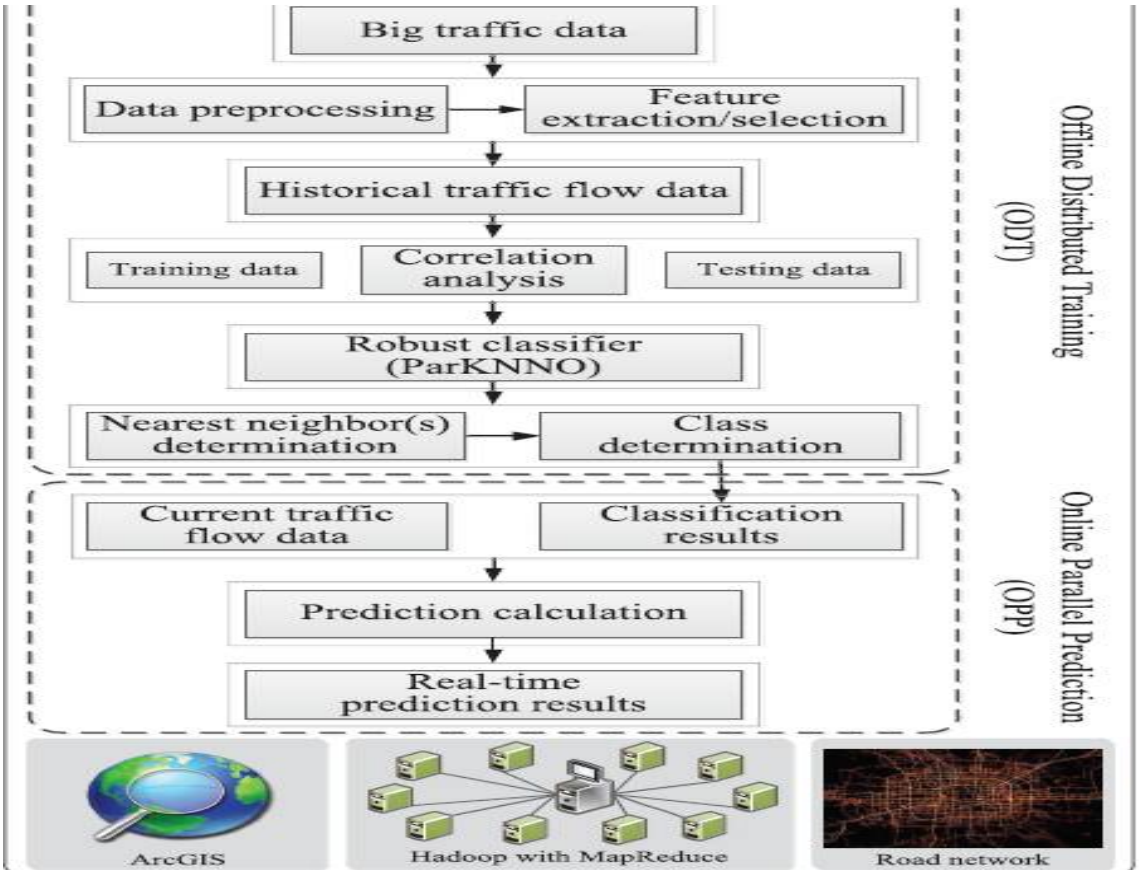

Fig.1 System Architecture

IV. SYSTEM IMPLEMENTATION

In this section we introduce some background information about the main components used in this project. A presents the k-NN algorithm as well as its weaknesses to deal with big data classification. Provides the description of the Map-Reduce paradigm and the implementation usedin this workA. k-NN and weaknesses to deal with big data. The k-NN algorithm is a non-parametric method that can be used for either classification and regression tasks.

This section defines the k-NN problem, its current trends and the drawbacks to manage big data. A formal notation for the k-NN algorithm is the following:LetTRbe a training dataset and TS a test set, they are formed by a determined number andof samples,respectively. Each sample xp is a tuple (xp1,xp2...,xpD,ω), where,xpfis the value of the f-th feature of the p-th sample.This sample belongs to a class ω, given by xωp, and a D- dimensional space. For the TRset

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

the classωis known, while it is unknown for TS. For each sample x test contained in the TSset, The k-NN model looks for the k closest samples in theTRset. To do this, it computes the distances between x test and all the samples of TR. The Euclidean distanceis commonly used for this purpose. Thekclosest samples(neigh1,neigh2, ...,neigh) are obtained by ranking (ascending order) the training samples according to the computed distance.

System Architecture:

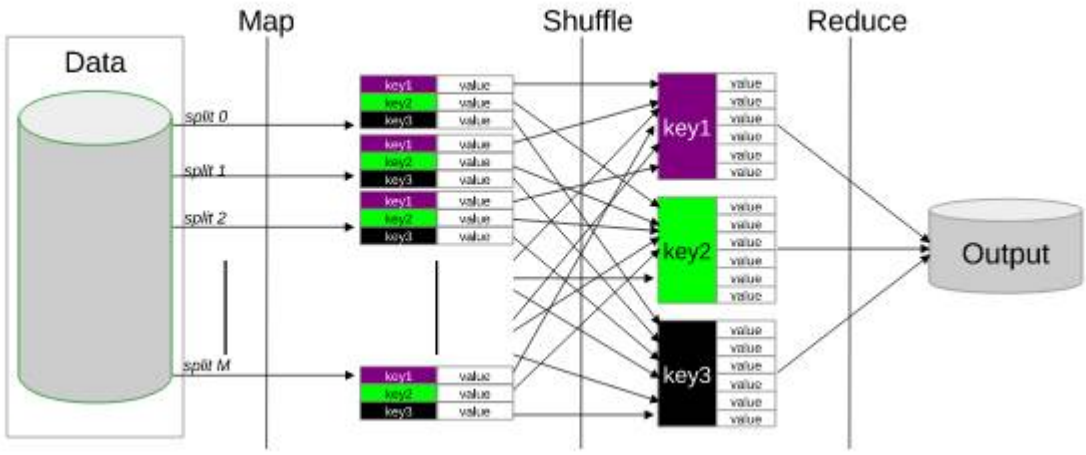

Fig2: The Map-Reduce workflow

V. EXPERIMENTAL RESULTS

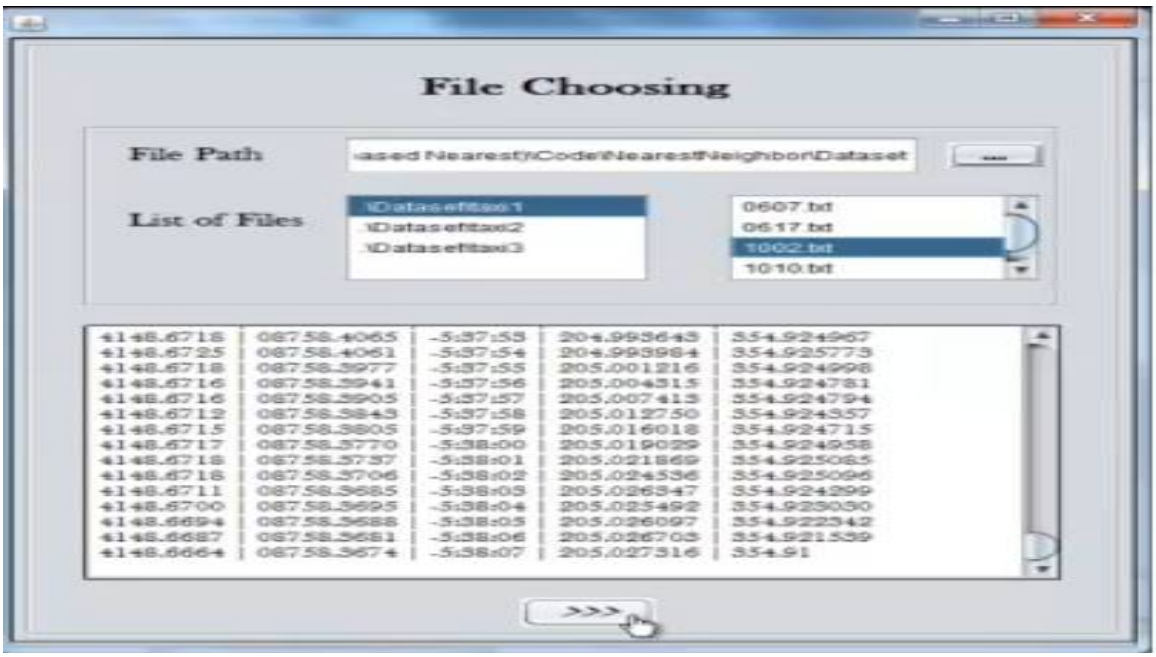

Fig3: File browse

User can browse dataset.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

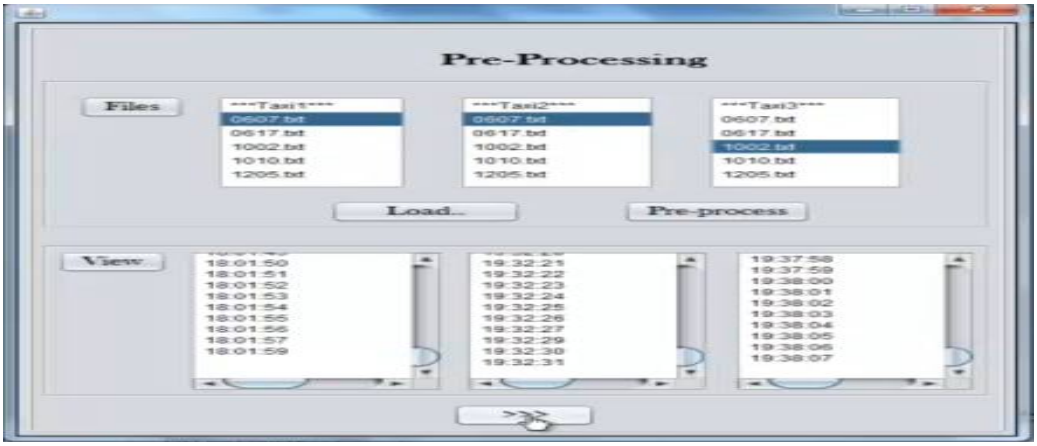

Fig4. File pre-processing

While user correct uploaded file, then goes to pre-process the file.

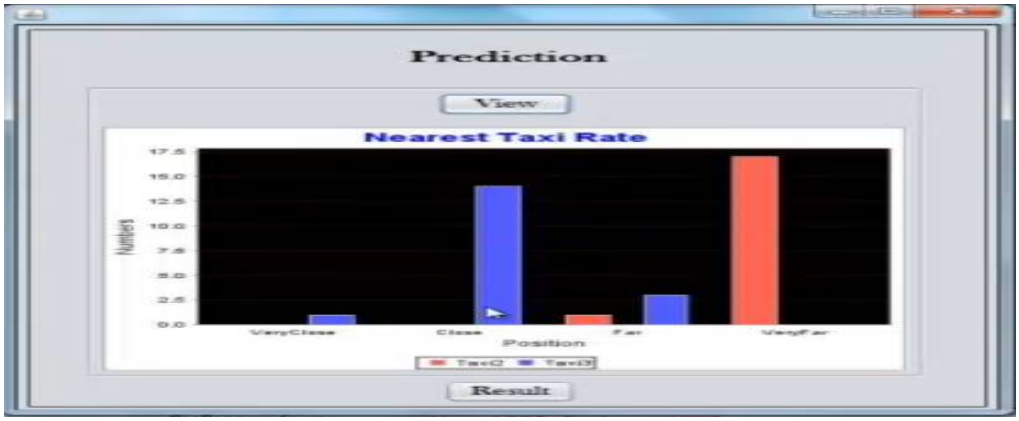

Fig5. Prediction

User see prediction graph of choose dataset.

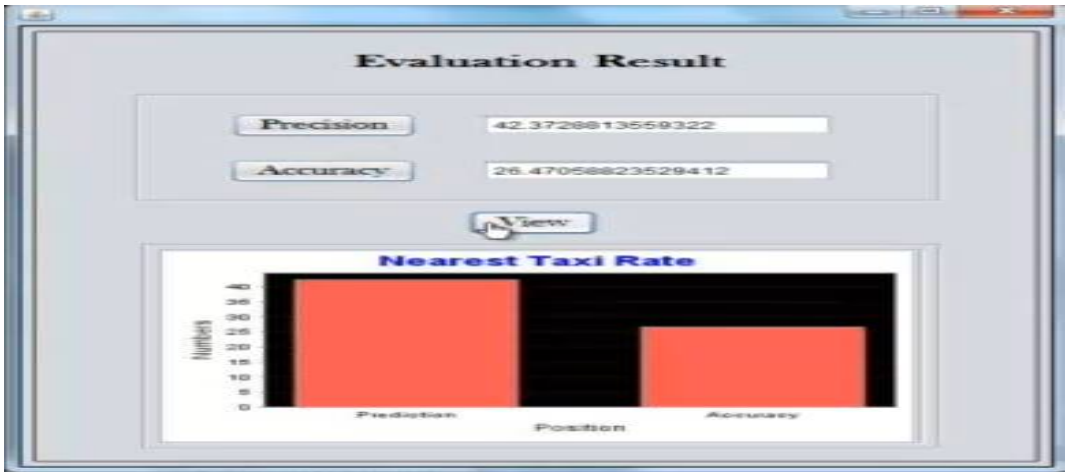

Fig6. Actual result

User see actual result graph of choose dataset.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

VI. CONCLUSION

From our project the prediction of the flow of traffic has been calculated earlier and that it will be helpful for the people to handle the future traffic that was created by this high population world. We proposed that minimizing the flow of vehicles will reduce the continuous traffic that was happed in the particular area or place which let the other peoplesto move quickly from that place. The Prediction process is updated continuous will lead the change in the map regularly where the user will aware of the place where they want to go. From our literal research we presented that the implementation of thisworking processwas not fullfilled by any countries. If it was implemented in any metro cities it will lead a rapid change in the flow of traffic in cities.For the future work, the traffic signal must be implemented by the Intelligent Traffic System (ITS).In order to calculate the flow of vehicles in the prediction process. And the data base have to be improved to store the large amount of data which contains all the video, images and other graphical elements have to store in that database..

REFERENCES

[1] Nikhil Apte, Mats Forssell, and A. Sidhwa, "Predicting Movie Revenue". 2011.
[2] Joshi, M., Das, D., Gimpel, K., & Smith, N. A. (2010). "Movie reviews and revenues: An experiment in text regression".In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 293-296). Association for Computational Linguistics.
[3] Bhave, A., Kulkarni, H., Biramane, V., & Kosamkar, P. (2015). "Role of different factors in predicting movie success".In Pervasive Computing (ICPC), 2015 International Conference on (pp. 1-4). IEEE.
[4] Mestyán, Márton, Taha Yasseri, and János Kertész. "Early prediction of movie box office success based on Wikipedia activity big data." PloS one 8.8 (2013): e71226.
[5] Jain, Vasu. "Prediction of Movie Success using Sentiment Analysis of Tweets." The International Journal of Soft Computing and Software Engineering3.3 (2013): 308-313.
[6] Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
[7] Rui, Huaxia, Yizao Liu, and Andrew Whinston. "Whose and what chatter matters? The effect of tweets on movie sales." Decision Support Systems 55.4 (2013): 863-870.
[8] Apala, Krushikanth R., et al. "Prediction of movies box office performance using social media." Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on. IEEE, 2013.
[9] Sharda, R., & Delen, D. (2006): "Predicting box-office success of motion pictures with neural networks". Expert Systems with Applications, 30(2), 243-254.