

An Approach To Analyze Different Route Factors Using Hadoop Framework

D. Swetha Priya^{1*}, M. Humera khanam²

^{1,2}Department of computer science and Engineering, Sri Venkateswara University College of Engineering, Sri Venkateswara University, Tirupati, India

**Corresponding Author: devarapuswethapriya942@gmail.com, Tel.: 8074841434*

Available online at: www.ijcseonline.org

Accepted: 19/Dec/2018, Published: 31/Dec/2018

Abstract— Whenever a person travelling from one place to another place. There are different routes between different regions. A person travelling from source to destination chooses a path based on different factors. Route choices from source to destination play an important role. It helps the user to choose the best route from many routes present based on different considerations. The traveller chooses the route with best factors like less time and distance. Such types of route factors are the main reasons to choose the route. We here develop a visual analytic system to display few more route choices. Here, based on the route factors the route with best factors is chosen as the best path and viewed to the user. We study and analyse route factors based on dataset. We analyse the dataset and a system with best and multiple route factors is developed using hadoop Framework.

Keywords— Route factors , Hadoop Framework, Big data.

I. INTRODUCTION

Now-a-days a variety of large amount of data has been produced from the sensing technologies [1]. Traffic data is large amount of data. The data collected from these can be grouped under BigData. To travel from one region to other region there are multiple routes. Route choices are the powerful measures that make wise declarations on making a decision to choose a route to travel. These systems not only help taxi drivers, it also helps new city planners in making route choices. This work mainly concentrates on routes which can be chosen for best travelling .There are multiple factors that affect making route choices. As multiple people are settling down in the city the crowd in the city also increases [2]. They are different considerations to choose a route. Among those city crowd is one of the factor. However factors like time and distance factors miserably fail to make perfect decision to choose a route for travelling. This project mainly concentrates on multiple route factors for making perfect travelling decisions.

There is a fact that users travelling from origin to destination do not always choose the same path for travelling. They choose routes depending upon their comfort level and satisfaction [3]. Route switching is one of the factors for route choice analysis. When-ever there is construction of road, the transport of large vehicles with

hazardous material cannot be travelled in that route. There will be need to select the preferred routes for the purpose of transportation [4]. There are so many factors that help in analysing route choice behaviour. Road network data is large amount of data. The best way to implement bigdata is through hadoop framework. This project concentrates on implementation of large amount of road network data using hadoop framework.

For the Rest of paper is organised as follows, Section II contain the related work where we analyse different route factors. Section III contains input dataset collection and its implementation. Section IV contains details about methodology with flowchart and algorithm. Section V describes results and discussion and Section VI concludes research work with future directions.

II. RELATED WORK

We, analyse the route related choices of different regions by using filtering techniques. There are multiple factors related to each and every path from the source to destination. These attributes help the traveller to choose the best route from all the routes present. The objectives of the study will determine the amount of detail needed in the various parts of the project. We make a study of multiple factors like traffic light count: gives the number of traffic lights on the way to travel.

Latitude and longitude values are taken into consideration in order to identify the regions of interest. Also we make a study of total time and total distance, Transport of heavy vehicles is other factor it helps heavy load vehicles in making route choice decision, details of crowded areas in the city. The route preference is the factor which analyses the route which is most likely to be chosen by most of the travellers.

Another predominant factor is economic considerations. That is transport Operational Costs Travel distance, time and the transportation and operational costs of alternative routes must be taken into account when considering the various options available. In road networks, the congested area i.e., heavy crowded areas can be represented by statistical methods. Clustering could be used for finding the areas with heavy crowd. Classification and clustering techniques are used to identify the models for route safety.

Here we, go through few stages. In stage 1, we have a interactive query for the traveller to understand the system and one can enter the source current location) and destination of interest. In stage 2, system analyses and performs filtering techniques in order to extract the routes. In stage 3, an overview of multiple routes is viewed along with best path among all the specific paths from origin to destination of traveller's interest.

The main aim is to find the effective factors that directly or indirectly inflicts the route choices in states of Andhra Pradesh based on the dataset using hadoop framework. The unstructured data is tuned into structured, which will be in the user understandable format.

III. DATA SET COLLECTION

Input Data, We take the dataset collected form internet taken as our experiment data. Here we collected the GPS dataset of the state Andhra Pradesh in India. The data collected is divided into two parts. Each dataset contains a large number of sampling points. As per dataset 1 contains 12048576 sampling points and dataset 2 contains 13048421 sampling points. Size of the dataset1 1GB and dataset2 size is 1GB Each sampling point contains the following attributes: Total time, Total distance, Number of traffic lights, Route Preference, safe/unsafe node, latitude and longitude values the route, crowded areas, Heavy loaded vehicles, Day, .

Attribute 1 is time. It is in the format of Hours: minutes: seconds, it helps the user to view the total time to travel from the chosen source to destination of different routes. Attribute 2 is total distance of the journey; Attribute 3 is count of Traffic light junctions. As we travel from source to destination there will be traffic signal junctions. Where there is intersection point of four roads there will be a traffic signal point. Whenever, traffic signal comes on our way to destination. We have to wait at least for few minutes which may increase our travel time. With the count of number of traffic lights user can analyse the time and can choose the route with less number of traffic signals and can reduce

travelling time. Attribute 4 is Route preferred. This attribute concentrates on the route safety which means the route which is most preferred by the travellers. This is implemented based on ranking method. Rank based technique is that in which rank is assigned to each route. Rank may be a numerical value. The route much preferred by the person is given a high rank expect to be as 10, the route which preferred less by the travellers can be assumed can be given rank which is less than the highest assigned value. Based on the rank the traveller/user can decide the path to travel from desired source to destination.

Attribute5 is safe/unsafe node this attribute tells the user about the route which is safe and which is not safe. It is a Boolean value if Boolean value is 1 then user can choose the route to travel safely. If Boolean value is 0 then it indicates that route is not safe to travel it tells whether the route is well developed or not. Attribute 6 is routes for heavy vehicles. Routes which are highly constructed can be allowed for movement of heavy loaded vehicles. Here in Hadoop it can be taken as a Boolean value. If it is 0 then Heavy loaded vehicles cannot be allowed in the route. Else they can be allowed to move. Day is another attribute which says about which day is better to travel in particular route. Latitude and longitude values are also recorded. Another attribute is high crowd areas here the routes which have high traffic is verified.

Data must be pre-processed and should be analysed in order to implement it. Pre-processing should be made to convert the unstructured data to structured data. Pre-processing is a task to analyse the data. Structured data is easy understandable by the end-user.

It is easy to use and can be processed faster .Unstructured data is which is not in textual format it contains images, video, audio files. The collected raw data will be saved in CSV format i.e; Comma Separated Value format and it can be opened in XL sheet. Missing values means that no data stored in the variable and unnecessary data should be eliminated from the dataset. This processed data should be uploaded to HDFS(hadoop file system). Based on the factors a person travelling, form a set of attractive alternatives. From this set one can make choices [6].

IV. METHODOLOGY

The data collected is the raw data. This data is unstructured data. This raw data is not in required format to analyze the system. The dataset must be preprocessed and should be changed to a structured data. We have to remove the duplicates present.

Data pre-processing:

Data pre-processing is one of most used technique to analyze the data in data mining. Here the missing values should be cleared and noisy data i.e; unwanted data must be removed to turn this raw data into data we required. It should be made into structured data which can be easily processed.

The unrelated data must be eliminated from the raw data. This data should be uploaded to Hadoop File System. Hadoop is a framework. Hadoop is that works on Map-Reduce strategy. Map Reduce is a software Framework for processing large amounts of data. Hadoop helps in processing of large amount of data in much less time. In Map-reduce Map performs sorting and filtering and reduce perform summary operations. Hadoop runs with the command prompt in Ubuntu operating system. After that, the dataset is uploaded to Hadoop File System. Then it is undergone Filtering technique.

Trajectory filtering and route extraction:

Filtering technique is applied and trajectories are filtered based on the source and destination. There are many paths from the source to destination. The routes are filtered and the route related factors are added to every route from source to destination. Where the source matches with user given source and destination matches with user specified destination, then the number of routes with different route factors is viewed using the filtering technique. Similarly destination also verified.

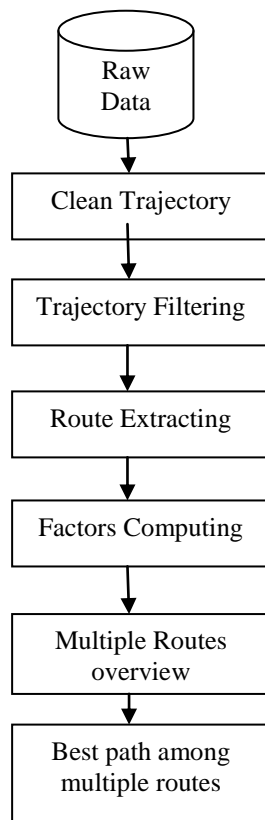


Figure1: Flowchart

Here we collected data is route related data of Andhra Pradesh state in India. This data contains gps dataset.

Different routes from source to destination are viewed i.e routes are extracted. Various route related factors for each route are computed. The different route related factors are Total Time, Total Distance, Route importance i.e the route most preferred, Day, it says which route can be chosen on specific day etc factors are computed. After route factor exploration multiple routes are viewed.

A review of the road hierarchy for the Study Area must be evolved. If a road hierarchy has not been formulated then it should be developed by analyzing the dataset. Identification of routes suitable for the movement of trucks or heavy loaded vehicles is essential. Thus multiple routes have a specific overview. Here for path or route generation we use a well-known method called as K-Shortest path algorithm which generates k-shortest paths from source to destination [7].

After the multiples routes overview, we consider the factor that predominantly affects the route choice i.e., we assumed it to be as less distance, less number of traffic lights, heavy loaded vehicles pass through, which are analyzed by conducting and observing the dataset. Depending upon these factors the route best for vehicles is being computed and best path is generated.

3.3.3 Algorithm:

Trajectory filtering:

Step1: Initialize the Source(s) and Destination (d)

Step2: Read the Dataset (da)

Step3: for each line do

Step4: if source(s) and destination (d) belongs to line

Step5: store the path in sp

Step6: end if

Step7: end for

Step8: return sp

Algorithm outlines the mining process

V. RESULTS AND DISCUSSION

Result and Analysis:

Results views the multiple paths from source to destination showing the attributes and also views the best path among multiple paths. The visual analytic system is mainly developed in java with hadoop framework. Various implementation details are discussed with in figures. We develop code for the mapper function and also for reduce function. Hadoop mainly runs with commands. We here use Hadoop5.2 for the project.

Figure4 shows the comparison of different lengths. It compares the two routes between source and destination. Here source is Tirupati and destination is Anantapur. Route 1 has a less route length than route 2 comparatively. Based on this factors user could be able to choose route with less Routelength.

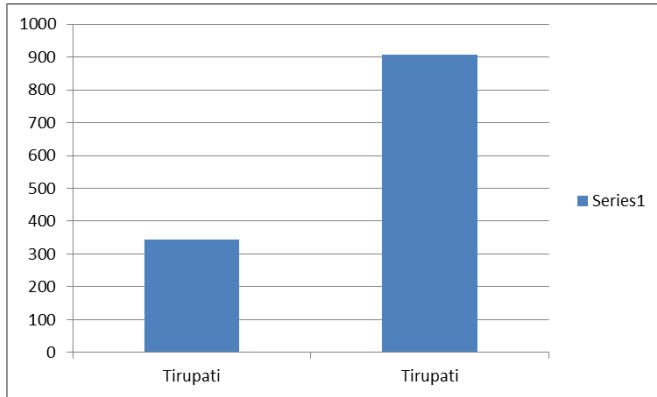


Figure 2: comparison of routes length.

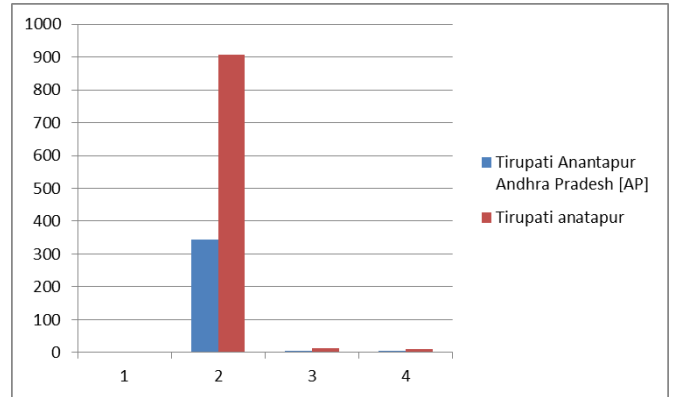


Figure5 : comparison of different route factors among routes.

Figure3 shows the comparison of routes which are highly preferred by the travellers depending upon the ranking method the route. Rank is assigned to every route figure shows route1 is less preferred compared to route 2. Here the source is Tirupati and destination is Anantapur

Figure 5 shows the comparison of different factors of two routes between the source:Tirupati and Destination:Anantapur. Different route factors are compared and the user can choose the route1 or route 2 based on the factors.

Along with these factors we also have another factors which can viewed as Boolean values as 0 and 1. Another factor Day is that which route can be chosen on a particular day for safe and comfortable journey. Thus route factors among different routes from origin to destination are compared and analysed for the convenience and best travelling experience. Users thus analyse and choose the routes with best factors.

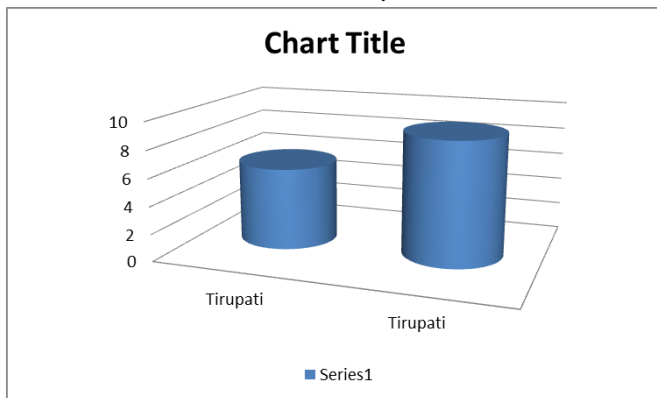


Figure 3: Figure shows the routes to tirupati which has highest preference.

VI. CONCLUSION AND FUTURE SCOPE

Here various route factors that affect route choice decision making are been analysed based on the collected dataset. We list the factors that are drawn from travel movements of a traveller .This project concentrates handling large datasets based on hadoop framework.

In the future, is about to combine large datasets even for different regions based on GPS dataset. We can analyse and few more factors that affect the route choice analysis can be added for the better route choice decision making.

REFERENCES

- [1] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 5, no. 3, p. 38, 2014.
- [2] Vaishali Y. Chandurkar "A Survey on Evaluation of Traffic Mobility Using Clustering" International Journal of Innovative Research in Computer and Communication Engineering.
- [3] Alessandro Vacca a, Italo Meloni "Understanding route switch behavior: an analysis using GPS based data" Transportation Research Procedia 5 (2015) 56 – 65
- [4] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatiotemporal causal interactions in traffic data streams," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 1010–1018

Count of number of traffic lights. The following figure shows the comparison of number of traffic lights between two routes from the source to destination.

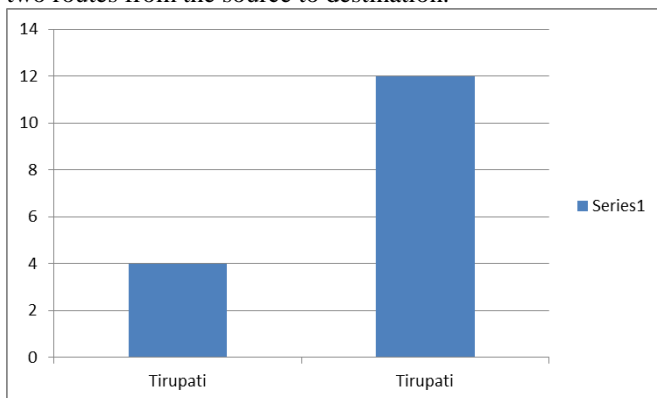


Figure 4: comparison of number of traffic lights.

- [5] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-Drive: enhancing driving directions with taxi drivers' intelligence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 220–232, 2013. Carlo Giacomo Prato, Victoria Gitelman and Shlomo, "Mapping patterns of pedestrian fatal accidents in Israel", *Accident Analysis and Prevention*, pp.54–62, Jan 2012.
- [6] H. Li, R. Guensler, and J. Ogle, "Analysis of morning commute route choice patterns using global positioning system-based vehicle activity data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1926.1, pp. 162–170, 2005.
- [7] Shlomo Bekhor, Moshe E. Ben-Akiva, M.Scott Ramming, "Evaluation of choice set generation algorithms," Springer science, 2006.