International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 4210–4215

A Unified And Innovative Route Prediction Using Bigdata Trajectory Clustering

Auk Pula Mounika , S.Vasundra2

1. M.Tech Scholar, Department of Computer Science and Engineering, JNTUA College of Engineering, JNTUA University, Anantapuramu, Andhra Pradesh

2. Department of Computer Science and Engineering, JNTUA College of Engineering, JNTUA University, Anantapuramu, Andhra Pradesh.

ABSTRACT

Clustering large number of trajectories is an important aspect in intelligent transportation system. It reveals valuable information about traffic like how traffic changes from time to time and day to day.

The existing system like Flow-based trajectory clustering schemes is not capable for analyze routes on a city road network. In existing NEAT[1] system, it is very tough to determine trajectories count on each cluster since it is one of the crucial reasons for planning routes in city road network[2]. In existing NETSCAN [3] system, first it scans the traffic over a network [4] one trajectory can assign to two or more clusters. Therefore, it is very tough to identify the correct trajectory count on each cluster.

The proposed system have capability to predict route in dense city road network by using A algorithm [5] with heuristic function.*

KEYWORDS: *Clustering, Dijkstra, Neat, Net scan, Prediction, Road network.*

INTRODUCTION

The process of analyzing trajectories in road networks is an important aspect for predicting routes. If analyzed effectively it reveals valuable information about traffic how traffic changes from time to time and day to day, it is also used to determine the highly prevalent routes[6] taken by the passengers for planning trips[7] and also reduces the traffic and accidental issues. The main aim of project is to predict route in dense city road networks using A* algorithm with heuristic function.

A* is same as Dijkstra Algorithm these two are used for finding best path. A* algorithm can find path by using heuristic function [8]. Heuristic function can be calculated as h(n)=a(n)+b(n) where a(n) represents distance between starting point to reached one and b(n) represents distance between reached point to destination point. Sum of the a(n)and b(n) gives the result of heuristic function.

The rest of the paper organized as follows: section 2 discusses related work, section 3 discusses methodology, section 4 discusses experimental results, section 6 discuss conclusion.

II. RELATED WORK

In urban areas, vehicle trajectories are typically obliged to a mind boggling road connect with many parallel and perpendicular road segments and intersections, which makes their time movement extremely unpredictable. Because of the vulnerability of moving objects [9], the vast majority of the current direction expectation strategies as it were concentrate on anticipating momentary halfway trajectories. They have poor expectation exactness for long haul direction forecasts and they don't function admirably for assessing constant and complete trajectories existing trajectory clustering is limited in its success by

ISSN: 2233-7857 IJFGCN

complex conditions such as application scenarios and data dimensions. The existing system is based on two Road network constrained trajectory clustering approaches: 1. Traditional clustering approaches are used for clustering trajectories [10] 2. Road segments are clustered based on their density and flow there are two approaches of flow-based trajectory clustering NEAT and NET SCAN these techniques are fast and efficient, but these are not appropriate in city road network, why because in net scan one trajectory assigns to multiple number of clusters difficult to count what present trajectory is using NEAT it is difficult to count trajectories in each cluster counting trajectory is important for finding frequency of traffic flow. In city road networks it contains multiple numbers of trajectories so that these techniques are not suitable. It is Due to this A* can gives efficient results like exact distance and unique route or predicting a route between two points.

III. METHODOLOGY

A * algorithm is used for predicting route in dense city road network based on the heuristic function h(n)=a(n)+b(n) where a(n) represents distance between starting point to reached one and b(n) represents distance between reached point to destination point. Fig 1 shows the schematic view of whole process for predicting route in city road networks.

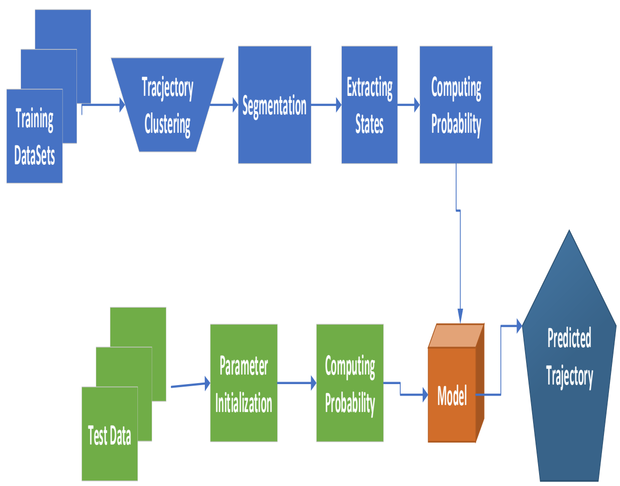


Fig 1: schematic view of proposed work

In proposed system first creates a dataset after that clustering can be performed by using K means clustering algorithm. K means is a unsupervised algorithm segmentation is performed among the trajectories, by extracting the states probability is computed once model is build it can be tested for knowing the performance.

K mean clustering algorithm is used For dividing the datasets. Here K indicates number of divided clusters [11]. Dataset can have set of data items for creating n clusters then k value become n and centroid is also n, centroid is nothing but mean, this can be chosen as randomly, difference can be calculated between centroid and data items which data item having minimum value with particular centroid that can be placed on one cluster else placed on other cluster. The work flow of K means clustering algorithm is follows

ISSN: 2233-7857 IJFGCN

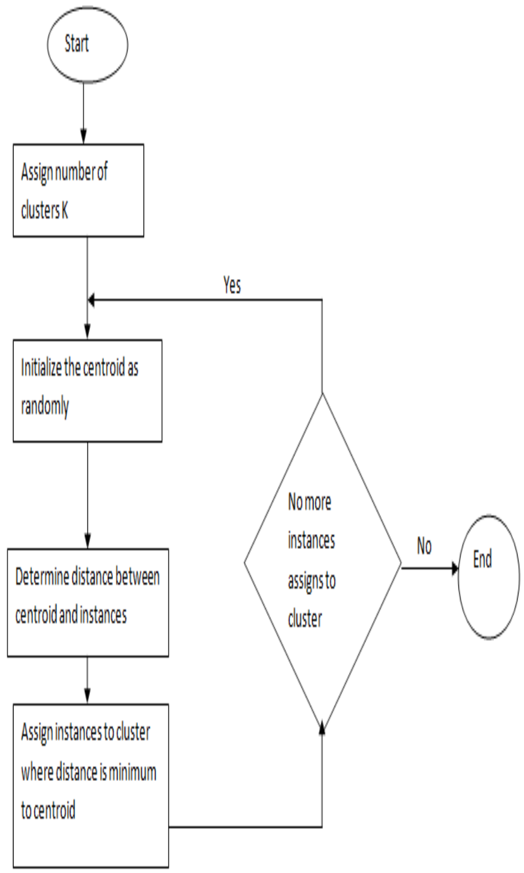


Fig 2: workflow of K means algorithm

ALGORITIHM:

Step1: Data evaluation is the critical initial step to analyzing the data by using statistics or graphs

Step2: preprocessing the data

Step3: calculating trip distance and trip time

Step4: predicting the route by using A* algorithm.

In Data evaluation EDA (exploratory data analysis) was used to analyze new data sets the outcome of this analysis is summarized by statistics or graphs. Data preprocessing is a crucial step to prepare the data, there are many methods for preparing the data they are data cleaning, data transformation and feature selection. Trajectory cleaning and transformation methods are used to remove outliers and standardize the data to create a model. Distance can be calculated between every two correlative sets of directions in kilometers. Each pair of directions are isolated 15 seconds in time. At that point, the trip time in minutes will be: (number of pair of directions - 1) x 15 seconds x 1 moment/60 seconds. The normal speed is basically the trip distance separated by the trip time in hours. The outcome is in km/h. The top speed is the best speed of the considerable number of segments of the trip. The speed of one segment is determined as the distance of the segment separated by 15 seconds and changed over to km/h. For predicting the route A* algorithm was used. A* is same as Dijkstra Algorithm used for finding a shortest path. A* is same as Greedy Best-First-Search it is also uses a heuristic to guide itself. In the simple case, it is as fast as Greedy Best-First-Search. A* algorithm can find path by using heuristic function. Heuristic

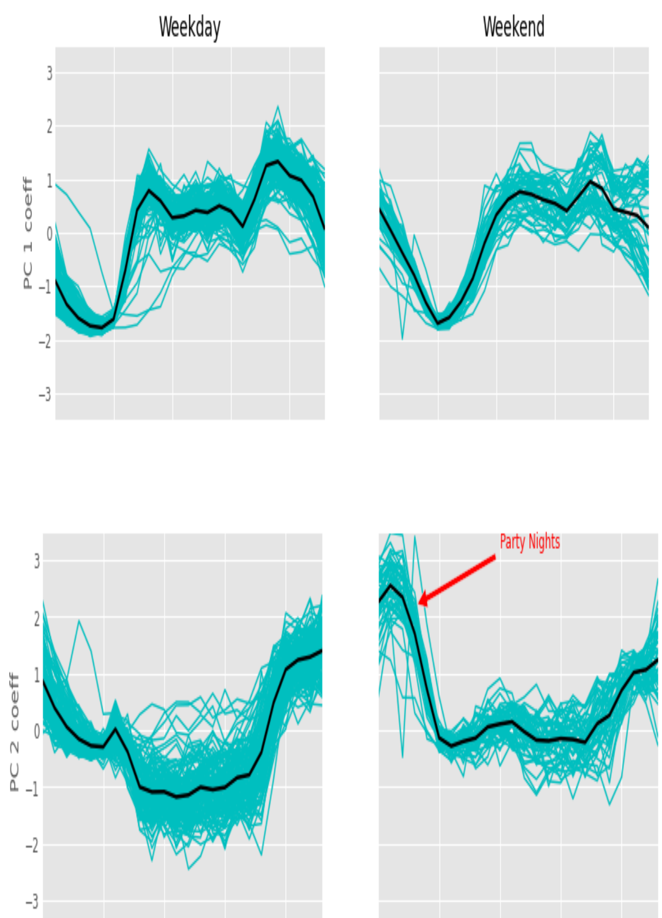
ISSN: 2233-7857 IJFGCN

function can be calculated as h(n)=a(n)+b(n) where a(n) represents distance between starting point to reached one and b(n) represents distance between reached point to destination point. Sum of the a(n)and b(n) gives the result of heuristic function.

EXPERIMENTAL RESULTS:

The below graph fig:2 demonstrate how traffic can changes in weekday weekend day party nights early risers and late nights based on the traffic flow prediction of route can be done by applying A* algorithm shown in fig:3

In below fig2 x axis shows hours of day, and y axis shows PC (principle components) coefficient. PC comes under principle component analysis PCA which reduces the dimensionality of large datasets and gives the more important information. In below fig3(a) x axis is hours of day and y axis is PC coefficient, thick black color line represents flow of traffic in weekday and weekend, the traffic raises up to 1.3 of coefficient value and in weekend traffic raises up to 1 coefficient value . In fig3 (b) x axis is hours of day and y axis is PC coefficient. Thick black line represents the flow of traffic in party nights; the flow of traffic in party nights reaches upto2.5 of coefficient value. In fig3(c) x axis shows hours of day and y axis show PC coefficient, thick black color line represents flow of traffic in early risers and late mornings. In early risers the traffic floe reaches 2.5 of coefficient value and in late mornings the traffic raises up to 1.4 of coefficient value.



ISSN: 2233-7857 IJFGCN

International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 4210–4215

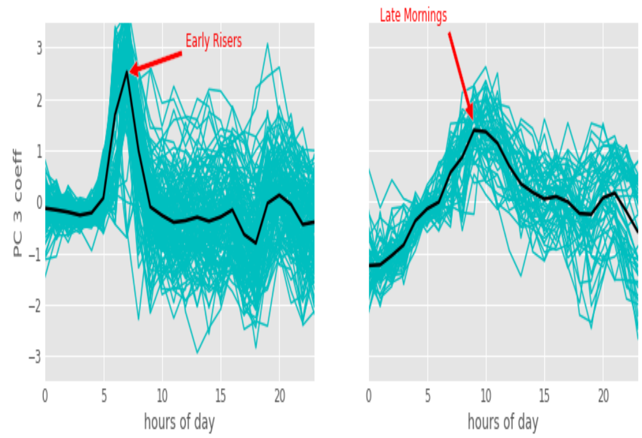


FIG 3: graphical representation of Flow of traffic in (a) weekend days, week days, (b)party nights ,(c)early risers and late mornings.

Prediction of routes using A* algorithm can be showed in fig(4) pink color point represents source and green color point represents destination point

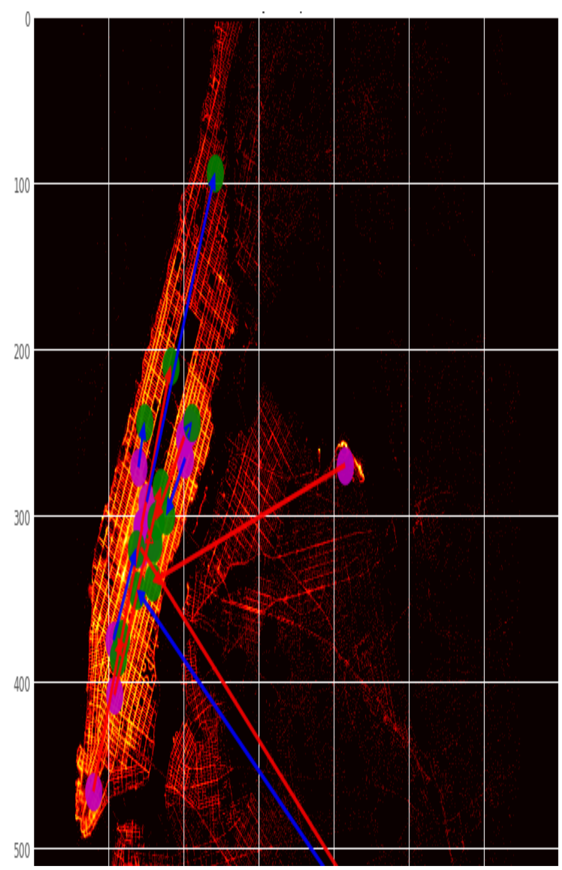


Fig4: Predicted routes

ISSN: 2233-7857 IJFGCN

Conclusion:

Clustering large number of trajectories in city road network is difficult to find paths using trajectory clustering schemes due to the difficulty to finding best paths with less amount of time and planning the trips proposed system use A* by using heuristic function route can be predicted in dense city road network.

References:

[1] B. Han, L. Liu, and E. Omiecinski, "Road-network aware trajectory clustering: Integrating locality, flow, and density," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 416–429, Feb. 2015.
[2]D. Kumar, H. Wu, S. Rajasegarar, C. Leckie, S. Krishnaswamy and M. Palaniswami, "Fast and Scalable Big Data Trajectory Clustering for Understanding Urban Mobility," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 11, pp. 3709-3722, Nov. 2018, doi: 10.1109/TITS.2018.2854775.
[3]P. Rathore, D. Kumar, S. Rajasegarar, M. Palaniswami and J. C. Bezdek, "A Scalable Framework for Trajectory Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 10, pp. 3860-3874, Oct. 2019, doi: 10.1109/TITS.2019.2899179.
[4] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 29, 2015.
[5] Xiao Cui and Hao Shi, "A*-based Path-finding in Modern Computer Games" Melbourne, Australia: IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.1, January 2011.finding in Modern Computer Games" Melbourne, Australia: IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.1, January 2011.
[6] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data,"*IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, Sep. 2013.
[7] Z. Hong, Y. Chen, and H. S. Mahmassani, "Recognizing network trip patterns using a spatio-temporal vehicle trajectory clustering algorithm," *IEEE Trans. Intell. Transp. Syst.*, to be published.
[Online]. Available: https://ieeexplore.ieee.org/document/8082114/,doi: 10.1109/TITS.2017.2754401.
[8]Dr. S. Vasundra *et.al,* CSE, JNTUACEA, Optimizing technique to reduce traffic load in peer-to-peer systems. International journal of advanced research in computer engineering and technology, Vol-4, issue-6, pp ISSN: 2278-1323, July 2015. IJARCET.

[9] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, "Online discovery of gathering patterns over trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1974–1988, Aug. 2014.

[10]D. Kumar, M. Palaniswami, S. Rajasegarar, C. Leckie, J. C. Bezdek and T. C. Havens, "clusiVAT: A mixed visual/numerical clustering algorithm for big data," 2013 IEEE International Conference on Big Data, Silicon Valley, CA, 2013, pp. 112-117, doi: 10.1109/BigData.2013.6691561.
[11]J. Xie and S. Jiang, "A Simple and Fast Algorithm for Global K-means Clustering," 2010 Second International Workshop on Education Technology and Computer Science, Wuhan, 2010, pp. 36-40, doi: 10.1109/ETCS.2010.347.

ISSN: 2233-7857 IJFGCN