11th International Conference on Recent Development in Engineering Science, Humanities and Management
Nutan Maharashtra Institute of Engineering & Technology, Talegaon Dabhade, Pune ESHM-18
24th-25th March 2018 www.conferenceworld.in ISBN : 978-93-87793-14-9

To Predict The Time-Location-Relationship Combined Service Recommendation For Taxi Drivers Using Clustering Techniques.

Narendra Patil[1], Ashwin Tate[2], Yogesh Kalbhor[3], Vaibhav Gawale[4]

Prof. Krishna Tayade[5]

Department of Computer Engineering,
Nutan Maharashtra Institute of Engineering and `Technology, Pune, (India)

ABSTRACT

Recently, Urban traffic management unaware about traffic situations in various area (Rush area) on time basis that why they travelling empty without passenger and got less profit at the end of the day. We propose a Time-Location-Relationship combined taxi service recommendation model (TLR) to improve taxi drivers' profits, uncover the knowledge of human mobility patterns, and enhance passengers' travel experience. Moreover, TLR model uses Gaussian Process Regression and statistical approaches to acquire passenger volume, mean trip distance and average trip time in functional regions during every period on weekdays and weekends, and allow drivers to pick up more passengers within a short time frame. Finally, we compare our proposed model with Auto Regressive Integrated Moving Average model (ARIMA), Back-Propagation Neural Network model (BPNN), Support Vector Machine model (SVM), and Gradient Boost Decision Tree model (GBDT) by using the real taxi GPS data.

We propose a density-based spatial-temporal clustering algorithm for geo-located data points, based on an extension of the SNN (Shared Nearest Neighbour) clustering. The proposed algorithm allows the integration of location, time and other semantic attributes in the clustering process. This algorithm can find clusters of different sizes, shapes, and densities in noisy data that will help us to predict rush hours in various area. We evaluate the effectiveness of our algorithm through a case study involving a New York City taxi cab pickup data and Maryland crime data. The experimental results show that the proposed algorithm can discover interesting patterns and useful information from spatial-temporal data. By using this system we can predict that, in which area there is rush ours right now and according to taxi drivers geo-location system will suggest him, nearest traffic area to get more passengers. That will be more helpful to get more passengers and increase profit.

Keywords:LBS (Location Based Service), GPS, Recommendation, spatial temporal clustering, shared Nearest Neighbour clustering.

I.INTRODUCTION

Now a day's taxi drivers are unaware about traffic situations in various area (Rush area) on time basis that why they travelling empty without passenger and got less profit at the end of the day. Because of not having information they have to travel empty and got big loss in business. To solve this issue we develop such system which will help them to find nearest rush area on time basis from driver's location. Whenever driver visit to new location he have to start application to get passenger. Application will take driver's geo-location and search for nearest rush places at that time. Then Application will automatically make groups of nearest rush area and show results using clustering. If driver got or does not got passenger in that area then he has to update into application, this will help for next prediction.

A. Problem Statement

Now a days taxi drivers are unaware about traffic situations in various area (Rush area) on time basis that why they travelling empty without passenger and got less profit at the end of the day. In our project we can predict that, in which area there is rush hours right now and according to taxi driver's geo-location system will suggest him, nearest traffic area to get more passengers.

B. Existing System and Problem

Typically, for an existing system, the customers need to book a taxi by telephone/internet in advance, and it is usually not free of charge. Most passengers hail a taxi along the road or stand where the taxis are available instead of booking a taxi. Taxi drivers are unaware about traffic situations in various area (Rush area) on time basis that why they travelling empty without passenger and got less profit at the end of the day.

C. Proposed System

We propose a Time-Location-Relationship combined taxi service recommendation model (TLR) to improve taxi driver's profits, uncover the knowledge of human mobility patterns, and enhance passengers travel experience. We can predict that, in which area there is rush hours right now and according to taxi driver's geo-location system will suggest him, nearest traffic area to get more passengers. That will be more helpful to get more passengers and increase profit.

Whenever driver visit to new location he have to start application to get passenger. Application will take driver's geo-location and search for nearest rush places at that time. Then Application will automatically make groups of nearest rush area and show results using clustering. If driver got/ does not got passenger in that area then he has to update into application, this will help for next prediction.

II.RELATED WORK

The prediction of the destination location at the time of pickup is an important problem with potential for substantial impact on the efficiency of a GPS enabled taxi service. While this problem has been explored earlier in the batch data set-up, we propose in these paper new solutions in the streaming data set-up. We examine four incremental learning methods using a Damped window model namely, Multivariate multiple regression, Spherical-spherical regression, Randomized spherical K-NN regression and an Ensemble of these methods for their effectiveness in solving the destination prediction problem. The performance of these methods on several large data sets is evaluated using suitably chosen metrics and they were also compared with some other existing methods. The Multivariate multiple regression method and the Ensemble of the three methods are found to be the two best performers. The next pickup location problem is also considered and the aforementioned methods are examined for their suitability using real world datasets. As in the case of destination prediction problem, here also we find that the Multivariate multiple regression method and the Ensemble of the three methods gives better performance than the rest.

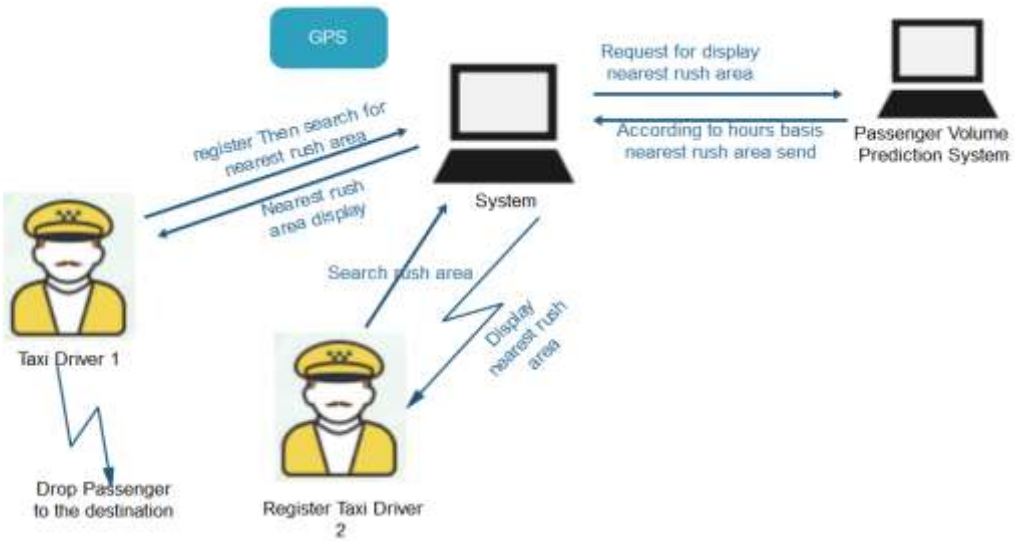
Density based clustering algorithm is one of the primary methods for clustering in data mining. The clusters which are formed based on the density are easy to understand and it does not limit itself to the shapes of clusters. This paper gives a detailed survey of the existing density based algorithms namely DBSCAN, VDBSCAN, DVBSCAN, ST-DBSCAN and DBCLASD based on the essential parameters needed for a good clustering algorithm. We analyse the algorithms in terms of the parameters essential for creating meaningful clusters.

Approximation algorithms are widely used in many engineering problems. To obtain a data set for approximation a factorial design of experiments is often used. In such case the size of the data set can be very large. Therefore, one of the most popular algorithms for approximation Gaussian Process regression can hardly be applied due to its computational complexity. In this paper a new approach for a Gaussian Process regression in case of a factorial design of experiments is proposed. It allows to efficiently compute exact inference and handle large multidimensional and anisotropic data sets.

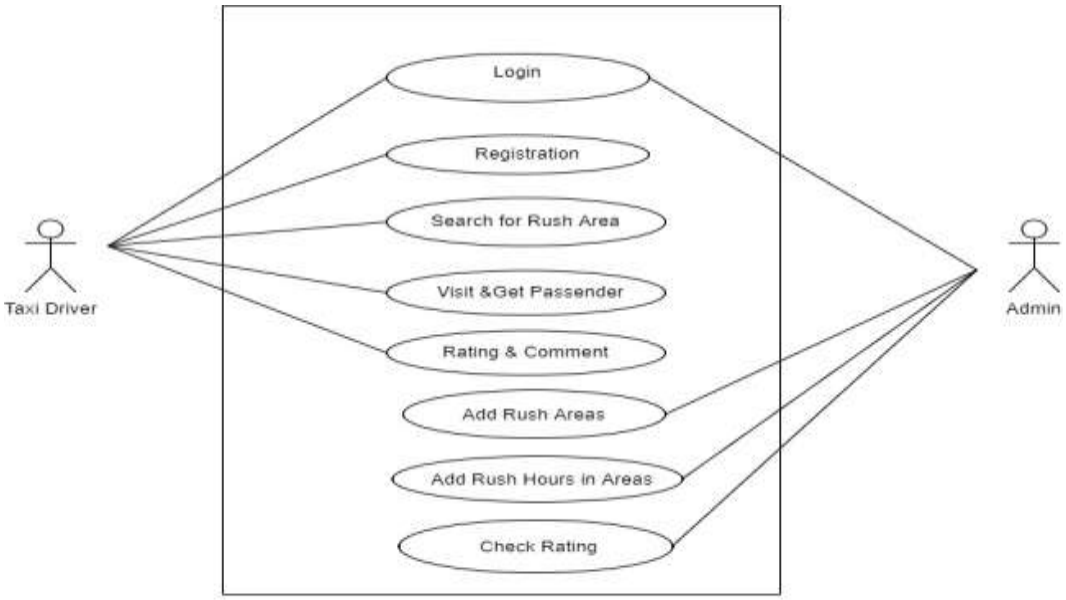
Detailed land use, which is difficult to obtain, is an integral part of urban planning. Currently, GPS traces of vehicles are becoming readily available. It conveys human mobility and activity information, which can be closely related to the land use of a region. This paper discusses the potential use of taxi traces for urban land-use classification, particularly for recognizing the social function of urban land by using one year's trace data from 4000 taxis. First, we found that pick-up/set-down dynamics, extracted from taxi traces, exhibited clear patterns corresponding to the land-use classes of these regions. Second, with six features designed to characterize the pick-up/set-down pattern, land-use classes of regions could be recognized. Classification results using the best combination of features achieved a recognition accuracy of 95%. Third, the classification results also highlighted regions that changed land-use class from one to anotherand such land-use class transition dynamics of regions revealed unusual real-world social events. Moreover, the pick-up/set-down dynamics could further reflect to what extent each region is used as a certain class.

In the taxicab industry, a long-standing challenge is how to reduce taxicabs' miles spent without fares, i.e., cruising miles. The current solutions for this challenge usually depend on passengers to actively provide their locations in advance for pickups. To address this challenge without the burden on passengers, in this paper, we propose a cruising system, pCruise, for taxicab drivers to find efficient routes to pick up passengers to reduce cruising miles. According to the real-time pick-up events from nearby taxicabs, pCruise characterizes a cruising process with a cruising graph, and assigns weights on edges of the cruising graph to indicate the utility of cruising corresponding road segments. Our weighting process considers the number of nearby passengers and taxicabs together in real-time, aiming at two scenarios where taxicabs are explicitly or implicitly coordinated with each other. Based on a weighted cruising graph, when a taxicab becomes vacant, pCruise provides a distributed online scheduling strategy to obtain and update an efficient cruising route with the minimum length and at least one arriving passenger. We evaluate pCruise based on a real-world GPS dataset from a Chinese city Shenzhen with 14,000 taxicabs. The evaluation results show that pCruise assists taxicab drivers to reduce cruising miles by 42 % on average.

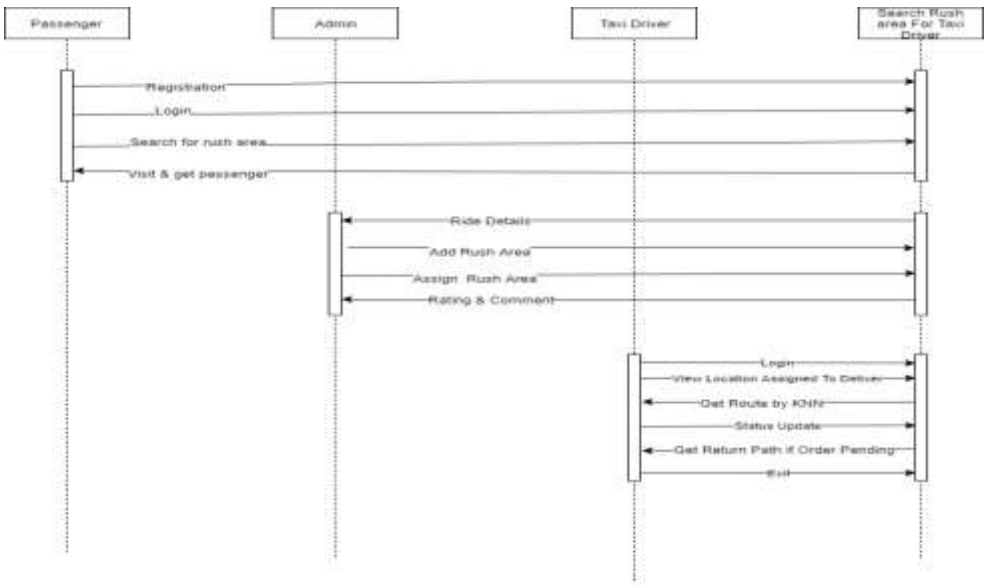
System Architecture



Use Case Diagram



Sequence Diagram



Mathematical Model

Mathematical Model for Proposed System

• Let S be a system that describes the System

• S={I,O,P,S,F}

Where,

I = Input of the system.

O=Output of the system.

P =Process of the system.

S =Desired outcome is generated.

F =Desired outcome is not generated due to system error.

• Identify input as I

S={I,..}

Let I ={i1,i2,i3,..id}

The input will be search nearest rush area.

• Identify output as O

S ={I,O,}

O= The taxi driver will receive the list of nearest rush area.

• Identify the processes as P

S= {I, O, P,..}

P= {E, D}

E= {parameter, Search nearest rush area}

D= {parameter, Exact location of rush area}

• Identify success case as S.

S={I,O,P,F,S,}

S=When data is accessed by authorized user.

• Identify failure cases as F

S={I,O,P,F,.}

F=Failure occurs when the data is accessed by an unauthorized user.

III.ALGORITHM

- **KNN Algorithm**

The KNN algorithm is a robust and versatile classifier that is often used as a benchmark for more complex classifiers such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Despite its simplicity, KNN can outperform more powerful classifiers and is used in a variety of applications such as economic forecasting, data compression and genetic.

KNN falls in the supervised learning family of algorithms. Informally, this means that we are given a labelled dataset consisting of training observations (x,y) and would like to capture the relationship between x and y. More formally, our goal is to learn a function $h{:}X\rightarrow Y$ so that given an unseen observation x, $h(x)$ can confidently predict the corresponding output y.

The KNN classifier is also a non parametric and instance-based learning algorithm.

Steps in KNN:

1.Compute the similarity matrix

This corresponds to a similarity graph with data points for nodes and edges whose weights are the similarities between data points.

2. **Sparsify the similarity matrix by keeping only the *k* most similar neighbors**

This corresponds to only keeping the *k* strongest links of the similarity graph.

3. Construct the shared nearest neighbor graph from the sparsified similarity matrix

At this point, we could apply a similarity threshold and find the connected components to obtain the clusters.

4. Find the KNN density of each point

Using a user specified parameter, *Eps*, find the number points that have an KNN similarity of *Eps* or greater to each point. This is the KNN density of the point.

5. Find the core points

Using user specified parameter, *MinPts*, find the core points, i.e., all points that have an KNN density greater than *MinPts*.

6. Form clusters from the core points

If two core points are within a radius, *Eps*, of each other they are placed in the same cluster.

7. Discard all noise points

All non-core points that are not within a radius of *Eps* of a core point are discarded.

8. Assign all non-noise, non-core points to clusters

This can be done by assigning such points to the nearest core point.

- **Haversine Formula**

Calculate geographic distance on earth. If you have two different latitude and longitude values of two different point on earth, then with the help of Haversine Formula , you can easily compute the great-circle distance. Haversine is very popular and frequently used formula when developing a GIS (Geographic Information System) application or analyzing path and fields.

The Haversine formula is perhaps the first equation to consider when understanding how to calculate distances on a sphere. The word "Haversine" comes from the function:

$$\text{haversine}(\theta) = \sin^2(\theta/2)$$

The following equation where φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371km) is how we translate the above formula to include latitude and longitude coordinates. The angles need to be in radians to pass to functions:

$$a = \sin^2(\varphi B - \varphi A/2) + \cos \varphi A * \cos \varphi B * \sin^2(\lambda B - \lambda A/2)$$
$$c = 2 * \text{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c$$

IV.EXPECTED RESULT

- Whenever driver visit to new location he have to start application to get passenger.
- Application will take drivers geo-location and search for nearest rush places at that time. Then Application will automatically make groups of nearest rush area and show results using clustering.
- If driver got/ does not got passenger in that area then he has to update into application, this will help for next prediction.

V.CONCLUSION

In this , we have proposed a taxi service recommendation model named TLR by analyzing the quantitative relationship between passengers getting on and off taxis in different functional regions during each period.
We introduced a new density-based spatial temporal clustering algorithm that can find clusters of different shapes, sizes, and densities. The algorithm can automatically determine the number of clusters.

VI.FUTURE SCOPE

We will consider different social properties and multi-source datasets to improve our prediction accuracy. We also plan to quantitatively evaluate our proposed model using bus drivers income data. Besides, we will focus

on supply-demand matching and recommendation between passengers and taxis, which makes passengers find vacant taxis in less time.

VII.ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on 'To Predict The Time-Location-Relationship Combined Service Recommendation For Taxi Drivers Using Clustering Techniques. '.
We would like to take this opportunity to thank my internal guide Prof. Krishna Tayade for giving me all the help and guidance we needed. We are really grateful to them for their kind support and valuable suggestions were very helpful.

REFERENCES

[1] Y. Elmehdwi, B. K. Samantha,W. Jiang ,"Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data".

[2] G. Ghinita, R. Rughinis, "Online Cruising Mile Reduction in Large-Scale Taxicab Networks", November 2017.

[3] Arnab Kumar Laha,Sayan Putatunda, "Real Time Location Prediction with Taxi-GPS Data Streams", W.P. No.2017-03-02, March 2017.

[4] Mikhail Belyaev, EvgenyBurnaev and Yermek Kapushev, "Gaussian Process Regression for Structured Data Sets."

[5] Gang Pan, Guande Qi, Zhaohui Wu, Daqing Zhang, and Shijian Li, " Land-Use Classification Using Taxi GPS Traces" ,IEEE, March 2017.

[6] X. Kong, Z. Xu, G. Shen, J.Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," Future Generation Computer Systems, vol. 61, pp. 97–107, 2016.

[7] T. Qin, X. Guan, W. Li, and P. Wang, "Monitoring abnormal traffic flows based on independent component analysis," in IEEE International Conference on, Dresden, Jun. 2009.

[8] F. Xia, A. Ahmed, L. Yang, and Z. Luo, "Community-based event dissemination with optimal load balancing," IEEE Transactions on Computers, vol. 64, no. 7, pp. 1857–1869, Jul. 2015.