Customer Pickup Location Optimizer

Anugrah S, Aishwarya Kaneri
Go-Jek

Abstract:- One of the major goals of ride hailing services is to provide a good user experience to its customers by allocating the nearest available driver to the customer, hence reducing the time to wait by the customer and by minimising the cancellation of rides by drivers or customers. In this paper we propose PiLO a novel approach for recommending the optimum pick up location for customer such that the acceptance rate can be increased.

Keywords:- *Clustering, optimization, GPS location, data mining, S2 cell ID, map.*

I. INTRODUCTION

In ride hailing services, ensuring customer satisfaction is dependent on whether a driver will accept the customer's request within a time period. In most real world scenarios, the driver will reject the orders if the customer is ordering from a place far from him. There are also scenarios like apartment complexes where the driver might not accept orders. In such cases it is much better if we could suggest the optimal pickup location to customer such that his overall experience with ride hailing service can be improved.

Consider a case in which a customer has ordered a cab while standing next to a one way while the cab driver is ahead of the customer. If driver would pick the customer from the location he would have to cover a lot of distance as he would have to take a u turn and then go back to the location and then pick up the customer. Most drivers in this situation would actually decline to take the order. The customer would find the platform not very useful in providing him with services. These situations are quite common and reduces the trust that customers have in the platform.

The historical data that we collected indicates specific locations from which the acceptance rates are quite high. These could be areas like exit of mall, apartment complexes, mall etc. We call these areas positive areas. The data also indicates areas of high cancellation from drivers. We call these areas negative areas. These could be areas like an apartment within an enclosed compound or an area which would require the driver to travel additional distance. For ensuring that customer would have a flawless experience with the platform we would need him to move from negative areas to positive areas with minimal distance between them. To identify the positive areas we first cluster the customer pickup locations for the completed ride bookings. Similar approach is made for identifying the negative areas by clustering the customer pickup locations for the cancelled or the driver not allocated bookings. When a customer makes a booking from a location near to the negative area he is recommended to move to nearest area.

Compared to optimisations around assigning drivers who are willing to take customers at negative areas, we found that recommending customers to move to an area near to positive areas resulted in better acceptance rate.

II. DATA COLLECTION

A. Selection of S2 Cell

We selected a S2 cell which showed a low convergence rate of customers booking the ride and the drivers completing the ride. The reasons for a low convergence rate can be: (1) No driver available near the customer pickup location, (2) driver cancelled the ride due to some reason, (3) customer cancelled the ride due to reasons like long time to wait for the driver to reach the pickup location.

To pick up such a S2 cell, we collected all the negative ride bookings done where the ride was either cancelled by the driver or customer or the nearest available driver was not allocated for the ride. A scatter plot is plotted for the pickup locations of these negative bookings (as shown in Fig. 1.) .

The clustering algorithm of MeanShift is run on this data to identify the cluster which is densely populated with the negative bookings and has a low convergence rate. The latitude and longitude of the centroid of this cluster is converted to the S2 cell ID (as shown in Fig. 3.)

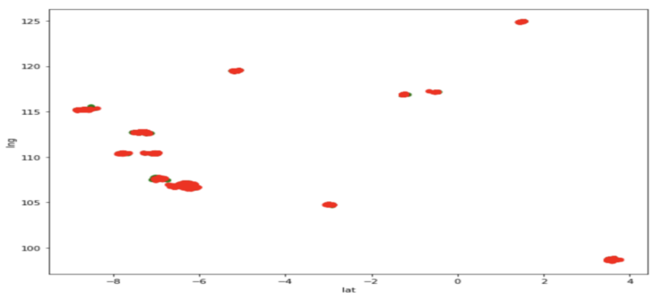

Fig 1:- Customer pickup locations for negative bookings

lat	lng	s2_id	class
-6.224313	106.816824	3.344470e+18	negatives
-6.135677	106.789150	3.344518e+18	negatives
3.571309	98.668853	3.472610e+18	negatives
-6.251063	106.793267	3.344469e+18	negatives
-6.224505	106.804043	3.344470e+18	negatives

Fig 2: Sample data of the customer pickup locations for negative bookings

```
def latlong_s2id(lat,lon,s2_size):
    pos = s2.LatLng.from_degrees(lat, lon)
    s2cell = s2.CellId.from_lat_lng(pos).parent(s2_size)
    return str(s2cell.id())
```

Fig 3:- Conversion of the latitude and longitude of the centroid of the cluster to the S2 cell ID.

B.Collection of Training Data for the S2 Cell

We collected the training data which consisted of ride bookings requested from the selected S2 cell using flink. This data consisted of the latitude and longitude of the customer pickup location of all the completed ride bookings.

III. APPROACH

A.Identifying the optimal pickup location for customer

When the customer makes a ride booking, the position of the customer pickup location is given as input to the trained model which suggests the nearest optimal pickup location having the high probability of the driver accepting the ride booking. To suggest such an optimal location the model is first trained with the training data collected from the selected S2 cell. The training data consists of the latitude and longitude of the customer pickup location of all the completed ride bookings, clusters of these locations are formed. The nearest cluster in which the incoming position of the customer pickup location input falls into is identified. If this input location is a negative pickup location from where the drivers have rejected the ride bookings, then the nearest positive pickup location is found using the haversine distance formula. This nearest positive pickup location is the optimal pickup location for the given input location.

```
import pandas
import numpy as np

pickup_locations = pandas.read_csv('./train.csv',
                                   dtype=np.float)
print(pickup_locations.head(10))

from sklearn.cluster import KMeans

clustering_algo = KMeans(n_clusters=8)

labels = clustering_algo.fit_predict(pickup_locations)

cluster_sizes = {i: len(np.where(clustering_algo.labels_ == i)[0])
                 for i in range(clustering_algo.n_clusters)}

cluster_centers = clustering_algo.cluster_centers_

labels_unique = np.unique(labels)
n_clusters_ = len(labels_unique)
```

Fig 4:- Training the model.

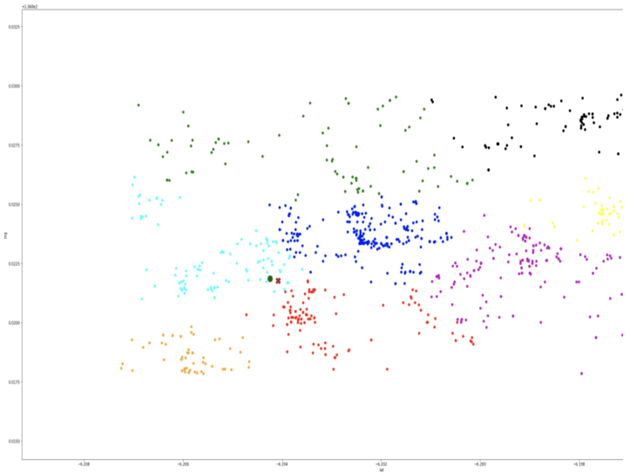

Fig 5:- Clusters formed after training the model.

IV. EMPIRICAL EVALUATION

A.Precision and Recall

We have experimented with Go-Jek, using the data of around 1 million ride bookings made across different S2 areas in Indonesia. We found that, our algorithm was recommending the optimal pickup location for the customer if the customer's ride booking was likely to be rejected or cancelled by the driver, with a precision of 0.79, recall of 0.77, f-measure of 0.78. The recommended optimal pickup locations were within the radius of 100m from the customer's current pickup location.

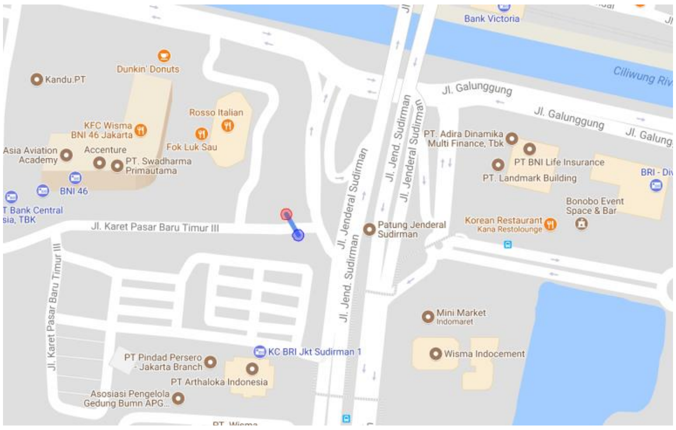

Fig 6:- Recommended optimal pickup location in blue for the customer with location in red.

V. CONCLUSIONS

In this paper, we present an algorithm to recommend the optimal pickup location for the customer booking a ride, based on the past history of completed, cancelled and rejected bookings. This recommendation of the optimal pickup location for the customer can be used to increase the probability of the driver accepting the ride booking, hence providing a good user experience to the customer.

ACKNOWLEDGEMENT

We acknowledge the support of Thiyagarajan Gnanasekaran, cofounder and big data architect at Fratics, Guo Jun, data scientist at Go-Jek and Go-Jek team for our work.

REFERENCES

[1]. Glaeser, Chloe and Fisher, Marshall and Su, Xuanming, Optimal Retail Location: Empirical Methodology and Application to Practice (May 6, 2017). Available at SSRN: https://ssrn.com/abstract=2842064 or http://dx.doi.org/10.2139/ssrn.2842064.

[2]. Xu Junjie, Wu Min, Convenient pickup point in e-commerce logistics: a theoretical framework for motivations and strategies (March 1, 2013). Available at http://www.cmnt.lv/upload-files/ns_39crt003%20CMNT18-0739_LG-ed-VG.pdf

[3]. http://blog.christianperone.com/2015/08/googles-s2-geometry-on-the-sphere-cells-and-hilbert-curve/

[4]. https://www.cc.gatech.edu/~isbell/reading/papers/berkhin02survey.pdf

[5]. Xu, D. & Tian, Y. Ann. Data. Sci. (2015) 2: 165. https://doi.org/10.1007/s40745-015-0040-1