A new site location approach based on Filtering and Clustering for car sharing system

Yang Yu[1, a], Juanjuan Zhang[2,b] and Jinglin Li[3,c]

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications

[a]xxxyangyu@gmail.com, [b]zhangjuanjuan_815@hotmail.com, [c]jlli@bupt.edu.cn

Keywords: Car sharing, Affinity Propagation, Filtering and Clustering.

Abstract. Car sharing service provides people with a new environmentally friendly way to travel. In car sharing system, one of the major challenges is how to determine the location of the car sharing sites. However, the existing method does not take into account people's demand in reality to select the car sharing sites position. In this paper, we focus on selecting the car sharing sites location by using real taxi GPS trajectories and propose a new method based on filtering and clustering for the location problem. Large-scale real-world experiments shows that the method we proposed in selecting car sharing sites has a better performance than other approaches.

Introduction

In order to reduce traffic pressure and effectively utilize idle vehicles, car sharing system, which allows people to rental vehicle for short distance journey, gradually shows its importance. As an alternative public transportation, car sharing system has been proved to have a positive effect on urban mobility[1], the mobile and flexible characteristic of car sharing system in the city's transportation system, particularly in the short distance journey has manifested the superiority which other transportation means do not have. With the development of the electric vehicles, car sharing system will have further significant benefits even have the potential to replace traditional private vehicles. Electric vehicles have been observed as the best candidate for shared vehicle systems because of the zero emission[2].

One of the major challenges encountered by the car sharing system providers is selecting the appropriate locations for car sharing sites. Regarding to the service which a car sharing site provide, it is crucial that the site which can serve more people who can reach in short time. In scientific literature on transportation networks, this problem is known as the depot location. The selection of car sharing site location should take quality of service, efficiency of service and the cost of service into consideration. Therefore, a future study of appropriate location for car sharing would be necessary.

To select the location of car sharing system sites, we should build a travel requirement module in the service area precisely. The requirement for car sharing usually refers to how much of vehicles is desired, hence we use the taxi trajectory data to present this requirement.

In this paper, we propose to locate the car sharing sites based on filtering and clustering. To evaluate this module, we use the real data source of taxi GPS trajectories which generated by over 12,000 taxis in Beijing.

The contribution of this paper includes: (1) We proposed a method to aggregate demand points to facilitate clustering based on Grid-distance. (2) We consider the size of each aggregation point as a parameter for the clustering algorithm. And we proposed that Affinity Propagation clustering algorithm is more proper in selecting location of car sharing sites. (3) Extensive experimental results on real data source of taxi GPS tracks collected from over 12,000 taxi in Beijing demonstrate that our location selection mechanism outperforms other compared mechanisms.

The remainder of this paper is organized as follows. In section II we review related work about car sharing system and Affinity Propagation. In Section III, firstly we formally define the problem and give an overview of our proposed method. Then, we discuss about the detailed model to locate the car

sharing sites. In section IV, we give the experimental result of our method and the results of the experiment. Finally, we conclude our work and discuss the future work.

Related works

To locate the car sharing site, Correia[3] proposed an optimization method to depot location in one-way car sharing system where the vehicle stock imbalance issues are addressed under three trip selection schemes. Martinez[4] optimizes the position of shard biking stations and measuring the bicycle relocation activities required in a regular operation day. Jorge[5] presents a simulation model that considers demand variability and one vehicle relocation policy and use MIP model to test the solution.

Frey and Dueck proposed a novel clustering method, named Affinity Propagation[6]. Affinity propagation is an exemplar-based clustering algorithm that finds a set of data-points that best exemplify the data, and associates each data-point with one exemplar. AP has been successfully applied to broad areas of computer science research because it has much better clustering performance than traditional clustering methods such as k-means. However，AP does not scale for large data sets because it requires quadratic CPU time in the number of data points to compute the results. Givoni[7] derived an inference algorithm that operates by propagating information up and down the hierarchy, and is efficient despite the high-order potentials required for the graphical model formulation.

Methodology definitions

We now define some notations that will be used throughout the paper.

Definition 1 (Vehicle Track): Vehicle Track is composed of sequence of GPS points represented as <latitude, longitude>, which corresponding to an occupied vehicle. For example, $s_1 \rightarrow s_2 \rightarrow \cdots \rightarrow s_N$, where s_k is a GPS point of a site consisting of a geospatial coordinate set and a timestamp. Respectively, s_1 and s_N are the origin and destination of the trajectory.

Definition 2 (Extraction Point): From vehicle tracks network, we can extract the GPS points which represent the pickup points and the drop points. An extracted GPS point S as a pickup or drop point can be represented as: *S= (l.id, l.lat, l.lon).* where, l.id is the identifier of a trip, *l.lat* and *l.lon* represent coordinates of latitude and longitude of a pickup or drop point.

Grid-distance based filtering for extraction points

After the extraction for the point in the vehicle tracks network, the points can represent the demand of vehicle to travel in selecting the sites of car sharing system. When creating maps with hundreds or thousands of points, it can be difficult to further select the locations of car sites. Hence, the information displayed on the map needs to be evenly distributed and clear. In addition, we should make a decision based on user's desire for car sharing sites when they search nearby places.

To solve above mentioned issues, in our method, we use grid-distance based filtering to aggregate the large numbers of extraction points in the map. The basic idea of grid-distance based filtering is to divide the map into a lot of regular grids and obtain the aggregation results for all extraction points fall in the grid. One aggregation result Ag_r_i contains the final aggregation point Ag_i and the extraction points set *{ S_set_i }* in the corresponding grid scope G_i, which can be represented as $Ag\_r_i = (Ag_i, S\_set_i, G_i)$. At first, there is no any known aggregation point in the map. As for each extraction point in the map, we identity the specific grid it belongs to by iterative computation. Then the aggregation point for each grid can be obtained by dividing and merging on extraction points. The detailed filtering process is illustrated in Algorithm 1.

Algorithm 1 Grid-distance based filtering

Input: initial extraction points *{S_1, S_2,..., S_N }*

Output: aggregation result set *Ag_r= {Ag_r₁, Ag_r ₂,..., Ag_r $_M$ }*
Ag_r= { };
Min_dis= MAX;
Disjoint = NULL;
for *i=1* to *N* **do**
    compute the specific grid G_i the point S_i belongs to
    for each G_{curr} *in Ag_r* **do** $a = a + b_2$
        if G_i disjoint with G_{curr} **AND** *Distance(S$_i$, Ag$_i$) < Min_dis* **do**
         Min_dis= Distance(S$_i$, Ag$_i$) ;
        Disjoint= Ag_r $_{curr}$;
       end if
    end for
end for
    if *Disjoint = NULL*
       add *(S$_i$, S$_i$, G$_i$)* to *Ag_r*
    else add *S$_i$* to the *S_set$_i$ Disjoint* point to
return *Ag_r*

Affinity Propagation clustering for selecting the sites

After filter the extraction points, we get *Ag_r* as the result. *Ag_r= {Ag_r₁, Ag_r₂,..., Ag_r$_M$ }*, in which *Ag_r$_i$* contains *Ag$_i$* , *S_set$_i$* and G_i. In car sharing system, *Ag$_i$* represents the location which customers want to take a car and size of *S_set$_i$* represents the demand degree of customers. In reality, limited by capital investment and geographical location, we cannot establish a site at each demand point. In order to reduce the number of sites, and to serve more customers, we need to use clustering algorithm.

To locate the sites of car sharing system using the result above and perform clustering accurately, we use AP (Affinity Propagation) algorithm proposed by Frey and Dueck at 2007 and it performs better than traditional clustering methods such as k-means. AP algorithm does not need to define the number of clusters K in advance and in the process of iteration, it searches for suitable clustering centers automatically, and identifies the position and number of cluster centers from the data points so that all data points are similar to the nearest clustering centers. The algorithm has been applied to many fields, such as facilities location[7], image recognition[8], image segmentation[9], image retrieval and text mining[10], and achieved good result.

The collection of similarities, *s(i, j)*, is the input for AP algorithm, whose similarities reflects how well the aggregation point *Ag$_j$* is suited to be the exemplar for *Ag$_i$*. The purpose of Affinity Propagation is to maximize the sum of similarities between the aggregation point *Ag$_i$* and its cluster center *Ag$_j$*. Each aggregation point *Ag$_i$* has a self-similarity, *s(i, i)*, which called preference P . And it will influence the number of cluster centers. The points which are initialized with a bigger P are more suitable to be the cluster centers. Usually, we choose the median of the pairwise similarities to be the initialized value for each point. In this way, each point has the same possibility to become the cluster center. However, in this paper, each aggregation result *Ag_r$_i$* contains not only *Ag$_i$* but also a set of extraction points *S_set$_i$*. The size of *S_set$_i$* will influence the possibility that *Ag$_i$* becomes the cluster center. Therefore we initialized *s(i, i)* with the negative number of linear normalization of each size of *S_set$_i$*. We defined function *size(S_set$_i$)* to represent the size number of *S_set$_i$*. The formulas for performance P are stated below:

$$\min\_val = \min\{size(S\_set_1), size(S\_set_2), \cdots, size(S\_set_m)\} \tag{1}$$

$$\max\_val = \max\{size(S\_set_1), size(S\_set_2), \cdots, size(S\_set_m)\} \tag{2}$$

$$s(i,i) = -\frac{size(S\_set_i) - min\_val}{max\_val - min\_val} \tag{3}$$

Meanwhile, we initialized $s(i, j)$ with the negative number of the square of Euclidean distance between each Ag using below formula:

$$s(i,j) = -dis^2(Ag_i, Ag_j) = -\|Ag_i - Ag_j\|^2 \tag{4}$$

In Affinity Propagation, there are two crucial parameters of information message. The responsibility, $r(i, j)$, represents how well suited Ag_j is to become a cluster center for Ag_i. The Availability, $a(i, j)$, represents how appropriate it will be for Ag_i to select Ag_j as its cluster center.

Algorithm 2 Affinity Propagation Clustering

Input:

s(i, j) the similarity for Ag_i to Ag_j.

Output: *c(i)*, the index of aggregation point j, center for each point Agi.

Step1: Assignment Availability *a(i, j)* to zero:

$$a(i, j) = 0 \tag{5}$$

Step2: Update the Responsibility *r(i, j)* with the formula:

$$r(i, j) = s(i, j) - \max_{j's.t.j' \neq j}\{a(i, j') + s(i, j')\} \tag{6}$$

Step3: Update the Availability with the formula:

$$\left\{ \begin{array}{l} a(j, j) = \sum\limits_{i's.t.i' \neq j} \max\{0, r(i', j)\} \\ a(i, j) = \min\{0, r(j, j) + \sum\limits_{i's.t.i \neq i',j} \max\{0, r(i', j)\}\} \end{array} \right\} \tag{7}$$

Step4: Another crucial parameter λ, called damping factor, in Affinity Propagation influences the iteration process. a(i, j) and r(i, j) are updated by weighting the current iteration value and previous iteration value for each loop using the formula below:

$$a(i, j)^{new} = \lambda a(i, j)^{old} + (1 - \lambda)a(i, j)^{new} \tag{8}$$

$$r(i, j)^{new} = \lambda r(i, j)^{old} + (1 - \lambda)r(i, j)^{new} \tag{9}$$

The procedure of message-passing would be terminated after a fixed number of iterations between formulas (8) and (9).

Step5: Calculate the index of the aggregation point, the cluster center of point i is:

$$c(i) = idx(\max_{j}\{a(i, j) + r(i, j)\}) \tag{10}$$

Experiments

To show the feasibility, effectiveness and benefits of the method we proposed in this paper, we completed our experiments with the data set of real taxi GPS trajectories. The data set is obtained from Beijing taxi GPS trajectories from July 17 to July 23 July, 2014. In order to quantitative analyze the performance of our method, we introduced two concepts coverage rate(CR) and average distance(*AD*).

CR represents the proportion of users within the server radius of the selected sites. In reality, site density enables consumers to reach a site within 30 minutes through walking. So it is reasonable to set service radius to 3 km.

$$CR = \frac{\sum_{i=1}^{N} servicePoint_i}{N}$$

(11)

where $servicePoint_i$ denotes the point which satisfies this formula $dis(S_i, c(i)) < serverRadius$. And N is the total number of all extraction point.

AD represents the average distance between each point and its car sharing site it belongs, that is the cluster center.

$$AD = \frac{\sum_{i=1}^{N} dis(S_i, c(i))}{N}$$

(12)

We compared the result of our method(FC) with the SAE model, radial basis function neural network(RBF), and the support vector machine method(SVM).
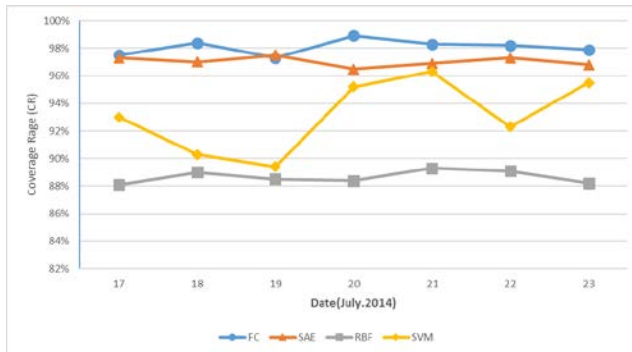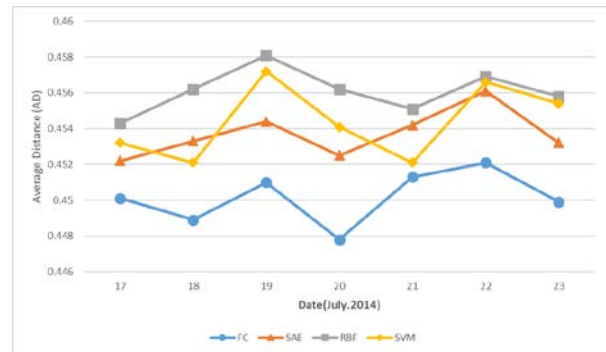

Fig. 1 Comparison of coverage rate   Fig. 2 Comparison of average distance

Fig. 1 shows the output of the comparison among the model above with the data set in seven days. We got a larger coverage rate than the others, which means that in the same condition the car sharing sites selected by our method will get more customers.

Fig. 2 shows the comparison of average distance based on the method above. Obviously, we got a smaller average distance than the other model. The shorter the average distance is, the better its performance is. Customers will spend a shorter time to reach a car sharing site. Through the above experiments, the method based on filtering and clustering in car sharing sites selection has a good performance.

Conclusion

In this paper, we proposed a new location approach based on Filtering and Clustering for car sharing system. We take the real travel demand into consideration. The comprehensive experimental analysis shows the efficiency and feasibility of our method. As a future work, we will analyze the relationship of traffic volume between car sharing sites.

Acknowledgement

References

[1] Litman, T., Evaluating carsharing benefits. Transportation Research Record: Journal of the Transportation Research Board, 2000. 1702(1): p. 31-35.

[2] Bruglieri M, Colorni A, Luè A. The relocation problem for the one-way electric vehicle sharing[J]. Networks, 2014, 64(4): 292-305.

[3] de Almeida Correia G H, Antunes A P. Optimization approach to depot location and trip selection in one-way carsharing systems[J]. Transportation Research Part E: Logistics and Transportation Review, 2012, 48(1): 233-247.

[4] Martinez L M, Caetano L, Eiró T, et al. An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon[J]. Procedia-Social and Behavioral Sciences, 2012, 54: 513-524.

[5] Jorge D, Correia G, Barnhart C. Testing the validity of the MIP approach for locating carsharing stations in one-way systems[J]. Procedia-Social and Behavioral Sciences, 2012, 54: 138-148.

[6] Frey B J, Dueck D. Clustering by passing messages between data points[J]. science, 2007, 315(5814): 972-976.

[7] Givoni I E, Frey B J. A binary variable model for affinity propagation[J]. Neural computation, 2009, 21(6): 1589-1600.

[8] Dueck D, Frey B J. Non-metric affinity propagation for unsupervised image categorization[C]//2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007: 1-8.

[9] Lu Z, Carreira-Perpinan M A. Constrained spectral clustering through affinity propagation[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[10] Guan R, Shi X, Marchese M, et al. Text clustering with seeds affinity propagation[J]. IEEE transactions on knowledge and data engineering, 2011, 23(4): 627-637.