Optimal model of locating charging stations with massive urban trajectories

Danhui Huang*, Yuan Chen and Xiaochen Pan
School of Big Data and Software Engineering, Chongqing University, Chongqing, China

*E-mail: hdh@cqu.edu.cn

Abstract. Popularization of electric vehicles (EVs) is a promising approach towards realizing environmentally friendly transportation. While government has implemented advantageous policy efforts to boost EVs, how to effectively build electric charging stations becomes a key link. A reasonable distribution should not only conform to user habits and potential behaviour patterns, but also satisfying the charging reliability. In this paper, vehicle track data and POI data are used to cluster with road network information. A clustering algorithm based on R serving radius is proposed to improve the results and make them accommodate the rationality of the real world. Finally, suggestions are made for each charging station power supply scale according to regional difference. Experiments show that the proposed model has practical guiding significance.

1. Introduction
In recent years, the aggravation of environmental pollution and the development of battery technology have promoted the development of electric vehicle industry. As the promotion of EVs is expected to reduce unhealthy pollution in the future, electric vehicle has drawn attention of automobile manufacturers and the government [1].

In view of the charging station location problem of EVs, many research institutions and researchers have done a lot of research works. From macroscopic aspect, the most k-location set was proposed, within an interactive mining processes involving users’ input [2]. From microscopic aspect, all sorts of methods had been tried to sitting these k points such as theoretical framework [3] and GIS-based [4]; In terms of distance and waiting time, OCSP and OCPA optimization components were designed to minimize the average time respectively [5]. Different from previous works, we take user’s potential charging behaviour into consideration, analyses their movement rule hide in trajectories to carry out charging station layout.

The rest of this paper is organized as follows. Section 2 gives a general description of the model. Section 3 describes charging station location optimization algorithm and scale calculation rule. Section 4 presents experimental results. Section 5 concludes the paper.

2. Model formulation
From the perspective of discovering people's potential charging habits, and meeting more charging needs, we apply the method of clustering POI and trajectory data to Chongqing taxi trajectory data [6].
Based on fusion data, a charging station location model for site selection, optimization and scale recommendation is proposed.

The main solutions of charging station location problem are as follows. The first step is data processing, the staying points are extracted from the taxi GPS track data, and the POI data is obtained from the social network at the same time. The second step is fusion data clustering, the center of cluster can be more convenient for users to reach. After fusion, site selection recommendation will be more close to business districts, office buildings, government agencies and so on, these locations are more user-friendly also. In the clustering process, density-based DBSCAN algorithm has been used [7].

The third step is to improve the algorithm. Due to the original result may produce a phenomenon of candidate sites gathering in one region, it’s unreasonable for the construction of the charging station. R-DBSCAN could find a compromise point to replace these gathering sites and optimize the model. The fourth step is scale recommendation. A charging station includes a variety of attributes, such as location, floor space, power supply capacity, scale, construction costs and so on. According to the research on regional charging demands, this model will recommend the number of charging piles in various charging stations by using scale conversion rule.

2.1. Data description

The dataset used in this paper is the GPS trajectory of more than 12,000 taxis in Chongqing (2017.03.01-2017.03.31), the frequency is 4 records per minute, the original data set is about 132 GB. This paper selects six attributes, a GPS point is represented as a six-tuple: P<date, time, id, lng, lat, speed>, details showed in table 1.

| Date    | Time | Id       | Lng    | Lat    | Speed |
|---------|------|----------|--------|--------|-------|
| 20170301 | 2001 | 0600307115 | 106.4851 | 29.53619 | 0     |
| 20170301 | 2435 | 0100504146 | 119.0178 | 33.59274 | 13.2  |
| 20170301 | 2249 | 0600305444 | 106.4894 | 29.60768 | 39.8  |

More than 20,000 POI data from Baidu Map was obtained, covering the nine districts of Chongqing City. There are eight categories of POI data: restaurants, hotels, companies, hospitals, parking lots, office buildings, government agencies, shopping centers. Each piece of data contains five attributes: ID, name, longitude, latitude, and POI category.

2.2. Scale definition and grid partition

Charging stations in different locations should have different scales according to charging needs. In prosperous areas, such as business gathering areas, the demand for EVs charging is far greater than that in dispersed residential areas [8]. For this reason, we propose the definition of charging station scale, and incorporates the construction scale of each charging station into the charging station selection planning.

Consider the heat value HValue and the threshold e as two factors that influence the scale S, set the threshold e to 10, the relationship between them is shown with Eq.(1).

$$S = \lfloor \frac{HValue}{e} \rfloor$$

Moreover, if a charger is available near the destination of the EV driver, the user may be willing to park the charging station and walk for a few minutes to the destination. Therefore, instead of using exact geographical locations, each staying point is allocated to a grid. In this paper, nine districts in the main city of Chongqing are divided into grids according to latitude and longitude. The size of grids is about 1.1Km*1.0Km, and the total number of grids is nearly 10,000.

2.3. Staying points identification

Assuming that a taxi has n GPS track sampling points, the taxi GPS track P=p1 p2...pn, is a continuous point sequence, composed of GPS sampling data points in time sequence[9]. There are
many methods for GPS staying point detection, ranging from the method based on missing point [10], direction feature [11], to merge trajectory points below a certain velocity threshold into candidate residence positions [12]. In this paper, the following definition of GPS staying point are selected [13].

A sub-trajectory is represented as:

\[ P_{k,m} = \{ (t_j, loc_j), j \in [k,k + m - 1] \} \]  

The length of sub-trajectory \( P_k \), m is m, m ≤ n and k ∈ (0, n-m). If there is a data point loci satisfied:

\[ \forall i \in [k,k + m], \text{when } t_{i+1} - t_i \leq \Delta t \text{ and } t_{m-1} - t_0 \geq \Delta T \]  

If two points have same latitude and longitude, and the time interval is less than \( \Delta t \), it indicates these two points are continuous points. Moreover, if the stay time at this point is greater than \( \Delta T \), then this position can be considered as a staying point. Because of the current technical limitations of EVs, in general, the charging time is more than one hour, so the stay time is set to 60 minutes.

3. Locating algorithm

DBSCAN is a density-based clustering algorithm which divides high density regions into clusters, and can find clusters of arbitrary shape in noisy spatial data. In this paper, the sample set is too large converge, so the KD tree is used to search for the nearest neighbour [14].

3.1. R-DBSCAN

In this model, the central point of cluster is selected as a candidate location in the final charging station layout. The results of the original DBSCAN algorithm may make the two cluster points too close to each other. To solve this problem and improve the adaptability in reality, we introduced R service radius, which can control the final site selection points won’t appear in the same area.

Each charging station will cover the serving area with a radius R, and if multiple site selection points are in the R scope, they will be merged into a new one.

As shown in figure 1(a), the radius ranges of points p1, p2 and p3 cover each other, while p4 only belong to its R range. In this paper, p1, p2 and p3 are merged into one, and their average coordinates are taken as the new point P1 in figure 1(b), as a result, four site selection points are finally optimized to two site selection points, and the location of P2 remains unchanged. Calculate the distance between two points as:

\[ d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 + y_2)^2} \]  

R-DBSCAN can be summarized as the following four steps:

- Save the original site selection points to set S, randomly select a point p and create another set T1 for p1.
- Calculate the distance d between other points in S and point p. If point d1m < R, save pm to T1 until there is no point in S.
- Remove the point that has been stores in T1 from S, and then repeat steps 1-3
- While S is empty, the average coordinates of the points in each set T1T2…Tn are the final site selection point. The algorithm ends.
3.2. HValue calculation

In the previous preparation work, we divided the nine districts in Chongqing into grids. Each cluster will occupy one or more grids, the sum of the HValue of these grids is also the HValue of the cluster. In figure 1(c), the red dot is the final site selection point generated by the three original points, add up the three HValue and the result is the HValue of red final site. This relationship can be expressed as:

\[
HValue_p = \sum_{j \in P} \sum_{i \in p_j} HValue(grid_i)
\]  

(5)

For all original sites j belong to P, and all grids i belong to pj, calculate the HValue of each original site and add them up, then we can get the HValup of the final site selection point P.

After we get the HValue, the scale and the number of charging piles for each charging station can be calculated to an exact value respectively.

4. Experiments

4.1. Visualization of experiment data R-DBSCAN validation

We extract the road network of Chongqing, which contains 11,311 roads and 1479 vertices, as shown in figure 2(a). figure 2(b) shows taxi trajectory spatial distribution after coordinate transformation. Figure 2(c) shows charging demands using a heat map, the lighter the denser. Figure 2(d) shows POI distribution on the map, by scaling the map, we can see the number and exact location in the region.

4.2. RDBSCAN validation

The original DBSCAN algorithm mainly contains two parameters, Eps and Minpts. These two parameters affect the distribution of charging stations by controlling the distance between two points in the cluster and the number of clustering points.

Figure 3 present the comparison between original DBSCAN and R-DBSCAN, as we can see in the red circle of figure 3(a), figure 3(c), the original DBSCAN algorithm regards the multiple clusters in adjacent areas as multiple site selection points, and suggests to build the same number of charging stations. However, considering the realistic factors such as land use restriction, capital restriction, business planning, the more reasonable way is to build only one charging station in this area. R-DBSCAN integrates the charging stations in the R radius region by setting R constraints, and finally gives a recommendation point that can cover all the charging requirements in the original region, as shown in figure 3 (b), figure 3 (d).
The heat value of the final recommendation point is the sum heat value of all the original cluster grids, that is to say, the scale of the final recommendation point will be larger and there will be more charging piles.

Figure 4 shows an example on real map in Jiefangbei Business Circle, Chongqing. The original DBSCAN system recommends 7 charging stations in this area as shown in figure 4 (a). figure 4 (b) projected these 7 points onto the real map, 5 recommended points outside the red circle cover several commercial locations such as Jie Fangbei Pedestrian Street, Dayuan Square and Hongyadong with reasonable distance. However, as shown in figure 4(c), the other two recommended points in the red circle, one is near Marriott hotel and the other is near internet cafe, it’s not necessary to build two charging stations on the same street. The R-DBSCAN algorithm re-integrates the two clusters and assigns them to a new recommendation point, which can cover the original two clusters. R-DBSCAN algorithm improves the adaptability of the reality, not only ensures convenience and power supply demand, but also contributes to avoid unnecessary waste of social resources.

4.3. HValue matrix and scale recommendation

The grid area in figure 5 (a) is part of the nine districts in the main city of Chongqing. The grid size is 16*16, HValue matrix calculates parking demand in each area, which is the numerical basis for making scale recommendations for charging station.

As shown in figure 5 (b), each charging station recommendation point has scale information, including the number of recommended charging piles, the heat value of the grid occupied by the charging station, the threshold and the number of charging piles per scale. The charging station recommendation point in Chongqing North Railway Station area has a high heat value , this is Chongqing's largest passenger station for high-speed train and bus traffic with a large number of parking bus and private car. Compared with other areas with business circle or park, its charging demand is larger than others, which is more in line with the reality.
Based on the user's behavioural habits, this paper analyses the trajectory data of the vehicle, and combines various facilities layout and road network distribution in the city to improve the clustering algorithm. The coverage radius \( R \) is used to control the number of charging stations in the area, and the final charging station location is optimized and adjusted. According to the difference in charging demand of each area, the proposal for the construction scale of each charging station is given. While guaranteeing convenience and satisfying charging demand, resources are reasonably allocated. Combined with the real map, we validate the recommendation site of nine districts charging station in the main city. The results can cover the crowded area of traffic vehicles in various streets, shopping districts, entertainment areas, etc. The distance between charging stations is much more reasonable, and the scale recommendation is also positively correlated with the density of local traffic flow, which fully illustrates the effectiveness of this model.

References
[1] Lu C, Rong K, You J, et al. Business ecosystem and stakeholders’ role transformation: Evidence from Chinese emerging electric vehicle industry[J]. Expert Systems with applications, 2014, 41(10): 4579-95.
[2] Li Y, Bao J, Li Y, et al. Mining the Most Influential k-Location Set from Massive Trajectories[J]. IEEE Transactions on Big Data, 2018, 4(4): 556-70.
[3] Xiong Y, Gan J, An B, et al. Optimal electric vehicle fast charging station placement based on game theoretical framework[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(8): 2493-2504.
[4] Erbaş M, Kabak M, Özceylan E, et al. Optimal siting of electric vehicle charging stations: A GIS-based fuzzy Multi-Criteria Decision Analysis[J], Energy, 2018, 163: 1017-31.
[5] Li Y, Luo J, Chow C Y, et al. Growing the charging station network for electric vehicles with trajectory data analytics[C]/2015 IEEE 31st International Conference on Data Engineering. IEEE, 2015: 1376-87.
[6] Morris B T, Trivedi M M. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(11): 2287-2301.
[7] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]/Kdd. 1996, 96(34): 226-31.
[8] Lingfeng K, Zifa L, Huan Z. Modeling algorithm of charging station planning for regional electric vehicle[J]. Modern Electric Power, 2010,
[9] Yan Z, Chakraborty D, Parent C, et al. SeMiTri: a framework for semantic annotation of heterogeneous trajectories[C], Proceedings of the 14th international conference on extending database technology. ACM, 2011: 259-70.
[10] Deng K, Xie K, Zheng K, et al. Trajectory indexing and retrieval[M]/Computing with spatial
trajectories. Springer, New York, NY, 2011: 35-60.

[11] Du J, Aultman-Hall L. Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues[J]. Transportation Research Part A: Policy and Practice, 2007, 41(3): 220-32.

[12] Wolf J, Guensler R, Bachman W. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data[J]. Transportation Research Record: Journal of the Transportation Research Board, 2001 (1768): 125-34.

[13] Kong Y, Wu J, Xu M, et al. Charging pile siting recommendations via the fusion of points of interest and vehicle trajectories[J]. China Communications, 2017, 14(11): 29-38.

[14] Bentley J L. Multidimensional binary search trees used for associative searching[J]. Communications of the ACM, 1975, 18(9): 509-17.