Planning Electric Vehicle Charging Stations Based on User Charging Behavior

Jinyang Li*, Xiaoshan Sun*, Qi Liu*, Wei Zheng†, Hengchang Liu* and John Stankovic‡

*School of Computer Science and Technology
University of Science and Technology of China, Anhui, China
Email: {ljyustc, sxs1166, liuqi100}@mail.ustc.edu.cn, hcliu@ustc.edu.cn
†Comprehend (Suzhou) Information Technology Inc.
Suzhou, Jiangsu
Email: 18.weizheng@gmail.com
‡Department of Computer Science
University of Virginia, Charlottesville, VA 22903, USA
Email: stankovic@virginia.edu

Abstract—Electric vehicles (EVs) as a green alternative of fossil-fuel vehicles (FFVs) have been promoted by many governments all over the world. As a result, constructing an efficient charging pile network has become a crucial task for governments and manufacturers to increase EV adoption, as well-planned charging sites can serve more EV users at a lower cost and improve user satisfaction. Unfortunately, most of existing planning approaches for EV charging stations estimate charging demand and optimize locations based on traffic patterns of FFVs, e.g., traffic flow and parking locations, and the patterns of charging behavior are overlooked causing an inefficient network layout for existing EV users.

In this paper, we propose and implement a novel algorithm to estimate charging demand and to plan new charging stations. The observations and analysis of the usage data of the charging mobile app developed by the official EV public service platform of Beijing and pile usage data of the charging pile network (CPN) of Beijing are presented. Users' charging-related search behavior and navigation behavior and the pile usage pattern are analyzed and modeled. A Bayesian-inference-based algorithm is proposed to fuse the three models to estimate charging demand. A flexible objective function is introduced to tune the benefit between serving the existing EV users well and attracting more FFV drivers. Finally, a reference system is developed using Beijing as a target city, and providing extensive experiments to demonstrate the performance of our system.

I. INTRODUCTION

Many governments are pushing hard to replace fossil-fuel vehicles (FFVs) with cleaner electric vehicles (EVs) since the wide use of EVs has the potential to reduce greenhouse gases emissions significantly. In Norway, alternative fuel vehicles accounted for 29% of all new car sales last year. As the largest EV market, China has ambitious medium-term goals for automotive efficiency and climate change. By the end of September 2017, according to the official statistics, there are 141,094 EVs running in Beijing (including private and commercial vehicles, e.g., taxis and rental cars) and 14,612 public charging piles have been deployed in the city [1]. With ever more EVs on the roads worldwide, how to build efficient charging networks to support them is an urgent issue.
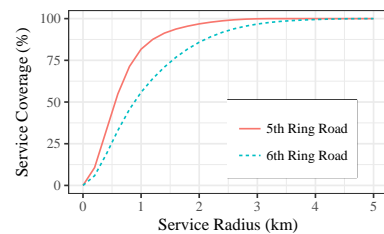

Figure 1.   Beijing CPN Coverage

To plan new charging stations, we first need to understand how charging demand is distributed. In the early literature, such knowledge is obtained through questionnaire-based surveys. Some studies use population data to estimate charging demand which is assumed to be proportional to the mobility between population centers. With the popularity of geomagnetic vehicle detectors and on-board GPS devices, follow-up work is mostly based on traffic data. A common approach is to estimate the charging demand based on origin-destination (OD) data. GPS track data allows researchers to explore more sophisticated driving behavior patterns, such as when and where the driver stops the car. As the EV charging time is significantly longer than the refueling time of the FFV, users usually tend to charge when parking. Therefore, many studies estimate charging demand based on the parking demand. As EV's adoption increases, there are studies using real-world EV operational data to explore users' charging behavior patterns. However, most of them are relatively small scale or target only electric taxis.

The planning problem is usually formulated as an optimization problem that tries to minimize the associate cost, e.g., the travel distance to the stations, or maximize the demand served. Most previous studies focus on improving the coverage of the charging pile network (CPN) which is indeed an important metric in the early construction of the charging infrastructure. However, these method may not provide good support to the decision makers when the

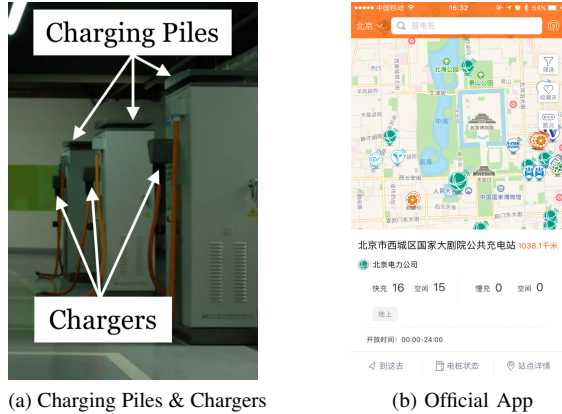(a) Charging Piles & Chargers   (b) Official App

Figure 2.   echarge Service Illustration

coverage of CPN achieves a certain level while the adoption rate of EV is still low. For example, by far the market share of EVs in Beijing is about 2.4%, but the urban area of the city is basically covered by the CPN. Figure 1 depicts the cumulative distribution function of CPN coverage at different service radii. It can be seen that the charging service has covered 95% of the area within the Fifth Ring Road with the service radius set to be 2km. So does the area within the Sixth Ring Road with service radius set to be 3km. In such a case, it is very difficult to use traffic data to capture the mobility of EVs because they only account for a small fraction of the traffic. In a sense, the charging demand estimated based on FFV traffic data reflects the potential needs for future EV drivers. How to capture the needs of existing EV drivers is an interesting and challenging issue.

To aide EV drivers, an official public EV services platform, *echarge*, is built with the support of the Beijing government [1]. A mobile phone app is developed by the platform to provide search, navigation and payment services for EV drivers. The platform also took the lead in drafting a national standard to unify the charging and information exchange protocol [2]. These two services have offered new research opportunities to exploit pile usage data and user's app activity for improving the understanding of EV driver's charging behavior. In particular, the pile usage data contains charging records on all public charging piles and user's search and navigation actions are recorded in the app activity dataset. Figure 2 shows the standard-charging piles and the mobile app. Such charging infrastructures are capable of capturing the needs of EV drivers accurately on an urban scale. However, until now our knowledge of the correlation and divergence in these data, and how to integrate them to conduct charging stations planning are still limited.

In this paper, we design and implement a novel demand sensing and charging station planning algorithm, which leverages the massive pile sensor and mobile application data to estimate charging demand and optimize siting and sizing for new charging stations. The proposed system contains three major components: 1) *Demand Estimation Server*, which filters and encodes the data from charging pile sensors and mobile applications, then uses a Bayesian-based method to fuse the models to obtain the charging demand estimation. 2) *Planning Server*, which takes decision maker's input, e.g., budget constraints and preference parameters, then uses demand distributions estimated by the *Demand Estimation Server* and external sources to provide charging station planning suggestions. 3) *GUI Frontend*, which displays the deployment suggestions and allows user to adjust parameters. The main contributions of this paper are:

- To the best of our knowledge, we conduct the first work to design and implement a charging demand estimation and charging station planning algorithm based on pile usage and user's app activity data. The search, navigation and charging behavior are analyzed and modeled. A Bayesian-based algorithm utilizes the multi-source data to provide reliable estimation to reflect the real needs of existing EV drivers.
- A preference parameter is introduced in the objective function to reflect the tradeoff between improving the convenience of current EV drivers and attracting more FFV drivers. We use a flexible objective function to unify different demand estimation results into the same optimization framework.
- The performance of our system is evaluated based on real-world pile usage and app activity datasets. The evaluation results show that our multi-model fusion estimation algorithm outperforms single model methods. The mean average percentage error of our method is 45.36% and 55.13% lower than the two baselines respectively. Our planning result can achieve better user satisfaction under the same budget and outperforms the stat-of-the-art methods by up to 16.57% and 36.49%, respectively.

The rest of the paper is organized as follows. Section II and Section III presents the related work and an overview. Section IV describes the charging demand estimation algorithm. Section V presents the planning method. Section VI evaluates our system followed by the conclusion in Section VII.

II. RELATED WORK

The charging pile planning problem has received significant attention with the growth of the EV market. The planning process usually involves estimating charging demand and determining the optimal location and station size. Various demand estimation and optimization methods have been proposed. The demand estimation methods that appear in previous literature can generally be divided into two categories: point-based and flow-based.

Point-based methods use point data, e.g., parking data, to estimate charging demand. [3] utilized parking information of personal trips and used a regression model containing

various variables, including population density and trip attributes, to estimate parking demand. Another parking-based method can be fount in [4]. [5] considered daytime activities or trips, including work, shopping, university and tourism, which is useful to serve daytime charging demand. These methods assume that charging demand is proportional to the parking demand. Another common approach is to use Origin-Destination (OD) trips. As the EV charging time is usually up to a few hours, EV drivers tend to charge at the beginning or the end of the trip. [6]–[8] treated the origins or destinations of trips as the potential locations of charging demand.

Flow-based methods use traffic flow to estimate charging demand. The idea is that when the share of EVs is high enough, there may be more charging demand where the traffic is heavy. [9] used OD trips to calculate traffic flow which is then used to estimated the charging demand. [10] assumed that the flow pattern of EVs follows that of FFV traffic and used a homogeneous EV adoption rate to estimate the charging demand. Whereas [11] considered heterogeneous EV adoption rate together with other factors such as driver income, vehicle ownership and other residential factors. [12] combined the parking-based approach with the flow-refueling location model. The authors used average stay duration of drivers to decide the number and location of slow-charging stations.

The siting and sizing problem is usually formulated as a optimization problem that minimizes the construction or charging cost or maximizes social welfare. [13] analyzed a large sample of GPS traces of privately-owned conventional fuel vehicles in Rome and proposed a siting method using the clustered trip destinations. The urban area of Rome was divided into subareas based on clusters of trip destinations. The centroids of the subareas were selected as the siting position of charging piles. [14] extracted stop events from the trajectory data of 11,880 taxis in Beijing for a month and evaluated the public charging opportunities which are assumed to exist in locations where many taxis choose to stop for a long duration. The existing gas stations were scored based on how well they are aligned with identified charging opportunities. Then a non-overlapping set of gas stations were selected to deploy charging piles based on different criteria, e.g., the maximum number of parking events, maximum daily parking time or average parking time per vehicle. [15] considered the parking locations of drivers in a day. The successive parking locations of drivers were used to associate different sites. The siting problem under a budget constraints is formulated as a mixed integer programming (MIP) problem. [16] used personal-trip data to extract parking information and a proportional relationship of charging and parking demand is assumed. The zone parking demand and trip-level parking duration were predicted by ordinary least squares regression models. The best sites for public charging stations were anticipated by solving an MIP problem to minimize the total access cost under a budget constraint. [17] used a genetic-algorithm-based method to find sub-optimized locations to deploy charging stations. The objective is to reduce the range anxiety of drivers which is measured by the number of trips and miles affected by a low battery. The GPS data of FFVs are used in the work and a similar driving pattern between gasoline-powered vehicle drivers and EVs drivers is assumed.

Most of these works either address the planning problem in the early stage of CPN construction or assume that the general traffic data can provide a good description of EV mobility. However, these methods may not provide good estimation when the charging service is basically covered the city but the general traffic cannot reflect the mobility of EV very well because the adoption rate of EV is still low. To the best of our knowledge, none of the previous works use charging pile usage data and mobile app usage data to estimate charging demand and provide planning suggestions.

III. PROBLEM STATEMENT AND SOLUTION OVERVIEW

In this section, we give an overview of the charging station siting and sizing problem and outline our system architecture.

A. Charging Demand Estimation

Recently a new charging and information exchange standard [2] has been established and adopted in the charging infrastructure in Beijing. The real-time status of charging piles are collected by on-pile sensors and uploaded to operators' servers. A mobile phone app, which provides search, navigation and on-line payment service, is developed by the city's public platform for EV services to promote EV usage. These two developments provides a new opportunity to study the charging behavior from both driver's and infrastructure's aspects. However, how to use these new data sources to estimate charging demand remains unexplored.

In this work, we study how to use charging events and app data to estimate charging demand distribution. Given a charging event observed on a charging pile, the possible locations where the corresponding charging demand is generated are estimated. Compared to pile usage data, app data are closer to charging demands. However, they may be still different from the real demand distribution. Our analysis shows that search and navigation behavior have different spatial distribution characteristics, and thus need to be treated differently. The charging demand estimation is discussed in detail in Section IV.

B. Charging Station Siting and Sizing

The goal of charging station siting and sizing is to maximize overall benefit which should (1) support existing EV drivers, and (2) cover more hotspot regions to attract more FFV drivers. Unfortunately, these two objectives may conflict since EV and FFV drivers have different driving
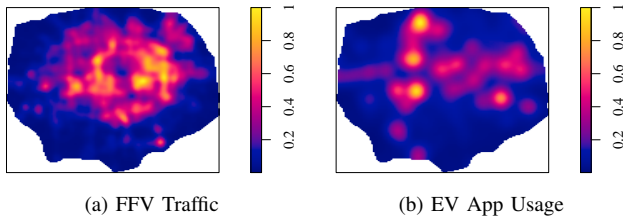
(a) FFV Traffic   (b) EV App Usage

Figure 3. Traffic and App Usage Density

Figure 4. System Architecture

behaviors. To provide such evidence, in Figure 3a, we give the traffic density within the Fifth Ring Road of Beijing based on FFV trajectories, and in Figure 3b, the app usage density in the same region.

The hotspots in Figure 3a are high traffic areas that are regarded as good places to deploy charging stations by previous work. The reason is that when the majority of vehicles are electrified these areas may generate many charging demands. The hotspots in Figure 3b are locations where many EV drivers use the app to search for or start navigation to charging piles, which reflects the charging demand to a certain extent. It can be seen that the hotspots in the two figures are overlapping, but not exactly matched. Interestingly, we note that the hottest areas of app usage are in the west, whereas the hottest areas of traffic are in the east. This pattern generally matches the characteristics of each district. Dongcheng (center east) and Haidian Districts (east) are more developed areas where shopping malls and CBDs are located. Many universities and IT companies exist in the Haidian District (northwest). Districts in south and southwest are relatively undeveloped areas that have less traffic.

To reflect the trade-off between the two objectives, we propose a flexible scoring function for decision makers to adjust their preference. The formulation and discussion are presented in Section V.

C. System Architecture

Our system consists of three components: 1) A *demand estimation server*, 2) a *planning server* and 3) an *user interface*. The system architecture is shown in Figure 4.

Demand Estimation Server. A data feeding mechanism is established to collect real-time app callbacks from the official platform and pile usage data from 36 charging service operators. The app callbacks include information of search and navigation events. When a new callback is received, the target address is converted to longitude and latitude by the geocoder. The origin and destination of the event are stored in the *App Activity Database*. The charging pile data follows the GBT32960.3 standard [2]. The received pile usage data is first filtered to remove unwanted fields and error data due to device failure. The cleaned occupancy status and charging event data are then
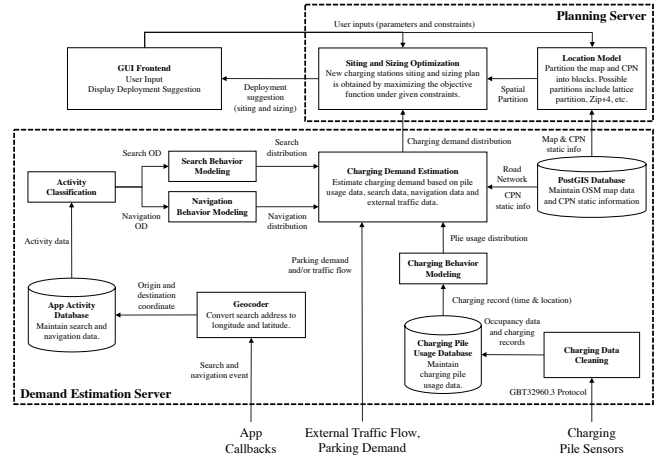
stored in the *Charging Pile Usage Database*. The *PostGIS Database* contains theOpenStreetMap [18] map data and static information of the charging pile network. The three databases together with external traffic data provide inputs to the *Charging Demand Estimation* algorithm.

Planning Server. The map data and CPN static information are used to construct a *Location Model*. In this paper, we use a lattice partition to divide the region of interest into uniform tiles. Other partition methods such as Zip+4 should work as well. Then the estimated demand distribution and spatial partition are fed to the *Siting and Sizing Optimization* algorithm. The optimization algorithm takes in user inputs, including a preference parameter and construction constraints, then optimizes siting and sizing by maximizing the objective function.

GUI Frontend. The deployment suggestion is presented to the users via a GUI interface. The decision makers can try out different parameters to find a satisfied deployment scheme interactively.

IV. DEMAND ESTIMATION SERVER

In this section, we first present our observations and analysis of the pile usage and app activity datasets, then introduce our demand estimation algorithm.

A. Pile Usage and App Activity Datasets

We have established a real-time data feeding mechanism to collect pile usage data from 36 different charging service operators and app activity from the official public platform for EV services of Beijing. The pile usage dataset contains detailed pile status including occupancy, current and voltage. In this work, we focus on the occupancy status and detect charging events when the occupancy status is changed from 0 (idle) to 1 (in use). The app activity dataset contains callbacks of activities of the app users. When an EV driver uses the app to search for charging piles, the search keywords

Date time	The start date and time of the charging event.
Pile ID	The unique ID of the pile.
Site ID	The unique ID of the charging station.

User ID	The unique ID of the app user
Date time	The date and time of the activity.
Action	The activity type. {search, navigation}
Origin	The longitude and latitude of the current location.
Destination	The longitude and latitude of the target.

Figure 5.   OD Distance of App Activity Data

(address) and the current location of the driver are recorded. The app also provides navigation service to charging piles. When the navigation starts, the current location (start point) and the destination (a known charging station) are recorded. Note that the navigation action does not necessarily happen after a search. An app user can click the icon of a charging station on the map to set the navigation target. Note that only the start and end points of the navigation are recorded. The specification of the two datasets are shown in Table I and Table II.

B. Search and Navigation Behavior Analysis

Compared to traffic data, the app activities are closely related to charging behavior and thus can serve as a better proxy of charging demand. However, an action involves two coordinates, the origin and the destination. Which one should we use?

A straightforward approach is to simply use the origin. However, this method discards at least half of the information and thus may produce a large estimate error. Figure 5 shows the origin-destination (OD) distance distribution of the app activity dataset. The OD distance distribution of navigation behavior (red bars) is concentrated at 3.89km and 75% of the search destinations are within 7.70km from the origins. While the OD distance distribution of search behavior is more dispersed with 14.82km as median and 32.23km as upper quartile. Some search destinations are as far as 105km away from the origins.

We further consider the meaning of the two behaviors as shown in Figure 6. Navigation behavior has a strong directivity because the destination is a specific charging station (Figure 6a). If it is not for charging, the driver is unlikely to use a charging app instead of a dedicated software to plan the route. The meaning of search behavior is more complicated. For example, as shown in Figure 6b, a driver in Chaoyang District searches with keywords "Daxing District Wild Animal Park" in the app. The straight-line distance between the target location and the driver's current location is about 51.2km. This behavior is more likely to mean that the driver wants to charge at the destination
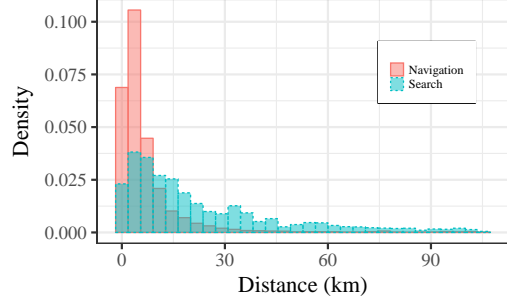
rather than the current location. Another example is shown in Figure 6c. A location is searched which is just 0.7km away from the driver. In this case, the driver may just want to find a charging pile near the current location.

Based on the observations and analysis, we present the search, navigation and pile usage modeling in the following subsections.

C. Search Behavior Modeling

Let $y_i^s = [y_i^{so}, y_i^{sd}]^T$ denote the ith observed search event, y_i^{so} denote the origin coordinate and y_i^{sd} denote the destination coordinate. The coordinate of the corresponding charging demand is represented by θ_i, with likelihood $p(y_i^s|\theta_i, \gamma_i)$, $\gamma_i \in \{0, 1\}$. We define a two-dimensional distribution:

$$\mathcal{F}(y_i|\theta_i, \gamma_i) = \mathcal{U}^{1-\gamma_i}(y_i)\mathcal{G}^{\gamma_i}(y_i|\theta_i, \Sigma) \qquad (1)$$

where $\mathcal{U}(\cdot)$ is the uniform distribution in the given area W and $\mathcal{G}(\cdot|\theta_i, \sigma)$ is the two-dimensional Gaussian distribution centered at θ_i with covariance matrix Σ. Then the likelihood function is assumed to have the following form

$$p(y_i^s|\theta_i, \gamma_i) = \mathcal{F}(y_i^{so}|\theta_i, \gamma_i)\mathcal{F}(y_i^{sd}|\theta_i, 1 - \gamma_i) \qquad (2)$$

The parameter γ_i indicates the type of search behavior. If $\gamma_i = 1$, the the origin of the search follows a Gaussian distribution centered at the θ_i, i.e., $\gamma_i = 1$ means the origin of search action is close to the charging demand. Whereas $\gamma_i = 0$ means the destination of the search is close to the charging demand.

Search events are assumed independent from each other. Given a series of N^s observed search events, the joint distribution is

$$p(y^s|\theta, \gamma) = \prod_i^{N^s} p(y_i^s|\theta_i, \gamma_i) \qquad (3)$$

D. Navigation Behavior Modeling

We consider a CPN with M charging stations. The location of the station $j \in \{1, \cdots, M\}$ is s_j and m_j is the station size in terms of the number of piles. Let $y_{i'}^{no}$ represent the origin, $y_{i'}^{nd}$ denote the destination of i'th navigation and

(a) Navigation (b) Search for Distant Target (c) Search for Nearby Target

Figure 6.   Origin and Destination Analysis

the observation $y_{i'}^n = [y_{i'}^{no}, y_{i'}^{nd}]^T$. The likelihood function has the form as follows:

$$p(y_{i'}^n|\theta_{i'}) = \mathcal{G}(y_{i'}^{no}|\theta_{i'}, \Sigma')\mathcal{H}(y_{i'}^{nd}|\theta_{i'}) \qquad (4)$$

and $\mathcal{H}(y_{i'}^{nd}|\theta_{i'})$ is defined as

$$\mathcal{H}(y_{i'}^{nd}|\theta_{i'}) = \frac{m_{i'}\exp\left(-\|y_{i'}^{nd} - \theta_{i'}\|\right)}{\sum_{j=1}^{M} m_j \exp\left(-\|s_j - \theta_{i'}\|\right)} \qquad (5)$$

where $y_{i'}^{nd} \in \{s_j|j = 1, \cdots, M\}$ because the destination of a navigation must be a known charging station. If $y_{i'}^{nd} = s_j$, then $m_{i'} = m_j$ is the corresponding station size.

Different from search behavior, navigation behavior is more likely to result in charging events. Since it is unlikely for a driver to use the app, which is designed to provide EV service related information, to navigate to other places, the origin of the navigation is more likely close to the charging demand. Given a charging demand θ_i, we assume the probability of a charging station being selected is positively correlated with the size of the site and negatively correlated with the distance to θ_i as shown in Equation (5).

Navigation events are regarded as independent samples as well. Given a series of N^n observed navigation events, the joint distribution is

$$p(y^n|\theta) = \prod_{i'}^{N^n} p(y_{i'}^n|\theta_{i'}) \qquad (6)$$

E. Pile Usage Modeling

Let $y_{i''}^p$ denote the location of the i''th observed charging event, since it must happen in a known charging station, $y_{i''}^p \in \{s_j|j = 1, \cdots, M\}$ where s_j is the coordinate of the charging station j. Given a charging demand $\theta_{i''}$, the likelihood function is

$$p(y_{i''}^p|\theta_{i''}) = \mathcal{H}(y_{i''}^p|\theta_{i''}) \qquad (7)$$

Both the charging event and destination component in Equation (4) are modeled by Equation (5). This is because they

are essentially the same process. Pile usage data can be seen as incomplete navigation data with origins missing. However, in our datasets, pile usage data has better coverage because drivers may find piles without using the app. To avoid repeated counting, records that have matched navigation events are removed from the pile usage dataset and search events are discarded if the corresponding navigation events occur immediately after the search, so that the three kinds of data are mutual exclusive. The charging events are seen as independent samples, given N^p observed charging events, thus the joint distribution is:

$$p(y^p|\theta) = \prod_{i''}^{N^p} p(y_{i''}^n|\theta_{i''}) \qquad (8)$$

F. Model Fusion and the Demand Estimation Algorithm

Considering the three datasets together there are $N = N^s + N^n + N^p$ observations. Let y_i denote the location of the ith observation, $i = 1, \cdots, N$ and θ_i denote the corresponding charging demand coordinates, with likelihood $p(y_i|\theta_i, \gamma_i)$. Each charging demand θ_i is regarded as an independent sample from a population distribution governed by a hyper-parameter vector ϕ; thus

$$p(\theta, \gamma|\phi) = \prod_{i=1}^{N} p(\theta_i, \gamma_i|\phi) \qquad (9)$$

Note that ϕ is not known and thus has its own prior distribution, $p(\phi)$. The appropriate Bayesian posterior distribution is of the vector (θ, γ, ϕ). The joint prior distribution is

$$p(\theta, \gamma, \phi) = p(\theta, \gamma|\phi)p(\phi) \qquad (10)$$

and the joint posterior distribution is

$$\begin{aligned} p(\theta, \gamma, \phi|y) &\propto p(y|\theta, \gamma, \phi)p(\theta, \gamma, \phi) & (11)\\ &= p(y|\theta, \gamma)p(\theta, \gamma, \phi) & (12) \end{aligned}$$

with the simplification from (11) to (12) holding because the data distribution, $p(y|\theta, \gamma, \phi)$, depends only on θ and γ. The

hyperparameters ϕ affects y only through (θ, γ). Then by (10) and (12), the joint posterior distribution can be written as:

$$p(\theta, \gamma, \phi|y) \propto p(y|\theta, \gamma)p(\theta, \gamma|\phi)p(\phi) \qquad (13)$$

Since y^n and y^p are independent from γ, and because the three kinds of data are mutually exclusive, their joint distribution can be given

$$p(y|\theta, \gamma) = p(y^s|\theta, \gamma)p(y^n|\theta, \gamma)p(y^p|\theta, \gamma) \qquad (14)$$

The charging demand $\theta_1, \cdots, \theta_N$ are assumed to be independent samples from a multivariate Gaussian Mixture distribution:

$$p(\theta, \gamma|\phi) = \prod_{i=1}^{N} \kappa^{\gamma_i} \kappa^{1-\gamma_i} \sum_{k=1}^{K} \frac{w_k e^{[-\frac{1}{2}(\theta_i - \mu_k)^T \Sigma_k^{-1}(\theta_i - \mu_k)]}}{2\pi|\Sigma_k|^{1/2}}$$
$$(15)$$

where K is the number of components of the mixture distribution, Σ_k is the covariant matrix of the kth component $\phi_k = [\mu_k, \Sigma_k]^T$, and $\phi = [\kappa, \phi_1, \cdots, \phi_K]^T$. K is determined based on Bayesian Information Criterion (BIC). The hyperprior distribution is assumed be an uniform distribution.

Given observations of search, navigation and charging events, by the joint posterior distribution (13), the joint likelihood (14) and the population distribution (15), the hyperparameter ϕ can be estimated via Markov Chain Monte Carlo sampling. Once ϕ is obtained, the charging demand distribution is given by (15).

G. Summary

In this section, we presented the observations from the pile usage and mobile application datasets, analyzed the divergence and integration of the two kinds of data and designed a Bayesian-inference-based algorithm to estimate charging demand distribution. Algorithm 1 presents the pseudo-code of our algorithm.

V. Planning Server

In this section we first describe the location model, then formulate the optimization problem. After that a flexible score function is introduced to reflect the preference of decision makers.

A. Location Model

The charging station planning problem is usually formulated as an optimization problem that tries to minimize the associated cost, e.g., travel time or distance to charging stations, or maximize the benefit score, e.g., demand served or trips covered. In general, most studies are either point-based models or flow-based models. Point-based models consider spatial points to locate charging demand. Since it usually takes more than an hour to charge an EV, the drivers prefer to charge at the beginning or the end of their trips. Therefore, origins and destinations are considered by many

---

Algorithm 1: Charging Demand Estimation

input : Search data y^s, navigation data y^n, pile usage data y^p, MCMC sample number N_{MC}
output: Hyperparameter ϕ, Charging Demand Samples S_{CD}

1   $y \leftarrow y^s \cup y^n \cup y^p$;
2   Initialize L, $\hat{L}$, θ, γ, ϕ, $\hat{\phi}$;
3   **for** $i = 1$ *to* N_{MC} **do**
4      $\phi \leftarrow$ Sample from $p(\phi)$;
5      $\theta, \gamma \leftarrow$ Sample from $p(\theta, \gamma|\phi)$ given in Eq. (15);
6      $L \leftarrow$ Use y, θ, γ and ϕ to calculate likelihood based on Eq. (13);
7   **if** $L > \hat{L}$ **then**
8          $\hat{L} \leftarrow L$;
9          $\hat{\phi} \leftarrow \phi$;
10   **end**
11 **end**
12 $\theta, \gamma \leftarrow$ Sample from $p(\theta, \gamma|\hat{\phi})$;
13 $S_{CD} \leftarrow \theta$;
14 **return** $\hat{\phi}$ *and* S_{CD};

(a) Beijing Capital International Airport
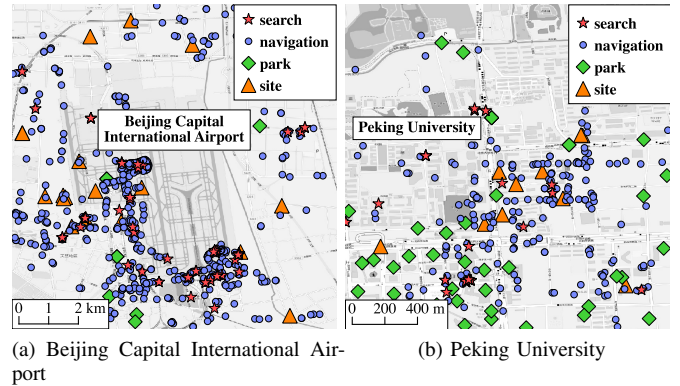
(b) Peking University

Figure 7. Granularity of Data

---

works. Parking zones are good locations to deploy charging infrastructure. With the increasing of EV adoption, parking demand can be easily turned into charging demand. Flow-based models consider traffic flow and select locations that can intercept the most flows as deployment locations. In this work, we base our system on the pile usage and app activity data that both are point observations. Therefore, our system falls into the point-based category.

The spatial partition of the region of interest is dependent on the spatial granularity of the collected data, as shown in Figure 7. In the two examples, it can be seen that app activities are concentrated at certain locations, such as locations around charging stations and the airport drop-off platform. Figure 8 shows the Stienen diagram of the charging stations within the Fifth Ring Road of Beijing.
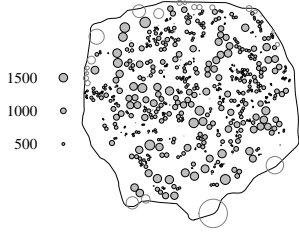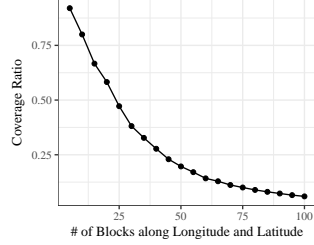
Figure 8. Stienen Diagram of Beijing CPN

Figure 9. Coverage Ratio

Each charging station is represented by a circle of diameter equal to its nearest neighbor distance. A gray solid circle means the nearest neighbor of the site is within the region of interest. Otherwise, the charging station is represented by a hollow circle. We find that existing charging stations tend to form small clusters and the spacing between the clusters ranges from a few hundred of meters to a few kilometers. Like many previous works, we divide the region of interest into uniform tiles. The coverage ratio defined as the number of tiles containing at least one charging station to the total number of tiles is shown in Figure 9. The horizontal axis is the number of tiles along the longitude and latitude. It can be seen that when the number of tiles along longitude and latitude is larger than 25, i.e., the spatial granularity is smaller than a tile of size 1522 meters by 1534 meters, more than half of the tiles are not equipped with a charging station.

B. Siting and Sizing Optimization

Given a region of interest that is partitioned into tiles, we formulate the charging station planning problem as a mixed integer programming (MIP) problem. Let $u \in U$ denote a tile where charging demands are generated and $v \in V$ denote a tile where demands are satisfied (i.e., charging stations). The objective is defined as follows:

$$\sum_{u \in U} \sum_{v \in V} f_{uv} l(c_{uv}) \tag{16}$$

where f_{uv} represent the number of demands that are generated at tile u and satisfied at tile v, c_{uv} is the driving distance from u to v, and $l(c_{uv}) = \max(0, 1 - a \cdot c_{uv})$ is a Hinge loss function. The goal is to maximize the objective function Equation (16) under the following constraints.

Budget Constraint. Let $v' \in V'$ denote a candidate tile to deploy new piles, $e_{v'}^p$ denote the monetary cost of deploying a charging pile in tile v' and $e_{v'}^s$ denote the site construction cost. Then the charging station deployment cost at tile v' is $e_{v'}^p \cdot m_{v'} + e_{v'}^s$, where $m_{v'} \in \mathbb{Z}^+$ is the number of piles. Assuming that the decision maker has an overall budget constraint B to building new charging stations, then the total cost of the construction cannot exceed the overall budget B,

as expressed in the Equation (17) below:

$$\sum_{v' \in V'} e_{v'}^p \cdot m_{v'} + e_{v'}^s \leq B. \tag{17}$$

The site construction cost $e_{v'}^s$ is considered separately instead of being shared by piles because it is more practical to build a few charging stations with many charging piles than lots of charging stations with a few (or even a single) charging piles.

Spacing Constraint. In some cases, the decision maker may want the charging stations to be sufficiently spaced out. For example, in an early construction stage, a spaced out layout can improve the coverage and publicity. If an uneven spatial partition is adopted, e.g., Zip+4 area partition, the spacing constraint is useful to avoid the final result clustering where small tiles are too intensive. An indicator $\delta_{vv'}$ is introduced to reflect the proximity status between a candidate tile v' and tile v (may be v' itself):

$$\delta_{vv'} = \begin{cases} 1, & \text{if } c_{vv'} < r \text{ and } m_v \times m_{v'} > 0 \\ 0, & \text{else} \end{cases} \tag{18}$$

where $c_{vv'}$ is the distance between the two tiles, $m_v \in \mathbb{Z}_0^+$ and $m_{v'} \in \mathbb{Z}^+$ are the numbers of piles of the two tiles, respectively, and r is the minimum spacing specified by the decision maker. Then the spacing constraint can written as:

$$\sum_{v \in V} \delta_{vv'} \leq 1, \forall v' \in V' \tag{19}$$

Demand Constraints. Let d_u denote the number of considered demands generated at tile u, then d_u has to be equal to the total number of satisfied demands, i.e. the "demand flow conservation":

$$\sum_{v \in V} f_{uv} = d_u, \forall u \in U \tag{20}$$

The total capacity of stations of a tile v is limited. The demands that can be served at tile v have to be less than the piles within the tile. This is another demand constraint:

$$\sum_{u \in U} f_{uv} \leq m_v, \forall v \in V \tag{21}$$

C. Preference Parameter

Previous studies mainly focus on improving the coverage of CPN using traffic data. These techniques are very useful in the early stage of CPN construction. However, when the city is basically covered by the CPN but the adoption of EV is still low, these techniques may not be able to provide a ideal solution since the traffic data is not enough to reflect the mobility of EVs due to the low adoption rate. On the other hand, our charging demand estimation method can better capture the needs of existing EV drivers, but cannot fully reflect the mobility of the potential EV users, i.e., FFV drivers. There is a tradeoff between attracting more FFV drivers and improving the convenience of existing EV users.
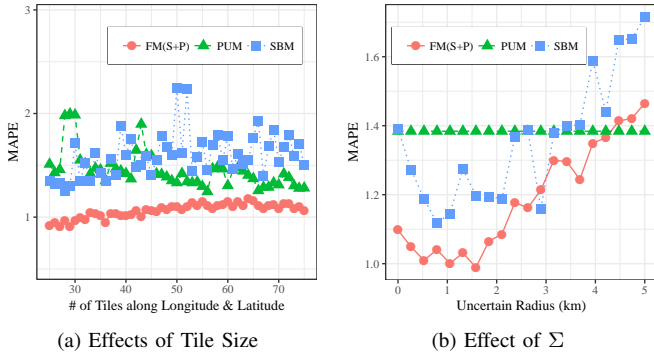
(a) Effects of Tile Size   (b) Effect of Σ

Figure 10.   Charging Demand Estimation

(a) Impact of Budget   (b) Impact of Tile Size

Figure 11.   Evaluation of Deployment Scheme

To reflect this tradeoff, we introduce a preference parameter β and rewrite the objective function (16) as follows

$$\sum_{u \in U} \sum_{v \in V} [\beta f_{uv} + (1-\beta) f'_{uv}] l(c_{uv}) \qquad (22)$$

where f_{uv} is the demand flow estimated based on pile usage and app activity data, and f'_{uv} is the demand flow estimated based on traffic data. $\beta \in [0,1]$ reflects how much the decision maker cares about the needs of current EV users. If $\beta = 1$, the charging stations planning is totally based on the data of EV users' activities. If $\beta = 0$, it means the decision maker bases the planning on FFVs' mobility. In this way, the planning problem in different situations are unified in the same framework and the decision makers can conduct charging station planning flexibly.

VI. EVALUATION

In this section, we conduct extensive experiments to evaluate the performance of our system. We first describe the evaluation methodology. Then the results of a series experiments are presented and discussed.

A. Methodology

We first evaluate the demand estimation algorithm. To the best of our knowledge we are the first work that estimates charging demand based on charging pile usage data and app activities, therefore we proposed two benchmark models to compare with our method. The baseline *SBM* (Search Behavior Model) estimates the demand based on search data and *PUM* (Pile Usage Model) only uses pile usage data. The proposed method, *FM* (Fusion Model), is compared against with *SBM* and *PUM* to show the effectiveness of multi-model fusion.

Note that it is very difficult to accurately obtain the ground truth of charging demand θ, unless the vehicle status and the driver's intention can be continuously measured. For evaluation purpose, the origins of navigation events are used as the ground truth because in general they are good proxy of where the charging demand is generated, as discussed
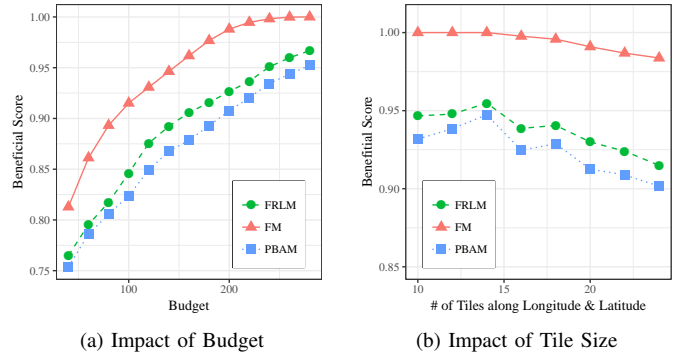
in Section IV-B. Therefore, in the experiment of charging demand estimation, navigation data are reserved for testing and only search data and pile usage data are used by *FM* and the baselines.

The estimate charging demand distributions given by *FM* and baselines are compared with the ground truth distribution by tile. The models are tested with the Mean Average Percent Error (MAPE) defined as

$$MAPE = \frac{1}{N_b} \sum_{i}^{N_b} \frac{\|T_i - \bar{T}_i\|}{\bar{T}_i}$$

where N_b is the total number of tiles, T_i is the total number of charging demands of the tile i and $\bar{T}_i$ is the ground truth of the total number of demands of the tile i.

Then we evaluate the deployment scheme given by our system with two start-of-the-art methods. The flow-refueling location model (FRLM) [19] locates p facilities as to intercept as many as trips as possible. A trip is considered captured if it is possible to travel from the origin to the destination and back without running out of electricity. FRLM requires the endurance mileage data to determine if a trip is possible. The nominal mileage data of 1,631 EVs are extracted from the vehicle register information on the official platform (Figure 13a). The charging demands are estimated based on traffic flow data. We use the taxi trajectory dataset published by Microsoft [20], [21] to calculate traffic flow. The dataset contains one-week trajectories of 10,375 taxis with total mileage of 9 million kilometers. Another baseline is a parking-based assignment method (PBAM) [16] that estimates the charging demand based on parking demand and tries to minimize the demand-weighted deviation for charging. A dataset containing the information of most of the parking zones in Beijing is used [22] to estimate the parking demand. In the evaluation, we add site capacity constraints to the models, since as most of previous studies the site capacity is undefined or assumed to be unlimited in the two original models.

In order to infer the ground truth, we introduce another related dataset for evaluation purposes which contains de-
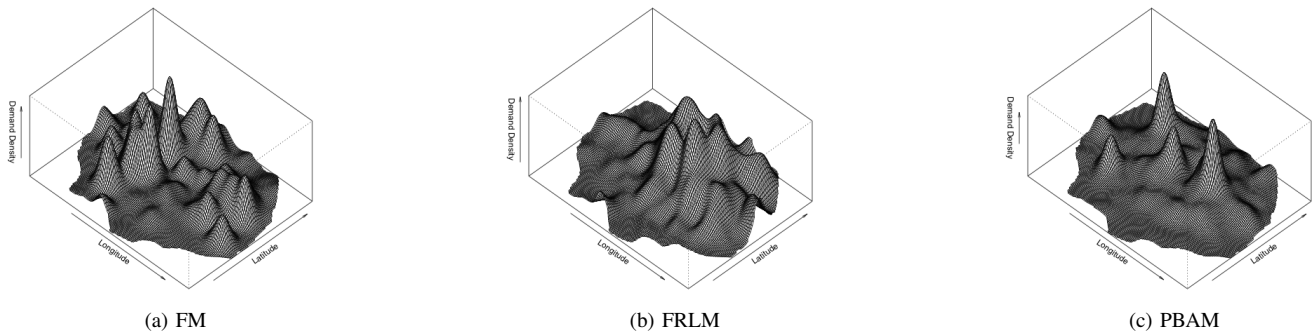
(a) FM   (b) FRLM   (c) PBAM

Figure 12.   Estimated Charging Demand Distribution

(a) EV Mileage Histogram   (b) User Charge Interval
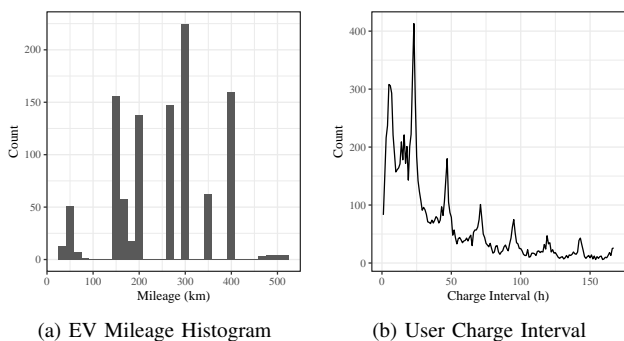
Figure 13.   EV User Profile

Figure 14.   Satisfaction

Table III
PILE USAGE DATASET SPECIFICATION

User ID	The unique ID of the user
Date time	The date and time of the request.
Request Location	The request location.

ployment requests from EV drivers. A request contains an user ID, the date, time and the coordinate of the location where the user wishes a pile to be deployed. The specification is shown in Table III.

B. Evaluation of Demand Estimation

We use the pile usage and search data from January 1st 2016 to August 31st 2017 to estimate charging demand. The navigation origins in the same period of time are used as ground truth. Note that the three datasets are filtered to be mutual exclusive. *FM* fuses pile usage and search data using the algorithm described in Section IV. Therefore, it is marked by *FM(S+P)* (Search + Pile). *SBM* only uses search data to estimate charging demand based on the model described in Subsection IV-C. While *PUM* uses the pile data and the method described in Subsection IV-E to estimate charging demand. The MAPE of estimation results are compared and shown in Figure 10.

Figure 10a gives the MAPE under different spatial partition granularity. In general, *FM* outperforms the other two
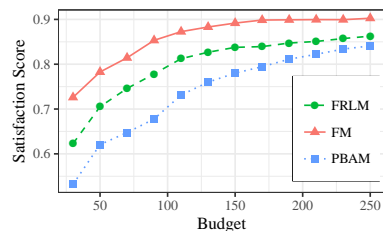
methods and the MAPE of *SBM* and *PUM* are 45.36% and 55.13% higher than the MAPE of *FM*, respectively. This is because pile usage data are gathered around charging stations while search data are more diversely distributed in the city (Figure 5). Relying only on one of the data sources leads to a biased estimation. *PUM* and *SBM* are more unstable because they only capture the charging demand either near the charging stations or far away from charging stations and a different partition may divide the estimated and ground truth demand that are in the same tile under the current partition into different tiles. *FM* is more stable because it combines the both sides and the effect is offset.

Figure 10b plots the MAPE under value of $\|\Sigma\|$. Σ is introduced in Equation (1) to reflect the uncertainty in the search origins. In the beginning, as the uncertain radius increase, the MAPE of *FM* and *SBM* decreases because $\|\Sigma\|$ gradually approaches the uncertainty in the search behavior. However, with the further increase of $\|\Sigma\|$, the uncertainty in the data is overestimated and the estimated distribution becomes dispersed leading to the increase of MAPE. Since *PUM* is not affected by Σ, its MAPE remains unchanged during the experiment.

C. Evaluation of Deployment Scheme

The pile usage and app activity data from January 1st 2016 to August 31st 2017 are used in these experiments. We leave out the data of the most recent three months, i.e. from June 1st to August 31st 2017, as the test set. The origins of navigation events in the test set are used as

| (a) FM | (b) FRLM | (c) PBAM |

Figure 15.   Deployment Scheme

ground truth charging demands. The deployment scheme is given by our system based on the the previous 17 months data. FRLM gives deployment scheme based on the traffic flow and EV mileage data. PBAM uses the parking data to give deployment suggestions. The study area is selected as the rectangle area from $N39.75658$ to $N40.02197$ and from $E116.202$ to $E116.5439$.

Figure 11a gives the social welfare defined as Equation (22) with $\beta = 1$, i.e., we only consider the needs of current EV users. The benefit score is normalized between 0 to 1. *FM* outperforms *FRLM* and *PBAM* up to 9.30% and 11.06% respectively. As the budget increases, the beneficial score gradually converges to a highest value where the budget is no longer the main constraint. When the budget is very large, each method can build enough stations to cover the needs. On the other hand, the beneficial score is also affected by the tile size because the differences of the demands within the same tile are not considered.

To show the impact of spatial granularity, Figure 11b gives the beneficial score of the three methods under different number of tiles along longitude and latitude. *FM* performs better than the other two methods during the experiment. As the tile size becomes smaller, the beneficial score of the three methods are gradually reduced. This is because smaller spatial granularity allows the finer structure of charging demand distribution to be considered.

In the two experiments, *FRLM* performs slightly better than *PBAM*. It is possibly because traffic flow data captures more charging demand than parking data does. For example, commercial EVs, such as electric taxis, are running for most of the time and rarely stop. However, taxi drivers charge their EV more frequently. Figure 13 shows the EV drivers' charge interval distribution derived based on pile usage data. It can be seen that almost half of the drivers recharge within 24 hours. Interestingly, except for the peak around 8 hours, we find that other peaks all appear at integer multiples of 24 hours. That suggests that many drivers charge their EV everyday or every few days. These drivers probably use EVs

to commute since commuters' routes are relatively fixed. The drivers who charge their EVs more than once per day are probably commercial EV drivers, e.g., taxi drivers or online-hailed car drivers. The peak around 8 hours matches the result from [23] which reports that the electric taxi drivers in Shenzhen charge three times a day on average.

Figure 12 shows the charging demand distribution estimated by *FM*, *FRLM* and *PBAM*. It can be seen that the three distributions show different patterns. Several peaks are observed in the west of the city in Figure 12a, whereas the peaks are gathered in the east in Figure 12b, which generally matches the pattern shown in Figure 3. All three methods estimate that there are many charging demand in the Chaoyang District. Both *FM* and *PBAM* estimate that there are hotspots in Haidian District in the northwest and the Fengtai District in the southwest.

Figure 15a shows the deployment scheme given by *FM*. Each of the red squares represents a selected location to deploy a new charging station. The deeper the color, the larger the size of the station (more piles). The number of tiles along longitude and latitude is set to 30. Four locations between the Fourth Ring Road and the Fifth Ring Road are selected to deploy large charging stations (two in the north and two in the southwest). Figure 15b gives the result of *FRLM*. Compared to the scheme given by *FM*, more sites in the center of the city are selected to deploy new stations and, unsurprisingly, many of them are near the main roads. The scheme given by *PBAM* is shown in Figure 15c. There are several locations selected by both *FRLM* and *PBAM*. This is probably because the two methods are both related to traffic conditions.

The deployment schemes are also evaluated in terms of the driver satisfaction and the result is shown in Figure 14. The percentage of requests that are satisfied is used as the satisfaction score. A request in a tile is considered satisfied if there is a new charging station deployed in the same tile. It can be seen that under the same budget, our approach can achieve higher customer satisfaction. When the budget

is low, *FM* outperforms *FRLM* and *PBAM* up to 16.57% and 36.49%, respectively.

VII. Conclusion

Building efficient charging networks in urban area to support the growing number of EVs is a very urgent and important issue, especially after the initial construction stage. This paper focuses on using users' charging behavior data to estimate charging demand distribution and provide charging station planning suggestions. Specifically, the search behavior of the charging mobile app users, the navigation behavior of the app users and the charging pile usage behavior are analyzed and modeled. A Bayesian-inference-based algorithm is proposed to fuse the three models. We formulate the charging station planning problem as a mixed integer programming problem with a flexible objective function.

With the help of our method, decision makers can better understand the real needs of the current EV drivers and adjust the parameters of the objective function to reflect their preference between improving the convenience of current EV drivers and attracting more FFV drivers. As a result, EV drivers will find it more convenient to find charging piles because charging stations are built at where they are most needed. The adoption rate of EV may be further improved because of the improvement of the quality of the charging service.

Acknowledgment

References

[1] "Electric vehicles and charging piles monitoring," http://www.evehicle.cn, Beijing Public Service Management Platform For Charging Facilities, 2017.

[2] "Technical specifications of remote service and management system for electric vehicles - part 3: Communication protocol and data forma," National Technical Committee of Auto Standardization, Tech. Rep., 8 2016.

[3] T. Chen, K. Kockelman, and M. Khan, "Locating electric vehicle charging stations: Parking-based assignment method for seattle, washington," *Transportation Research Record: Journal of the Transportation Research Board*, 2013.

[4] X. Xi, R. Sioshansi, and V. Marano, "Simulation–optimization model for location of a public electric vehicle charging infrastructure," *Transportation Research Part D: Transport and Environment*, 2013.

[5] I. Frade, A. Ribeiro, G. Gonçalves, and A. Antunes, "Optimal location of charging stations for electric vehicles in a neighborhood in lisbon, portugal," *Transportation research record: journal of the transportation research board*, 2011.

[6] Y.-W. Wang, "An optimal location choice model for recreation-oriented scooter recharge stations," *Transportation Research Part D: Transport and Environment*, 2007.

[7] Y.-W. Wang, "Locating battery exchange stations to serve tourism transport: A note," *Transportation Research Part D: Transport and Environment*, 2008.

[8] Y.-W. Wang, "Locating flow-recharging stations at tourist destinations to serve recreational travelers," *International Journal of Sustainable Transportation*, 2011.

[9] M. Kuby and S. Lim, "Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs," *Networks and Spatial Economics*, 2007.

[10] S. H. Chung and C. Kwon, "Multi-period planning for electric car charging station locations: A case of korean expressways," *European Journal of Operational Research*, 2015.

[11] M. Kuby, L. Lines, R. Schultz, Z. Xie, J.-G. Kim, and S. Lim, "Optimization of hydrogen stations in florida using the flow-refueling location model," *International journal of hydrogen energy*, 2009.

[12] Y.-W. Wang and C.-C. Lin, "Locating multiple types of recharging stations for battery-powered electric vehicle transport," *Transportation Research Part E: Logistics and Transportation Review*, 2013.

[13] N. Andrenacci, R. Ragona, and G. Valenti, "A demand-side approach to the optimal deployment of electric vehicle charging stations in metropolitan areas," *Applied Energy*, 2016.

[14] H. Cai, X. Jia, A. S. Chiu, X. Hu, and M. Xu, "Siting public electric vehicle charging stations in beijing using big-data informed travel patterns of the taxi fleet," *Transportation Research Part D: Transport and Environment*, 2014.

[15] J. Cavadas, G. H. de Almeida Correia, and J. Gouveia, "A mip model for locating slow-charging stations for electric vehicles in urban areas accounting for driver tours," *Transportation Research Part E: Logistics and Transportation Review*, 2015.

[16] T. D. Chen, K. M. Kockelman, M. Khan *et al.*, "The electric vehicle charging station location problem: a parking-based assignment method for seattle," in *Transportation Research Board 92nd Annual Meeting*, 2013.

[17] J. Dong, C. Liu, and Z. Lin, "Charging infrastructure planning for promoting battery electric vehicles: An activity-based approach using multiday travel data," *Transportation Research Part C: Emerging Technologies*, 2014.

[18] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.

[19] M. Kuby and S. Lim, "The flow-refueling location problem for alternative-fuel vehicles," *Socio-Economic Planning Sciences*, 2005.

[20] J. Yuan, Y. Zheng, C. Zhang, X. Xie, and G.-Z. Sun, "An interactive-voting based map matching algorithm," in *Mobile Data Management (MDM), 2010 Eleventh International Conference on*. IEEE, 2010.

[21] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*. ACM, 2010.

[22] B. C. Lab, "Parking places of beijing," http://www.beijingcitylab.com, 2014.

[23] Z. Tian, T. Jung, Y. Wang, F. Zhang, L. Tu, C. Xu, C. Tian, and X.-Y. Li, "Real-time charging station recommendation system for electric-vehicle taxis," *IEEE Transactions on Intelligent Transportation Systems*, 2016.