IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 16, 2020, accepted June 15, 2020, date of publication June 24, 2020, date of current version July 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004715

A Data-Driven Approach for Optimizing the EV Charging Stations Network

YU YANG, YONGKU ZHANG, AND XIANGFU MENG
School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China
Corresponding author: Xiangfu Meng (marxi@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 61772249.

ABSTRACT With the exhaustion of oil resources and the aggravation of environmental pollution, electric vehicles, as the main force of new energy consumption, have a more and more promising development prospect. In China, the utilization rate of charging facilities in the public is very low, and there is a large number of redundant charging stations that waste resources. Charging station congestion, meanwhile, is one of the reasons why it is difficult to charge for electric vehicles. This paper proposes a data-driven approach to optimize the existing charging station network by eliminating redundant charging stations, and to identify the charging station congestion areas in the original charging network to provide suggestions for further solving the difficulty of charging electric vehicles. Firstly, we infer that the fine-grained charging situation (consisting of the waiting time and the visiting rate) at different stations. Using a 3D tensor, we model the charging behavior of the electric vehicle, in which the three dimensions represent stations, hours, and days respectively. Secondly, for times and stations with sparse data, we use a context-aware tensor collaborative decomposition method to estimate the situation. For charging stations in a specific period of time, we separately set up a queue system for them to estimate their visiting rate and detect the distribution characteristics of EV charging hotspots in the city. Finally, we introduce a flexible scoring function to evaluate the usage benefits of charging stations and propose a heuristic network expansion algorithm to optimize the network. Applying the data-driven approach to Wuhan city, the results show that using our method can eliminate redundant sites while increasing utilization and find charging station congestion area to guide the government to further charging station planning. Our approach can be adapted for other optimal problems such as chain supermarket layout, public facility planning, and resources configuration using trajectory data.

INDEX TERMS Charging station layout, data mining, electric vehicles, tensor decomposition.

I. INTRODUCTION

In China, the government has invested a lot of money to build the infrastructure of charging stations [1], and improving the charging infrastructure system is the key to the large-scale implementation of electric vehicles [2], [3]. According to data released by China EV100 Forum in 2018, the utilization rate of the public charging facilities in China is less than 15%, and many charging stations even become zombie stations. The reason behind this phenomenon is that the layout of the charging stations network lacks rationality. More and more scholars began to study the optimization of site selection of charging stations for electric vehicles.

The optimal planning of EV charging stations has been studied from different perspectives. From EV charging demand, several methods have been proposed and utilized. In [4], a modeling framework for positioning multi-type BEV charging facilities is proposed to minimize public social costs and meet the demands of different types of BEV. Tu *et al.* [5] modeled the interaction between demands for electric taxis, electric taxis, and charging stations. To optimize the layout of charging stations for electric taxis, a spatial-temporal demand coverage method is proposed, which realizes a high-quality tradeoff between ET service level and charging service. To optimize charging stations, some studies have also begun to focus on the location and size of charging stations [6], [7]. Jian and Zhang [8] established a BASS model to predict the permeability and number of charging stations of electric vehicles in different years. Then, according to the queuing

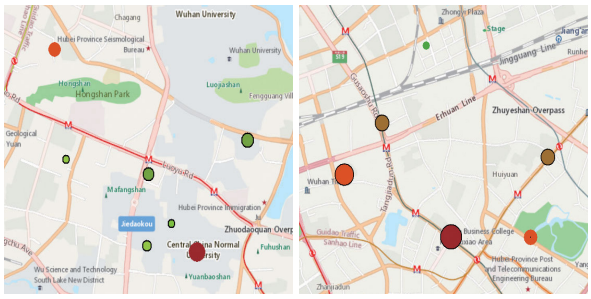
The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

theory, the optimal model of EV charging station location is proposed, which the total cost of the EV charging station is minimized by considering the constraint conditions such as charging station capacity. To enhance charging station's utilization and save corresponding investment costs by incorporating coordinated charging, Zhang *et al.* [9] proposed a new charging spot model, namely single output multiple cables charging spot (SOMC spot). The layout of charging stations for electric vehicles is a problem of location optimization, and many studies have revealed the optimal location of charging stations [10], [11]. Srinivasa *et al.* [12] proposes a two stage Grasshopper Optimization Algorithm (GOA) based Fuzzy multiobjective approach to determine the optimal location of EV charging stations and the number of vehicles at charging stations. Fortunately, the big data era has brought us unprecedented data in urban areas, providing complementary information for pinpointing the location and size of charging stations. Therefore, there are more and more data-driven methods in charging station planning, such as [5], [13], [14]. To improve the positioning strategy, a variety of data-based driving behavior and travel behavior analysis methods are adopted. In [15], the search behavior, navigation behavior, and usage patterns of charging related to electric vehicle users are analyzed and modeled, and a charging demand evaluation method based on Bayesian reasoning is proposed. Ge *et al.* [16] proposed a dynamic traffic simulation-based optimization planning technology for EV charging station in expressway network, aiming at the problem that the existing research could not consider the travel demand of vehicle space and time change. Some factors should be considered for network optimization of charging stations include:

(i) Charging behavior regularity. The optimization of the charging station network is directly affected by the charging behavior of EV. By analyzing the charging behavior of the whole city, the active degree of EV in each charging station can be mastered, and the distribution of charging hot spots of EV in the city can be found, which is the core of the charging station network.

(ii) The relationship between charging stations. When EV is in a lower state of charge, there is a certain competitive relationship between charging stations within the charging distance acceptable to electric vehicle users, because EV will choose the most suitable charging station within the range. Although the competitiveness strong charging station will attract more users to charge, once many users go to the station to charge for the same period of time (peak), the electric car charging needs to wait for a long time. At this time, nearby charging stations are often needed to relieve the charging pressure, and these charging stations will form a dependency relationship.

(iii) Required by the government. Figure 1 shows a local view of several charging stations. The size of the stations indicates the drivers' average arrival rate. We can see from figure 1(a) that charging stations are very dense in this area even though very few drivers have visited there (colored green). Therefore, the government could consider closing some of them to reduce waste. On the contrary, figure 1(b) indicates that a large amount of EVs has charged in the area, thus these charging stations have long waiting times (colored red). It might be worthwhile to consider building a new charging station nearby to relieve charging pressure. The government hopes to design an interactive optimization system based on the existing network, which users can set different tuning parameters according to the real requirements, discover and eliminate redundant charging stations, and identify the congestion area of the charging station to guide the next step of charging station planning.



(a) Demand surplus (b) Lack of demand

FIGURE 1. A local view of charging station.

There are prominent phenomena such as ''zombie charging station'' and ''charging difficulty.'' From the above literature review, it is obvious that the location optimization of charging stations has been discussed from many different angles. Few research results have fully solved the optimization of the actual charging station network under the background of charging station overflow, charging station competition, government requirements, and charging behavior. This is an important aspect of achieving a sustainable transport system. To better solve the above problems, this paper proposes a complete data-driven method to optimize the charging station network. The sources data explored include the driving trajectory of ET, POIs data, the distribution of charging stations, and various local features. Our main contributions are summarized as follows:

- Citywide EV's charging modeling: EVs GPS trajectories not only indicate the popularity of charging stations but also reflect the charging behavior of drivers during different periods of a day. We use a 3D tensor to model the correlation of time spent among different stations, the hourly period and the daily period.
- Dealing with data sparsity: Electric vehicle trajectory data is sparse, resulting in a sparse tensor. Due to the sparse limitation of data, it lacks the accuracy to fill the missing data only based on the non-zero component. To deal with tensor sparsity, we propose a context-aware tensor collaborative decomposition method to recover the time spent situation of different stations in urban areas by feeding the feature set during the tensor decomposition process. To solve the sparsity problem inside a

charging station, we estimate the total visiting rate by modeling each charging station as a queue system.

- New charging station usage benefit measurement method: We consider the charging station network optimization problem with various constraints and put forward a flexible charging station scoring function to evaluate the usage benefit of different stations with the distance and characteristics of the relationship of the charging stations (competition and dependence).
- Optimization method of charging station network: We propose a novel heuristic network expansion algorithm based on EVs' charging hot spots to solve the optimization problem of the existing network by eliminating redundant charging stations and identify the congestion areas of charging stations in the original network to guide the government in further charging station planning. To achieve better effectiveness, we also propose two different approaches to initialize the network expansion algorithm, which work well for less charging stations and more charging stations, respectively. We also design two different network expansion algorithms based on whether there is a quantity constraint of charging stations.

II. RELATED WORK

Data-Driven Urban Planning. The era of big data has arrived with the availability of massive amounts of mobility from vehicles, peoples, and transportation systems, and exploring urban movement regularity has gradually become an emerging and attractive field. Massive mobile data play an important role in urban planning, which can reflect the activity characteristics and various travel demands of people in the real world, and discover urban anomalies and human movement patterns [17], [18]. For example, based on the check-in and check-out data of users using shared bikes, the movement patterns of users are mined, and a hierarchical prediction model is proposed to predict the check-in and check-out situation of each station in the future, to accurately allocate the number of shared bikes at each site in advance [19]. According to taxi tracks, traffic patterns and POI distribution characteristics are mined to find different functional areas of the city [20]. In this paper, we focus on providing a data-driven method to optimize the existing charging station network, which is a more effective and economical planning method.

Charging station layout optimization. As a green transportation tool with low pollution, zero-emission, and high energy efficiency, electric vehicles are attracting more and more attention. Since the performance of electric vehicles will have a significant impact on future urban traffic behavior, a traffic model for charging demands needs to be considered. Mainul *et al.* [21] proposed a method to optimize the location of a quick charging station with comprehensive consideration of transportation loss, power grid loss, and construction cost, so as to meet the travel needs of travelers as much as possible. Ahmed *et al.* [22] focus on incorporating two important features into the optimization problem modeling: the level of decision-making over multiple time periods and the uncertainty of charging demands, including the number of electric vehicles charged and the long-distance travel set to be covered. Its goal is to determine the charging stations to be opened at each time period so as to maximize the expected value of the satisfied recharging demand over the entire planning horizon. In [23] and [9], EV charging behavior is divided into destination charging and emergency charging. According to the historical data of EV parking behavior, the specific charging demand corresponding to each charging behavior is predicted, and the prediction method is essentially a Monte Carlo simulation method [24]. The deployment of EV public charging stations can alleviate the range anxiety of EV drivers and ensure that the EV maintains its current activities. Pan *et al.* [25] developed a positioning model for electric vehicle public charging stations with the goal of maximizing the maintenance of the existing activities of electric vehicle drivers. First, the deterministic process of electric vehicle charging selection is proposed to simulate the electric vehicle driver's charging selection behavior. Secondly, an electric vehicle public charging station coverage positioning model is proposed to maximize the existing activities of electric vehicle drivers. Based on real trajectory data, Ahn *et al.* [26] analyzed the driving patterns of electric taxis under different constraints. Luo *et al.* [27] studied the interaction between travel patterns, EV driver behavior, urban road network, power grid network, and charging station layout, and proposed a multi-stage charging station layout method with different EV penetration rates. Many scholars proposed various optimization objective functions based on the moving data of vehicles to study the optimal location of urban charging stations [5], [13], [14], [28]. However, most of these optimization work ignored the existence of real charging stations to optimize the charging station network. On the contrary, our optimization work is based on the existing network, charging behavior characteristics of electric vehicles, and the relationship between charging stations (competition and dependency). We provide urban planners with a more economical and feasible optimization plan to eliminate redundant sites and reduce resource waste.

Tensor decomposition for urban computing. The problem of sparse data or the absence of any data is inevitable in the field of urban computing, where tensor decomposition methods containing multiple data sets have been widely applied [17]. Previous studies have proved the advantages of tensor decomposition in solving multidimensional data input [29]–[32]. In [33], noise data, user check-in data, road network, and POIs data were combined with a tensor decomposition algorithm to identify the noise distribution and noise composition of each region of the city. [34], [35] use a context-aware tensor decomposition that contains additional information (such as historical correlations between items, geographic and spatial-temporal characteristics) for estimation and recommendation. In this article, we use a similar

approach, our goals are different from theirs, and we include more data sources (for example, charging station data).

III. PROBLEM FORMULATION

In this section, we will clarify some definitions and terms used in this paper, and define the optimization problem of the charging station network, and outline our solution framework.

A. DEFINITIONS

Definition 1 (Road Network): A road network $RN = \{V, E\}$ (where the vertex set V represents intersections and the edge set E represents all relevant road segments.), and each road segment $e \in E$ has a level $e.level$ (e.g., a highway or a street), a length $e.len$.

Definition 2 (POI): A point of interest POI is a venue (like a school and shopping mall) in the physical world. It is described by a name, a latitude, a longitude, address, and category.

Definition 3 (Trajectory): A spatial trajectory Tr is a sequence of time-ordered spatial points, $Tr = \{p\prime_1, p\prime_2, \ldots.., p\prime_n\}$, where each point $p\prime$ is composed of a latitude, a longitude, and a timestamp.

Definition 4 (Charging Station (CS)): $S = \{s_1, s_2, \ldots.., s_k\}$ represents the set of charging stations. Charging stations are mainly divided into two levels, the parent level, and the child level. Comparing two charging stations in pairs, we define the parent charging station with higher usage benefit, while the smaller one is the child charging station. A parent charging station has many child charging stations, and similarly, the child charging stations also belong to multiple parent charging stations. This is a many-to-many existence relationship.

Definition 5 (Charging event (CE)): A charging event describes the phenomenon of an EV charging at charging stations, which consists of arrival time aT, the selected specific charging station s_j and departure time dT, $CE = \{aT, s_j, dT\}$. The difference between arrival time and departure time represents the duration of the charging event.

Definition 6 (Recording Cell and Recording Cube): A recording cell $W(i, j, k)$ is a spatial-temporal division for charging events, which stores the total number of charging events (who fall in this cell) of the timestamp of hour h_i and the timestamp of day d_k. The recording cell is a fine-grained charging situation, and we can use it to analyze the charging behavior of urban EVs. We only focus on the charging events that fall within the cell and the two indicators (i.e., time spent and visiting rate). Time spent refers to the average charging time of an electric vehicle in this cell during a certain period. The visiting rate indicates how many drivers have visited the cell. All recording cells are combined to form a cube, as shown in figure 2.

B. NAVIGATING EV CHARGING ON SOC

We use the loss model which is a simple linear model based on the maximum range know from data sheets of manufacturers,

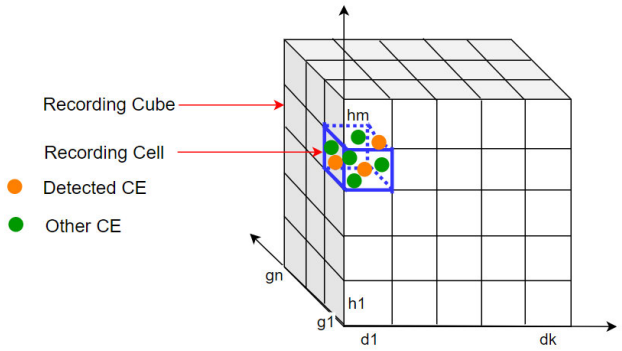


FIGURE 2. Recording cube and recording cell.

e.g., according to the performance of used EV BYD e6 in an urban area, 100% capacity can travel about 250 km. To ensure the service life of the battery, the depth of discharge (DOD) cannot be too large, so the safe state of charge needs to be reserved S_{safe}. We assume that all-electric cars have the same capacity, and the full charge state is denoted by E'. The maximum range of electric vehicle is D_{max} at the full state of charge, so when the electric vehicle is in S_{safe} state, the distance EV can continue to drive is

$$D_{rm} = \frac{S_{safe} \cdot D_{max}}{E'} \tag{1}$$

C. PROBLEM STATEMENT

Problem definition. Given a set of charging station $S = \{s_1, s_2, \ldots.., s_k\}$ with overflow status, the shortest distance matrix $D \in R^{k \times k}$, K charging hotspots, a tuning parameter γ, and a total optimal number of stations $\overline{P}$. Our data-driven optimization solution is designed to identify a subset of sites $S' \subseteq S$, that follows two criteria: (i) quantity constraint of stations, (ii) maximum usage benefit.

Quantity constraint of charging stations ($\overline{P}$). The construction and operation costs of charging stations are high. However, there is a large number of redundant charging stations in the actual CS network, resulting in a great waste of resources. At the same time, the government planning department has the overall budget constraint, here we convert the budget constraint into the quantity limit of CS to control the cost output and the range of the quantity of CS in the network optimization scheme.

Maximum usage benefit (U). The goal here is to make the best use of the deployed charging stations in a limited number of charging stations, which should follow three criteria: (i) In the process of network optimization, the charging pressure of each charging station must be considered. We regard a station as the overflow station when the peak visit and the average visit of the station deviate from the overall average visit of charging stations by 75%. The overflow status of a charging station is represented by 0 or 1, where 1 represents overflow and 0 indicates that there is no overflow at the charging station. It is worth noting that the size of the deviation value can be set according to the specific requirements of the experts.

TABLE 1. Nomenclature and symbol.

EV	Electric vehicle	f_G	Geographical feature matrix
ET	Electric taxi	f_T	Traffic feature matrix
CE	Charging event	f_A	Area feature matrix
CS	Charging station	$\frac{1}{\mu}$	Average service time
SOC	State-of-charge	W_s	Time spent of station
S_{safe}	Safe state of charge	λ	Arrival rate of station
T_r	Set of trajectory	$\overline{P}$	Quantity constraint of the optimal network
$W^{h\times s\times d}$	Three dimensional tensor	γ	A tuning parameter
S	Set of charging station	S'	Set of optimal charging station
$W(i,j,k)$	Recording cell	h_i	The ith hour of day
d_k	The kth day of month	s_j	The jth charging station
s_{ij}	The ith child station of the parent station s_j	$U_{s_{ij}}$	Usage benefit score of the child station s_{ij}
$V(s_{ij})$	Visit of station s_{ij}	$s_{ij}.l$	The distance between the parent station s_j and its child station s_{ij}
D	Shortest distance matrix	D_{rm}	Maximum charging distance acceptable to driver
D_{max}	maximum range at full state of charge	E'	Full state of charge
$min(S.l)$	The shortest distance between all stations	K	Number of starting sites

If the charging station is prone to overflow during the charging peak period, we should keep the charging station nearby as much as possible to relieve the local charging pressure in the optimization process. This kind of overflow problem that can be solved by the nearby charging station is a pseudo overflow that caused by the charging habits of EV users. (ii) As much as possible to provide charging convenience for more users, which requires the charging station to be built at a site with high visit (a competitive site), covering more users' charging demands. (iii) As long as possible, continuous travel routes of drivers are covered. Note that in charging station planning, continuous travel route coverage is critical as it can directly affect the driver's user experience.

D. USAGE BENEFIT SCORING FUNCTION

In this paper, charging stations that are preferentially expanded to the network are regarded as parent charging stations, and those that are not expanded or later expanded are called child charging stations. We put parent charging stations into the result set S'. The unselected child charging stations related to parent charging stations were put into the candidate set C as the new candidates. In order to better judge the usage benefit of child charging stations, we propose a flexible scoring function.

$$U_{s_{ij}} = \gamma^{\pm\frac{s_{ij}\cdot l}{min(S\cdot l)}} \times V(s_{ij}), \quad \gamma > 1 \tag{2}$$

where $U_{s_{ij}}$ represents the usage benefit score of the jth child charging station s_{ij} of the parent station s_i. U is the function to calculate the score of each station, where $\frac{s_{ij}\cdot l}{min(S\cdot l)}$ normalizes the distance between all child stations and their parent stations (where $min(S.l)$ is the shortest distance between all charging stations), with the guarantee that its value is no less than 1. $s_{ij}.l$ is the driving distance between the parent charging station s_i and its child charging station s_{ij}. $V(s_{ij})$ is the visit of the child station s_{ij}. γ is a tuning parameter to set the preference on the distance between parent and its child charging stations. The reason for designing a score function using the exponential function of the normalized length is that when $\gamma > 1$ and the parent charging stations in the result set overflows, We give priority to solving the overflow problem of the parent charging station, that is, to keep the charging station nearby the parent charging station as far as possible to expand the network, ease the local charging pressure, and improve the service quality of charging stations. So, the child charging stations within the scope D_{rm} of the overflow parent charging station get higher scores than other child charging stations outside the scope D_{rm} (i.e., the maximum charging distance acceptable for EV), that is, with the exponential function $\gamma^{-\frac{s_{ij}\cdot l}{min(S.l)}}$. When $\gamma > 1$ and the parent charging station in the result set has no overflow phenomenon, we prefer to cover more users and more continuous travel routes. Therefore, the child charging station with higher visiting rate outside the range D_{rm} of the parent charging station will get higher scores, that is, this term $\gamma^{\frac{s_{ij}\cdot l}{min(S.l)}}$ will be added in the calculation. Since the parent charging station is sufficient to meet the local charging demands, we should prefer the more popular child charging station far away from the parent charging station to expand the network to meet the charging demands of more EVs and more continuous travel routes.

E. FRAMEWORK

Figure 3 gives an overview of our system, which consists of two main components:

(i) **Pre-processing.** This component takes ET trajectories, road network, POIs and station data and performs three main tasks: 1) Trajectory data parsing that removes the outliner GPS points and trajectory map-matching that projects the drivers-generated trajectory onto the corresponding road segment to identify charging events. 2) Context-aware

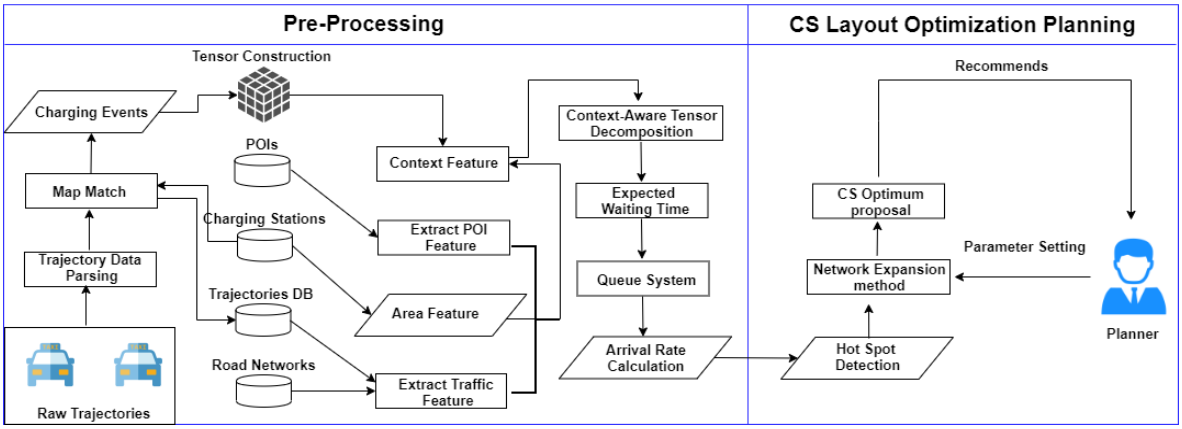


FIGURE 3. An overview of system.

TABLE 2. The format of taxi GPS data in the city of Wuhan, China.

VIN	Timestamp(/s)	Speed(m/s)	Charging status	SOC	Longitude	Latitude
LGJE13A0FM39****	1525909912	50.8	Not charging	44%	114.18493	30.48467
LGJE13A0FM40****	1525938594	19.7	Not charging	16%	114.22084	30.58308
...	...	...	...	...	...	...
LGJE13A0FM02****	1525959470	19.9	Not charging	81%	114.82207	30.66023
LGJE13A0FM04****	1525907271	0	Charging	96%	114.26301	30.53480
LGJE13A0FM60****	1525939456	0	Charging completion	100%	114.09734	30.72631

tensor collaborative decomposition, which estimates the waiting time at charging stations when data is sparse. 3) Queue system, which calculates the overall arrival rate at the charging station from the inferred waiting time during a period.

(ii) **CS Layout Optimization Planning.** This component takes charging hotspots and urban planner's parameters (e.g., a value $\overline{P}$ and tuning parameter γ) and outputs the charging stations layout recommendation results. This component contains three main components: 1) Charging hotspots detection. We compute average visits of stations by aggregating the record cells corresponding to the same charging station together and using these cells' average. Then, we rank all stations with their visits and extract the highest visit stations as charging hotspots for electric vehicles. Finally, we briefly analyze some realistic constraints and requirements of government planning departments that affect the layout of the charging station network. 2) A hotspots-based heuristic network expansion, where we propose a flexible charging station scoring function. And we initialize the greedy-based network expansion with the charging hotspots and continuously expand to select the station with the highest usage benefit score until the stop constraint condition is satisfied.

IV. WAITING TIME LEARNING

We detect charging events according to the change of charging status of electric vehicles. Each ET has been installed with a smart terminal connected with a GPS receiver, which records data concerning the vehicle identification, time, speed, position, charging status, the state of Charge (SOC), longitude, and latitude, and its sampling interval is 10 seconds. Table 2 describes the GPS format of ETs and examples. Figure 4 shows the two cut edges of cube W in the charging station dimension. If there are enough CEs contained, we can use their average duration to estimate the waiting time (including both waiting time for charging and service time) of the cell. The shades of color represent different numbers of charging events (CE). Some cells with sparse data and the reason behind this may be due to the lack of trajectories data. Tensor factorization is applied to the high dimensional prediction problem [33], [36], [37]. To predict the expected waiting time (i.e., time spent) of the cells with sparse data accurately, we propose a context-aware tensor collaborative decomposition method to solve the data sparsity problem. The value of each entry in tensor W is normalized to [0,1] for decomposition.

A. CONTEXTUAL FEATURES EXTRACTION

To deal with the data sparsity problem, we first extract three categories of features of charging stations, geographical feature, traffic feature and area feature.

(i) Geographical feature f_G. Feature f_G is extracted from POIs falling in a region near to stations. To detect the correlation between POI categories C and charging stations, we applied the standard proposed by Jensen *et al.* [38], which is given as

$$F_C = \frac{\#t\_location\{C, S\}}{\#C} \tag{3}$$

Algorithm 1 Context-Aware Tensor Collaborative Decomposition

Input: tensor W, matrix F, an error threshold ε.
Output: S, H, D, Z.

Initialize $z \in R^{d_H \times d_S \times d_D}$, $S \in R^{s \times d_S}$, $H \in R^{h \times d_H}$, $D \in R^{d \times d_D}$, $U \in R^{d_S \times R}$ with small random values;
while $Loss_l - Loss_{l+1} > \varepsilon$ **do**
  foreach $W_{ijk} \neq 0$ **do**
    Y_{ijk}=$Z \times_H H_{i*} \times_S S_{j*} \times_D D_{k*}$;
    $H_{i*} \leftarrow H_{i*}$-$\eta\lambda_2 H_{i*}$-$\eta(Y_{ijk} - W_{ijk}) \times Z \times_S S_{j*} \times_D D_{k*}$;
    $S_{j*} \leftarrow S_{j*} - \eta\lambda_2 S_{j*} - \eta(Y_{ijk} - W_{ijk}) \times Z \times_H H_{i*} \times_D D_{k*} - \eta\lambda_1(S_{j*} \times U - F_{j*}) \times U$;
    $D_{k*} \leftarrow D_{k*} - \eta\lambda_2 D_{k*} - \eta(Y_{ijk} - W_{ijk}) \times Z \times_H H_{i*} \times_S S_{j*}$;
    $Z \leftarrow Z - \eta\lambda_2 z - \eta(Y_{ijk} - W_{ijk}) \times H_{i*} \otimes S_{j*} \otimes D_{k*}$;
    $U \leftarrow U - \eta\lambda_2 U - \eta\lambda_1(S_{j*} \times U - F_{j*}) \times S_{j*}$;
Return H, S, U, D, Z.

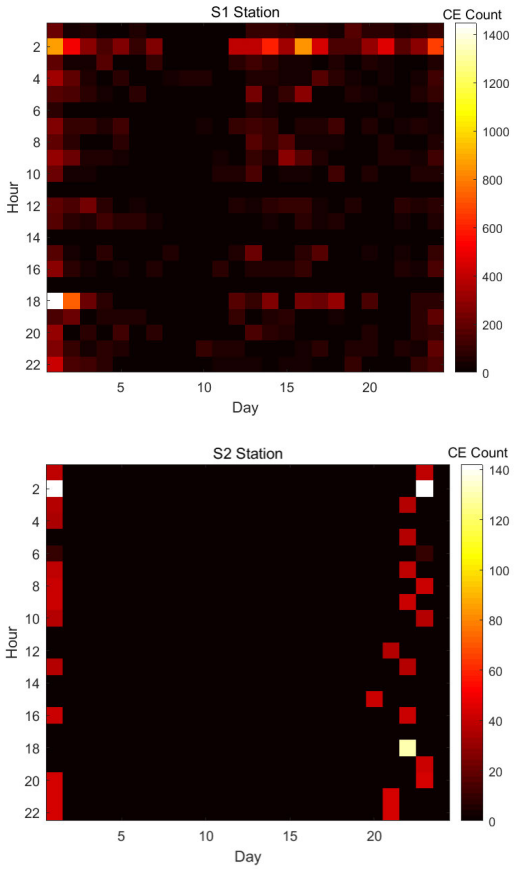


FIGURE 4. Detected CEs' heatmap in two charging stations

where $\#t_location(C, S)$ denotes the frequency of co-location for category C with the charging station, and $\#C$ refers to the individual frequency. According to statistics, the top 5 discovered POIs are {Transportation Facilities, Shopping, Food& Dining, Home& Family, Education&Arts, Vehicle Service}. The geographical feature of the charging station is given as

$$f_G(S_i) = \sum_C f(C, S_i) \cdot F_C \tag{4}$$

where $f(C, S_i)$ represents the frequency of category C near to the specific charging station S_i.

(ii) Traffic feature f_T. The traffic feature depends on the traffic flow and competition around charging stations. We estimated the traffic flow on the road near to the charging station by mapping a mass of trajectories data generated by ETs to the road network. The influence of traffic flow on the surrounding station is measured by referring to the method defined in [39].

$$TF(e \rightarrow s_i) = TF_e \cdot \frac{\frac{1}{dist(e,s_i)}}{\sum_{s_j} \frac{1}{dist(e,s_j)}} \tag{5}$$

where $dist(e, s_i)$ represents the distance between the road segement e and charging station s_i. TF_e represents the traffic flow in the road segement e. Finally, the traffic feature of the charging station is

$$f_T = \sum_e TF(e \rightarrow S_i) \tag{6}$$

(iii) Area feature f_A. The area feature of the charging station will directly affect its passenger capability, thus affecting the waiting time at the station. We get the area of each charging station by manual marking.

Ultimately, by putting together the geographical feature, traffic feature, and area feature of the station into a vector, we formulate a matrix $F^{S \times R}$ (R denotes the dimension of contextual features and S is the number of station) that incorporates the similarity among different stations. Intuitively, stations with similar contextual features could have a similar waiting time situation.

B. CONTEXT-AWARE TENSOR COLLABORATIVE DECOMPOSITION

We model the time spent at each station using a tensor, $W^{h \times s \times d}$ with three dimensions denoting h hour time slots, s charging stations, and d day time slots, respectively. $W(i, j, k)$ stores the time spent of stations. A common way to fill in the missing tensors is to decompose W into the product of some (low-rank) matrix and a core tensor, based on W's

non-zero entries. For example, we can decomposition W into the multiplication of three matrices, $H \in R^{h \times d_H}$, $S \in R^{s \times d_S}$, $D \in R^{d \times d_D}$ and a core tensor $Z \in R^{d_H \times d_S \times d_D}$, using a tucker decomposition model [40]. We can recover the missing values in W by equation 7.

$$W = Z \times_H H \times_S S \times_D D \tag{7}$$

The Symbol "$\times_H$" denotes the matrix multiplication, where the subscript stands for the direction, e.g., $T = Z \times_H H$ is $T_{ijk} = \sum_{i=1}^{d_H} Z_{ijk} \times Hij$.

Economists have found that competition among charging stations is influenced by both external and internal factors (e.g., location, nearby traffic flow, and its area, etc.). Therefore, these factors may affect the time spent on charging stations in a certain period. To achieve a higher accuracy of filling in the missing entries of W, we decomposition W with contextual feature matrices F collaboratively. Matrix F can be factorized into the multiplication of two low-rank matrices, $F = S \times U$, where $S \in R^{S \times d_S}$ and $U \in R^{d_S \times R}$. The objective function control the error of the decomposition is defined as equation 8:

$$\begin{aligned} L(Z, H, S, D, U) = &\frac{1}{2}||W - Z \times_H H \times_S S \times_D D||^2 \\ &+ \frac{\lambda_1}{2}||F - SU||^2 \\ &+ \frac{\lambda_2}{2}(||Z||^2 + ||H||^2 + ||S||^2 + ||D||^2) \end{aligned} \tag{8}$$

where $||\cdot||$ denotes the l_2 norm. $||W - Z \times_H H \times_S S \times_D D||^2$ is to control the error of decomposition W. $||Z||^2 + ||H||^2 + ||S||^2 + ||D||^2$ is a regularization penalty to avoid over-fitting. $||F - SU||^2$ is to control the error of factorization F. λ_1 and λ_2 are parameters controlling the contribution of each part during the collaborative decomposition. By minimizing the objective function, we can get optimized H, S and D. In our model, W and F share matrix S. The dense representation of F contributes to the generation of a relatively accurate S, which reduce the decomposition error of W. In other words, the knowledge from geographical features, traffic features, and area feature is propagated into tensor W. Equation 8 is to guarantees our model could reconstruct the value as accurately as possible. Algorithm 1 shows the pseudocode of the context-aware tensor collaborative decomposition. We use stochastic gradient descent to find a local optimization.

V. VISITING RATE CALCULATION AND CHARGING HOTSPOTS

We want to know how many EVs have visited the CS, form which we can analyze urban EVs charging behavior and obtain their charging hotspots. However, our data cover only about 800 electric taxis, a small fraction of the total number of electric vehicles in the city. To solve the problem of sparsity inside the charging stations, we regard the charging station of each cell as a queue system to estimate the visiting rate.

Queue System. There is a great deal of randomness when EVs enter a CS for charging. Literature [41] shows that this process conforms to the $M/M/c$ queue system. To reduce the complexity of the system, we made some simplifications. First, we assume that a CS is a queue system, ignoring the transmission from one queue to another. Moreover, when a vehicle arrives at the charging station, if there are idle charging piles, the vehicle always joins the queue immediately for charging. If no idle charging piles are available, the EV enters the waiting state. In each cell, the arrival flow of vehicles in a queue satisfies a homogeneous Poisson process $N(t, \lambda_i)$ in the period [0,t]. For the ith queue Q_i, we assume it has c_i servers. In the queue system, given the random process of vehicles' arrival and the allocation of server service time, The equilibrium indicators, waiting time, system time and other equilibrium indicators can be obtained as follows:

$$\begin{aligned} W_s = &\frac{\frac{\lambda_i^{c_i}}{\mu^{c_i+1}(c_i-1)!}}{(1 - \frac{\lambda_i}{c_i\mu})^2} \cdot [\sum_{k=0}^{c_i-1} \frac{1}{k!}(\frac{\lambda_i}{\mu})^k \\ &+ \frac{1}{c_i! - \frac{\lambda_i(c_i-1!)}{\mu}}(\frac{\lambda_i}{\mu})^{c_i}]^{-1} + \frac{1}{\mu} \end{aligned} \tag{9}$$

where λ_i is the the ith queue's average visiting rate, and $\frac{1}{\mu}$ is average service time expressed as the average of the top 800 shorted duration. W_s is the equilibrium system time (incorporating both the waiting time and service time), the length of time the vehicle is expected to stay, and we use each cell's expected time spent to denote it. Intuitively, once given μ, c_i, and W_s, it is easy to calculate the visiting rate (i.e, arrival rate) of each cell directly by numerical algorithms, such as the Newton method. Finally, we calculate the arrival rate λ_i of the charging station within a certain time through equation 9. Finally, charging stations are sorted in descending order according to visit, and those near to the top are regarded as urban EV charging hotspots.

VI. NETWORK EXPANSION ALGORITHM

Main idea. The intuition of the heuristic network expansion algorithm is to expand a set of K starting charging hot spots in the station network. This is inspired by the charging behavior characteristics of EVs, namely, spatial charging hot spots and star-shaped mobile modes.

Spatial charging hot spots. The charging hot spots of EVs are formed by the higher visitor volume of the station. Figure 5 shows the charging hot spot with the higher visit, where illustrates a popular shopping mall. The reason behind the observation is simple: because the site is located between three different urban functional areas (Entertainment, Business to Business and Food&Dining), where the higher passenger flow which indicates the higher charging demand.

Star-shaped mobility patterns. We further studied the travel patterns of electric vehicles after charging at the hotspot, and found that EVs will start from the same starting point and reach different destinations, and present a star-shaped mobility pattern, as the red arrows in figure 5.

Algorithm 2 The Network Expansion Algorithm Based on Spatial Hotspots

Input: charging station set S, charging station visits set V, matrix D, a value K, the tuning parameter γ and a value $\overline{P}$.
Output: Optimal station set S' and redundant charging station set I.

```
1  //Initialization;
2  Parent station set S' ← K starting spatial hot spots;
3  Candidate set C ← select child charging stations of parent charging stations in S';
4  Remaining number of station P̄ ← P̄ - S';
5  //Network Expansion;
6  while number of station P̄ > 0 do
7      s_next ← φ, C' ← φ;
8      if an overflow site exists in the result set S' then
9          Retrieve the child stations of all overflow parent stations form C and insert them into the set C';
10         for s_ij ∈ candidate set C' do
11             Get the usage benefit score U(s_ij) by equation (2);
12             List L ← s_ij and its usage score;
13         Traverse L to find the maximum value of the usage benefit score and the corresponding child station s_next;
14         S' ← S' ∪ s_next;
15         P̄ ← P̄ − S';
16         C ← C∪ none-selected child station of s_next;
17         Calculate the visit of s_next after users charging at the none-selected stations within the scope D_rm of s_next and
           overflowed users of the parent station are allocated to the new site s_next;
18         if s_next don't overflow then
19             I ← child stations within the scope D_rm of s_next;
20     else
21         for s_ij ∈ candidate set C do
22             Get the usage benefit score U(s_ij) by equation (2);
23             List L ← s_ij and its usage score;
24         Traverse L to find the maximum value of the usage benefit score and the corresponding child station s_next;
25         S' ← S' ∪ s_next;
26         P̄ ← P̄ − S';
27         C ← C∪ none-selected child station of s_next;
28         Calculate the visit of s_next after the users charging at the none-selected stations within the scope D_rm of s_next;
29         if s_next don't overflow then
30             I ← child stations within the scope D_rm of s_next;
31 I ← (S − S') ∪ I;
32 //Termination;
33 Return S', I.
```

In this section, we have designed two network expansion algorithms to achieve the optimization of the charging station network in different situations.

A. THE NETWORK EXPANSION ALGORITHM BASED ON SPATIAL HOTSPOTS

Based on the observation of charging hotspots and star-shaped mobility patterns of EVs, and the quantity constraint of charging stations, a network expansion algorithm based on spatial hot spots is proposed to optimize the network of charging stations. The algorithm starts from charging hotspots and expands greedily the optimal site, until reach the stop condition. The algorithm actually extends the incremental network in the road network, e.g., [42].

Algorithm Design. Algorithm 2 gives the pseudo-code of the network expansion method with the quantity constraint of charging stations $\overline{P}$. In the initialization stage, the algorithm first selects K charging hot spots into the result set S' and they will become parent charging stations. Put the unselected child charging stations corresponding to the parent charging stations into the candidate set C. Then, we update the value of $\overline{P}$ by $\overline{P} - S'$ (Lines $2 - 4$). In each iteration of network expansion, when $\overline{P} > 0$, we iterate over the result set S' to determine whether there is an overflow site. If there is an overflow site in S', we put all the child charging stations of overflow sites into the set C' and calculate their usage benefit score based on equation 2 and then put the calculated results into the list L. We traverse the list L and select the child charging stations s_{next} with the maximum usage benefit

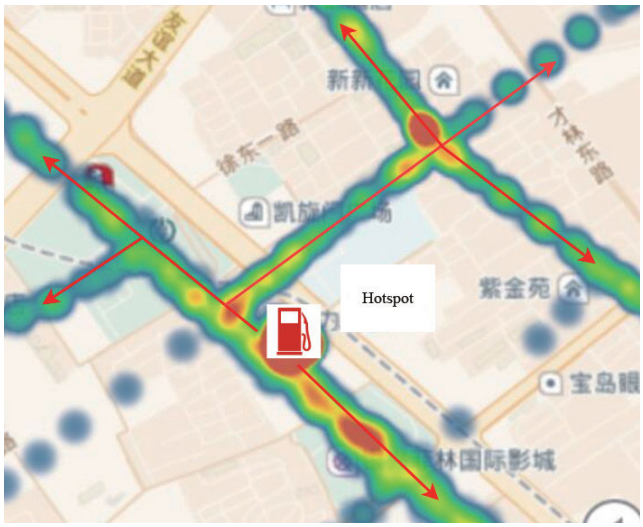


FIGURE 5. Spatial hot spot and star-shaped mobility pattern.

score and inserted it into the result set S'. Then we put the unselected child charging stations of s_{next} into the candidate set C (Lines $6-16$). As shown in lines $17-19$, we calculate the overflow status of the newly expanded optimal site s_{next}, after the users charging at the unselected sites within the scope D_{rm} of s_{next} and overflowed users of the parent station are redistributed to s_{next}. If s_{next} is no overflow, the unselected charging station within the scope D_{rm} can be overwritten by s_{next}. We insert these stations into the redundant site set I. If s_{next} overflow occurs after redistribution, it means that the charging station within the scope D_{rm} cannot be covered by s_{next}. If there are no overflow parent sites in the result set S', we directly calculate the usage benefit score of all child charging stations of candidate set C, expand the site with the maximum use benefit score to the charging station network, and insert its unselected child charging stations into candidate set C as new candidates (Line $21-23$). Lines $24-30$ are processed in the same way as lines $13-19$. Update the value of I by $(S-S') \cup I$ (line 31). When $\overline{P} == 0$, the loop is terminated and the site set S' with overflow status is returned as the optimization scheme and the redundant charging set I (Line 33).

Update the overflow status of CS. The overflow status is mainly divided into two types: (1) true overflow status. At this time, the charging station overflow is mainly caused by the overall lack of piles in the area. Such overflow is irreversible, unless the government planning department adds a new station in the area to relieve the local charging pressure. (2) pseudo overflow status. At this time, the overflow of charging station is mainly caused by the habits of EV users. The users there can completely go to other charging stations in the area for charging. This overflow is reversible and there is no need to add new stations. During the expansion process, if the parent charging station overflows and there is a child charging station within the range D_{rm} that can relieve the charging pressure, then the overflowed users transfers to the child charging station to charge and updates the overflow status of the parent charging station to be 0, otherwise to 1.

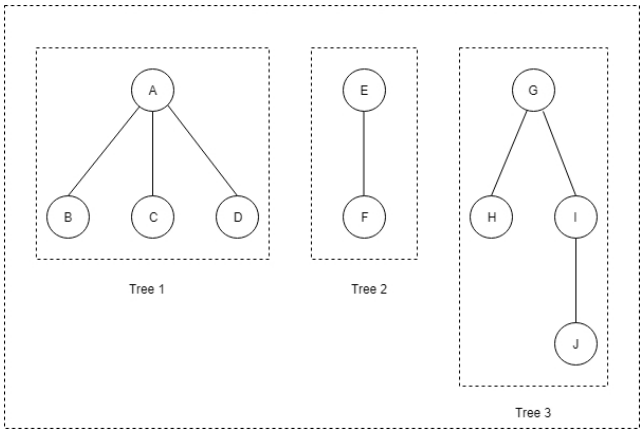


FIGURE 6. Forest structure of the charging station network.

Update the visiting rate of CS. With each expansion, the algorithm greedily selects the most usage benefit charging station s_{next} to expand the network, and updates the visiting rate of s_{next}. There are two ways to update the visiting rate of the new site s_{next}: 1) No overflow occurred at the parent charging station of the new site s_{next}. We calculate the overflow status of s_{next} when users charging at none-selected stations within the scope D_{rm} of s_{next} to charge at s_{next}. If overflow occurs at the new site s_{next}, it means that none-selected stations within the scope cannot be covered by the new site, and the visiting rate of s_{next} doesn't need to be updated. If no overflow occurs at the new site due to redistribution, it indicates that none-selected stations within the scope D_{rm} can be covered by s_{next}, and all users charging at none-selected stations within the scope will be allocated to the new site s_{next}, and then update the visiting rate of s_{next}. 2) Overflow occurred at the parent charging station of the new site s_{next}. We calculate the overflow status of s_{next} when users charging at none-selected stations within the scope D_{rm} of s_{next} and overflowed users in the peak period of the parent charging station to charge at s_{next}. If overflow occurs at the new site s_{next}, it means that the new site cannot cover none-selected stations within the scope D_{rm} and solve the overflow problem of its parent charging station, so its visiting rate will not be updated at this time. If no overflow occurs at the new site due to redistribution, it indicates that none-selected stations within the scope D_{rm} can be covered and the overflow problem of the parent charging station can be solved. At this time, the overflow of the parent charging station is regarded as pseudo overflow, and its overflow status is updated to 0. Overflowed users and the users charging at none-selected stations within the scope D_{rm} are all assigned to the new site s_{next}, and the updated visiting rate of s_{next} is also provided.

The representation of the parent-child relationship. In the optimization process of EV charging station network, in order to better demonstrate the parent-child relationship between charging stations in the optimal network and show the distribution area of parent and child charging stations, we adopted the forest structure to represent the relationship (figure 6). The forest is made up of some trees, and the

Algorithm 3 The New Network Expansion Algorithm

Input: charging station set S, charging station visits set V, matrix D, a value K and the tuning parameter γ.
Output: Optimal station set S' and redundant charging station set I.

1 //Initialization;
2 Parent station set $S' \leftarrow K$ starting spatial hot spots;
3 Candidate set $C \leftarrow$ select child charging stations of parent charging stations in S';
4 $S \leftarrow S - S'$;
5 //Network Expansion;
6 **while** $S! = \phi$ **do**
7 $s_{next} \leftarrow \phi$, $C' \leftarrow \phi$;
8 **if** *an overflow site exists in the result set* S' **then**
9 Retrieve the child stations of all overflow parent stations form C and insert them into the set C';
10 **for** $s_{ij} \in$ *candidate set* C' **do**
11 Get the usage benefit score $U(s_{ij})$ by equation (2);
12 List $L \leftarrow s_{ij}$ and its usage score;
13 Traverse L to find the maximum value of the usage benefit score and the corresponding child station s_{next};
14 $S' \leftarrow S' \cup s_{next}$;
15 $C \leftarrow C \cup$ none-selected child station of s_{next};
16 Calculate the visit of s_{next} after users charging at the none-selected stations within the scope D_{rm} of s_{next} and overflowed users of the parent station are allocated to the new site s_{next};
17 **if** s_{next} *don't overflow* **then**
18 $I \leftarrow$ child stations within the scope D_{rm} of s_{next};
19 $S \leftarrow S - s_{next} - I$;
20 **else**
21 $S \leftarrow S - s_{next}$;
22 **else**
23 **for** $s_{ij} \in$ *candidate set* C **do**
24 Get the usage benefit score $U(s_{ij})$ by equation (2);
25 List $L \leftarrow s_{ij}$ and its usage score;
26 Traverse L to find the maximum value of the usage benefit score and the corresponding child station s_{next};
27 $S' \leftarrow S' \cup s_{next}$;
28 $C \leftarrow C \cup$ none-selected child station of s_{next};
29 Calculate the visit of s_{next} after the users charging at the none-selected stations within the scope D_{rm} of s_{next};
30 **if** s_{next} *don't overflow* **then**
31 $I \leftarrow$ child stations within the scope D_{rm} of s_{next};
32 $S \leftarrow S - s_{next} - I$;
33 **else**
34 $S \leftarrow S - s_{next}$;
35 //Termination;
36 Return S', I.

root node of each tree is regarded as the charging hotspot of network expansion algorithm. The upper layer of each tree represents the parent charging station, and the lower layer is the child charging station. By analogy, the parent-child relationship between charging stations can be obtained.

B. THE NEW NETWORK EXPANSION ALGORITHM

Algorithm 2 realizes the optimization of the existing charging station network under the constraint of the number $\overline{P}$ of charging stations. To get the optimal network of the entire original charging station network, we develop the new network expansion algorithm shown in algorithm 3. The main difference between algorithm 3 and algorithm 2 is the different termination conditions for network expansion.

Algorithm Design. Algorithm 3 gives the pseudo-code of the new network expansion algorithm. In line 2, select K charging hotspots as the starting sites of the algorithm, and plug them into the network optimization set S' to become the parent charging stations. Then, we put unselected child charging stations of the parent charging station of S' into the candidate set C, and update the charging station set S by $S - S'$ (lines $3 - 4$). In each iteration, lines $7 - 16$ of algorithm 3 are processed in the same way as lines $7 - 17$ of algorithm 2. As shown in lines $17 - 21$, we calculate the

TABLE 3. Statistics of datasets.

Data sets	Period	Scales
ET Trajectory	6/5/2018-27/5/2018	Number of ET: 800 Number of GPS points: 9202012
Charging Station	2018	Number of stations: 156 Attributes (number of charging piles, area, visits, waiting time, latitude and longitude)
POIs	2018	Number of POIs: 37990 Number of categories: 20 categories
Road Network	2018	Wuhan, China (longitude:114.1192-114.4947, Latitude:30.4398-30.7289)

overflow status of the newly expanded optimal site s_{next}, after the users charging at the unselected sites within the scope D_{rm} are redistributed to s_{next}. If s_{next} is no overflow, the unselected charging station within the scope D_{rm} can be overwritten by S_{next}. We insert these stations into the redundant site set I and then update set S by removing these stations and s_{next} from S. If s_{next} overflow occurs after redistribution, it means that the charging station within the scope D_{rm} cannot be covered by s_{next}, and the update S by $S - S_{next}$. If there are no overflow sites in the result set S', we directly calculate the usage benefit score of all child charging stations of candidate set C, expand the site with the maximum use benefit score to the charging station network, and insert its unselected child charging stations into candidate set C as new candidates (Lines $23 - 29$). Lines $30 - 34$ are processed in the same way as lines $17 - 21$. When $S == \varnothing$, the loop is terminated and the optimal charging station set S' with overflow status is returned as the optimization scheme of the entire charging station network and the redundant charging set I (Line 36).

C. INITIALIZATION METHODS

It is worth noting that the selection of initialization sites has a great influence on the recommended result, so how to choose an efficient initialization method becomes the key of the network expansion algorithm. Hence, we will discuss the two available initialization methods.

Top-k based initialization. The most direct method is top-k initialization, which essentially selects charging stations (i.e., charging hotspots) where EV charging is frequent as the starting sites for the network expansion. The reason behind the method is that charging hotspots should always be included in the optimization scheme to cover more users, ensuring that the algorithm does not miss any charging stations with the highest usage benefit. However, most of the top-k ranked stations may be very close in spatial distribution, which may cause some important areas to be missed, especially when $\overline{P}$ is relatively small.

Clustering-based initialization. In order to include more spatially diverse starting sites in the initialization phase and more effective when the number of charging stations is small, we adopted the spatial clustering method to select starting sites. The intuition behind the spatial clustering-based initialization is from the observation of the charging heat map (i.e., figure 9), which visually has some cluster over the space. In this study, we adopted the clustering method based on the hierarchical clustering, because it does not need to adjust the clustering parameters, and always produces relatively stable results. Then, we select the highest-ranked charging station in each cluster as the starting site. Compared with the results generated by the top-k initialization method, the clustering-based method has significantly better diversity and user coverage when the number of charging stations $\overline{P}$ is small. The main reason is that after the spatial clustering step, the starting site is no longer a site with a high overflow rate, which can only expand with the charging station near to the site with lower visit to relieve the local charging pressure.

VII. EXPERIMENT

In this section, we conducted extensive experiments to evaluate the performance of our proposed method. This study is based on the area of about 290 square kilometers in Wuhan, which mainly includes the central area of Wuhan (like Jiangan district, Wuchang district, Qiaokou district and Hanjiang district).

A. DATASET AND SYSTEM SETTING

Table 3 summarizes the information of four data sources. The transportation data set includes the data about the station system and the road network in Wuhan, China. EVs' charging behavior has two important indicators: time spent and visiting rate. We set 1 hour as a time slot, the size of the tensor is $24 \times 156 \times 22$, where 24 is the number of hours in a day, 156 the number of charging stations, and 22 the number of days. By feeding the trajectory data of 22 days and the detected charging events into the tensor, we can obtain 6.39% non-zero entries (i.e., the entry's value $>= 1$) as showed in table 4. However, it is not reliable for a charging station to have only one charging event within a recording cell, which may be an incorrect record. Setting a high threshold to determine non-zero entries can improve the quality of the value of a single entry, but it will lead to worse data sparsity. Consider the trade-off, we set threshold $= 2$ here. Whenever a recording cell covers more than 2 detected CEs, we regard $W(i, j, k)$ as being observed and use the CEs' (who fall in this recording cell) average duration to denote its waiting time. Thus, 80786 record cells need supplementing by the inference.

TABLE 4. The sparseness of the tensor with different thresholds.

Data set	Threshold=1	Threshold=2	Threshold=3
Charging Events	6.39%	1.92%	1.03%

System Settings. Our model is executed in Python, MatLab and ArcMap 10.3. All the experiments are performed

TABLE 5. Results of Context-aware learning w.r.t. baselines.

Methods	RMSE	MAE
AWD	7.577	4.487
AWH	6.766	4.024
AWS	5.650	3.366
Kriging	5.347	5.14
TD	5.221	5.214
TD+F	3.250	3.243

in a 3.4 GHz Titan Cluster with 96 GB memory. The system allows users to interact with it using different parameters and get charging station layout suggestions in a short time.

B. EXPERIMENTS FOR EXPECTED WAITING TIME

There are total number of 82368 cells (24 hours × 156 stations × 22 days) in the recording cube, and each cell contains 0.643 charging events average, which indicated the number of cells was lack of enough detected CEs to estimate the expected waiting time. We evaluate the context-aware collaborative tensor decomposition model in the approach. We first randomly remove 30% of the non-zero entries from the tensor and use our proposed context-aware tensor co-decomposition model to fill in these entries. We then evaluated the model using the original values of these entries as the ground truth to measure inferred values.

Baseline methods. Table 5 shows the results of the performance comparison of different methods. We compare our model with four baselines: 1) AWD (Average within Day) fills a missing entry with the average expected duration of the same day timestamp. 2) AWH (Average within Hour) finds the same hour timestamp and fills in a missing entry with their average expected durations. 3) AWS (Average within Station) is similar to the previous two methods, which uses the average value within the same charging station. 4) Kriging interpolates the durations of a missing entry with the non-zero entries geospatially nearby. We also study the contribution of matrix F in helping supplement the missing entries. To show the effectiveness of the proposed model, we use two metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), where n is the number of instances; y_i is the ground truth and $\widehat{y_i}$ is an inference.

$$RMSE = \sqrt{\frac{\sum_i (y_i - \widehat{y_i})^2}{n}} \tag{10}$$

$$MAE = \frac{\sum_i |y_i - \widehat{y_i}|}{n} \tag{11}$$

C. EXPERIMENTS FOR ARRIVAL RATE CALCULATE

In this subsection, we discuss the experiment with the calculation of the recording cells' arrival rate, then we sorted by the arrival rate of the stations and found the charging hotspots of EVs. During the experiment, we set up a queue system for each charging station in the cell, and the number of servers in the queue is expressed as c_i. We selected the top 800 shortest durations among all the detected CEs to estimate $\frac{1}{\mu}$, and

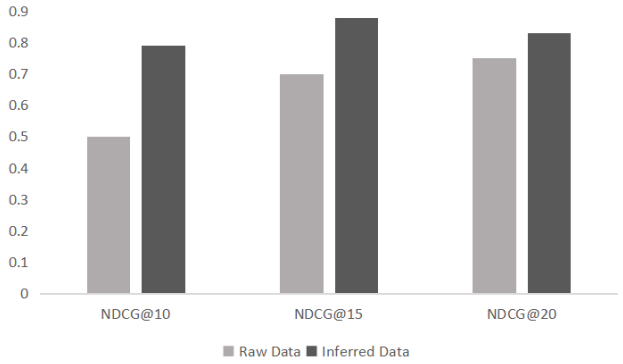


FIGURE 7. Performance of in-the-field study.

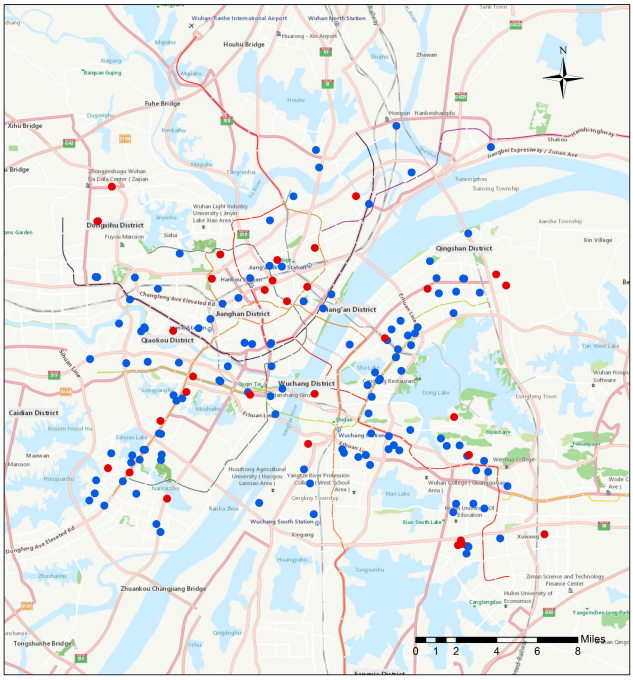


FIGURE 8. Stations' distribution of the central area of Wuhan, China.

finally obtain $\mu = 1.42$ hours. We make use of each cell's expected duration to estimate W_s (time spent), and equation 9 is used to calculate the arrival rate of each charging station in a certain period. To better evaluate the performance of the queue system, we perform an in-the-field study in 30 charging stations in Wuhan, collecting the real arrival rate. We then rank these charging stations in terms of the real arrival rate and the inferred values respectively, measuring the closeness of the two ranks using NDCG (Normalized Discounted cumulative gain). Figure 7 presents the performance of the NDCG evaluation approach, where our method outperforms the method only using ET trajectory data. The higher NDCG is better ranking performance is. Experimental results verify the performance of the queue system to predict the arrival rate of different charging stations in the same period.

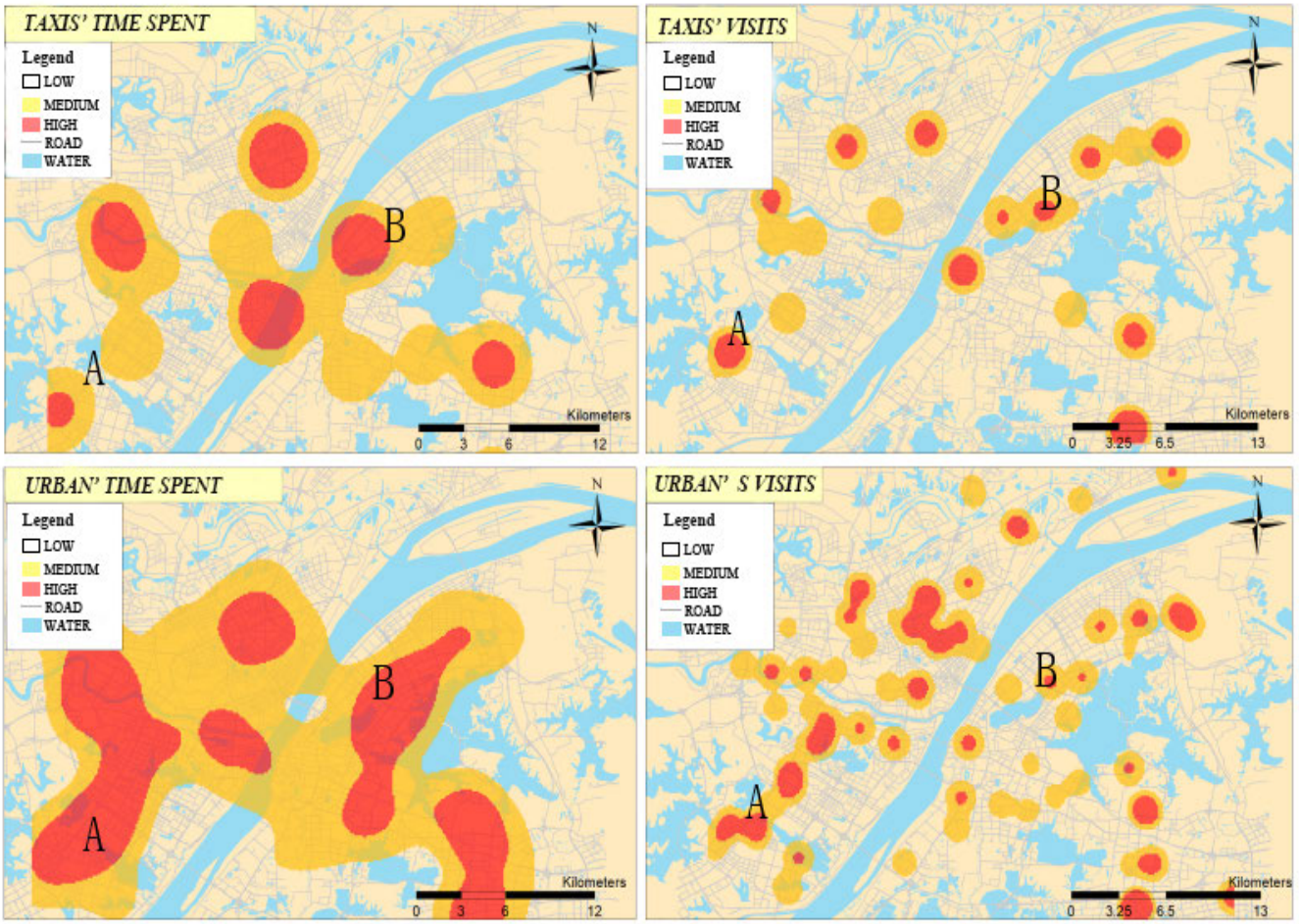
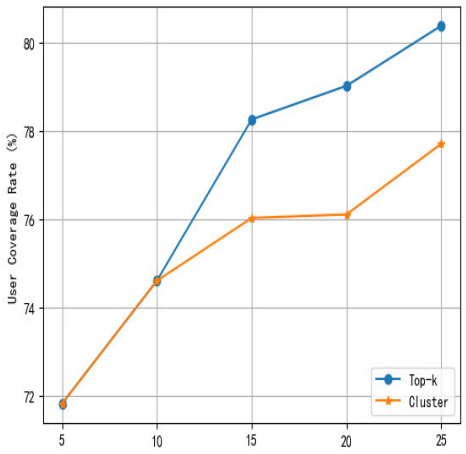


FIGURE 9. Spatial distribution of charging behavior of electric taxis and electric vehicles.

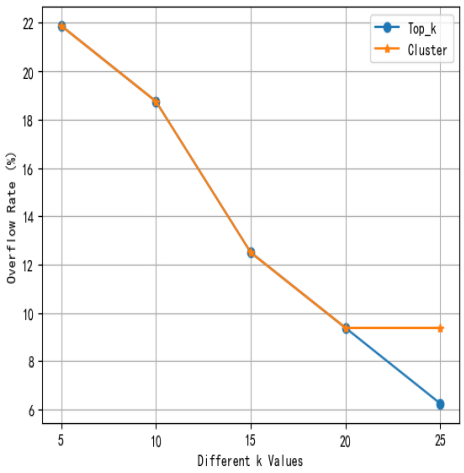
In the previous experiment, we obtained expected time spent (expected waiting time) and arrival rate of different stations. These information reveals ET drivers charging behavior and at the same time shows the whole city's charging behavior from spatial prospective. For the entire city's charging behavior, we combined all the corresponding cells according to the dimension of charging station direction and used their average waiting time and average visiting rate to represent the time spent and visits of the corresponding charging stations. Figure 8 depicts the spatial distribution of charging stations in the central area of Wuhan city, and the locations of charging stations are shown in red and blue solid circle (156 charging stations). The red solid circle indicates that the charging station overflow during the peak period, while the blue circle indicates that the charging station does not overflow. Figure 9 the upper side presents the spatial distribution of ETs time spent and the distribution of their visits. Redder color refers to longer time spent or higher visit. The low side describes the time spent and the visiting rate of EVs in the study area. Most areas frequented by EVs have been granted longer time pent. However, although taxi drivers rarely visit the charging stations in area A, they still need a long time spent, which means that there are many other types of vehicles charging here. According to our survey, charging stations in area A are close to urban highways, so many private EVs choose to charge here. ETs are often charged in prosperous areas. For instance, charging hotspots area B is near to recreational areas where have a great advantage in attracting taxis. Many ETs serve customers in these places, and they are likely to charge at nearby charging stations. At last, we sorted the visits of all urban charging stations in descending order and selected the top K charging stations that ranked the front as the spatial hotspots for EV charging in the area.

D. EXPERIMENTS FOR CS LAYOUT NETWORK OPTIMIZATION

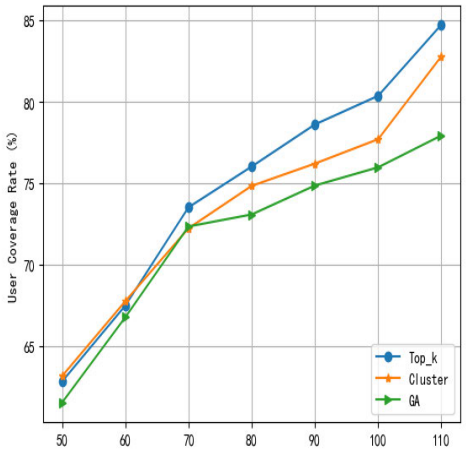
In this subsection, we study the effects of different parameters in our optimization system. Note that we didn't compare the effectiveness of other related works ([5], [28]) because of the differences in constraints and optimization target. For example, an optimal solution of the charging infrastructure is derived based on mean trip times of electric vehicles [14].



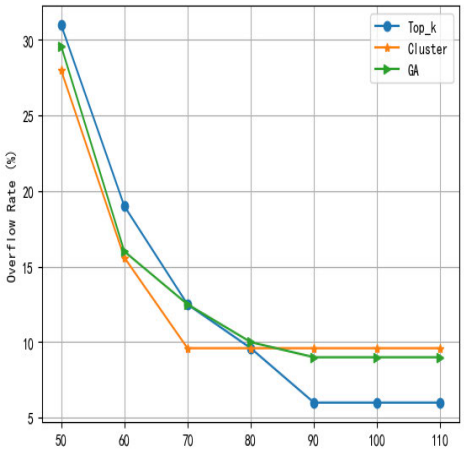
(a) User coverage rate for different K values



(b) Overflow rate for different K values



(c) User coverage rate for different $\overline{P}$ values



(d) Overflow rate for different $\overline{P}$ values

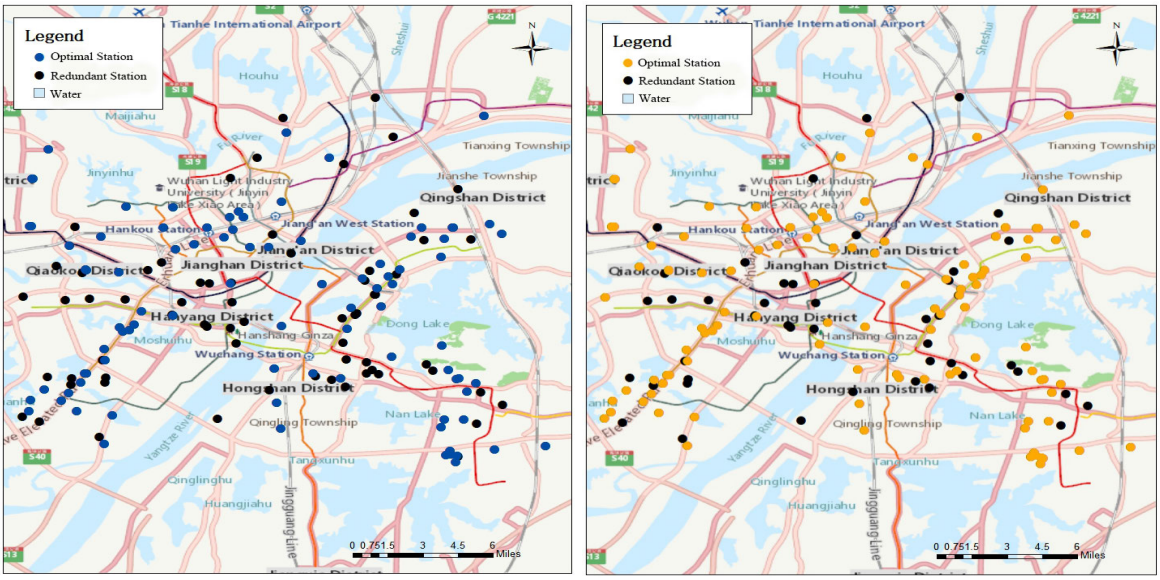
FIGURE 10. **Effectiveness evaluation.**

1) OPTIMIZATION RESULTS OF CHARGING STATION QUANTITY CONSTRAINT

In order to evaluate the effectiveness of our method, we did a lot of experiments to study the impact of three factors to the optimization results, including the different number of station $\overline{P}$, the different number of charging hot spots K and the value of γ. Unless mentioned explicitly, the default parameters used in the following experiments are $\overline{P} = 100$, $K = 25$, and $\gamma = 1.03$. For the sake of solution algorithms comparison, we also apply the genetic algorithm (GA) to solve the optimal problem. Before using GA to optimize the CS network, the parameter setting method of Coy *et al.* [43] was used to conduct indepth experiments to determine the parameter values of the genetic algorithm, such as population size p, selection rate a, mutation rate b.

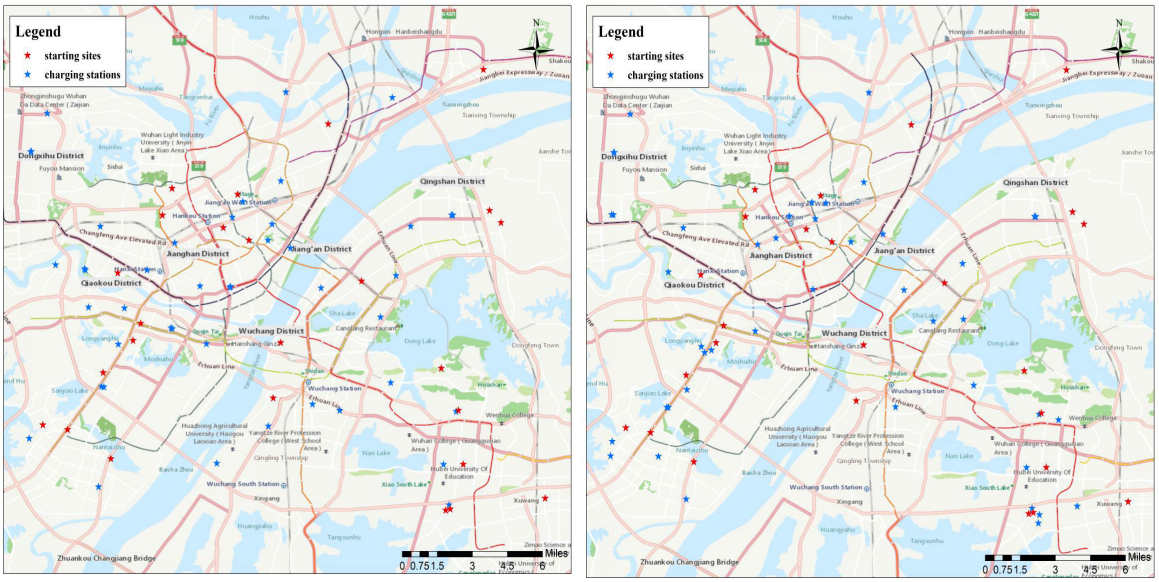
Different K Values. Figure 10 (a,b) illustrate the total user coverage rate of choosing different numbers of starting sites (i.e., K value). As a result, we make the following observation: 1) In most cases, the top-K initialization method gets a higher user coverage rate, and also gets a lower overflow rate. 2) when K is in the range of 5 to 20, the overflow rate of the two initialization methods remains the same, but the user coverage rate of top-k is much larger than the clustering-based initialization method, up to 3%. 3) when K value is small, two methods are similar. This is because in these cases the starting charging stations of top-k results are the same as clustering.

Different number of stations ($\overline{P}$). Figure 10 (c,d) illustrate the total user coverage and overflow rate with the different number of stations $\overline{P}$, from 50 stations to 110 stations. From these two figures, we can see that the proposed optimal algorithms outperform the GA approach for both performance metrics. And then, we make the following observations: 1) the top-k initialization method performs better when the optimal number of stations $\overline{P}$ is larger. 2) when the value $\overline{P}$ is small clustering-based method is better than the top-k



(a) top-25, $\overline{P} = 90$ (b) top-25, $\overline{P} = 110$

FIGURE 11. Effect of $\overline{P}$ values.



(a) $\gamma = 1, \overline{P} = 70$ (b) $\gamma = 1.03, \overline{P} = 70$

FIGURE 12. Effects of γ values.

method. This because, when the number of stations is small, the best strategy is to ensure that users in every major area have a charging station to charge (essentially the intuition of clustering-based). 3) As can be seen from figure 10(d), the overflow rate of the top-k method remains unchanged within the range of 90-110, and the user coverage rate continues to rise. Figure 11 shows the optimized layout results of $\overline{P} = 90$ (blue solid circle) and $\overline{P} = 110$ (yellow solid circle), in which the black solid circle represents the distribution of redundant charging stations under different conditions. As can be seen from figure 10(d), under the condition of $\overline{P} > 90$, the overflow rate is basically unchanged. In order to continue to optimize the charging station network, we can only improve the user coverage rate and cover more continuous user travel routes by selecting charging stations with the higher visit.

Different γ values. Figure 12 provides the results with different γ settings, with the top-k method, where the red stars is the starting sites and the blue stars is the recommended network layout of charging stations. When $\gamma = 1$, intuitively,

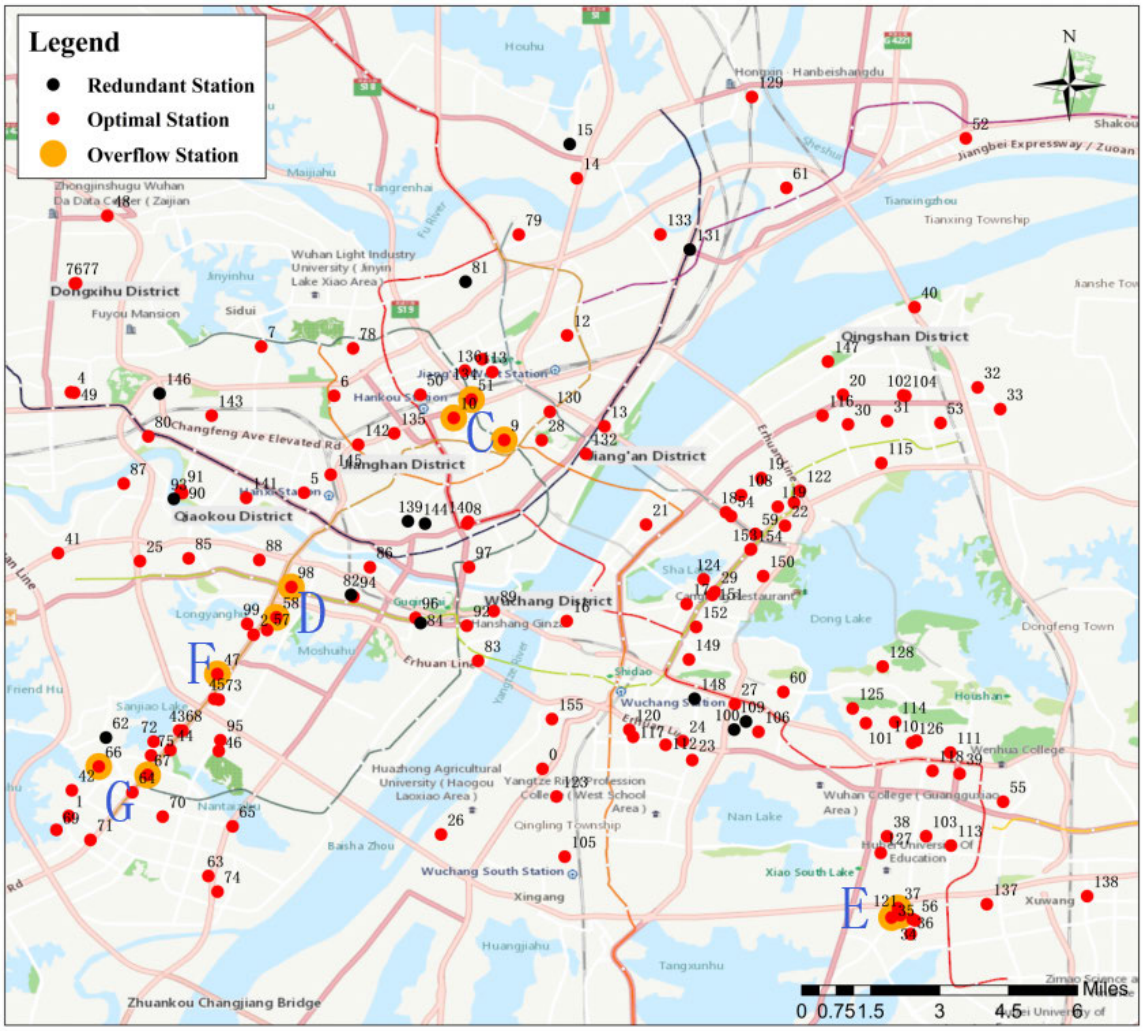


FIGURE 13. The optimal charging station network for Wuhan central region.

it is that the result of an expansion goes further always on heavily visited charging stations, they don't take into account overflow problems of charging stations and users' continuous trip coverage. When $\gamma > 1$, if the charging station is in an overflow status, the algorithm will give priority to the more popular charging stations near to the overflow charging station to relieve the local charging pressure and achieve more user coverage. If the overflow status of the current charging station has been basically stable without a downward trend, the higher usage benefit score is given for the station covering more users and more continuous user trips.

2) RESULT OF THE OPTIMAL CHARGING STATION NETWORK

There are 156 charging stations in the original station network of Wuhan central region, the average utilization rate of charging station is about 61.04%. The original network has 32 overflow sites. The utilization rate of charging station in one day (24 hours) is an intuitive indicator to measure the usage of the entire charging station network. The calculation formula of the average utilization rate of charging stations is as follows:

$$AU = \frac{\sum_{j=1}^{k} CT(s_j)}{24} \qquad (12)$$

where $\sum_{j=1}^{k} CT(s_j)$ is the total charging duration of all stations in one day. Num(S) is the total number of charging stations in the network.

The optimal network of charging stations is shown in figure 13, where the number represents the charging station ID. Table 6 shows the experimental results of the optimized network, in which the utilization rate of the charging station is 7.72% higher than the previous one, and the user coverage rate is as high as 97.63%. Compared with the original network, there are 21 fewer overflow sites in the optimized network. The reason is that some areas are just caused by the user's charging habit, which can be alleviated by other charging stations in the area, and the overflow of these 21 charging stations is a pseudo overflow. However, in the case of regional

TABLE 6. Result of the optimal network.

Original Network		Optimal Network	
Number of charging stations	156	Number of charging stations	143
Average utilization of charging stations	61.04%	Average utilization of charging stations	68.76%
-	-	Number of redundant sites	13
Number of overflow sites	32	Number of overflow sites	11
-	-	User coverage rate	97.63%

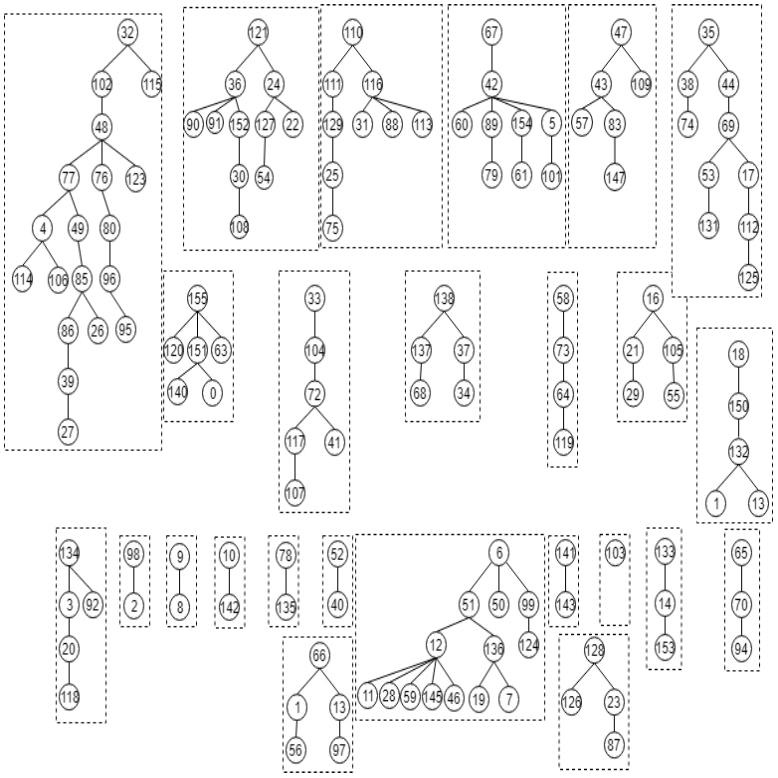
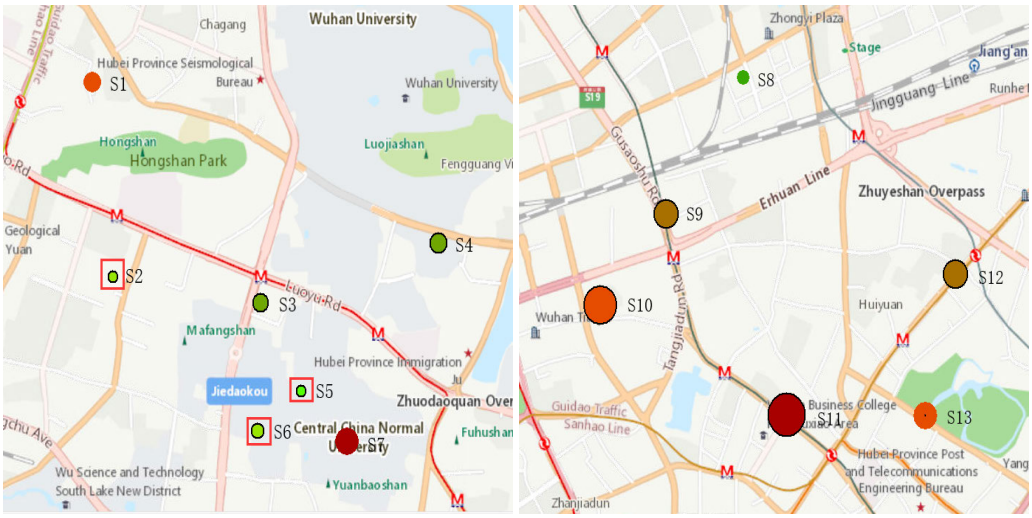


FIGURE 14. The parent-child schematic diagram for the optimal network.

overall charging pile shortage, the government could only relieve the pressure of local charging by adding new sites in the region, and the overflow was irreversible. The 11 overflow charging stations in the optimized network are distributed in the region overall charging pile shortage. It can be seen from the optimization results that there are charging station congestion problems (occurring in zone C, zone D, zone E, zone F and zone G) and charging station redundancy problem (occurring in 13 stations, i.e., site 15, site 62, site 81, site 82, site 84, site 93, site 100, site 109, site 131, site 139, site 144, site 146, and site 148) in the downtown area of Wuhan. The government could add new stations in areas with congestion to meet local EVs' charging demands. Then remove redundant sites to reduce waste of resources, thus improving the utilization of charging stations. Figure 14 shows the parent-child relationship between charging stations in the optimal network. The upper layer represents the parent charging station and the lower layer represents the child charging station. The root nodes of trees (i.e., 32, 121, 110, 67, 47, 35, 155, 33, 138, 58, 16, 18, 134, 98, 9, 10, 78, 52, 6, 141, 103, 133, 65, 66 and 128) regard as the starting sites of the expansion algorithm, and expand them as individual trees. The starting stations are usually located in the area frequently visited by EVs. The algorithm regards them as the parent charging stations and calculates the usage benefit score of other charging stations not selected into the optimized network (calculated by formula 2) according to parent stations situations (overflow status), and greedily selects the optimal charging station to expand the network. For example, if there is an overflow station in the parent charging station, the algorithm will give priority to the overflow problem, and expand the popular child charging station near to the overflow parent station as much as possible to alleviate the local charging pressure. Until all overflow sites are traversed, the more popular charging stations, which are further from the parent stations, are expanded to cover more users and more



(a) Optimization results of demand surplus

(b) Optimization results of lack of demand

FIGURE 15. Optimization results of different situations.

continuous travel routes. The algorithm continuously selects the optimal charging station in a cycle until it has traversed all charging stations, and returns the current network optimal scheme and the overflow charging stations to guide the government planning department in the next step of charging station planning.

Case analysis. Different from the traditional charging station optimal location problem that decides whether to build the station in each candidate location, our problem further considers the overflow problem of charging station in the peak period, and at the same time considers more user coverage and more continuous travel routes for the overall network optimization, which is based on the actual charging station layout. To understand the effectiveness of our charging station planning recommendations, we carried out case analysis on the demand surplus (figure 15(a)) and the lack of demand zone C (figure 15(b)), as these two areas appeared in our recommendation, regardless of parameters.

Figure 15 shows the optimization results of these two cases, where the sites in the red rectangle are redundant stations, which will be removed during the network optimization of charging stations. Intuitively, the area shown in figure 15(a) is very dense with charging stations, although there is rarely visited by drivers (colored green). There are many redundant sites (i.e., $S2$, $S5$, and $S6$) in this area, which is the reason for the low utilization rate of the original charging station network. We use the top-k initialization method to expand the network. First, two initial sites are selected (i.e., $S1$ and $S7$), and all unselected sites are put into the candidate set as their child charging stations. In the network expansion stage, we calculated the usage benefit scores of the child charging stations of $S1$ and $S7$ respectively according to equation 2 (as shown in 16) and selected the optimal site with the highest scores to expand the network, which is $S4$. Then put the remaining unselected charging stations

S1's child stations	S2	S3	S4	S5	S6
Usage benefit score	603.54	999.45	9883.36	700.32	760.89

S7's child stations	S2	S3	S4	S5	S6
Usage benefit score	340.69	513.78	720.57	11.57	43.93

FIGURE 16. Greedy network expansion example.

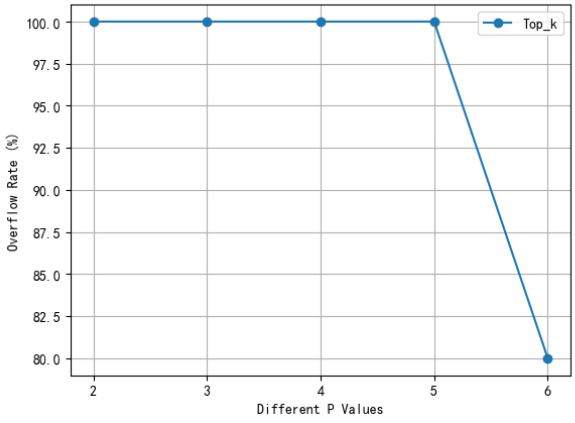


FIGURE 17. At different number $\overline{P}$ of stations, the overflow rate of the case with a lack of demand.

(i.e., $S2$, $S3$, $S5$, and $S6$) as the child of $S4$ into the candidate set as new candidates and recalculate their usage benefit scores. The algorithm terminates when $\overline{P}$ is equal to 4 (because there are no redundant sites in this area), the optimization scheme is obtained ($S1$, $S3$, $S4$, $S7$) and the redundant sites are $S2$ (i.e., site 148), $S5$ (i.e., site 109), and $S6$ (i.e., site 100).

With a similar algorithm operation, we can get the optimization result of lack of demand, as shown in figure 15(b). There are no redundant sites in the *C* zone with overall charging pile shortage. As can be seen from figure 17, the overflow rate of the *C* zone is very high, and it is not until the network expands to the last site *S*8 in this region that the overflow rate has a downward trend. The government may consider adding some new charging stations in area *C* to ease the local charging pressure. As a result, based on our analysis, we conclude that the current network layout of charging stations is not reasonable enough. With the real EV track and data-driven analysis, we suggest that some sites should be built in areas with high charging pressure to relieve the charging pressure and reduce unnecessary charging stations to avoid the waste of resources. The experimental results show that our optimization algorithm can effectively eliminate redundant charging stations and improve the utilization rate of charging stations, identify the congestion areas of the network, and give suggestions to the government for further improvement of charging station planning.

VIII. CONCLUSION

In this paper, we propose a data-driven approach to optimize the layout of the existing charging stations network. Our method can address the optimization problem of charging station layout in a more realistic way, considering the constraints and requirements from the government planning department's perspective: 1) charging station quantity constraint, 2) overflow, and 3)utilization. We also present a framework for discovering urban charging behavior regularity using four data sources: ET trajectory dataset, POIs, station data and road network data. Our work currently relies solely on the trajectory of the electric taxi, however, other vehicles can be seamlessly integrated into our system. During the charging events detection phase, we actually only rely on the detected CEs' time spent, which is the result of charging event detection and is independent on whether this vehicle is an ET or not. We model citywide charging behavior with a 3D tensor, filling in the missing entries of times and stations with sparse data using a context-aware tensor collaborative decomposition approach to estimate the time spent at station, and develop queue systems to calculate the visits among different stations. The two indexes not only reflect EV charging regularity but also the popularity of stations.

With the research of electric vehicle charging behavior characteristics of the city, the relationship between stations and geographical location, we can develop a flexible usage benefit score function to adjust and measure the usability between charging stations. In the course of our research, we found two extreme phenomena: 1) lack of demand, and 2) demand surplus. These phenomena show that in the actual charging station layout, there is an imbalance phenomenon for the charging stations, that is, supply exceeds demand and demand exceeds supply. Therefore, We propose a heuristic network expansion algorithm based on spatial charging hot spots to address the network optimization problem. The experimental results show that our method can effectively remove the redundant charging stations and identify the congestion areas of the charging stations. Therefore, our next research work is how to add a new charging station in the case of regional overall charging pile shortage.

REFERENCES

[1] L. Zhe and M. Ouyang, "The pricing of charging for electric vehicles in China—Dilemma and solution," *Energy*, vol. 36, no. 9, pp. 5765–5778, 2011.

[2] C. E. Hatton, S. K. Beella, J. C. Brezet, and Y. C. Wijnia, "Charging stations for urban settings the design of a product platform for electric vehicle infrastructure in Dutch cities," *World Electr. Vehicle J.*, vol. 3, no. 1, pp. 134–146, 2009.

[3] T. Sweda and D. Klabjan, "An agent-based decision support system for electric vehicle charging infrastructure deployment," in *Proc. IEEE Vehicle Power Propuls. Conf.*, Sep. 2011, pp. 1–5.

[4] H. Liu and D. Z. W. Wang, "Locating multiple types of charging facilities for battery electric vehicles," *Transp. Res. B, Methodol.*, vol. 103, pp. 30–55, Sep. 2017.

[5] W. Tu, Q. Li, Z. Fang, S.-L. Shaw, B. Zhou, and X. Chang, "Optimizing the locations of electric taxi charging stations: A spatial–temporal demand coverage approach," *Transp. Res. C, Emerg. Technol.*, vol. 65, pp. 172–189, Apr. 2016.

[6] S. Davidov and M. Pantoš, "Planning of electric vehicle infrastructure based on charging reliability and quality of service," *Energy*, vol. 118, pp. 1156–1167, Jan. 2017.

[7] A. Awasthi, K. Venkitusamy, S. Padmanaban, R. Selvamuthukumaran, F. Blaabjerg, and A. K. Singh, "Optimal planning of electric vehicle charging station at the distribution system using hybrid optimization algorithm," *Energy*, vol. 133, pp. 70–78, Aug. 2017.

[8] J. Ma and L. Zhang, "A deploying method for predicting the size and optimizing the location of an electric vehicle charging stations," *Information*, vol. 9, no. 7, p. 170, 2018.

[9] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "Optimal planning of PEV charging station with single output multiple cables charging spots," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2119–2128, Sep. 2017.

[10] L. Luo, W. Gu, S. Zhou, H. Huang, S. Gao, J. Han, Z. Wu, and X. Dou, "Optimal planning of electric vehicle charging stations comprising multi-types of charging facilities," *Appl. Energy*, vol. 226, pp. 1087–1099, Sep. 2018.

[11] J. A. Domínguez-Navarro, R. Dufo-López, J. M. Yusta-Loyo, J. S. Artal-Sevil, and J. L. Bernal-Agustín, "Design of an electric vehicle fast-charging station with integration of renewable energy and storage systems," *Int. J. Electr. Power Energy Syst.*, vol. 105, pp. 46–58, Feb. 2019.

[12] S. R. Gampa, K. Jasthi, P. Goli, D. Das, and R. C. Bansal, "Grasshopper optimization algorithm based two stage fuzzy multiobjective approach for optimum sizing and placement of distributed generations, shunt capacitors and electric vehicle charging stations," *J. Energy Storage*, vol. 27, Feb. 2020, Art. no. 101117.

[13] Y. Dong, S. Qian, J. Liu, L. Zhang, and K. Zhang, "Optimal placement of charging stations for electric taxis in urban area with profit maximization," in *Proc. 17th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. (SNPD)*, May 2016, pp. 177–182.

[14] A. Hess, F. Malandrino, M. B. Reinhardt, C. Casetti, K. A. Hummel, and J. M. Barceló-Ordinas, "Optimal deployment of charging stations for electric vehicular networks," in *Proc. 1st Workshop Urban Netw. (UrbaNe)*, 2012, pp. 1–6.

[15] J. Li, X. Sun, Q. Liu, W. Zheng, H. Liu, and J. A. Stankovic, "Planning electric vehicle charging stations based on user charging behavior," in *Proc. IEEE/ACM 3rd Int. Conf. Internet Things Design Implement. (IoTDI)*, Apr. 2018, pp. 225–236.

[16] S. Ge, L. Zhu, L. Hong, L. Teng, and L. Chang, "Optimal deployment of electric vehicle charging stations on the highway based on dynamic traffic simulation," *Trans. China Electrotech. Soc.*, vol. 33, no. 3, pp. 91–101, 2018.

[17] Z. Yu, C. Licia, W. Ouri, and Y. Hai, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Sep. 2014.

[18] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou, "Detecting urban black holes based on human mobility data," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2015, pp. 1–10.
[19] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2015, pp. 1–10.
[20] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2012, pp. 186–194.
[21] M. M. Islam, H. Shareef, and A. Mohamed, "Optimal location and sizing of fast charging stations for electric vehicles by incorporating traffic and power networks," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 947–957, Oct. 2018.
[22] A. A. Kadri, R. Perrouault, M. K. Boujelben, and C. Gicquel, "A multi-stage stochastic integer programming approach for locating electric vehicle charging stations," *Comput. Oper. Res.*, vol. 117, May 2020, Art. no. 104888.
[23] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "An integrated planning framework for different types of PEV charging facilities in urban area," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2273–2284, Sep. 2016.
[24] H. Zhang, W. Tang, Z. Hu, Y. Song, Z. Xu, and L. Wang, "A method for forecasting the spatial and temporal distribution of PEV charging load," in *Proc. IEEE PES Gen. Meeting Conf. Exposit.*, Jul. 2014, pp. 1–5.
[25] L. Pan, E. Yao, Y. Yang, and R. Zhang, "A location model for electric vehicle (EV) public charging stations based on drivers' existing activities," *Sustain. Cities Soc.*, vol. 59, Aug. 2020, Art. no. 102192.
[26] Y. Ahn, C. Jun, and H. Yeo, "Analysing driving patterns of electric taxi based on the location of charging station in urban area," in *Proc. IEEE Int. Smart Cities Conf. (ISC)*, Sep. 2016, pp. 1–6.
[27] C. Luo, Y. Huang, and V. Gupta, "A consumer behavior based approach to multi-stage EV charging station placement," *CoRR*, vol. abs/1801.02135, pp. 1–6, Oct. 2018.
[28] Z. Yi and P. H. Bauer, "Optimization models for placement of an energy-aware electric vehicle charging infrastructure," *Transp. Res. E, Logistics Transp. Rev.*, vol. 91, pp. 227–244, Jul. 2016.
[29] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transp. Res. C, Emerg. Technol.*, vol. 98, pp. 73–84, Jan. 2019.
[30] K. Tang, S. Chen, Z. Liu, and A. J. Khattak, "A tensor-based Bayesian probabilistic model for citywide personalized travel time estimation," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 260–280, May 2018.
[31] K. Takeuchi, H. Kashima, and N. Ueda, "Autoregressive tensor factorization for spatio-temporal predictions," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1105–1110.
[32] Y. Wang, Y. Zhang, X. Piao, H. Liu, and K. Zhang, "Traffic data reconstruction via adaptive spatial-temporal correlations," *IEEE Trans. Intell. Transport. Syst.*, vol. 20, no. 4, pp. 1531–1543, Apr. 2019.
[33] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing new york city's noises with ubiquitous data," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. (UbiComp Adjunct)*, Seattle, WA, USA, Sep. 2014, pp. 715–725.
[34] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2014, pp. 25–34.
[35] K. Tang, S. Chen, and A. J. Khattak, "Personalized travel time estimation for urban road networks: A tensor-based context-aware approach," *Expert Syst. Appl.*, vol. 103, pp. 118–132, Aug. 2018.
[36] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering," in *Proc. 4th ACM Conf. Recommender Syst. (RecSys)*, Barcelona, Spain, Sep. 2010, pp. 79–86.
[37] F. Zhang, D. Wilkie, Z. Yu, and X. Xing, "Sensing the pulse of urban refueling behavior," in *Proc. 2013 ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2013, pp. 13–22.
[38] P. Jensen, "Network-based predictions of retail store commercial categories and optimal locations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, Sep. 2006, Art. no. 035101.
[39] D. L. Huff, "Defining and estimating a trading area," *J. Marketing*, vol. 28, no. 3, pp. 34–38, Jul. 1964.
[40] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
[41] S. Bae and A. Kwasinski, "Spatial and temporal model of electric vehicle charging demand," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 394–403, Mar. 2012.
[42] J. Bao, C.-Y. Chow, M. F. Mokbel, and W.-S. Ku, "Efficient evaluation of k-range nearest neighbor queries in road networks," in *Proc. 11th Int. Conf. Mobile Data Manage.*, Kanas City, MO, USA, May 2010, pp. 115–124.
[43] S. P. Coy, B. L. Golden, G. C. Runger, and E. A. Wasil, "Using experimental design to find effective parameter settings for heuristics," *J. Heuristics*, vol. 7, no. 1, pp. 77–97, 2001.

YU YANG was born in 1994. She received the master's degree from the School of Electronic and Information Engineering, Liaoning Technical University, China. Her research interests include urban computing and data mining.

YONGKU ZHANG was born in 1974. He received the master's degree from Liaoning Technical University. His research interests include database systems and data mining.

XIANGFU MENG was born in 1981. He received the Ph.D. degree from Northeastern University, China, in 2010. He is currently a Full Professor and the Ph.D. Supervisor with Liaoning Technical University, China. His research interests include spatial data management, ecommender systems, and Web database query.