Transportation Research Part D 106 (2022) 103264

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Electric vehicle demand estimation and charging station allocation using urban informatics

Zhiyan Yi [a], Xiaoyue Cathy Liu [b,*], Ran Wei [c]

[a] Department of Civil & Environmental Engineering, University of Utah, 110 Central Campus Dr. RM 1650, Salt Lake City, UT 84112, United States
[b] Department of Civil & Environmental Engineering, University of Utah, 110 Central Campus Dr. RM 2137, Salt Lake City, UT 84112, United States
[c] School of Public Policy, University of California, Riverside, United States

ARTICLE INFO

Keywords:
Electric vehicles
PageRank model
Charging infrastructure optimization
Spatiotemporal travel patterns

ABSTRACT

This paper performs a novel data-driven approach to optimize electric vehicle (EV) public charging. We translate the study area into a directed graph by partitioning it into discrete grids. A modified geographical PageRank (MGPR) model is developed to estimate EV charging demand, built upon trip origin–destination (OD) and social dimension features, and validated against real-world charging data. The results are fed into the capacitated maximal coverage location problem (CMCLP) model to optimize the spatial layout of public charging stations by maximizing their utilization. It is shown that MGPR can effectively quantify the EV charging demand with satisfactory accuracy. Optimized EV charging stations based on the CMCLP model can remedy the spatial mismatch between the EV demand and the existing charging station allocations. The developed methodological framework is highly generalizable and can be extended to other regions for EV charging demand estimation and optimal charging infrastructure siting.

1. Introduction

With the incentives and policy support from governmental agencies and Electric Vehicle (EV) manufactures, EV markets are progressively growing in the recent decade (Huo et al., 2015). EV sales reached more than 2 million units globally in 2018 with an increase of 63% on a year-on-year basis (McKinsey, 2019). China and the United States are the two major EV markets accounting for over 20% of total sales worldwide (Gersdorf et al., 2020). San Jose, San Francisco, and Los Angeles metropolitan areas, have some of the highest EV sales and market shares, where there are already more than a quarter-million EVs on the roads (Hall and Lutsey, 2017). The rapid growth of EV adoption results in the increase of charging demand as well. EV charging events can be mainly classified into home charging, workplace charging, and public charging depending on the locations (Grote et al., 2019). In 2010, ECOtality and Idaho National Laboratory conducted EV charging units analysis (Smart and Schey, 2012). According to their survey of 2,903 private EV owners in the United States, 80% of charging events occurred at the participants' home, and over 70% of the vehicles had charged at locations other than home, such as shopping malls, restaurants, and work offices. Therefore, effectively satisfying EV users' public charging needs is crucial. In fact, a recent study shows that public charging infrastructure is key to the growth of EV market (Hall and Lutsey, 2017). Another reason for promoting public charging infrastructure is that although large portion of EV adopters in the United States and Europe have home charging facilities, countries like China still have low penetration of home charging due to fewer single-family dwellings. It is estimated that in China public charging (as compared against home charging option) will increase from 55% to

* Corresponding author.
E-mail addresses: zhiyan.yi@utah.edu (Z. Yi), cathy.liu@utah.edu (X.C. Liu), ran.wei@ucr.edu (R. Wei).

https://doi.org/10.1016/j.trd.2022.103264

Available online 1 April 2022
1361-9209/

80% by 2030 (Engel et al., 2018). Yet, two major barriers exist for most countries in terms of public charging infrastructure expansion. First, some ill-chosen charging stations can be significantly underutilized due to their inconvenience of access. Distance plays an important role for EV users when choosing charging stations because of range anxiety (Xu, Yang and Wang, 2020). Second, insufficient charging networks in certain regions could fail to meet the charging demand (Csiszár et al., 2019). To this end, how to optimally allocate the public charging stations to improve the charging coverage and effectively exploit their utilization are the main challenges for siting public electric vehicle supply equipment (EVSE).

Solving EVSE allocation problem generally involves two steps: estimate spatial distribution of charging demand; and apply mathematical modeling to obtain optimal locations with specific objectives. When estimating charging demand, a common approach is to model the EV battery usage, and simulate EVs' energy consumption under different traffic scenarios (Qian et al., 2010; Leou et al., 2014; Chaudhari et al., 2019). Such simulation-based approaches would be difficult to scale to large urban context due to computational expense. In remedy to that, data-driven methodologies have gained more and more interest in recent years with the proliferation of urban informatics and mobility data. For instance, geographical features including point of interest (POI), traffic flow, and population density were collected as inputs to spatial statistical models to infer public charging demand (Wagner et al., 2013; Dong et al., 2019; Yi et al., 2021). Traffic data such as GPS trajectory are available for exploring driver' travel patterns and further identifying hot spots for public charging. The general procedure for conducting such type of analysis involves formulating it as a discrete problem by partitioning the study area into sub-regions (or cells), extracting drivers' travel patterns from travel mobility data, and finally inferring public charging demand for each cell (Vazifeh et al, 2019; Kontou et al., 2019; Dong et al., 2019). Existing studies on EV charging demand estimation highlighted three major gaps to fill. First, previous research mostly applied geospatial statistical models to quantify charging demand (Wagner et al., 2013; Tu et al., 2016; Dong et al., 2019). Such models assume that charging demand varies by geographical proximity – demand is likely to be higher if the neighboring region has higher demand. This assumption could be biased since charging events usually occur at the trips' origins or destinations (ODs), and the distance in between may not be necessarily close. Therefore, it would be far-fetched to measure the charging demand variation by Haversine distance. To address this, we propose to translate the study area into a directed graph linked by ODs, and then apply graph theory to model the demand linkage. PageRank is a website ranking algorithm that quantifies the importance of a web page via its linkage to other pages on the internet graph. It is uniquely suited for this problem in that travel behavior to a certain extent resembles people's internet browsing process (treated as a random walk). Locations with high charging demand and their intrinsic correlation can be studied by PageRank algorithm. Second, very few studies have attempted to integrate geographical features with GPS trajectory together for model construction due to the distinction between these two types of data (static vs. dynamic). We therefore develop a modified geographical PageRank (MGPR), which is capable of incorporating geographical features to estimate the EV charging demand within a region. Third, a common issue with the existing charging demand estimation is the lack of real-world data to either construct or validate proposed models (Wagner et al., 2013; Dong et al., 2019). Public charging infrastructures are typically managed by governmental agencies or private companies, and the charging event data are usually not publicly accessible. In this project, we propose a novel approach to obtain the real-world public charging data that can be generalizable to other sites in the future. In sum, the main contributions of this study are as follows:

- We develop an innovative approach leveraging PageRank algorithm, graph theory, geographical features, and trajectory data to estimate the spatial distribution of EV charging demand. The proposed model brings insights on how driving patterns and urban informatics simultaneously influence charging demand on the urban scale;
- The proposed MGPR is validated using real-world public charging data. To overcome the challenges of data availability, we build an automated crawling pipeline using Amazon Web Services (AWS). Such pipeline allows for the dynamic fetch of public charging information via multiple cloud desktops, and automatically stores it into the local database. Moreover, the entire framework can be easily transferable to other regions where public charging station information is available on Google Map; and
- The estimated EV charging demand is then fused into a capacitated maximal coverage location problem (CMCLP) model to optimize the EVSE distribution by maximizing the utilization of charging stations. The framework is beneficial to transportation agencies and could provide insightful guidance for future public EVSE installation.

2. Literature review

2.1. EV charging demand estimation

A myriad of studies have utilized GPS trajectory data to explore trip purposes and spatiotemporal travel patterns to infer EV charging demand. Hu et al. (2018) used drivers' travel activities to evaluate the feasibility of replacing the gasoline yellow taxi with EVs in New York City. GPS data of 13,587 taxis spanning the entire year of 2013 were analyzed to extract spatiotemporal driving patterns, travel demand, dwelling and other features. Their proposed EV feasibility model indicated that only 8% of current taxis can be freely charged under the constraints of mileage range and pickup activity. Similarly, Tu et al. (2016) employed an optimization algorithm to optimize the location of electric taxi in Shenzhen, China. Dynamic pickup demands were estimated using trip data in combination with the corresponding transportation network information first, and then a spatial–temporal demand coverage location model was applied to maximize the taxi service coverage while minimizing the charging wait time. Their results indicate that downtown area, airport, and railway stations have intensive charging demand for electric taxis. The aforementioned studies offer insights on exploring the charging demand of electric taxis based on GPS data and trip activity. However, travel patterns and charging demand can be drastically different between taxis and private vehicles. Kontou et al. (2019) explored the relationship between charging demand and people's daily activities for private vehicles. They identified places with high trip destination densities and prioritized those regions for charging infrastructure installation. The result showed that if top 10% most frequently visited grid cells have installed charging stations in the Puget Sound region, then EV users will be able to access public charging on 71% of their trips. This study suggested that charging probability is highly associated with people's daily travel activities and trip destinations. For EV drivers, they prefer to leave their EVs charging at nearby stations while conducting other activities, e.g., working, shopping, etc. For people who plan to purchase EV in the future, they are less willing to compromise their daily routines to go to distant charging stations. Meanwhile, Vazifeh et al. (2019) reconstructed trajectory using cellular data to track individual movement patterns in the Boston area. A discrete optimization model was formulated by minimizing the total number of charging stations and the average travel distance on those routes.

Apart from GPS trajectory data, urban informatics such as POIs are utilized to analyze the charging demand. Wagner et al. (2013) built a linear regression model using POIs to fit the usage data from more than 32,000 charging sessions in Amsterdam. Results indicated that POI imposes a significant influence on the charging behavior of EV users. Likewise, Dong et al. (2019) applied spatial features including traffic flows, population density, and POI to model the charging demand based on the distribution of current charging stations in London using Bayesian spatial log-Gaussian Cox process model. Statistical analysis showed that transport, retail, and commercial POIs significantly influence the charging demand in the urban area.

2.2. PageRank model and its application

PageRank is one of the most widely used webpage ranking algorithms, developed by Google (Page et al., 1999). PageRank model treats the internet as a large directed graph, where each website represents a node and the hyperlinks are the edges connecting those nodes. Each node is assigned a PageRank value, denoting the importance of the website. PageRank models users' internet browsing behavior as a random walk process. The underlying assumption of PageRank is that more popular web pages are likely to be linked to other web pages, and their importance tends to propagate via hyperlinks. Nodes that are more frequently visited will receive higher PageRank scores and are subsequently deemed more important. PageRank is proved to be highly efficient and simple enough to solve complex graph problems. Yet, a few strategies could be applied to improve its performance. The original PageRank algorithm assumes that the transition probability from one node to all linked nodes is equal. However, that is not always the case in reality. For example, it is likely that a user jumps to a more popular website over the less popular ones. To fix this issue, Xing and Ghorbani (2004) proposed a weighted PageRank algorithm. The core concept of this extended model is to assign higher transition probability to more popular pages instead of distributing them equally. Another deficiency of the original PageRank is that it does not consider the content of web pages. In many situations, users jump to other web pages with similar content. Haveliwala (2003) proposed a topic-sensitive PageRank algorithm by clustering the web pages into a set of topics and biased the original PageRank with those topics. Such modification makes it possible for the PageRank model to incorporate additional features for augmented model performance.

Although PageRank was originally used for ranking websites, it is quite effective in capturing a variety of relations among vertices of graphs (Chung and Zhao, 2010). For instance, PageRank model has been employed to infer traffic states in urban region. Kim et al. (2015) explored the traffic congestion at 57 intersections in Cheongju city. Specifically, they generated a network graph to connect intersections by roads, and then applied PageRank to extract intrinsic relationship of traffic conditions across intersections. The result indicated that the intersections with higher PageRank scores generally have higher traffic density and thus are prone to congestion. Wang et al. (2017) studied the traffic states in urban area of Beijing by partitioning the area into 62 by 65 grids and constructing the network using 12,000 taxi GPS trajectories. The traffic volume between two adjacent grids are used as the weight of the link. It is found that there exists a positive correlation between the PageRank value and congestion index for most regions, and the PageRank value can therefore predict the upcoming congestion. Besides the traffic states inference, PageRank model has been applied to analyze other geographical-related problems, such as revealing urban structure through roads connectivity (Agryzkov et al., 2012).

To the best of our knowledge, PageRank model has not been utilized for EV charging demand estimation to date. Yet, with the proliferation of big data (e.g. GPS trajectory, POI, etc.) as reviewed in Section 2.1, PageRank is well suited for modeling EV charging demand from a graph-theory perspective, as an EV users traveling from an origin and a destination can be treated as Markov process, and the charging demand is highly correlated with the features of trip destinations (Dong et al., 2019). A modified PageRank model has the potential to not only consider EV users' travel habits for demand estimation, but also incorporate social dimension features (e.g. POI, land use) to improve model accuracy.

2.3. Optimization of EVSE location

Public charging infrastructure deployment problem can be deemed as optimally siting EVSE on a landscape. A variety of optimization algorithms have been employed to attempt this from multiple angles. Among them, maximal coverage location problem (MCLP) is a classic model to optimally assign facilities (Church and ReVelle, 1974). Dong et al. (2019) applied a standard MCLP to maximize the coverage of EV charging demand by assigning a fixed amount of charging stations. The model did not consider the constraint of charging stations' capacity, yet in reality charging stations are constrained by energy load. The CMCLP model was subsequently proposed to optimize EVSE location by taking into account the capacity constraints (Chung, Schilling and Carbone, 1983; Current and Storbeck, 1988). A more sophisticated approach to optimizing EVSE location is to formulate it as a multi-objective optimization problem (Zhou et al., 2021). Wang and Wang (2010) proposed a mixed integer programming model to site refueling stations to serve intercity and intra-city travel with the goals of minimizing siting cost and maximizing population coverage. Vazifeh et al. (2019) treated this as a set covering problem with dual objectives of minimum number of charging stations required and minimum average distance for drivers to the nearest accessible charging stations. Extended upon that, Kınay et al. (2021) developed a full cover modeling framework with a novel objective function, which optimizes the charging station locations and determines the optimal OD routes so that the total en-route recharging is minimal for each trip. Yet most of the multi-objective optimization problems are computationally infeasible due to large amount of intricate constraints. Such optimization problems mostly require heuristic algorithms to obtain near-optimal solutions.

To this end, the CMCLP model appears to be computationally efficient and suited for the EV charging infrastructure allocation problem. The flexibility of having the capacity constraint is uniquely aligned with the EV charging problem, as agencies when siting the charging stations, have to consider the energy load capacity for the specific area. Further, the CMCLP is not a binary allocation, in that the model allows partial charging demand assignment. In case where one charging station reaches its capacity, the rest of the charging demand can be allocated to another eligible station nearby.

3. Methodology

In this section, we present the formulation of MGPR in details. Each cell within study area will receive a PageRank score, describing how appealing that cell is to EV drivers. Following that, we build a regression model to map PageRank score to actual charging demand using data from existing charging stations. This trained model is further fed into the CMCLP to optimize the allocation of EVSE for maximized coverage.

3.1. Modified geographical PageRank (MGPR) model

Given a weighted directed graph $\boldsymbol{Q} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{W})$, where $\boldsymbol{V}$ represents the set of nodes, $\boldsymbol{E}$ represents the set of edges, and $\boldsymbol{W}$ is the set of weight corresponding to each edge. The simplified version of PageRank model is defined as follows:

$$\boldsymbol{R}_{t+1} = \boldsymbol{M}\boldsymbol{R}_t \tag{1}$$

where $\boldsymbol{R}_t$ is the vector indicating the PageRank value of each node at step t. Formally, $\boldsymbol{R}_t = [PR(v_1), \cdots, PR(v_n)]^T$, and n is the total number of nodes. $\boldsymbol{M}$ is the stochastic matrix that describes the transition probability from one node to another. For each node j, the transition probability has the following two properties:

$$M_{ij} \geq 0 \tag{2}$$

$$\sum_{i=1}^{n} M_{ij} = 1 \tag{3}$$

After infinite steps of long walk, the PageRank value for each node will converge to a stationary probability denoted by the following equation:

$$\boldsymbol{M}\boldsymbol{R} = \boldsymbol{R} \tag{4}$$

To guarantee convergence, the graph is required to be a strongly connected graph, and an aperiodic one. However, not all network graphs can meet the aforementioned requirements. To address this problem, the simplified PageRank incorporates a random term to make the graph strongly connected and aperiodic. Eq. (4) is therefore modified as follows:

$$\boldsymbol{R} = d\boldsymbol{M}\boldsymbol{R} + \frac{(1-d)}{n}1 \tag{5}$$

where the second term allows each node to have a certain probability to transfer to all other nodes, and d is the damping factor that controls the tradeoff between the first and the second terms.

In the original PageRank model, the network graph does not consider the weight of links. Instead, it assumes all links have equal transition probabilities. Specifically, for node j, M_{ij} shares equal transition probability for each incoming node i, if there is an edge between nodes i and j, and otherwise 0. As noted by Kontou et al. (2019), trip destination density can be used as a surrogate to measure potential EV charging demand. Therefore, in MGPR, instead of assuming equal transition probability across nodes, we use the trip counts (derived from vehicle trajectories) to construct the transition matrix. The transitional probability M_{ij} is thus re-defined as:

$$M_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}} \tag{6}$$

where w_{ij} is the trip count from node i to node j. Another flaw of the original PageRank is the inability to incorporate other information such as web content. Inspired by the topic-sensitive PageRank algorithm (Haveliwala, 2003), we adopted similar concept to our MGPR model such that social dimension that might influence the EV charging demand could be considered. The core idea is to add a third term in the PageRank model. Different from the second term in Eq. (5) which has equal transition probability to all nodes, the third term will direct the drivers to other nodes with varying probabilities. Grid cells that are more favorable to charging activities would receive higher transitional probabilities. For instance, if a grid cell contains large number of commercial buildings, EV drivers are more likely to charge in that cell during the day (while they work). The proposed MGPR model is presented below:

$$\boldsymbol{R} = \alpha_1 \boldsymbol{MR} + \frac{\alpha_2}{n} 1 + \alpha_3 \boldsymbol{G} \tag{7}$$

where $\boldsymbol{G}$ represents the social dimension term, α_1, α_2, and α_3 are the weights of PageRank term, random transferring term, and social dimension term, respectively. The sum of α_1, α_2, and α_3 should equal to 1. In this study, POI data, land-use type information, and socio-economic factors are used to describe urban and geographical features. The social dimension term $\boldsymbol{G}$ is therefore expressed as:

$$\boldsymbol{G} = \frac{1}{3}(\boldsymbol{G}_{POI} + \boldsymbol{G}_{land-use} + \boldsymbol{G}_{socio-economics}) \tag{8}$$

Detailed information and definition with respect to the social dimension features will be introduced in the *Data* and *Results* sections.

3.2. Charging demand estimation

Although PageRank score can quantify the intensity of charging demand, it cannot be directly fed into CMCLP due to its abstraction. We bridge the gap by building a regression to map PageRank score to estimated EV energy consumption using data from existing charging stations. In the regression model, we also include the number of charging ports in each cell as a variable, since larger number of ports is more likely to bring higher energy consumption for the cell. The polynomial regression model is defined as follows:

$$D_i = \max(0, \beta_1 PR(v_i)^k + \beta_2 Y_i + \beta_0) \tag{9}$$

where D_i is the estimated charging demand at cell i; $PR(v_i)$ and Y_i denote the normalized PageRank value and the number of charging ports at cell i, respectively; k is the degree of PageRank score; β_0, β_1, and β_2 are corresponding coefficients;

3.3. CMCLP optimization model

Once the charging demand is estimated, we formulate the optimal public EVSE allocation problem as a CMCLP following similar ideas from (Chung, Schilling and Carbone, 1983; Current and Storbeck, 1988). Given the total investment budget of EVSE, the CMCLP optimizes the construction of charging stations and associated charging ports by maximizing the coverage of total public charging demand under the constraint of charging power capacity. The mathematical formulation is defined as follows:

Objective function:

$$Maximize \sum_i \sum_{j \in N_i} Z_{ij} \tag{10}$$

Subject to:

$$\sum_j Z_{ij} \leq D_i, \forall i \tag{11}$$

$$\sum_j C_s X_j + C_p Y_j \leq P \tag{12}$$

$$\sum_i Z_{ij} \leq c Y_j, \forall j \tag{13}$$

$$Y_j \leq p_{max} X_j, \forall j \tag{14}$$

$$X_j = \{0, 1\}, Y_j \in \mathbb{N}, \forall j \tag{15}$$

where

i is the index of grid cells.
j is the index of candidate grid cells that can be assigned with new charging stations.
Z_{ij} indicates the amount of charging demand that can be covered in grid i by the neighboring charging stations j.

$$X_j = \begin{cases} 1; if \ \ charging \ \ station \ \ j \ \ is \ \ sited \\ 0; otherwise \end{cases}$$

Y_j is the number of charging ports built at j.
C_s and C_p are the cost of installing a charging station and equipment cost of one charging port, respectively.
P is the total investment budget for public charging stations.
D_i is the estimated charging demand in each grid i.
N_i is the set of potential neighboring charging stations for grid cell i.
c is the capacity of each charging port.

p_{max} is the maximum number of charging ports allowed for each charging station.

The objective function (10) is maximizing the coverage of charging demand for all grids. Constraints (11) guarantee neighboring charging stations will only provide the charging energy up to the estimated charging demand for grid i. Constraint (12) imposes the total budget limit for installing public charging stations and ports. Constraints (13) define that the allowed charging energy each charging station serves to the neighboring areas cannot be greater than its capacity. Constraints (14) guarantee charging ports are allowed to be sited only if the grid is selected to build charging stations. Constraints (15) ensure that X_j is a binary variable, and Y_j is a natural number.

According to Seneviratne (1985), people's willingness to walk from parking location to their activity place would become extremely low when the distance is over 3000 feet (0.91 km). In our study, the grid cell is set as 1 km by 1 km. The area that EV users are willing to walk to a charging station can be roughly estimated as a ring buffer from the grid centroid (center of activity) with a radius of 0.91 km, which intersects with all neighboring grids. For this reason, we define that the neighboring charging stations for a grid cell refer to charging stations located in its adjacent 8-directional grid cells. It is also important to note that the D_i is associated with decision variable Y_j, the number of charging ports. Once the coefficients in Eq. (9) are estimated using existing charging station usage, the D_i becomes a linear function of the number of charging ports and is integrated into the optimization model.

The CMCLP is solved using a commercial optimization solver Gurobi (Gurobi, 2021) in this study.

4. Data processing and analytics

4.1. Data

4.1.1. OD data

As noted earlier, in order to apply the MGPR for charging demand inference, the problem is treated as a directed graph, where each grid cell within the study area is treated as a node and the edge is characterized by the number of trip ODs. The OD data are obtained from probe vehicle trajectories provided by Inrix (Inrix, 2021). The trajectories were extracted from a portion of vehicle stream using probe sensors, such as cell phone and automated vehicle location (AVL). The raw data contains 2.5 million trips distributed in the State of Utah during September of 2018. We further extracted trips that are enclosed within the Salt Lake City metropolitan area, as it is the boundary of this study. As described in Section 3, we employ a grid-based approach to partition the region and model EV charging demand. Such level of granularity can remedy the GPS reading errors, while providing sufficient resolution for EVSE planning purpose. Note that determining the size of grid cells is an empirical process. If the size is set too large, it would be difficult to pinpoint the optimal locations of charging stations with fine granularity. On the contrary, small grid cell size may fail to capture hot spots of high charging demand. Previous studies (Tu et al., 2016; Dong et al., 2019; Vazifeh et al., 2019) suggest the appropriate size of grid cells set as 1 km by 1 km. After grid segmentation, there are 756,303 trip OD pairs in total within the region. The maximum number of origin count in a grid cell is 9,281, and the maximum number of destination count in a grid cell is 9,268.

4.1.2. POI data

POI data can effectively reflect urban context and infer people's trip purposes (Wagner et al., 2013). In this study, we use Google Place API (Google Place API, 2021) to extract POIs in our study area. There are 62,673 POIs in total obtained from Google Place API with 103 different labels. In fact, many labels share similar denotations. For instance, both hospital and doctor refer to health-related POI. For simplicity and practical concerns, we further classify the 103 labels into eleven categories. The detailed information of classified POI data is shown in Table 1.

Among the eleven categories, a few do not have apparent association with public charging behaviors. To reduce the noise that might be incurred, Health, NGO, and Service POIs are eliminated from further analysis.

4.1.3. Social dimension features

Socioeconomics describe the relationship between social behavior and economics, while land-use information reveals the human

Table 1
Description of POI data.

POI ID	Category	Label Examples	Total Number
1	Business	office, personal business	23,472
2	Health	hospital, health, doctor	8982
3	Finance	agency, finance building	6691
4	Retail	supermarket, grocery store	10,066
5	Restaurant	restaurant, food delivery	2181
6	Transportation	bus station, train station	3140
7	Education	school, university	1290
8	NGO	church, government building	1591
9	Entertainments	park, salon, bar, zoo	2422
10	Service	post office, gas station, laundry	2427
11	Hotel	hotel, lodging	410

use of land. These two types of geographical-based features are highly associated with people's parking behaviors and subsequently could impact the potential EV charging demand (Csiszár et al., 2019). For this reason, we incorporate socioeconomic and land use features in our model to infer the public charging need. These two datasets are obtained from Wasatch Front Regional Council (WFRC), the metropolitan planning organization that synthesizes a variety of data sources for transportation planning in the study region (WFRC, 2021). Population data is obtained at traffic analysis zone (TAZ) level. Land-use is categorized into agriculture, commercial area, residential area, recreation, and transportation. For simplicity, we reclassified the land-use into commercial vs. non-commercial region only, since most public charging facilities are inclined to be installed in commercial places.

4.1.4. Real-world public charging data

The real-world charging data is crawled from ChargePoint, an online application that assists EV users to navigate and review nearby charging sites (Charge Point, 2021). ChargePoint operates the largest online network of independently owned EV charging stations, operating in fourteen countries worldwide. Directly accessing the utilization information of public EV charging stations is challenging, since it requires authorization from owners of EVSE. Alternatively, such information can be accessed via Google Map Service. We search all public charging stations in the study area first, and then retrieve the real-time data (e.g. the number of in-use ports) constantly for each charging station. Meanwhile, associated features (e.g. power, location) for each charging station are collected. There are 126 public charging stations with 576 charging ports recorded by ChargePoint in the Salt Lake City metropolitan area. Among them, 109 charging stations (516 ports) broadcast real-time utilization information (i.e. number of in-use ports at current time point), indicating charging status. A sample of collected features of public charging stations is displayed in Table 2.

In order to obtain the real-time charging station utilization, we applied *beautifulsoup* API, a Python package for crawling on HTML. The program was further deployed on Amazon Web Service (AWS) Elastic Compute Cloud (EC2) and Simple Storage Service (S3) for dynamic crawling (AWS, 2021). The dynamic crawling framework and database schemas are displayed in Fig. 1(a) and (b), separately. Three virtual machines were rented on EC2 to cover all charging stations in the study area, and the crawling was triggered every 10-minute for each charging station. The data collection spanned from Nov 5th, 2020 to Dec 12th, 2020. The dataset contains 656,179 records for the 109 charging stations.

There are three tables in our database to store the public charging data, as shown in Fig. 1(b). *Charging Station* and *Charging Port* tables record the associated features for charging stations and corresponding charging ports. *Charging Session* table documents the dynamic crawling records. The *in-use ports* feature reflects the number of ports being occupied at a charging station at a specific time point. We further aggregated the crawled records to obtain the charging energy at each station. In a nutshell, once the crawler detects that the charging ports in a station are in use, the energy consumption at that time point is calculated as the total number of in-use ports multiplied by the corresponding power of the ports and 0.167 h (the crawling interval). The accumulative energy consumption is then summed up across the entire crawling period as the total charging energy consumption.

4.2. Spatiotemporal analysis

We examine the spatiotemporal distribution of the charging energy consumption based on the data collected. Fig. 2 indicates the accumulative energy consumption at each charging station within the study period.

In Fig. 2, two charging stations highlighted by green show very high energy consumption. These two charging stations are equipped with Level 3 charging ports (CSS and CHAdeMO) which have much higher power than level 2 ports. The energy consumptions are therefore larger. A cluster of charging stations with high demand exists around the Salt Lake City downtown area (marked by purple). Although none of them exhibit substantial energy consumption, the average usage frequency is relatively high compared to stations in other areas. It is worthy to note that the Salt Lake City international airport (marked by yellow) also indicates high energy consumption. EV users might charge their vehicles while waiting to pick up someone or travel to other places while leaving their vehicles charged at the airport.

We further aggregate the charging station utilization data by different time resolutions to examine the temporal patterns. Fig. 3(a) shows that most of the charging activities occur during the day, between 7:00 am and 7:00 pm. Fig. 3(b) indicates that the average charging demand on Monday and Tuesday is relatively low, while the highest charging demand is observed on Sunday. There is a significant difference between charging patterns across the day-of-week. The average number of daily in-use ports is 145 Monday through Thursday, and 185 Friday through Sunday.

Subsequently, we explore the difference in charging patterns by day-of-week. Specifically, for each station, we calculate the difference between the average daily charging energy consumption from Friday to Sunday and that value from Monday to Thursday. In fact, charging behaviors during Monday to Thursday are more likely to be linked with work trips, while charging behaviors during

Table 2
A sample of stationary features of five public charging stations.

Station ID	Number of Ports	Address	Latitude	Longitude	Power (kW/port)	Status
0	2	425S Orchard Dr, SLC	40.83**	−111.91**	7.2	Available
1	4	2280 Rose Park Ln, SLC	40.77**	−111.94**	7.2	Unknown
2	2	210 N 1950 W, SLC	40.77**	−111.94**	7.2	Available
3	8	168 N 1950 W, SLC	40.77**	−111.94**	7.2	Available
4	6	195 N 1950 W, SLC	40.77**	−111.95**	7.2	Available

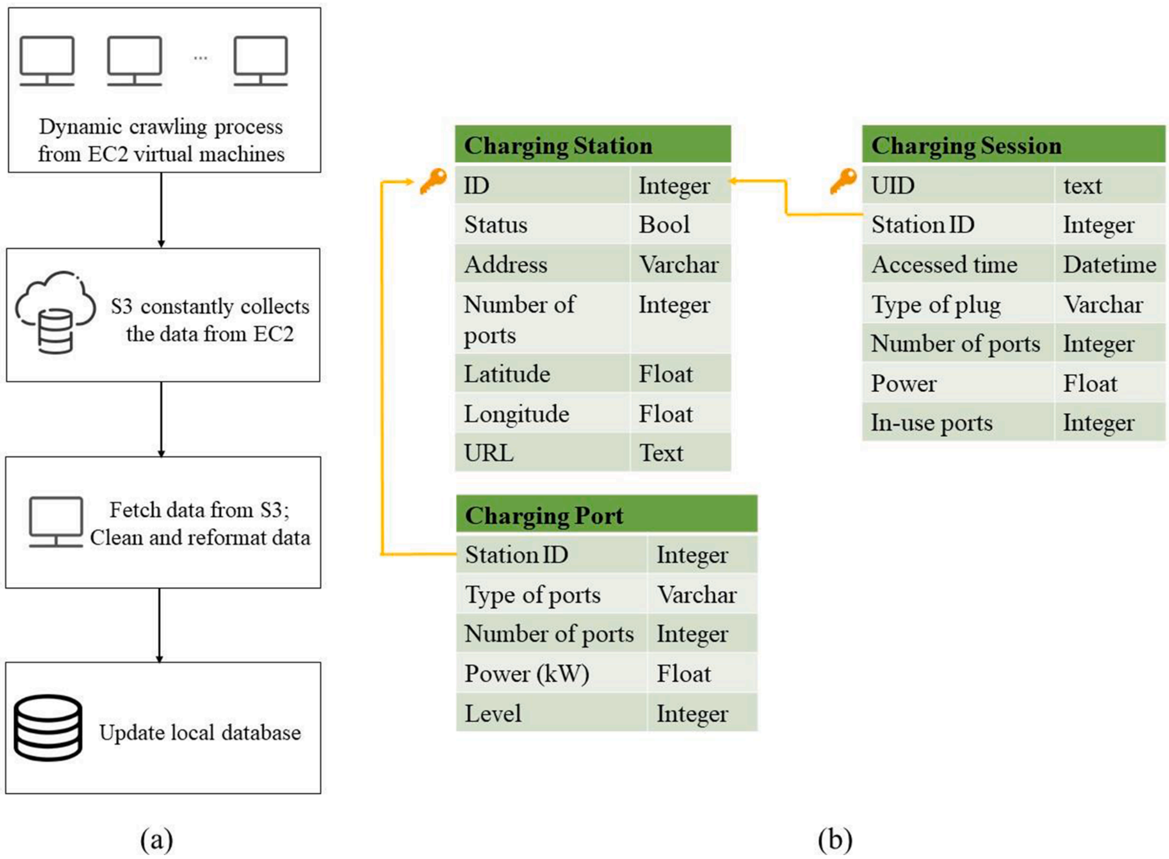


Fig. 1. (a) Framework of dynamic crawling; and (b) the SQL schema for the database of EV charging information.

Friday to Sunday are more likely to be linked with non-work trips. We further overlay the POI data to infer the nature of trip purposes around the charging stations. For POI data, we combine the finance, business, and education into one group to infer work trips. Retail, entertainment, and restaurant POIs are grouped together to infer the non-work trips. The charging differences by day-of-week (i.e. average daily charging energy consumption from Friday to Sunday minus that value from Monday to Thursday) are further overlaid with different POI types in Fig. 4 (a) and (b), respectively.

In Fig. 4(a), it is found that charging stations that are more frequently used during weekdays are located around regions with large number of POIs associated with workplace. In contrast, charging stations that are more frequently visited during weekends are scattered within the area as shown in Fig. 4(b). POIs such as parks, grocery stores are identified in neighboring regions. This distinction of charging patterns with respect to day-of-week validates the aforementioned hypothesis. Note that the Salt Lake City downtown area is mixed land use with both commercial buildings and recreational places. As a result, there is a mix of usage both on weekdays and weekends.

5. Result and analysis

5.1. MGPR model

The MGPR model incorporates travel mobility and urban informatics data to quantify the importance of each grid cell connected in the directed graph. When the model converges, each grid cell receives a PageRank score, inferencing the intensity of public charging demand. The normalized PageRank score ranges between 0 and 1 with higher score implying more importance and therefore higher charging demand. In MGPR, the OD matrix $\boldsymbol{M}$ is defined in Eq. (6). The social dimension term $\boldsymbol{G}$ includes $\boldsymbol{G}_{POI}$, $\boldsymbol{G}_{socio-economics}$, and $\boldsymbol{G}_{land-use}$.$\boldsymbol{G}_{POI}$ for grid cell i is calculated as the number of POIs in the grid cell i divided by the total number of POIs in the study area. Similarly, population density for grid cell i is calculated to represent $\boldsymbol{G}_{socio-economics}$ value for grid cell i. Lastly, dummy variable is used to indicate whether grid cell i is a commercial area. The dummy value is then normalized to represent $\boldsymbol{G}_{land-use}$ value for grid cell i. The coefficients α_1, α_2, and α_3 control the OD matrix, random effect, and social dimension, respectively. For example, large α_1 will cause OD matrix dominates PageRank score. In this study, the parameter values are determined empirically. α_2 is set as 0.05 since random effect is expected to be marginal. Meanwhile, it is found that when α_1 ranges between 0.5 and 0.7, not much variation on results is detected. The optimal result is observed when α_1, α_2, and α_3 are set as 0.6, 0.05, and 0.35, separately. Once the coefficients are

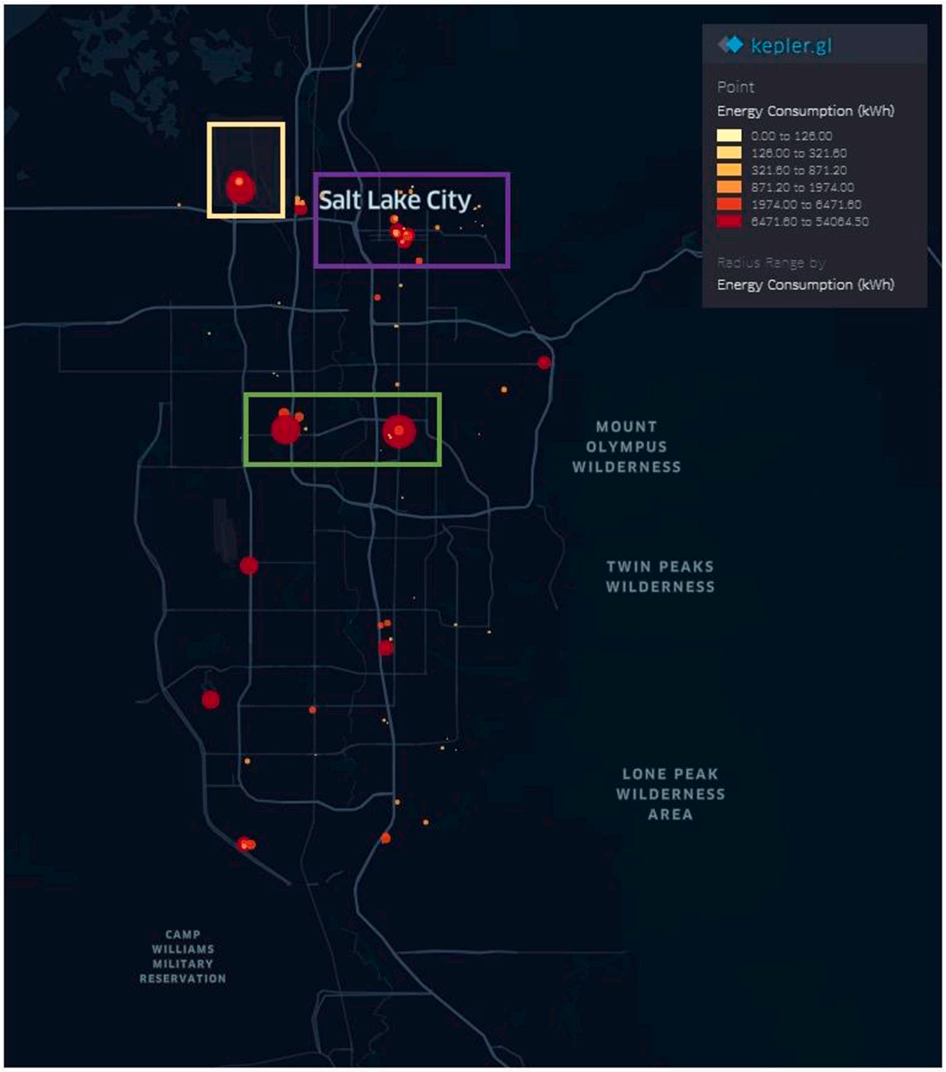


Fig. 2. Total energy consumption of each public charging stations across entire period. Each dot represents a charging station. Radius and color indicate the level of energy consumption.

determined, PageRank value is determined via an iterative process. We keep updating the PageRank vector $\boldsymbol{R}_t$ in Eq. (7) at each step t until it reaches convergence. The total number of iteration steps is set as 500 for the MGPR. Numeric result indicates that when the step reaches 20, the PageRank value almost converges.

To assess the effectiveness of MGPR model, we compare the PageRank score with actual charging energy using different metrics. Ideally, grid cells with lower energy consumption should receive lower score, and grid cells with larger amount of energy consumption should receive higher score instead. Four prevailing methods from existing studies on public charging demand prediction are implemented for comparison purpose:

1. $\mathrm{LR}_{\mathrm{POI}}$: a linear regression model using the count of POI in each grid to infer public charging energy proposed by Wagner et al. (2013);
2. $\mathrm{LR}_{\mathrm{trip\_destionation}}$: a linear regression model using the count of trip destination in each grid to infer public charging energy. Kontou et al. (2019) used the count of trip destination in each grid cell to quantify the public charging demand;
3. Spatial Log-Gaussian Cox process: a spatial point process statistical model. Dong et al. (2019) used POI, population density, and traffic flow to infer public charging demand in London with spatial Log-Gaussian Cox process model;
4. XGBoost: Almaghrebi et al. (2020) applied XGBoost (Chen and Guestrin, 2016) to predict public charging demand. In our study, the count of trip destination, the count of POI, land-use type, and commercial type for each grid are used as features. We split our data by 60% as training data, and the rest as test data, and perform it five times by randomly shuffling.

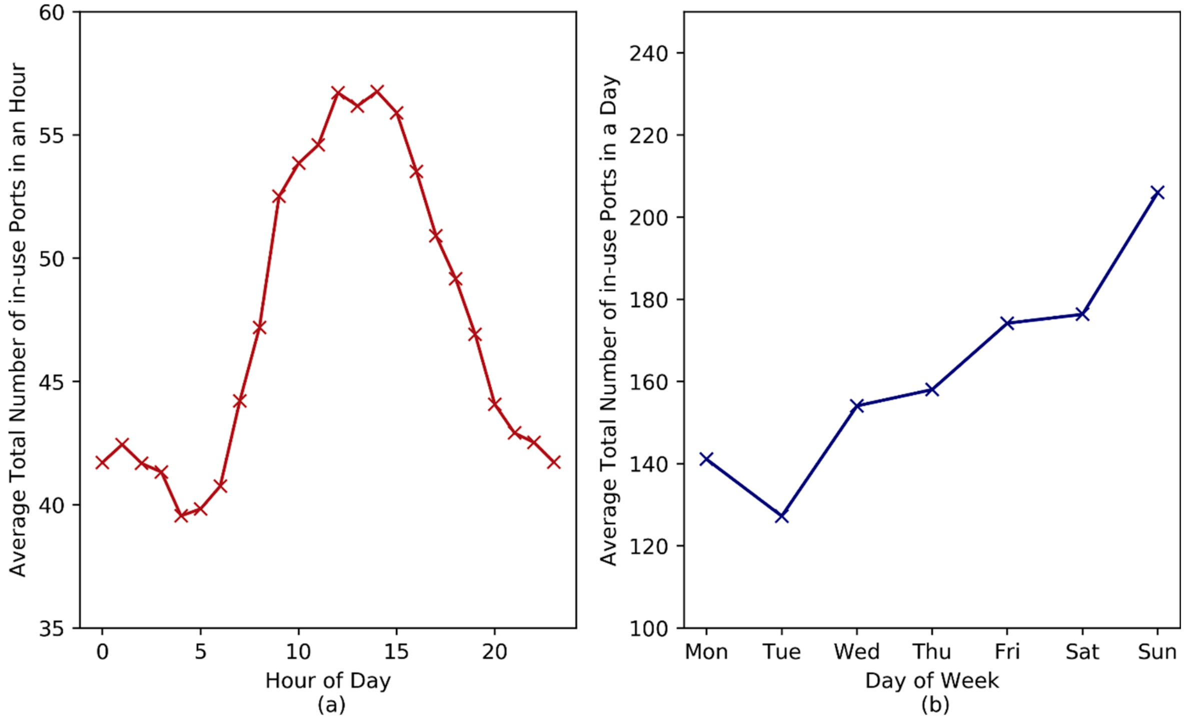


Fig. 3. (a) The average total number of in-use ports in each hour-of-day; and (b) the average total number of in-use ports in each day-of-week.

Moreover, the original PageRank and weighted PageRank models are implemented. For original PageRank and weighted PageRank models, we adopt 0.85 for the dampening factor in Eq. (5) as seen in (Page et al., 1999; Xing and Ghorbani, 2004). The computation of PageRank values for original PageRank and weighted PageRank models follows similar process. To assess model performances, the following metrics are calculated:

Pearson correlation coefficient (r_{xy}):

$$r_{xy} = \frac{\sum_{i \in S}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i \in S}(x_i - \overline{x})^2}\sqrt{\sum_{i \in S}(y_i - \overline{y})^2}} \tag{17}$$

where S is the set of grid cells with charging facilities; x_i is the predicted value of the i^{th} grid cell by each model, and y_i is the actual charging energy of the i^{th} grid cell. r_{xy} measures the linear correlation between model's output value and actual charging energy. The coefficient ranges from −1 to 1, with the absolute value closer to 1 indicating a stronger linear correlation.

Ranked total energy (RTE):

$$RTE = \sum_{i \in S} Rank\_x_i * y_i \tag{18}$$

$Rank\_x_i$ is the rank of the i^{th} grid cell sorted by model's output in ascending order. It measures the ability of the model in capturing regions of high charging demand (Xing and Ghorbani, 2004). If the ranking of grid cells with large charging energy is high, the RTE would be large as well.

Mean absolute rank difference (MARD):

MARD is calculated by averaging the absolute difference of the rank of model's output and the rank of actual charging energy for each grid cell equipped with charging facilities. It is expressed as follows:

$$MARD = \frac{1}{N}\sum_{i \in S}|Rank\_x_i - Rank\_y_i| \tag{19}$$

where N is the total number of grid cells with EVSE; $Rank\_y_i$ is the ranking of actual charging energy consumption for the i^{th} grid cell. The smaller the MARD, the better the result. MARD value reflects how well the model can distinguish high and low charging demand regions. The results of all models are shown in Table 3.

It is worthy to mention that the models for comparison do not generate PageRank score. Specifically, linear regression models and XGBoost generate the estimated charging energy for each grid cell, and Spatial Log-Gaussian Cox process model outputs the intensity value which is similar to PageRank score. However, the proposed metrics can be well applied across all models for performance

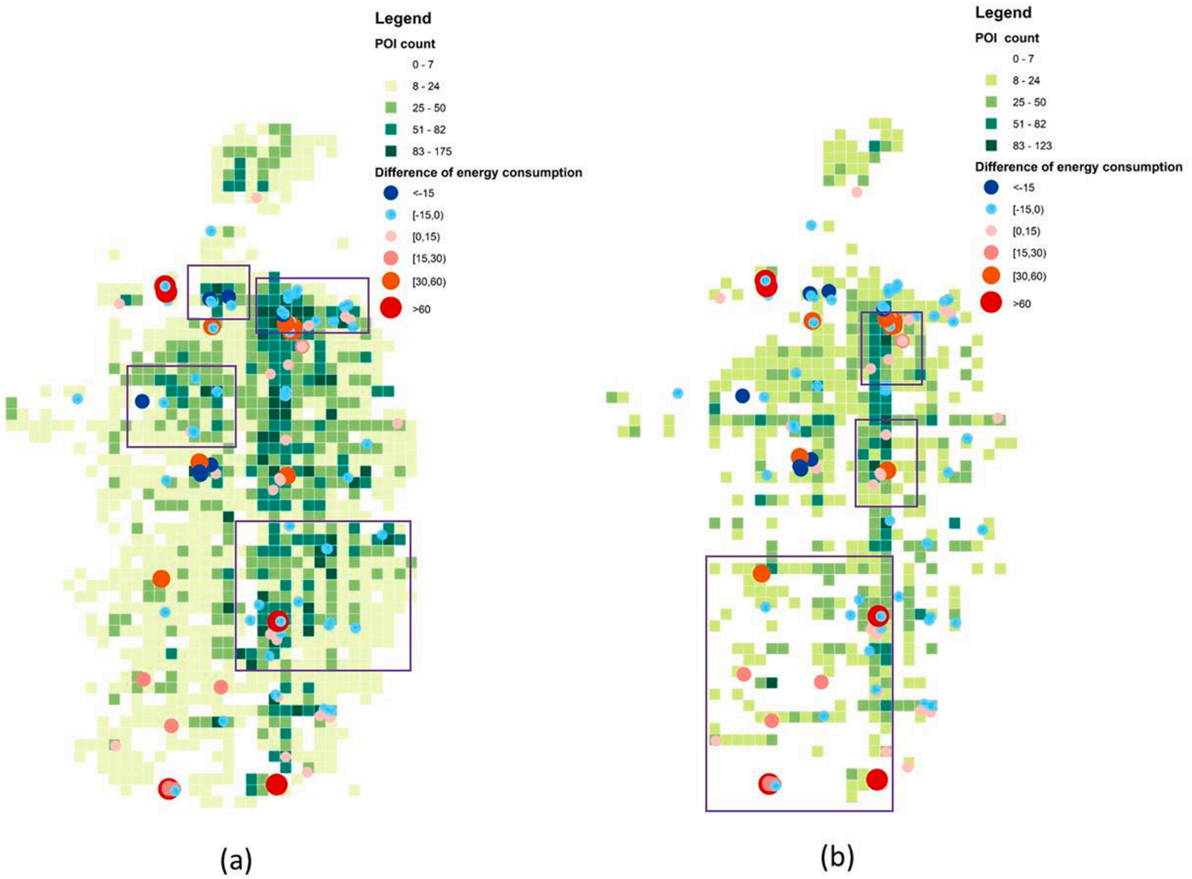


Fig. 4. (a) The work-related POIs (financial buildings, business, and education) distribution and public charging stations' charging patterns; (b) the non-work-related POIs (entertainment places, retails, and restaurants) distribution and public charging stations' charging patterns.

evaluation. In Table 3, it is observed that the linear regression model using POI data yields the worst performance among all models with Pearson's r being only 0.21. This result is consistent with Wagner's (2013) experiment, where the R^2 was merely 0.16 in his study. The relatively low correlation reveals the fact that inferencing public charging demand only with POIs might achieve rather limited effectiveness. In contrast, the linear regression model using the count of trip destination and spatial Log-Gaussian Cox Process model obtain satisfactory predicting results. It demonstrates that the density of trip destinations is a key indicator of the public charging demand, and urban informatics can potentially impact drivers' charging preference. The conclusions are aligned with Kontou's (2019) and Dong's (2019) findings. The spatial Log-Gaussian Cox process is essentially a spatial Bayesian model. The model assumes stronger intensity of charging demand in area where the number of charging facilities is large. However, this is not the case in reality, since some charging facilities might be underutilized. Such assumption can consequentially lead to the decrease of predicting power. As for the XGBoost model, it achieves relatively low Pearson's r (0.47) compared to spatial Log-Gaussian Cox process, since each grid is treated as a discrete unit without spatial correlation. Another issue that constrains the performance of XGBoost is the small amount of training samples. Note that RTE and MARD cannot be computed for XGBoost because a portion of grid cells are used as training samples. For the PageRank-series models, weighted PageRank marginally improves RTE and MARD values by 4% and 3% respectively compared to original PageRank. Further, MGPR effectively augments the model performance over weighted PageRank with 9% increase of RTE

Table 3
The overall performance of MGPR and methods from existing studies.

Models	r_{xy}	RTE	MARD
Linear Regression$_{\text{POI}}$	0.21	$7.5*10^6$	17.5
Linear Regression$_{\text{Trip Destination}}$	0.53	$1.70*10^7$	16.1
Spatial Log-Gaussian Cox process	0.50	$9.6*10^6$	12.2
XGBoost	0.47	NA	NA
Original PageRank	0.39	$1.63*10^7$	18.8
Weighted PageRank	0.53	$1.70*10^7$	18.2
MGPR	0.70	$1.85*10^7$	11.8

from $1.70*10^7$ to $1.85*10^7$, and 35% decrease of MARD from 18.2 to 11.8. MGPR also achieves the highest Pearson's r value (0.70) among all models. The significant reduction of MARD by incorporating social dimension features illustrates that for some of regions with high trip density their charging demands are not necessarily high, e.g., residential neighborhoods. To better visualize the results geographically, we present the grid plot for each model, separately, in Fig. 5.

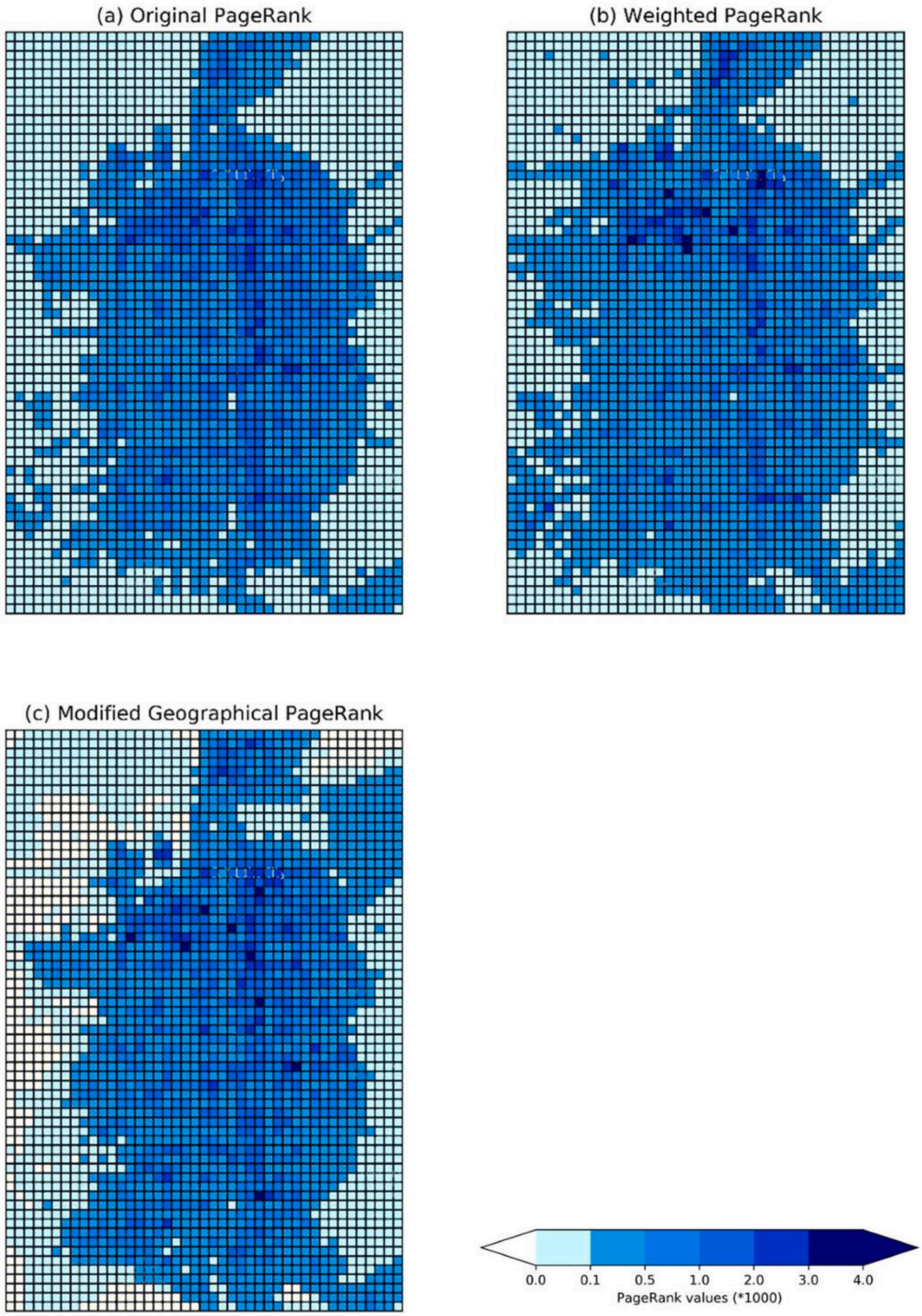


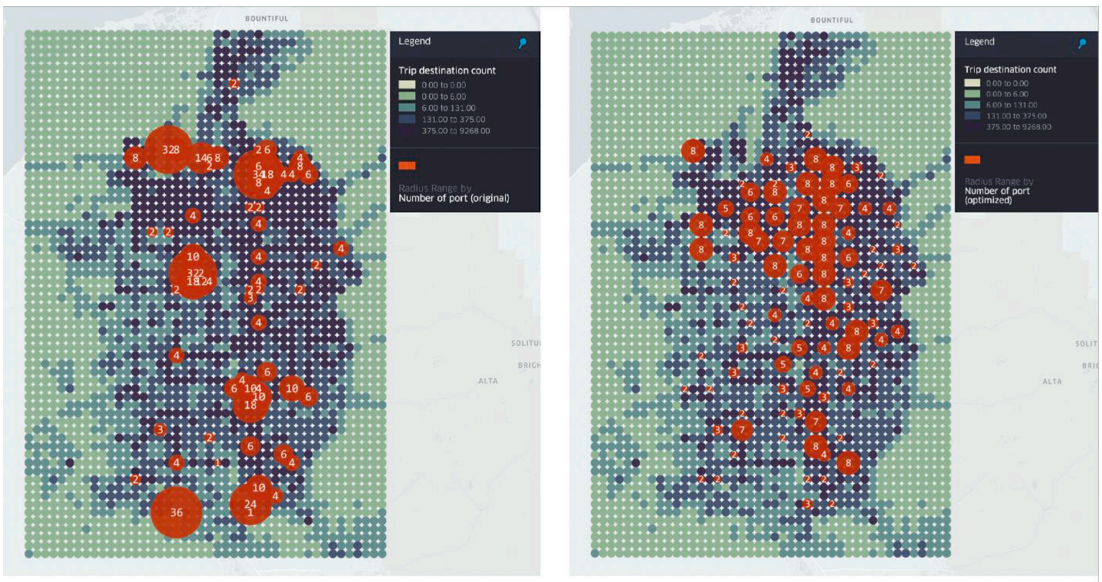
Fig. 5. PageRank values distribution of (a) original PageRank model; (b) weighted PageRank model; and (c) MGPR model.

The original PageRank and weighted PageRank indicate similar public charging demand distributions as illustrated in Fig. 5(a) and (b). However, some cells receive higher PageRank values by employing weighted PageRank. Those cells have higher trip density, and subsequently more appealing to charging activities. In Fig. 5(c), it is observed that the PageRank values of large portion of cells located in the outskirts decrease to 0. We also notice that the PageRank values for some cells increase on the contrary, which is attributable to the POIs, population density, and land-use type. In summary, MGPR presents a more distinct pattern compared to original PageRank, and weighted PageRank.

5.2. Optimizing public charging stations

The least square method is applied to fit Eq. (9) using data of 63 grid cells with existing charging stations. The degree of PageRank score k is determined as 2, and the values of β_0, β_1 and β_2 are -107.73, $1.25*10^7$, and 15.46, respectively. The value of R^2 being 0.78 indicates a satisfactory model fitness. The trained regression model is integrated into the CMCLP to explore the optimal siting of EVSE. While there are 2,816 grid cells, not all grid cells are suitable for siting new public charging stations due to their unique land-use types. We assume that the new public charging stations can be only installed in grids that have public parking lots. The public parking lots data is available from WFRC (WFRC, 2021), and there are 463 grid cells in total with public parking lots. Without loss of generality, we assume each grid cell is allowed to install at most one public charging station, with the maximum number of allowed charging ports p_{max} being 8 (empirical data shows that existing charging stations have no more than 8 ports with only a few exceptions). The power of each L2 commercial charging port is 7.2 kW, and the daily capacity c is set as 7.2*24 kWh per port, assuming all public charging stations can be freely accessed 24/7. In general, the cost of building one charging station C_s is approximately $5500 including labor cost, materials and cable connection, and the average price for one L2 commercial port C_p is approximately $2500 (Borlaug et al., 2020). In order to optimize the layout of existing charging infrastructure, the total budget P is calculated using the current 109 charging stations with 516 ports, which is approximated at $1.89 million ($5500*109 + $2500*516). The layout before and after optimization is displayed in Fig. 6.

There are currently 109 charging stations with 516 ports in total. By using the CMCLP to maximize the total coverage of public charging demand, the total number of charging stations expands to 130 with 470 charging ports using the same budget. The wider distribution of charging stations shown in Fig. 6(b) allows for expanded coverage where charging facilities are unavailable in current setting. In comparison, the existing charging stations can only cover $4.0*10^4$ kWh public charging demand, while the optimized layout can significantly improve the coverage to $7.2*10^4$ kWh. Meanwhile, it is observed that the optimally allocated charging stations are clustered in Salt Lake City downtown (northeastern), indicating high charging demand in that area. In fact, a large portion of existing charging stations are deployed in downtown already. Such consistency suggests that most of charging stations in this region are effectively exploited. Yet there are areas (e.g. southern Salt Lake County) where a congregated number of EVSEs are currently present that might need to be reallocated. One possible explanation is that those EVSEs are located in residential neighborhoods where people can just charge their EVs at home instead. Another possible explanation is that those EVSEs are distant from drivers' activity places,



(a) (b)

Fig. 6. (a) The layout of current charging stations aggregated by grid cells; and (b) the layout of charging stations after optimization, where each circle represents a charging station and the radius indicates the number of ports for each charging station. The bottom layer shows the count of trip destinations in each grid cell.

and people are not willing to walk for charging. It is also found that many optimal charging stations are distributed along the freeways, such as I-15. Chen et al. (2016) mentioned that freeways generally have excessive vehicle flow and correspondingly high fast-charging demand. Future deployment of public charging stations can be considered to be located in the close vicinity of freeway entrance or exit.

When combining with the trip destination data, we find that the optimized layout of charging stations is more concentrated on grids with dense trip destination counts, which is beneficial to drivers accessing nearby charging stations, because the walking distance from charging station to activity place plays a crucial role considering EV drivers' willingness of charging publicly. To quantify the improvement on access convenience, we calculate the average distance for all trip destinations to their nearest charging stations. The result shows that the average distance is decreased by 39% from 2.3 km to 1.4 km after optimization. EV drivers may opt for alternative solutions such as home charging if walking distance is above 0.91 km (Seneviratne, 1985). Based on this assumption, we further determine the number of trips that are feasible for public charging. We define that a trip is feasible for public charging only if charging station exists in a trip's destination grid or adjacent ones. It is found that in the current layout, only 45.5% of 756,303 trips are feasible within such defined threshold. However, such rate is effectively augmented to 70% by optimally reallocating charging stations. The improvement on access convenience can in return incentivize more charging opportunities from potential EV drivers and consequently increase the revenue.

Current spatial configuration of EVSEs indicates access inconvenience and unbalanced charging utilization. In fact, siting additional EVSEs using CMCLP optimization based on current layout instead of reallocating existing EVSEs is more meaningful for both short-term and/or long-term investment. Planners could make more informed evaluation on how many additional charging stations

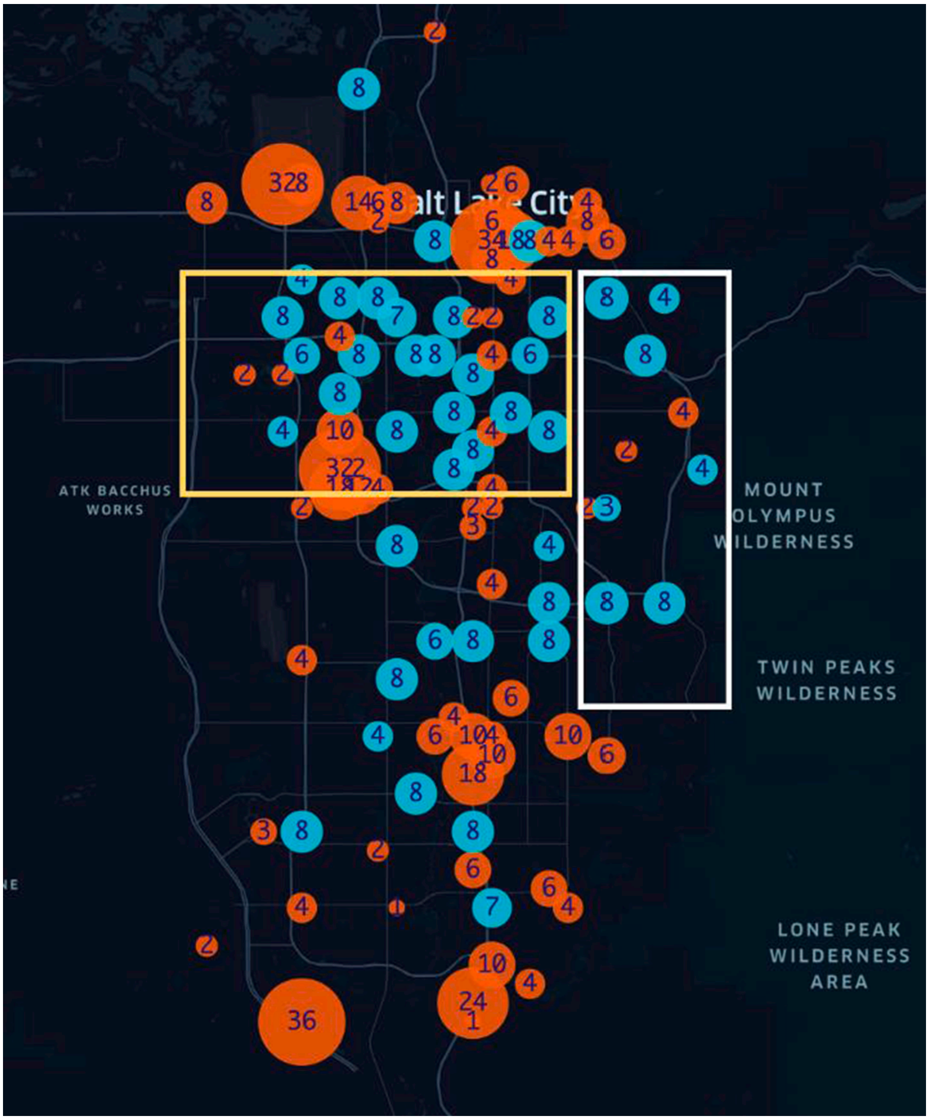


Fig. 7. The optimized layout based on existing public charging stations with additional $1 million investment cost, where red dots represent existing charging stations and blue dots represent newly installed charging stations.

are required to satisfy public charging needs. For short-term planning, an example of adding $1 million investment budget is presented to identify areas that are the most desirable to site new charging stations. Fig. 7 displays the geographical layout of the newly sited 42 charging stations (307 ports) with the additional $1 million investment. Only two charging stations need to be installed in Salt Lake City downtown as it is already equipped with abundant charging facilities. On the contrary, other urban areas close to downtown such as West Valley City and South Salt Lake (highlighted by yellow) are suggested to site more charging stations to cover potential charging demand. Apart from that, tourist attraction sites with parks and trials (highlighted by white) are encouraged to build more charging stations due to large volume of trips during weekends.

As mentioned earlier, distance from activity place to charging station plays a crucial role in EV users' willingness to charge publicly. Shorter distance in accessing the nearest charging station improves charging convenience and makes public charging more appealing to users. For this reason, we explore the average distance to the nearest charging stations of all trips in response to the increase of investment. The corresponding number of trips that are feasible for public charging is calculated as well. Specifically, we change the investment with an increment of $1 million dollars until the average distance does not improve. The result is displayed in Fig. 8.

It is found that when the investment exceeds $10 million dollars, the average distance from the destination to the nearest charging station will not improve any longer. The proposed metric drops from 2.3 km to 0.6 km with the corresponding percentage of trips that are feasible for public charging increases from 45.5% to 88.2%. However, the improvement becomes marginal after the investment exceeds $2 million dollars.

For long-term planning, EV adoption growth and the consequently increased demand need to be taken into account. Existing EV adoption forecast models are difficult to replicate due to the unavailability of micro-level variables such as consumer price index needed for the models (Duan, Gutierrez and Wang, 2014; Zhang et al., 2017). For simplicity, the assumption of linear growth of charging demand by 10% every year is adopted for short-term forecast in our study (Block et al., 2015). Specifically, we solve the CMCLP by increasing the total budget by every $1 million until coverage of public charging demand is maximized. This set of experiments is performed across different years from 2020 to 2025 with 10% annual demand growth as shown in Fig. 9.

In Fig. 9, we observe that the investment of the first $1 million brings the most significant enhancement of coverage rate by 80% from $4.0*10^4$ kWh to $7.2*10^4$ kWh. The growth rate of coverage slows down after the investment cost exceeds $3 million, indicating the newly installed charging stations are not fully exploited. When investment cost is over $10 million, the coverage of public charging demand reaches maximum value ($12.1*10^4$ kWh), and installing additional charging stations may only bring very marginal benefit. Meanwhile, we identify a non-linear increase of investment cost with years due to the annual surge of public charging demand. To reach the maximum public charging demand, the total budget should be increased from $10 million to $12 million in 2025.

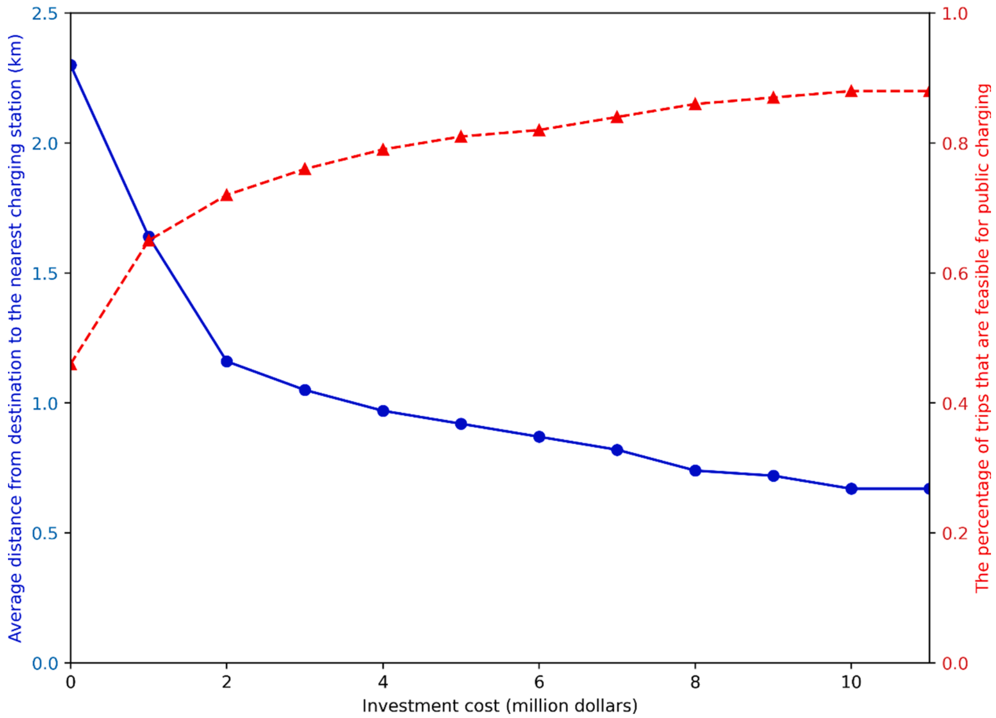


Fig. 8. The average distance from destination to the nearest charging station and the percentage of trips that are feasible for public charging with increasing investment budget.

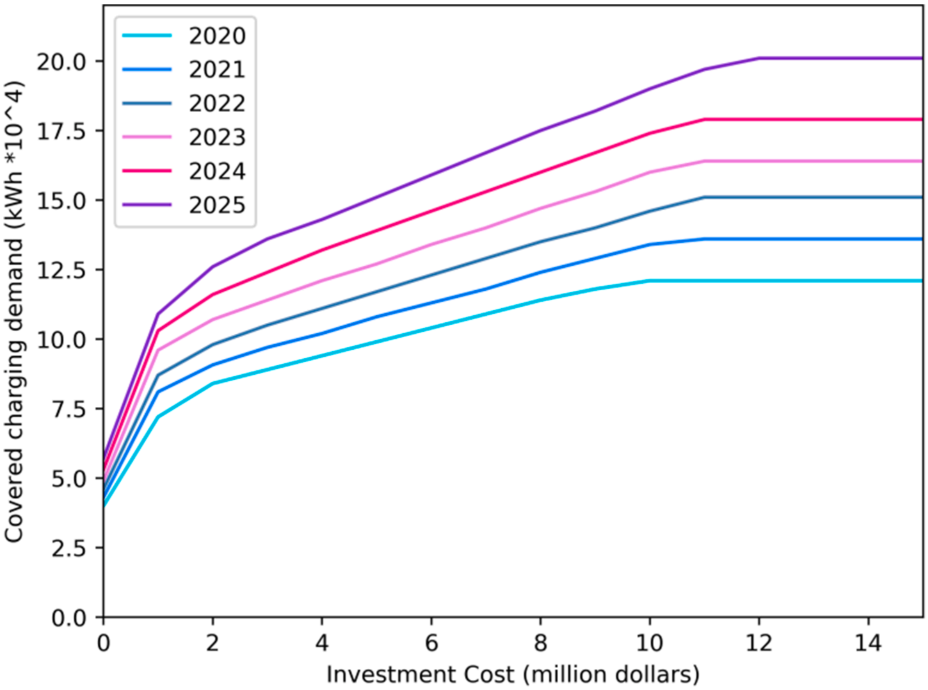


Fig. 9. Covered public charging demand with increasing investment budget from 2020 to 2025.

6. Conclusion

In this paper, we present a methodological framework for EV charging demand estimation and EVSE optimization using advanced graph-theory based approach and mathematical modeling. First, we developed a MGPR model based on the classic web ranking algorithm-PageRank-to explore potential charging demand by translating the problem into a directed graph and utilizing trip ODs and social dimension features. Second, an optimization model CMCLP is employed to optimize the locations of EVSEs by maximizing the utilization of charging stations. In addition, a crawling pipeline is created to retrieve real-world public charging data for modeling purpose. Such pipeline framework can be widely generalizable to other sites.

The Salt Lake City metropolitan area is chosen to demonstrate the effectiveness of the framework. Real-world data were obtained from 109 public charging stations from Nov 5th, 2020 to Dec 12th, 2020. The proposed MGPR is compared against four prevailing methods from existing studies on public charging demand prediction. Metrics including Pearson's r, RTE, and MARD indicate that MGPR achieves the most desirable performance, further validating that trip ODs and social dimension features can effectively infer the public charging demand. Once potential charging demand for each cell is estimated by regression model, CMCLP is employed to optimize the EVSE locations. It is found that most of existing charging stations located at Salt Lake City downtown are effectively exploited; while some charging stations located in Southern Salt Lake are underused. Meanwhile, it is observed that there are mismatches between the currently deployed charging infrastructures and charging demand. More charging stations are encouraged to be sited along the interstate highways and tourist attraction sites for future planning. By optimally reallocating existing charging stations, the average distance for all trip destinations to the nearest EVSEs is decreased from 2.3 km to 1.4 km. We further examine the optimization based on current layout for both short-term and long-term planning. Results indicate that by investing $1 million with 42 new stations, the coverage of public charging demand can be significantly increased by 80%. The total budget should raise from $10 million to $12 million if full public charging demand needs to be covered in year 2025. Due to the small amount of Level 3 charging stations (34 out of 576), we simplified the scenarios by assuming all charging stations are Level 2 in the CMCLP model. However, cities with higher EV adoptions such as Los Angeles have much higher proportion of fast charging (Level 3) stations. Future work therefore should consider the scenario of integrating Level 3 stations in the planning process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This article is based upon work partially supported by the National Science Foundation under Grant No. 2051226, and partially supported by the National Institute for Transportation and Communities (NITC, grant number 1519), a U.S. DOT University Transportation Center. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s).

References

Agryzkov, T., Oliver, J.L., Tortosa, L., Vicent, J.F., 2012. An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector. Appl. Math. Comput. 219 (4), 2186–2193.

Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K., Alahmad, M., 2020. Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods. Energies 13 (16), 4231. https://doi.org/10.3390/en13164231.

AWS, 2021. Available at: https://aws.amazon.com/.

Block, D., et al., 2015. Electric Vehicle Sales for 2014 and Future Projections. (March 2015), pp. 1–9.

Borlaug, B., Salisbury, S., Gerdes, M., Muratori, M., 2020. Levelized cost of charging electric vehicles in the United States. Joule. Elsevier Inc. 4 (7), 1470–1485. https://doi.org/10.1016/j.joule.2020.05.013.

Charge Point, 2021. Available at: https://na.chargepoint.com/charge_point.

Chaudhari, K., Kandasamy, N.K., Krishnan, A., Ukil, A., Gooi, H.B., 2019. Agent-based aggregated behavior modeling for electric vehicle charging load. IEEE Trans. Ind. Inf. IEEE 15 (2), 856–868.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17-Augu, pp. 785–794. doi: 10.1145/2939672.2939785.

Chung, C.H., Schilling, D.A., Carbone, R., 1983. The capacitated maximal covering problem: A heuristic. Proceedings of Fourteenth Annual Pittsburgh Conference on Modeling and Simulation 1423–1428.

Chung, F., Zhao, W., 2010. PageRank and random walks on graphs. Bolyai Soc. Math. Stud. 20, 43–62. https://doi.org/10.1007/978-3-642-13580-4_3.

Church, R., ReVelle, C., 1974. The maximal covering location problem. Pap. Reg. Sci. Assoc. 32 (1), 101–118.

Csiszár, C., Csonka, B., Földes, D., Wirth, E., Lovas, T., 2019. (2019) 'Urban public charging station locating method for electric vehicles based on land use approach'. J. Transp. Geogr. Elsevier 74, 173–180. https://doi.org/10.1016/j.jtrangeo.2018.11.016.

Current, J.R., Storbeck, J.E., 1988. Capacitated covering models. Environ. Plan. B: Plan. Des. SAGE Publications Sage UK: London, England 15 (2), 153–163.

Dong, G., Ma, J., Wei, R., Haycox, J., 2019. (2019) Electric vehicle charging point placement optimisation by exploiting spatial statistics and maximal coverage location models. Transp. Res. Part D: Transp. Environ. Elsevier 67, 77–88. https://doi.org/10.1016/j.trd.2018.11.005.

Duan, Z., Gutierrez, B., Wang, L., 2014. Forecasting plug-in electric vehicle sales and the diurnal recharging load curve. IEEE Trans. Smart Grid. IEEE 5 (1), 527–535. https://doi.org/10.1109/TSG.2013.2294436.

Engel, H., et al., 2018. Charging ahead: electric- vehicle infrastructure. McKinsey & Company, (Exhibit 1), 1–8.

Gersdorf, T., et al., 2020. McKinsey Electric Vehicle Index: Europe Cushions a Global Plunge in EV Sales. McKinsey & Company, July(July), pp. 1–12.

Google Place API, 2021. Available at: https://developers.google.com/maps/documentation/places/web-service/search.

Grote, M., Preston, J., Cherrett, T., Tuck, N., 2019. Locating residential on-street electric vehicle charging infrastructure: A practical methodology. Transp. Res. Part D: Transp. Environ. Elsevier 74, 15–27. https://doi.org/10.1016/j.trd.2019.07.017.

Gurobi, 2021. Available at: https://www.gurobi.com/.

Hall, D., Lutsey, N., 2017. Emerging Best Pratices for Electric Vehicle Charging Infrastructure. ICCT White Paper, (October).

Haveliwala, T.H., 2003. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. IEEE Trans. Knowl. Data Eng. 15 (4), 784–796. https://doi.org/10.1109/TKDE.2003.1208999.

Hu, L., Dong, J., Lin, Z., Yang, J., 2018. Analyzing battery electric vehicle feasibility from taxi travel patterns: The case study of New York City, USA. Transp. Res. Part C: Emerg. Technol. Elsevier 87, 91–104. https://doi.org/10.1016/j.trc.2017.12.017.

Huo, H., Cai, H., Zhang, Q., Liu, F., He, K., 2015. Life-cycle assessment of greenhouse gas and air emissions of electric vehicles: A comparison between China and the U.S. Atmos. Environ. Elsevier Ltd 108, 107–116. https://doi.org/10.1016/j.atmosenv.2015.02.073.

Inrix, 2021. Available at: https://inrix.com/.

Kim, Y.Y., et al., 2015. Analysis on the transportation point in Cheongju City using Pagerank algorithm. ACM International Conference Proceeding Series 20–23-Octo, 165–169. https://doi.org/10.1145/2837060.2837087.

Kınay, Ö.B., Gzara, F., Alumur, S.A., 2021. Full cover charging station location problem with routing. Transp. Res. Part B: Methodol. 144, 1–22. https://doi.org/10.1016/j.trb.2020.12.001.

Kontou, E., Liu, C., Xie, F., Wu, X., Lin, Z., 2019. Understanding the linkage between electric vehicle charging network coverage and charging opportunity using GPS travel data. Transp. Res. Part C: Emerg. Technol. Elsevier 98, 1–13. https://doi.org/10.1016/j.trc.2018.11.008.

Leou, R.-C., Su, C.-L., Lu, C.-N., 2014. Stochastic analyses of electric vehicle charging impacts on distribution network. IEEE Trans. Power Syst. IEEE 29 (3), 1055–1063.

McKinsey, 2019. Expanding electric - vehicle adoption despite early growing pains. McKinsey & Company, (August). Available at: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/expanding-electric-vehicle-adoption-despite-early-growing-pains.

Page, L., et al., 1999. The PageRank citation ranking: Bringing order to the web.

Qian, K., et al., 2010. Load model for prediction of electric vehicle charging demand. In: 2010 International Conference on Power System Technology, pp. 1–6.

Seneviratne, P.N., 1985. Acceptable walking distances in central areas. J. Transp. Eng. American Society of Civil Engineers 111 (4), 365–376.

Smart, J., Schey, S., 2012. Battery electric vehicle driving and charging behavior observed early in the EV project. SAE Technical Papers 1 (1), 27–33. https://doi.org/10.4271/2012-01-0199.

Tu, W., Li, Q., Fang, Z., Shaw, S.-L., Zhou, B., Chang, X., 2016. 'Optimizing the locations of electric taxi charging stations: A spatial–temporal demand coverage approach',. Transp. Res. Part C: Emerg. Technol. Elsevier Ltd 65, 172–189. https://doi.org/10.1016/j.trc.2015.10.004.

Vazifeh, M.M., Zhang, H., Santi, P., Ratti, C., 2019. Optimizing the deployment of electric vehicle charging stations using pervasive mobility data. Transp. Res. Part A: Policy Pract. Elsevier 121, 75–91. https://doi.org/10.1016/j.tra.2019.01.002.

Wagner, S., Götzinger, M., Neumann, D., 2013. Optimal location of charging stations in smart cities: A point of interest based approach. International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design, 3, pp. 2838–2855.

Wang, M., Yang, S.u., Sun, Y.i., Gao, J., 2017. Discovering urban mobility patterns with PageRank based traffic modeling and prediction. Physica A. Elsevier B.V. 485, 23–34.

Wang, Y.W., Wang, C.R., 2010. Locating passenger vehicle refueling stations. Transp. Res. Part E: Logist. Transp. Rev. Elsevier Ltd 46 (5), 791–801. https://doi.org/10.1016/j.tre.2009.12.001.

WFRC, 2021. Available at: https://wfrc.org/.

Xing, W., Ghorbani, A., 2004. Weighted PageRank algorithm. In: Proceedings - Second Annual Conference on Communication Networks and Services Research. IEEE, pp. 305–314. https://doi.org/10.1109/dnsr.2004.1344743.

Xu, M., Yang, H., Wang, S., 2020. Mitigate the range anxiety: Siting battery charging stations for electric vehicle drivers. Transp. Res. Part C: Emerg. Technol. Elsevier 114 (January), 164–188. https://doi.org/10.1016/j.trc.2020.02.001.

Yi, Z., et al., 2021. Electric vehicle charging demand forecasting using deep learning model. J. Intell. Transp. Syst. Taylor & Francis 1–14.
Zhang, Y., Zhong, M., Geng, N., Jiang, Y., Hu, X., 2017. Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. PLoS ONE 12 (5), e0176729. https://doi.org/10.1371/journal.pone.0176729.
Zhou, Y., Liu, X.C., Wei, R., Golub, A., 2021. Bi-objective optimization for battery electric bus deployment considering cost and environmental equity. IEEE Trans. Intell. Transp. Syst. IEEE 22 (4), 2487–2497.