This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house’s archive system.
Optimizing the deployment of electric vehicle charging stations using pervasive mobility data

Mohammad M. Vazifeh\textsuperscript{a}, Hongmou Zhang\textsuperscript{a,c,*}, Paolo Santi\textsuperscript{a,b}, Carlo Ratti\textsuperscript{a}

\textsuperscript{a}Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
\textsuperscript{b}Istituto di Informatica e Telematica del CNR, Pisa, Italy
\textsuperscript{c}Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Call detail records
Data-driven planning
Electric vehicle
Genetic algorithm

\textbf{ABSTRACT}

With the recent advances in battery technology and the resulting decrease in the charging times, public charging stations are becoming a viable option for Electric Vehicle (EV) drivers. Concurrently, emergence and the widespread use of location-tracking devices in mobile phones and wearable devices has paved the way to track individual-level human movements to an unprecedented spatial and temporal grain. Motivated by these developments, we propose a novel methodology to perform data-driven optimization of EV charging station locations. We formulate the problem as a discrete optimization problem on a geographical grid, with the objective of covering the entire demand region while minimizing a measure of drivers’ total excess driving distance to reach charging stations, the related energy overhead, and the number of charging stations. Since optimally solving the problem is computationally infeasible, we present computationally efficient solutions based on the genetic algorithm. We then apply the proposed methodology to optimize EV charging stations layout in the city of Boston, starting from Call Detail Records (CDR) of one million users over the span of 4 months. The results show that the genetic algorithm provides solutions that significantly reduce drivers’ excess driving distance to charging stations, energy overhead, and the number of charging stations required compared to both a locally-optimized feasible solution and the current charging station deployment in the Boston metro area. We further investigate the robustness of the proposed methodology and show that building upon well-known regularity of aggregate human mobility patterns, the layout computed for demands based on the single day movements preserves its advantage also in later days and months. When collectively considered, the results presented in this paper indicate the potential of data-driven approaches for optimally placing public charging facilities at urban scale.

1. Introduction

The level of discomfort drivers experience, alongside high costs, is one of the factors contributing to the slow adoption of plug-in all-electric vehicles (PEVs) (Chan, 1993, 2002; Deluchi et al., 1989; Wirasingha et al., 2008). Driver discomfort depends on a number of factors ranging from the limitations imposed by the short vehicle range to the significant amount of time and distance required to recharge batteries at charging stations. Current technological advances have increased the mileage and reduced the time required for a single full recharging to around 30 min (Chan et al., 2008; Foley et al., 2010; Dickerman and Harrison, 2010; Aggeler et al., 2010; Etezadi-Amoli et al., 2010). These trends, coupled with the increased push for environment friendly transportation modes, signal the

\textsuperscript{*}Corresponding author.

\textit{E-mail address:} hongmou@mit.edu (H. Zhang).

https://doi.org/10.1016/j.tra.2019.01.002
Received 2 January 2017; Received in revised form 19 September 2018; Accepted 3 January 2019
0965-8564/ © 2019 Elsevier Ltd. All rights reserved.
rapidly growing importance of public charging stations as a viable and alternative option to home-stations for charging EVs to further incentivize adoption (Eberle and von Helmolt, 2010; Nie et al., 2016; Lieven et al., 2011; Axsen and Kurani, 2013; Graham-Rowe et al., 2012). An important question to address in this regard is how a network of public charging stations should be deployed in order to minimize the drivers’ excess driving distance to charging stations in an urban environment and how big urban data can be used to make this planning more efficient.

As discussed in detail in the next section, there have been a number of studies aimed at introducing optimization frameworks for the deployment of charging stations. These studies use traditional methodologies developed in the field of facility location optimization, which are based on computer simulations of the traffic flows, travel surveys, or small sensor-based datasets. Both computer simulations and travel surveys are known to have major limitations in terms of scalability, accuracy, and/or cost—the computer simulations are modelled on agents’ behaviours from theoretical rules and aggregated data, while travel surveys are expensive to conduct, have lower spatial and temporal resolutions, and can only cover a small fraction of the population of a city. This explains the high interest in the possibility of accessing and analyzing large-scale real anonymized individual trajectories to feed the optimization process. This is possible nowadays given the availability of data sets of digital traces of individual-level human movements, such as those provided by cell phone records. These and other large-scale data sets collected by urban sensing devices have been used in a number of recent papers (Lee and Gerla, 2010; Miluzzo et al., 2010; Kwapisz et al., 2011; Ni, 2006; Boulos et al., 2011; Wang et al., 2011; Calabrese et al., 2010, 2013; Alexander et al., 2015; Jiang et al., 2013; Wang et al., 2012). However, to our best knowledge approaches based on large-scale trajectory data for optimally locating EV charging stations have not been systematically explored so far.

This motivated us to develop a data-driven optimization methodology to unravel the implications big data of human movement can have for planning at urban scale. The methodology is based on first constructing a model for charging station energy demand that takes into account the mobility patterns of real individual movements recorded in the region surrounding Boston. More specifically, we have used a massive data set of over one million cell phone users whose activities have been recorded over a 4-month period (Calabrese et al., 2013). The second step of the methodology consists in formulating a discrete optimization problem to find an efficient configuration for a network of charging stations. To this end, the region of interest is divided into a geographical grid, and an objective function is defined to simultaneously minimize the overall number of charging stations required and the aggregate distance EV drivers need to travel to reach the closest charging station. Minimizing aggregate distance not only reduces drivers’ excess driving distance to charging stations, but also minimizes the additional energy the EVs need to use to reach the charging station from their actual destination. We call this the energy overhead in the following.

Since optimally solving the problem at hand is computationally infeasible, we present computational efficient, near-optimal solutions based on greedy and genetic algorithms. The third and final step is assessing the performance and robustness of the presented approach, which is done by: (i) comparing quality of the obtained solutions vs. those provided by local optimization approaches and by current layout of stations in Boston; and (ii) showing that the best solution computed via genetic algorithm using single day movements preserves its properties also in later days and months.

The rest of the paper is organized as follows: In Section 2, we present an overview of related work. In Section 3, we formulate the problem and introduce the data-driven optimization framework. The optimization framework is then described in Sections 4–7: we describe the mobile phone data set used in the analysis, and how it has been processed to feed the optimization framework. In Section 8, we present and discuss the results of the proposed optimization methodology, including a comprehensive robustness analysis. Finally, Section 10 presents conclusions and discusses future research directions.

2. Literature review

Facility Location Optimization (FLO) problems have been extensively studied in the field of logistics and transportation planning over the past few decades (Aikens, 1985). With the exception of a few works mentioned below, to our best knowledge the methodologies used for FLO in the context of EV charging stations were based on theoretical models, simulations, and/or aggregate transportation data obtained from census tracts or travel surveys which often do not accurately model and represent the real-world scenarios of the spatial distribution of the demand. Some of these approaches were based on simulations of individual trip trajectories which required theoretical modelling of the behaviour of the service seeker agents or limited travel-surveys. A parking-based assignment introduced in Chen et al. (2013) is based on minimizing EV users’ stations access costs while penalizing the unsatisfied demand. The case study relied on 30,000 personal trip records based on household travel surveys in Seattle, US. In Liu et al. (2012), a multi-objective optimization problem is introduced considering various costs associated with introducing stations (including construction, operational costs, charging related costs). The aim of this study was to minimize the costs with consideration of charging demand distribution obtained from aggregate but static transportation data for Beijing.

In another work, the authors of Lam et al. (2013) proposed an optimization problem from the EV driver’s perspective, where the objective is to minimize users’ discomfort measured as the distance between the charging demand spots and the location of the stations.1 This study used static population and income-level in Hong Kong as an input to model the spatial distribution of charging demand. In Hess et al. (2012), EV traffic and drivers’ behaviour have been simulated on the real road network of Vienna. Based on a linear battery depletion model, an approach is proposed to optimize the location of charging stations for a limited number of EVs with

---

1 We also follow the same definition of discomfort and only consider the discomfort due to the excess driving distance to reach charging stations. To clarify the concept, we only use the words ‘excess driving distance to charging stations’ instead of discomfort.
the objective of minimizing the overall travel time. Sweda et al. (2011) used an agent-based decision support model to identify patterns in EV ownership and driving activities. Using this model they assessed the efficiency of various charging stations deployment by quantifying drivers’ inconvenience defined as the distance they must travel to visit nearest charging stations.

In Frade et al. (2012), a study on optimizing the location of EV charging stations for a neighbourhood in Lisbon is presented based on the estimation of the refuelling demand through the application of a maximal covering model. In this study, the refuelling demand estimation was based on static census data. In Ip et al. (2010) a simple assignment model is formulated to locate charging stations to demand clusters of various sizes identified through a hierarchical clustering analysis based on the aggregate traffic data obtained from traffic monitoring devices. There has been a few studies based on limited individual-level trajectory analytics. Ahn and Yeo (2015) proposed a simplified fuzzy model in which an optimal density map of charging stations is obtained based on taxi trajectories to help efficient allocation of the charging stations to minimize the driving-range anxiety. In another work, an optimization framework was introduced to deploy charging stations and to assign optimized number of plugs based on real EV taxi-trajectory data, with the objective of minimizing the average time to find a charging station and waiting time before charging (Li et al., 2015). Using LP-Rounding, the authors provided approximate solutions to the NP-hard optimization problem. Dong et al. (2014) proposed an activity-based approach based on a small GPS based travel survey data to optimize the location of the charging stations in metro Seattle. Cai et al. (2014) examined taxi trajectory data to study public charging stations planning through estimation of the demand based on the taxi stops in Beijing. The data sets used in these studies represent only a small fraction of the private and public mobility demand and/or they often are non-representative of the entire demand, which can be better characterized by means of cell phone data (Alexander et al., 2015) as it is done in this paper.

Other studies have proposed multi-objective optimization problems to locate charging stations. In Wang and Wang (2010) the authors proposed a model based on set-cover with dual objectives of minimizing the overall costs of opening new charging stations and maximizing their overall coverage on Taiwan road network based on static population density data. Wang and Lin (2013) used set- and maximum coverage to formulate capacitated multi-type charging stations allocation model using a demand simulation based on a vehicle-refuelling logic. In Xi et al. (2013), a linear model is used to estimate EV penetration in various sub-regions to determine the volume of EV flows between the sub-regions based on demographic data. This simulation model is then fed into an optimization problem to obtain the location of charging stations considering an objective function which maximizes the EVs that charge at the new public stations. Other simulation-based studies have considered multi-objective planning by considering both the impact on the power grid and drivers satisfaction (Bayram et al., 2013; Zhang et al., 2014; Wang et al., 2013; AU Dashora et al., 2010).

A number of studies have focused on the coupling between the transportation and power networks. In He et al. (2013) an optimization problem is introduced to allocate public charging stations to maximize the social welfare associated with the coupled transportation and power network. In Bayram et al. (2015) an optimization framework is proposed to maximize the quality of service for drivers while preventing potential grid failures using optimal distribution of resources. While most of the above mentioned studies have relied on static demand simulation, there has been a few studies focusing on simulation-optimization allocation of charging stations by considering stochastic dynamic demand simulations. In Jung et al. (2014) an itinerary-interception allocation model is proposed and tested on a fleet of 600 shared-taxis in Seoul to minimize the queue delay and maximize overall level of service. In the cases where charging times are very short (ξ ≤ 5 min), to optimize locations of refuelling stations, Kuby and Lim (2007; Kim and Kuby (2012) have introduced Flow Refuelling Location Model (FRLM), a flow interception approach to locate refuelling stations in a way to maximize the captured flow by modelling flows on the road network starting from origin-destination (O-D) matrices. Although these flow-interception approaches are applicable to optimize the station locations for a fixed number of stations when charging times are short (e.g., for hydrogen stations, with charging times of around 5 min), they may not be suitable for EV station locations optimization and to find the minimum number of required charging stations to cover the demand where charging times are still regarded significant (ξ ≥ 30 min).

The aforementioned works consider various aspects and are important first steps towards constructing a comprehensive optimization framework for efficient planning of EV charging stations. In this work, we take a step further by proposing the first optimization framework based on modelling the demand via CDR. In recent years, large-scale CDRs have been shown to be accurate sources for modelling the individual trip origin/destination flows in urban settings (Alexander et al., 2015; Calabrese et al., 2010; Jiang et al., 2013; Wang et al., 2012). Further, Pappalardo et al. (2015) investigated CDR along with GPS trajectories, finding evidences of two typical human mobility profiles called returners, for whom a small number of locations could approximate their overall mobility patterns, and explorers, whose location visits are very sparse and diverse. This and the other mentioned works showed the validity and prospective of using CDR for human mobility and travel demand prediction. Our paper provides evidences that data-driven optimization based on CDR has the potential to emerge as a prominent approach for many urban facility location problems, of which EV charging station is a first example.

3. Introducing the framework

We introduce our data-driven optimization framework using metro Boston as a case study and by analyzing the individual movement patterns of around one million individuals (see Fig. 1) obtained in this region through CDR over the span of 4 months. For each individual record, the data set contains the location and time information when the person makes a call, text or uses data service. Using the road network and based on the shortest path, individual trajectories can be reconstructed with good accuracy. Starting from these trajectories and by dividing the region of interest into a geographical grid, we formulate a discrete optimization problem equivalent to the well-known set cover problem (Cormen et al., 2009). The objective is to minimize the total number of charging stations needed and at the same time to minimize the average distance travelled by drivers on their routes from the end of
their intended trips to closest available charging stations in order to help reduce “range anxiety.” As commented in the introduction, the aggregate distance is also a measure of the energy overhead induced by a specific charging station layout. Since the set cover problem is NP-hard (Cormen et al., 2009), we use a combination of Chvatal’s greedy (Chvatal, 1979) search and Genetic Algorithm (GA) (Beasley and Chu, 1996) to approximate a solution for EV charging station locations that significantly improves the average distance while reducing the number of charging stations.

The overall optimization framework can be divided into the following stages:

(A) **Demand Modelling:** We construct an energy demand model for EVs based on real individual trip trajectories and the road network.

(B) **Formulating the Optimization Model:** We formulate the location optimization problem as a set cover problem.

(C) **Analyzing the Data:** We process the raw data to estimate EV energy demand based on a simple assumption for EV adoption rates.

(D) **Optimization Methods:** We perform a combination of Chvatal greedy and GA meta-heuristic search methods to find efficient layouts of EV charging stations.

---

**Fig. 1.** Individual-level location traces: The individual location traces across space shown for two distinct users (red and green dots). Each dot represents the location and time of the logged network connectivity. The location of user is estimated based on the AirSage’s Wireless Signal Extraction Technology. The vertical dimension represents time (an entire week in July 2009) in increasing order starting from the plane. These individual-level mobility patterns are the basis of the optimization framework introduced in this work. The optimization objective is to minimize the total traveled distance to closest charging stations from the demand spots modelled based on where people stay for long enough time to allow charging. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

---

**Table 1**

Notations used in this paper.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_j \in C/C_N$</td>
<td>Partitioned cells in the study area $C$, and the corresponding node in the cell network $C_N$</td>
</tr>
<tr>
<td>$\Omega_{ij}$</td>
<td>Number of trip segments ending in cell $C_i$ during time interval $[t, t+\Delta t]$</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of trip segments</td>
</tr>
<tr>
<td>$h$</td>
<td>Trip distance traveled and the minimum distance that a trip needs to pass before charging</td>
</tr>
<tr>
<td>$l_{\text{min}}$</td>
<td>EV penetration rate, i.e., the share of total travel demand realized by EVs</td>
</tr>
<tr>
<td>$D_{ji}$</td>
<td>Shortest path distance on the road network from node $C_j$ to $C_i$</td>
</tr>
<tr>
<td>$P(h)$</td>
<td>Index set of cells which can be covered by $C_i$ with distance $h$</td>
</tr>
<tr>
<td>$w_i$</td>
<td>Excess driving distance to charging stations imposed on drivers by locating charging station at $C_i$, measured by additional distance traveled</td>
</tr>
<tr>
<td>$\mathcal{U}$</td>
<td>The node set of network $C_N$</td>
</tr>
<tr>
<td>$x_i$</td>
<td>The set of nodes that $C_i$ can cover</td>
</tr>
<tr>
<td>$S$</td>
<td>The total set of nodes in $C_N$</td>
</tr>
<tr>
<td>$w(s_i)$</td>
<td>Total excess driving distance to charging stations of the drivers in the cells covered by $C_i$</td>
</tr>
<tr>
<td>$w_0$</td>
<td>System parameter to reflect the cost of deployment of a new charging station</td>
</tr>
<tr>
<td>$x_i$</td>
<td>Binary variable in the optimization problem indicating whether there is a charging station at $C_i$</td>
</tr>
<tr>
<td>$n_t$</td>
<td>The number of EVs that a charging station can serve at a time</td>
</tr>
<tr>
<td>$k_i$</td>
<td>The minimum number of charging stations needed in the h-adjacency of cell $C_i$</td>
</tr>
<tr>
<td>$SCP[i]$</td>
<td>Binary matrix indicating whether $C_i$ covers $C_j$ with distance $h$</td>
</tr>
<tr>
<td>$c_t$</td>
<td>A cellphone usage record at time $t$ in cell $C_i$</td>
</tr>
<tr>
<td>$u_i$</td>
<td>User index</td>
</tr>
<tr>
<td>$H(u_i)$</td>
<td>Home location of user $u_i$</td>
</tr>
<tr>
<td>$D_k$</td>
<td>The total traveled distance from last charging to the current cell $k$</td>
</tr>
</tbody>
</table>
Each step will be described in a separate section. To begin with, Table 1 lists all the notations used in this paper.

4. Demand modelling

We assume the urban area is partitioned into a number of non-overlapping square cells \( C = \{C_1, C_2, \ldots, C_N\} \), e.g., corresponding to a square grid partitioning. Similar to space, time is also divided into a number of non-overlapping intervals, corresponding to, e.g., \( \Delta t = 30 \text{ min} \) slots. In the following, we use \( i \) as a generic cell index and \( t \) as a generic time slot index. Equipped with the above definitions, we now define the following quantity for each spatio-temporal slot \((i, t)\):

\[
\Omega_{i,t} = \text{number of trip segments ending in cell } C_i \text{ during time interval } [t, t + \Delta t)
\]

A trip segment is the travel between two consecutive stays identified from CDR data. For example, if from a trajectory we observe a person to stop at \( C_i \) and after 5 min at \( C_j \), without other intermediary stops at other locations, then we consider the displacement of the user to be a trip segment ending in \( C_j \). We concatenate consecutive trip segments if the stay time between them is less than \( \tau_{\text{min}} \), where \( \tau_{\text{min}} \) represents the minimum time interval needed for completely charging an EV, and is set to 60 min in the following. The trip segments counted in \( \Omega_{i,t} \) are then those at which ends it is possible to completely charge the EV.

Note that, if we assume that the duration of a time slot \( \Delta t \) is \( \leq \tau_{\text{min}} \), all segments contributing to \( \Omega_{i,t} \) belongs to different vehicles. \( \delta \) is EVs penetration rate, i.e., a fraction \( 0 < \delta \leq 1 \) of \( \Omega_{i,t} \), i.e., \( \delta \Omega_{i,t} \), can be defined as the number of candidate EVs arriving and potentially demanding fast refuelling of their batteries at \((i, t)\). We construct this refuelling demand model based on the durations of the stays at various cells, and by assuming a linear energy consumption model with the trip length \( l \) computed by considering the real road network and based on the shortest path between the origin and the destination of each trip segment. More specifically, \( u_c(t) = (u_c(0) - c \cdot l) \Theta(u_c(0) - c \cdot l) \) where \( u_c = 1 \) when the EV is fully charged and 0 when the battery is empty of charge and \( \Theta(x) \) is a step function which is zero for \( x \leq 0 \) and 1 when \( x > 0 \). The constant \( c \) depends on a number of parameters including road conditions, traffic flow, driver behaviour, and vehicle type (Peaure et al., 2011; Nie and Ghamami, 2013). For simplicity, we assume that the battery depletion on average is a linear function with distance travelled, and that the range is \( 250 \text{ km} \). In order to have a better estimate of refuelling demand at each cell, at each time slot we only count trip segment end points with trip lengths equal to or larger than \( \tau_{\text{min}} \sim 100 \text{ km} \). Also, we exclude trips of length above the EV range of 250 km. We obtain the number \( \Omega_{i,t} \) of needed “chargeable” parking spots at each cell according to these considerations.

It is important to note that, the penetration factor \( \delta \) is actually a compound of three distinct factors: (1) the ratio between CDR-derived travel demand and the actual travel demand; (2) the share of private vehicle trips in all travel demand; and (3) the share of EVs in all private vehicles. Further, \( \delta \) may also vary at different locations in a city due to the spatial heterogeneity of the aforementioned three factors. Nonetheless, we would like to point out that we need \( \delta \) only as inputs of our algorithm for an estimate of the EVs that needs to charge, and the model we propose will be independent of, and can work with any given \( \delta \) or \( \delta \Omega \). A more detailed approach would be to consider all the three factors, and the spatial heterogeneity in the penetration rate (Xi et al., 2013), which can be studied as an extension to this work in future. Moreover, the penetration rate of EVs will also change due to the technological advance and better public acceptance, leading to an increasing \( \delta \). In the rest of the paper we assume \( \delta = 0.25 \), while acknowledging that in practice of implementing this algorithm a better estimate of the parameter can be based on the best knowledge of the implemented city and time.

5. Formulating the optimization model

Based on the topology of the cell partitioning, we build a cell network \( C_N \) as follows. We add a node for each cell \( C_i \in C \), and we add a link between \( C_i \) and \( C_j \) if the two cells are spatially adjacent to each other. For instance, if \( C \) corresponds to a square grid partitioning, the node corresponding to cell \( C_i \) in \( C_N \) is connected to the nodes corresponding to the cells adjacent to \( C_i \) in the North, East, South, and West direction if there is a road segment in the road network that connects these cells. We now introduce a parameter \( h \) to notate how far a driver is willing to drive to reach a charging station from their destination point. Having set \( h \), we say that a charging station located at \( C_i \) covers a cell \( C_j \) if the respective nodes in \( C_N \) are at a distance equal to at most \( h \). In turn, we say that \( C_i \) is covered if there exists at least one cell at within distance \( h \) from \( C_i \) that hosts a charging station. The tradeoff between drivers’ excess driving distance to reach charging stations and the coverage is now clear: if \( h = 0 \), drivers’ excess driving distance to charging stations is minimum, but a charging station located at \( C_i \) covers only the cell itself. With increasing values of \( h \), we have a higher excess driving distance to charging stations, but coverage of charging stations increases as well, and less charging stations are required to cover the entire city. Each node in \( C_N \) is then weighted as follows. Let \( N^h \) be the set of nodes at distance \( j \) from node \( C_i \) in \( C_N \). The weight \( w_i \) of node \( C_i \) is computed as

\[
w_i = \frac{1}{m} \sum_{j=1}^{m} \sum_{j \in P(h)} (\Omega_{j,t} - h) \cdot \delta \Omega_{j,t}.
\]

Weight \( w_i \) is designed to model the excess driving distance to reach charging stations imposed on drivers by locating charging station at node \( i \). \( P(h) \) is the set of indices of cells within distance \( h \) from \( C_i \). For instance, it is 0 for drivers arriving in \( C_i \), 1 for drivers

\footnote{The shortest path is computed considering the origin and destination cell centroids as starting and ending point.}

\footnote{To simplify notation, in the following \( C_i \) denotes both a cell and its corresponding node in the cell network \( C_N \).}
arriving in a cell \( C_j \) at a distance 1 km from \( C_i \), etc. Therefore, the total driving distance to charging stations is obtained by summing up the number of drivers arriving in cells at a certain distance from \( C_j \), up to the maximum value of \( h \). The distance matrix components \( D_{j,i} \) are computed by considering the shortest path on the real road network between the centroid of \( C_i \) and \( C_j \). Furthermore, this quantity is averaged across the total number of time slots \( k \) considered in the analysis. Note that as we only consider cells in the coverage range of \( C_i \) in the computation of \( w_i \), all the weights are negative considering the fact that \( D_{j,i} \leq h \) for all \( j \in \mathcal{P}(h) \).

We are now ready to formulate the charging station optimization problem. The problem of finding the optimal location of charging stations can be formulated as a weighted set-covering problem (Gorman et al., 2009) in which the universe set, \( \mathcal{U} \), and the set of subsets, \( \mathcal{S} \), are given as follows

\[
\mathcal{U} = \{ C_1, C_2, \ldots, C_N \},
\]

\[
\mathcal{S} = \{ s_1, s_2, \ldots, s_N \},
\]

where \( s_i \) is defined as

\[
s_i = \{ C_j | j \in \mathcal{P}(h) \},
\]

i.e., the set of all cells within distance \( h \) from \( C_i \) in \( C_N \). Each \( s_i \) is weighted with \( w(s_i) = w_l + w_o \), in which \( w_0 \) is a measure of the total excess driving distance to reach charging stations by the drivers in cells covered by \( C_j \) and \( w_0 > 0 \) is an offset reflecting the cost of deployment of a new charging station facility. As discussed in the Results section, \( w_0 \) can be tweaked starting from 0 and by increasing its value in order to reduce the number of stations significantly without compromising the minimization of the average traveled distance to closest charging stations from demand spots. The optimization problem is to find a subset \( S_{opt} \) of \( S \) such that all elements of \( \mathcal{U} \) are covered, and the sum of the weights in \( S_{opt} \) is minimized. Note that when \( h = \infty \), each of the \( s_i \)'s coincides with the universe set, therefore the optimal \( C \) is the \( s_i \) that has the lowest \( w_0 \). When \( h = 0 \), \( s_i = \{ C_i \} \) and all the weights are zero, therefore optimal \( S_{opt} = S \) itself, implying that there must be a charging station at every cell. We are then interested in investigating the non-trivial case in which \( 0 < h < \infty \).

The weighted set covering problem described above can be formulated in terms of the following constrained integer linear optimization problem (ILP):

Minimize \( \sum_{i=1}^{N} w(s_i) \cdot x_i \),

subject to:

\[
p_i = \sum_{j \in \mathcal{P}(h)} x_j \geq 1 \quad (i = 1, \ldots, N),
\]

\[
x_i \in [0, 1].
\]

Note that \( p_i \) is the number of available stations in the \( h \)-proximity of \( C_i \). In practice, there is a finite capacity on the number of EVs a charging station can handle at a time, i.e., \( n_c \). If we require that the capacity is not exceeded, then one needs to make sure that the following condition is satisfied for every cell:

\[
p_i \geq \text{ceiling} \left( \frac{\max(\Omega_{i,t})}{n_c} \right) = k_i.
\]

where \( \max(\Omega_{i,t}) = \max(\Omega_{i,t}) \) is the maximum value of \( \Omega_{i,t} \) during the observed time interval. Interestingly, this modifies the inequality constraint in a simple way. The difference is that instead of \( \geq 1 \), we have \( \geq k_i \). Therefore, the problem becomes:

Minimize \( \sum_{i=1}^{N} w(s_i) \cdot x_i \),

subject to:

\[
p_i = \sum_{j \in \mathcal{P}(h)} x_j \geq k_i \quad (i = 1, \ldots, N),
\]

\[
x_i \in [0, 1].
\]

In the following, we assume that the capacity of charging stations is large enough to address all energy demand in the coverage range, which is equivalent to setting \( k_i = 1 \) for all \( C_i \)’s. As discussed in Appendix B, the model can be easily generalized to consider capacitated case for the charging stations. In this way, the number of EV charging plugs required in each cell can be optimized to maximize the coverage and to distribute the load evenly between the charging stations. After finding the GA-optimal stations layout it is possible to estimate the average number of plugs required by analyzing the temporal patterns of the number of requests each charging station receives for charging as discussed in Appendix B.

The set cover problem can be represented also in a binary matrix form—useful for GA implementation—as follows. An \( N \times N \) binary matrix is formed where rows are associated with cells that need to be covered, and columns are associated with stations at various cell locations. Now the \((i, j)\)-element of this matrix is either 1 or 0 depending on whether the station at \( C_j \) covers \( C_i \) or not. This
binary matrix, which we call SCP, along with a weights array \( W \) uniquely represents the set cover problem:

\[
SCP_{ij}^h = \begin{cases} 
1 & j \in \mathcal{P}(h) \\
0 & \text{otherwise}
\end{cases}
\]

(13)

Each configuration of charging stations in this binary representation is a binary vector in which the indices of nonzero elements correspond to indices of the cells which have charging stations in them.

It is also important to note that some of the cells in the partitioning have zero energy demand in the observed time period. Therefore, we can reduce the problem size by removing the associated rows/columns.

6. Analyzing the data

6.1. Data

The data set consists of estimated anonymous location traces from approximately one million users in the Boston Metropolitan Areas, Massachusetts, USA collected by a cell-phone provider over the span of 4 months in 2009. The CDR contains connection events, which are recorded when a text message is sent or received, when a call is placed or received, or when the user connects to the Internet through the cell-phone network. The location estimation is generated through triangulation by means of the AirSage’s Wireless Signal Extraction Technology. We closely follow the noise reduction method technique used in Calabrese et al. (2013) by fusing points that are within certain distance threshold (< 500 m) consistent with the accuracy of the distance measurements based on tower communication signal triangulation. We choose the cell size 1 km \( \times \) 1 km to be consistent with the highest spatial resolution that can be achieved considering the amount of uncertainty in the location measurements using CDR (±500 m). The average inter-event time measured for the whole population in the time window of interest is approximately 260 min. However, the distribution of the inter-event times has an arithmetic average of medians which is around 84 min. These inter-event times are much shorter after filtering outliers and during the period of high activities during the day in a way that it allows reconstruction of individual trajectories with high accuracy for the majority of users in the time scales relevant to our demand modelling (see Fig. 2). This temporal resolution enables us to detect significant portion of the events associated with location changes of the users and therefore reconstruct individual trajectories based on the shortest path approximation on the road network. The data set is then suitable to characterize individuals’ mobility behaviour, which is needed to feed into the proposed optimization framework. It is important to note that the closeness of the reconstructed paths to the actual paths depends on the temporal resolution/granularity of the sampling. As the sampling rate increases due to increase in the frequency of network connectivity especially for internet usage, higher resolution CDR becomes available and therefore reconstructed individual trajectories better represent actual paths.

6.2. Processing call detail records

The first data processing step is to identify mobility from the detailed records of cellphone usage. If a continuous sequence of records occurring in the same cell are observed, the user is assumed to stay in that cell from the earliest time in the sequence to the

Fig. 2. Temporal analysis: The left panel represents the two dimensional distribution of the average inter-event times \( \Delta t \) and the number of events for each individual \( N_{\text{events}} \) during the span of one week (for more than 500,000 users). The right panel represents the cumulative distribution of users versus the average inter-event times \( \langle \Delta t \rangle \) for various lower bounds used for the number of events. This plot shows for significant number of users in this dataset we have relatively high sampling rates to reconstruct individual trajectories. The vertical lines represent the mean corresponding to each curve (with the same colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
latest. This implies that the records in the middle of the sequence are not relevant for our analysis. Based on this observation, we reduce the list of records from

\[ c_{t_0}^{i_0}, ..., c_{t_k}^{i_p}, ..., c_{t_{k+m}}^{i_p}, ..., c_{t_{n-1}}^{i_q}, ..., c_{t_n}^{i_q} \]

where \( c_{t_k}^{i_p} \) represents a cellphone usage recorded at time point \( t_k \) in cell \( C_{i_p} \), to a shorter list

\[ c_{t_0}^{i_0}, c_{t_k}^{i_p}, ..., c_{t_{k+m}}^{i_p}, ..., c_{t_{n-1}}^{i_q}, c_{t_n}^{i_q} \]

obtained by removing in-cell records. The list is then transformed into the following list of tuples

\[ (i_0, t_0, t_w), ..., (i_p, t_k, t_{k+m}), ..., (i_q, t_{n-1}, t_n) \].

To simplify notation, we relabel this list with continuous subscripts according to their orders in the list, and label the starting and ending times of a location with superscripts ‘start’ and ‘end’

\[ (i_0, t_{0}^{\text{start}}, t_{0}^{\text{end}}), (i_1, t_{1}^{\text{start}}, t_{1}^{\text{end}}), ..., (i_w, t_{w}^{\text{start}}, t_{w}^{\text{end}}). \]

6.3. Home detection

For each user in the data set, we then identify a home cell amongst the list of visited cells. This step is necessary because we assume that a user does not need to recharge their EV at a public charging station if they are at home. Therefore, the arrivals at home location cells will not be counted in \( \Omega_{it} \).

Fig. 1 illustrates two sample users’ location traces during a week using the vertical axis as time. Each cellphone record is a point on the map based on its location (x,y)-axes) and the time of occurrence (z-axis).

Following the methodology proposed by Calabrese et al. (2013), the home location \( H(u_j) \) of a user \( u_j \) is determined by finding the cell where he/she cumulatively spends most of the time from 8 p.m. to 6 a.m. in a week, i.e.,

\[ H(u_j) = C_q, \quad q = \arg \max_p \sum_{x_{s_{p}}=a_p} (t_{s_{p}}^{\text{end}} - t_{s_{p}}^{\text{start}}) \]

where \( t_{s_{p}}^{\text{start}}, t_{s_{p}}^{\text{end}} \) are only considered if they are between 8 p.m. and 6 a.m. In Fig. 3 we show the spatial distribution of the identified home locations with a sample of a week’s data.

6.4. Counting the number of trips ending in a cell

The process of counting \( \Omega_{it} \) is implemented with a linear search. For each user, starting from the beginning of the reduced list of records, we track the total distance the user has traveled from the last charging cell to the current cell through a path along all the intermediate cells without a charging opportunity, i.e., with stays shorter than 60 min. For example, if the last charging cell is \( C_{i_s} \), since then the user has gone through \( C_{i_s}, ..., C_{i_{k+1}} \), while the current cell is \( C_{i_{k+1}} \), the total tracked distance is

\[ D_{i_{k+1}} = \sum_{p=3}^{x+k-1} D_{i_p|i_{p+1}} \]

where \( D_{i_p|i_{p+1}} \) is the shortest path distance from the centroid of \( C_p \) to the centroid of \( C_{i_{p+1}} \).

If during the tracking process, the user stays for longer than 60 min at a non-home cell \( C_i \) and the total traveled distance is at least

![Fig. 3. Current charging stations layout and the home population: The blue cells in the left panel corresponds to grid cells in which at least one charging station exists as obtained from http://openchargemap.org/ for a region centred around Boston. The colour in the right panel represents the number of individuals who reside in each cell which has been estimated using the home-detection algorithm used throughout this work. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
$l_{\text{min}} = 100$ km, the trip is counted in $\Omega_{i,t}$ for the respective time period $t$. Then the tracking distance will start from 0 again, until next charging cell is found.

This process is repeated for all the users in the data set, so that the $\Omega_{i,t}$ values of all cells for all time periods are determined.

7. Optimization methods

Since optimally solving the set-covering problem is computationally infeasible, in the following we present a combination of two optimization methods to produce efficient solutions.

7.1. Greedy search

The first approach is based on a greedy search (Chvatal, 1979), which is used to find a set of column indices that cover all the rows in the binary matrix representation. The algorithm is summarized in the pseudo-code reported in Algorithm 1. In this greedy search method, at each optimization step the location of a charging station is chosen by sorting all the available cells according to their average weight per uncovered cell in their coverage range, and a station is placed in a cell corresponding to the lowest value of the average weight per uncovered cell. This process is repeated until all cells are covered. Once the greedy coverage phase is terminated, the algorithm then checks whether cells hosting redundant charging stations exist or not, and, in case they exist, it removes them. By eliminating redundant cells, the number of stations will be subsequently reduced.

7.2. Generating initial GA population

The solution provided by the greedy search method is known to be inferior to those provided by more sophisticated heuristic/meta-heuristic methods (Beasley and Chu, 1996). In this section, we present a Genetic Algorithm (GA) approach to improve the average distance drivers need to travel to the charging stations from their demand points while minimizing the number of charging stations that are required to address all the demands.

The first step is randomly constructing a population of possible layouts by stochastically and uniformly distributing charging stations until all the cells are covered, and then removing the redundant cells that most contribute to the overall objective function. This way, we can generate a population of locally optimized layouts for charging stations that can be used as the starting point of the GA. Another approach to generate a population of the solutions is a generalization of Chvatal’s search method by adding stochasticity to the search scheme. In this probabilistic version of Chvatal’s algorithm, deterministic choice of the best possible cell at each stage is replaced by random choice of a cell from a small subset of $k$ elite locations, which are the best $k$ possibilities out of all available cells at each step according to how locations are sorted in Chvatal’s search. The required change in the algorithm is explained in pseudo-code reported in Algorithm 2. This way, the solution is no longer deterministic, and each execution of the algorithm results in a different solution. We repeat this process until a population of a desired size is produced. As one increases $k$, the search space becomes more diverse; however, the average quality of the solutions degrades. On the other hand, decreasing $k$ reduces the size of the search space; however, the average quality of the population is higher. In the following, we set $k = 5$, which represents a good compromise between size of the search space and population quality.

After the initial population is generated, we need to define rules to replace less fit members of the population with new offsprings. For this purpose, we closely follow the strategy presented by Beasley and Chu (1996), which uses a binary tournament selection and a fusion crossover operator to evolve the population. The relevant pseudo-code is reported in Algorithms 3 and 4. This way the average weight of the population improves and finally the population converges to a high quality set of solutions with a significantly lower average distance to the charging stations than that of the best member of the initial population in the majority of cases – see next section for details. It is important to note that new binary vectors produced using the fusion crossover operation are not necessarily feasible layouts as they may contain redundant cells or uncovered cells. We need to perform a local search to find a close feasible solution in which all the demand cells are covered and there is no redundancy in the layout. This part is the most computationally intensive part of the algorithm, and it is considered as a necessary step to make sure that the minimum number of stations is used to cover the demand. What we do in this step is simply finding all the uncovered cells, and then performing a greedy Chvatal’s approximation to cover the remaining cells as we described in the previous section. Then, we remove redundant stations in exactly the same fashion as described in the redundant cell removal stage of Algorithm 1. To diversify our search space further, we use constant mutation operator which mutates one of the components in the newly generated vector (corresponding to a charging station configuration) in the binary vector representation. The component is chosen among the ones that two parents share with each other. We run the GA by generating 20,000 children and replacing them with the less fit members of the population. The initial population evolves to a new population after each iteration, till the total number of 20,000 iterations is reached.

8. Results

Fig. 4 shows the actual charging stations layout in Boston in comparison with the best configuration obtained using the optimization framework developed here for various values of the coverage range $h$, where the $\Omega_{i,t}$ values used in the optimization are computed from the movement data of a single day in July 2009. As one expects, by increasing the value of $h$, a lower number of charging stations is required to cover all the demand spots. Interestingly, we observe that, while the solutions obtained for a relatively larger value of $h$ contains relatively less stations as expected, the computed charging station location sets are not subsets of those
representing previous solutions, indicating the non-trivial combinatorial structure of the problem at hand. However, there are certain cells with a charging station in all the layouts indicating a robustness with respect to change in optimization and demand modelling parameters.

We can also compute the average distance to a charging station from each demand spot, and compare the initial and final

Fig. 4. The current layout vs optimization layout: Layout of charging stations as computed by the GA for different values of stations coverage \( h = 1.5, 2.0, 2.5, 3.0 \) km in comparison with the actual layout for a geographical region centred around Boston (data obtained from \( \text{http://openchargemap.org/} \)). The red colour represent locations where according to the optimization there must be a charging station. The parameters used are \( l_{\text{min}} = 100 \) km, \( \tau_{\text{min}} = 60 \) min, \( w_0 = 12|\bar{w}| \) where \( \bar{w} \) is the average weight of the cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. The average distance vs the number of charging stations: The average distance from the demand spots to the charging stations versus the number of charging stations. Each point represents a charging stations layout. The clusters with darker colours correspond to layouts obtained using GA. The other clusters are those obtained using a local optimization. Left and right panels differ only in the value of the offset used in the objective function (left: \( w_0 = 0 \), right: \( w_0 = 12|\bar{w}| \) where \( \bar{w} \) is the average weight of the cells). Increasing \( w_0 \) decreases the number of charging stations required in the layout as it increases the cost of adding a new station. The black circle represents the average distance and the number of charging stations for the current layout of charging stations according to the data obtained from \( \text{http://openchargemap.org/} \). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
populations of configuration layouts. In Fig. 5, we report a scatter plot where each point represents a charging station configuration, and the horizontal and vertical axes represent the average distance and the number of charging stations required to cover the demand spots. Different colors correspond to different values of $h$. As one can see, for any value of $h$ the final population (smaller clusters) has significantly lower average distance to charging stations compared to initial populations (larger clusters of points) while requiring significantly less number of charging stations when the offset value is properly set. Considering $w_0 = 0$ leads to maximum reduction in the average distance to proximate charging stations, however, the number of stations is not necessarily minimum. In order to achieve a significant reduction in the number of stations without significantly compromising the achieved reduction in the average traveled distance to closest charging stations, we tweak $w_0$ parameter to achieve the highest reduction in the number of stations while the drop in the distance gain compared to the theoretical value for $h = 2$ km remains under 25% of the gain at $w_0 = 0$.

To further quantify how much improvement can be gained using demand-based, data-driven optimization framework introduced here in terms of drivers’ excess driving distance to charging stations, we consider the cumulative distribution of the distance from each demand spot to the closest charging station and compare it with the current layout of charging stations obtained from the openchargemap.com open database. Fig. 6 shows that all the demand spots in the optimization case are covered with a station within the coverage distance $h = 2$ km, while in the current layout of charging stations, a significant fraction of the demand spots are not covered by any station in their close proximity, even though more charging stations are used in the current layout as opposed to one obtained from optimization (158 vs. 154—see Fig. 5). This result supports the argument that data mining approaches in which individual level human movement are analyzed in cities may reveal non-trivial travel patterns which cannot be predicted by simply considering the static residential population distribution data which most likely have been used in the process of planning the current deployment of charging stations, as it can be seen from higher density of charging stations in the regions with higher population.

An important aspect to investigate is whether or not the charging station layout computed using data from a single day remains near-optimal in the following days and months. In other words, we would like to assess the robustness of the computed solution in presence of variation of the input data, which is especially important in infrastructure planning processes given the high costs and long lifetime of infrastructure. To this purpose, we have used the best charging station configuration obtained using a single reference day in July 2009 – call it the reference layout, kept the obtained number and location of charging stations fixed, and assessed the...
quality of this solution when the $\Omega_i$ values are computed from data obtained from different days than that used to compute the reference layout. In particular, we considered various days in a week selected from the months of July, September, and October 2009. To estimate the quality of the reference charging station layout in the different days, we have computed the average distance to the closest charging station and its variance for each day, and compared to the average distance traveled in the reference day. Fig. 7 reports the results of the analysis. The plot clearly shows that the average traveled distance to the charging station remains extremely stable over time. While it is possible that a slightly better layout for charging stations existed for each specific day, leading to slightly lower average distances, the fact that the traveled distances to the charging stations in the reference layout remain stable, combined with the fact that aggregate mobility demand is also relatively stable over time (Sagarra et al., 2015), clearly indicate near-optimality of the reference layout also for the other days.

Another way to investigate the robustness of our optimization approach is to tune the parameters of the EV demand to see how sensitive the average driven distance is with respect to a change in these parameters. In Fig. 8, we take the $h = 1.5$ km case as a reference, consider the layout configuration obtained for $r_{\text{min}} = 60$ min and $l_{\text{min}} = 100$ km, and recompute the average distance based on the demand for different values of $r_{\text{min}}$ and $l_{\text{min}}$. Again, the gap in the average distance between a locally optimized layout and the best layout obtained using GA shows significant robustness in the space of these parameters. This shows that the computed charging station configuration remains stable not only across time, but also in case technological development significantly change charging times and EV ranges. Finally, based on the analysis of the temporal patterns of charging requests each station receives on average during the day, we can show how the load is distributed more evenly for GA-optimized layout in comparison with the current layout of charging stations as it can be seen in Fig. 9.

9. Benefit and limitations of using CDR

The use of Call Detail Records (CDR) in transportation research has seen a significant increase over the past years. This is mainly due to the promise that it provides for large-scale analysis of individual movement patterns for use in transportation planning decisions. However, similar to any other available urban movement dataset, CDR comes with certain limitations such as relatively...
high uncertainty in the exact location estimation compared to more accurate methods such as GPS and the spatio-temporal sparsity due to the large inter-event times as well as the difficulty in identifying the transportation modes used in each movement activity. The good news is that recent methodological developments promise to at least partially address some of these challenges associated with accurately inferring individual movements and the modes of transportation from mobile phone data at urban and regional scales. Furthermore, better location tracking methods and data are becoming available with significantly higher temporal and spatial resolutions thanks to pervasive sensor technologies. The same methodologies introduced in this paper could be used off-line and/or in real-time to optimize infrastructure planning and/or to evaluate the performance of a deployed/to-be deployed layout. We acknowledge that creating better models of demand for EV charging requires accurately identifying modes of transportation used in individual movements, most importantly being able to distinguish between public and driving transportation modes. For example the approach developed in Wang et al. (2010) is an attempt to address this problem using an unsupervised machine learning method, k-means clustering. The assessment and optimization framework introduced in this work does not rely on merely using raw CDR data and can be used in combination with a transportation mode-inference approach or with any other urban-scale mobility data collected at individual level.

10. Conclusion

Efficient deployment of the network of public charging stations is an important matter which will play a significant role in further increasing the market share of the EVs in the near future. Motivated by this, we have proposed a modelling–optimization framework to find efficient layout of charging stations to minimize overall energy overhead and EV drivers’ excess driving distance to charging stations, which is reflected in the distance required to travel to reach the closest charging station starting from their refuelling demand spots. Our minimalistic model can be easily generalized to include more objectives. For instance, spatial variations in the cost of constructing charging stations and the capacity of each station can be considered as an optimization variable without changing the problem structure and the required optimization method. We also showed that the computed near-optimal layout is robust to variation in input data, in EV penetration rate, as well as in EV and charging station technology. This aspect is particularly important...
given the high costs related to infrastructure deployment, and promote data-driven planning as a viable solution for optimal EV charging station location.

Acknowledgment

The authors would like to gratefully thank; Allianz, the Amsterdam Institute for Advanced Metropolitan Solutions, the Fraunhofer Institute, the Kuwait-MIT Center for Natural Resources and the Environment, Singapore-MIT Alliance for Research and Technology (SMART), UBER, ANAS, and all the members of the MIT Senseable City Lab Consortium for supporting this research. The authors also benefited from discussions with Thomas J. Matarazzo and Chaewon Ahn.

Appendix A. Details of optimization algorithms

Algorithm 1. Chvatal greedy approximation

Initialization

\begin{align*}
W & \leftarrow \text{get the weights array} \\
S & \leftarrow \text{set of all cell indices at which the stations can be located} \\
U & \leftarrow \text{set of all cell indices which must be covered} \\
\Gamma(s) & \leftarrow \text{set of cell indices in the coverage range of the element } s \text{ in } S, \text{ i.e., row indices corresponding to the nonzero elements on the } i\text{-th column of } S^T_C \\
\Omega(u) & \leftarrow \text{set of all cell indices which can provide coverage for element } u \text{ in } U, \text{ i.e., column indices corresponding to the nonzero components on the } i\text{-th row of } S^T_C \\
U_c & \leftarrow \text{set of the so far covered cell indices} \\
S_c & \leftarrow \text{set of so far picked columns} \\
\end{align*}

Greedy coverage phase

\begin{algorithm}
\begin{algorithmic}
\While{not }\(U \neq U_c\)
\State pick an item \(s\) in \(S\) that has the minimum weight per its still uncovered rows
\State \(S \leftarrow S - \{s\}\)
\State \(S_c \leftarrow S_c \cup \{s\}\)
\State \(U_c \leftarrow U_c \cup \text{previously uncovered rows covered by } s\)
\EndWhile
\end{algorithmic}
\end{algorithm}

Redundant cell removal

\begin{algorithm}
\begin{algorithmic}
\For{\(s\) in \(S_c\)}
\State \(\text{counter} \leftarrow 0\)
\For{\(c\) in \(\Gamma(s)\)}
\If{\(\Omega(c) \cap \{Q - \{s\}\} = \emptyset\)}
\State \(\text{counter} \leftarrow \text{counter} + 1\)
\EndIf
\EndFor
\If{\(\text{counter} = 0\)}
\State \(Q = Q - \{s\}\)
\EndIf
\EndFor
\end{algorithmic}
\end{algorithm}

\begin{algorithm}
\State \(Q\) \text{ list of picked column indices sorted according to their weights in decreasing order}
\State \(Q \leftarrow \{S_c\}\)
\end{algorithm}

\begin{algorithm}
\State return \(Q\)
\end{algorithm}

Algorithm 2. Stochastic Chvatal

Replace the while loop in the Chvatal greedy approximation with the following

\begin{algorithm}
\While{not }\(U_c = U\)
\State \(S_k\) \leftarrow \text{pick } k \text{ item from } S \text{ that has the least weight per their uncovered rows}
\State \(s\) \leftarrow \text{choose one item randomly from } S_k
\State \(S \leftarrow S - \{s\}\)
\State \(S_c \leftarrow S_c + \{s\}\)
\State \(U_c \leftarrow U_c + \text{previously uncovered rows covered by } s\)
\EndWhile
\end{algorithm}

Algorithm 3. Binary Tournament Selection and Fusion Crossover

\begin{algorithm}
\State \(N\) \text{ Dimension of the binary vectors in the population, which is the same as the number of cells allowed to accommodate charging stations}
\State \((v_1, v_2)\) \leftarrow \text{Randomly select two subset of binary vector members each of size } T = 2, \text{ and select the fittest in each subset.}
\State \((w_1, w_2)\) \leftarrow \text{Compute the corresponding weights for each of the selected members}
\State \(P_i \leftarrow \frac{w_1}{w_1 + w_2}\)
\State \(v_{new} \leftarrow v_1\)
\For{\(i \in \{1, 2, \ldots, N\}\)}
\If{\(v_1(i) \neq v_2(i)\)}
\end{algorithm}
excess driving distance to charging stations of cell with these modifications. One way to incorporate this optimization dimension in the context of the set-covering problem is to duplicate the corresponding stations. The idea is to model the optimization problem in terms of the spatial pattern of the demand while the current layout is more focused on regions with higher population density.

The interesting point is that the problem is still a set-covering problem and can be handled with the framework introduced here. In a similar way, we can address the case of charging stations with different numbers of available plugs. The idea is to model the optimization problem in terms of “unitary stations” that contains only a single plug. Then, heterogeneous number of stations (plugs) can be required in each cell using the method described above. Whether the required “unitary stations” (plugs) are aggregated into a single charging station, or multiple charging stations across the cell becomes then not relevant, as the smaller unit of analysis in our study is that of a cell.

In Fig. 9, by plotting the cumulative distribution of the average requests each charging station receives during the day, we show how the average load on each charging station is distributed more evenly for the optimized layout compared to current layout of charging stations in the city of Boston. We also sort charging stations for each layout based on the number of requests each station receives in the time window of interest (one day in this case) and show that the average number of requests each charging station receives for the top 20 charging stations is significantly lower for the optimal layout throughout the day (average over 5 weekdays of a week in July 2009). This is due to the increase in utilization of charging stations by placing them in better locations in comparison with currently located charging stations that based on the modelled demand have a low utilization (see Fig. 9). The spatial pattern of the average distance to the closest charging station is also shown in Fig. 6 and confirms that the GA-optimized layout closely follows the spatial pattern of the demand while the current layout is more focused on regions with higher population density.

References


This holds true under the assumption that drivers have access to real time information about the availability of the stations.

Algorithm 4. Genetic Algorithm

\[
p \leftarrow \text{draw uniformly from } [0, 1]
\]

\[
\text{if } p > p_i \text{ then } v_{new} \leftarrow v_j \text{ end if}
\]

\[
\text{end if}
\]

\[
\text{end for}
\]

Appendix B. Charging load distribution and capacitated layout

Assume that a charging station can handle \( l \) EVs at a time. Now if it turns out that, frequently, there are \( q > l \) EVs in a particular cell over a significant time period, then one station is not enough for that cell and that cell needs to be covered by more than one stations. One way to incorporate this optimization dimension in the context of the set-covering problem is to duplicate the corresponding \( C_i \) in the universe set. In other words, if it turns out that cell \( C_i \) needs to be covered by at least \( k \) stations, we can redefine the universe set \( \mathcal{U} \) by replacing the \( C_i \) element in it with \( C_i^1, C_i^2, \ldots, C_i^k \). The weight needs also to be modified accordingly: \( \Omega_{kl} \rightarrow \Omega_{kl}/k_i \). With these modifications, the universe set grows, and more subsets are required to cover the whole set. Also, the contribution of excess driving distance to charging stations of cell \( C_i \) decreases for a station located at distance \( j \), because its load is now divided between the other \( k_i - 1 \) available stations.\(^4\) The interesting point is that the problem is still a set covering problem and can be handled with the framework introduced here. In a similar way, we can address the case of charging stations with different number of available plugs. The idea is to model the optimization problem in terms of “unitary stations” that contains only a single plug. Then, heterogeneous number of stations (plugs) can be required in each cell using the method described above. Whether the required “unitary stations” (plugs) are aggregated into a single charging station, or multiple charging stations across the cell becomes then not relevant, as the smaller unit of analysis in our study is that of a cell.

In Fig. 9, by plotting the cumulative distribution of the average requests each charging station receives during the day, we show how the average load on each charging station is distributed more evenly for the optimized layout compared to current layout of charging stations in the city of Boston. We also sort charging stations for each layout based on the number of requests each station receives in the time window of interest (one day in this case) and show that the average number of requests each charging station receives for the top 20 charging stations is significantly lower for the optimal layout throughout the day (average over 5 weekdays of a week in July 2009). This is due to the increase in utilization of charging stations by placing them in better locations in comparison with currently located charging stations that based on the modelled demand have a low utilization (see Fig. 9). The spatial pattern of the average distance to the closest charging station is also shown in Fig. 6 and confirms that the GA-optimized layout closely follows the spatial pattern of the demand while the current layout is more focused on regions with higher population density.

\(^4\)This holds true under the assumption that drivers have access to real time information about the availability of the stations.