Demand Prediction and Placement Optimization for Electric Vehicle Charging Stations

Ragavendran Gopalakrishnan

Xerox Research Centre India, Bangalore, Karnataka 560103,
Ragavendran.Gopalakrishnan@xerox.com

Arpita Biswas

Xerox Research Centre India, Bangalore, Karnataka 560103,
Arpita.Biswas@xerox.com

Alefiya Lightwala

Xerox Research Centre India, Bangalore, Karnataka 560103,
Alefiya.Lightwala@xerox.com

Skanda Vasudevan

Xerox Research Centre India, Bangalore, Karnataka 560103,
Skanda.Vasudevan@xerox.com

Abhishek Tripathi

Xerox Research Centre India, Bangalore, Karnataka 560103,
Abishek.Tripathi3@xerox.com

Partha Dutta

Xerox Research Centre India, Bangalore, Karnataka 560103,
Partha.Dutta@xerox.com

Due to the environmental impact of fossil fuels and high variability in their prices, there is rising interest in adopting electric vehicles (EVs) by both individuals and governments. Despite the advances in vehicle efficiency and battery capacity, a key hurdle is the inherent interdependence between EV adoption and charging station deployment–EV adoption (and hence, charging demand) increases with the availability of charging stations (operated by service providers) and vice versa. Thus, effective placement of charging stations plays a key role in EV adoption. In the placement problem, given a set of candidate sites, an optimal subset needs to be selected with respect to the concerns of both (a) the charging station service provider, such as the demand at the candidate sites and the budget for deployment, and (b) the EV user, such as charging station reachability and short waiting times at the station. This work addresses these concerns, making the following three novel contributions: (i) a supervised multi-view learning framework using Canonical Correlation Analysis (CCA) for demand prediction at candidate sites, using multiple datasets such as points of interest information, traffic density, and the historical usage at existing charging stations; (ii) a "mixed-packing-and-covering" optimization framework that models competing concerns of the service provider and EV users, which is also extended to optimize government grant allocation to multiple service providers; (iii) an iterative heuristic to solve these problems by alternately invoking knapsack and set cover algorithms. The performance of the demand prediction model and the placement optimization heuristic are evaluated using real world data. In most cases, and especially when budget is scarce, our heuristic achieves an improvement of 10-20% over a naive heuristic, both in finding feasible solutions and maximizing demand.

1. Introduction.

The environmental impact of fossil fuels and the high variability in their prices have led to rising adoption of Electric Vehicles (EVs), which is supported by ambitious government policies for promoting EVs [31, 15]. Despite the technological advances in vehicle efficiency and battery capacity, a key hurdle in EV adoption is that, barring a few pockets of densely
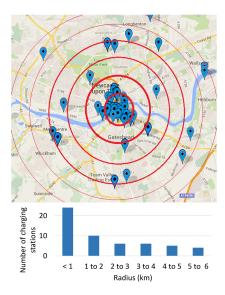
FIGURE 1. Charging stations in North East England.

populated areas, the distribution of EV charging stations is sparse in most regions.[1] Because of this, EV owners and potential buyers frequently worry about whether the vehicle will have sufficient charge to travel to their trip destinations or an intermediate charging station. On the other hand, given the high cost of building a charging station and currently low[2] number of EVs, charging station operators would only want to place stations where there is sufficient demand for charging. This results in a situation where the charging station density is concentrated near the city centre and rapidly decrease as we get farther, e.g., Figure 1 illustrates this for North East England.

This work is an attempt to break this deadlock by improving the distribution of new charging stations so as to mitigate the range anxiety of EV users. To be effective, such a solution must simultaneously address the concerns of both EV users and charging station operators. Charging station operators are concerned about (a) *demand*: charging stations must be distributed to serve the maximum total demand for charging, and (b) *budget*: there is a limited budget available for setting up charging stations. EV users are concerned about (c) *reachability*: there must be a charging station that is reachable within a short driving distance from most locations, and (d) *waiting time*: the waiting time to begin charging at charging stations should not be prohibitively large, given that the charging event itself is time-consuming.[3]

This paper models the above placement problem within a framework that is inspired from the facility location literature [14]. Given a set of candidate sites for charging stations, e.g., parking lots in a city, the objective is to select an optimal subset of these locations that maximizes the total demand satisfied, subject to the total cost not exceeding the available budget ("packing" constraint), and given a set of locations of interest, there is at least one charging station within a specified driving range of every location in this set ("covering" constraint). This "mixed-packing-and-covering" formulation (Section 3.1) covers concerns (a)-(c). In order to take care of (d), a queueing model is used to translate the constraint on waiting time and the charging demand at a

---

[1] Recent government mandates [27, 1] focus specifically on the expansion of refueling infrastructure to cover entire geographical regions.

[2] Nevertheless, this number is rapidly growing, e.g., in the U. S., EV market share has risen [9] from 0.14% in 2011 to 0.72% in 2014.

[3] There may be several other concerns for charging station operators and EV users, including alternate interpretations of *reachability*, e.g., [16], but in this work, we restrict our attention to modeling just these four.

location to a constraint on the minimum number of charging slots needed at a charging station at that location, which is then factored into the setup costs in a pre-processing step.

Before tackling the optimization problem itself, it is necessary to predict the demand for charging at candidate sites. Although demand prediction has been extensively studied for planning transportation infrastructure, the EV charging station demand prediction problem is difficult because of sparse deployment of such stations, and lack of sufficient and reliable historic data. Thus, in addition to the limited historical demand data for existing charging stations in other locations, two other types of features of the location which may have a strong impact on the usage are considered. First, nearby Points of Interest (PoI), e.g., shopping malls, institutions, restaurants, hospitals etc., indicate the frequency of visits that could be made by travelers to the given location, and hence, are useful for estimating the demand. Second, traffic density data at nearby road junctions are also useful as features in demand prediction. Using these features, in Section 2 proposes a supervised multi-view learning framework using Canonical Correlation Analysis (CCA) for EV charging station demand prediction.

Once the demands are estimated using the demand prediction algorithm, the next step is to solve the mixed packing and covering problem. We identify a family of heuristics in Section 3.2 that seeks to iteratively find the optimal allocation of the available budget between satisfying the packing and the covering constraints, by alternately invoking algorithms that solve reduced knapsack and set-cover subproblems. For example, choosing the well known greedy algorithms for the knapsack and set-cover problems [33] yields one instance in this family.[4] In Section 3.3, we present results from an experimental evaluation of this instance for the EV charging station placement problem using charging data from the UK. In most cases, and especially when budget is scarce, our heuristic achieves an improvement of 10-20% over a naive heuristic, both in terms of finding feasible solutions and maximizing demand.

The modeling and heuristic frameworks above can be easily extended to the following alternate scenarios: (a) incremental placement of charging stations when the budget is progressively released over a period of time, and (b) when there are multiple charging station operators, each with their respective budgets, and a government agency (subject to its own budget) seeks to optimally allocate grants to incentivize these providers to set up charging stations at selected locations (e.g., where the demand is low).

1.1. Related work. With the recent increase in the adoption of EVs, the charging station placement problem has received significant attention. One of the prerequisites of an effective charging station placement is the availability of estimated charging demand at candidate sites. In this context, prior literature has studied the impact of external information on charging station demand. For example, parking demand is combined with facility location problem in [7], and the effect of demographic features is studied using regression models in [35]. However, these approaches lack a principled multi-view learning framework to integrate heterogeneous data sets to predict EV charging demand. In the machine learning literature, CCA-based models have been used for multivariate regression [28, 30, 20]; in this paper, we propose a novel variant of CCA-based multivariate regression.

A useful charging station placement solution needs to take into account the charging demand, budget, coverage, and waiting time at the stations. Existing work has not considered all of these factors simultaneously. For example, placement of charging stations based on predicted demand has been studied in [12, 19, 35, 7, 22, 8], but they do not simultaneously consider constraints on budget, coverage, and waiting times. Coverage requirements in charging station placement has

---

[4] Since the primary mixed packing and covering problem, and the associated knapsack and set-cover problems are NP-Hard, it is unlikely that any polynomial-time solution, including this instance, would be optimal.

been modeled as path cover in [16], and reachability cover in [29] over the city road network, but they do not consider the demand satisfied by the deployment and the waiting time faced by the users.

The optimization problem considered in this work contains both packing and covering constraints, with integer decision variables. Mixed packing and covering problems have been studied for fractional decision variables [36, 3], as well as for integer decision variables [5, 23]. However, this work introduces an optimization problem with a knapsack packing constraint and a set covering constraint, which has not been studied earlier.

2. Demand prediction at candidate sites. Understanding the EV charging demand is key to optimal placement of charging stations. Since the distribution of charging stations is typically sparse, there is insufficient data to efficiently model the demand. Thus, it becomes important to model the relationship of EV charging demand with auxiliary data such as Point of Interest (PoI) and traffic density. Such relationships help in better understanding not only the charging demand, but also the factors behind it. Additionally, the expected demand at a candidate site where no historical demand is known, can be predicted based on external factors. A multi-view learning framework based on Canonical Correlation Analysis (CCA) is used to jointly model the charging demand data and other external data sources, and to predict the demand at new candidate sites given the external factors. Next, we briefly describe CCA and then present the CCA-based regression framework.

2.0.1. Canonical correlation analysis. Canonical correlation analysis is a classical method to find a linear relationship between two sets of random variables. Let X and Y be data matrices of size $n \times d_x$ and $n \times d_y$ respectively, where each row corresponds to a realization of the random variables. CCA finds linear projections $u_x \in \mathbb{R}^{d_x}$ and $u_y \in \mathbb{R}^{d_y}$ such that their correlation, $corr(Xu_x, Yu_y)$, is maximized. It can be shown that maximizing correlation is equivalent to minimizing $\|Xu_x - Yu_y\|^2$ subject to $\|Xu_x\|^2 = 1$ and $\|Yu_y\|^2 = 1$. A set of $k \leq min(rank(X), rank(Y))$ linear projections can be computed by solving a generalized eigenvalue problem; [18] provides a detailed review of classical CCA.

2.1. Multi-view regression using CCA. In this section, a multi-view learning based regression framework for the charging demand prediction at candidate sites is presented. The regression framework uses CCA in a supervised setting as a multivariate regression tool. CCA is, in fact, closely related to multivariate linear regression; [4, 30] present least-square formulations. CCA-based approaches model the statistical dependence between two or more data sets and assume a single latent factor to describe information shared between all datasets. In a supervised setting such as multivariate regression task, the output variable is considered as one of the random variables and CCA models the statistical dependence between the output variable (dependent) and a covariate (independent variable). However, when there are more than two datasets (output variable and multiple covariates), modeling shared information may not always be a good choice, since it would discard dataset-specific information which may be crucial to predicting the output variable.

To address this deficiency, a multi-view learning regression framework is proposed when there are $n > 2$ covariates, all of which can be used to predict the output variable. The idea is to model the statistical dependence of the output variable with each of the covariates separately and then learn a weighted combination for each such model that maximizes the prediction performance. We call this model Multiple Dependent Regression (MDR), which can be formulated as the ensemble regression function $\mathbf{f} = \sum_i w_i \mathbf{f_i}$, where $\sum_i w_i = 1$. Here, $\mathbf{f_i}$ is a regression function learned from covariate X_i to the output variable Y, and w_i are the weights denoting the importance of covariate X_i in the regression task. The regression function $\mathbf{f_i}$ for covariate X_i can be learnt according to a

CCA-based regression, given by $\arg\min_{U_x^i, U_y^i} \|X_i U_x^i - Y U_y^i\|^2$, where U_x^i and U_y^i are k-dimensional linear projections for each (covariate, output variable) pair (X_i, Y). Hence, MDR is a two step framework of multiple regression:

(1) For each covariate X_i, apply CCA for X_i and Y, and compute the prediction error e_i.

(2) For each i, compute $w_i = \frac{1}{n-1}(1 - \frac{e_i}{\sum_j e_j})$.

The values of parameters U_x, U_y, w_i are learnt from the training data as described in the steps above. The model thus trained can be used to predict the output variable given the input covariates. In our case, the task is to predict the demand of EV charging at a new candidate location given the covariates as explained earlier.

MDR is analogous to ensemble learning approaches where multiple learning models are trained to solve a given problem. In our case, we train multiple CCA-based regression models in the first step and learn a weighted combination of models based on training error in the second step. The weights w_i explain the importance of each covariate in the prediction task. This procedure is flexible–any CCA-based solution can be used in Step (1). In this work, we use a Bayesian solution with group-wise sparsity proposed in [20].

By learning a separate CCA model for each covariate, MDR ensures that dataset-specific information of any covariate (with respect to other covariates) is not ignored if it is shared with the output variable. MDR provides an intuitive, flexible, and yet, a simple approach to multi-view regression. It can be easily extended to non-linear projections using non-linear solutions to CCA.

2.2. Experimental results for demand prediction. MDR is evaluated on EV charging data obtained from 252 public charging points in North East England through UK's Plugged-In-Places program [26]. The location of charging points was obtained from UK's National Charge Point Registry data [24]. PoI information is extracted from OpenStreetMap [17] API for 11 categories: sustenance, education, transportation, financial, healthcare, entertainment, sports, gardens, places of worship, shops and public buildings. Finally, we use traffic data are for each junction-to-junction link on major road networks, provided by [13].

The following data matrices are created: (i) $Y \in \mathbb{R}^{252 \times 24}$ represents hour-wise charging demand for each of 252 charging points, where $Y[i,j]$ is the average energy consumed at charging point i in hour j; (ii) $X_1 \in \mathbb{R}^{252 \times 11}$ represents the PoIs, where $X_1[i,j]$ is the frequency of PoI category j within a radius of 500 meters around charging point i; (iii) $X_2 \in \mathbb{R}^{252 \times 5}$ represents the traffic densities at the 5 nearest traffic junctions to each charging point; and (iv) $X_3 \in \mathbb{R}^{252 \times 5}$ represents the charging demands at the 5 nearest charging points to each charging point. Since charging data is sparse during early mornings and late nights, only data from 07:00 hrs. till 23:00 hrs. is used.

MDR is evaluated in a leave-one-out cross-validation manner by training on all but one instance (charging point) and testing on the left-out instance. Prediction performance is measured as the average Root Mean Square Error (RMSE) over all test instances. MDR is compared against (i) Linear Regression(LR) - all covariates are concatenated feature-wise similar to [35], (ii) Multiple Linear Regression (MLR) - weighted sum of multiple LR analogous to MDR, and (iii) Bayesian CCA (bCCA) - group factor analysis [34]. Figure 2 shows the average error for each hour of the day over all charging stations. It can be observed that MDR clearly outperforms the other three baselines giving an overall reduction in error by 27% compared to simple LR, 21% compared to multiple LR and 18% compared to bCCA.

3. Optimal selection from candidate sites for placing charging stations. In this section, we explain the mathematical model for the charging station placement problem, and detail the methods to solve the resulting mixed packing and covering optimization problem.
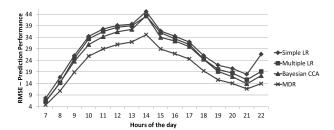
FIGURE 2. Comparison of MDR against baselines.

3.1. Preliminaries. Let $\mathcal{L} = \{1, \ldots, |\mathcal{L}|\}$ denote the set of all candidate sites for placing a charging station. Let $r \in \mathbb{R}_+$ denote the desired "reachability radius," that is, the maximum distance to be travelled in order to reach a charging station. Let $\mathcal{I}$ denote the set of all locations of interest that are desired to be "covered," that is, lie within the reachability radius from at least one charging station. Let B denote the total budget available.

For each candidate site $i \in \mathcal{L}$: $x_i \in \{0,1\}$, $d_i \in \mathbb{R}_+$, and $c_i \in \mathbb{R}_+$ denote, respectively, whether a charging station is placed, the demand for service, and the cost of setting up a charging station at i. $S_i^r \subseteq \mathcal{I}$ is the *cover set* of i, that is, the set of locations of interest that would be "covered" (within a driving distance of r) if a charging station is placed at i.

3.1.1. Optimizing for demand. For a given reachability radius r, set of candidate sites $\mathcal{L}$, and set of locations of interest $\mathcal{I}$, the cover sets $\{S_i^r\}_{i \in \mathcal{L}}$ can be precomputed. Given predicted demand and charging duration data, a queueing model [21] can estimate the number of charging slots necessary at each candidate site to meet a specified constraint on waiting times, which is then used to precompute the costs $\{c_i\}_{i \in \mathcal{L}}$. (See Appendix 3.2.3 for details.) Then, the optimization problem to be solved is the Mixed Pack & Cover (MPC) problem shown in Figure 3(a).[5]

3.1.2. Approximation algorithm for a transformed problem. MPC (Figure 3(a)) is NP-Hard as it contains as a special case, the set-cover problem [33]. For a closely related pure covering problem called Demand & Set-Cover (DSC, Figure 3(e)), an approximation algorithm is described next.

DSC has as subproblems, two well-known NP-Hard problems, Set-Cover (SC) [33] and Minimization Knapsack (MinKP) [11], which are given in Figures 3(c) and 3(d), respectively. Consider, an f_{sc} factor approximation algorithm A_{sc} for set cover and an f_{kp} factor approximation algorithm A_{kp} for minimization knapsack. Let $\mathcal{L}_{sc}$ and $\mathcal{L}_{kp}$ (subsets of $\mathcal{L}$) be the solutions returned by A_{sc} and A_{kp}, respectively. Consider algorithm A_{dsc} for DSC that uses A_{sc} and A_{kp} as subroutines, and returns the union of the solutions returned by the two subroutines, that is, $\mathcal{L}_{sc} \cup \mathcal{L}_{kp}$, as a solution for DSC. The following lemma (proved in Appendix A) establishes that A_{dsc} is an $(f_{sc} + f_{kp})$ factor approximation algorithm for DSC.

LEMMA 1. *A_{dsc} returns a feasible solution for DSC whose cost is at most $(f_{sc} + f_{kp})$ times the optimal cost.*

The heuristic framework for MPC presented next is inspired by A_{dsc} and the observation that, similar to DSC, a solution for MPC can be obtained by combining set cover and knapsack algorithms.

---

[5] The total demand satisfied by installing charging stations at a subset of candidate sites may be less than the sum of the predicted demands at those sites due to overlapping reachability regions. We ignore this effect in order to keep the optimization problem simple.

$$\max_{\{x_i : i \in \mathcal{L}\}} \sum_{i \in \mathcal{L}} d_i x_i$$

$$\text{subject to } \sum_{i \in \mathcal{L}} c_i x_i \leq B$$

$$\sum_{i \in \mathcal{L} : \ell \in S_i^r} x_i \geq 1 \quad \forall \ell \in \mathcal{I}$$

(a) Mixed Pack & Cover (MPC)

$$\max_{\{x_i : i \in \mathcal{L}\}} \sum_{i \in \mathcal{L}} d_i x_i$$

$$\text{subject to } \sum_{i \in \mathcal{L}} c_i x_i \leq B$$

(b) Knapsack (KP)

$$\min_{\{x_i : i \in \mathcal{L}\}} \sum_{i \in \mathcal{L}} c_i x_i$$

$$\text{subject to } \sum_{i \in \mathcal{L} : \ell \in S_i} x_i \geq 1 \quad \forall \ell \in \mathcal{I}$$

(c) Set-Cover (SC)

$$\min_{\{x_i : i \in \mathcal{L}\}} \sum_{i \in \mathcal{L}} c_i x_i$$

$$\text{subject to } \sum_{i \in \mathcal{L}} d_i x_i \geq D$$

(d) Min-Knapsack (MinKP)

$$\min_{\{x_i : i \in \mathcal{L}\}} \sum_{i \in \mathcal{L}} c_i x_i$$

$$\text{subject to } \sum_{i \in \mathcal{L} : \ell \in S_i^r} x_i \geq 1 \quad \forall \ell \in \mathcal{I}$$

$$\sum_{i \in \mathcal{L}} d_i x_i \geq D$$

(e) Demand & Set-Cover (DSC)

FIGURE 3. Important optimization problems.

3.2. Solving the mixed-pack-and-cover problem. Our goal is to introduce a general methodology for algorithms to solve MPC (Figure 3(a)), which is NP-Hard, by breaking it down into the following two subproblems:

1. *The Knapsack Problem (KP):* This is the packing problem in Figure 3(b). Let $\text{KP}\left(\mathcal{L}, \{d_i\}_{i \in \mathcal{L}}, \{c_i\}_{i \in \mathcal{L}}, B\right)$ denote any algorithm that solves this problem.
2. *The Set-Cover Problem (SC):* This is the covering problem in Figure 3(c). Let $\text{SC}\left(\mathcal{L}, \mathcal{I}, \{c_i\}_{i \in \mathcal{L}}, \{S_i\}_{i \in \mathcal{L}}\right)$ denote any algorithm that solves this problem.

The algorithm KP() (respectively, SC()) need not be optimal, but we require that it be nontrivial in the sense that its solution must be *maximal* (respectively, *minimal*). That is, it should not be possible to add (remove) an item and still satisfy the packing (covering) constraint(s).

3.2.1. The iterative-pack-and-cover (IPAC) framework. The available budget could be used very differently if the problem had been a pure packing problem (where maximizing the demand is the only concern) or a pure covering problem (where satisfying the covering constraints is the only concern). Thus, a good solution to the mixed packing and covering problem should achieve an appropriate balance by dividing the available budget between these two concerns. The core idea behind the Iterative Pack And Cover (IPAC) framework is to iteratively search for such an optimal split.

In each iteration, the budget B can be thought of as comprising of three parts: B^p, B^c, and some excess (due to the integrality constraint). Here, B^c is the portion of the budget used by a solution to a covering problem, and B^p is the portion of the budget used by a solution to a packing problem when constrained by a reduced budget of $B - B^c$. A check is performed using SC() to determine whether the remaining budget $B - B^p$ is sufficient to satisfy the covering constraints left unsatisfied by the solution to the packing problem. Starting with $B^c = 0$ (pure packing) and the corresponding solution from KP(), during each iteration, B^c is increased until the covering check passes, at which point the solutions of the packing and covering problems obtained in the last iteration are merged.[6] The resulting solution is guaranteed to be a feasible solution to the MPC problem (Figure 3(a)). Optionally, KP() can be invoked one final time to use up any remaining

---

[6] As an extension, note that B^c need not necessarily increase in each iteration, as long as it can be shown that the procedure will terminate in finite time.

portion of the budget. In the worst case, the iterations continue until $B^c = B$, at which point it becomes a pure covering problem, and if the covering check still fails, then IPAC fails to find a feasible solution to MPC. (But if this happens, it cannot necessarily be concluded that MPC is infeasible, unless `SC()` is optimal, which is unlikely in polynomial time since the set-cover problem is NP-Hard.)

The above framework is quite generic and encompasses a large class of algorithms. Next, we describe a particular instantiation of the framework which runs in polynomial time. The corresponding pseudocode is detailed in Algorithm 1.

1. Set $B^c = 0$ and invoke `KP()` to solve the pure packing problem. Let B^p be the portion of the budget used by the chosen items. $B^{free} = B - B^p$ is the remainder.
2. Update B^c to be the minimum budget required to satisfy the unsatisfied covering constraints, computed by invoking `SC()` to solve the residual covering problem.
3. Repeat the following two steps until either $B^c \leq B^{free}$ or $B^c > B$:
   (a) *Packing:* In response to the reduced budget $B - B^c$, remove one or more chosen items. In order to do so, use a ranking method, `RANK()`, that ranks the currently chosen items according to some measure of their importance. Keep removing items that are least important until the remaining items satisfy the reduced budget. Let B^p be the portion of the budget used. $B^{free} = B - B^p$ is the new remainder.
   (b) *Covering:* Since the removal of some items in the previous step might have resulted in more unsatisfied covering constraints, invoke `SC()` on the new residual covering problem and update B^c accordingly.
4. If $B^c > B$, then a feasible solution cannot be found. Otherwise, $B^c \leq B^{free}$, and so, add the items from the solution to the last instance of `SC()` to the remaining chosen items to obtain a feasible solution.
5. Invoke `KP()` to fill any unused portion of the budget using unallocated items.

Algorithm 1 `IPAC`

1: **Input:** $\mathcal{L}, \mathcal{I}, \{d_i\}_{i \in \mathcal{L}}, \{c_i\}_{i \in \mathcal{L}}, B, \{S_i^r\}_{i \in \mathcal{L}}, $`KP()`,`SC()`,`RANK()`
2: **Output:** $\{x_i\}_{i \in \mathcal{L}}$

3: $\{x_i\}_{i \in \mathcal{L}} \leftarrow $`KP`$(\mathcal{L}, \{d_i\}_{i \in \mathcal{L}}, \{c_i\}_{i \in \mathcal{L}}, B)$;
4: $\mathcal{L}^p \leftarrow \{i \in \mathcal{L} \ : \ x_i = 1\}$; $B^p \leftarrow \sum_{i \in \mathcal{L}} c_i x_i$;
5: $\mathcal{I}^p \leftarrow \bigcup_{i \in \mathcal{L}^p} S_i^r$;
6: // $\mathcal{I}^p \leftarrow$ elements of $\mathcal{I}$ covered by $\mathcal{L}^p$.
7: $\{x_i^c\}_{i \in \mathcal{L} \setminus \mathcal{L}^p} \leftarrow $`SC`$(\mathcal{L} \setminus \mathcal{L}^p, \mathcal{I} \setminus \mathcal{I}^p, \{c_i\}_{i \in \mathcal{L} \setminus \mathcal{L}^p}, \{S_i^r\}_{i \in \mathcal{L} \setminus \mathcal{L}^p})$;
8: $B^c \leftarrow \sum_{i \in \mathcal{L} \setminus \mathcal{L}^p} c_i x_i^c$; $B^{free} \leftarrow B - B^p$;
9: **while** $(B^c > B^{free}$ and $B^c \leq B)$ **do**
10:     // ρ is $\mathcal{L}^p$ in increasing order of importance.
11:     $\rho \leftarrow $`RANK`$(\mathcal{L}^p, \{d_i\}_{i \in \mathcal{L}^p}, \{c_i\}_{i \in \mathcal{L}^p}, \{S_i^r\}_{i \in \mathcal{L}^p})$;
12:     // Remove least important items.
13:     $j \leftarrow 0$;
14:     **while** $B^{free} < B^c$ **do**
15:       $j \leftarrow j + 1$; $x_{\rho[j]} \leftarrow 0$; $B^{free} \leftarrow B^{free} + c_{\rho[j]}$;
16:     // Recompute quantities.
17:     $\mathcal{L}^p \leftarrow \{i \in \mathcal{L} \ : \ x_i = 1\}$; $\mathcal{I}^p \leftarrow \bigcup_{i \in \mathcal{L}^p} S_i^r$;
18:     $\{x_i^c\}_{i \in \mathcal{L} \setminus \mathcal{L}^p} \leftarrow $`SC`$(\mathcal{L} \setminus \mathcal{L}^p, \mathcal{I} \setminus \mathcal{I}^p, \{c_i\}_{i \in \mathcal{L} \setminus \mathcal{L}^p}, \{S_i^r\}_{i \in \mathcal{L} \setminus \mathcal{L}^p})$;
19:     $B^c \leftarrow \sum_{i \in \mathcal{L} \setminus \mathcal{L}^p} c_i x_i^c$;
20: **if** $B^c > B$ **then** EXIT;
21: // Add required items to knapsack.
22: **for** i in $\mathcal{L} \setminus \mathcal{L}^p$ **do** $x_i \leftarrow x_i + x_i^c$;
23: // Add more items if possible.
24: $\mathcal{L}^p \leftarrow \{i \in \mathcal{L} \ : \ x_i = 1\}$; $B^p \leftarrow \sum_{i \in \mathcal{L}} c_i x_i$;
25: $\{x_i\}_{i \in \mathcal{L} \setminus \mathcal{L}^p} \leftarrow $`KP`$(\mathcal{L} \setminus \mathcal{L}^p, \{d_i\}_{i \in \mathcal{L} \setminus \mathcal{L}^p}, \{c_i\}_{i \in \mathcal{L} \setminus \mathcal{L}^p}, B - B^p)$;

3.2.2. The `RANK()` function. The effectiveness of IPAC depends on the choice of methods for `KP()`, `SC()`, and `RANK()`. There are several choices in the literature for the first two [33], so we briefly discuss one possible choice for the third. A general observation is that an item $i \in \mathcal{L}$ is more important or desirable if its demand d_i is high, or its cost c_i is low, or if the number of elements it covers $|S_i|$ is high. Based on this, a viable candidate for `RANK`$(\mathcal{L}, \{d_i\}_{i \in \mathcal{L}}, \{c_i\}_{i \in \mathcal{L}}, \{S_i\}_{i \in \mathcal{L}})$ would be a method that ranks items in increasing order according to the value $v_i = \left(\frac{d_i}{\sum_{j \in \mathcal{L}} d_j} + \frac{|S_i|}{|\mathcal{I}|} \right) / c_i$, where $\mathcal{I} = \bigcup_{i \in \mathcal{L}} S_i$ denotes the set of all elements covered by items in $\mathcal{L}$.

3.2.3. Computing charging station costs. In this section, we briefly address estimating the setup costs c_i, which depend on (i) the number of charging slots necessary at i to satisfy any desired SLA on waiting times, and (ii) per-unit setup cost, which can include infrastructure as well as land/licensing costs. Since data on the per-unit costs are generally available, we focus on (i), for which we model candidate charging stations as a multi-server queue.

One possible way to model each candidate charging station at location i in order to estimate its size (and hence cost) is using a multi-server queue that follows an $M/M/N_i$ discipline, where $N_i \in \mathbb{Z}_+$ is the number of charging slots to be set up. Customers arrive into the queue according to a Poisson process with rate λ_i per time unit. The service time for each customer is exponentially distributed with mean $1/\mu_i$ time units. λ_i can be derived from the demand d_i, whereas μ_i can be derived from existing data on the average customer service time in nearby existing charging stations. The average time a customer waits before service is then given by $\mathbb{E}[W] = \frac{\text{ErlC}\left(N_i, \frac{\lambda_i}{\mu_i}\right)}{N_i \mu_i - \lambda_i}$, where $\text{ErlC}(N, \rho) = \left(\frac{\rho^N}{N!} \frac{N}{N-\rho} \right) / \left(\sum_{j=0}^{N-1} \frac{\rho^j}{j!} + \frac{\rho^N}{N!} \frac{N}{N-\rho} \right)$ [21]. Suppose the SLA to be met is given by $\mathbb{E}[W] \leq t_i$. Then, because it is well-known [21] that $\mathbb{E}[W]$ is a decreasing function of N_i, we simply choose the smallest N_i (e.g., using binary search) such that the SLA is satisfied.

3.2.4. Optimizing for reachability. In addition to maximizing the demand for a given reachability radius, it may also be desirable to minimize the reachability radius itself. Without loss of generality, we assume that $r \in [R^{\min}, R^{\max}]$, where $R^{\min} = \max_{\ell \in \mathcal{I}} \min_{i \in \mathcal{L}} \text{dist}(\ell, i)$, and $R^{\max} = \max_{\ell \in \mathcal{I}} \max_{i \in \mathcal{L}} \text{dist}(\ell, i)$. Here, $\text{dist}(\ell, i)$ denotes the distance between locations ℓ and i according to an underlying transportation network.[7] Let $D^*(r)$ denote the maximum demand covered (obtained from a solution to the MPC problem (Figure 3(a), perhaps by using Algorithm 1) for a given reachability radius r. (For convenience, we define $D^*(r) = 0$ for those r for which MPC is infeasible.) If $\alpha \in [0,1]$ is a given trade-off parameter, then the objective can be defined as:

$$\max_{r \in [R^{\min}, R^{\max}]} \alpha \frac{D^*(r)}{\sum_{i \in \mathcal{L}} d_i} + (1 - \alpha) \frac{R^{\max} - r}{R^{\max} - R^{\min}}. \tag{1}$$

Let $\mathcal{R} = \{ \text{dist}(\ell, i) | \ell \in \mathcal{I}, i \in \mathcal{L}, \text{dist}(\ell, i) \geq R^{\min} \}$ denote the set of all distances from a location of interest to a candidate site that is at least $R^{\min}$, and suppose we represent this set as $\mathcal{R} = \{ r_1, r_2, \ldots, r_{|\mathcal{R}|} \}$, where $R^{\min} = r_1 < r_2 < \ldots < r_{|\mathcal{R}|} = R^{\max}$. The following lemma presents two observations concerning $D^*(r)$:

LEMMA 2. $D^*(r)$ *is nondecreasing,*[8] *and for all* $1 \leq m < |\mathcal{R}|$, *for all* $r_m \leq r < r_{m+1}$, $D^*(r) = D^*(r_m)$.

---

[7] The lower bound for r stems from the observation that when $r < R^{\min}$, even if charging stations are placed at all candidate sites, there would be at least one uncovered location of interest; so, no feasible solution exists for such r. The upper bound for r follows from the fact that even if there is only a single selected site where a charging station is placed, and the solution is feasible, $R^{\max}$ is, by definition, the maximum distance to be travelled in order to reach that charging station.

[8] This is because, for any two radii r_1, r_2 with $r_1 \leq r_2$, the feasible set of MPC for r_1 is a subset of that for r_2.

In other words, $D^*(r)$ is a nondecreasing step function and remains unchanged between values of r that do not correspond to $\texttt{dist}(\ell, i)$ for some (ℓ, i). Thus, it can be seen that the value(s) of r at which the objective function in (1) attains its maximum must be among the elements in $\mathcal{R}$. Therefore, solving (1) is equivalent to solving:
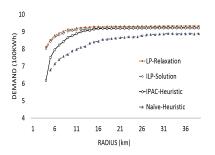
$$\max_{r \in \mathcal{R}} \alpha \frac{D^*(r)}{\sum_{i \in \mathcal{L}} d_i} + (1 - \alpha) \frac{R^{\max} - r}{R^{\max} - R^{\min}}. \tag{2}$$
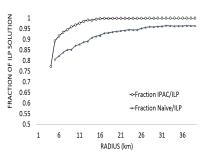
3.3. Experimental results. We conducted two sets of experiments to evaluate the performance of IPAC using real world data – one for North East England (presented in this section), and the other for the western region of East Anglia and southern region of East Midland (presented in Appendix C).

3.3.1. Setup. We use 1305 parking locations in North East England for both the candidate sites for placing charging stations, as well as the locations of interest that need to be covered. This is equivalent to a requirement that each candidate site must either have a charging station on-site, or one that is a short driving distance away. Accordingly, the cover set S_i^r for each candidate site i consists of the candidate sites that are within a distance of r from i. The charging demands at candidate sites are predicted using the MDR model trained in Section 2.2. The costs are computed as $c_i = N_i(L_i + F_i)$; we estimate each of these components as follows:

1. L_i, *the per-unit land cost*: Points of interest around candidate sites play vital roles in determining their land values. In addition, different types of facilities affect the land value differently. Thus, we assume $L_i = p + \sum_{j \in P_i^\delta} \frac{\texttt{Score}_{ij}}{\texttt{dist}(i,j)}$, where $p = \$4000$ is the minimum per-unit land cost across all the locations in North East England (computed using data obtained from [32]), P_i^δ is the set of points of interest that are within a radius $\delta = 1$ km from candidate site i, and $\texttt{Score}_{ij}$ is the score assigned to the point of interest j according to its type, as follows. Airports and railway stations have the highest score of 800, whereas schools, restaurants and hospitals have a lower score of 300. $\texttt{Score}_{ij}$ is then normalized by $\texttt{dist}(i, j)$.
2. F_i, *the per-unit infrastructure cost*: For Level 2 charging, we set $F_i = \$1852$ from Table 6 of [6].
3. N_i, *the number of charging spots*: We assumed a Level 2 charging rate of 6.4kW and set, for each candidate site, N_i to be the minimum number of Level 2 charging spots necessary (using the queueing model described in Section 3.2.3) to ensure that the average "peak-demand" waiting time (taken as the estimated maximum hourly demand at the candidate site over two years) is less than 5 minutes.
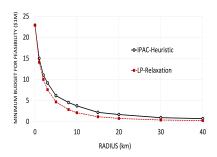
The predicted demands and estimated costs of the candidate sites are then used to find the optimal locations using Algorithm 1 (IPAC), where, for both KP() and SC(), we choose the well-known greedy approximation algorithms introduced in [33], and for RANK(), we use the function proposed in Section 3.2.1.

3.3.2. Performance. To evaluate the performance of IPAC when it can locate feasible solutions, we compare it with (i) a naive heuristic that first solves the covering problem using SC(), and then invokes KP() on the remaining budget to add unselected candidate sites, (ii) the optimal ILP solution of the MPC problem (Figure 3(a)), and (iii) a solution to the LP relaxation of MPC problem (Figure 3(a)), which gives an upper bound on the actual (integer LP) optimization problem. We use a budget of $B = \$6$M. The number of pairwise distances between candidate sites is prohibitively large for our experiment; so, we used values of r (in units of km), from the set $\{5.1, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40\}$. The results are plotted in Figures 4(a)-4(b).
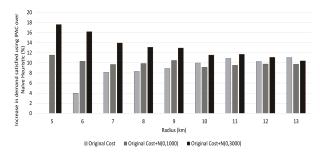
As we increase the reachability radius r, the set of allocations that satisfy the covering constraints steadily gets larger, until, for a large enough radius, any allocation would satisfy the covering constraints, reducing the optimization to a pure packing problem. Since the feasibility set only
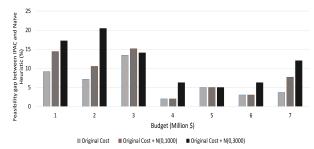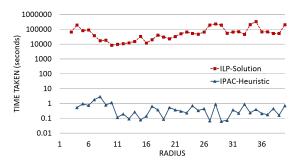
(a) Demand satisfied by LP-relaxation, Optimal ILP-solution, IPAC-heuristic, and Naive-heuristic.

(b) Fraction of ILP-solution demand satisfied by IPAC-heuristic and Naive-heuristic.

(c) Minimum required budget for feasibility by IPAC-heuristic and LP-relaxation.

(d) Increase in demand satisfied by IPAC-heuristic over Naive-heuristic.

(e) Feasibility gap between IPAC-heuristic and Naive-heuristic.

(f) Running time of IPAC as compared to directly solving the ILP using CPLEX.

FIGURE 4. Experimental results for North East England.

gets larger with r, the demand covered should also increase accordingly. The graphs validate this expected behavior. In addition, we observe that the demand captured by IPAC is far closer to that of the LP relaxation than that captured by the naive heuristic. In particular, from Figure 4(b), we can see that when $r = 9$km, the IPAC heuristic already captures almost 90% of the LP relaxation demand and its performance is always better than the naive heuristic. Also, the time taken to obtain the ILP solution for a given radius using CPLEX [10],[9] ranges from $23 - 30$ hours for most instances, whereas IPAC finished in less than 0.5 seconds; see Figure 4(f).

3.3.3. Feasibility. Since IPAC is a heuristic, it may falsely deem an instance of the MPC problem (Figure 3(a)) infeasible, when in fact, feasible solutions exist. Thus, it is important to first analyze the extent of this limitation. It is easy to observe that an instance of MPC is feasible

---

[9] Intel(R) Xeon (R) CPU @ 2.2 GHz (16 cores), 32GB RAM, 64-bit Windows.

if and only if the available budget B is greater than or equal to the minimum required budget determined by solving the corresponding instance of the set-cover subproblem (Figure 3(c)). In the case of IPAC (Algorithm 3.2.1), it reduces to whether the available budget B is greater than or equal to the minimum required budget as determined by SC().

Therefore, given an instance of MPC, we look at the corresponding set-cover subproblem and compare its solutions (the minimum required budgets) as obtained by (i) using the greedy approximation algorithm in [33], and (ii) directly solving its relaxed LP.[10] We use different values of r (in units of km) to generate instances. The results are plotted in Figure 4(c). (Since the LP relaxation allows for fractional allocations, the corresponding solution is only a lower bound on the minimum budget required for feasibility.)

3.3.4. Accounting for noisy costs. Since our model for estimating the costs may not be accurate, we perform experiments for two additional scenarios, where we add zero-mean gaussian noise to the costs — one with a standard deviation of \$1000 and another with a standard deviation of \$3000. In each case, we compare IPAC and the naive heuristic:

• *Performance:* Using a budget of \$6M, we calculate the difference in the demands satisfied by the solution obtained using IPAC and the naive heuristic. Figure 4(d) shows this difference as a percentage, for several values of the reachability radius. It can be seen, once again, that IPAC's advantage is significant (8-18% in most cases).

• *Feasibility:* We calculate the minimum reachability radius for which IPAC and the naive heuristic are able to find a feasible solution to MPC, and find that IPAC is always able to find feasible solutions for smaller radii. Figure 4(e) shows this difference as a percentage, for several values of the available budget. It can be seen that IPAC's advantage over the naive heuristic is particularly remarkable (10-20% in most cases) when the budget is scarce.

Thus, experimental results show that IPAC's feasibility gap as compared to the LP-Relaxation is not significant, and IPAC's performance quickly approaches that of the LP-Relaxation demand. Further, in terms of both feasibility and performance, it can be seen that the advantage of IPAC over the naive heuristic is significant (10-20% in most cases) and this advantage is fairly robust to noise in the estimated costs.

4. Extensions and future work. In this section, we explain how the optimization framework we introduced in Section 3 can be extended to a situation where the deployment of charging stations is incremental, and conclude with some open questions.[11] To highlight another important alternate scenario, consider a set of private charging station providers that would like to set up charging stations, but would need government subsidies to incentivize them to do so, especially in regions of low demand. The government agency would then run a grant program [2, 25], where the providers specify the subsidies they need to set up charging stations at candidate sites they are interested in. The government has a budget from which to allocate grants; moreover, the providers come with their own budget constraints. In Appendix B, we expand the placement optimization problem (MPC) to accommodate these additional packing constraints and explain how to extend the IPAC framework accordingly by solving a multi-dimensional knapsack subproblem.

---

[10] Computing ILP solutions for feasibility analysis exactly is prohibitively time-consuming, since it must be computed for various r and B values.

[11] Even though the work in this paper is motivated by EV charging stations, the problem formulation, modeling and solution frameworks presented here can be extended to other facility location problems which have demand, budget and coverage requirements, such as placement of bus-stop shelters, parking lots, and healthcare kiosks.

4.1. Incremental (multi-period) placement. In practice, the budget for the deployment of charging stations may be released over time, and the demands at candidate sites and charging station installation costs may change over time. We now extend the problem formulation and the heuristic framework presented in Section 3 for this case. Let there be T time periods, $t = 1, \ldots, T$. We first extend earlier notation to include time periods, by superscripting them with t. Then, the single-period formulation in Figure 3(a) and Equation (2) can be combined and generalized for the multi-period placement optimization problem as follows:

$$
\begin{aligned}
\max_{\substack{r^t \in [R^{\min}, R^{\max}] \\ x_i^t \in \{0,1\}}} &\sum_{t=1}^{T} \left(\alpha \frac{\sum_{i \in \mathcal{L}^t} d_i^t x_i^t}{\sum_{i \in \mathcal{L}^t} d_i^t} + (1 - \alpha) \frac{R^{\max} - r^t}{R^{\max} - R^{\min}} \right) \\
\text{subject to } &x_i^t \geq x_i^{t-1} \qquad \forall\, i \in \mathcal{L}^{t-1}; 1 \leq t \leq T \\
&\sum_{t=1}^{\tau} \sum_{i \in \mathcal{L}^t} c_i^t (x_i^t - x_i^{t-1}) \leq \sum_{t=1}^{\tau} B^t \quad \forall\, 1 \leq \tau \leq T \\
&\sum_{i \in \mathcal{L}: \ell \in S_i^{r^t}} x_i^t \geq 1 \qquad \forall\, \ell \in \mathcal{I}^t; 1 \leq t \leq T
\end{aligned}
\tag{3}
$$

In the above formulation, initially, $x_i^0 = 0$ at all candidate sites $i \in \mathcal{L}^0$. If a charging station is placed at location i at time $t > 0$, then x_i^t is set to 1, and the first constraint ensures that the charging station remains installed for all subsequent periods. The second constraint ensures that the cost of all (new) installations done at time τ is within the budget released at τ plus any leftover budget from previous time periods, and the third set of constraints is the covering constraints for each time period t, as a function of the reachability radius r^t. The objective is to maximize the fraction of demand satisfied while minimizing the reachability radius, summed over all time periods. A straightforward "greedy" heuristic for this problem is to predict demands and estimate costs at the remaining candidate sites at the beginning of each time period t, and then using Algorithm 1 (IPAC) to locate additional charging station installations.

A common and important scenario that any solution for incremental placement must handle is future increase in demand at previously deployed charging stations. The greedy heuristic above only indirectly addresses this concern by ensuring that the increased demand at such charging stations would be taken into account by the demand prediction algorithm in subsequent time periods. Still, a simple modification can allow for the option of expanding existing charging stations by adding more charging slots in response to increasing demand—let the set of candidate sites at time period $t + 1$ be a superset of the set of candidate sites at time period t, i.e., $\mathcal{L}^{t+1} \supseteq \mathcal{L}^t$, but the costs c_i^{t+1} at previously selected candidate sites i would now be the cost of adding additional charging slots to maintain the waiting time constraints. This would let IPAC decide whether the right thing to do in response to increasing demand is to expand an existing charging station or set up a new charging station nearby. While our greedy heuristic is a starting point, more effective algorithms likely exist, especially when future demands and installation costs can be predicted with reasonable accuracy. We leave such directions to future work.

4.2. Extending the IPAC framework. Further analysis of the class of algorithms defined by the IPAC framework would be of interest. In particular, it is worth exploring provable performance guarantees of an IPAC algorithm, e.g., Algorithm 1, in terms of those of its constituent knapsack and set-cover algorithms. Another interesting direction would be to investigate if the demand prediction model can be integrated into the optimization framework for the placement of charging stations, since the placement of a charging station at one candidate site would likely affect the demand at nearby candidate sites.

References

[1] 114th Congress. 2015. H.R.22 – Surface Transportation Reauthorization and Reform Act. http://www.congress.gov/bill/114th-congress/house-bill/22.

[2] Armistead, Thomas F. 2015. California Offers \$10-Million Grant for Electric Vehicle Charging Stations. http://www.enr.com/articles/4435-california-offers-10-million-grant-for-electric-vehicle-charging-stations.

[3] Azar, Yossi, Umang Bhaskar, Lisa Fleischer, Debmalya Panigrahi. 2013. Online mixed packing and covering. *ACM-SIAM SODA*. 85–100.

[4] Borga, Magnus. 2001. Canonical Correlation: A Tutorial. http://people.imt.liu.se/magnus/cca.

[5] Chakaravarthy, Venkatesan T., Anamitra R. Choudhury, Sivaramakrishnan R. Natarajan, Sambuddha Roy. 2013. Knapsack cover subject to a matroid constraint. *IARCS FSTTCS*. 275–286.

[6] Chang, Daniel, Daniel Erstad, Ellen Lin, Alicia Falken Rice, Chia Tzun Goh, An-An (Angel) Tsao, Jason Snyder. 2012. Financial viability of non-residential electric vehicle charging stations. http://luskin.ucla.edu/sites/default/files/Non-Residential%20Charging%20Stations.pdf.

[7] Chen, T. Donna, Kara M. Kockelman, Moby Khan. 2013. Locating electric vehicle charging stations: Parking-based assignment method for Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board* **2385** 28–36.

[8] Chung, Sung Hoon, Changhyun Kwon. 2015. Multi-period planning for electric car charging station locations: A case of korean expressways. *European Journal of Operational Research* **242**(2) 677–687.

[9] Cobb, Jeff. 2014. December 2014 Dashboard. http://www.hybridcars.com/december-2014-dashboard.

[10] CPLEX, IBM ILOG. 2009. V12. 1: Users manual for CPLEX. *International Business Machines Corporation* .

[11] Csirik, J., J. B. G. Frenk, M. Labbé, S. Zhang. 1991. Heuristics for the 0&Ndash;1 Min-knapsack Problem. *Acta Cybernetica* **10**(1-2) 15–20.

[12] Dai, Jing D., Ajay A. Deshpande, Younghun Kim, Ming Li, Pavankumar Murali, Sambit Sahu. 2013. Electric vehicle (ev) charging infrastructure with charging stations optimumally sited. US Patent Publication No. US20130222158 A1.

[13] Department for Transport. 2013. Traffic Counts. http://www.dft.gov.uk/traffic-counts/area.php?region=North+East.

[14] Drezner, Zvi, ed. 1995. *Facility Location: A Survey of Applications and Methods*. Springer.

[15] European Union. 2014. Directive 2014/94/EU of the European Parliament and of the Council. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0094.

[16] Funke, Stefan, André Nusser, Sabine Storandt. 2015. Placement of Loading Stations for Electric Vehicles: No Detours Necessary! *Journal of Artificial Intelligence Research* **53** 633–658.

[17] Haklay, Mordechai (Muki), Patrick Weber. 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing* **7**(4) 12–18.

[18] Hardoon, David R., Sandor Szedmak, John Shawe-Taylor. 2004. Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neural computation* **16**(12) 2639–2664.

[19] Klabjan, Diego. 2012. Information systems and methods for deployment of charging infrastructure in support of electric vehicles. US Patent Publication No. US20120203726 A1.

[20] Klami, Arto, Seppo Virtanen, Samuel Kaski. 2013. Bayesian Canonical Correlation Analysis. *Journal of Machine Learning Research* **14**(1) 965–1003.

[21] Kleinrock, Leonard, ed. 1975. *Queueing Systems Volume 1: Theory*. Wiley–Interscience.

[22] Lam, Albert Y. S., Yiu-Wing Leung, Xiaowen Chu. 2013. Electric vehicle charging station placement. *IEEE Smart Grid Communications*. 510–515.

[23] Mukherjee, Koyel, Partha Dutta, Gurulingesh Raravi, Thangaraj Balasubramaniam, Koustuv Dasgupta, Atul Singh. 2015. Fair resource allocation for heterogeneous tasks. *IEEE IPDPS*. 1087–1096.

[24] Office for Low Emission Vehicles. 2012. National Charge Point Registry. http://data.gov.uk/dataset/national-charge-point-registry.

[25] Office for Low Emission Vehicles. 2013. Plug-in vehicle charge point grants. http://www.gov.uk/government/collections/plug-in-vehicle-chargepoint-grants.

[26] Office for Low Emission Vehicles. 2013. Plugged-in Places. http://www.gov.uk/government/publications/plugged-in-places.

[27] Press Release. 2014. Clean fuels for transport: Member States now obliged to ensure minimum coverage of refuelling points for EU-wide mobility. http://europa.eu/rapid/press-release_IP-14-1053_en.htm.

[28] Rai, Piyush, Hal Daume. 2009. Multi-Label Prediction via Sparse Infinite CCA. *NIPS*. 1518–1526.

[29] Storandt, Sabine, Stefan Funke. 2013. Enabling E-Mobility: Facility Location for Battery Loading Stations. *AAAI*.

[30] Sun, Liang, Shuiwang Ji, Jieping Ye. 2011. Canonical Correlation Analysis for Multilabel Classification: A Least-Squares Formulation, Extensions, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(1) 194–200.

[31] U. S. Department of Energy. 2012. EV Everywhere Grand Challenge: DOE's 10-Year Vision for Plug-In Electric Vehicles. http://energy.gov/eere/vehicles/ev-everywhere-grand-challenge-does-10-year-vision-plug-electric-vehicles.

[32] Valuation Office Agency. 2011. Property market report. http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371470/pmr_2011.pdf.

[33] Vazirani, Vijay V. 2001. *Approximation Algorithms*. Springer.

[34] Virtanen, Seppo, Arto Klami, Suleiman A. Khan, Samuel Kaski. 2012. Bayesian Group Factor Analysis. *AISTATS*. 1269–1277.

[35] Wagner, Sebastian, Tobias Brandt, Dirk Neumann. 2014. Smart City Planning – Developing an Urban Charging Infrastructure for Electric Vehicles. *ECIS*. http://aisel.aisnet.org/ecis2014/proceedings/track08/7/.

[36] Young, Neal E. 2001. Sequential and parallel algorithms for mixed packing and covering. *IEEE FOCS*. 538–546.

Appendix A: Approximation algorithm for a transformed problem. This section is an expanded version of the discussion towards the end of Section 3.1 which includes a proof of Lemma 1. We present an approximation algorithm for a pure covering problem that is closely related to the Mixed Pack & Cover (MPC) problem (Figure 3(a)), which serves as a basis for designing a heuristic framework for solving MPC.

Consider the following Demand and Set-Cover (DSC) problem, which does not contain any packing constraint, but is closely related to the MPC problem:

$$
\begin{aligned}
\min_{\{x_i : i \in \mathcal{L}\}} \quad & \sum_{i \in \mathcal{L}} c_i x_i \\
\text{subject to} \quad & \sum_{i \in \mathcal{L} : \ell \in S_i^r} x_i \geq 1 \qquad \forall \, \ell \in \mathcal{I} \\
& \sum_{i \in \mathcal{L}} d_i x_i \geq D
\end{aligned}
\tag{4}
$$

Intuitively, an algorithm for DSC can be instantiated with different values of total demand D, to search for a solution for MPC, e.g., with increasing values of total demand D (between $\min_{d_i}$ and $\sum d_i$) until the total cost exceeds B. The IPAC framework for MPC presented in Section 3.2.1 is inspired by this observation. Now, to develop an approximation algorithm for DSC, consider its two subproblems:

$$
\begin{aligned}
\min_{\{x_i : i \in \mathcal{L}\}} \quad & \sum_{i \in \mathcal{L}} c_i x_i \\
\text{subject to} \quad & \sum_{i \in \mathcal{L} : \ell \in S_i^r} x_i \geq 1 \qquad \forall \, \ell \in \mathcal{I}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
\min_{\{x_i : i \in \mathcal{L}\}} \quad & \sum_{i \in \mathcal{L}} c_i x_i \\
\text{subject to} \quad & \sum_{i \in \mathcal{L}} d_i x_i \geq D
\end{aligned}
\tag{6}
$$

Problems (5) and (6) are the well-known NP-Hard problems Set-Cover (SC) [33] and Minimization Knapsack (MinKP) [11], respectively. Let, OPT_{dsc}, OPT_{sc} and OPT_{MinKP} be the costs of an optimal solution of problems (4), (5), and (6), respectively. Since every feasible solution for problem (4) is a feasible solution for problem (5) as well as problem (6), $OPT_{sc} \leq OPT_{dsc}$ and $OPT_{MinKP} \leq OPT_{dsc}$.

Consider an f_{sc} factor approximation algorithm A_{sc} for set-cover and f_{kp} factor approximation algorithm A_{kp} for minimization knapsack. Let, $\mathcal{L}_{sc}$ and $\mathcal{L}_{kp}$ (subsets of $\mathcal{L}$) be the solutions returned by A_{sc} and A_{kp}, respectively. Consider algorithm A_{dsc} for DSC that uses A_{sc} and A_{kp} as subroutines, and returns the union of the solutions returned by the two subroutines, i.e., $\mathcal{L}_{sc} \cup \mathcal{L}_{kp}$, for problem (4). Clearly, $\mathcal{L}_{sc} \cup \mathcal{L}_{kp}$ is a feasible solution for DSC since locations in $\mathcal{L}_{sc}$ help satisfy the set-cover constraints, and locations in $\mathcal{L}_{kp}$ help satisfy the demand-cover constraint. Moreover, $c(\mathcal{L}_{sc} \cup \mathcal{L}_{kp}) \leq c(\mathcal{L}_{sc}) + c(\mathcal{L}_{kp}) \leq f_{sc}.OPT_{sc} + f_{kp}.OPT_{MinKP} \leq (f_{sc} + f_{kp}).OPT_{dsc}$. Here $c(\cdot)$ denotes the cost of a solution, that is, the sum of the costs of all locations that appear in a solution.

Thus, A_{dsc} is an $(f_{sc} + f_{kp})$ factor approximation algorithms for DSC. Substituting A_{sc} and A_{kp} with known set cover and minimization knapsack algorithms from literature gives different approximation algorithms for DSC. For example, using the f-factor algorithm for set-cover from [33] (where f is the frequency of the most frequent element, i.e., maximum number of reachability sets in which a candidate site appears), and the 2 factor algorithm for minimization knapsack from [11], gives an $(f + 2)$ factor algorithm for DSC.

Appendix B: Alternate scenario: Granting subsidies to charging station providers.
The problem modeled in Section 3 and its extension in Section 4.1 focused on a situation in which a central planner such as a government agency would use funds to construct charging stations in a geographical region. Hence, c_i referred to the cost to the government agency of setting up a charging station at i. An alternate scenario could be one where there is a set $\mathcal{P} = \{1, 2, \ldots, |\mathcal{P}|\}$ of private charging station providers that would like to set up charging stations, but would need government subsidies to incentivize them to do so, especially in regions of low demand. In this case, the government agency runs a grant program, where the providers specify the subsidies they need to set up charging stations at candidate sites they are interested in. Instead of c_i, we now have c_{ij} denoting the subsidy to be paid to provider $j \in \mathcal{P}$ at candidate site $i \in \mathcal{L}$, if selected.

We solicit the following additional inputs, from the private providers:[12] (a) preferences regarding which candidate sites the providers are interested in setting up and operating charging stations, (b) estimates of costs and budgets, and (c) bids on desired subsidies. We then determine the subset of candidate sites selected for deployment, the "winning" providers for each of the selected sites, and the subsidy amounts allocated to each of them. In addition to demand maximization, coverage and government agency budget constraints, the additional constraints that should be taken into consideration are the site preferences and budget constraints of the providers.

B.1. Optimal placement with subsidy allocation. A couple of concerns arise in allowing multiple private providers to compete for setting up and operating charging stations:

• Providers can express interest in multiple sites, even though they may not have the means to build charging stations at all those sites. Each provider $j \in \mathcal{P}$ would therefore also specify their budget B_j, and their estimated costs p_{ij} (after taking into account the subsidy c_{ij}) of building a charging station at candidate site $i \in \mathcal{L}$. It is then the government agency's responsibility to ensure that the outcome does not violate any provider's budget constraint. This adds additional *packing* constraints to the problem. Note that we allow for providers to specify *site-specific* subsidies, because the amount of money they are willing to spend setting up a charging station at a given site is likely to depend on that site's demand.

• There could be candidate sites (e.g., with relatively low demand) that no provider is interested in, or, even if there is interest, providers might ask for very high subsidies, so that even the least of these would cost the government agency more than what it would cost to build a charging station by itself. Furthermore, it is likely that some such sites with low demand are the only ones that cover a remote area, one of which must therefore be selected in any feasible allocation. To handle such scenarios, we model the government agency to be a "provider" as well, who, for every site, asks for a subsidy of c_i, which is the cost to the government agency of constructing its own charging station at that site.[13] Therefore, we add the government agency to the set of providers $\mathcal{P}$ and denote it with the index 0, to obtain the set of participants, $\mathcal{B} = \{0, 1, 2, \ldots, |\mathcal{P}|\}$.

To recap our notation for this scenario, $\mathcal{L} = \{1, 2, \ldots, |\mathcal{L}|\}$ denotes the set of all candidate sites for charging stations, $r \in \mathbb{R}_+$ denotes the reachability radius, $\mathcal{I}$ denotes the set of all locations of interest to be covered, for each candidate site $i \in \mathcal{L}$, $d_i \in \mathbb{R}_+$ is the demand for charging at i, and $S_i^r \subseteq \mathcal{I}$ is the set of locations of interest that would be covered if a charging station is placed at i.

Let $\mathcal{P} = \{1, 2, \ldots, |\mathcal{P}|\}$ denote the set of charging station providers. For each provider $j \in \mathcal{P}$,

1. $B_j \in \mathbb{R}_+$ is the total budget available,
2. $c_{ij} \in \mathbb{R}_+$ is the subsidy that the provider specifies they need at site $i \in \mathcal{L}$, and

---

[12] The procedure usually involves a screening stage where the proposals submitted by the providers are evaluated for the accuracy/legitimacy of their cost estimation and project feasibility by a grant committee in order to determine their eligibility for further consideration.

[13] This is equivalent to setting a *reserve* subsidy at each site.

3. $p_{ij} \in \mathbb{R}_+$ is the cost to the provider of building a charging station at site $i \in \mathcal{L}$, after taking into account the subsidy c_{ij}.

$\mathcal{B} = \mathcal{P} \cup \{0\}$ denotes the set of participants after adding the government agency to the set of providers. $B \in \mathbb{R}_+$ denotes the total budget available to the government agency, and $c_{i0} = c_i$ is the government agency's subsidy specification for site $i \in \mathcal{L}$. For each site $i \in \mathcal{L}$, each participant $j \in \mathcal{B}$, the outcome $x_{ij} \in \{0, 1\}$ denotes whether participant j "won" site i.

Given a set of prices, budgets and subsidy specifications from the participants, the outcome is determined by the following optimization problem:

$$
\begin{aligned}
\max_{\{x_{ij}:i\in\mathcal{L},j\in\mathcal{B}\}} & \sum_{i\in\mathcal{L}} d_i \sum_{j\in\mathcal{B}} x_{ij} \\
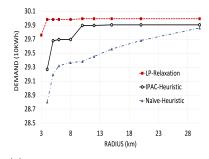\text{subject to} \ & \sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{B}} c_{ij} x_{ij} \leq B \\
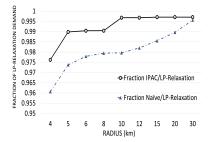& \sum_{i\in\mathcal{L}} p_{ij} x_{ij} \leq B_j \qquad \forall\, j \in \mathcal{P} \\
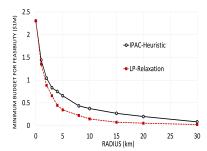& \sum_{j\in\mathcal{B}} x_{ij} \leq 1 \qquad\quad \forall\, i \in \mathcal{L} \\
& \sum_{k:\ell\in S_k^r}\sum_{j\in\mathcal{B}} x_{kj} \geq 1 \quad \forall\, \ell \in \mathcal{I}
\end{aligned}
\tag{7}
$$

The IPAC framework (and its instantiation in Algorithm 1) presented in Section 3.2.1 can be extended by (a) replacing calls to KP() with calls to MULTIKP(), a method that solves the multi-dimensional knapsack problem, and (b) incorporating the multiple knapsack costs of an item into the value function used in RANK(), and (c) within each iteration, reducing multiple knapsack budgets and making sure that items are removed until *all* reduced budget constraints are satisfied.

Appendix C: Second experiment for optimal placement: Western part of East Anglia and southern part of East Midland. We also used IPAC to find the locations for optimal placement of charging stations across the western region of East Anglia and southern region of East Midland, where there are very few existing charging stations. We use parking locations in these areas as candidate sites for placing charging stations. We predict the demand and estimate the cost for the candidate sites as we did for North East England and used a budget of $B = \$800,000$. To evaluate the performance IPAC with the naive heuristic and the LP relaxation, we used values of r (in units of km) from the set $\{4, 5, 6, 8, 10, 12, 15, 20, 30\}$. The corresponding results are shown in Figure 5. It can be seen that qualitatively, the results are very similar to those obtained for North East England (Section 3.3).

(a) Demand satisfied by LP-relaxation, IPAC-heuristic, and Naive-heuristic.

(b) Fraction of ILP-solution demand satisfied by IPAC-heuristic and Naive-heuristic.

(c) Minimum required budget for feasibility by IPAC-heuristic and LP-relaxation.

FIGURE 5. Experimental results for western part of East Anglia and southern part of East Midland.