

A Machine Learning Methodology for Optimal Big Data Processing in Advanced Smart City Environments

Alfredo Cuzzocrea[1,2(✉)], Luigi Canadè[1], Riccardo Nicolicchia[3], and Luca Roldo[3]

[1] iDEA Lab, University of Calabria, Rende, Italy
alfredo.cuzzocrea@unical.it
[2] Department of Computer Science, University Paris City, Paris, France
[3] ISIRES, Turin, Italy
{riccardo.nicolicchia,luca.roldo}@isires.org

Abstract. This paper introduces a *Machine Learning methodology for supporting optimal big data processing in advanced smart city environments*. In particular, the proposed work focuses the attention on the issue of efficiently charging *Electric Vehicles* (EVs) in an urban area, given various specific constraints. Our preliminary evaluation and analysis proofs the quality of our heuristics.

Keywords: Network Algorithms · Theory of Computing · Heuristic Algorithms

1 Introduction

Given the growing need and attention given to electric vehicles (EV), in the last decade many researchers have focused their attention on the optimal placement of EV chargers. Many have focused on the placement on large scale networks, to provide optimal placement considering multiple cities. These models consider a set of traffic flows with specific routes and locate charging stations along the routes so that the percentage of traffic flows performing their round trip without running out of charge is maximized, as proposed in [15,22,23,26].

Within the context of optimal location of charging stations for EV, the aim of this paper is exactly that of finding an optimal placement, in terms of location and quantity, of fast and slow charging stations within the parking lots of a urban environment. In particular, this project focuses on the optimal placement of EV charging stations in a urban environment under demand uncertainty conditions, given by variable characteristics of the traffic flows inside the city. Given the complexity of the EV charging station location problem when applied to a real case scenario, some effort has been spent into developing heuristic algorithms. Some related work can be found in [1,4,11]. In particular, the work in [1] was

A. Cuzzocrea—This research has been made in the context of the Excellence Chair in Big Data Management and Analytics at University of Paris City, Paris, France.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023, LNCS 13956, pp. 713–730, 2023.
https://doi.org/10.1007/978-3-031-36805-9_46

analyzed in depth to develop an heuristic algorithm for this paper. Another inspiring line of research has been represented by the emerging *big data call for arms* (e.g., [8,20]).

This paper thus introduces a *Machine Learning methodology for supporting optimal big data processing in advanced smart city environments*. Our preliminary evaluation and analysis proofs the quality of our heuristics. This paper extends the preliminary paper presented in [9].

The model takes as input a map containing a set of buildings and parking lots. The EV users' destination is defined by a specific building, therefore the public parking lots surrounding the buildings are considered as potential location for the charger installation. The model is characterized by the following assumptions:

- Drivers visit two buildings during the day.
- Drivers park their vehicle in a parking lot near the building of interest. Parking lots are considered to be near a building based on a specific distance parameter, that represent an approximate maximum distance that a user is willing to walk.
- EV users are divided in groups, representing traffic flows. Each group is characterized by: the two buildings they visit during the day, the arrival time at the first building, the dwell times in each building and the time they spend travelling from the first building to the second.
- The parameters of each group are deterministic and known with certainty.
- For each group, one single recharging is sufficient to complete the whole path from home to the first building, then to the second one and finally back home.
- The charging behaviour (duration and choice between fast and slow chargers) depends on the state of charge when arriving at the first building. Users of each group can therefore be divided in three different classes, described in Table 1, in which β_i represent the ratio of users of each group belonging to class i. The "Charge state" column represents the charging state of the battery when arriving at the first building and it is divided in three states: *low*, with 1/3 of charge remaining; *middle*, with 2/3 of charge remaining; *high*, with full battery.
- The dwell time in each building is divided in three cases: *short* (θ_1), *medium* (θ_2) and *long* (θ_3) duration.
- Drivers prefer to have the battery fully charged after recharging. Drivers will select the charging mode (*fast/slow*) based on their class (*EV charge state*) and the dwell time in each building. All the possible charging modes selections

Table 1. Classification of Drivers Based on Charge State.

Class	Charge state	Characteristics	Ratio
1	Low	Charge the EV when in the first building	β_1
2	Middle	Can charge the EV when in the first or second building	β_2
3	High	Charge the EV when in the second building	β_3

are listed in Table 2. A combination of class, choice of charging when in the first or second building, dwell time and consequent charging mode is called "type".

Table 2. Type Definition: Charging Modes Selection by Different Classes of Drivers.

Type	Class	Building	Dwell Time	Charging mode
1	1	First	Short, Medium	Fast
2	1	First	Long	Slow
3	2	First	Short	Fast
4	2	First	Medium, Long	Slow
5	2	Second	Short, Medium	Fast
6	2	Second	Long	Slow
7	3	Second	Short	Fast
8	3	Second	Medium, Long	Slow

- The cost of installing a fast or slow charger in a parking lot as well as the purchase cost of a single charger are known with certainty.
- A limited number of chargers can be installed in the whole considered environment. This assumption addresses a budget restriction.
- There is a limitation also on the number of chargers that can be installed in a single parking lot, to set a limit in the physical capacity as well as in the energy consumption of each parking lot.
- There is an upper bound on the number of parking lots that can be equipped with chargers.
- The number of drivers of each group is uncertain and its variation generates multiple possible scenarios.

Based on the listed assumption, the problem is two-stage stochastic programming model. In the first stage parking lots that should be equipped with chargers are determined, as well as the number of fast and slow chargers for each parking. In the second stage, for each scenario, the number of drivers in each group is computed and EVs are assigned to parking lots accordingly. The objective function must take into account both the cost of purchasing and installing the chargers as well as maximizing the coverage of demands, that can be interpreted also as the total income achieved from users charging their vehicles.

For convenience, time is divided in slots, with each slot assumed to be 1 hour long. Let $\mathbb{T}$ (indexed by t) be the set of time-slots, $\mathbb{B}$ (indexed by b) be the set of buildings, $\mathbb{P}$ (indexed by p) be the set of parking lots and $\mathbb{G}$ (indexed by g) be the set of groups. Assume that $\bar{b}_{1,g}, \bar{b}_{2,g} \in \mathbb{B}$ are the first and second buildings visited by EV users of group g, and $\bar{t}_g$ is the time slot at which users of group g arrive at the first building $\bar{b}_{1,g}$. Dwell times, expressed in number of slots, of group g at buildings $\bar{b}_{1,g}$ and $\bar{b}_{2,g}$ are denoted by $d_{1,g}$ and $d_{2,g}$, respectively. The travel time from $\bar{b}_{1,g}$ to $\bar{b}_{2,g}$ is denoted by d'_g. Parameter α represents the maximum distance that a user of every group is willing to walk to reach a building from a parking lot. The fixed costs of equipping a parking lot with fast and slow

chargers are represented by F' and F'', respectively, and the cost of purchasing one fast or slow charger is denoted by C' and C''. At most, M parking lots can be equipped with chargers, and a total of M' chargers are installable. Moreover, the number of chargers that can be installed in any parking lot is limited by M''. The number of drivers of each group is uncertain and in this regard, $\mathbb{S}$ (indexed by s) represents the set of possible scenarios. Variable dem_g^s represents the number of drivers of group g under scenario s, while π_s represents the probability of the occurrence of scenario s. The meaning of each described set and parameter is summarized in Table 3.

Table 3. Summary of Sets and Parameters

Name	Description
$\mathbb{T}$	Set of time slots (indexed by t)
$\mathbb{B}$	Set of buildings (indexed by b)
$\mathbb{G}$	Set of groups of EV users (indexed by g)
$\mathbb{P}$	Set of parking lots (indexed by p)
$\mathbb{S}$	Set of scenarios (indexed by s)
π_s	Occurrence prob. of scenario s
$\bar{b}_{i,g}$	i^{th} building visited by users of group g ($i = 1, 2$)
$\bar{t}_g$	Time-slot in which users of group g arrive at first building
$d_{i,g}$	Dwell-time (in time slots) of group g at building $\bar{b}_{i,g}$
d'_g	Time (in time slots) spent by group g to travel from $\bar{b}_{1,g}$ to $\bar{b}_{2,g}$
dem_g^s	Number of drivers of group g under scenario s
α	Maximum walking distance
β_i	Fraction of drivers in a group belonging to class i ($i = 1, 2, 3$)
θ_i	Indicating the duration (in time-slots) of short, medium and
	long dwell times ($i = 1, 2, 3$)
E'	Income achieved by charging an EV with a fast charger in a unit of time
E''	Income achieved by charging an EV with a slow charger in a unit of time
F'	Fixed cost of equipping a parking lot with fast chargers
F''	Fixed cost of equipping a parking lot with slow chargers
C'	Cost of purchasing a fast charger
C''	Cost of purchasing a slow charger
M	Maximum number of parking lots that can be equipped with chargers
M'	Total number of available chargers
M''	Maximum number of chargers that can be installed in any parking lot

1.1 Map Generation

In order to work, the model needs a set of parking lots, buildings and the distance among them. To do so, an algorithm was built that generates pseudo-random city environments and provides realistic and coherent data to the model. The algorithm takes as input the number of buildings and parking lots to place, the minimum distance among buildings and the desired length of the side of the square map. The algorithm works as follows: (i) Creates a square matrix whose dimensions are those specified by the input settings. This matrix represents a physical map in which each location is identified by the coordinates of the cell. It is assumed

that each cell is $10m \times 10m$. Therefore each provided distance is specified in units of 10 m. (*ii*) Places each building, one at the time. The x and y coordinates are randomly chosen, taking however into consideration that no building is already present in that cell or within a certain radius from that cell; the value of the radius is specified in the input settings. (*iii*) Places each parking, in the same fashion of the building placement. Parking lots can be placed within the input radius of a building, but not within an input radius of another parking lot.

In order to check the connectivity between building and parking lots (i.e. which parking lots can be used by EV users to access a building) the map can be translated in a graph, in which parking lots and buildings are nodes and a link between a parking lot and a building is present if the distance among them is lower than the maximum walking distance α. An example with 20 buildings, 10 parking lots, a map side length of 20000 meters, a minimum distance between buildings of 1000meters is shown in Fig. 1. Building indicators have different sizes based on their popularity, which is randomly assigned to buildings when the map is created and whose usage will be described in the next section. The connectivity between parking and buildings is shown in Fig. 2. It is noted that some buildings result inaccessible from any parking lot; groups accessing these buildings will not be considered, for the time they stay in the inaccessible building, in the calculation of the optimal placement of the EV chargers since they will not occupy any present parking lot.

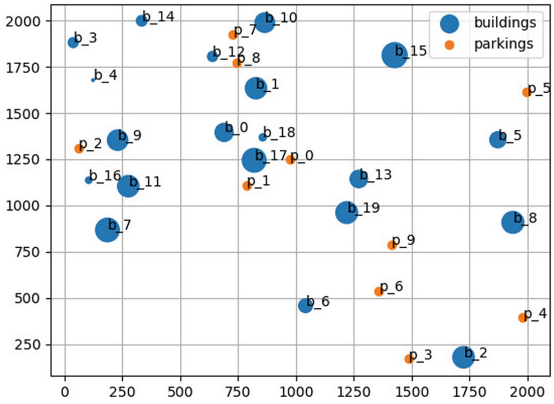


Fig. 1. Urban Environment Map.

1.2 Group Generation

A group is characterized by EV users visiting the same first and second building, have the same dwell time in both the first and second building, arrive at the first building at the same time and take the same time to travel from the first building to the second. Given these characteristics, many groups have to be generated in

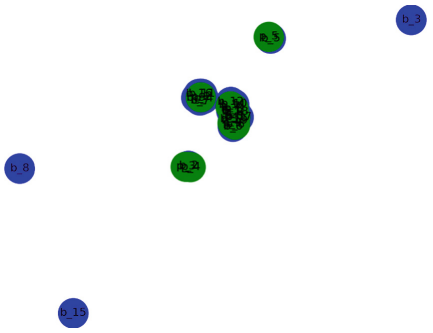


Fig. 2. Connectivity between Buildings and Parking Lots.

order to have a realistic flow of EV users. In this problem, group generation is performed by taking random (and different) first and second building, a random arrival time slot, two random dwell times for both the first and second building, and a travel time proportional to the distance between the first and second building. The algorithm also takes care not to generate two identical groups.

1.3 Group Sizing

The number of drivers in each group g, denoted by dem_g^s, is scenario dependent and therefore affected by uncertainty. In each scenario, the number of drivers is assigned to each group using the following procedure: (i) A popularity index q_b, in the range $[0, 1]$, is assigned to each building. (ii) A total amount of population is assigned to each building taking into account the popularity index and the mean (m_{pop}) and variance (σ_{pop}) of one building's population, provided in the input settings. In particular, to each building is assigned an amount of people equal to a realization of $\mathcal{N}(q_b m_{pop}, \sigma^2_{pop})$. (iii) Then, the total population of each building is randomly distributed to all those groups accessing that building, as first or second. Since a group accesses two buildings, its value of dem_g^s will be the result of the sum of this random assignment from the first building and that from the second building.

2 Mathematical Model

Looking at the actual literature, some deterministic approaches in urban environments consider several different factors to make the model more realistic and accurate. The model in [6] considers travelers having two main stops during the day (as was considered also in this paper) and drivers that cannot charge their EV privately. The works in [12] and [13] consider the drivers' spontaneous adjustments and interactions of travel and recharging decision. The model in [21] considers also environmental factors for siting and sizing of EV charging stations. [24] considers real vehicle travel patterns. The study in [14] considers capacitated chargers and flow deviations due to recharging needs.

Turning back to our main problem, in order to properly define constraints and objective function, two subsets are introduced. Subsets $\mathbb{G}'_{i,j,p}$ contain all those groups whose dwell time in their i^{th} building (with $i = 1, 2$) is equals to θ_j (with $\theta_j = 1, 2, 3$)and parking lot p is a possible parking lot when visiting that in building. In particular:

$$\mathbb{G}'_{i,j,p} = \{g \in \mathbb{G} : d(p, \bar{b}_{i,g}) < \alpha \text{ and } d_{i,g} = \theta_j\} \tag{1}$$

where $d(p, \bar{b}_{i,g})$ is the physical distance between parking p and building $\bar{b}_{i,g}$.

Furthermore, subsets $\mathbb{G}''_{t,p,r}$ contain those groups that use the charging mode r (defined in row r in Table 2) and hence may occupy parking lot p in time slot t. All the decision variables of the problem are reported in Table 4.

Table 4. Decision Variables

Name	Description
δ_p	Binary variable that is 1 if parking lot p is equipped with chargers, 0 otherwise
δ'_p	Binary variable that is 1 if parking lot p is equipped with at least one fast charger, 0 otherwise
δ''_p	Binary variable that is 1 if parking lot p is equipped with at least one slow charger, 0 otherwise
γ'_p	Number of fast chargers installed in parking lot p
γ''_p	Number of slow chargers installed in parking lot p
$y^s_{p,g}$	Number of drivers of group g belonging to class 1 and charge their EV in parking lot p during the time they are in the first building under scenario s
$x^s_{p,g}$	Number of drivers of group g belonging to class 2 and charge their EV in parking lot p during the time they are in the first building under scenario s
$x'^s_{p,g}$	Number of drivers of group g belonging to class 2 and charge their EV in parking lot p during the time they are in the second building under scenario s
$w^s_{p,g}$	Number of drivers of group g belonging to class 3 and charge their EV in parking lot p during the time they are in the second building under scenario s

Given the defined parameters and decision variables, the objective function is formulated as follows:

$$\begin{aligned}
\max & -\sum_{p\in\mathbb{P}}\left(F'\delta'_p + F''\delta''_p + C'\gamma'_p + C''\gamma''_p\right) \\
& + \sum_{s\in\mathbb{S}}\pi^s \sum_{p\in\mathbb{P}}\left(E'\left[\theta_1\left(\sum_{p\in\mathbb{G}'_{1,1,p}}\left(y^s_{p,g}+x^s_{p,g}\right)+\sum_{p\in\mathbb{G}'_{2,1,p}}\left(w^s_{p,g}+x'^s_{p,g}\right)\right)\right.\right. \\
& \left.\left. + \theta_2\left(\sum_{p\in\mathbb{G}'_{1,2,p}} y^s_{p,g} + \sum_{p\in\mathbb{G}'_{2,2,p}} x'^s_{p,g}\right)\right] + E''\left[\theta_2\left(\sum_{p\in\mathbb{G}'_{1,2,p}} x^s_{p,g} + \sum_{p\in\mathbb{G}'_{2,2,p}} w^s_{p,g}\right)\right.\right. \\
& \left.\left. + \theta_3\left(\sum_{p\in\mathbb{G}'_{1,3,p}}\left(y^s_{p,g}+x^s_{p,g}\right)+\sum_{p\in\mathbb{G}'_{2,3,p}}\left(w^s_{p,g}+x'^s_{p,g}\right)\right)\right]\right)
\end{aligned} \tag{2}$$

As previously mentioned, the objective function makes a trade-off between minimizing the costs of purchasing and equipping chargers and maximizing the expected coverage demands (earnings obtained by charging their EV). The constraints of the problem are defined in Table 5. Descriptions of constraints are reported in Table 5.

3 Heuristic Model

Two different heuristic models are considered. The first solves the problem presented in the Sect. 1. The second solves an improved version of the former problem because it allows the EV users whose charging mode is slow to park their vehicle in an available fast charger if all the slow ones are occupied. By using this second method, it is possible to obtain a higher value of the objective function.

Both heuristic models are divided into a first and a second stage and operate on the same data generated for the math model. The first stage gives as output the placement of the charging stations (both fast and slow chargers) in the parking lots. The second stage returns the number of EV users that recharge their vehicle in the installed chargers. The drivers are divided according to: (i) the type of used charger; (ii) the dwell time in the building in which they recharge their EV.

Those values are used to evaluate the resulting objective function. Given the variables reported in Table 6, the objective function is evaluated as follow:

$$\begin{aligned}
\max & -\sum_{p\in\mathbb{P}}\left(F'\delta'_p + F''\delta''_p + C'\gamma'_p + C''\gamma''_p\right) \\
& + \sum_{s\in\mathbb{S}}\pi^s\left(E'\left[\theta_1 nF_1 + \theta_2 nF_2 + \theta_3 nF_3\right]\right. \\
& \left. + E''\left[\theta_1 nS_1 + \theta_2 nS_2 + \theta_3 nS_3\right]\right)
\end{aligned} \tag{3}$$

If the resulting value of the objective function is negative, then the heuristic model does not place any charger at all.

Table 5. Constraints

#	Constraint	Description
1	$\sum_{p\in\mathbb{P}} \delta_p \leq M$	Ensures that at most M parking lots are equipped with chargers
2	$\delta'_p + \delta''_p \leq 2\delta_p \quad \forall p \in \mathbb{P}$	Ensures that if $\delta_p = 0$, also δ'_p and δ''_p are zero
3	$\gamma'_p \leq M''\delta'_p \quad \forall p \in \mathbb{P}$	Ensures that if $\delta'_p = 0$, also $\gamma'_p = 0$
4	$\gamma''_p \leq M''\delta''_p \quad \forall p \in \mathbb{P}$	Ensures that if $\delta''_p = 0$, also $\gamma''_p = 0$
5	$\gamma'_p + \gamma''_p \leq M'' \quad \forall p \in \mathbb{P}$	Restricts the total number of chargers installed in parking lot p
6	$\sum_{p\in\mathbb{P}} (\gamma'_p + \gamma''_p) \leq M'$	Restricts the total number of chargers installed in the network
7	$\sum_{p\in\mathbb{P}:d(p,\bar{b}_{1,g})\leq\alpha} y^s_{p,g} \leq \beta_1 dem^s_g \quad \forall g \in \mathbb{G}, \forall s \in \mathbb{S}$	Restricts the number of class 1 drivers in each group
8	$\sum_{p\in\mathbb{P}:d(p,\bar{b}_{1,g})\leq\alpha} x^s_{p,g} + \sum_{p\in\mathbb{P}:d(p,\bar{b}_{2,g})\leq\alpha} x'^s_{p,g} \leq \beta_2 dem^s_g \quad \forall g \in \mathbb{G}, \forall s \in \mathbb{S}$	Restricts the number of class 2 drivers in each group
9	$\sum_{p\in\mathbb{P}:d(p,\bar{b}_{2,g})\leq\alpha} w^s_{p,g} \leq \beta_3 dem^s_g \quad \forall g \in \mathbb{G}, \forall s \in \mathbb{S}$	Restricts the number of class 3 drivers in each group
10	$\sum_{g\in\mathbb{G}''_{t,p,1}} y^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,3}} x^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,5}} x'^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,7}} w^s_{p,g} \leq \gamma'_p$ $\forall p \in \mathbb{P}\ \forall t \in \mathbb{T}, \forall s \in \mathbb{S}$	The number of EVs using fast chargers of parking lot p at time t cannot exceed the number of installed fast chargers
11	$\sum_{g\in\mathbb{G}''_{t,p,2}} y^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,4}} x^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,6}} x'^s_{p,g} + \sum_{g\in\mathbb{G}''_{t,p,8}} w^s_{p,g} \leq \gamma''_p$ $\forall p \in \mathbb{P}\ \forall t \in \mathbb{T}, \forall s \in \mathbb{S}$	The number of EVs using slow chargers of parking lot p at time t cannot exceed the number of installed slow chargers
14	$\delta_p, \delta'_p, \delta''_p \in \{0,1\}, \gamma'_p, \gamma''_p \geq 0, \text{Integer} \quad \forall p \in \mathbb{P}$	Type of variables definition
15	$y^s_{p,g}, x^s_{p,g} \geq 0 \quad \forall g \in \mathbb{G}, \forall p \in \mathbb{P} : d(p,\bar{b}_{1,g}) \leq \alpha, \forall s \in \mathbb{S}$	Type of variables definition
16	$x'^s_{p,g}, w^s_{p,g} \geq 0 \quad \forall g \in \mathbb{G}, \forall p \in \mathbb{P} : d(p,\bar{b}_{2,g}) \leq \alpha, \forall s \in \mathbb{S}$	Type of variables definition

Table 6. Heuristic Variables

Name	Description
nF_1	Number of drivers that recharge their vehicle in a fast charger and whose dwell time is short
nF_2	Number of drivers that recharge their vehicle in a fast charger and whose dwell time is medium
nF_3	Number of drivers that recharge their vehicle in a fast charger and whose dwell time is long
nS_1	Number of drivers that recharge their vehicle in a slow charger and whose dwell time is short
nS_2	Number of drivers that recharge their vehicle in a slow charger and whose dwell time is medium
nS_3	Number of drivers that recharge their vehicle in a slow charger and whose dwell time is long

3.1 First Stage

The first stage of the heuristic model determines the parking lots that should be equipped with chargers and the number of fast and slow chargers for each parking. The variables used are reported in Table 8 and in Algorithm 1 is given the pseudocode. From the fraction of drivers belonging to each class and the cost of purchasing and positioning the chargers, it is computed the number of fast and slow chargers available. Afterward, the maximum number of chargers that can be installed in a parking lot (M") is assigned to the most popular parking lots on the map, firstly fast chargers are positioned and then slow. The outputs of the first stage are two lists, one for slow chargers and the other for fast, where, for each parking lot equipped, it is reported the number of chargers installed (Table 7).

Table 7. Heuristic First Stage Sets

Name	Description
P_{sc}	The set that contains the number of slow chargers for each parking lot
P_{fc}	The set that contains the number of fast chargers for each parking lot

3.2 Second Stage

The second stage of the heuristic model considers the main concepts of [1] and reuse them by adapting them to the different model: (*i*) Connection Matrix; (*ii*) Charging Site Degree; (*iii*) Electric Vehicle Degree.

The sets and parameters to be added to those already present in the mathematical model are given in Table 9 (i can take the value f if it refers to fast chargers, s for slow chargers).

Algorithm 1. Heuristic Model - First Stage

Initialize P_{sc} and P_{fc}
$perc_s = \beta_2/2 + \beta_1/3 + 2\beta_3/3$
$perc_f = \beta_2/2 + 2\beta_1/3 + \beta_3/3$
$price_f = E'/(C' + F')$
$price_s = E''/(C'' + F'')$
$tot_price = price_s + price_f$
$price_s = price_s/tot_price$
$price_f = price_f/tot_price$
$Tot_{charg} = min(M', M \cdot M'')$
$av_i = floor(Tot_{charg} \cdot (perc_i + price_i)/2)$
Sort the set $\mathbb{B}_P$ according to the sum of all the popularity indexes
for $(p \in arg(\mathbb{B}_P))$ **do**
  if $(N_e \geq M)$ **then**
    STOP
  else
    $N_e = N_e + 1$
    if $(av_f \geq M'')$ **then**
      $P_{fc}[p] = M''$
      $av_f = av_f - M''$
    else if $(av_f < M'' \wedge av_f > 0 \wedge av_s \geq (M'' - av_f))$ **then**
      $P_{fc}[p] = av_f$, $P_{sc}[p] = M'' - av_f$
      $av_s = av_s - (M'' - av_f)$, $av_f = 0$
    else if $(av_f = 0 \wedge av_s \geq M'')$ **then**
      $P_{sc}[p] = M''$
      $av_s = av_s - M''$
    else if $(av_f = 0 \wedge av_s < M'' \wedge av_s > 0)$ **then**
      $P_{sc}[p] = av_s$
      $av_s = 0$
    else
      STOP
    end if
  end if
end for
return P_{sc}, P_{fc}

Table 8. Heuristic First Stage Variables

Name	Description
$perc_i$	The percentage of chargers of type i based on the value of $\beta_1, \beta_2, \beta_3$
$price_i$	The percentage of chargers of type i based on the price and on the revenue of the i chargers
av_i	The number of available chargers of type i
$\mathbb{B}_P[p]$	The set of the buildings that can be reached from the parking lot p
N_e	The number of equipped parking lots
Tot_{charg}	The maximum number of chargers that can be placed

Table 9. Heuristic Additional Sets and Parameters

Name	Description
S_i	The set of all the activated charging sites
R_i	The set of all the activated charging sites that are not completely full
O_i	The set of all the occupied chargers per parking lots
K_i	The set of users that arrive in the selected time slot and that need the charging of type i
P_i	The set of all the available charging sites of type i that are not analyzed yet
AR	The set that keeps track of the number of people of the second class that have already recharged their EV in the first building, so they do not need to charge it in the second one
TR_i	The set that contains all the users that must leave for every time slot
$\mathbb{P}_i$	The set of the parking lots containing chargers of type i

The pseudocode of the steps of the second stage is reported in Algorithm 2. Before launching the algorithm, the input structure that contains the groups is modified to divide the people belonging to the groups into subgroups according to the following parameters: (*i*) the visited building: note that the first and the second building are considered separately; (*ii*) the arrival time slot to the visited building; (*iii*) the charging mode needed; (*iv*) the dwell time inside the building; (*v*) the cardinality of the considered subgroup; (*vi*) the class of the considered subgroup; (*vii*) the ID number of the group.

Then this new set of subgroups is grouped according to other two parameters: (*i*) the scenario; *ii*) the time slot of arrival to the building.

Connection Matrix. The Connection Matrix C_i is a matrix ($\mathbb{P}_i$ x $\mathbb{B}$) and its element $C_i(p, b)$ is equal to 1 if the building b can be reached from the parking lot p, 0 otherwise. It happens when the euclidean distance between the building and the parking lot is smaller than the maximum walking distance α.

Charging Site Degree. The Charging Site Degree $CSD(p)$ is the total number of users that can park in the selected parking lot$p$.

Algorithm 2. Heuristic Model - Second Stage

Initialization
Calculate the Connection Matrices C_i
Initialize the variables in Table 6
for ($s \in \mathbb{S}$) **do**
  Initialize S_i, R_i, TR_i, O_i, AR
  for ($t \in \mathbb{T}$) **do**
    Initialize K_i, P_i
    Calculate EVD_i ▷ considering C_i and $g \in K_i$
    $O_i, S_i, R_i = remove_users_function[TR_i, O_i, S_i, R_i]$
(Algorithm 3) ▷ Delete all the users that should leave in the current time slot
    $K_i = K_i - AR$ ▷ Delete the users already served
    $K_f, S_f, R_f, O_f = solve_second_stage_subproblem_function$
$[K_f, P_f, S_f, R_f, O_f, EVD_f, C_f]$
(Algorithm 4) ▷ Add the users that arrive in the current time slot who need fast chargers to the available parking lots' chargers
    $K_s, S_s, R_s, O_s = solve_second_stage_subproblem_function$
$[K_s, P_s, S_s, R_s, O_s, EVD_s, C_s]$ (Algorithm 4) ▷ Add the users that arrive in the current time slot who need slow chargers to the available parking lots' chargers
    – **The following lines are added to obtain the second heuristic model** –
    Calculate EVD_{fs} ▷ considering C_f and $g \in K_s$
    $K_s, S_f, R_f, O_f = solve_second_stage_subproblem_function$
$[K_s, P_f, S_f, R_f, O_f, EVD_{fs}, C_f]$ (Algorithm 4) ▷ Add the users that arrive in the current time slot that need slow chargers, but that find them all occupied, to the fast chargers
  end for
end for

Algorithm 3. Heuristic Model - $remove_users_function[TR_i, O_i, S_i, R_i]$

if ($TR_i[t] \neq \emptyset$) **then**
  for ($item \in TR_i[t]$) **do**
    Assign to p the parking lot where the charging stations become free
    $O_i[p] = O_i[p] - \{item\}$
    if ($p \in S_i \wedge O_i[p] = \emptyset$) **then**
      $S_i = S_i - \{p\}$
    end if
    if ($p \in R_i \wedge O_i[p] = \emptyset$) **then**
      $R_i = R_i - \{p\}$
    end if
    if ($p \notin R_i \wedge O_i[p] \neq \emptyset$) **then**
      $R_i = R_i \cup \{p\}$
    end if
  end for
end if
return O_i, S_i, R_i

$$CSD(p) = \sum_{g \in \mathbb{G}} C_i(p, b) * n \tag{4}$$

where b is the building related to the group g and n is its cardinality.

Electric Vehicle Degree. The Electric Vehicle Degree $EVD(g)$ is the number of parking lots in which the group g can park.

$$EVD(g) = \sum_{p \in \mathbb{P}} C_i(p, b) \tag{5}$$

where b is the building related to the group g.

Algorithm 4. Heuristic Model - *solve_second_stage_subproblem_function* $[K_i, P_i, S_i, R_i, O_i, EVD_i, C_i]$

while $(K_i \neq \emptyset \wedge P_i \neq \emptyset)$ **do**
  Calculate CSD_i
  for $(value \in CSD_i)$ **do**
    if $(value = 0)$ **then**
      Assign to p the parking lot value 0
      $P_i = P_i - \{p\}$
    end if
  end for
  $k = argmax(CSD_i(p))$ ▷ Take the parking lot in which more users can park
  Assign to $slots$ the number of chargers available in the parking lot k
  $P_i = P_i - \{k\}$
  if $(k \notin S_i)$ **then**
    $S_i = S_i \cup \{k\}$
  end if
  Initialize $parked_EV = 0$
  if $(slots - O_i[k] = 0)$ **then**
    pass ▷ There are not available slots
  else if $(slots - O_i[k] < CSD_i[k])$ **then**
    while $(parked_EV < slots - O_i[k])$ **do**
      $j = argmin(EVD_i(g))$ ▷ Take the group that can charge its EVs in the smallest number of parking lots
      Assign to n the cardinality of the group j
      if $(n > slots - O_i[k])$ **then**
        $n = n - (slots - O_i[k])$ ▷ Update the cardinality of the group j
        $parked_EV, O_i, TR_i[t], AR, nI_1, nI_2, nI_3 = update_variable_function(slots - O_i[k])$ (Algorithm 5)
      else
        $parked_EV, O_i, TR_i[t], AR, nI_1, nI_2, nI_3 = update_variable_function(n)$ (Algorithm 5)
        $K_i = K_i - \{j\}$, $EVD_i = EVD_i - \{EVD_i[j]\}$
      end if
    end while
  else
    $O_i[k] = O_i[k] + CSD_i[k]$
    if $(k \notin R_i)$ **then**
      $R_i = R_i \cup \{k\}$
    end if
    Initialize $parked_group = \emptyset$
    for $(g \in K_i)$ **do**
      Assign to b the building related to the group g
      if $(C_i[k][b] = 1)$ **then**
        Assign to n the cardinality of the group g
        $parked_group = parked_group \cup g$
        $TR_i[t] = TR_i[t] \cup \{g\}$
        if (g class is 2) **then**
          $AR = AR + \{g\}$
        end if
        if (dwell time of g is $short$) **then**
          $nI_1 = nI_1 + n$
        else if (dwell time of g is $medium$) **then**
          $nI_2 = nI_2 + n$
        else
          $nI_3 = nI_3 + n$
        end if
      end if
    end for
    for $(g \in parked_group)$ **do**
      $K_i = K_i - \{g\}$, $EVD_i = EVD_i - \{EVD_i[g]\}$
    end for
  end if
end while
return K_i, S_i, R_i, O_i

Algorithm 5. Heuristic Model - *update_variable_function*(*num*)

$parked_EV = parked_EV + num$
$O_i[k] = O_i[k] + num$
$TR_i[t] = TR_i[t] \cup \{j\}$
if (j's class is 2) **then**
  $AR = AR + \{j\}$
end if
if (dwell time of j is *short*) **then**
  $nI_1 = nI_1 + num$
else if (dwell time of j is *medium*) **then**
  $nI_2 = nI_2 + num$
else
  $nI_3 = nI_3 + num$
end if
return $parked_EV, O_i, TR_i[t], AR, nI_1, nI_2, nI_3$

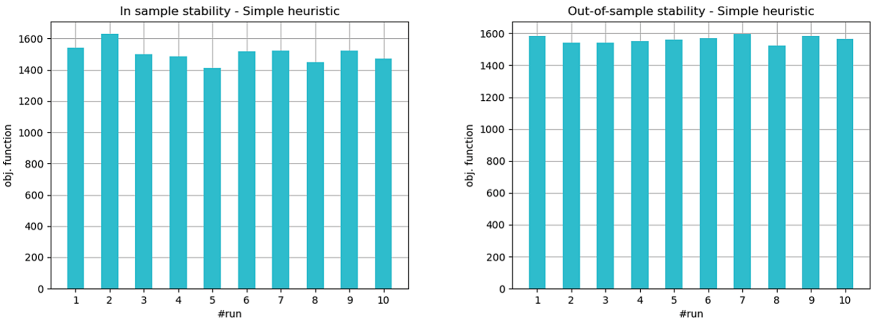
3.3 Example Result

In order to give a practical example of the obtained results, the heuristic model was launched considering the map in Fig. 1 and the default settings. The heuristic model places 5 fast chargers each in parking lots with ID 0, 1, 2 and 8, while it places 5 slow chargers only in parking lot with ID 7. Comparing the result with what was obtained in the results of the mathematical model with the same settings, the heuristic identifies the same parking lots that need to be equipped. However, it places 5 fast chargers in parking lot with ID 1 instead of 5 slow chargers. In fact, for how the heuristic algorithm is structured, it tends to favor the placement of fast chargers. However, after a number of simulations, it was discovered that this preference suit well cases with a high number of EV users with high diversity (i.e. number of groups), which represents a realistic situation.

3.4 Stability Testing

Both the in sample and out-of-sample stability of the heuristic model were tested. The in sample stability (3a) was tested over 10 different scenario trees, each composed by 50 different scenarios, while the out-of-sample stability (3b) was tested over 10 different scenario trees composed by 200 scenarios. Given the number of repetitions, the settings of the instance were lowered. The results of the in sample ad out-of-sample stability testing are shown in Fig. 3.

It can be stated that the stochastic heuristic model is stable and consistent. The in sample stability presents some fluctuations, but with still in an acceptable amount.



(a) In Sample Stability (b) Out of Sample Stability

Fig. 3. Sample Stability Heuristic

4 Conclusions and Future Work

Starting from open research challenges of the research community, in this paper we have proposed three innovative heuristic algorithms that solve the optimal charging-station location problem for electric vehicles, whose main benefit consists in achieving good approximate solutions while keeping acceptable computational time. Future work is mainly oriented towards making our algorithms compliant with emerging *big data trends* (e.g., [2,3,5,7,10,16–19,25,27]).

References

1. Adacher, L.: Heuristics for electric vehicle charging station allocation problem. In: 2018 5th International Conference on Mathematics and Computers in Sciences and Industry (MCSI), pp. 72–76. IEEE (2018)
2. Barkwell, K.E., et al.: Big data visualisation and visual analytics for music data mining. In: 22nd International Conference Information Visualisation, IV 2018, Fisciano, Italy, 10–13 July 2018, pp. 235–240. IEEE Computer Society (2018)
3. Bellatreche, L., Cuzzocrea, A., Benkrid, S.: F& A: A methodology for effectively and efficiently designing parallel relational data warehouses on heterogenous database clusters. In: DAWAK 2010, Bilbao, Spain, August/September 2010. Proceedings, pp. 89–104 (2010)
4. Bi, R., Xiao, J., Pelzer, D., Ciechanowicz, D., Eckhoff, D., Knoll, A.C.: A simulation-based heuristic for city-scale electric vehicle charging station placement. In: 20th IEEE International Conference on Intelligent Transportation Systems, ITSC 2017, Yokohama, Japan, 16–19 October 2017, pp. 1–7. IEEE (2017)
5. Camara, R.C., Cuzzocrea, A., Grasso, G.M., Leung, C.K., Powell, S.B., Souza, J., Tang, B.: Fuzzy logic-based data analytics on predicting the effect of hurricanes on the stock market. In: 2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio de Janeiro, Brazil, 8–13 July 2018. pp. 1–8. IEEE (2018)
6. Cavadas, J., de A Correia, G.H., Gouveia, J.: A mip model for locating slow-charging stations for electric vehicles in urban areas accounting for driver tours. Trans. Res. Part E: Logist. Trans. R. **75**, 188–201 (2015)

7. Coronato, A., Cuzzocrea, A.: An innovative risk assessment methodology for medical information systems. IEEE Trans. Knowl. Data Eng. **34**(7), 3095–3110 (2022)
8. Cuzzocrea, A.: Analytics over big data: Exploring the convergence of datawarehousing, OLAP and data-intensive cloud infrastructures. In: COMPSAC 2013, Kyoto, Japan, 22–26 July 2013, pp. 481–483. IEEE Computer Society (2013)
9. Cuzzocrea, A., Canadé, L., Nicolicchia, R., Roldo, L.: Optimal location of charging stations for electric vehicles: A theoretically-sound heuristic approach. In: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC 2023, Tallinn, Estonia, 27–31 March 2023. ACM (2023)
10. Demchenko, Y., De Laat, C., Membrey, P.: Defining architecture components of the big data ecosystem. In: 2014 International Conference On Collaboration Technologies and Systems (CTS), pp. 104–112 (2014)
11. Gatica, G., Ahumada, G., Escobar, J., Linfati, R.: Efficient heuristic algorithms for location of charging stations in electric vehicle routing problems. Stud. Inf. Control **27**, 73–82 (2018)
12. He, F., Yin, Y., Zhou, J.: Deploying public charging stations for electric vehicles on urban road networks. Trans. Res. Part C: Emerging Technol. **60**, 227–240 (2015)
13. He, J., Yang, H., Tang, T., Huang, H.: An optimal charging station location model with the consideration of electric vehicle's driving range. Trans. Res. Part C: Emerg. Technol. **86**, 641–654 (2018)
14. Hosseini, M., MirHassani, S., Hooshmand, F.: Deviation-flow refueling location problem with capacitated facilities: Model and algorithm. Transp. Res. Part D: Transp. Environ. **54**, 269–281 (2017)
15. Kuby, M., Lim, S.: The flow-refueling location problem for alternative-fuel vehicles. Socioecon. Plann. Sci. **39**(2), 125–145 (2005)
16. Leung, C.K., Braun, P., Hoi, C.S.H., Souza, J., Cuzzocrea, A.: Urban analytics of big transportation data for supporting smart cities. In: Ordonez, C., Song, I.-Y., Anderst-Kotsis, G., Tjoa, A.M., Khalil, I. (eds.) DaWaK 2019. LNCS, vol. 11708, pp. 24–33. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27520-4_3
17. Leung, C.K., Chen, Y., Hoi, C.S.H., Shang, S., Cuzzocrea, A.: Machine learning and OLAP on big COVID-19 data. In: 2020 IEEE International Conference on Big Data (IEEE BigData 2020), Atlanta, GA, USA, 10–13 December 2020, pp. 5118–5127. IEEE (2020)
18. Leung, C.K., Chen, Y., Hoi, C.S.H., Shang, S., Wen, Y., Cuzzocrea, A.: Big data visualization and visual analytics of COVID-19 data. In: 24th International Conference on Information Visualisation, IV 2020, Melbourne, Australia, 7–11 September 2020, pp. 415–420. IEEE (2020)
19. Leung, C.K., Cuzzocrea, A., Mai, J.J., Deng, D., Jiang, F.: Personalized deepinf: Enhanced social influence prediction with deep learning and transfer learning. In: 2019 IEEE International Conference on Big Data (IEEE BigData), Los Angeles, CA, USA, 9–12 December 2019, pp. 2871–2880. IEEE (2019)
20. Li, K., Jiang, H., Yang, L.T., Cuzzocrea, A. (eds.): Big Data - Algorithms, Analytics, and Applications. Chapman and Hall/CRC (2015)
21. Liu, Z., Wen, F., Ledwich, G.: Optimal planning of electric-vehicle charging stations in distribution systems. IEEE Trans. Power Delivery **28**(1), 102–110 (2013)
22. Miralinaghi, M., Keskin, B., Lou, Y., Roshandeh, A.: Capacitated refueling station location problem with traffic deviations over multiple time periods. Netw. Spatial Econom. **17**, 129–151 (2017)
23. MirHassani, S.A., Ebrazi, R.: A flexible reformulation of the refueling station location problem. Transp. Sci. **47**(4), 617–628 (2013)

24. Shahraki, N., Cai, H., Turkay, M., Xu, M.: Optimal locations of electric public charging stations using real world vehicle travel patterns. Transp. Res. Part D: Transp. Environ. **41**, 165–176 (2015)
25. Souza, J., Leung, C.K., Cuzzocrea, A.: An innovative big data predictive analytics framework over hybrid big data sources with an application for disease analytics. In: AINA 2020, Caserta, Italy, 15–17 April, pp. 669–680 (2020)
26. Wang, Y., Shi, J., Wang, R., Liu, Z., Wang, L.: Siting and sizing of fast charging stations in highway network with budget constraint. Appl. Energy **228**, 1255–1271 (2018)
27. White, L., Burger, K., Yearworth, M.: Big data and behavior in operational research: towards a "smart or". In: Behavioral Operational Research, pp. 177–193 (2016)