Predicting Vehicles Parking Behaviour for EV Recharge Optimization

Vinicius Monteiro de Lira[1], Fabiano Pallonetto[2], Lorenzo Gabrielli[1] and Chiara Renso[1]

[1]Institute of Information Science and Technologies, Italian National Research Council, Pisa, Italy, {vinicius.monteirodelira, lorenzo.gabrielli,chiara.renso}@isti.cnr.it

[2]School of Business, Maynooth University, Kildare, Ireland, fabiano.pallonetto@mu.ie

Abstract

The global electric car sales in 2020 continued to exceed the expectations climbing to over 3 millions and reaching a market share of over 4%. However, uncertainty of generation caused by higher penetration of renewable energies and the advent of Electrical Vehicles (EV) with their additional electricity demand could cause strains to the power system, both at distribution and transmission levels. The present work fits this context in supporting charging optimization for EV in parking premises assuming a incumbent high penetration of EVs in the system. We propose a methodology to predict an estimation of the parking duration in shared parking premises with the objective of estimating the energy requirement of a specific parking lot, evaluate optimal EVs charging schedule and integrate the scheduling into a smart controller. We formalize the prediction problem as a supervised machine learning task to predict the duration of the parking event before the car leaves the slot. This predicted duration feeds the energy management system that will allocate the power over the duration reducing the overall peak electricity demand. We experiment different algorithms and features combination for 4 datasets from 2 different campus facilities in Italy and Brazil. Using both contextual and time of the day features, the overall results of the models shows an higher accuracy compared to a statistical analysis based on frequency, indicating a viable route for the development of accurate predictors for sharing parking premises energy management systems

Keywords

parking prediction, electrical vehicles, machine learning, EV recharge optimization

1. Introduction

The advent of Electrical Vehicles (EV) are in increasing spreading in our society. According to MCkinsley report[1] in our society EV sales rose 65 percent from 2017 to 2018 and Europe has seen the strongest growth in EVs.

The concerns as we move to EVs is that, firstly, there will not be enough charge points to meet consumer demand and, secondly, this additional load on the electricity grid will cause partial and total failure of specific electrical plant due to overloading.

The present work fits this context supporting optimization for EV charging and assuming a incumbent high penetration of EVs in the system. We propose a methodology to predict an estimation of the parking duration in shared parking premises. This is essential for estimating

SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy

[1]https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/mckinsey-electric-vehicle-index-europe-cushions-a-global-plunge-in-ev-sales#

the energy requirement of a specific parking lot, evaluate optimal EVs charging schedule and integrate the scheduling into a smart controller.

The specific behaviour of parking lots of campuses refereed to EV charge is peculiar since it substantially differs from the general parking lots available in the streets. In campus-like facilities (Universities, large industries, etc) we can observe regular patterns of parking behaviour that mainly include staff working hours besides a part of other visitor [5]. This can be an advantage when trying to predict general behavioral patterns of parking habits and thus reach an optimal recharge plan for EVs.

Given this context, the specific objective of this work is to predict the duration of each *parking event* in a campus-like parking lot, where the *parking event* is the actual parking action of a car in a slot.

We formalize the prediction problem as a supervised machine learning task that, given a parking event at a given time, tries to predict the duration of the parking event. The reason for this event-based formulation is to be able to feed the energy management system with the duration prediction each time a car is parked. This will allow the energy management system to decide when to start the actual charge based on the prediction. In paper [1], that is the full version of this extended abstract, we detail all experiments into 5 datasets from 2 different campus facilities in Italy and Brazil. We show that using both contextual and time of the day features, the overall results of the models shows an higher accuracy compared to a statistical analysis based on frequency, indicating a viable route for the development of accurate predictors for sharing parking premises energy management systems.

2. The parking duration prediction problem

The objective of our approach is to exploit historical data on parking usage and additional contextual data like weather conditions and parking lot occupancy levels, to predict the duration of a parking slot occupancy. Differently from many state of the art approaches that want to predict if a giving parking lot will be free in a next period of time [3, 6], here we focus on the prediction of the temporal duration of the occupancy of a car in a slot. We recall that our approach, to be suitably integrated with an Energy Management System, focuses on specific parking context that we call of shared premises (e.g. parking lots of universities, workplaces, supermarkets, etc), not focusing on fee-based street parking.

It is worth noticing that the parking behaviour in a campus-like facility reflects a different parking behaviour compared to fare-based streets parking lots. In campus-like parking, the parking duration is expected to be longer than on street parking, since these premises are used by people parking to go to work, or study or perform an activity with a minimal temporal duration, while a fare-based parking in a street is generally affected by the parking fees that tend to encourage the reduction of the parking duration.

2.1. Problem formulation

Given a parking area a *car parking event* represents an event where a driver parks at a given timestamp in one of the available slots. The vehicle stays parked for a certain temporal duration until it leaves the slot. It is assumed that the vehicle can be charged while parked. The charging time can start as soon as the car arrives, or can start later on, or again, can start, interrupt and start again.

Having the prediction of the parking duration when a vehicle arrives at the parking is essential to properly schedule the starting of the charge avoiding energy usage peaks.

We define a car parking event e as a tuple $e =< s_{id}, t_{start}, d_e >$, where s_{id} represents the parking slot identifier where the car is parked, t_{start} represents the timestamp indicating when a car has started the parking and d_e is the temporal duration of the car park until it leaves the slot. We want to predict the parking duration d_e of a car parking event e, given a $slot_{id}$ and the parking event starting time t_{start}. This prediction is modelled as a classification problem where the objective is to assign, for each car parking event e, a class representing the predicted duration interval. More formally, we have the following definition of the problem. Given a parking event e where it is known the slot identifier s_id and the start time t_{start} but not duration d_e, we want to define a function $f(s\_id, t_{start}) = c$ where the class c represents a temporal interval such that $d_e \in c$.

We can observe that our target variable c represents ordinal categories. An ordinal variable is a categorical variable, where there is a clear ordering of the categories. For example, our variable could assume ordinal categories like: short, medium or long duration.

2.2. Predicting parking duration with Machine Learning

We propose to use supervised machine learning approaches to predict the parking duration based on an historical dataset of car parking events and contextual features.

The learning task is based on a three types of features: *single event-related*, *spatial* and *contextual* features. The event-related features represent the features that we can extract directly from the sets of parking events like the time of the parking event or the weather conditions.

The spatial features are based on the location of the parking slots inside the car parking area, while the contextual features representing the occupancy of the different zones of the parking area.

From the timestamp t_{start}, we derive three features: the day of week dw, hour of the day h, and the minutes m rounded to 5 minutes. The motivation of these temporal features is to enable the predictive model to learn the correlation between the time when the car parks and the relative parking temporal duration. We also include in this category of features the weather condition wr at the moment of the car parking event starts, t_{start}, using this as extra information to feed the predictive models.

To improve the prediction performance we also add some spatial features and specifically the parking spatial cluster. We therefore focus on the spatial distribution of the parking slots: we split the whole parking lot into smaller areas using different clustering approaches. Then, we include these spatial features in our predictive models to learn if a parking area can correlate with the slot occupancy duration.

Another aspect that we investigate for the parking duration prediction is the context. In our case the context is represented by the status of occupancy of the slots in the spatial clusters and relationship of this occupancy with the duration of a given parking event.

Specifically, we want to discover if the occupancy status of an area (e.g 100%, means totally full, while 0% totally empty) where a driver parks has relationship with the parking duration.

The overall idea is to investigate how to train the predictive models using different information that might have a predictive power on the parking duration. In the next section we detail the experimental setting and results on exploiting these features in a machine learning task for predicting the parking duration of a given event.

3. Experimental evaluation

The experimental evaluation aims at studying how accurately a supervised machine learning approach can predict the duration of a parking event in a campus-like parking lot. We compare the performance results of our machine learning based approach against several baselines and we investigate different machine learning approaches to tackle this problem as a supervised task: Classification, Ordinal Regression, and Regression.

Datasets. We selected two public datasets of parking occupancy in campus-like parking lots: PKlot [4] and CNRPark [2]. Both datasets contain the occupancy information detected by video cameras for each slots of parking areas of two academic institutions: the research area of the National Research Council of Pisa[2], in Italy and the parking area of the two Brazilian universities. In both cases the whole parking lot is split in different parking areas with a variable number of parking slots. In both datasets, a car parking event occurs when a car parks in a parking slot of the area. In this case, the event starts at the timestamp of the frame that detects a car in the slot. The car parking event ends at the timestamp of the frame showing: (1) an empty parking slot, or (2) a different car parked in the same slot. The duration of the parking event is then computed as the difference of the timestamps of the two image frames, the start and the end.

Details on the CNR park and PKlot datsets are reported in the full version of this paper [1]

The PKlot dataset contains the occupancy information for each slot of the parking areas of two academic institutions: (1) the Federal University of Parana (UFPR) and (2) the Pontifical Catholic University of Parana (PUCPR), both located in Curitiba, Brazil. The dataset includes a total of three different parking lots represented by PUCPR, UFPR04, and UFPR05. The occupancy information is detected by a number of cameras taking images of the parking slots and detecting the change of the car or the slot becoming empty. This dataset contains 12.417 images captured in three different parking areas with different weather conditions for a total of 168 slots in the period between 11 September 2012 and 16 April 2013. Specifically, dataset PUCPR has 100 parking slots, UFPR04 has 28 and UFPR05 has 45 slots. PKLot is larger than CNRPark and contains images spanning across months.

We have considered two different scenarios of classes for the predictive variable (i.e. the car parking event duration): (a) *Lower sensitivity*, with longer time intervals having a total of 3 classes with discrete values in minutes: $Short \leq 60, 60 < Mid \leq 240, Long > 240$; and (b) *Higher sensitivity* having shorter time intervals with a total of 6 classes: $Short1, \leq 30, 30 > Short2 \leq 60, 60 < Mid1 \leq 120, 120 < Mid2 \leq 240, 240 < Long1 \leq 480$, and

[2]http://www.area.pi.cnr.it

$Long2 > 480$. With these two scenarios, we want to illustrate applications requirements with different sensitivity for the predictive variable.

We used three training approaches: Classification, Regression and Ordinal Regression and used as measure the *micro-fscore and* mean square error (MAE)

Algorithms. For the Classification and Regression tasks we used the following algorithms: Random Forest (RF), XGBoosting (XGB), AdaBoosting (AB), Logistic Regression (LR) and Support Vector Machine (SVM). For the Ordinal Regression task we selected: Random Forest (RF), XGBoosting (XGB), AdaBoosting (AB) and Logistic Regression (LR) To compute the spatial features, we have used the K-means and the DBScan clustering algorithms. For all algorithms, we used the implementation available in the scikit-learn library[3].

Features. The following features are extracted and used to feed the ML algorithms. The event-related features include hour of the day h, time stamp minutes m, day of week dw, slot id s, and weather condition wr; the spatial features include the spatial cluster id spt; the occupancy features include the spatial cluster occupancy ocy. We use different feature combinations to train the models, specifically:

1. Single event-related feature. We train the model using only one event-related feature.
2. All event-related features together. We train the model using all single event-related features at once. We refer to all when we use all the event-related features to train the ML model.

For both cases, we perform two further combinations: using and not using the spatial and occupancy features to feed the models.

Baselines. To be able to evaluate the performance of our approach we have used the following baselines: (a) Random: randomly choose a class; (b) Longest Class: always select the longest interval; (c) Shortest Interval: always choose the shortest interval; (d) Majority Class: always choose the class with highest frequency in the training data. Furthermore, for each training approach, we also use specific baselines. For the classification and ordinal approaches, we use (e) Gaussian Naive Bayes (GNB) and (f) Multinomial Naive Bayes (MNB) as additional baseline algorithms. While for regression, we compare with the (f) Linear Regression (LN). Naive Bayes and Linear Regression are both simple ML models with high bias. They are used here as baselines given their easy interpretation.

ML model training process. For each dataset, we split the car parking events into train and test with 0.8 and 0.2 ratio respectively without shuffle the data. To avoid data leakage, we ordered the car parking events using their timestamps before split. When training the models on the training data, we use a stratified cross-validation with 5 folds. After the training, for each algorithm, the best configuration of hyper-parameters is used to retrain the model using the whole training data and then assess its performance now using the test set.

Evaluation metrics. To evaluate the experiment results we have used the following measures: micro f1-score ($F1_{micro}$), macro f1-score ($F1_{macro}$) and mean absolute error (MAE). These measures give some clues about the precision and recall of the models on predicting the true positives.

[3]https://scikit-learn.org/

3.1. Discussion of experimental results

We analyse the performance of each ML approach ($Ordinal$, $Classification$, and $Regression$) when predicting parking events duration. The experiment details are reported in the full version of the present paper [1].

We study the feature importance scores of the best models for each dataset and sensitivity scenario. Feature importance scores can provide insights of how the ML models works and what can be further improved. The relative scores can highlight which features are most relevant for the model to predict the target values, and the converse, which features are the least relevant.

Figure 1 shows the accuracy and the mean average error of the three machine learning approaches tested against the four datasets covering the lower sensitivity categories (3 categories).

Both classification and regression algorithms produce a similar performance for long term parking; however, classification is more accurate on the short term forecasting while regression has an overall lower mean average error in the medium range. The parking prediction module based on classification could provide a better user experience to drivers because accurate identification of short-term parking will force the controller to guarantee a higher energy share to short-term park events. However, it will reduce the peak shaving capabilities of the parking area. On the other hand, a regression model could facilitate demand response measures because of forecasted parking events shifted towards long parking time.

Figure 2 shows the predictive algorithms' results on the higher sensitivity categories (6 categories). The results confirm the challenges in forecasting short term parking events for the CNR dataset while it confirms the best prediction performance for the UPFR05 data. In this context, both classification and ordinal approaches are the most accurate for short term parking events ($<$30 mins), while regression is the most accurate for long term parking events ($>$ 8hr). In the latter case, the ordinal approach is the more accurate for three datasets over four, while predicting UPFR04 shows a high percentage of errors. All the approaches result in low accuracy for categories with a smaller number of events, such as categories 2 (30-60 min) and 3 (1hr-2hr). Such a lower score depends on the limited amount of data available for training. In comparison, high-frequency events are most likely to be correctly predicted, as illustrated in the CNR dataset, category 5 (4-8 hrs) and in UPFR04, UPFR05 and PCUPR for category 1 ($<$30 min). Overall, the lower sensitivity models are more accurate that to be integrated in an optimisation module for energy management systems. The high sensitivity predictions result suffering from low accuracy and high mean average error that could lead to system malfunctions and uncertainty. Therefore, to integrate the parking prediction module in an energy management system, further improvement of the forecasts are necessary, especially for high sensitivity experiments. Additional data could be used to improve the prediction's accuracy, such as higher picture resolution that could better identify users by reading the licence plates or identify unique marks (stickers, internal objects or scratches). Additionally, other data sources can be employed to forecast the number of car at an aggregated level and compare it with similar works. The current work aims to provide an overall design of a smart charging energy management system to optimally integrate the distributed energy systems and EVs into the power grid by developing a parking prediction module to estimate the vehicles' parking time using machine learning algorithms. The proposed system can capture EVs users' aggregated uncertain behaviour to obtain an optimised solution for both the capital expenditures (CAPEX)

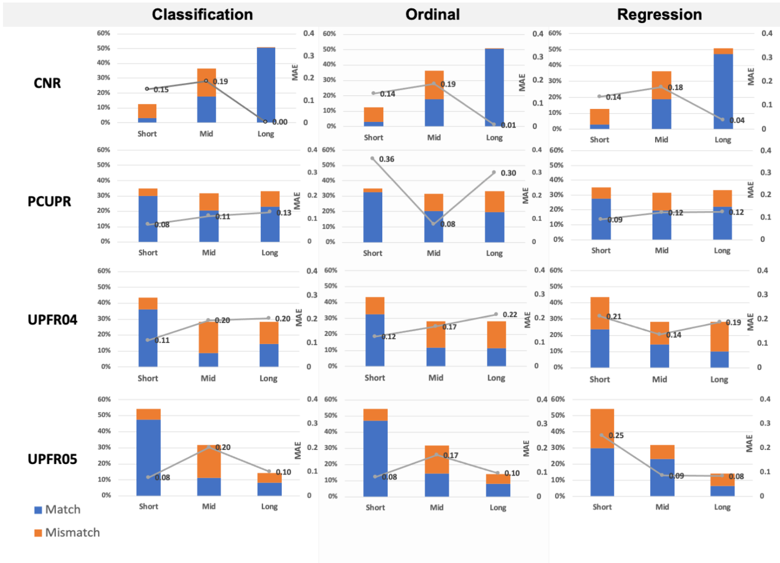

Figure 1: For lower sensitivity (3 intervals) and for the four datasets analysed, comparison of the accuracy and MAE for the prediction

and operational expenditures (OPEX) at the network's planning and operation phases. CAPEX can be minimised by optimising and distributing distributed energy resources and charging stations for electric vehicles under economic and social constraints. At the same time, optimal power flow solutions considering technical constraints can lead to OPEX minimisation.

4. Conclusions and Future Works

The current work aims to develop a parking prediction module to estimate the vehicles' parking time in shared premises using machine learning algorithms. Future works include the use of anonymised user profiles to reach more accurate predictions based on the single user habits, as well as having more dense and richer datasets to improve the accuracy of the models. Another direction is the proper integration of the prediction module into an Energy Management System.

Acknowledgment

The work is supported by the ERA-NET Smart Energy System, Sustainable Energy Authority Ireland and Italian Ministry of Research with project N. ENSGPLUSREGSYS18_00013. This publication has emanated from research conducted with the financial support of of the EVCHIP project under grant agreement 19/RDD/579, EVCHIP.

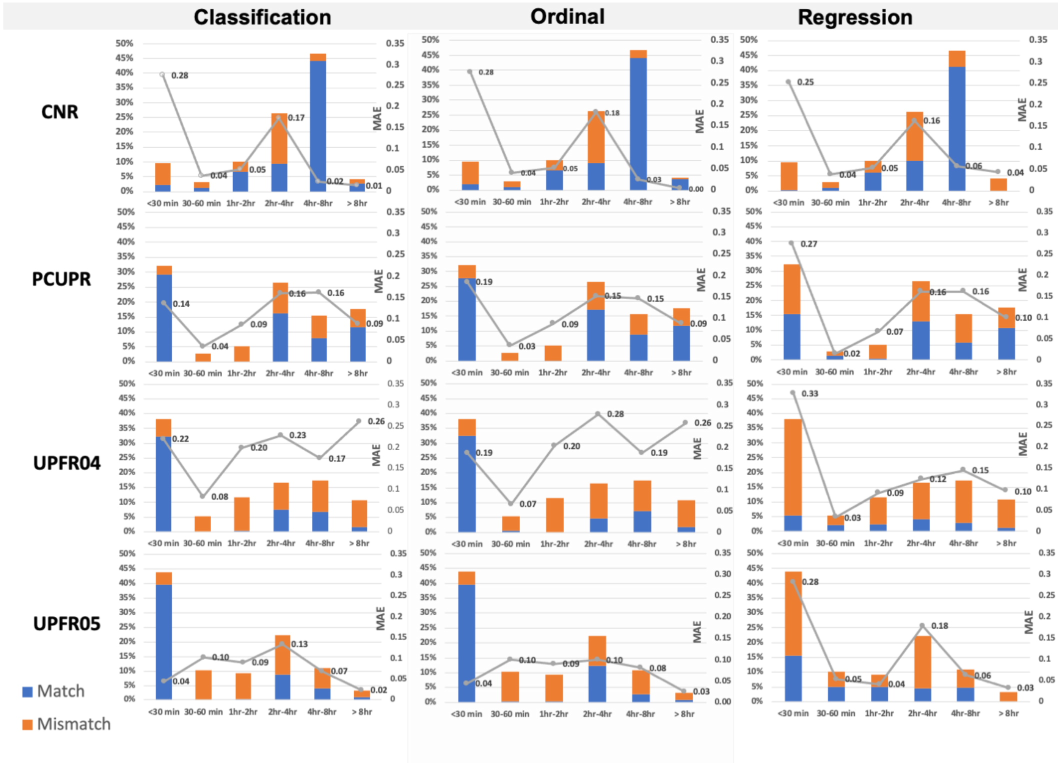

Figure 2: For higher sensitivity (6 intervals) and for the four datasets analysed, comparison of the accuracy and MAE for the prediction

References

[1] Vinicius Monteiro de Lira,Fabiano Pallonetto, Lorenzo Gabrielli, Chiara Renso Predicting vehicles parking behaviour in shared premises for aggregated EV electricity demand response programs, CoRR, abs/2109.09666 ,https://arxiv.org/abs/2109.09666,

[2] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo. Car parking occupancy detection using smart camera networks and deep learning. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 1212–1217, 2016.

[3] Felix Caicedo, Carola Blazquez, and Pablo Miranda. Prediction of parking space availability in real time. *Expert Systems with Applications*, 39(8):7281 – 7290, 2012.

[4] Paulo RL De Almeida, Luiz S Oliveira, Alceu S Britto Jr, Eunelson J Silva Jr, and Alessandro L Koerich. Pklot–a robust dataset for parking lot classification. *Expert Systems with Applications*, 42(11):4937–4949, 2015.

[5] Duda-Wiertel, U. and Szarata, A. The analysis of transport-related behaviours of drivers in highly occupied paid parking zones. Advances in Transportation Studies, Volume 47, 2019

[6] Eric Hsueh-Chan Lu and Chen-Hao Liao. Prediction-based parking allocation framework in urban environments. *International Journal of Geographical Information Science*, 34(9):1873–1901, 2020.