Data-driven prediction of fine-grained EV charging behaviors in public charging stations

Fuli Qiao
Shanghai Jiao Tong University
qiaoflii@sjtu.edu.cn

Shan Lin
Stony Brook University
Shan.X.Lin@stonybrook.edu

ABSTRACT

With the rapid growth of electrical vehicle public charging stations, accurate predictions of local charging demand enable many prospective applications. In this paper, we explore a data-driven approach to predict future charging demand, and build predictive models to characterize behaviors of both registered long-term users and unregistered short-term users. With a real-world dataset of 28053 records over 798 days at multiple locations, evaluation results demonstrate that our model with XGBoost outperforms existing solutions, reducing the prediction error up to 40.8% at the finest time granularity (15-minute interval).

CCS CONCEPTS

• **Computing methodologies → Machine learning**.

KEYWORDS

EV charging prediction, user behaviors, machine learning

ACM Reference Format:
Fuli Qiao and Shan Lin. 2021. Data-driven prediction of fine-grained EV charging behaviors in public charging stations. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Prediction of charging demand is very important for both electric vehicle (EV) drivers and charging station owners to manage and optimize their operations[4]. However, a main challenge is how to understand the aggregated charging behaviors of both registered users and unregistered users in the system[2]. The registered users tend to use the system for longer terms; whereas the unregistered users typically are short-term users who use charging stations occasionally. To address this issue, we design a data-driven approach to predict usage patterns of EV charging piles at public stations. Our prediction design includes one predictive model for registered users and another one for unregistered users. Working on a real-world charging record dataset collected in Caltech[3], we apply supervised learning based algorithms, specifically XGBoost, Support Vector Regression (SVR), and Gradient Boost Decision Tree (GBDT), to predict sequences of future availability.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 EV CHARGING DEMAND PREDICTION MODELS

Goal: Obtain a function f that inputs a sequence of the occupied number of charging piles in a previous continuous time period $OP_{TG,M} = \{o_1, o_2, ..., o_m\}$ and outputs a sequence of the available number of charging piles in a future continuous time period $AP_{TG,N} = \{a_1, a_2, ..., a_n\}$. Here, TG presents time granularity. The element in $OP_{TG,M}$ and $AP_{TG,N}$ respectively mean the occupied and available number of charging piles at the time divided by TG. Besides, M and N present the number of TG.

$$f(OP_{TG,M}) = AP_{TG,N} \tag{1}$$

Considering different charging patterns of registered users R and unregistered users U, the improved function is as follows.

$$f(OP_{TG,M}) = f_{re}(OP^R_{TG,M}) + f_{unre}(OP^U_{TG,M}) \tag{2}$$

Method: As Figure 1 shown, for registered/unregistered users, use XGBoost to train two predictive models separately and then combine middle prediction results to obtain final prediction results.

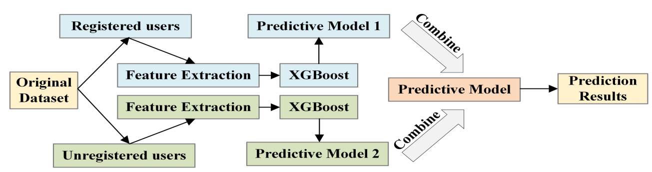


Figure 1: Proposed predictive model

Data processing: Calculate the availability (unavailability: 0, availability: 1) of each charging pile to obtain total occupied number in a certain period. Figure 2 shows data samples of unregistered/registered users from 20:00 to 24:00 of a day. In order to predict fine-grained, we divide the time into 15, 30, 45, 60, 80 minutes.

Prediction process: To predict $AP_{TG,S-M} = \{a_{m+1}, a_{m+2}, ..., a_s\}$, use $OP_{TG,M} = \{o_1, o_2..., o_m\}$. First, use $\{o_1, o_2, ..., o_m\}$ to fit a function and get $\{a_{m+1}\}$. Then, use $\{o_2, o_3, ..., a_{m+1}\}$ and get $\{a_{m+2}\}$. Conduct this multi-step iterative process until obtaining $\{a_s\}$.

Prediction results: As Figure 1 shown, for registered users, use predictive model 1 to get prediction results:

$$AP^R_{TG,N} = \{R_1, R_2, ..., R_N\} \tag{3}$$

Use predictive model 2 to obtain prediction results for unregistered users:

$$AP^U_{TG,N} = \{U_1, U_2, ..., U_N\} \tag{4}$$

Finally, combine these two prediction results to get the available number of EV charging piles in a future time period:

$$AP_{TG} = AP^R_{TG,N} + AP^U_{TG,N} = \{R_1 + U_1, ..., R_N + U_N\} \tag{5}$$

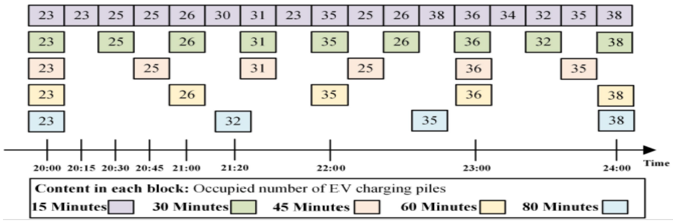


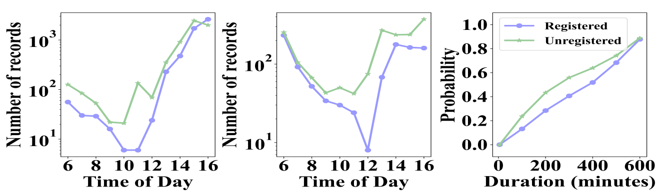
Figure 2: Data samples of unregistered/registered users with different time granularity

3 EXPERIMENTS

The dataset contains a total of 798 days from Apr. 2018 to Jul. 2020[1] and 52 charging piles in one charging station. The selected features are shown in Table 1. The distribution of charging connection/disconnection time is shown in Figure 3 (a) and (b). The CDF (cumulative distribution function) of charging duration of sections is shown in Figure 3 (c). It shows when the duration is shorter, the CDF of unregistered users is larger than that of registered users, which means the majority of unregistered users have shorter durations. In each time granularity, use 75 percent of the data as training samples and 25 percent as testing samples.

Name&Description	Example
connectionTime(time when plugged in)	23 May 2018 20:04:39
disconnectTime(time when unplugged)	23 May 2018 23:08:50
spaceID(Identifier of a charging pile)	CA-495
userID(Identifier of a user)	68

Table 1: Selected features from the dataset



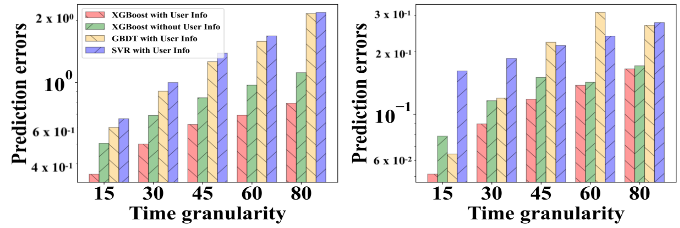
(a) Connection time **(b) Disconnect time** **(c) Charging Duration**

Figure 3: Distribution of dis/connection time and CDF of charging duration of sections between different users

With vs. without user information We compare the proposed prediction model with XGBoost and the primitive model with XGBoost, which is no distinction between users. As Figure 4 shown, the performance of both RMSE and MAPE in our proposed prediction model with XGBoost is better than the primitive model with XGBoost in each time granularity. Especially, the difference is significant when the time granularity is 15-minute.

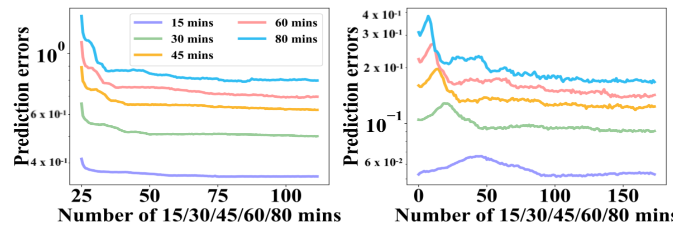
Comparison of different machine learning methods We compare our proposed prediction model with SVR, GBDT, and XGBoost. As Figure 4 shown, for RMSE and MAPE, XGBoost performs the best in each time granularity. For MAPE, GBDT performs better than SVR when time granularity is 15, 30, and 80 minutes, while SVR performs better than GBDT in other cases.

Prediction of proposed prediction model with XGBoost for different time granularity We try to find a better scheme, including time granularity and the length of training data. As Figure 5 shown, the smaller the time granularity, the better RMSE and MAPE. It shows 15-minute is better and its length of training data is between 25 and 50. In each time granularity, RMSE appears a downward trend from the beginning, while MAPE firstly increases then decreases as the number of time granularity increases.



(a) RMSE **(b) MAPE**

Figure 4: Prediction errors of primitive models with XGBoost, GBDT, and SVR with different time granularities



(a) RMSE **(b) MAPE**

Figure 5: Prediction accuracy v.s. size of training dataset

4 CONCLUSION AND FUTURE WORK

We proposed a data-driven approach for predicting EV charging availability at workplace by considering user behaviors. Experiments on ACN-Data showed that the accuracy of our model trained with XGBoost outperforms other machine learning solutions. A more comprehensive study is needed to understand the spatiotemporal charging patterns and their context embedded in the data.

5 ACKNOWLEDGEMENTS

This work was supported in part by NSF CNS-1952096 and NSF CNS-1553273 (CAREER).

REFERENCES

[1] 2020. *Adaptive Charging Network Dataset.* Retrieved Jul. 20, 2020 from https://ev.caltech.edu/dataset
[2] Ahmad Almaghrebi, Fares Aljuheshi, Mostafa Rafaie, Kevin James, and Mahmoud Alahmad. 2020. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods. *Energies* 13, 16 (2020), 4231.
[3] Zachary J. Lee, Tongxin Li, and Steven H. Low. 2019. ACN-Data: Analysis and Applications of an Open EV Charging Dataset. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems* (Phoenix, AZ, USA). 139–149.
[4] Yukun Yuan, Desheng Zhang, Fei Miao, Jimin Chen, Tian He, and Shan Lin. 2019. p^2 Charging: Proactive Partial Charging for Electric Taxi Systems. In *IEEE ICDCS*. 688–699.