Electric Vehicle User Behavior Prediction Using Gaussian Mixture Models and Soft Information

Adam, Rebecca; Qian, Kun; Brehm, Robert

Published in:
202110th IEEE PES Innovative Smart Grid Technologies Asia (ISGT)

DOI:
10.1109/ISGTAsia49270.2021.9715580

Publication date:
2021

Document version:
Accepted manuscript

Citation for pulished version (APA):
Adam, R., Qian, K., & Brehm, R. (2021). Electric Vehicle User Behavior Prediction Using Gaussian Mixture Models and Soft Information. In *202110th IEEE PES Innovative Smart Grid Technologies Asia (ISGT)* IEEE. https://doi.org/10.1109/ISGTAsia49270.2021.9715580

Electric Vehicle User Behavior Prediction Using Gaussian Mixture Models and Soft Information

Rebecca Adam*†, Kun Qian† and Robert Brehm†
*Corresponding author
†Department of Mechanical and Electrical Engineering
Centre for Industrial Electronics, University of Southern Denmark (SDU)
Email: rebeccaadam@sdu.dk*, kqian@sdu.dk, brehm@sdu.dk

***Abstract*—Intelligent scheduling algorithms pave the way to vehicular mobility electrification. Besides enabling charging grid providers to upscale the charging grid infrastructure without costly changes, they enable optimal green or price-scheduled charging. However, scheduling always requires a priori user behavior parameter knowledge, like the arrival time, departure time, and energy demand. Recent comparisons indicate that machine learning-based predictions are more accurate than direct user prediction. Therefore, we exploit a time-series measurement dataset to predict the departure time and energy demand for intelligent scheduling. The collected data belongs to a charging grid used by nurses working at an elderly home in Denmark. Our data analysis reveals a clustered distribution for the departure time versus arrival time and the energy demand versus arrival time. Targeting on arrival prediction, we propose a sub-clustering strategy employing Gaussian mixture models in combination with the expectation-maximization or the variational Bayesian method to learn the departure times and energy demand. We demonstrate that the investigated method inherently determines different soft information metrics enabling us to propose different prediction strategies adaptable to the users' desired reliability degree for charge scheduling. Moreover, we show that our proposed clustering approach outperforms the assessed regression-based machine learning approaches.**

***Index Terms*—Gaussian Mixture Models, Intelligent Charge Scheduling, Clustering, User behaviour prediction, machine learning, soft information, unsupervised learning,**

I. Introduction

Upscaled and reliable charging infrastructures are a mandatory prerequisite for the transformation to climate-friendly e-mobility. Charge scheduling algorithms, like peak shaving, target an optimal vehicle connection time exploitation and yield a flattened load profile. Therefore, they are suitable to cost-efficiently upscale the charging grid sizes. However, intelligent charge scheduling algorithms' objectives are diverse. They range from price-determined scheduling over green-energy optimized scheduling to charging in a way that allows an increased grid capacity [1]–[5]. Note that the user behavior parameters, namely the arrival times (AT), the departure times (DT), and the energy demand (ED), constitute a mandatory input for intelligent charge scheduling algorithms. Either the user can provide this information, or machine learning algorithms can learn user behavior. The analyses carried out by the authors of [6] implicate that predicting the user behavior yields higher accuracy than using the user-specified information. Hence, we assess DT and ED prediction for intelligent charging on vehicle arrival.

We thank the Brødrene Hartmanns Fond for the project "Intelligent Charge Management Algorithms for EV" for their financial support.

A. Open Challenges

1) Reliability or Soft Information Usage: The authors of [7] state that user-specific patterns and high stochasticity aggravate user behavior prediction. These challenges also implicate significant prediction uncertainties. Therefore, employing soft information promises to improve charge scheduling. However, to our best knowledge, contributions addressing user behavior prediction for intelligent charge scheduling like [6]–[8] consider hard estimates only for prediction.

2) Lack of Diverse User Behavior Patterns: Different user behavior patterns require different machine learning solutions. Consequently, a broad data set range similar to the sets [6], [9] is necessary to thoroughly investigate prediction and scheduling algorithms. Especially, to substantiate prediction applicability the assessment of different diverse datasets gains new knowledge.

3) Data Analysis and Hyper-parameter-tuning Necessity: The authors in [7] stress that there is no single optimal method for all data sets. Thus, user behavior prediction not requiring data analysis and careful hyper-parameter-tuning pose open research challenges.

B. Related Contributions

1) Regression-based Contributions: The linear dependencies between DT, ED, and AT suggest linear regression as the intuitive choice for the DT and ED versus AT prediction. Linear regression, support vector regression, kernel density methods, and ensemble machine learning were assessed in and compared in [10] yielding the ensemble approach as the optimal solution. Since, the data used in this contribution is clustered, we use clustering and regression-based methods as a benchmark.

2) Kernel Method-based Contributions: A hybrid kernel density-based user behavior prediction was proposed in [7], [10] employing a novelty detection method to choose the most appropriate kernel density approach.

3) Clustering-based Contributions: Similar to this work, [6], [11] investigated a cluster-based DT and ED approach. Compared to the ACN-data [6] our measured dataset indicates less linear dependencies and distinct clusters.

4) Deep Learning-based Contributions: The authors in [8] studied deep learning, long-term short-term (LSTM) memory networks, for user behavior prediction based on the data from [6]. LSTMs successfully capture time-series dependencies and so can be used to exploit the weekdays' specific pattern in the dataset. Our assessed dataset exhibits a daily seasonality that does not vary over the weekdays.

C. Novel contributions

Our novel contributions for DT and ED prediction are:

- assessing a novel vehicle user behavior dataset.
- clustering and regression-based prediction comparison.
- flexible soft-information-based user prediction strategies

D. Notational Conventions

We denote vectors and matrices by lower and upper case boldfaced letters, respectively. Let $\hat{\bullet}$, $\tilde{\bullet}$, $\bar{\bullet}$, $\bullet^+$ and $\bullet^-$ denote predictions, hypothesis, mean values, test and training sizes, respectively. We use $[\bullet]_{:,m}$ for the mth matrix column.

II. The measured data

We preprocessed a time series dataset from a danish nursing home measuring voltages, currents, and the charging grid station's connection statuses, extracting four features per identified charging session, the AT, the DT, the ED, and the charging station identifier, the ID. Targeting a reasonable data size for scheduling, the chosen timeframe is limited to two months' data, 15th April to 15th June 2021. Preliminary analysis shows Gaussian distributed AT/DT clusters (explainable by the nurses work shifts).

III. Problem formulation

We target the DT and ED on AT prediction. Let U denote the number of grid's charging stations, D the number of days, $K_{u,d}$ the number of charging sessions for all $u \in \{1,\dots,U\}$ stations and for all $d \in \{1,\dots,D\}$ days. Then the overall session number is $N = \sum_{u=1}^{U}\sum_{d=1}^{D} K_{u,d}$. Let a_n, d_n, e_n with $n \in \{1,\dots,N^-\}$ denote the training AT, DT, and ED and α_n, δ_n, ε_n the test AT, DT, and ED with

$$n = d \cdot U \cdot K_{u,d} + u \cdot K_{u,d} + k,\ n \in \{1,\dots,N^+\}, \quad (1)$$

for $u \in \{1,\dots,U\}$, $d \in \{1,\dots,D\}$ and $k \in \{1,\dots,K_{u,d}\}$. Stacking in training and test vectors for all $n \in \{1\dots,N\}$ yields $\mathbf{a}$, $\mathbf{d}$, $\mathbf{e} \in \mathbb{R}^{N^- \times 1}$ and $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon} \in \mathbb{R}^{N^+ \times 1}$. We use the same approach to predict the DT and the ED separately. Thus, we prepare a 2-dimensional template for the size $N_{D^-} \times 2$ training matrix $\mathbf{X} \in \{[\mathbf{a},\mathbf{d}],[\mathbf{a},\mathbf{e}]\}$ and size N_{D^+} test matrices $\boldsymbol{\mathcal{X}} \in \{[\boldsymbol{\alpha},\boldsymbol{\delta}],[\boldsymbol{\alpha},\boldsymbol{\varepsilon}]\}$. Here, let $D^- = 46$ (training days), $D^+ = 14$ (test days) and $U = 4$ stations. For DT and ED prediction, we propose the following processing procedure:

- Perform 1-dimensional AT clustering.
- Choose AT cluster label and subcluster in 2 dimensions.
- Predict the DTs and EDs using hard and soft estimates.

IV. Clustering via Gaussian Mixture models

A. Gaussian Mixture Models

Let $\mathbf{x}_n \in \mathbb{R}^{2\times 1}$ denote the data matrix $\mathbf{X}$'s nth row. The Gaussian mixture model (GMM)

$$p(\mathbf{x}_n) = \sum_{c=1}^{C} \pi_c \frac{\exp(-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_c)\boldsymbol{\Sigma}_c^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_c))^T}{\sqrt{(2\pi)^{\mathcal{D}}|\boldsymbol{\Sigma}_c|}} \quad (2)$$

is a weighted superposition of C Gaussian distributions $\mathcal{N}_{\mathcal{D}}(\mathbf{x}_n|\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c)$ with mean $\boldsymbol{\mu}_c \in \mathbb{R}^{2\times 1}$ and covariances $\boldsymbol{\Sigma}_c \in \mathbb{R}^{2\times 2}$ for all $c \in \{1\dots C\}$. The mixing coefficients $\pi_c = p(c)$, the a priori probabilities, have the properties $\sum_{c=1}^{C}\pi_c = 1$ and $0 \le \pi_c \le 1$. Let $\mathbf{y} = [y_1,\dots,y_N]$ hold the latent variables relating the data points to the clusters. Iterative algorithms determine the cluster parameters in (2) by updating the responsibilities $\gamma_{c,n}$, or the posterior probability that c is responsible for $\mathbf{x}_n$ [12]:

$$\gamma_{c,n} = P(y_n = c|\mathbf{x}_n,\pi_c,\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c) = \frac{\pi_c p(\mathbf{x}_n|\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c)}{\sum\limits_{c=1}^{C} p(\mathbf{x}_n|\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_{\mathbf{c}})}. \quad (3)$$

In this work, a 1-D GMM model variant of (2) is first used to determine the 1-D AT cluster parameters, and then (2) is used a second time given a test AT to determine 2-D subclusters. Let the cluster tupels $\mathcal{C} = (k,c)$ denote the subclustering indices $k \in \{1,\dots,C^{\text{AT}}\}$ and $c \in \{1,\dots,C\}$.

B. Expectation Maximization and Variational Bayes

The expectation-maximization (EM) algorithm [13] is tailored to iterate via two subsequent steps, an expectation and a maximization step, to determine $\hat{\pi}_c$, $\hat{\boldsymbol{\mu}}_c$, $\hat{\boldsymbol{\Sigma}}_c$ and $\hat{\mathbf{y}}$ for all $c \in \{1,\dots,C\}$ by updating and using the responsibilities in (3). The variational Bayesian (VB) approach uses the same iterative updating strategy for the EM steps. The difference to the classic EM approach comes from the VB specific assumption that all parameters are modeled as random distributions, with conjugate prior distributions (Dirichlet distribution for the mixing coefficients and a Gaussian Wishart distribution for all Gaussian components). Note that we can determine slightly different responsibilities for EM and VB and mixing coefficients. We used the methods explained in [12], [14], [15] implemented in [16] and refer to these works for a detailed algorithm description that would exceed the scope of this paper. To initialize we use the K-means solution as suggested in [12]. Variational Bayesian GMM clustering automatically determines the number of clusters $\hat{C} \le C$ as comprehensively explained in [12, pp. 483-485]. Alternative approaches employ information-theoretic criteria [17], [18], requiring large sample sizes or cross-validation. The sub-clusters obtained via GMM and VB including confidence regions and weightings are visualized in Fig. 1.

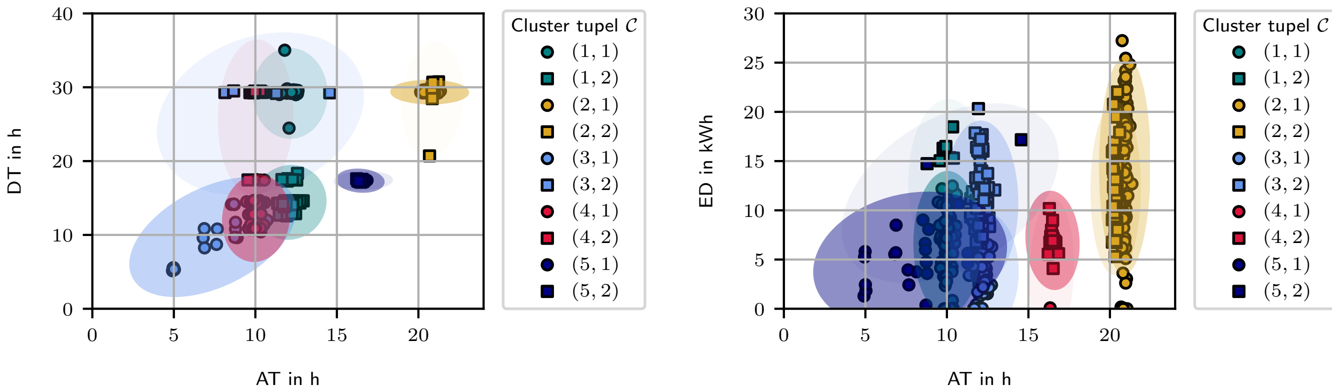


Fig. 1: The plots a) and b) show an example clustering solution and 2-dimensional data distributions. The linear dependency between ED and AT seems stronger than between DT and AT. From a) we see that given a single AT, two DT clusters are often present due to the nurses' work shifts. The confidence regions' opacities are scaled by the determined mixture weights, depicting the effective sub-cluster probabilities.

C. *Soft Information*

The EM and the VB algorithm inherently yield soft information as a measure for reliability, namely the responsibilities, the mixture coefficients, the covariances, and confidence regions. Exploiting this soft information can improve intelligent charge scheduling. For example, a schedule could benefit from predictions targeting a full state-of-charge at DT.

1) The Responsibilities and mixture weights: Note that the mixture weights for the cth component is determined as the average responsibility

$$\pi_c = \frac{1}{N}\sum_{n=1}^{N}\gamma_{c,n}. \tag{4}$$

Both parameters, therefore, are soft information and can improve the user behavior prediction. The mixture weight π_c indicates the a priori probability for cluster c. From (4), we see that it uses a sum of data points and hence, potentially indicates unlikely clusters. The responsibilities are determined by both the a priori weights and the associated data point. The responsibilities are preciser and more sensitive to prediction errors than the weights.

2) The Covariance Matrix: The covariance $\boldsymbol{\Sigma}_c$ is updated in EM and VB. We use the covariance matrix to determine the DT and ED predictions (see Section V). Its main diagonal is a direct soft reliability indicator for the hard DT and ED predictions. Furthermore, the confidence regions depend on the covariance matrix (see Section IV-C3).

3) The Confidence Regions: Confidence regions depend on a specified probability P and define the region for which the true parameter lies in with probability P. Thus, they enable outlier detection. These regions (here ellipses) can be determined by following the instructions in [19].

V. Departure Time and Demand Prediction

Now we assume that all GMM cluster parameters $\pi_c, \boldsymbol{\mu}_c^T, \boldsymbol{\Sigma}_c$ with $c \in \{1,\dots,C\}$ have been estimated via EM or VB, that the ATs $\boldsymbol{\alpha}$ are given (on arrival charge scheduling), and we have predicted the AT cluster label $\hat{k}$. For each sub-cluster component c we predict the DT and ED $\hat{\delta}_n^c := [\hat{\boldsymbol{\mathcal{X}}}]_{n,2}^c$ and $\hat{\varepsilon}_n^c := [\hat{\boldsymbol{\mathcal{X}}}]_{n,2}^c$ for given α_n and $\mu_{c,2} = \overline{\alpha}$ by setting the negative loglikelihood derivative to zero and solving for $\boldsymbol{\mathcal{X}}_{n,2}^c$ similar to [6]:

$$[\hat{\boldsymbol{\mathcal{X}}}]_{n,2}^c = \mu_{c,2} + (\alpha_n - \overline{\alpha}_c)\frac{[\boldsymbol{\Sigma}]_{1,2}}{[\boldsymbol{\Sigma}]_{1,1}}. \tag{5}$$

Different prediction policies apply, as we show subsequently.

A. *Tradeoff Prediction*

Superimposing the candidate predictions in (5) weighted by $\gamma_{c,n}$ yields a tradeoff (T-GMM) solution similar as in [6]:

$$\hat{\boldsymbol{\mathcal{X}}}_{n,2}^{\text{T-GMM}} = \sum_{c=1}^{C}\gamma_c[\hat{\boldsymbol{\mathcal{X}}}]_{n,2}^c. \tag{6}$$

B. *Likeliest Prediction*

Choosing the candidate prediction in (5) with the highest $\gamma_{c,n}$ or π_c yields (LR-GMM and LW-GMM)

$$\hat{\boldsymbol{\mathcal{X}}}_{n,2}^{\text{L-GMM}} = [\hat{\boldsymbol{\mathcal{X}}}]_{n,2}^{\hat{c}} \tag{7}$$

$$\hat{c} = \begin{cases} \arg\max_{\tilde{c}}\{\gamma_{\tilde{c},n}\} & \text{for LR-GMM} \\ \arg\max_{\tilde{c}}\{\pi_{\tilde{c}}\} & \text{for LW-GMM} \end{cases}. \tag{8}$$

C. *Secure Prediction*

The secure prediction (S-GMM) prefers the likeliest solution only, if it exceeds a specified probability r, with $0 \leq r \leq 1$. Otherwise it chooses the secure solution (the smallest DT and the largest ED)

$$\hat{\boldsymbol{\mathcal{X}}}_{n,2}^{\text{S-GMM}} = \begin{cases} \hat{\boldsymbol{\mathcal{X}}}_{n,2}^{\text{L-GMM}} & \text{if } \gamma_{c,n} > r \\ \min\{\tilde{\boldsymbol{\delta}}\} & \text{if } \gamma_{c,n} \leq r,\ \boldsymbol{\mathcal{X}}_{:,2} = \boldsymbol{\delta} \\ \max\{\tilde{\boldsymbol{\varepsilon}}\} & \text{if } \gamma_{c,n} \leq r,\ \boldsymbol{X}_{:,2} = \boldsymbol{\varepsilon} \end{cases}. \tag{9}$$

VI. Regression-based User behaviour prediction

For this work a straightforward benchmark approach determines the DT and ED on arrival by exploiting functional relationships found via regression $f(x) = d(a)$ or $g(x) = e(a)$. If $f(x)$ and $g(x)$ are modelled as a linear relationship, we can determine the linear regression parameters following [12] or use the Scikitlearn API [16]. The support vector regression (SVR) models non-linearities and promises a better fit than the linear regression for the data assessed in this work. Moreover, the SVR is tailored to reject outliers by constructing an ϵ-tube decision-region around the function that defines which data points to use and which not to use to model the function. Note that this can be interpreted as using soft information. Following [20], [21], and [12, p. 339-345] a support vector regression solves a constrained optimization problem using Lagrange multipliers to determine the support vectors and the model function by employing a kernel function, which is here a radial basis function. In this contribution, we employed the Scikitlearn SVR API to apply SVR [16].

VII. Results and Discussion

We employ the absolute error (AE)

$$\mathrm{AE}_n = |\mathcal{X}_{n,2} - \hat{\mathcal{X}}_{n,2}| \tag{10}$$

and the median absolute error (MdAE) that determines the median for all AE_n $n \in \{1, \ldots, N\}$ errors. Targeting fully charged vehicles on DT is crucial for intelligent scheduling. Consequently, in this work, we assess the error part that risks incomplete charging: DT overestimation (negative DT error) and ED underestimation (positive ED error). To distinguish this undesirable case from the desirable case we associate the labels "sorry" and "safe" with them. Hence, to assess the overestimated DT predictions and the underestimated ED predictions, we define the "sorry" session data subset $\mathcal{S}^{\text{sorry}} \subset \mathcal{S} = \{1, \ldots, N\}$ including those sessions n with a negative AE for the DT and and a positive AE for the ED prediction. Then the complementary "safe" session data is in the subset $\mathcal{S}^{\text{safe}} \subset \mathcal{S}$ as the set including those sessions n that yield positive errors for the DT and negative errors for the ED prediction with $\mathcal{S}^{\text{safe}} \cap \mathcal{S}^{\text{sorry}} = \emptyset$ and $\mathcal{S} = \mathcal{S}^{\text{safe}} \cup \mathcal{S}^{\text{sorry}}$. We define for the "sorry scenario" an undesirable error $\mathrm{AE}^{\text{sorry}}$ for all $n \in \mathcal{S}^{sorry}$ and for the "safe" scenario an error $\mathrm{AE}^{\text{safe}}$ for all $n \in \mathcal{S}^{\text{safe}}$. For the secure prediction, we set the probability $r = 1.0$. We use rolling cross-validation, increasing the training size of each step by 1 day more data for 14 days for testing and present the error distributions in Table I and Fig. 2. Note that, for DT prediction, as expected due to the clustered data distribution the reliability in terms of targeting the fully charged status on DT is larger for the GMM than for LR or SVR. For ED prediction SVR and GMM perform comparably well, since the linear dependency is stronger between ED and AT than between DT and AT. The MdAEs in Table I and Fig. 2 indicate that the soft-information exploitation for the GMM-based solutions improves reliability.

A. Discussion and Reflection

Table I shows that GMMs with the EM approach outperforms regression methods for the assessed data. This effect is more pronounced for the DT than for the ED prediction. The result is explainable by the nurses working shifts, rather revealing clusters than linear behavior. The ED prediction suffers from higher randomness (wider ED distributions). Compared to regression-based methods the clustering approach has a higher computational complexity. The reliability information exploitation in our approach addresses a question that is relevant for intelligent scheduling user acceptance: "How can reliability measures be exploited flexibly depending on the users' wishes to improve intelligent scheduling?" The three strategies, the tradeoff, the likeliest (chooses likeliest and hence) and the secure prediction (prevents undesired "sorry cases" most often - see I), encode different user objectives for intelligent scheduling. We propose to include all available soft-information metrics into the scheduling formulation.

VIII. Conclusion

We introduced a novel dataset exhibiting clustered charging user-behavior parameters: arrival, departure time, and energy demand. We propose a method to increase the user behavior prediction reliability by exploiting inherently available soft information from clustering. Moreover, the soft information potentially allows flexible scheduling adaptable to users' safety desires. We show that the soft clustering-based method yields a higher reliability and prediction accuracy than the regression-based methods.

References

[1] N. Chen, C. W. Tan, and T. Q. S. Quek, "Electric vehicle charging in smart grid: Optimality and valley-filling algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1073–1083, Dec. 2014.

[2] B. Khaki, C. Chu, and R. Gadh, "A hierarchical ADMM based framework for EV charging scheduling," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition*. IEEE, Apr. 2018, pp. 1–9.

[3] G. Putrus, P. Suwanapingkarl, D. Johnston, E. Bentley, and M. Narayana, "Impact of electric vehicles on power distribution networks," in *2009 IEEE Vehicle Power and Propulsion Conference*. IEEE, Sep. 2009, pp. 827–831.

[4] J. de Hoog, T. Alpcan, M. Brazil, D. A. Thomas, and I. Mareels, "Optimal charging of electric vehicles taking distribution network constraints into account," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 365–375, Jan. 2015.

[5] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 371–380, Feb. 2010.

[6] Z. J. Lee, T. Li, and S. H. Low, "ACN-data," in *Proceedings of the Tenth ACM International Conference on Future Energy Systems*. New York, NY, USA: ACM, Jun. 2019, pp. 139–149.

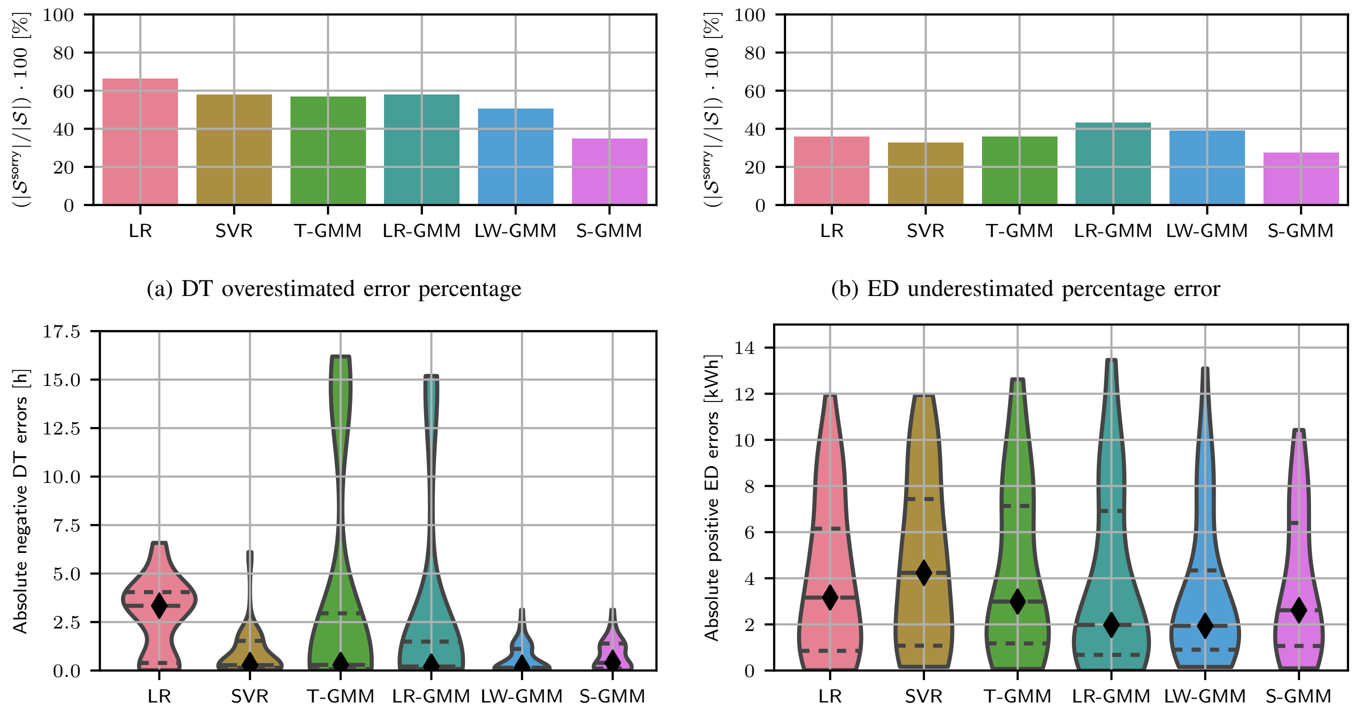
[7] Y.-W. Chung, B. Khaki, C. Chu, and R. Gadh, "Electric vehicle user behavior prediction using hybrid kernel density estimator," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, Jun. 2018, pp. 1–6.

[8] S. Khan, B. Brandherm, and A. Swamy, "Electric vehicle user behavior prediction using learning-based approaches," in *2020 IEEE Electric Power and Energy Conference (EPEC)*. IEEE, Nov. 2020, pp. 1–5.

[9] E. Xydas, C. Marmaras, L. M. Cipcigan, N. Jenkins, S. Carroll, and M. Barker, "A data-driven approach for characterising the charging demand of electric vehicles: A UK case study," *Applied Energy*, vol. 162, pp. 763–771, Jan. 2016.

TABLE I: Table summarizing error metrics (Clustering outperforms regression, bold indicates minima)

	DT prediction errors				ED prediction errors			
Method	MdAE(d) in h	MdAE$^{\text{sorry}}(d)$ in h	MdAE$^{\text{safe}}(d)$ in h	$\frac{\lvert\mathcal{S}^{\text{sorry}}\rvert}{\lvert\mathcal{S}\rvert} \cdot 100$ in %	MdAE(e) in kWh	MdAE$^{\text{sorry}}(e)$ in h	MdAE$^{\text{safe}}(e)$ in kWh	$\frac{\lvert\mathcal{S}^{\text{sorry}}\rvert}{\lvert\mathcal{S}\rvert} \cdot 100$ in %
LR	3.47	3.34	10.86	66.32	4.55	3.16	5.43	35.79
SVR	0.41	0.28	2.75	57.89	4.62	4.24	4.69	32.63
T-GMM	0.39	0.29	0.60	56.84	4.49	2.99	4.75	35.79
LR-GMM	**0.29**	0.21	**0.40**	57.89	**3.77**	1.98	**4.56**	43.16
LW-GMM	**0.29**	**0.14**	0.75	50.53	**3.77**	**1.94**	5.51	38.95
S-GMM	1.27	0.39	2.63	**34.74**	6.40	2.61	7.07	**27.37**



(a) DT overestimated error percentage

(b) ED underestimated percentage error

(c) overestimated DT error distributions

(d) underestimated ED error distributions

Fig. 2: The plots a)-d) illuminate the challenging cases, DT overestimation, and ED underestimation. This "sorry"-labeled error part is responsible for incomplete EV charging and hence exciting to assess in detail. Subplots a) and b) indicate that the "sorry"-scenario percentage is minimal for the GMM-type algorithms, particularly the S-GMM. Furthermore, the violin plots c) and d) show the absolute error distributions (rotated kernel density plots symmetrically drawn to both sides). They indicate GMM-based prediction algorithms outperform the regression-based DT algorithms and that SVR-based and GMM-based ED prediction perform comparably well. The ♦ denotes the median, and the dashed lines denote the 25% and the 75% percentiles.

[10] Y.-W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, "Ensemble machine learning-based algorithm for electric vehicle user behavior prediction," *Applied Energy*, vol. 254, p. 113732, Nov. 2019.

[11] Y. Xiong, B. Wang, C.-c. Chu, and R. Gadh, "Vehicle grid integration for demand response with mixture user model and decentralized optimization," *Applied Energy*, vol. 231, pp. 481–493, Dec. 2018.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Jan. 2006, no. 4.

[13] A. P. Dempster and N. M. Laird, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[14] D. M. Blei and M. I. Jordan, "Variational inference for dirichlet process mixtures," *Bayesian Analysis*, vol. 1, no. 1 A, pp. 121–144, 2006.

[15] H. Attias, "A variational bayesian framework for graphical models," in *Proceedings of the 12th International Conference on Neural Information Processing Systems*. Dever, CO: MIT Press, 1999, pp. 209–215.

[16] F. Pedegrosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blodel, P. Prettenhofer, R. Weiss, V. Duborg, J. Vanderplass, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn, machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[18] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.

[20] V. Vapnik, *The nature of statistical learning*. Springer, 1995.

[21] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, Aug. 2004.