

A Deep Generative Model for Non-intrusive Identification of EV Charging Profiles

Shengyi Wang, *Student Member, IEEE*, Liang Du, *Senior Member, IEEE*, Jin Ye, *Senior Member, IEEE*, Dongbo Zhao, *Senior Member, IEEE*

Abstract—The proliferation of electric vehicles (EVs) brings environmental benefits and technical challenges to power grids. An identification algorithm which can accurately extract individual EV charging profiles out of widely available smart meter measurements has attracted great interests. This paper proposes a non-intrusive identification framework for EV charging profile extraction, which is driven by deep generative models (DGM). First, the proposed DGM is designed as a representation layer embedded into the Markov process and used to model the joint probability distribution of available time-series data. A novel contribution is to approximate posterior distributions by neural networks whose parameters are obtained by variational inference and supervised learning. Second, the EV charging status is inferred from the DGM via dynamic programming. Lastly, the desired EV charging profile can be reconstructed by the rated power of EV models and inferred status. Compared with the benchmark Hidden Markov Models, the proposed framework can better handle noise in data with less computational complexity and better overall accuracy performances with smaller recall. The proposed framework is validated by numerical experiments on the Pecan Street dataset.

Index Terms—electric vehicles charging, energy disaggregation, load modeling, deep learning, statistical inference

NOMENCLATURE

π	prior probability vector of initial EV charging status
π_i	probability of initial EV charging status being i
π_i^*	estimated π_i
$\mathbf{1}_{ij}$	indicator function
A	probability transition matrix
A_{ij}	probability of EV charging status transits from i to j
A_{ij}^*	estimated A_{ij}
N	number of labeled data
N_b	minibatch size
P	EV charging profile
$p(y_t)$	probability distribution of EV charging status at t
P_t	charging power of an EV charging profile P at t
x	aggregate power consumption profile by AMIs

x_t	value of x at time t
y_t	EV charging status at t
$y_t^{(n)}$	EV charging status of the n -th labeled data at t
z	layer to represent the aggregate power consumption.

I. INTRODUCTION

The worldwide electricity demand profile is experiencing a paradigm shift with increasing penetration of electrified transportation. In the U.S., it is expected that transportation electrification will drive domestic electricity demand rise through 2050 [1], by when over 2.3 million new light-duty electric vehicles (EVs) will be sold annually [2]. Across the globe, many major economies have announced their intentions to end the sale of internal combustion engine vehicles [3] within several decades. The impact of high volume of EVs on power grids has been extensively studied in literature [4]. In general, EVs have been considered as active loads which could provide flexibility in terms of grid services [5] through vehicle-to-grid (V2G) modes [6] or transactive controls [7].

In the literature, aggregated EV charging demands are modeled as a stochastic part of the overall load model. However, the uncertainty in individual EV charging profiles (i.e., start charging time, initial state-of-charge (SOC), charging power, and charging duration) [8] and traffic conditions [9] makes it difficult to accurately derive real-time EV charging demand models under various scenarios. Therefore, probabilistic distributions are typically assumed. In [10] and [11], the charging start time is represented by the normal distribution. Similar, a truncated normal distribution is suggested to represent the arriving time and parking time at commercial buildings [12] for EV charging duration. Furthermore, in [13], EV charging duration is assumed to be exponentially distributed. Moreover, the initial SOC is modeled as a random variable under log-normal distribution [8]. However, it is questionable whether these assumptions from locational models can be used in other regions. For example, charging start times in rural residential, urban residential, and commercial districts at different seasons are unlikely to be the same. Therefore, in recent years, pilot projects have been carried out globally to collect and analyze EV charging profiles in the Netherlands [14], U.K. [15], Australia [16], and California [17].

However, most of the historical data is only small-scale and sampled at commercial charging stations. For residential applications, it is costly to (intrusively) install additional sampling devices into existing residential EV chargers and (more importantly) unrealistic to sample and communicate

Manuscript received September 4, 2019; revised January 13, 2020 and April 9, 2020; accepted May 24, 2020. Date of publication XXXXXX, 2020; date of current version XXXXXX, 2020. This work was supported in part by Oak Ridge Associated Universities (ORAU) under Ralph E. Powe Junior Faculty Enhancement Award (L. Du). The work of J. Ye was supported by the National Science Foundation under Grant ECCS-1946057. (Corresponding author: L. Du.)

S. Wang and L. Du are with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122 USA e-mail: {shengyi.wang, ldu}@temple.edu.

J. Ye is with the School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602 USA e-mail: jin.ye@uga.edu.

D. Zhao is with Argonne National Laboratory, Lemont, IL 60439 USA e-mail: dongbo.zhao@anl.gov.

EV charging information to system operators, another recent research effort [18]–[22] focuses on utilizing widely-available smart meter data to non-intrusively, locally, and reliably estimate EV charging profiles in real-time to preserve privacy and avoid unnecessary investment in additional infrastructure.

To conclude above discussions, it is of great interests for system operators and planners to extract EV charging profiles from smart meter data in a non-intrusive manner such that 1) unrealistic and uncertain assumptions (as pointed out in the above discussions) can be alleviated; and 2) EV charging profiles can be accurately extracted in real-time to support both short-term system operations and long-term planning.

To our best knowledge, reference [19] is probably the first to adopt non-intrusive load monitoring (NILM) and apply benchmark algorithms such as the Hidden Markov Models (HMMs) [23] to detect events and disaggregate EV charging profiles from low-frequency smart meter readings. Reference [20] presents an unsupervised algorithm to extract EV charging loads non-intrusively from the smart meter data using independent component analysis. Reference [21] aims at identifying EV models to determine charging power. Finally, reference [22] proposes a training-free non-intrusive algorithm based on bounding-box fitting and load signatures. To summarize, HMMs are probably the most popular identification models, which are relatively easy to train but require detailed prior information of all appliances in the aggregated power consumption profiles (and thus cannot tackle unknown appliances). Moreover, computational complexity of an exact inference in HMMs grows exponentially with the sequence lengths and the number of appliances. Therefore, it is desired to design an algorithm to mitigate the aforementioned issues.

This paper retains the Markov property in HMMs but only utilizes one Markov chain to involve only the aggregated and partial EV charging data points as known information. Without involving other appliances' power consumption data, the computational complexity of an exact inference is greatly reduced. Specifically, this paper proposes a deep generative model (DGM)-driven inference framework for non-intrusive, real-time identification of EV charging profiles. Firstly, the joint probability distribution for available smart meter data (which can actually be considered as time series) is modeled by deep generative models (DGMs). A novel contribution by this paper is to approximate posterior distributions by neural networks whose parameters are obtained by variational inference and supervised learning. Secondly, the EV charging status is inferred from DGMs via dynamic programming. Finally, the target EV charging profile can be reconstructed according to the rated power of EV models and inferred status.

The main contributions of this paper are listed as follows.

- Compared to the existing literature in which most works need to manually define features, the proposed DGM with convolutional neural layers can automatically extract features and represent highly nonlinear features of aggregated power consumption profile with less weights.
- Compared to the existing literature in which assumptions on prior knowledge are typically made, the proposed framework makes full use of smart meter data to extract

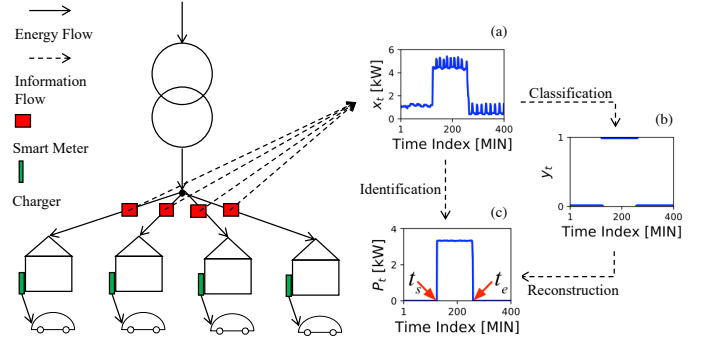


Fig. 1. Overview of EV charging profile identification. (a): a sample aggregated power consumption profile; (b): its corresponding EV charging status (charging started at t_s and ended at t_e); and (c) its corresponding EV charging profile.

EV charging profiles without any need of prior knowledge in other appliances being used at the same time as EVs.

- Two different schemes with both transfer and non-transfer learning settings have been studied, and results show that the proposed framework possesses good robustness against noise and error in data as well as generalization capability to unseen data.

The remainder of this paper is organized as follows. Section II defines the EV charging profile identification problem considered in this paper and then formulates it with in the architecture of NILM. Next, Section III reviews the framework of HMM, which will be used as a benchmark algorithm. Furthermore, Section IV proposes a DGM to model the joint probability distribution of the available aggregated consumption data, of which parameters are obtained by variational inference and supervised learning. Section V utilizes dynamic programming to perform exact inference of the DGM for the EV charging status. Moreover, Section VI discusses numerical validation setup and results. Finally, Section VII presents conclusions and future work.

II. PROBLEM FORMULATION

The EV charging profile identification problem considered in this paper is presented in Fig. 1, in which a sample aggregated power consumption profile is shown in Fig. 1(a), with its corresponding EV charging profile shown in Fig. 1(c). Furthermore, Fig. 1(b) shows the corresponding EV charging status (charging started at time t_s and ended at time t_e).

A. Definitions

Given an aggregated power consumption profile $x = (x_1, \dots, x_T)$, i.e., a timed sequence of a total of T power consumption data points, determine its corresponding EV charging profile P (or $P(x)$ if the source power consumption profile x is relevant). Note that the power consumption profile is called aggregated as most smart meters measure the power consumption of the whole household and thus include all loads (i.e., aggregated). An EV charging profile is thus a timed sequence (of the same length T) of EV charging power consumption data points. In other words, the value P_t of P at time step t denotes the amount of power by EV charging.

Moreover, at time step t , the charging status y_t of an EV is binary, i.e., either ON (i.e., $y_t = 1$ if P_t is greater than a pre-defined threshold P_{th}) or OFF ($y_t = 0$ otherwise). Furthermore, the probability of an EV at its y_t is denoted by $p(y_t)$. When $y_t = 1$, $p(y_t = 1) = 1$ and $p(y_t = 0) = 0$. When $y_t = 0$, $p(y_t = 1) = 0$ and $p(y_t = 0) = 1$.

B. EV Charging Profile Identification as NILM

The objective of the EV charging profile identification problem considered in this paper is to determine EV charging profile P given aggregated power consumption profile x . Therefore, the scope of this work falls within the framework of a NILM problem. Most techniques used for NILM problems in the literature consist of two sub-tasks: 1) classification and 2) reconstruction. The former task aims at classifying the load operation status into known categories, and the latter task is to reconstruct load consumption profiles using classification results. For example, if the first task returns that the charging status of a certain model of EV at time step t is classified to be ON (i.e., $y_t = 1$) with rated power consumption around 6.7 kW (i.e., $P_t = 6.7$), then the latter task would focus on reconstructing its corresponding EV charging profile. Therefore, this paper follows [21], [24] to assume that EV charging power level and corresponding models can be identified separately and mainly focuses on EV charging status classification and converts the EV charging profile identification problem into a binary EV charging status classification task.

Therefore, the EV charging profile identification problem this paper aims to solve can be formulated as follows: given an aggregated power consumption profile $x = (x_1, \dots, x_T)$, determine y_t of an EV at each time step $t = 1, \dots, T$. The general procedure of how the proposed EV charging profile identification problem is studied in this paper is presented in Fig. 2. First of all, each generative process for time-series data is modeled by a joint probability distribution in Step 1. Secondly, each component (e.g., evidence and transition probabilities) in the joint distribution from Step 1 is approximated by a common parametric density function in Step 2. Thirdly, parameters of the density function (in the joint distribution) from Step 2 are learned by maximum likelihood estimation in Step 3. Finally, the above-defined identification problem is converted to a Bayesian inference process by the proposed DGM in Step 4. The above steps are further specified in Sections IV and V.

III. BENCHMARK ALGORITHM: HMMs

Hidden Markov Models (HMMs) are a widely acknowledged tool for modeling time series, which have been utilized for the EV charging profile identification problem in the following manner. Given the aggregated power consumption profiles and their corresponding EV charging profiles, their joint probability distribution can be modeled by a standard HMM [25]. The graphical illustration of such an HMM is given by Fig. 3, in which an aggregated power consumption profile $x = (x_1, \dots, x_T)$ and its corresponding EV charging status profile $y = (y_1, \dots, y_T)$ (depicted by nodes) are modeled as random variables. Under this model, the aggregated power

consumption at each step x_t only depends on some unobserved or latent EV charging status y_t , which are depicted by direct edges. Therefore, the former can be generated/emitted by the latter. This conditional generative process can be modeled by $p(x_t|y_t, \theta)$, which is typically known as emission probability [25]. The Markovianity is defined such that y_t only depends on y_{t-1} . Note that the bottom representation layer (shown in blue) in Fig. 3 is the proposed model, which will be discussed in the next section.

Such an HMM can model the following joint probability distribution reflecting the generative process of the time series x and y :

$$p(x, y|\phi) = \underbrace{p(y_1|\pi)}_{\text{Prior}} \underbrace{\left[\prod_{t=2}^T p(y_t|y_{t-1}, A) \right]}_{\text{Transition Probability}} \underbrace{\left[\prod_{t=1}^T p(x_t|y_t, \theta) \right]}_{\text{Emission Probability}} \quad (1)$$

where $\phi = \{\pi, A, \theta\}$ denotes parameters of HMMs, π is the categorical distribution of the initial status, A is the status transition matrix, and θ is the mean and deviation if the emission probability is Gaussian, which can be obtained by counting frequencies through supervised learning. With ϕ learned, the EV charging status y can be inferred by maximizing the *a posteriori* (MAP) $p(y|x, \phi)$ which can be solved by the Viterbi Algorithm [26, p. 629], which is a well-known method for exact inference.

IV. THE PROPOSED DGM: REPRESENTATION AND PARAMETER LEARNING

This section presents in details the proposed DGM-driven framework, which simplifies the classification complexity using only a single (but enhanced) Markov chain and utilizes deep neural networks to approximate posterior probability distributions with weights trained via supervised learning.

A. Representation Layer Embedded Markov Chain

A major innovation of the proposed framework for EV charging status classification is to embed a representation layer (denoted by z) into the Markov chain. The z can be seen as abstract but meaningful features from any aggregated power consumption profile x . As shown in Fig. 5, the left figure shows a standardized aggregated power consumption profile x , and the right figure shows the value of the corresponding representation layer z . The motivation behind the proposed representation layer can be summarized as follows.

- Firstly, although a single Markov chain can classify EV charging status, the representation capability by raw, aggregated power consumption profiles (i.e., inputs) is relatively weak due to similar and ambiguous features;
- Secondly, nonlinear dynamics in raw time-series inputs might present higher non-stationary variances and thus fail to provide useful features;
- Finally, the learned likelihood distributions $p(x_t|y_t = 0)$ and $p(x_t|y_t = 1)$ from raw inputs may have significant overlaps, which could cause mis-classification issues in later steps. As illustrated in Fig. 4, the learned (Gaussian mixture) likelihood distributions for $p(x_t|y_t = 0)$ and

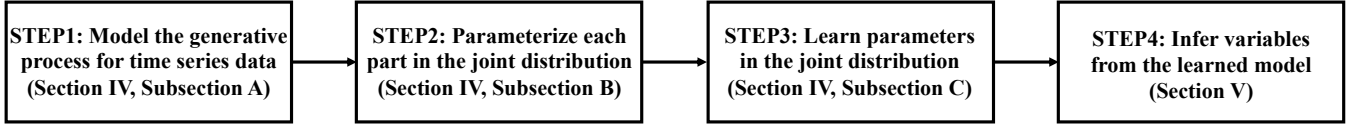


Fig. 2. The general procedure of the proposed identification algorithm under the architecture of NILM can be divided into four steps.

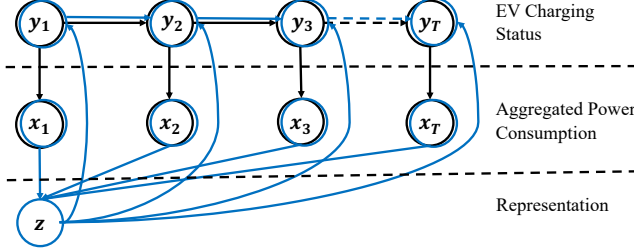


Fig. 3. Graphical illustration of how aggregated profiles and EV charging profiles are represented by an HMM (black) vs. the proposed representation-layer-embedded Markov model (blue), which introduces an additional layer of representations.

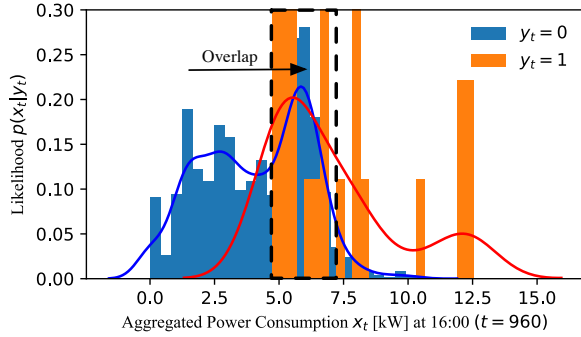


Fig. 4. An example likelihood function $p(x_t|y_t)$ in HMM, where the learned likelihood distributions for the two distinctive scenarios $p(x_t|y_t = 0)$ and $p(x_t|y_t = 1)$ are represented by the blue and red lines, respectively.

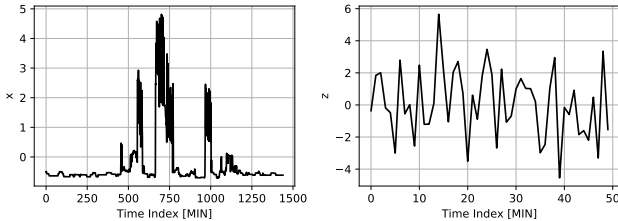


Fig. 5. Illustration of a sample standardized aggregated power consumption profile (left) and its corresponding representation layer (right).

$p(x_t|y_t = 1)$ are represented by the blue and red lines respectively and overlap in the dash rectangle area.

The proposed representation-layer-embedded Markov model is illustrated by Fig. 3, which depicts the generative process of x and y with z (depicted by nodes) also modeled as a random variable. Considered as measurements, x can always be observed, while y can only be partially observed and z always cannot be observed. Compared with the conventional HMM which only contains the upper and middle layers, the proposed representation-layer-embedded Markov model adds a representation layer with the same nodes as the HMM but without directed paths from y to x , which models likelihood

distributions $p(x_t|y_t)$. Both methods generate the same x and y from different perspectives. The HMM assumes that x can be generated by or correlated with y . Moreover, the proposed model assumes that the feature z can generate y (depicted by directed edges). Generating such a z needs to have a powerful feature extractor, which is one of reasons to use deep neural networks.

The two directed edges connecting y_{t-1} to y_t and z to y_t meet the Markov property, i.e., current charging status only relies on the previous charging status and representation layer. The advantage of adding a representation layer is to alleviate the above-discussed likelihood distributions overlapping issue (as shown in Fig. 4), i.e., the proposed model learns the posterior distributions $p(z|x)$ and $p(y|z) = \prod_{t=1}^T p(y_t|z)$ instead of all $p(x_t|y_t)$, which will be studied in details in later sections.

B. Probability Distribution Approximation in DGM

In this work, a DGM is developed based on the proposed representation-layer-embedded Markov model with the following characteristics:

- 1) **Deep**: using deep neural networks to approximate distributions;
- 2) **Generative**: taking x as input to generate z and then using z to generate a sequence of y_t , which can be abstracted into the following joint probability distribution

$$p(x, y, z|\phi) = \underbrace{p(x)}_{\text{Evidence}} \underbrace{p(y_1|\pi)}_{\text{Prior}} \underbrace{\left[\prod_{t=2}^T p(y_t|y_{t-1}, A) \right]}_{\text{Transition Probability}} \underbrace{p(z|x, \theta)}_{\text{Posterior of } z} \underbrace{\left[\prod_{t=1}^T p(y_t|z, \theta) \right]}_{\text{Posterior of } y_t} \quad (2)$$

where $\phi = \{\pi, A, \theta\}$ denotes parameters of the proposed representation-layer-embedded Markov model.

The probability distributions in the right-hand side of (2) are discussed in details as follows.

1) $p(y_1|\pi)$: $\pi \in [0, 1]^2$ is prior probability distribution of the initial EV charging status, in which $\pi_i := p(y_1 = i)$, $i \in \{0, 1\}$, represents the probability that initially an EV is at charging status i . With N labelled initial $y_1^{(n)}$, $n = 1, \dots, N$, π_i can be estimated by

$$\pi_i^* = \begin{cases} \frac{\sum_{n=1}^N y_1^{(n)}}{N} & \text{if } i = 1; \\ 1 - \frac{\sum_{n=1}^N y_1^{(n)}}{N} & \text{if } i = 0. \end{cases} \quad (3)$$

2) $p(y_t|y_{t-1}, A)$: $A \in \mathbb{R}^{2 \times 2}$ is the probability transition matrix, in which A_{ij} represents the probability that the EV charging status transits from i at $t-1$ to j at t :

$$A_{ij} = p(y_t = j|y_{t-1} = i) \quad \text{where } i, j \in \{0, 1\}. \quad (4)$$

With N labelled $y_t^{(n)}$, $n = 1, \dots, N$, A_{ij} can be estimated by

$$A_{ij}^* = \frac{\sum_{n=1}^N \sum_{t=2}^T \mathbf{1}_{ij}(y_t^{(n)} = j | y_{t-1}^{(n)} = i)}{N(T-1)}, \quad (5)$$

where $\mathbf{1}_{ij}(y_t = \cdot | y_{t-1} = \cdot)$ is an indicator function whose output is 1 if and only if $y_t^{(n)} = j$ and $y_{t-1}^{(n)} = i$.

3) $p(z|x, \theta)$ and $p(y_t|z, \theta)$: Recall that an abstract but meaningful feature, z is always unobservable. Therefore, the true posterior distribution of z given x is unknown. This paper follows [27], [28] to assume that $p(z|x, \theta)$ takes on an approximate Gaussian form, i.e., a multivariate Gaussian with a diagonal covariance, given as

$$\log p(z|x, \theta) = \log N(z | \mu_z(x), \sigma_z^2(x) \mathbf{I}) \quad (6)$$

where $\mu_z(x)$ and $\sigma_z(x)$ are the mean and standard deviation of z , respectively. Because y_t is a binary variable, it is assumed that y_t follows a Bernoulli distribution

$$\log p(y_t|z, \theta) = \log(\mu_{y_t}(z)^{y_t} (1 - \mu_{y_t}(z))^{1-y_t}) \quad (7)$$

where $\mu_{y_t}(z)$ is the mean of y_t , which can also be interpreted as the probability that an EV is at ON charging status given z at t , i.e., $\mu_{y_t}(z) = p(y_t = 1|z, \theta)$. Thus the $p(y|z, \theta)$ is a multivariate Bernoulli distribution that is a product of Bernoulli distribution of each y_t .

In this work, deep neural networks (DNNs) [29, Ch. 6 and 9] are used to approximate $p(z|x, \theta)$ and $p(y_t|z, \theta)$. Compared to linear regression models, DNNs can better extract features in a robust manner (against noise and error in data) and generalize to new data. Moreover, recurrent layers are often used in DNNs to process time-series data. However, recurrent layers have difficulties in handling long sequences, such as sequences more than 100 time steps [29, Ch. 10]. Therefore, instead of recurrent layers, this paper utilizes fully connected (FC) and convolutional neural network (CNN) layers. The former layer is the weighted sum of inputs transformed by nonlinear activations into outputs. The latter layer employs convolution operations to extract nonlinear features [30]. The CNN layer is demonstrated in Fig. 6, where

- **Conv1D** is a one-dimensional (1D) convolutional layer applied to time series. Its parameters consist of a set of learnable filters, which can capture features such as the spatial structure (e.g., change points) and local information (e.g., magnitudes) of input time series. During the forward pass, each filter is convolved from the beginning of a time series towards its end, computing the dot product between the entries of the filter and the input and producing a feature map of that filter. For instance, an example time series 1, 1, 2, -1, 1, -2, 1 with a filter 1, 0, -1 is given in Fig. 6. The convolution operation can be considered as the dot product, e.g., the first output -1 after the convolution operation can be calculated as $1 \times 1 + 1 \times 0 + 2 \times (-1) = -1$. Note that the output -1, 2, 1, 1, 0 after the convolution operation is called a feature map.
- **UpSampling1D** is an operation to upsample (i.e., repeat and resize) the feature map. For instance, given a feature map -1, 2, 1, 1, 0 as shown in Fig. 6, the

output after the upsampling operation with factor 2 is -1, -1, 2, 2, 1, 1, 1, 1, 0, 0.

- **MaxPooling1D** is an operation to reduce the input size by taking the maximum value of sliding windows in the original input. For instance, considering the feature map -1, 2, 1, 1, 0, the output after **MaxPooling1D** with size 3 can be calculated as 2, 2, 1, where the corresponding sliding windows are [-1, 2, 1], [2, 1, 1], and [1, 1, 0].

The above three operations can be performed with or without overlaps and paddings, which would imply different final output sizes. The complete network architecture and parameter settings of the proposed DGM are shown in Table I, which presents approximation actions of $p(z|x, \theta)$ and $p(y_t|z, \theta)$ in each layer. The approximation of $p(z|x, \theta)$ is carried out by two convolutional layers with the same padding, two pooling layers with stride size of 2, and a FC layer for $\mu_z(x)$ and $\log \sigma_z(x)^2$. Moreover, the approximation of $p(y_t|z, \theta)$ is conducted by a FC layer, two upsampling layers, and two convolutional layers with the same padding. Since $\mu_{y_t}(z)$ is a probability, the sigmoid function is selected to be the activation of the last layer.

Compared with FC layers, the proposed DGM can benefit from convolutional layers. On one hand, convolutional layers can automatically capture useful features from local patterns. As shown in Fig. 7, an one-dimensional convolutional layer with 16 filters is applied to a sample aggregated power consumption profile to illustrate this advantage. Each subplot shows the identical power consumption profile (lines in black), EV charging operation status (lines in red), and feature map outcomes of one filter (lines in blue). Each trained filter aims at extracting both the spatial structure (e.g., change points) and local information from the input and providing feature maps for its downstream layer. In this illustration, filters #3, #4, #6, #7, #8, #12, #13, and #16 can effectively extract feature maps that match the EV charging status change points (step changes in red lines). The other filters also introduce positive impacts, but not as strong as these ones. For different inputs, in general there would be a different subset of filters that perform dominating roles. On the other hand, convolutional layers with weight sharing can reduce the number of neural network weights and indirectly prevent overfitting.

Remark 1: The number of weights in convolutional layers depends on the product of the number and the size of convolution filters (e.g., 16×3), while the number of weights in FC layers relies on the product of the input and output size (e.g., 360×50). In general, the latter is much larger than the former. Therefore, given the same number of layers, a DNN with more convolution layers requires less weights.

Remark 2: Though it is generally true that recurrent layers perform better in capturing temporal information than convolutional layers, they tend to have difficulties in handling very long time series. Some recent results [30], [31] have shown that convolutional layers can achieve better performances than recurrent layers in some applications. Moreover, this paper utilizes convolutional layers for different purposes, such as extracting spatial structures. As a result, the temporal information is not processed in the neural network part of the proposed model. However, since the proposed DGM retains

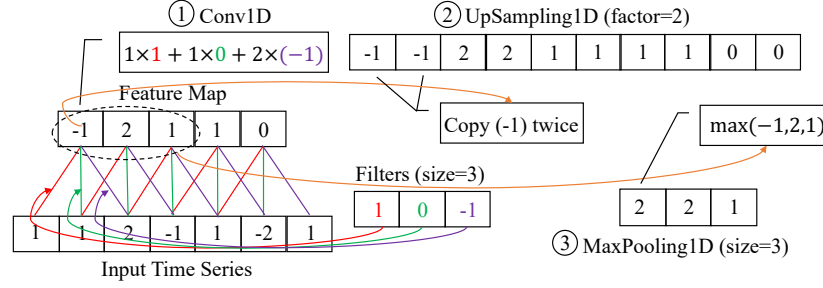


Fig. 6. Demonstration of a CNN layer, which consists of three sequential operations: Conv1D, UpSampling1D, and MaxPooling1D.

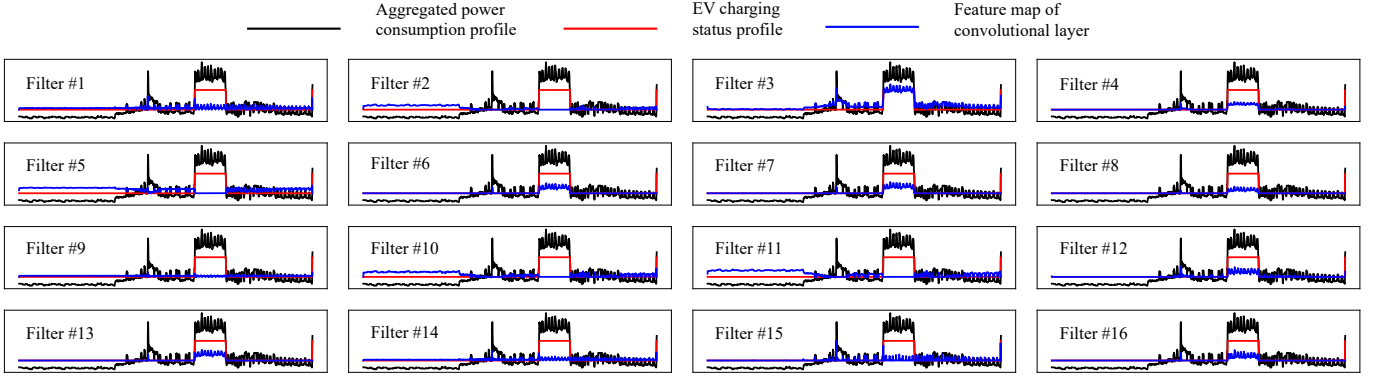


Fig. 7. Visualization of feature maps extracted from convolutional layers with sixteen filters.

TABLE I
THE PROPOSED DGM ARCHITECTURE AND PARAMETER SETTINGS

Layer Name	$p(z x, \theta)$	$p(y z, \theta)$
Input	1440	50
Layer1	Conv1D,16,3,ReLU	FC,360,ReLU
Layer2	MaxPooling1D,2	UpSampling1D,2
Layer3	Conv1D,1,3,ReLU	Conv1D,16,3,ReLU
Layer4	MaxPooling1D,2	UpSampling1D,2
Layer5	FC,50/50,-	Conv1D,1,3,sigmoid

*Conv1D denotes 1D convolution layer followed by number and size of filters and an activation layer; MaxPooling1D denotes 1D max pooling layer followed by size of the max pooling windows; UpSampling1D denotes 1D upsampling layer followed by upsampling factors; FC denotes a fully connected layer followed by number of neurons and an activation layer

the Markov property of HMMs, the corresponding temporal information is addressed by transition probabilities. In other words, convolutional layers with Markov property is proposed here as an alternative of recurrent layers for time series.

C. Supervised Learning in DGM

With labelled dataset $(\mathcal{X}, \mathcal{Y})$, θ in (6) and (7) can be determined by maximum likelihood [26]. The marginal distribution of each $(x, y) \in (\mathcal{X}, \mathcal{Y})$ is obtained from the joint distribution (2) by marginalizing over the latent variable z

$$\log p(x, y|\theta) = \log \int_z p(x, y, z|\theta) dz \quad (8)$$

Maximizing (8) could lead to complicated expressions with no closed-form solutions since 1) the integral of the marginal

distribution is intractable when $p(z|x, \theta)$ and $p(y|z, \theta)$ are approximated by DNNs with nonlinear hidden layers and 2) batch optimization is costly for large amount of data. Following recent advances in variational inference [32], the proposed DGM can be trained by maximizing the evidence lower bound (ELBO) under data distribution. A lower bound $\mathcal{L}(\theta|x, y)$ on the marginal distribution of (x, y) is given by

$$\begin{aligned} \mathcal{L}(\theta|x, y) = & \log p(x) + \log p(y_1) + \sum_{t=2}^T \log p(y_t|y_{t-1}) \\ & + \sum_{t=1}^T E_{z \sim p(z|x, \theta)} [\log p(y_t|z, \theta)], \end{aligned} \quad (9)$$

with proof given in the Appendix. Therefore, the ELBO under data distribution $x, y \sim p_{data}$ can be written as

$$\mathcal{L}(\theta|\mathcal{X}, \mathcal{Y}) = E_{x, y \sim p_{data}} [\mathcal{L}(\theta|x, y)]. \quad (10)$$

Then (6) and (7) can be trained by

$$\theta^* = \arg \min_{\theta} -\mathcal{L}(\theta|\mathcal{X}, \mathcal{Y}) \quad (11)$$

Since z is stochastic and thus gradients cannot be backpropagated, reparameterization is used to sample z , i.e., given x and a unit Gaussian noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $z = \mu_z(x) + \epsilon \sigma_z(x)$. Note that the noises injected into the representation layer enables the proposed DGM to learn continuous feature representations. Note that such a sampling process of z is similar to the variational autoencoder (VAE) [27]. Therefore, in this paper the number of samples z is set to be 1 with a large minibatch size N_b in accordance with the experimental setting in the

VAE. Based on (6), (7), and (11), it can be concluded that

$$\theta^* = \arg \min_{\theta} - \sum_{n_b=1}^{N_b} \sum_{t=1}^T y_t^{(n_b)} \log \mu_{y_t}(z^{(n_b)}) + (1 - y_t^{(n_b)}) \log(1 - \mu_{y_t}(z^{(n_b)})) \leftarrow \text{Loss} \quad (12)$$

where $z^{(n_b)} = \mu_z(x^{(n_b)}) + \epsilon^{(n_b)} \sigma_z(x^{(n_b)})$ and $\epsilon^{(n_b)} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$, $x^{(n_b)}$ and $y^{(n_b)}$ are the n_b -th instance from minibatch, $z^{(n_b)}$ is generated from $x^{(n_b)}$, and $y_t^{(n_b)}$ is the t -th element of $y^{(n_b)}$. From (12), it can be seen that the training objective is to minimize the binary multi-label classification loss. In this paper, $p(z|x, \theta)$ and $p(y_t|z, \theta)$ are both differentiable functions containing different neural layers composed of multilayer perceptrons, convolution, max-pooling, upsampling, Rectified Linear Units (ReLU), and sigmoid. Therefore, the gradient-descent based training methods Adam [33] is applied, which is fairly insensitive to the choice of hyperparameters.

Furthermore, this paper utilizes the minibatch training, also known as the minibatch gradient descent, which is a variation of the gradient descent algorithm that splits the training dataset into small batches. The implementation flowchart of the minibatch training in DGM is shown in Fig. 8. For each batch, the forward propagation first generates z and outputs $\mu_{y_t}(z)$, and then the back propagation calculates model loss and update model weights.

V. THE PROPOSED DGM: EXACT INFERENCE

Once the proposed model is trained with ϕ^* , the next step is to infer EV charging status y^* given aggregated consumption profile x via maximizing *a posteriori* (MAP), i.e.,

$$y^* = \arg \max_y p(y|x, \phi^*), \quad (13)$$

which is approximated (with z sampled from $p(z|x, \theta^*)$) by

$$y^* = \arg \max_y \log p(y_1) + \sum_{t=2}^T \log p(y_t|y_{t-1}) + \sum_{t=1}^T \log p(y_t|z) \quad (14)$$

Note that y^* in (14) can be further inferred via DP in two stages. For forward induction, at each time step t , the first step is to solve the following

$$F_c(t, y_t) = \min_{y_{t-1}} \{F_c(t-1, y_{t-1}) - \log p(y_t|z) - \log p(y_t|y_{t-1})\} \quad (15)$$

where $F_c(t, y_t)$ is the optimal cost function over time steps t and $t-1$ given y_t , and $F(1, y_1) = -\log p(y_1) - \log p(y_1|z)$. An example is shown in Fig. 9 to demonstrate the calculation of $F_c(t, y_t)$, where the values on the nodes are the cost (or values of the negative logarithm of posterior) and the values on the directed edges are the cost of transporting a unit from one node to the other (i.e., the negative logarithm of transition probability). Therefore, the overall process of forward induction is to find the minimum cost. Note that the complexity of this implementation is $O(4T)$.

For backward induction, the second step is solve the following and find the minimum cost route,

$$y_{t-1}^* = \arg \min_{y_{t-1}} \{F_c(t-1, y_{t-1}) - \log p(y_t^*|z) - \log p(y_t^*|y_{t-1})\} \quad (16)$$

where $y_T^* = \arg \min_{y_T} F(T, y_T)$. The complexity of this implementation is $O(2T)$. Fig. 10 shows four typical inference results corresponding to (a) once-charging, (b) twice-charging in day and night, (c) twice-charging in two nights, and (d) multiple-charging, respectively. It can be observed that the measured and inferred EV charging status are almost identical, which validates the effectiveness of the proposed framework. The incorporation of $p(y_1)$ and $p(y_t|y_{t-1})$ into the graph enables the model to consider the past events at the expense of increased computational complexity of inference.

VI. NUMERIC RESULTS

In this section, the proposed algorithm is validated on the Pecan Street dataset [34], which consists of measurement of circuit-level household electricity consumption data from nearly 1,000 homes across the U.S. Each such home have eight extra channels to record power consumption by major appliances such as HVAC, refrigerators, and EVs.

A. Experiment Setup and Evaluation Metrics

The DGM in this paper is trained using Adam with an epoch of 20, a mini-batch size of 100, and a learning rate of 0.001. All neuron weights are initialized using Glorot initialization [35]. After data pre-cleaning with removal of bad data points, the aggregated power consumption profiles and EV charging profiles are then standardized and binarized, respectively. The main program is executed on an Intel i7-7820X 8-Core CPU while the training of the proposed DGM including the forward and backward propagation is implemented on a TITAN Xp GPU using TensorFlow as the computational framework. It is observed that the loss of the model goes to convergence as the epoch increases, as shown in Fig. 11. To evaluate performance of the proposed algorithm, the following classification metrics are employed,

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{F1} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned}$$

where

- TP is the *true positive* indicator, i.e., is the number of cases where the DGM classifies the EV charging status as ON and the actual status is indeed ON;
- TN is the *true negative* indicator, i.e., the number of cases where the DGM classifies the EV charging status as OFF and the actual status is indeed OFF;
- FP is the *false positive* indicator, i.e., the number of cases where the DGM classifies the EV charging status as ON but the actual status is OFF; and

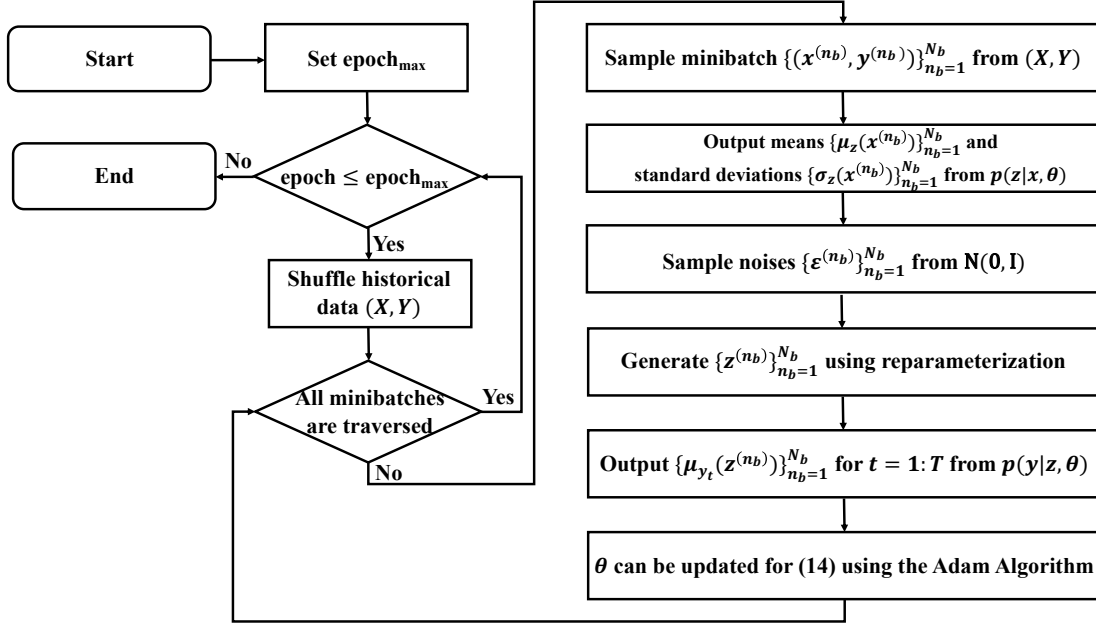


Fig. 8. Implementation flowchart of the minibatch training in DGM, including both forward and backward propagation.

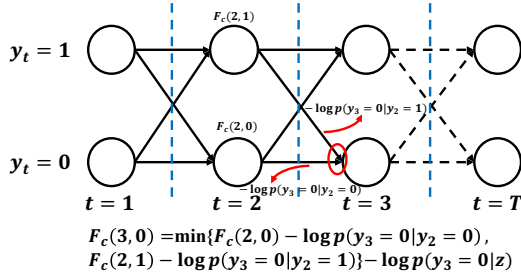


Fig. 9. Example demonstration of how to calculate $F_c(t=3, y_t=0)$ where the red sign highlights the possible minimum cost paths (the directed edges) from the last time step ($t=2$) to the current time step ($t=3$).

- FN is the *false negative* indicator, i.e., the number of cases where the DGM classifies the EV charging status as OFF but the actual status is ON.

B. EV Charging Status Classification

There are 93 houses with EV charging activities in the Pecan Street dataset, with one data point per minute per house. With each aggregated power consumption profile defined to be of 24 hours, i.e., daily profiles. In this work, both transfer-learning-based and non-transfer-learning-based settings are utilized for the purpose of comparison on performances. The difference between these two settings is that for non-transfer learning, the set of houses used in training is typically the same as the set of houses used in testing. On the contrast, for transfer learning, the set of houses used in testing is typically different from those used in training, which is a powerful method to check whether a certain model can “transfer” knowledge from one dataset to another.

For the non-transfer learning setting, to reduce bias and variance caused by the source data and better evaluate the effectiveness of the proposed DGM, the five-fold cross-

validation (i.e., all available data is first shuffled and divided into five subsets, and each trial takes one subset for testing and the other four subsets for training) [26] is performed. As shown in Fig. 12, the box plot and the green triangle are used to visualize the variance and mean of evaluation results for the following three scenarios, respectively.

- The first scenario is the proposed DGM without any noise injected into the test sets, denoted as “DGM w/o noise”;
- The second scenario is the proposed DGM with a Gaussian noise (zero mean and half standard deviation, i.e., $\pm 0.5\text{kW}$), denoted as “DGM w/ noise”;
- The third scenario is the HMM without any noise, denoted as “HMM w/o noise”.

It can be observed that the variance of evaluation results is small for all five data partitions, and thus the proposed DGM is reasonably stable. On average, the proposed DGM increases accuracy, precision, and F1 by 8.20%, 134.39%, 56.43%, respectively, at the cost of reducing recall by 19.31% compared with the HMM on the five different data partitions. In terms of F1, the proposed DGM is better than the HMM with better average performance of accuracy.

Furthermore, another comparative experiment is performed to demonstrate the robustness of the proposed DGM against noise in the data. From Fig. 12, the performance of the proposed DGM could be slightly affected by noise and error in the data (the accuracy, precision, recall, and F1 drop by 0.82%, 2.83%, 19.02%, and 13.61%). However, at a reasonable noise level, the proposed DGM still greatly outperforms HMM. Detailed evaluation results are provided in Table. II

For transfer learning setting, all data are split into the training dataset of 73 households and testing dataset shown in the first column of Table III. For household with dataid 6871, it is an extreme case where there is no any EV charging events. From Table III, it is shown that the proposed DGM increases

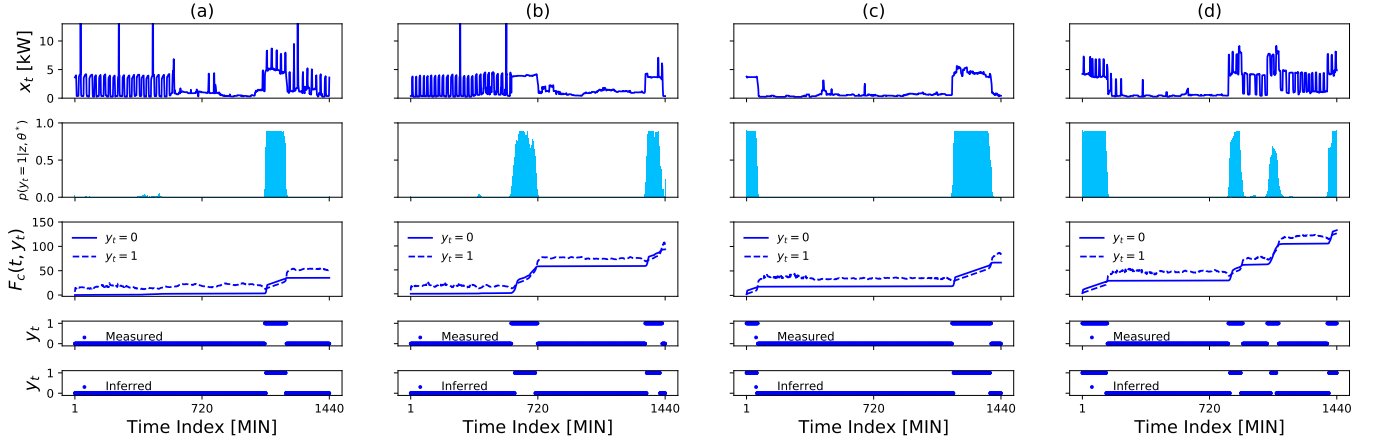


Fig. 10. Selected typical examples of inference results including (a) once-charging, (b) twice-charging in day and night, (c) twice-charging in two nights, and (d) multiple-charging, respectively. Each row shows the aggregated power consumption profile, the probability of an EV at ON, the optimal cost over time steps t and $t - 1$ given y_t , the measured EV charging status profile, and the inferred EV charging status profile, respectively.

TABLE II
PERFORMANCE COMPARISON USING FIVE-FOLD CROSS-VALIDATION AND NON-TRANSFER LEARNING SETTING

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Overall
Accuracy	0.9794 / 0.9692 / 0.9055	0.9801 / 0.9724 / 0.9077	0.9802 / 0.9763 / 0.9084	0.9794 / 0.9728 / 0.9061	0.9809 / 0.9692 / 0.9058	0.9800±0.0006 / 0.9720±0.0027 / 0.9057±0.0011
Precision	0.7683 / 0.6348 / 0.3409	0.8369 / 0.8373 / 0.3494	0.8306 / 0.8275 / 0.3527	0.7751 / 0.7913 / 0.3466	0.8169 / 0.8233 / 0.3417	0.8056±0.0285 / 0.7828±0.0756 / 0.3437±0.0045
Recall	0.8318 / 0.8818 / 0.9822	0.7462 / 0.5548 / 0.9827	0.7613 / 0.6700 / 0.9820	0.8301 / 0.6224 / 0.9826	0.7894 / 0.4765 / 0.9819	0.7917±0.0035 / 0.6411±0.1369 / 0.9812±0.0004
F1 Score	0.7978 / 0.7370 / 0.5061	0.7879 / 0.6661 / 0.5155	0.7935 / 0.7394 / 0.5190	0.8006 / 0.6953 / 0.5124	0.8021 / 0.6021 / 0.5070	0.7964±0.0052 / 0.6880±0.0509 / 0.5091±0.0049

Note that top/middle/bottom numbers in each cell represent metrics of DGM w/o noise, DGM w/ noise, and HMM w/o noise, respectively.

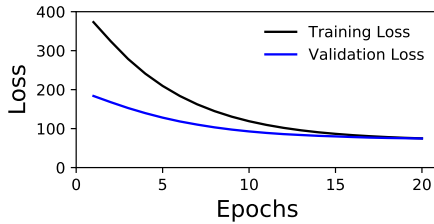


Fig. 11. Illustration of convergence of the training (black line) and validation (blue line) losses as the epochs increase during training.

accuracy, precision, and F1 by 10.8%, 89% and 42.8% at the cost of reducing recall by 17.2% compared with the HMM. In terms of F1, DGM is better than HMM with high average performance of accuracy.

Compared to HMMs, it can be observed that the proposed DGM achieved better performance in *accuracy*, *precision*, and *F1* but only received a lower score in *recall* under both settings. That is because that HMMs cannot accurately classify the EV charging status as “OFF”, i.e., HMMs classifies most “OFF” statuses as “ON” wrongly while the proposed DGM method can mitigate this issue of the HMM method at the cost

of classifying a small number of “ON” statuses as “OFF”.

C. EV Charging Profile Elements Analyses

Four houses (dataid 3036 (a), 370 (b), 1782 (c), and 2018 (d)) are selected according to their different F1 values from high to low. Four elements of EV charging profiles are extracted from measured and corresponding classified EV charging status. Once the extracted results are collected, the distribution of EV charging profile elements can be visualized by Gaussian mixture. As shown in Fig.13, the distribution of EV charging profile elements from measured and classified results are represented by the blue and red lines respectively. For each household, it can be seen that the red line is almost identical to the blue line. That is, the distribution of the measured elements can be well approximated by the classified elements. Therefore, the proposed framework is accurate and effective. According to the distribution of the classified elements, it can be summarized that most households charge their EVs after work around 6 p.m and also tend to charge their EVs for one hour and once per day. So these information can be further analyzed to achieve more accurate EV charging profiles.

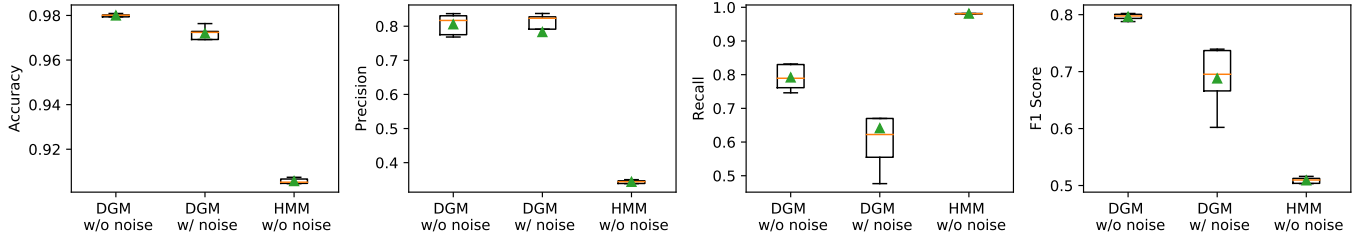


Fig. 12. Performance comparison using five-fold cross-validation and non-transfer learning settings.

TABLE III
PERFORMANCE COMPARISON USING TRANSFER LEARNING SETTING

Dataid	Accuracy [HMM]	Accuracy [DGM]	Precision [HMM]	Precision [DGM]	Recall [HMM]	Recall [DGM]	F1 Score [HMM]	F1 Score [DGM]
370	97.49	98.54	74.50	95.49	98.04	83.28	84.67	88.97
545	94.63	98.68	50.86	95.64	96.40	79.93	66.59	87.08
1185	87.96	97.17	31.64	71.84	97.21	82.34	47.74	76.73
1782	85.85	97.70	28.40	86.04	98.23	70.94	44.06	77.77
2018	77.79	96.11	21.69	67.25	99.30	72.33	35.61	69.70
2335	88.90	96.30	20.48	43.48	96.52	86.88	33.79	57.96
2769	90.58	98.34	31.16	79.26	96.69	83.55	47.13	81.35
3036	98.09	99.28	79.74	96.55	97.11	92.98	87.57	94.73
3367	88.59	98.41	35.54	87.67	98.05	87.28	52.17	87.48
4373	93.40	95.00	67.21	95.96	94.86	63.71	78.68	76.58
4641	97.13	96.18	86.22	92.49	93.27	77.56	89.61	84.37
4957	89.28	98.38	30.25	86.72	99.77	77.05	46.42	81.60
5357	53.17	91.67	6.53	22.62	99.85	63.80	12.25	33.40
5749	92.53	98.34	49.72	84.18	99.87	95.40	66.39	89.44
5786	81.68	96.20	2.15	8.29	99.23	83.30	4.20	15.08
6139	93.16	98.56	45.36	86.13	96.48	89.22	61.70	87.65
6871	91.29	92.66	-	-	-	-	-	-
7863	94.61	98.86	50.10	94.04	98.35	84.20	66.38	88.85
8197	67.49	98.11	12.94	79.00	99.62	83.06	22.90	80.98
8669	88.25	96.20	42.71	77.35	98.64	80.32	59.61	78.81
Overall	87.59±10.58	97.03±2.05	40.38±24.03	76.32±25.07	97.76±1.79	80.90±8.58	53.02±24.34	75.71±20.15

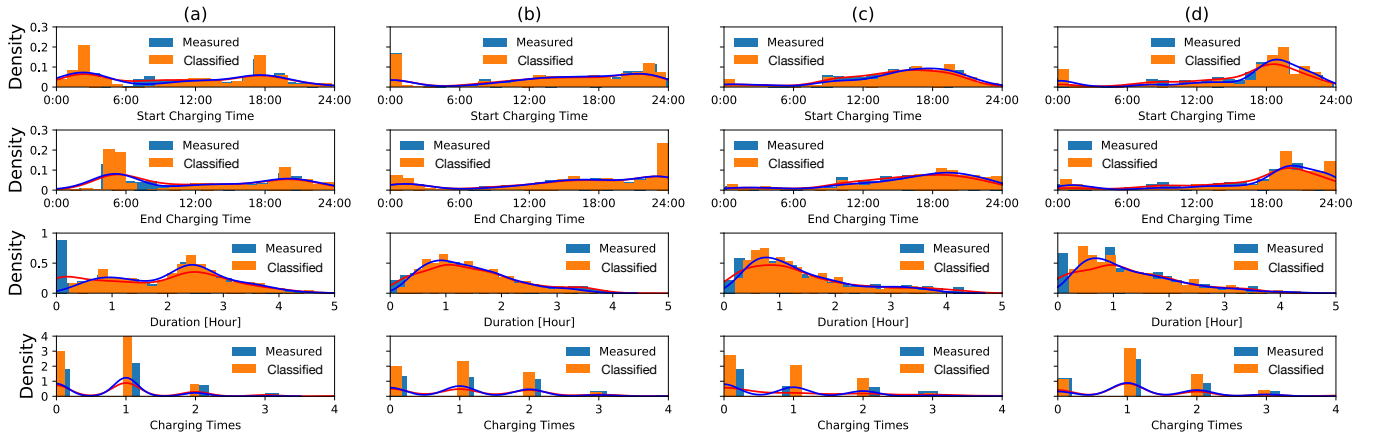


Fig. 13. The analysis of some example EV charging profile elements including start charging time, end charging time, duration, and charging times for four representative houses in terms of their F1 values, where the blue and red lines are the measured and classified distributions, respectively.

VII. CONCLUSION

This paper proposed a DGM driven non-intrusive identification framework for EV charging profile. With the capability of complex density estimation by DGMs, the EV charging status can be modeled and inferred from DGMs via DP. Then EV charging profiles can be reconstructed according to the rated power of EV models and inferred status. Experiments on Pecan Street datasets were conducted to validate the feasibility and effectiveness of the proposed framework. The numerical results show that the proposed method can improve the overall performance compared with the state-of-art HMMs, though a decrease in the recall was observed. In addition, the proposed framework can well handle noisy and unseen data and thus possesses improved robustness and generalization capabilities. For future research, the proposed framework can be extended to more general multi-class multi-label classification tasks.

APPENDIX A PROOF OF ELBO (9)

Proof: Inserting (2) into (8) implies

$$\begin{aligned} \log p(x, y|\theta) &= \log \int_z p(x)p(y_1) \prod_{t=2}^T p(y_t|y_{t-1}) \\ &\quad p(z|x, \theta) \prod_{t=1}^T p(y_t|z, \theta) dz \\ &= \log p(x) + \log p(y_1) + \sum_{t=2}^T \log p(y_t|y_{t-1}) \\ &\quad + \sum_{t=1}^T \log(E_{z \sim p(z|x, \theta)}[p(y_t|z, \theta)]) \end{aligned} \quad (17)$$

According to Jensen's Inequality, (17) implies

$$\begin{aligned} \log p(x, y|\theta) &\geq \log p(x) + \log p(y_1) + \sum_{t=2}^T \log p(y_t|y_{t-1}) \\ &\quad + \sum_{t=1}^T E_{z \sim p(z|x, \theta)}[\log p(y_t|z, \theta)]. \quad \square \end{aligned} \quad (18)$$

REFERENCES

- [1] U.S. Energy Information Admin., "Annual Energy Outlook," 2019.
- [2] D. Gohlke and Y. Zhou, "Impacts of electrification of light-duty vehicles in the U.S., 2010-2017," Argonne National Lab, Tech. Rep., 2018.
- [3] P. G. Pereirinha, M. González, I. Carrilero, D. Anseán, J. Alonso, and J. C. Viera, "Main trends and challenges in road transportation electrification," *Transport. Res. Procedia*, vol. 33, pp. 235–242, 2018.
- [4] A. Dubey and S. Santoso, "Electric vehicle charging on residential distribution systems: Impacts and mitigations," *IEEE Access*, vol. 3, pp. 1871–1893, 2015.
- [5] S. Huang and Q. Wu, "Dynamic tariff-subsidy method for pv and v2g congestion management in distribution networks," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5851–5860, Sep. 2019.
- [6] B. Zhang and M. Kezunovic, "Impact on power system flexibility by electric vehicle participation in ramp market," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1285–1294, May 2016.
- [7] Z. Liu, Q. Wu, K. Ma, M. Shahidehpour, Y. Xue, and S. Huang, "Two-stage optimal scheduling of electric vehicle charging based on transactive control," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2948–2958, 2019.
- [8] K. Qian, C. Zhou, M. Allan, and Y. Yuan, "Modeling of load demand due to ev battery charging in distribution systems," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 802–810, May 2011.
- [9] D. Tang and P. Wang, "Nodal impact assessment and alleviation of moving electric vehicle loads: From traffic flow to power flow," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4231–4242, 2016.
- [10] Y. Cao, S. Tang, C. Li, P. Zhang, Y. Tan, Z. Zhang, and J. Li, "An optimized ev charging model considering tou price and soc curve," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 388–393, 2011.
- [11] S. Shao, M. Pipattanasomporn, and S. Rahman, "Grid integration of electric vehicles and demand response with customer choice," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 543–550, 2012.
- [12] J. Zheng, X. Wang, K. Men, C. Zhu, and S. Zhu, "Aggregation model-based optimization for electric vehicle charging strategy," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1058–1066, June 2013.
- [13] A. Y. Lam, K.-C. Leung, and V. O. Li, "Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 156–166, 2015.
- [14] M. G. Flammini, G. Pretticco, A. Julea, G. Fulli, A. Mazza, and G. Chicco, "Statistical characterisation of the real transaction data gathered from electric vehicle charging stations," *Electric Power Systems Research*, vol. 166, pp. 136–150, 2019.
- [15] E. Xydias, 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, 2016.
- [16] Y. B. Khoo, C.-H. Wang, P. Paevere, and A. Higgins, "Statistical modeling of electric vehicle electricity consumption in the Victorian EV trial, Australia," *Transport. Res. Part D*, vol. 32, pp. 263–277, 2014.
- [17] E. C. Kara, J. S. Macdonald, D. Black, M. Bérge, G. Hug, and S. Kiliccote, "Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach," *Applied Energy*, vol. 155, pp. 515–525, 2015.
- [18] S. Wang, L. Du, J. Ye, and D. Zhao, "Robust identification of EV charging profiles," in *2018 IEEE Transportation Electrification Conf. Expo. (ITEC)*. IEEE, 2018.
- [19] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, and J. K. Moon, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," in *40th Annual Conference of the IEEE Industrial Electronics Society*, Oct 2014, pp. 5419–5425.
- [20] A. A. Munshi and Y. A.-R. I. Mohamed, "Unsupervised nonintrusive extraction of electrical vehicle charging load patterns," *IEEE Trans Ind. Informat.*, vol. 15, no. 1, pp. 266–279, 2019.
- [21] Q. Dang, Y. Huo, and C. Sun, "Privacy preservation needed for smart meter system: A methodology to recognize electric vehicle (EV) models," in *2018 IEEE ISGT-ASIA*. IEEE, 2018, pp. 1016–1020.
- [22] H. Zhao, X. Yan, and L. Ma, "Training-free non-intrusive load extracting of residential electric vehicle charging loads," *IEEE Access*, vol. 7, pp. 117 044–117 053, 2019.
- [23] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [24] V. Hoffmann, B. I. Fesche, K. Ingebrigtsen, I. N. Christie, and M. Punnarud Engelstad, "Automated detection of electric vehicles in hourly smart meter data," *IEEE Access*, 2019.
- [25] Z. Ghahramani, "An introduction to hidden markov models and bayesian networks," in *Hidden Markov models: applications in computer vision*. World Scientific, 2001, pp. 9–41.
- [26] C. M. Bishop, *Pattern recognition & machine learning*. Springer, 2006.
- [27] D. P. Kingma and M. Welling, "Stochastic gradient vb and the variational auto-encoder," in *Second International Conference on Learning Representations, ICLR*, vol. 19, 2014.
- [28] D. I. J. Im, S. Ahn, R. Memisevic, and Y. Bengio, "Denoising criterion for variational auto-encoding framework," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [30] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.
- [31] J. Kelly and W. Knottenbelt, "Neural nlm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015, pp. 55–64.
- [32] C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.

- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [34] "Pecan street database," <http://www.pecanstreet.org>. [Online]. Available: <http://www.pecanstreet.org/>
- [35] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th international conf. artificial intelligence and statistics*, 2010, pp. 249–256.



Shengyi Wang (S'17) received the B.S. and M.S. degrees in electrical engineering from Shanghai University of Electric Power, China, and Clarkson University, Potsdam, NY, in 2016 and 2017, respectively. He is currently pursuing his Ph.D. degree at Temple University, Philadelphia. He received Outstanding Undergraduate Thesis Award in 2016. He was a summer Research Intern at Global Energy Interconnection Research Institute North America (San Jose, CA) and at Mitsubishi Electric Research Labs (Cambridge, MA), in 2019 and 2020, respectively.

His research interests include game-theoretic control for multi-agent systems, data-driven optimization for power systems, and non-intrusive load monitoring.



Liang Du (S'09–M'13–SM'18) received the Ph.D. degree in electrical engineering from Georgia Institute of Technology, Atlanta, GA in 2013. He was a Research Intern at Eaton Corp. Innovation Center (Milwaukee, WI), Mitsubishi Electric Research Labs (Cambridge, MA), and Philips Research N.A. (Briarcliff Manor, NY) in 2011, 2012, and 2013, respectively. He was also an Electrical Engineer with Schlumberger, Sugar Land, TX, from 2013 to 2017. He is currently an Assistant Professor with Temple University, Philadelphia. Dr. Du received the

Ralph E. Powe Junior Faculty Enhancement Award from ORAU in 2018 and currently serve as an associate editor for IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS and IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION.



Jin Ye (S'13–M'14–SM'16) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, China, in 2008 and 2011, respectively. She received her Ph.D. degree in electrical engineering from McMaster University, Canada in 2014. She is currently an assistant professor of electrical engineering and the director of intelligent power electronics and electric machines (IPEM) laboratory at the University of Georgia. Her main research areas include power electronics, electric machines, energy management systems, and cyber-

physical security and resilience with applications to smart grids and electrified transportation.

Dr. Ye was the general chair of 2019 IEEE Transportation Electrification Conference and Expo (ITEC) and a publication chair and women in engineering chair of 2019 IEEE Energy Conversion Congress and Expo (ECCE). She serves as a steering committee member for the cyber-physical security initiative for the IEEE Power Electronics Society (PELS), a PELS representative for the Transportation Electrification Community (TEC), and TEC representative for PELS Technical Committee on Transportation and Vehicle Systems. She is an associate editor for IEEE TRANSACTIONS ON POWER ELECTRONICS, IEEE OPEN JOURNAL OF POWER ELECTRONICS, IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.



Dongbo Zhao (SM'16) received his B.S. degrees from Tsinghua University, Beijing, China, the M.S. degree from Texas A&M University, College Station, Texas, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, Georgia, all in electrical engineering. He has worked with Eaton Corporation from 2014 to 2016 as a Lead Engineer in its Corporate Research and Technology Division, and with ABB in its US Corporate Research Center from 2010 to 2011. Currently he is a Principal Energy System Scientist with Argonne National Laboratory, Lemont, IL. He is also an Institute Fellow of Northwestern Argonne Institute of Science and Engineering of Northwestern University. His research interests include power system control, protection, reliability analysis, transmission and distribution automation, and electric market optimization.

Dr. Zhao is the editor of IEEE TRANSACTIONS ON POWER DELIVERY, IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, and IEEE POWER ENGINEERING LETTERS. He is the subject editor of subject "Power system operation and planning with renewable power generation" of IET Renewable Power Generation and the Associate Editor of IEEE ACCESS.