Abstract

To achieve long-term temperature goal of Paris Agreement and carbon neutrality by the mid-century, nonintrusive load monitoring (NILM) provides a promising approach to reducing electricity usage and carbon emission for residential and commercial buildings. Existing methods of NILM in literatures generally suffer from high computational complexity and/or low accuracy in identifying working household appliances. This paper proposes an event-driven Factorial Hidden Markov model (eFHMM) for multiple appliances with multiple states in a household, aiming for low computational complexity and high load disaggregation accuracy on with high-resolution electrical measurements, which fills the research gaps in terms of complexity and accuracy. The proposed eFHMM decreases the computational complexity to be linear to the event number, which ensures online load disaggregation. Furthermore, the eFHMM is solved in two stages, where the first stage identifies state-changing appliance using transient signatures and the second stage confirms the inferred states using steady-state signatures. The combination of transient and steady-state signatures, which are extracted from transient and steady periods segmented by detected events, enhances the uniqueness of each state transition and associated appliances, which ensures accurate load disaggregation. The event-driven two-stage NILM solution, termed as eFHMM-TS, is naturally fit into an edge-cloud framework, which makes possible the real-world application of NILM. The proposed eFHMM-TS method is validated on the high-resolution datasets such as PLAID, LIFTED and synD and low-resolution datasets such as AMPds, REFIT, REDD and compares with 20 different methods. Results demonstrate that the eFHMM-TS method outperforms other methods on the high-resolution datasets and can be applied in practice.

Introduction

To achieve the long-term temperature goal of the Paris Agreement, countries aim to reach global peaking of greenhouse gas emission and carbon neutrality by mid-century [1]. As the world's largest CO_2 emitter, China aims to peak its carbon dioxide emissions before 2030 and achieve carbon neutrality by 2060 [2]. Similarly, the U.S. wants to be carbon neutral by 2050 [3]. In 2020, emission of carbon dioxide by the U.S. electric power sector was 1448 million metric tons (MMmt), about 32% of total U.S. energy-related CO_2 emission. Residential and commercial buildings contribute 75% of electricity-related CO_2 emission [4].

Research shows that solar and wind energy are two of the most important options to mitigate climate change [5], [6], and their annual growth rates over 2010–50 period would be nearly 10% [7]. This will lead to high penetration of renewable energy such as solar and wind energy that may not generate enough electricity to meet peak demand at night. Besides, it is projected that the global electricity demand in 2040 would increase by 58% compared with that in 2018 [8]. Therefore, effective power management would increase the efficiency of energy conservation and renewable energy utilization. Residential and commercial building accounts for more than 40% of electricity consumption in developed countries [9], even 73.5% in the U.S [10]. So, building load management should be paid more attention to reduce CO2 emission and optimize power usage. Specifically, demand response (DR) technology, which utilizes the flexibility of specific loads to support distributed renewable energy utilization and provides ancillary services for distribution network, enables efficient power usage. Research shows that real-time feedback of detailed power information of individual appliances can save power consumption [73] through direct load control (DLC). In DLC, utility is able to shape customer power profile by remotely controlling appliances such as heat, air conditioning and ventilation system. It adjusts the noncritical loads by shifting and reducing their consumption so that available generation, especially wind and solar power, can be employed efficiently.

To estimate the DR potential of each customer, it is necessary to monitor appliances' daily activities. Both intrusive load monitoring (ILM) and nonintrusive load monitoring (NILM) are used to acquire appliance-level power details. Out of consideration for budget and privacy, ILM is ruled out as it requires one sensor connecting to each appliance. NILM is to decompose the aggregated load measured at the household level from smart meters into individual contribution of each appliance, which was initiated by Hart in 1990s [11]. After the past thirty years of development, and especially over the last decade, NILM has become an active area of research. More promising solutions presented in literatures are based on machine learning algorithms, and general scheme includes extracting signatures of individual appliances, identifying the associated appliances and estimating their power consumption using supervised and unsupervised algorithms [12], [13], [14]. Supervised algorithms require appliance ground truth to train models or require support of users to sequentially switch on and off the appliances of interest in practice [11]. It is reported that the requirements have been partially reduced by extracting signatures for selected appliances while other appliances remain on [12]. Unsupervised algorithms could automate the learning process without users' involvement [13] and dynamically adapt to the power system changes over time. However, these algorithms struggle to apply an appropriate label to disaggregated signals of different appliances. Such difficulties are being solved by building a generic labelled dataset that can be used for unseen household data [14]. Generally, NILM should develop from supervised algorithms to unsupervised ones by learning and updating appliance models.

In the past years, different methods have been developed to achieve significant disaggregation performance. K-Nearest Neighbors [15], Wavelet transform [16], Decision Trees [54], Graph Signal Processing [17], [18], Karhunen-Loeve expansion (KLE) based subspace separation method [43], Deep Neural Network (DNN) [19], [20], [21], [22], [23], [24], and Hidden Markov models (HMM) [26], [27], [28], [29], [30], [31], [32] have been successfully employed for NILM. With the development of deep learning techniques, DNN-based methods have been paid much more attention. Neural-NILM [19] proposes three neural network architectures for energy disaggregation, being the first of its kind. Subsequent works further develop neural network based NILM to enhance the performance. Ref. [20] applies gated recurrent unit (GRU) based on recurrent neural network (RNN) to classify the activations of target appliances. A convolutional neural network (CNN) based method classifies the switching on events in [21]. Ref. [22] modifies seq2seq learning to seq2point learning in CNN structure to decrease computational cost. Their results indicate that these models struggle to achieve high accuracy for multi-state appliances with repeatable switching patterns. It is further developed to apply the trained model to other households using transfer learning [23]. Ref. [24] uses long short-term memory (LSTM) to classify appliances based on denoising autoencoder. Refs. [59], [60] use feature extraction method to introduce novel signatures for LSTM-based model to improve the appliance identification performance. Ref. [55] extends the study of LSTM-based model on load disaggregation. It generates a novel transient feature APF (Amplitude-Phase-Frequency) based on a Hilbert transform and applies Sequence-to-Sequence LSTM (Seq2Seq LSTM) to identify appliances by using the extracted APF features, which achieves good performance on high-resolution datasets. Two basic properties of NILM problems (non-causality and adaptivity to contextual factors) are considered in [57]. It introduces a Bayesian-optimized framework to select the best configuration of the proposed CoBiLSTM model that could address multi-dimensionality issues when the number of appliances increases. Ref. [61] proposes a causal 1-D convolutional neural network model for low-resolution datasets. It also studies the use of various components of power signal for load disaggregation which demonstrates that using all components (Voltage, Current, Active Power and Reactive Power) could achieve the best performance. This conclusion also proves that more features can lead to better load disaggregation results. Ref. [58] presents a Generative Adversarial Network (GAN) model that could produce more detailed load estimations for specific appliances, where the Generator and Discriminator of the model are appropriately adapted to fit the NILM application. As DNN-based NILM methods require a large training dataset such as ImageNet [25] for visual object recognition and tuning parameters is accompanied by higher computational complexity than other methods, it deserves further careful consideration and analysis.

HMM is another major type of solution to NILM. It is based on Markov chain that describes transition probability among multiple different states. A Markov chain can be used to model the transitions among different states of an appliance. Appliance states of interest are hidden so they cannot be observed directly. HMM aims to infer an appliance's latent states from observed measurements, which makes HMM naturally suited for modeling an appliance. Factorial HMM (FHMM) is an extension of HMM, which includes multiple independent Markov chains of hidden states, and each chain represents one appliance. A household with many appliances can be modeled using FHMM.

Recently, many variants of HMM have been reported to better model appliances in a household. Ref. [26] uses an iterative hidden model to separate individual appliances from the aggregated load iteratively with tuned appliance models. Ref. [27] adopts two FHMM variants, i.e., additive and difference FHMM, to model energy disaggregation problem. It develops a convex formulation of approximation inference instead of exact inference to separate appliances. Load disaggregation is performed using an alternative formulation of additive Factorial Approximate Maximum a Posteriori (AFAMAP) [28]. An Adaptive Density Peak Clustering-Factorial Hidden Markov model (ADP-FHMM) reduces the dependence of prior information and automatically determines the working appliances based on power consumption [29]. A hybrid signature-based iterative disaggregation based on the combination of FHMMs is proposed to improve performance when multiple appliances are operated simultaneously [30]. A segmented integer quadratic constraint programming (SIQCP) is proposed to model home appliances as HMMs and identify appliances [31]. A novel load disaggregation method that uses a super-state hidden Markov model and a sparse Viterbi algorithm variant, SparseHMM, is proposed for real-time load disaggregation [56]. This method preserves dependencies between loads and can perform computationally efficient exact inference which outperforms the previous HMM-based approximate methods. Ref. [32] propose a conditional factorial hidden semi-Markov model (CFHSMM) integrating additional time-related features to disaggregate load using maximum likelihood estimation. To better evaluate performance of different NILM methods, NILM-TK [74] is proposed as an open-source toolkit to enable comparison among different NILM methods in a reproducible way. It gives a complete pipeline from multiple datasets to evaluation metrics. It also includes several benchmark disaggregation algorithms such as combinatorial optimization (CO) and FHMM. These HMM variants lead to further development of HMM-based NILM, but some associated problems have not been completely solved.

At first, the most urgent problem is to disaggregate loads with a high accuracy even with scores of appliances having similar signatures. Previous studies use active or reactive power [31], [33], current, voltage [34] and their combination V-I trajectories [35] as well as current harmonics with Fourier transform [36] to extract signatures. Typically, these signatures are extracted from low-resolution [31], [37] and high-resolution datasets. It is impossible to extract transient signatures from low-resolution datasets thus performance of corresponding methods decreases when appliance number is increasing. With transient signatures such as transient power waveform (TPW), it could identify operating states through transient identification [38], [39], [40], even to separate appliances with similar steady-state signatures [41], [42]. Even though each appliance has its own deterministic transient pattern and there is one-to-one corresponding between its TPW and state transition, it may fail to work and misidentify appliances when multiple appliances simultaneously change states. There is no doubt that studies using high-resolution dataset could achieve better performance due to the diversity of load signatures.

Second, high computational complexity is another problem to be solved and it is essential to achieve real-time disaggregation result so that DR can be better conducted. However, the existing solutions [26], [27], [28], [31], [32] are restricted to the modelling of multiple appliances and solution algorithm, and it is very hard to achieve real-time appliance identification. Event-based NILM that separates appliances by identifying operating state transitions makes it possible to achieve real-time disaggregation. Ref. [21] is an event-based load disaggregation method, but it only works for type I appliances as in [24] that proposes an event matching algorithm for ON/OFF events. Ref. [25] models each appliance with triangle and rectangle signatures, and it classifies appliances using range of power change that may misidentify appliances with similar signatures. An unsupervised event-based method in [26] is designed using additive FHMM, but it does not consider the transient signatures to infer associated appliances. Therefore, even though event-based NILM methods have opened up a whole new area to facilitate the progress of NILM, a novel comprehensive scheme should be proposed to make its adoption easier in practice.

The third problem is accompanied by event-based methods, i.e., big data transmission. Most of literatures consider design of NILM algorithms [26], [27], [28], [29], [30], [32], [33], but not consider their practical applications [31]. Event-based methods usually require higher sampling rate (no less than 1 Hz) and more complex and costly hardware, which limits their attractiveness to researchers. A novel solution should be proposed to promote the development of NILM toward event-based NILM without considering problems brought by high-resolution data and to decrease the costs of hardware.

As aforementioned, a real-time NILM solution with good performance can be implemented in FHMM, which could solve the problems of high computational complexity and low accuracy. An event-driven NILM solution combining FHMM is specifically designed in this paper to tackle such problems. Corresponding solutions to different challenges in data acquisition, event detection, signature extraction, appliance modeling and load disaggregation are specifically proposed based on the challenges that have not solved yet in other literatures as shown in Fig. 1 where the blue dashed boxes refer to the parts that are focused on in this paper and the red boxes include the proposed solutions. Generally, as the change of appliances' working state is always accompanied by the occurrence of an event, it is natural to reformulate the conventional time-driven FHMM to an event-driven FHMM, termed as eFHMM in this paper. In fact, it is not necessary to frequently infer appliances' states in steady periods. Doing that otherwise might be the main reason why most NILM methods based on HMM take a long even exponential time to obtain solution. In the proposed eFHMM framework, the inference times are as limited as the appliances' state-changing times. Furthermore, an event-driven two-stage NILM solution, termed as eFHMM-TS, is proposed to improve the accuracy of load disaggregation, where transient load signatures are adopted to infer which appliance is changing state in the first stage and steady-state load signatures are used to confirm the obtained results by comparing composite total and estimated total load in the second stage.

Using high-resolution data with decisive transient signatures can further improve the performance of NILM. In general, an event-driven NILM solution requires a data sampling rate larger than 1 Hz to guarantee the usability of transient signatures. It puts huge burden on communication network bandwidth to make associated data available to the remote server. To solve this problem, eFHMM-TS can be configured naturally in an edge-cloud framework by dividing the whole NILM task into two parts that are completed in edge and cloud sides to further mitigate the burden of network bandwidth and local computing system in a reliable way.

Besides, the paper solves challenges of integrating transient signatures into mathematical formulas and proposes a novel way to model home appliances. Therefore, this paper provides a systematic and comprehensive solution with focused measures to enhance the performance of NILM in each of its steps. The main contributions of this paper are three-fold.

- •
  First, this paper proposes an event-driven Factorial Hidden Markov Model (eFHMM) to model the appliances in a household. The proposed eFHMM enables online load disaggregation by decreasing inference times to be linear to detected event number thus greatly reducing computational complexity. Compared with time-driven FHMM, eFHMM can be used for high-resolution data with decisive transient information for appliance identification.

- •
  Second, this paper specifically designs a two-stage NILM solution to eFHMM, with the first stage inferring changing states using transient signatures, which

is the first time of its kind to be modelled as mathematical distribution, and the second stage confirming the estimated states with steady-state signatures. The combination of transient and steady-state signatures ensures the uniqueness of transition states thus leading to a higher load disaggregation accuracy.

- •
  Third, this paper proposes the implementation of eFHMM-TS in an edge-cloud framework, where the data-intensive event detection is done at the edge side and the model-intensive load disaggregation is completed at the cloud side. Such an implementation can minimize the requirements on computing capacity, storage space, and available communication bandwidth for resource-poor IoT devices, and along with the inherent protection of data privacy, making it possible to apply NILM in real world.

The rest of this paper is organized as follows. The eFHMM is introduced in Section 2. Section 3 presents the two-stage solution to eFHMM. Section 4 integrates the eFHMM-TS into the edge-cloud framework. Section 5 presents and discusses experimental results. Finally, Section 6 concludes this paper.

Section snippets

Event-driven factorial HMM

Modeling of individual appliances and the combination of appliances in a household is the core of NILM. This section will briefly introduce the motivation of event-driven NILM and the modeling methodology of sparse HMM and eFHMM.

Two-stage solution to event-driven factorial HMM

Reformulation of the basic FHMM into the event-driven FHMM significantly decreases the computational complexity by transferring point-to-point inference to the event-driven inference. The performance of load disaggregation can be further improved by a two-stage NILM solution, termed as eFHMM-TS, which uses the extracted transient and steady-state signatures specifically designed for the proposed eFHMM. Computational complexity of the two-stage NILM solution for eFHMM will be described in detail.

Integrating eFHMM-TS into edge-cloud framework

Typical NILM methods require high computational capacity to deal with complex processes. It is hard to disaggregate the composite load on the end side that has limited computational capacity, or it is a huge resource waste even if it is done on the end side that has ample computational capacity. The centralized model, i.e., using cloud service, faces three significant challenges: 1) Congestion of big data: in centralized model, massive data needs to be transferred from end devices to remote

Case study

Testing on high-resolution datasets with appliance-level ground truth can verify how an NILM approach such as eFHMM-TS takes advantage of event detection and transient information to improve performance. In this section, experiments will be conducted on three datasets with different resolutions including 50 Hz LIFTED, 10 Hz LIFTED, and 5 Hz synD [50] to verify the event-driven two-stage solution method (eFHMM-TS) and compare the eFHMM-TS method with the SIQCP method and different methods in

Conclusions

High-resolution load measurements contain rich information to help disaggregate the load thus leading to a high-performance NILM solution. However, existing methods have not paid enough attention to the transient load signatures and usually suffer from high computational complexity when using high-resolution data. This paper proposes an event-driven NILM solution to solve this problem. It studies the characteristics of individual appliances and their components and finds out the significance of

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.