Dynamic IFFSM Modeling Using IFHMM-Based Bayesian Non-parametric Learning for Energy Disaggregation in Smart Solar Home System

Kalthoum Zaouali[1](✉), Mohamed Lassaad Ammari[2,3], Amine Chouaieb[3,4], and Ridha Bouallegue[4]

[1] Ecole Nationale d'Ingénieurs de Tunis, Innov'Com Laboratory, Université Tunis El Manar, Tunis, Tunisia
zaoualikalthoum@gmail.com
[2] Department of Electrical and Computer Engineering, Laval University, Quebec, Canada
mlammari@gel.ulaval.ca
[3] Chifco Company, Tunis, Tunisia
amine.chouaieb@chifco.com
[4] Sup'Com, Innov'Com Laboratory, Carthage University, Tunis, Tunisia
ridha.bouallegue@supcom.tn

Abstract. Recently, the analysis and recognition of each appliance's energy consumption are fundamental in smart homes and smart buildings systems. Our paper presents a novel Non-Intrusive Load Monitoring (NILM) recognition method based on Bayesian Non-Parametric (BNP) learning approach to solve the problem of energy disaggregation for smart Solar Home System (SHS). Several researches assumed that there is prior information about the household appliances in order to restrict those that do not hold the maximum expectation for inference. Therefore, to deal with the unknown number of electrical appliances in a SHS, we have adapted a dynamic Infinite Factorial Hidden Markov Model (IFHMM) -based Infinite Factorial Finite State Machine (IFFSM) to our NILM times-series modeling as an unsupervised BNP learning method. Our suggested method can grip with few or nappropriate learning data as well as to standardize electrical appliance modeling. Our proposed method outperforms FHMM-based FSM modeling results illustrated in literature.

Keywords: Bayesian Non-Parametric (BNP) · Energy disaggregation · Infinite Factorial Finite State Machine (IFFSM) · Infinite Factorial Hidden Markov Model (IFHMM) · Non-Intrusive Load Monitoring (NILM) · Solar Home System (SHS)

1 Introduction

Nowadays, many countries have widely adopted renewable energy sources to reduce the adverse effects of traditional fossil-fueled electricity generation on the

Q. Li et al. (Eds.): BROADNETS 2019, LNICST 303, pp. 156–175, 2019.
https://doi.org/10.1007/978-3-030-36442-7_11

environment and cut down on their bills. The social and technical developments, as well as economic profits can be ensured by energy saving and Internet of Things (IoT) technologies improvements especially for renewable energy.

In fact, understanding daily activities and energy consumption behaviors can contributes to energy bills reducing, a fine management and an optimal usage of electrical appliances in a Solar Home System (SHS). IoT and smart home technologies are able to guarantee an efficient management of energy and can satisfy the SHS implementation requirements. Furthermore, sensor-based technologies and data science permit the solar field automation for electricity saving and participate for green energy awards. Thanks to IoT technology, smart metering infrastructure, sensors and analytics tools, it is possible to connect solar panels into one system and manage and remote the smart home via mobile or Web applications. Our SHS described in Fig. 1 is extensively considered to lead to energy efficiency goals that require the ability to monitor electric energy appliances via smart meter technology. This smart device promotes aggregated electric energy signal acquisition and ingestion before analysis and data mining.

Currently, there has been a wide interest in the field of Non-Intrusive Load Monitoring (NILM), which involves several methods for monitoring electric appliances and providing appropriate notifications on usage patterns to home-owners. NILM approach gives rise to energy disaggregation by discovering the energy behavior of each household appliance using a single smart meter.

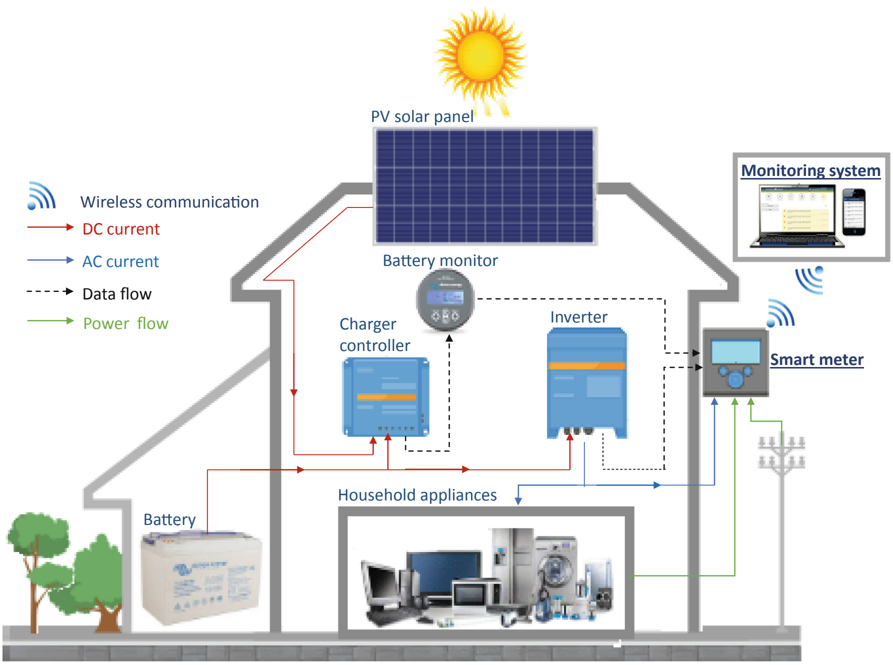

Fig. 1. Solar home system based on smart metering for an efficient green energy management

Since the number and the states of electrical appliances differ from one household to another, several researches assumed that there is prior information about the household appliances in order to restrict those that do not hold the maximum expectation for inference. One of the key areas of the NILM study is the use of unsupervised learning for energy disaggregation which can be an efficient solution to the problem of dependence on a set of training data. Applying unsupervised learning in NILM can reduce the IT complexity of smart homes solution deployment.

In this paper, we suggest a BNP model as an unsupervised learning method to deal with the problem of energy disaggregation with unknown number of electrical appliances in a SHS. We may apply this unsupervised model to solve the problem of a few or an inappropriate learning data as well as to standardize the electrical appliance modeling. To address the NILM challenge for energy disaggregation based on unsupervised learning, we have implemented our suggested Infinite Factorial Finite State Machine (IFFSM) building an Infinite Factorial Hidden Markov Model (IFHMM). This practical method contributes to inform homeowners about each household appliances energy consumption. The results of our proposed BNP model are compared with the Factorial Hidden Markov Model (FHMM) model results illustrated in [32].

The rest of this paper is organized as follows. Section 2 discuss some related work. Section 3 exposes our proposed method focusing on the adaptation of time series modeling by IFFSM to the energy disaggregation problem. The source separation technique using IFHMM-based BNP model, the inference algorithms as well as their different related process and models are detailed in followed subsections. In Sect. 4, we discuss the experimental results of the suggested method. The conclusions and perspective work are given in Sect. 5.

2 Related Work

According to Hart [20], indirect monitoring methods can measure the non-electrical characteristics, from which the power demand of each device is deduced. Device labeling is one of those indirect methods reported in the literature. The signals are detected by a central hub which estimates the energy consumption of each device. Nevertheless, this approach requires the characterization of each device as well as the installation of a central signature detector. This method is expensive and takes a long time to install. For this reason, researchers have been thinking of Wireless Sensor Network (WSN) to identify the power consumed by each device. These sensors make it possible to monitor the human behavior and to control the functioning of the device according to the indicators brought on the temperature, the luminosity, the movements of the inhabitants [28]. This approach also requires the installation of several sensors which reflects the same high cost constraint. In addition, the Conditional Demand Analysis (CDA) technique is emphasized since it does not require the installation of additional meters. However, the CDA requires a large participant base, in which each participant must complete a detailed questionnaire and unusual cases will not be examined. In order to overcome the high cost and intrusive complexity of the WSN and its installation, some emphasis has shifted to NILM methods which automatically deduce the energy signals from each device by a single sensor at a single measurement point.

Since then, several research work has been suggested with improvements to the initial design and different approaches. Taking advantage of NILM approach to infer individual appliance's consumption through a smart meter, several researches have demonstrated its utility for appliance classification, demand response, energy feedback, as well as activity recognition [45,49,54]. Other work that incorporates consumer behavior information has begun to explore learning algorithms that operate on a large number of samples obtained from many homes over extended periods of time. The real-time identification of faulty appliance behavior is a desired technique to save energy by analyzing wasted energy reasons as well as servicing appliance in failure [47]. Much work has been developed for use with NILM systems based on electrical signature detection and classification utilizing different machine learning as shown in Table 1 that summarizes the state of art of some NILM techniques and their characterizations.

As a non-parametric machine learning, the Hidden Markov Model (HMM) was used to describe and identify electrical uses by modeling the combination of stationary stochastic processes that translate the steady-state power level of the different combined waveforms into the total load curve [55]. As an extended version of HMM model, FHMM model was widely used for NILM approach. Kolter et al. [32] proposed a device-level power sub-meter learning model using the prediction maximization technique and then performed rough identification by Gibbs sampling. Such model does not tolerate non-stationary noise, and therefore requires training data to be collected on all devices in the home. For a Conditional FHMM (CFHMM), the state of each hidden variable is dependent on the state of each variable of all other Markov strings in the previous time interval [28].

The Finite State Machine (FSM) is an extension of the HMM model that has a set of input and output vectors, a transition matrix that considers the current input and previous state and returns the future state, as well as a send matrix that treats the input and the current state and returns the output. The FSM has been used as an unsupervised modelling framework for NILM in multiple research [24,39]. Under this model, the future hidden state depends only on a finite number of previous entries.

The source separation problem was closely related to independent component analysis (ICA) based on Discriminative Disaggregation Sparse Coding (DDSC) or Source-separation via Tensor and Matrix Factorizations (STMF) methods. Kolter [30] suggested a NILM based on unsupervised machine learning structured by the DDSC algorithm that builds a base matrix dictionary corresponding to a small subset of device types that contains devices with similar functionality. In order to overcome the limitations of the DDSC learning model which does not take into account the dependence between the devices, Figueiredo [16] presented a new model based on the separation of the global electrical signal sources by the

Table 1. State of art of different NILM based on non-parametric techniques via supervised and unsupervised machine learning

Acronyms	References	NILM features	Network	Sampling frequency	Simpling interval	Event-based	Scal-ability	Disaggregation accuracy	Applications
DDSC	[30]	Source separation	PLC	0.1 Hz	15 min	No	No	+	The vast majority of appliances
STMF	[16,46]	Source separation	PLC	0.1 Hz	15 min	No	No	++	The vast majority of appliances
CO	[20]	Steady states	HAN	0.1 KHz	10 s–1 s	No	No	++	Appliances with power > 50 W
HMM	[29,33]	Steady states	HAN	0.1 KHz	10 s–1 s	No	No	+++	Appliances with power > 50 W
NN	[7,49]	Steady states	HAN	0.1 KHz	10 s–1 s	–	Yes	+++	Appliances with power > 50 W
DL	[26,27,34]	Steady states	HAN	0.1 KHz	10 s–1 s	–	Yes	+++++	Appliances with power > 50 W
FHMM	[31]	Steady states	HAN	0.1 KHz	10 s–1 s	No	No	+++	Appliances with power > 50 W
AFHMM	[31]	Steady states	HAN	0.1 KHz	10 s–1 s	No	No	+++	Appliances with power > 50 W
DTW	[23,40]	Steady states Transient states	HAN	0.1 KHz	10 s–1 s	Yes	No	+++++	Appliances with power > 50 W
CFHMM	[29]	Steady states Transient states	WSN	0.1 KHz–1 KHz	1 s–1 ms	No	No	++++	Appliances with power > 50 W
RF	[35,42]	Steady states Transient states	WSN	1 KHz–20 KHz	1 ms–50 µs	No	Yes	+++++	Appliances with power > 50 W
DT	[19]	Steady states Transient states	WSN	1 KHz–20 KHz	1 ms–50 µs	No	Yes	+++++	Appliances with power > 50 W
CDM	[4]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	++++++	Appliances with power > 50 W
SVM	[10]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	+++++++	Appliances with power > 50 W
NFL	[11,12]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	+++++++	Appliances with power > 50 W
MLC	[50]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	+++++++	Appliances with power > 50 W
GSP	[21,56]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	+++++++	Appliances with power > 50 W
EMI	[1]		WSN	100 KHz	10 ms	Yes	No	++++++++	Some appliance models
GMM	[3]	Steady states Transient states	WSN	1 KHz–100 KHz	1 ms–10 ms	Yes	Yes	+++++++	Appliances with power > 50 W

factorization in non-negative tensors. This method has generated great interest in the concept of blind separation of sources. However, these tow methods assume a fixed and known number of latent sources.

Many other unsupervised machine learning have been introduced in recent research work to address NILM requirements [22,41]. NILM based on Bayesian Non-Parametrics (BNP) dynamical system is introduced to solve the problem of inferring the operational state of individual electrical appliances though aggregate measurements. An unbiased algorithm for neural variational identification and filtering was investigated in [36]. BNP techniques are also used in different fields, such as driving-styles analysis and recognition for smart transportation and vehicle calibration [52], intelligent dynamic spectrum access notification in cognitive radio environments [1,2,17,51,53], detection and estimation of sparse acoustic channels [9,25], video and image segmentation, reconstruction and recognition [13,15,43,44], data clustering and classification [6,14,37], etc. Numerous surveys have been carried out to describe the different learning models and the extractions features methods for NILM and home monitoring systems [3,8]. The advantages of the BNP models are that they can take a complex problem and create a model describing the appropriate solution. Due to its non-parametric nature with infinite memory chains length bonus, BNP models are able to deal with an unbounded number of states. Our proposed NILM method focuses on BNP models which are used most often in source separation problem.

3 Adaptation of Time Series Modeling by IFFSM to the Energy Disaggregation Problem

The approach adopted in our energy disaggregation problem is summarized in Fig. 2 which introduces the NILM approach based on an automatic unsupervised learning. In particular, the sources separation method allowing the processing of data from the smart meter, as well as the use of BNP model in conjunction with the adopted automatic learning method. Our proposed approach for energy disaggregation problem is described in detail in the following sections and subsections.

3.1 Source Separation Using IFHMM-based BNP Model

In our study, the FSM relies on a finite memory denoted $\mathbf{L}$ and a finite set denoted $\mathcal{X}$. Each new entry x_t is able to modify the FSM state as well as the observed output. The future state and the output depend only on the current state and the input. The FSM can be modeled as single HMM where the vector containing the last n inputs can characterize each state. Making an inference on this model has a $\mathcal{O}(T|\mathbf{S}|^{2L})$ complexity, but it can be reduced to $\mathcal{O}(T|\mathbf{S}|^{L+1})$ by exploiting the suggested IFFSM machine learning based on IFHMM model that requires approximation inference methods to avoid dependence on memory length L.

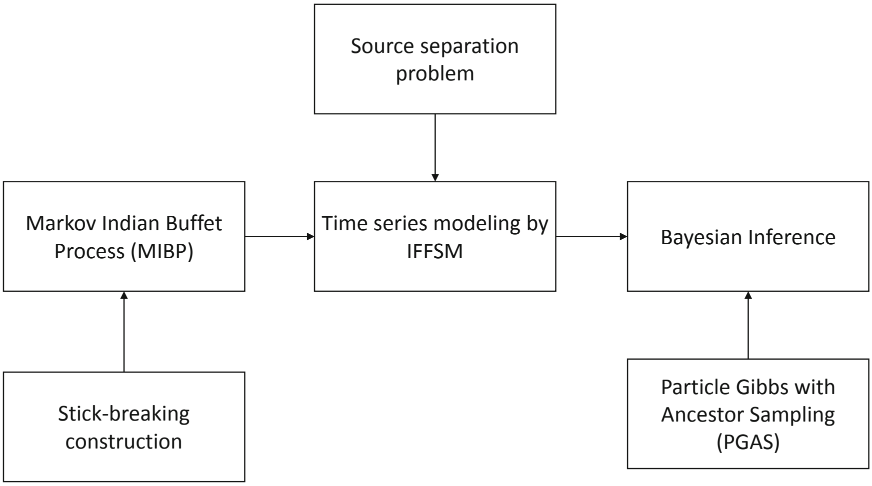

Fig. 2. Adopted NILM approach based on source separation

In the proposed IFHMM model, we assume binary input variables $x_{tn} = 0$ for $t \leq 0$. This model can be generalized to obtain additional properties concerning the fundamental structure of the model. We consider two methods for generalization that our inference algorithm can handle with little or no modifications. First, we assume that the input vectors x_{tk} belong to a finite set X, so that we can give $|X|^L$ possible states in each parallel Markov chain. However, we can also consider that the set X is a countable infinite set, which implies that the input vectors do not necessarily contain discrete values. The resulting model is no longer an FSM model, but an infinite factorial model in which the hidden variables affect present observations, as well as future observations [18].

To solve the energy disaggregation problem using the suggested unsupervised learning method, we assume that the observation y_t represents a sample of the general smart meter signal, which depends on the signals generated by the different active electrical appliances. We suppose that there is theoretically an infinite number of sources that display the observed sequence $\{\mathbf{y}_t\}_{t=i}^{T}$, where T is the number of time steps. Every source is modeled by a dynamic system model wherein the input symbol correspond to the $n'th$ source at time t is denoted by $x_{tn} \in \mathcal{X}$ setting the first-order Markov chain, where $\mathcal{X}$ may be a discrete or continuous state space.

Each element of the auxiliary binary matrix $\mathbf{S}$, denoted by $s_{tn} \in \{0, 1\}$, reveals the source state (active or inactive) at time instant t, and can be expressed as:

$$x_{tn}|s_{tn} \sim \begin{cases} \delta_0(x_{tn}) & \text{if } s_{tn} = 0 \ ; \\ \mathcal{U}(\mathcal{X}) & \text{if } s_{tn} = 1 \end{cases} \tag{1}$$

where $delta_0(.)$ designates the Dirac measure at $t = 0$, and $\mathcal{U}(\mathcal{X})$ is the uniform law on the set $\mathcal{X}$. The entries x_{tn} are independent and identically distributed conditionally on the auxiliary variables s_{tn}.

Therefore, the proposed dynamic model depends basically on this conditional probability $p(x_{tn}|s_{tn}, x_{(t-1)n}) >$ with $s_{tn} = 0$ for $T \leq 0$. This transition model performs the active states x_{tn} evolving over-time as dynamics of the global smart meter signal. During the multi-channel propagation of the individual signals, the electric waves can be reflected and that may cause reception delays. Considering this memory effect, we can note that the hidden state x_{tn} affects the observations y_t as well as the last future observations $y_{t+1} \dots y_{t+L-1}$ if $s_{tn} \neq 0$, where L denotes the last states of overall Markov chains.

The expression of y_t likelihood is as follows:

$$p(y_t|\mathcal{X}, \mathbf{S}) = p(y_t|\{x_{tn}, s_{tn}, x_{(t-1)n}, s_{(t-1)n}, \cdots x_{(t-L+1)n}, s_{(t-L+1)n}\}_{n=1}^{N}) \quad (2)$$

The considered IFFSM graphic model is shown in Fig. 3.

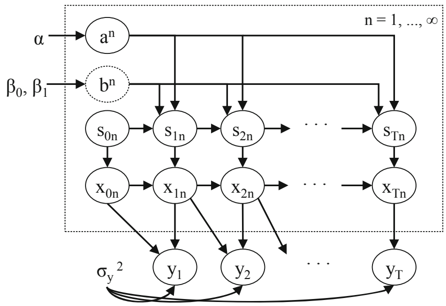

Fig. 3. IFFSM modeling with memory length L = 2.

Markov Indian Buffet Process. Among the BNP models, the Markov Indian Buffet Process (MIBP) is the fundamental building block of the Infinite Factorial Hidden Markov Model (IFHMM) as an IFFSM. To process an infinite number of sources, the binary matrix $\mathbf{S}$ is distributed as MlBP priors with distribution parameters α, β_0, β_1 [48] as:

$$\mathbf{S} \sim lBP(\alpha, \beta_0, \beta_1) \quad (3)$$

This MlBP prior distribution ensures that, for any finite number of time instant T, only a finite number of Markov chain N become active, while the rest of them remain in the zero state without affecting the observations.

We consider the total consumption of the household, the energy disaggregation concept is to predict the number of the active electrical appliances as well as their corresponding consumption. We treat a 24-hour segment for 6 different houses. Each electrical appliance holds 4 different states: one inactive state and 3 active states pointing different values of energy consumption.

A symmetric Dirichlet a prior distribution is placed on the vectors of the transition probabilities as a_i^n $Dirichlet(1)$, where $a_{ij}^n = p(x_{tn} = j | s_{tn} = 1, x_{(t-1)n} = i, s_{(t-1)n}$. The energy consumption of the electrical appliance n at the time instant t is null when $x_{tn} = 0$ ($s_{tn} = 0$), and the total energy consumption is given by:

$$\mathbf{y}_t = \sum_{n=1}^{N} P_{x_{tn}}^n + \varepsilon_t \tag{4}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma_y^2 I)$ represents additive Gaussian noise with the hyper-parameter σ_y^2 is the noise variance and I is the identity matrix.

To get a maximum accuracy, we associate to each electrical appliance an estimated chain. Thus, the accuracy of each estimated appliance consumption evaluating the performance of our method is determined by:

$$accuracy = 1 - \frac{\sum_{t=1}^{T} \sum_{n=1}^{N} |x_t^{(n)} - \hat{x}_t^{(n)}|}{2 \sum_{t=1}^{T} \sum_{n=1}^{N} x_t^{(n)}} \tag{5}$$

where $\hat{x}_t^{(n)} = P_{x_{tn}}^n$ is the estimated consumption of each n electrical appliance at time instant t. In the case where the inferred number of electrical appliances is less than the available electrical appliances number, the additional chains will be grouped in a category of unknown electrical appliances $x_t^{(unknown)}$.

Stick-Breaking Construction. The Stick-Breaking construction can be adapted to the MlBP to efficiency improve inference algorithms by introducing the transition probability from inactive state to active one denoted by a^n, as well as the self-transition probability of the *n-th* Markov chain active state denoted by b^n. These two hidden variables a^n and b^n are described as follows:

$$a^n = p(s_{tn} = 0 | s_{(t-1)n} = 0) \tag{6}$$

$$b^n = p(s_{tn} = 0 | s_{(t-1)n} = 1) \tag{7}$$

The matrix of the transition probabilities of the *n-th* Markov chain can be written as follows:

$$\mathbf{A}^n = \begin{pmatrix} 1 - a^n & a^n \\ 1 - b^n & b^n \end{pmatrix} \tag{8}$$

The distribution on the variables a^n is given by:

$$a^1 \sim Beta(\alpha, 1) \tag{9}$$

and the distribution probability on the variables a^n can be given by:

$$p(a^{(n)} | a^{(n-1)}) \propto (a^{(n-1)})^{(-\alpha)} (a^{(n)})^{(\alpha-1)} I(0 \leq a^{(n)} \leq a^{(n-1)}) \tag{10}$$

with $I(.)$ characterizes the indicator function which is worth 1 if its argument is true and 0 if its argument is false, α is the concentration parameter that controls the number of active Markov chains. Independently of n, the MIBP prior over variables b^n, distributed according to a Beta process, is defined by:

$$b^n \sim Beta(\beta_0, \beta_1) \tag{11}$$

3.2 Inference Algorithms

Each Bayesian model looks for a posterior inference calculated according to the posterior distribution of hidden variables. Several time series BNP models take a proximate inference algorithm if the posterior distribution cannot be obtained directly. Such inference algorithm can be based on Markov Chain Monte Carlo (MCMC) methods. In particular, we have used a Gibbs Sampling inference algorithm that combines MCMC and Sequential Monte Carlo (SMC) standard tools. The suggested IFHMM model is incorporated with the Gibbs sampling-based inference algorithm to treat the parallel chains number and transition variables. The Inference algorithm based on MIBP prior starts with introducing new inactive chains N_{new} utilizing a slice sampling method and an auxiliary slice variable to provide a finite factorial model. Thus, the number of parallel chains is increased from N_+ to $N^\ddagger = N_+ + N_{news}$ and consequently the N_+ chains number cannot be updated. The first sampled auxiliary slice variable $\mathcal{V}$ is distributed as:

$$\mathcal{V}|S, \{a^n\} \sim Uniform(0, a_{minimale}) \tag{12}$$

where $a_{minimal} = \min_{n:\exists t, s_{tn} \neq 0} a^n$ can be a Beta distribution.

The next new variables a^n make the following sampling iterations until $a^n < \mathcal{V}$:

$$p(a^n|a^n) \propto \exp\left(\beta_0 \sum_{t-1}^{T} \frac{1}{t}(1-a^n)^t\right) \times (a^n)^{-\beta_0 - 1}(1-a^n)^T, (0 \leq a^n \leq a^{n-1}) \tag{13}$$

Moreover, the next step of the inference algorithm is focused on sampling the states s_{tn} and the input symbols x_{tn} of all chains of the IFFSM model. This compacted sampling eliminates the chains that remain inactive throughout the observation time, which allows the update of N_+ chains number. We put forward the use of Particle Gibbs algorithm for inference in non-Markovian latent variable models.

Despite the efficiency and the simplicity of the Gibbs sampling model of each element $x_{tn}|s_{tn}$, this technique cannot provide good mixing properties owning to the high coupling of successive steps. The Gibbs sampling model can utilize the Forward Filtering Backward Sampling (FFBS) method to treat the successive sampling of chains according a complexity of $\mathcal{O}(TN^\ddagger|\mathcal{X}|^{L+1})$ to our suggested IFFSM model while $L > 1$. However, the exponential dependence on L can prevent convergence of the FFBS calculation. Thus, to deal with this problem, a Particle Gibbs with Ancestor Sampling (PGAS) algorithm is combined with the inference algorithm to jointly sample the matrices $\mathcal{X}$ and $\mathbf{S}$. If P particles are used for the PGAS kernel, the complexity of our adopted algorithm is about $\mathcal{O}(PTN^{\ddagger}L^2)$.

Finally, inference algorithm will sample the global variables joining the transition and the emission probabilities depending on their posterior distribution so as to evaluate the likelihood $p(y_t|\mathcal{X}, \mathbf{S})$.

This final step of the inference algorithm is focused on the sampling of the global variables of the model from their full conditional distributions under the Stick-Breaking construction is expressed as follows:

$$p(a^n|\mathbf{S}) = Beta(1 + tr_{00}^k, tr_{01}^k) \tag{14}$$

with tr the number of transitions between states in the N column of $\mathbf{S}$.

With regard to the transition probabilities of the active state to the inactive state b^n, we have:

$$p(b^n|\mathbf{S}) = Beta(\beta_1 + tr_{00}^k, \beta_2 + tr_{01}^n) \tag{15}$$

We also define the extended matrix $\mathcal{X}^{extended}$ of size $TxLN_+$ with:

$$\mathcal{X}^{(n)} = \begin{pmatrix} x_{1n} & 0 & \cdots & 0 \\ x_{2n} & x_{1n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{Tn} & x_{(T-1)n} & \cdots & x_{(T-L+1)n} \end{pmatrix} \tag{16}$$

Particle Gibbs with Ancestor Sampling. Compared to the FFBS method, the PGAS algorithm offers several benefits in addition to its non-exponential complexity. Notably, it can be applied independently of $\mathbf{X}$ proprieties as finite or infinite set. Furthermore, its inference properties are better than those of FFBS algorithm which cannot contribute to the $N^{\ddagger} - 1$ propriety of the observation chains. For energy disaggregation problems, the FFBS is often limited to later local modes in which several Markov chains correspond to a single hidden source. However, for each simultaneous instant t, the PGAS algorithm can treat and sample all chains in parallel. This characterization does not exist in FFBS and FHMM models [5]. The integration of such algorithm involves the elimination of certain a^n and b^n variable as well as the updated chains N_+.

In order to better adapt the PGAS algorithm to our energy disaggregation problem based on source separation approach, we can refer to [38] describing the appropriate steps as well as the theoretical justification of the PGAS algorithm.

We assume that the suggested PGAS model is composed of a set of P particles which represent the hidden states $\{x_{tn}\}_{n=1}^{N^{\ddagger}}$ at an instant t. Let defining the i-th particle state at time t by x_t^i of length $N^{\ddagger}$ as well as its ancestor indexes $a_t^i \in 1, \ldots, P$. Given $\mathbf{x}_{1:t}^i$ as the ancestral trajectory of the particle $\mathbf{x}_t^i$, the recursive form of the particle trajectory is described as follows:

$$\mathbf{x}_{1:t}^{i} = \left(\mathbf{x}_{1:t-1}^{a_t^i}, \mathbf{x}_t^i\right) \tag{17}$$

A fixed reference particle denotes by $\mathbf{x}_t^*$ is generated by previous iteration outputs is required for the PGAS algorithm extension in order to introduce novel inactive chains. Thus, the corresponding ancestor indexes a_t^P of the fixed particle $\mathbf{x}_t^P$ are variables and randomly picked. Hence, a distribution form $q_t\left(\mathbf{x}_t|\mathbf{x}_{1:t-1}^{a_t}\right)$ is required to specify the propagate way of the particles, so that, we assume the following expression:

$$q_t\left(\mathbf{x}_t|\mathbf{x}_{1:t-1}^{a_t}\right) = p\left(\mathbf{x}_t|\mathbf{x}_{t-1}^{a_t}\right) \tag{18}$$

$$= \prod_{n=1}^{N^\ddagger} p\left(x_{tn}|s_{tn}, x_{(t-1)n}^{a_t}\right) p\left(s_{tn}|s_{(t-1)n}^{a_t}\right) \tag{19}$$

For each instant t, the PGAS inference algorithm samples the particles at instant $t-1$ taking into account their weight specification w_{t-1}^i and their distributed propagation $q_t\left(\mathbf{x}_t|\mathbf{x}_{t-1}^{a_t}\right)$.

The following equations develop the weights expression as well as the ancestor weights:

$$W_t(\mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)}{p\left(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}\right) q_t\left(\mathbf{x}_t|\mathbf{x}_{1:t-1}\right)} \tag{20}$$

$$\propto \frac{p\left(\mathbf{y}_{1:t}|\mathbf{x}_{1:t}\right) p\left(\mathbf{x}_{1:t}\right)}{p\left(\mathbf{y}_{1:t-1}|\mathbf{x}_{1:t-1}\right) p\left(\mathbf{x}_{1:t-1}\right) p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)} \tag{21}$$

$$\propto p\left(\mathbf{y}_t|\mathbf{x}_{t-L+1:t}\right) \tag{22}$$

with $\mathbf{y}_{\tau_1:\tau_2}$ denotes the set of observations $\{\mathbf{y}_t\}_{t=\tau_1}^{\tau_2}$. The $W_t(\mathbf{x}_{1:t})$ equation involves that obtained weights w_{t-1}^i depending essentially on the likelihood evaluation at time t. The estimated ancestor weights $\tilde{w}_{t-1|T}^i$ of the reference particle are given by:

$$\tilde{w}_{t-1|T}^i = w_{t-1}^i \frac{p\left(\mathbf{x}_{1:t-1}^i, \mathbf{x}_{t:T}^*|\mathbf{y}_{1:T}\right)}{p\left(\mathbf{x}_{1:t}^i|\mathbf{y}_{1:t-1}\right)} \tag{23}$$

$$\propto w_{t-1}^i \frac{p\left(\mathbf{y}_{1:T}|\mathbf{x}_{1:t}^i, \mathbf{x}_{t:T}^*\right) p\left(\mathbf{x}_{1:t-1}^i, \mathbf{x}_{t:T}^*\right)}{p\left(\mathbf{y}_{1:t-1}|\mathbf{x}_{1:t-1}^i\right) p\left(\mathbf{x}_{1:t-1}^i\right)} \tag{24}$$

$$\propto w_{t-1}^i p\left(\mathbf{x}_t^*|\mathbf{x}_{t-1}^i\right) \prod_{T=t}^{t+L-2} p\left(\mathbf{y}_T|\mathbf{x}_{1:t}^i, \mathbf{x}_{t:T}^*\right) \tag{25}$$

Calculating the weights $\tilde{w}_{t-1|T}^i$ for $L > 1$ and $i = l, \ldots, P$ increasing the model complexity by $\mathcal{O}(PN^\ddagger L^2)$.

To deal with the last equation, we have to take only the IFHMM factors which depending on the particles index i, without memory ($L = 1$) and with a transition probability $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ factored in parallel through the IFHMM model. The weights expression can be written as follows:

$$\tilde{w}_{t-1|T}^i \propto w_{t-1}^i p\left(\mathbf{x}_t^*|\mathbf{x}_{t-1}^i\right) \tag{26}$$

4 Simulations and Results

We have implemented the proposed IFFSM based on IFHMM model with the following parameters to obtain the desired results:

- Length of the observation sequence T.
- Number of possible states $Q = 4$
- Iteration number $i = 10000$
- Hyper-parameters of our IFFSM model:
  - $\lambda_1 = 15$ and $\lambda_2 = 10$ are the parameters of the Gaussian distribution on consumption, i.e. $P_q^n \sim \mathcal{N}(15.10)$
  - $\gamma_1 = 0$ and $\gamma_2 = 0.5$ are the hyper-parameters of the Gaussian distribution on the white noise denoted as $\varepsilon_t \sim \mathcal{N}(0, 0.5)$.
  - $\beta_0 = 1$ is the hyper-parameter of the distribution on the self-transition of the inactive state $a_i^n \sim Dirichlet(1)$.
  - $\beta_1 = 0.1$ and $\beta_2 = 2$ are the Beta distribution hyper-parameters involving the transition from an active state to an inactive state b^n, i.e. $b^n \sim Beta(0.1, 2)$.
  - MlBP distribution on the activation binary matrix $\mathbf{S}$ is defined as $\mathbf{S} \sim MIBP(l, 0.1, 2)$.

The visualization of the general load curve which represents the total consumption of the electrical appliances is given by the Fig. 4.

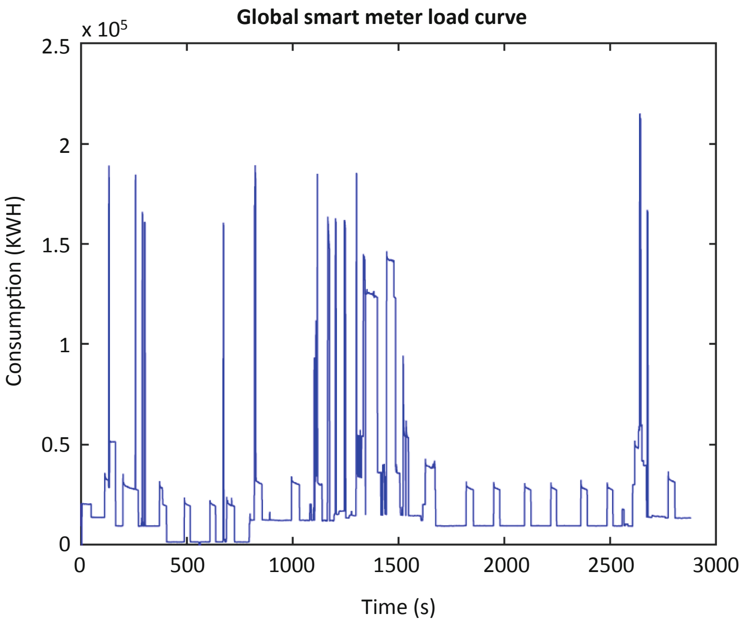

Fig. 4. The global smart meter load curve

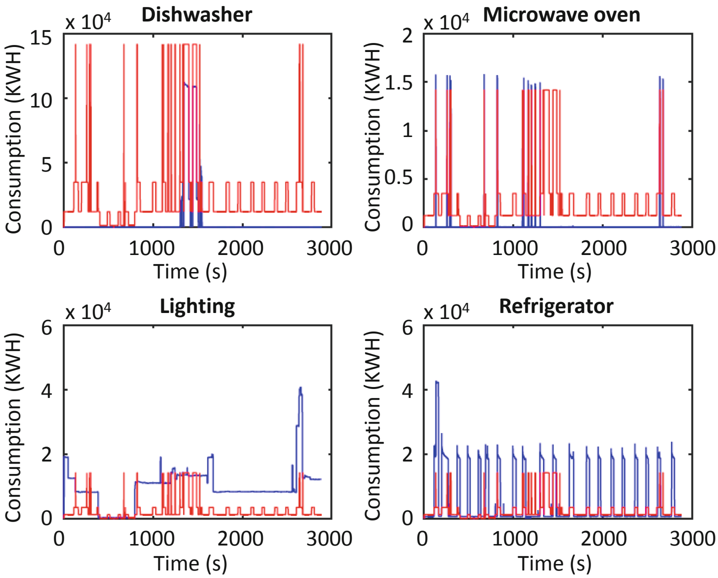

Fig. 5. The estimated load curve of each electrical appliance

After implementing our suggested IFFSM model adapted to the energy disaggregation problem, we obtain the load curves for a specific electrical appliance. Each load curve represents the evolution of the consumption of each electrical appliance over time. A visualization on Matlab allows to obtain the results presented by the Fig. 5.

5 Performance Evaluation and Discussion

To evaluate the performance of our model, we apply the formula of the equation:

$$accuracy = 1 - \frac{\sum_{t=1}^{T}\sum_{n=1}^{N}|x_t^n - \hat{x}_t^n|}{2\sum_{t=1}^{T}\sum_{n=1}^{N}x_t^n} \tag{27}$$

This formula allows to test the performance of our IFFSM model to estimate the individual energy consumption of each appliance installed in smart homes. The performance of our model is calculated for the 4 appliances that consume the most energy. Table 2 shows the different performances of our applied model.

We have to test the performance of our IFFSM model and compare the accuracy with an existing Markov model. Thus, we have used the Reference Energy Disaggregation Dataset (REDD) in order to be able to evaluate our model. We have compared the obtained results with the FHMM model results implemented in [32]. Table 3 visualizes the performance of each model for 6 smart homes described by the REDD database.

Table 2. Performances evaluation: the accuracy of the estimated energy consumption for 4 household appliances situated in a smart home environment

	Smart home electrical appliances			
	Dishwasher	Microwave oven	Refrigerator	Lighting
Accuracy (%)	66.3	67.2	70.1	61.4
Average	66.75			

Table 3. Comparison of our IFFSM model performances to the FHMM model

	Accuracy (%)	
	Proposed IFFSM model (IFHMM)	Compared model (FHMM)
Smart home 1	66.7	71.5
Smart home 2	66.3	59.6
Smart home 3	66.2	59.6
Smart home 4	66.7	69.0
Smart home 5	66.4	62.9
Smart home 6	66.7	54.4
Average	66.5	64.5

We demonstrate that the proposed model slightly exceeds the FHMM model in the energy disaggregation problem. Thus, the accuracy of the model does not change substantially in different houses, this is because the model is completely unsupervised and does not require learning data but can determine consumption from prior observations and distributions.

Thus, the proposed model is more practical because it is not realistic to believe that learning data can be obtained from all the homes where we will implement our solution and we should not expect to have a model for each electrical appliance installed in each home.

6 Conclusion

In this paper we have proposed a NILM approach based on an unsupervised learning to solve the energy disaggregation problem. In fact, we have adapted an IFFSM- based time series modeling to the source separation problem with an unknown number of sources. We have developed a BNP model which aims to disaggregate the general load curve at the end of a smart meter. We have implemented the suggested IFFSM model building an IFHMM model and we have obtained satisfactory results compared to the FHMM model which requires learning data and test data. The proposed model builds on the MIBP to recognize an infinite number of hidden Markov chains with either discrete or continuous states. We have put forward a PGAS algorithm for posterior inference to deal with the FFBS complexity. We have successfully implemented our proposed model and we have visualized the experimental results using the REDD data-set.

However, our proposed approach is inappropriate if we want to use it on a large number of SHS deployed in smart cities or smart grids. Being a blind method, it is difficult to recognize each estimated chain of a specific device with the lack of sufficient individual information from each appliance for every SHS. Being a blind method, it is difficult to recognize each estimated chain of a specific device for lack of sufficient individual information from each appliance for every SHS. To deal will this limitation, we can improve our suggested IFHMM model by dividing the chains between the houses in a hierarchical way while calculating each activation function individually for every house and inferring the common features between different houses. In further research, we can improve our inference algorithm scalability to contribute for both a larger number of appliances and larger observation sequences. Furthermore, we can transform our model from a time-invariant load model using an offline static database to an on-line unsupervised model for autonomous household database construction in order to recognize the real-time behavior of the power consumption [22].

Abbreviations

The following abbreviations are used in this manuscript:

AFHMM	Additive Factorial Hidden Markov Models
ANN	Artificial Neural Network
BNP	Bayesian Non-Parametric
CDA	Conditional Demand Analysis
CFHMM	Conditional Factorial Hidden Markov Models
CO	Combinatorial Optimization
DDSC	Discriminative Disaggregation Sparse Coding
DL	Deep Learning
DT	Decision Tree
DTW	Dynamic Time Warping
CDM	Committee Decision Mechanism
EMI	Electromagnetic Interference
FFBS	Forward Filtering Backward Sampling
FHMM	Factorial Hidden Markov Models
FSM	Finite State Machine
GMM	Gaussian Mixture Model
GSP	Graph Signal Processing techniques
HMM	Hidden Markov Model
IFFSM	Infinite Factorial Finite State Machine
IFHMM	Infinite Factorial Hidden Markov Model
IoT	Internet of Things
KNN	k-nearest neighbor
MCMC	Markov Chain Monte Carlo
MIBP	Markov Indian Buffet Process
MLC	Multi-Label Classification
NFL	Neuro-Fuzzy Logic algorithm

NILM	Non Intrusive Load Monitoring
PGAS	Particle Gibbs with Ancestor Sampling
REDD	Reference Energy Disaggregation Dataset
RF	Random Forest
SHS	Solar Home System
SMC	Sequential Monte Carlo
STMF	Source-separation via Tensor and Matrix Factorizations
SVM	Support Vector Machine
WSN	Wireless Sensor Network

References

1. Aboulian, A., Donnal, J.S., Leeb, S.B.: Autonomous calibration of non-contact power monitors. IEEE Sens. J. **18**(13), 5376–5385 (2018)
2. Ahmed, M.E., Kim, D.I., Kim, J.Y., Shin, Y.: Energy-arrival-aware detection threshold in wireless-powered cognitive radio networks. IEEE Trans. Veh. Technol. **66**(10), 9201–9213 (2017)
3. Alcalá, J., Ureña, J., Hernández, Á., Gualda, D.: Event-based energy disaggregation algorithm for activity monitoring from a single-point sensor. IEEE Trans. Instrum. Meas. **66**(10), 2615–2626 (2017)
4. Andrean, V., Zhao, X., Teshome, D.F., Huang, T., Lian, K.: A hybrid method of cascade-filtering and committee decision mechanism for non-intrusive load monitoring. IEEE Access **6**, 41212–41223 (2018)
5. Aueb, M.T.R., Yau, C.: Hamming ball auxiliary sampling for factorial hidden Markov models. In: Advances in Neural Information Processing Systems, pp. 2960–2968 (2014)
6. Campbell, T., Kulis, B., How, J.: Dynamic clustering algorithms via small-variance analysis of Markov chain mixture models. IEEE Trans. Pattern Anal. Mach. Intell. **41**(6), 1338–1352 (2019)
7. Chang, H., Lee, M., Lee, W., Chien, C., Chen, N.: Feature extraction-based hellinger distance algorithm for nonintrusive aging load identification in residential buildings. IEEE Trans. Ind. Appl. **52**(3), 2031–2039 (2016)
8. Dan, W., Li, H.X., Ce, Y.S.: Review of non-intrusive load appliance monitoring. In: 2018 Proceedings of the IEEE 3rd Advanced Information Technology Electronic and Automation Control Conference (IAEAC), pp. 18–23, October 2018
9. Diamantis, K., Dermitzakis, A., Hopgood, J.R., Sboros, V.: Super-resolved ultrasound echo spectra with simultaneous localization using parametric statistical estimation. IEEE Access **6**, 14188–14203 (2018)
10. Duarte, C., Delmar, P., Goossen, K.W., Barner, K., Gomez-Luna, E.: Non-intrusive load monitoring based on switching voltage transients and wavelet transforms. In: 2012 Future of Instrumentation International Workshop (FIIW), pp. 1–4. IEEE (2012)
11. Ducange, P., Marcelloni, F., Antonelli, M.: A novel approach based on finite-state machines with fuzzy transitions for nonintrusive home appliance monitoring. IEEE Trans. Ind. Inform. **10**(2), 1185–1197 (2014)
12. Egarter, D., Bhuvana, V.P., Elmenreich, W.: PALDi: online load disaggregation via particle filtering. IEEE Trans. Instrum. Meas. **64**(2), 467–477 (2015)

13. Erdil, E., Ghani, M.U., Rada, L., Argunsah, A.O., Unay, D., Tasdizen, T., Cetin, M.: Nonparametric joint shape and feature priors for image segmentation. IEEE Trans. Image Process. **26**(11), 5312–5323 (2017)
14. Fan, W., Bouguila, N., Du, J., Liu, X.: Axially symmetric data clustering through dirichlet process mixture models of Watson distributions. IEEE Trans. Neural Netw. Learn. Syst. **30**(6), 1683–1694 (2019)
15. Fan, W., Sallay, H., Bouguila, N.: Online learning of hierarchical Pitman-Yor process mixture of generalized dirichlet distributions with feature selection. IEEE Trans. Neural Netw. Learn. Syst. **28**(9), 2048–2061 (2017)
16. Figueiredo, M., Ribeiro, B., de Almeida, A.: Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home. IEEE Trans. Instrum. Meas. **63**(2), 364–373 (2014)
17. Ford, G., et al.: Wireless network traffic disaggregation using Bayesian nonparametric techniques. In: 52nd Annual Conference on Information Sciences and Systems, CISS 2018, Princeton, NJ, USA, 21–23 March 2018, pp. 1–6. IEEE (2018)
18. Gael, J.V., Teh, Y.W., Ghahramani, Z.: The infinite factorial hidden Markov model. In: Advances in Neural Information Processing Systems, pp. 1697–1704 (2009)
19. Gillis, J.M., Alshareef, S.M., Morsi, W.G.: Nonintrusive load monitoring using wavelet design and machine learning. IEEE Trans. Smart Grid **7**(1), 320–328 (2016)
20. Hart, G.W.: Nonintrusive appliance load monitoring. Proc. IEEE **80**(12), 1870–1891 (1992)
21. He, K., Stankovic, L., Liao, J., Stankovic, V.: Non-intrusive load disaggregation using graph signal processing. IEEE Trans. Smart Grid **PP**(99), 1 (2017)
22. Hosseini, S.S., Kelouwani, S., Agbossou, K., Cardenas, A., Henao, N.: Adaptive on-line unsupervised appliance modeling for autonomous household database construction. Electr. Power Energy Syst. J. **112**, 156–168 (2019)
23. Iwayemi, A., Zhou, C.: SARAA: semi-supervised learning for automated residential appliance annotation. IEEE Trans. Smart Grid **8**(2), 779–786 (2017)
24. Jia, R., Gao, Y., Spanos, C.J.: A fully unsupervised non-intrusive load monitoring framework. In: Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 872–878, November 2015
25. Jing, L., He, C., Huang, J., Ding, Z.: Joint channel estimation and detection using Markov chain Monte Carlo method over sparse underwater acoustic channels. IET Commun. **11**(11), 1789–1796 (2017)
26. Kaselimi, M., Protopapadakis, E., Voulodimos, A., Doulamis, N., Doulamis, A.: Multi-channel recurrent convolutional neural networks for energy disaggregation. IEEE Access **7**, 81047–81056 (2019)
27. Khodayar, M., Mohammadi, S., Khodayar, M.E., Wang, J., Liu, G.: Convolutional graph autoencoder: a generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. IEEE Trans. Sustain. Energy 1 (2019)
28. Kim, H., Marwah, M., Arlitt, M., Lyon, G., Han, J.: Unsupervised disaggregation of low frequency power measurements. In: Proceedings of the 2011 SIAM International Conference on Data Mining, pp. 747–758. SIAM (2011)
29. Kim, M.S., Kim, S.R., Kim, J., Yoo, Y.: Design and implementation of MAC protocol for SmartGrid HAN environment. In: Proceedings of the IEEE 11th International Conference on Computer and Information Technology, pp. 212–217, August 2011
30. Kolter, J.Z., Batra, S., Ng, A.Y.: Energy disaggregation via discriminative sparse coding. In: Advances in Neural Information Processing Systems, pp. 1153–1161 (2010)

31. Kolter, J.Z., Jaakkola, T.: Approximate inference in additive factorial HMMs with application to energy disaggregation. In: Artificial Intelligence and Statistics, pp. 1472–1482 (2012)
32. Kolter, J.Z., Johnson, M.J.: REDD: a public data set for energy disaggregation research. In: Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, vol. 25, pp. 59–62 (2011)
33. Kong, W., Dong, Z.Y., Ma, J., Hill, D.J., Zhao, J., Luo, F.: An extensible approach for non-intrusive load disaggregation with smart meter data. IEEE Trans. Smart Grid **9**(4), 3362–3372 (2018)
34. Kong, W., Dong, Z.Y., Wang, B., Zhao, J., Huang, J.: A practical solution for non-intrusive type II load monitoring based on deep learning and post-processing. IEEE Trans. Smart Grid 1 (2019)
35. Kramer, O.: Non-intrusive appliance load monitoring with bagging classifiers (2015)
36. Lange, H., Bergés, M., Kolter, Z.: Neural variational identification and filtering for stochastic non-linear dynamical systems with application to non-intrusive load monitoring. In: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ICASSP 2019, pp. 8340–8344. IEEE (2019)
37. Lhéritier, A., Cazals, F.: A sequential non-parametric multivariate two-sample test. IEEE Trans. Inf. Theory **64**(5), 3361–3370 (2018)
38. Lindsten, F., Jordan, M.I., Schön, T.B.: Particle Gibbs with ancestor sampling. J. Mach. Learn. Res. **15**(1), 2145–2184 (2014)
39. Liu, B., Yu, Y., Luan, W., Zeng, B.: An unsupervised electrical appliance modeling framework for non-intrusive load monitoring. In: Proceedings of the IEEE Power Energy Society General Meeting, pp. 1–5, July 2017
40. Liu, B., Luan, W., Yu, Y.: Dynamic time warping based non-intrusive load transient identification. Appl. Energy **195**, 634–645 (2017)
41. Liu, Q., Kamoto, K.M., Liu, X., Sun, M., Linge, N.: Low-complexity non-intrusive load monitoring using unsupervised learning and generalized appliance models. IEEE Trans. Consum. Electron. **65**(1), 28–37 (2019)
42. Mei, J., He, D., Harley, R.G., Habetler, T.G.: Random forest based adaptive non-intrusive load identification. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN), pp. 1978–1983, July 2014
43. Meng, N., Sun, X., So, H.K., Lam, E.Y.: Computational light field generation using deep nonparametric Bayesian learning. IEEE Access **7**, 24990–25000 (2019)
44. Mesadi, F., Erdil, E., Cetin, M., Tasdizen, T.: Image segmentation using disjunctive normal Bayesian shape and appearance models. IEEE Trans. Med. Imaging **37**(1), 293–305 (2018)
45. Nalmpantis, C., Vrakas, D.: Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. Artif. Intell. Rev. **52**(1), 217–243 (2019)
46. Rahimpour, A., Qi, H., Fugate, D., Kuruganti, T.: Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint. IEEE Trans. Power Syst. **32**(6), 4430–4441 (2017)
47. Rashid, H., Singh, P., Stankovic, V., Stankovic, L.: Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour? Appl. Energy **238**, 796–805 (2019)
48. Ruiz, F.J.R., Valera, I., Svensson, L., Perez-Cruz, F.: Infinite factorial finite state machine for blind multiuser channel estimation. IEEE Trans. Cogn. Commun. Netw. **4**(2), 177–191 (2018)

49. Shin, C., Rho, S., Lee, H., Rhee, W.: Data requirements for applying machine learning to energy disaggregation. Energies **12**(9), 1696 (2019)
50. Tabatabaei, S.M., Dick, S., Xu, W.: Toward non-intrusive load monitoring via multi-label classification. IEEE Trans. Smart Grid **8**(1), 26–40 (2017)
51. Varela, P.M., Hong, J., Ohtsuki, T., Qin, X.: IGMM-based co-localization of mobile users with ambient radio signals. IEEE Internet Things J. **4**(2), 308–319 (2017)
52. Wang, W., Xi, J., Zhao, D.: Driving style analysis using primitive driving patterns with Bayesian nonparametric approaches. IEEE Trans. Intell. Transp. Syst. 1–13 (2018)
53. Xu, Y., Cheng, P., Chen, Z., Li, Y., Vucetic, B.: Mobile collaborative spectrum sensing for heterogeneous networks: a Bayesian machine learning approach. IEEE Trans. Sig. Process. **66**(21), 5634–5647 (2018)
54. Yan, D., et al.: Household appliance recognition through a Bayes classification model. Sustain. Cities Soc. **46**, 101393 (2019)
55. Zaidi, A.A., Kupzog, F., Zia, T., Palensky, P.: Load recognition for automated demand response in microgrids. In: Proceedings of the IECON 2010–36th Annual Conference on IEEE Industrial Electronics Society, pp. 2442–2447, November 2010
56. Zhao, B., Stankovic, L., Stankovic, V.: On a training-less solution for non-intrusive appliance load monitoring using graph signal processing. IEEE Access **4**, 1784–1799 (2016)