A Semi-Supervised and Online Learning Approach for Non-Intrusive Load Monitoring

Hajer Salem[1,2][0000−0002−7433−7226] ✉ and Moamar Sayed-Mouchaweh[1][0000−0002−6929−986X]

[1] Institut Mines-Télécom Lille Douai, France,
Department of Computer Science and Automatic Control
hajer.salem@imt-lille-douai.fr
moamar.sayed-mouchaweh@imt-lille-douai.fr
[2] National School of Computer Science ENSI, Manouba University, Tunisia

Abstract. Non-Intrusive Load Monitoring (NILM) approaches aim at identifying the consumption of a single appliance from the total load provided by smart meters. Several research works based on Hidden Markov Models (HMM) were developed for NILM where training is performed offline. However, these approaches suffer from different issues: First, they fail to generalize to unseen appliances with different configurations or brands than the ones used for training. Second, obtaining data about all active states of each appliance requires long time, which is impractical for residents. Third, offline training requires storage of huge amount of data, yielding to share resident consumption data with external servers and causing privacy issues. Therefore, in this paper, a new approach is proposed in order to tackle these issues. This approach is based on the use of a HMM conditioned on discriminant contextual features (e.g., time of usage, duration of usage). The conditional HMM (CHMM) is trained online using data related to a single appliance consumption extracted from aggregated load in order to adapt its parameters to the appliance specificity's (e.g., brand, configuration, etc.). Experiments are performed using real data from publicly available data sets and comparative evaluation are performed on a publicly available NILM framework.

Keywords: Non Intrusive Load Monitoring (NILM) · Load Disaggregation · Hidden Markov Model · Online learning · Online Expectation Maximization algorithm.

1 Introduction

Non-intrusive load monitoring or power disaggregation refers to the problem of disaggregating single appliance consumption from the total electrical load in a house. NILM has many practical applications in the smart grid development in order to solve many challenges. For instance, it helps to reduce a consumer electricity bill by providing details to consumers about the consumption of each used appliance. Recently, many smart meters have been deployed in Europe. However, resident feedback following the smart meter installation in households

point out the need to meet protecting privacy requirement. A big issue facing NILM deployment is the privacy loss because residents usually complain about sharing their personal data with utility companies. A potential solution is to perform data processing at the household level where personal data are not shared with external parts. A major advantage of the proposed approach in this paper is to fully protect a consumer privacy. Indeed, the learning and disaggregation modules could both be performed at the level of the smart meter thanks to the low complexity of the proposed approach and on the fly treatment of the data.
Generally NILM systems include three main modules that are data acquisition, model learning and load disaggregation [26]. In this paper, the learning module is enhanced in order to process online. Data is gathered from smart meter installed in households. A low sampling rate of 1/60 Hz (1 minute interval) is considered in this paper because it is the real world sampling rate existing in households. Indeed, smart meters with higher sampling rate are expensive for a deployment in residential sector [21].
The NILM framework proposed in this paper is depicted in Figure 1. The proposed approach is semi-supervised, performs online learning and disaggregation using data gathered from low sampling rate smart meters in order to overcome the discussed challenges in NILM. It pursues the following steps: First data preprocessing is performed and data is analyzed in order to extract discriminating features and prior generic models are created; Second, a method for appropriate edge change detection and a new approach for selecting training windows of a single appliance consumption extracted from the total load is proposed (blocks 2 and 3); Third, a new HMM conditioned on contextual features that we call CHMM for NILM is proposed. Besides, an extended version of an online Expectation Maximization (EM) algorithm for estimating the CHMM parameters is developed (block 4). The algorithm is adapted for NILM and propose a solution to estimate correctly under represented states. Finally, disaggregation is performed using the updated CHMM.
The reminder of this paper is organized as follows: section 2 is a discussion of recent NILM works, their advantages and limits. Section 3 explains data preprocessing performed and the generic models taken as prior. Section 4 develops the edge change detection and training windows selection approach for NILM. The proposed conditional HMM is formalized in section 5. The proposed online parameter estimation approach is developed in section 6. Experiment results are depicted and discussed in section 7. Section 8 concludes the paper and presents future work.

2 Related work

A common category of NILM approaches assumes that sub-metered ground truth data is available for training prior to performing disaggregation [15] [25]. These approaches show promising accuracy results but they violate two requirements of NILM that are generalization and unsupervised disaggregation. Another category of approaches are often based on a signature database of appliances and

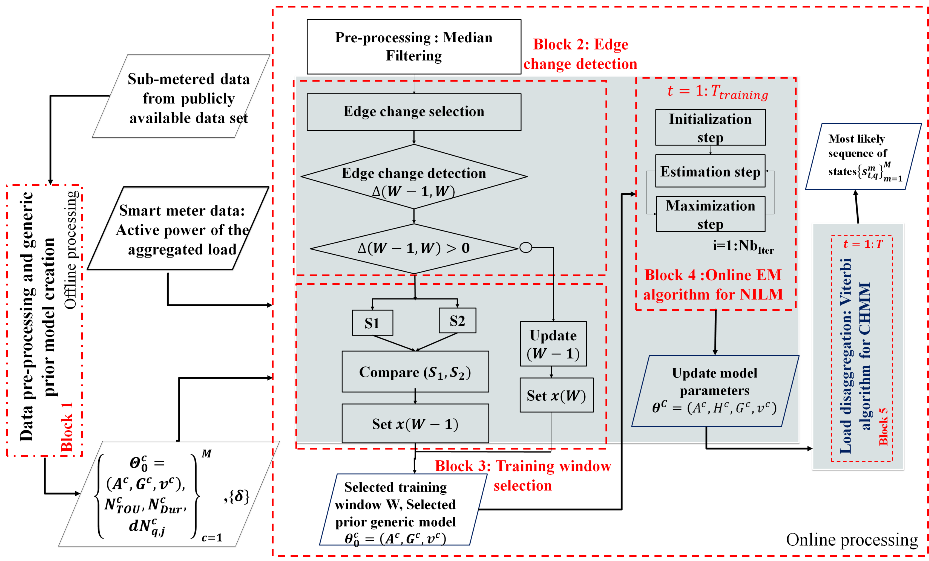

Fig. 1: Proposed approach global diagram

classify appliances according to these signatures [2] [23]. However, it is impossible and very expensive to collect all appliances signatures in a database because many instances (specific size, model, etc.) exist for each appliance category.
Several deep learning based approaches [9] [13] [24] have been proposed for NILM. They showed very promising results regarding generalization to unseen appliances and presented major improvement in terms of complexity. However this performance still prone to: the need to high frequency data which is impractical in residential sector due to return of investment issues; the difficult process of training; and sometimes dependency to labelled data. Unsupervised and semi-supervised HMM based approaches, [15], [14], present the best compromise in terms of meeting NILM requirements and disaggregation accuracy [17]. A bench of unsupervised approaches based on variants of Factorial Hidden Markov Models (FHMM) have been proposed [11][8][18] [3]. These approaches share almost the same scalability issues. Indeed, these approaches are not applicable for a number of appliances greater than ten [17]. However, integrating non traditional based features seems to alleviate this limitation [18]. Non traditional features refer to contextual based and behavioural based features. Hour of the day, day of the week and duration have been proposed as additional features for FHMM [10] which improved the disaggregation accuracy, at the expense of increasing the model complexity. Furthermore, time of the day and seasonal context-based patterns have been incorporated to a recent NILM approach [7]. A whole year of usage data have been used for training in order to build usage patterns which is impractical for real world application. Power consumption patterns of appliances and user presence have been investigated in [18]. The

proposed inference algorithm is an extension to the AFAMAP (Additive Factorial Approximate Maximum a Posteriori) disaggregation algorithm proposed in [11] with contextual features. This extension shows promising results where the precision increased by approximately 15%. Nevertheless, the approach is considered as supervised because the same appliances are used for training and testing and its performance in the case of unseen appliances cannot be evaluated. As a matter of conclusion, contextual features may be an important lever for power disaggregation accuracy. However there is a lack in the state of the art of an online algorithm for learning these features from a specific household.

3 Data pre-processing and prior generic models creation

The proposed approach takes advantage of publicly available appliance consumption in databases in order to create prior generic models for each category of appliances (i.e, fridge, microwave, stove,...). For each category denoted by c, several instances exist (e.g., fridge 1, fridge 2,..., fridge N). These prior models are used to help labelling after disaggregation and are updated based on data stream readings from smart meters to best fit a particular household appliance. Mainly, data pre-processing is performed on the data sets and M generic prior models are created. The prior generic model for each appliance, is modeled as a Hidden Markov Model (HMM) with parameters $\theta_0^c = (A_0^c, G_0^c, v_0^c)$.

3.1 Number of states per appliance and Generic power profiles

Clustering analysis have been performed on the publicly available data sets in order to set the number of states Q per appliance and approximate emission distributions. Generic prior emission distributions have been approximated to Gaussian distribution using different instances for each appliance category and setting a prior on the mean of these distributions.

3.2 Generic contextual features distributions

Three additional distributions are considered as additional information that are the distribution of usage duration per state for cooling appliances, the distribution of time of usage per hour of the day for entertainment appliances and some kitchen devices and the difference between state consumption for states within the same appliance. These distributions are obtained as the following:

- **Generic Duration distribution:** For cooling appliances, the state duration is not related to a user's habits but related to the appliance internal operational behavior. Duration represents an interesting discriminating generalized feature which could be generalized to unseen appliances especially to distinguish between states from different appliances that have similar active power consumption distribution. However, for these appliances, the usage time is not a discriminating feature because cooling appliances consume continuously. Analysis have been performed and showed that state duration of cooling appliances could be generalized to Gaussian density with a Gaussian prior on the mean of these distributions where:

$$\begin{aligned} Dur_q^m &\sim N\left(\frac{\sum_{i=1}^{N} \mu_{q,i}^c}{N}, \max_{i=1:N} \sigma_{q,i}^c\right) \\ \mu_q^c &\sim N\left(0, \sqrt{\frac{\sum_{j=1:N} (\mu_{q,i}^c - \mu_q^c)^2}{N}}\right) \end{aligned} \tag{1}$$

where $q \in [1:Q]$ and $i \in [1:N]$ is the number of instances per appliance category.

- **Generic Usage time distribution:** Data have been analyzed in function of usage time. Time of usage have been extracted from the timestamp reported with active power at sampling instant. Data suggest that some appliances such as entertaining appliances (e.g., Laptop, TV, Stereo) are often ON during specific hours of the day. These appliances' times of usage could be generalized over different households because it is concentrated over some hours of the day (after work, in the evening and before sleeping). Generalization over different devices within the same appliance category has been performed in the same manner as for duration.
- **Difference Distribution between each appliance Power Profiles:** Difference between power consumption of two different states of the same device has been studied. Difference between two states remains always the same even if further appliances consume simultaneously [11][19]. Using the consumption difference aims at detecting when states of the same appliance are consuming successively in the aggregated load. The consumption difference between two emission probability distributions $X_i \sim N_i$ and $X_j \sim N_j$, $i, j \in \{1, Q\}$ has been approximated to a Gaussian distribution.

3.3 Prior adaptive Edge change thresholds

An adaptive edge change threshold is proposed. Edge change detection consists in detecting if an appliance changes its state from one to another. A naive approach deployed in almost power disaggregation approaches [22][11] is to monitor power consumption readings and to flag an event when the power change deviates beyond a fixed threshold. However, a fixed threshold could be within an appliance variance when it is too small and hence detects false state changes. Besides, choosing a large threshold may lead the system to loop state transitions with small consumption. For this current work, the edge change is detected based on an adaptive threshold δ_W that dynamically changes according to the detected appliance state operating during the last observed window of observations. Several thresholds are proposed and are computed according to the variance of the emission distributions of the generic prior models. Indeed, a clustering is performed on the different variances of appliances states consumption. The minimal variance within each cluster $i \in [1, Nb_{clusters}]$ has been set as an event detection threshold δ_i.

4 Edge change detection and online training windows selection

4.1 Problem formulation and complexity

Power consumption readings $\{y_t\}$ arrive in data streams. An edge change ($\Delta_{t,t-1}$) is detected if the active power consumption difference between y_t and y_{t-1} is greater than a threshold δ_i (see Figure 2). Let W denotes the current dynamic window of observations. Its size is determined by the number of observations received between the two last successive detected edge changes. δ_i is chosen according to the appliance state denoted by x_q^m identified as operating during window $W-1$. Indeed, for each appliance state identified as operating during $W-1$, the proposed approach selects a different threshold δ_i where $i \in [1, Nb_{clusters}]$ as explained above in section 3.3 and illustrated in Figure 2.

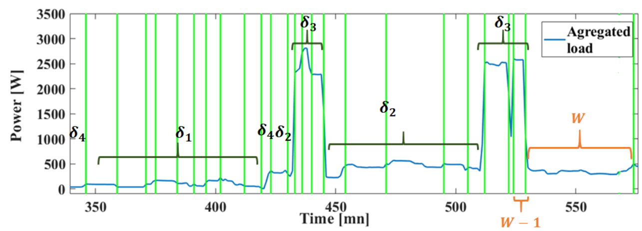

Fig. 2: Event detection example using the adaptive threshold δ_i

The proposed training window selection approach aims at extracting single appliance consumption from the aggregated load to be used as training samples for the HMM parameter estimation. The proposed training module is able to run online and process data on the fly. It succeeds to overcome several shortcomings facing offline training approaches in the state of the art as follows:

- Space complexity: The proposed approach requires the storage of only data within W. Its space complexity is equal to $O(|W| *D)$ where $|W|$ is the window length and D is the observations dimension; The length of W is usually equal to fzw minutes of consumption data. The spatial complexity of this approach is negligible compared to the spatial complexity of offline training approaches [19] [10] [11] [8] which is equal to $O(T*D)$. Indeed, these approaches require the storage of the full training data (T) which is equal to several days of consumption data.
- Time complexity: The proposed approach is based on the Kullback Leibler divergence and a posteriori probability estimation that have at the worst case a temporal complexity of $O(1)$. Besides, the proposed online EM algorithm performs on the fly data processing and has a temporal complexity equal to $O(Q^2)$ where Q is the number of an appliance states. This proposed

approach has significantly a lower complexity compared to the state of the art approaches based on the Forward Backward algorithm, which iterates over the full training data T and has a temporal complexity of $O(Q^2T)$ per iteration.

4.2 The proposed training window selection method

As mentioned above, only data within the last window W is registered. Besides, the training module memorizes sufficient statistics related to: *i)* the consumption distribution denoted by N_{est}, *ii)* the difference distribution between $(W, W-1)$ denoted by dN_{est} and *iii)* the contextual features (time of the day, duration). In the sequel, we enumerate the different scenarios that might explain an edge change detection and propose a method that evaluates each scenario. The steps of the proposed method are enumerated in block (3) of Figure 1. Two types of edge changes $\Delta_{W,W-1}$ could be observed: a positive edge change and a negative one.

On the one hand, a negative $\Delta_{W,W-1}$, is interpreted as the appliance state x_i^m, active within $(W-1)$, turns off when $\Delta_{W,W-1}$ is detected. In order to identify x_i^m, the proposed method pursues the following steps: First, the distribution of observations within $(W-1)$ is approximated to a Gaussian distribution N_{est}, with a mean equal to $|\Delta_{W,W-1}|$; Second, N_{est} is compared to all prior generic model emission distributions $\{g_i^m\}_{m=1}^M$ based on Kullback Leibler divergence. A subset of most similar prior generic appliance states are selected and are compared further by computing a posteriori probability of contextual features; A score is computed for each potential active appliance state and the state having the highest score is identified as the state operating during window $(W-1)$. Finally, observations within $W-1$ are feeded to the online EM algorithm to update the parameters of the appliance m model.

On the other hand, a positive $\Delta_{W,W-1}$ could be explained by two scenarios as follows:

First scenario: the edge change corresponds to an appliance state q of appliance m'' that was OFF within $(W-1)$ and turns ON during W. Let's denote by $x_i^{m'}$ the state appliance that was identified active within $W-1$. In this case the mean power consumption of the new activated state $x_q^{m''} \approx \Delta_{W,W-1}$ and its consumption distribution is approximated to $N_{est}(\Delta_{W,W-1}, \sigma_W)$. The proposed evaluation procedure adopted in this scenario consists in comparing the Kullback Leibler divergence between the estimated distribution N_{est} within W and all prior generic models emission distributions of appliances categories. A subset of most similar states is selected. Then a posteriori probability of the observed contextual feature is calculated. The state having the maximal probability is selected as potentially consuming during W.

Second scenario: The edge change corresponds to the same appliance m that changes its active state (q_1) during $W-1$ to its second active state (q_2) during W. In this case the mean power consumption of x_{q2}^m is approximated to $\mu_{q2}^m \approx \Delta_{W,W-1} + \mu_{q1}^m$ and its consumption distribution is approximated to

$N_{est}(\mu^{m}_{q2}", \sigma_W)$. The difference distribution dN_{est} observed within W is compared to all prior generic difference distributions dx^m_i defined in section 3. Kullback Leibler divergence is computed between dN_{est} and each prior generic difference distribution. A subset of most similar states is selected. Further comparison is performed based on a posteriori probability of contextual features and a score is computed. The state having the highest score is selected.
The two selected states from each potential scenario are compared according to their mutual scores and the one having the highest score is identified as active. For some observation windows, the proposed approach may not select any appliance state as consuming during W. Indeed, the proposed algorithm only selects a window of observations where it confidently recognize an appliance is operating based on both Kullback Leibler and Maximum a Posterior. Finally, the samples within a selected window are used as training data for the proposed CHMM.

5 Conditional hidden Markov model(CHMM)

In this work, we suggest to condition the hidden state on the probability of usage during hours of the day. The likelihood of appliance state q being active during an hour of the day td $p(x = q|t = td)$ follows a Multinomial distribution of probabilities $(h_{1,q}, ..., h_{td,q}, ..., h_{24,q})$ and a number of trials n_h that is the number of observations per hour where $td \in \{1, ..., 24\}$, $\forall q \in \{1, ..., Q\}$ and $\sum_{td=1}^{24} h_{td,q} = 1$. Indeed, as shown in figure 3, appliances states have higher probability to be active during some hours of the day than others. For a particular household, the time of usage of appliances encapsulates the user habits and it is interesting to learn these habits online. Therefore, the hidden states X

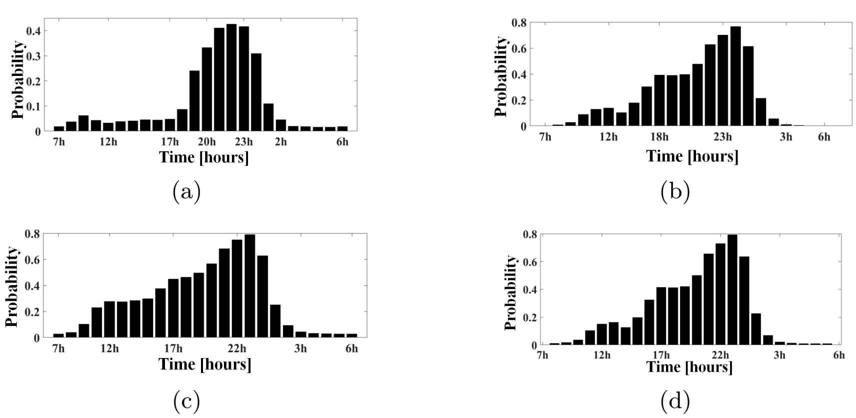

Fig. 3: Histograms of usage time probabilities over hours of the day: a) Lamp, b) Television, c) Stereo, d) Entertainment

are conditioned on both the transition matrix A and a usage time matrix H.

Henceforth, the proposed joint probability distribution over latent variables and observed ones is given by equation (2) where $\theta = (v, g, A, H)$:

$$p\left(X_{1:T}, Y_{1:T}|\theta\right) = p(x_1|v)p(y_1|x_1)p(x_1|h_{td})\prod_{t=2}^{T}\left[p(x_t|x_{t-1}, A)p(x_t|td, H)p(y_t|x_t, g)\right] \tag{2}$$

An example of a washing machine model is represented in figure 4(a) and the graphical representation of the CHMM is depicted in figure 4(b).

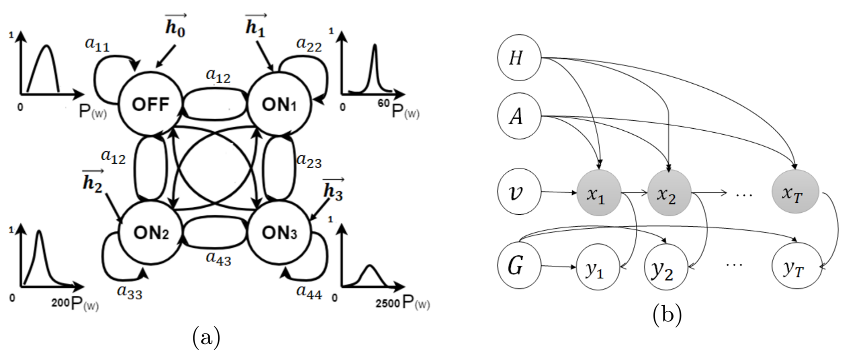

Fig. 4: (a): Example of a CHMM modeling a washing machine with four states, (b): Conditional HMM graphical representation

6 Online Parameter Estimation: Online EM for NILM

The EM algorithm is one of the most popular algorithms for parameter inference in HMMs due to its convergence properties and robustness. An online EM algorithm has been proposed by Cappé et al. [4]. The algorithm main objective is to train HMM with continuous observations on data streams. Data are processed just one time and never stored. We opted to adapt this algorithm to NILM and extend it to CHMM. In this work, this algorithm is enriched and modified according to two main points: 1) A new approximate filter $\psi(x)$ and intermediate quantity ρ^h are added to the model in order to estimate recursively the time of usage parameters; 2) A new decreasing sequence γ_n is proposed. This new sequence ensures the learning of parameters related to under represented states in training data. Indeed, some appliances states are rarely observed. A second challenge facing the application of Online EM [4] to NILM, is that this algorithm is designed to train only one HMM from data generated continuously from the same model. This is not the case in NILM. Indeed, data are generated from different models. Observations y_t and y_{t+1} may be generated from different models HMM_i and HMM_j where $i, j \in \{1, M\}$.

6.1 The new time of usage parameter estimation

The proposed online EM algorithm for CHMM aims at learning a new matrix H of size $24 \times Q$. Each row vector of H denotes the probabilities of each state q to be active during the hour of the day $td \in [1:24]$. The sum of each row has to be equal to one. In the online EM considered by [4], the leaned parameter θ is composed of the state transition matrix A and the mean vector and covariance matrix associated with each of the q emission distribution g_q. In such a model, there are two distinct types of EM complete-data sufficient statistics which give rise to two separate forms of the auxiliary function ρ^q_{n+1} and ρ^g_{n+1} and an approximate filter ϕ_{n+1}. However, in the proposed CHMM, a new type of sufficient statistic related to the hour of the day usage probability is added. Therefore, in addition to the two previous forms of the auxiliary function, a new one is introduced and formulated according to equation (9) in algorithm 1 and computes the expectation of the state j being active during the hour of the day td according to the intermediate quantity 4 as follows:

$$\rho^h_{n+1}(j,k,td;\theta) = \frac{1}{n}\mathbb{E}_{v,\theta}\left[\sum_{t=1}^{n} \mathbb{1}\{X_t = j, t = td\}|X_n = k,\right], \tag{3}$$

Besides, a new approximate filter denoted by ψ_{n+1} and a backward retrospective probability denoted by r^h_{n+1} are defined as formulated respectively in equations (7) and (11) of algorithm 1. Finally, a new parameter update formula is proposed for the maximization step in order to update h_{θ_n} (Equation (13), algorithm 1).

6.2 New proposed weights

A well known problem of EM algorithm is the convergence rate. Indeed, the latter depends on the initial parameters θ_0 [6]. Besides, the weights γ_{n+1} in equations (8,9 and 10) have crucial importance for the rate of convergence. It is usually chosen to form a positive decreasing sequence. A step size of $\gamma_n = n^{-\alpha}$ is used in [4] where $\alpha \in [0.5, 0.8]$. New studies showed that a decreasing sequence equal to $Cn^{-\alpha}$ (C is a constant) reaches better rates of convergences [16]. Still, all these fully decreasing sequences are proposed based on the assumption that observed data are generated from all HMM states continuously which is not the case in NILM. The decreasing sequence adopted in almost EM algorithms applies the same factor (α) regardless of the observations. However, for under represented states (case of unbalanced observations), the left hand side of equations (8,9 and 10) could be omitted due to a small value of γ and a small number of observations.

In this current work, the decreasing sequence γ_n is considered as an adaptive cooling schedule of a simulated annealing algorithm that could increase. The application of a different cooling rate, that depends on the representation of states, would allow the algorithm to best fit to rare observations. Hence, γ_{n+1} continue to decrease if the consequent observations are generated from a state for which training windows have been selected. We memorize a flag per appliance state to denote if a training window has been found for the considered state or not. In the case of a first training window selected for an appliance state (a HMM latent variable), γ_{n+1} increases in order to give more important weights to the observed data from the new observed state.

6.3 Online EM algorithm for Conditioned HMM

The auxiliary EM function is defined for the CHMM as follows:

$$\begin{aligned} Q(\theta,\theta') = &\sum_{i=1}^{Q}\sum_{j=1}^{Q} S_n^q(i,j)\log a(i,j) \times S_n^h(j)\log h(td,j) \\ &- \frac{1}{2}\sum_{i=1}^{K}\left(\frac{S_{n,2}^g(i,\theta)}{v(i)} - \frac{2\mu(i)S_{n,1}^g(i)}{v(i)} + S_{n,0}^g\left(\frac{\mu^2(i)}{v(i)} + \log v(i)\right)\right) \end{aligned} \quad (4)$$

The proposed online EM for CHMM applied to NILM is described in algorithm 1. First prior parameters $\theta_0 = (v_0, G_{\theta_0}, A_{\theta_0}, H_{\theta_0})$ are initialized where A_{θ_0} is the initial transition matrix, H_{θ_0} is the initial hour usage probability matrix. Then both initial values of the approximated filters $\hat{\phi}_0(x)$ and $\hat{\psi}_0(x)$ are initialized [algorithm 1, line 2]. The auxiliary quantities (8, 9, 10) are initialized to zero. Then for each new observation within a selected training window W, the approximate filters ϕ and ψ are updated as formulated in (6) and (7). A test is performed in order to verify the representation of states. The parameters update formulas of the maximization step are calculated by maximizing equation (4) with respect to a_θ, h_θ μ_θ and σ_θ and satisfying the constraints $\sum_j^Q a(i,j) = 1$, $\sum_{td=1}^{24} h(j,td) = 1$ as formulated in (12)-(15).

7 Experimental Evaluation

We implemented the proposed approach using Matlab 2017b on a 64-bit windows 10 PC with core intel(R) i5-3320M CPU processor and 8.00 GB of memory. Experiments have been carried out on Nilm-Eval [5] framework for disaggregation of real-world electricity consumption data. We conducted our experiments using three publicly available data sets for real world electricity consumption. Two European databases named Tracebase [20] and Electricity Consumption and Occupancy (ECO) [1] and an American database named the Reference Energy Disaggregation Data Set (REDD) [12]. All the conducted experiments in this paper are based on low frequency sampling data of 1/60 Hz (1 minute interval) sampling rate. Each obtained result is an average of 10 independent runs of the algorithm. Prior models have been built using both Tracebase and REDD data sets to cover a maximal number of appliance categories. For each appliance category, seven different appliance instances are used to build the prior models. For each instance, the same number of samples (data) are used. We carried out three kinds of experiments in order to test the proposed approach. All the carried experiments are based on data from household 2 of ECO data set because it contains the maximal number of appliances.

Algorithm 1 Online Expectation Maximization (EM)

Initialization
Compute for $x \in X$

$$\phi_0(x) = \frac{v(x) g_{\hat{\theta}_0}(x, y_0)}{\sum_{x' \in X} v(x') g_{\hat{\theta}_0}(x', y_0)}, \quad \psi_0(x, td) = \frac{h_{\theta_0}(td, x)}{\sum_{td' \in [1:24]} h_{\theta_0}(td', x)} \tag{5}$$

$\hat{\rho}_0^q(x) = 0,\ \hat{\rho}_0^h(x) = 0,\ \hat{\rho}_0^g(x) = 0$
Recursion
for $n \geq 0$ **do**
  Compute, for $x \in X$
  Approximate filter Update

$$\hat{\phi}_{n+1}(x) = \frac{\sum\limits_{x' \in X} \hat{\phi}_n(x') a_{\hat{\theta}_n}(x', x) g_{\hat{\theta}_n}(x, y_{n+1})}{\sum\limits_{x', x" \in X} \hat{\phi}_n(x') a_{\hat{\theta}_n}(x', x") g_{\hat{\theta}_n}(x", y_{n+1})}, \tag{6}$$

$$\hat{\psi}_{n+1}(x, td) = \frac{\sum\limits_{x' \in X} \hat{\psi}_n(x', td_n) h_{\hat{\theta}_n, i}(td, x)}{\sum\limits_{td' \in [1:24]} \sum\limits_{x' \in X} \hat{\psi}_n(x') h_{\hat{\theta}_n, i}(td', x')} \tag{7}$$

  Stochastic Approximation E-step
  if flag $==$ true **then**
    $\gamma_{n+1} = \gamma_1$
  else
    $\gamma_{n+1} = (n+1)^{-\alpha}$
  end if

$$\rho_{n+1}^q(i, j, k; \theta) = \gamma_{n+1} \delta(j-k) \hat{r}_{n+1}(i|j) + (1 - \gamma_{n+1}) \sum_{k'=1}^{Q} \hat{\rho}_n^q(i, j, k') \hat{r}_{n+1}(k'|k), \tag{8}$$

$$\rho_{n+1}^h(j, k, td; \theta) = \gamma_{n+1} \delta(j-k) \delta(t-td) \hat{r}_{n+1}^h(j|td) + (1-\gamma_{n+1}) \sum_{k'=1}^{Q} \hat{\rho}_n^h(j, k', td) \hat{r}_{n+1}^h(j|td), \tag{9}$$

7.1 Evaluation of the proposed training window selection approach

The main goal of the first experiment is to evaluate the accuracy of the selected training windows. Evaluations have been carried out in terms of True Positive rate (TPR) and False Positives rate (FPR). True Positive rate measures the proportion of actual appliance activation in the ground truth data (plug level data) identified correctly by the proposed approach. False positive rate measures the proportion of appliance activation identified by the proposed approach and do not exist in the ground truth data. The intended metrics verify that the selected windows for training correspond truly to windows where a state of a specific appliance is active. The obtained results shown in Table 1 highlight that the proposed approach succeeds to find online training windows for all appliances. The evaluation have been performed on data from two weeks of consumption.

$$\rho^g_{n+1}(i,k) = \gamma_{n+1}\delta(i-k)s(Y_{n+1}) + (1-\gamma_{n+1})\sum_{k'=1}^{Q}\hat{\rho}^g_n(i,k',\theta)\hat{r}_{n+1}(k'|k), \tag{10}$$

Where:

$$r_{n+1}(i|j) = \frac{\hat{\phi}_n(i)q_{\hat{\theta}_n}(i,j)}{\sum_{i'=1}^{Q}\hat{\phi}_n(i')q_{\hat{\theta}_n}(i',j)}, \quad r^h_{n+1}(j|td) = \frac{\hat{\psi}_n(j,td)h_{\hat{\theta}_n}(td,j)}{\sum_{td'=1}^{24}\hat{\psi}_n(j,td')h_{\hat{\theta}_n}(td',j)} \tag{11}$$

if $n \geq n_{min}$ **then** Update the parameters:

$$a_{\theta_n}(i,j) = \frac{S^q_n(i,j)}{\sum_{j=1}^{Q} S^q_n(i,j)} \tag{12}$$

$$h_{\theta_n}(td,i) = \frac{S^h_n(td,i)}{\sum_{j=1}^{Q} S^h_n(td,j)} \tag{13}$$

$$\mu_{\theta_n}(i) = \frac{S^g_{n,1}(i)}{S^g_{n,0}(i)} \tag{14}$$

$$\sigma_{\theta_n}(i) = \mu^2_{\theta_n}(i) + \frac{S^g_{n,2}(i) - S^g_{n,1}(i)\times\mu_{\theta_n}(i)}{S^g_{n,0}(i)} \tag{15}$$

else set $\hat{\theta}_{n+1} = \hat{\theta}_n$ **end if**=0

The approach shows also robustness to noise. For instance, appliances that have never been ON (i.e., the kettle) during the test period, have not been identified as active (FP=0). Unfortunately, the fridge and freezer false positive rates are important. However, this is a common problem because these appliances are consuming continuously and their steady state is frequently confused with states from different appliances. These results also reveal that duration feature considered for cooling appliances is not discriminating enough.

Table 1: Accuracy evaluation of detected training windows

	Fridge	Dishwasher	Laptop	Freezer	TV	Air Exhaust	Lamp	Kettle	Stereo	Tablet	Entertainment	Stove
TPR	0.42	0.38	0.49	0.41	0.43	0.29	0.35	0	0.27	0	0.44	0.56
FPR	0.31	0.06	0.16	0.68	0.02	0.01	0	0	0	0	0.06	0.01

7.2 Evaluation of the proposed Online EM

The second kind of experiment evaluates the impact of the new proposed weights γ on the accuracy of the parameter estimation. A comparison between the effect of γ used in the online EM algorithm [4] and the one proposed in this work is conducted on the entertainment appliance. Indeed, the entertainment appliance is a typical example of appliances that have imbalanced represented states. Figure **??** (c) shows data observed within the two first days of consumption of the appliance in household 2. We can observe that the state ON 2 is under represented compared to the two other states and is rarely active. For the scope of this experiment, we focused on the mean update of the appliance state emission distributions. We can visually distinguish three different states where the OFF state has a mean around 0 watts, the ON 1 state has a mean around 160 watts and the ON 2 state has a mean around 40 watts. Figure **??**(a) shows that the online EM algorithm proposed in [4] converges to a mean around 100 watts for the case of the state ON2. This is explained by the fact that the algorithm confuses between ON 1 and ON 2 states. However, the adaptive γ sequence proposed in this work helps the algorithm to converge accurately to the real state mean as shown in figure **??**(b) because it gives more important weights to observations generated from the new detected state.

7.3 Load disaggregation evaluation

The third kind of experiments intend to evaluate the impact of conditioning on the time of usage probabilities on the disaggregation results. Accuracy has been evaluated using the same data from household 2. Evaluation have been performed on 90 days of consumption data. Accuracy reports how much power is being consumed by an appliance compared to its actual consumption and is computed as follows [11]:

$$Acc = 1 - \frac{\sum_{t=1}^{T} |\hat{y}_t^m - y_t^m|}{2\sum_{t=1}^{T} y_t^m}$$

where y_t^m is the real consumption of appliance m and $\hat{y}_t^m$ is the estimated consumption of appliance m during time t.
Three variants of HMM are compared where the first variant is a HMM trained on sub-metered data denoted by "Supervised HMM". The second variant, is a HMM trained using sample data selected online using the method developed in section 4.2 and using the online EM proposed in [4]. This second variant is denoted by "Online HMM". The third variant is the CHMM formalized in section 5, trained on data samples selected using the method developed in section 4.2 and the online EM for CHMM developed in section 6. This third variant is denoted by "Online CHMM". Results are depicted in table 2. The proposed CHMM outperforms the two other models mainly on the lamp, dishwasher, stereo and laptop appliances. This improvement highlights that time of usage is an important discriminating feature for the aforementioned appliances. However, accuracy results obtained for cooling appliances are approximately the same for the three models. This could be interpreted as the time of usage has no additional impact on disaggregation for these appliances. Moreover, the results obtained by comparing the first two models "Supervised HMM" and "Online HMM" show that both models give similar accuracy results which confirms that training the model online gives the same results as using sub-metered data.

Table 2: Accuracy results using the proposed CHMM trained online, a supervised HMM trained on sub-metered data and a HMM trained online

	Fridge	Freezer	Lamp	Dishwasher	Stereo	Laptop	Tablet	Air Exhaust	Entertainment	Kettle	Stove	TV
Supervised HMM	0.46 ± 0.02	0.49 ± 0.04	0.59 ± 0.12	0.56 ±0.11	0.58 ±0.08	0.52 ±0.09	0.49 ± 0.05	0.51 ± 0.08	0.57 ± 0.07	0.55 ± 0.04	0.49 ± 0.07	0.53 ± 0.0
Online HMM	0.47 ±0.03	0.5 ±0.03	0.58 ±0.09	0.55 ±0.15	0.51 ±0.08	0.33 ±0.12	0.49 ± 0.01	0.53 ± 0.09	0.53 ± 0.05	0.54 ± 0.01	0.49 ± 0.03	0.49 ± 0.0
Online CHMM	0.57 ±0.02	0.52 ±0.03	0.85 ±0.02	0.83 ± 0.02	0.64 ±0.03	0.77 ±0.01	0.5 ± 0.04	0.69 ± 0.04	0.75 ± 0.08	0.68 ± 0.02	0.81 ±0.03	0.86 ± 0.0

8 Conclusion

In this paper, a semi-supervised approach performing online learning for Non-intrusive Load Monitoring (NILM) was proposed. The aim is to develop an algorithm that can be embedded in a smart meter in order to alleviate the privacy issues facing NILM. The proposed approach succeeded to extract training samples from the aggregated load online for each appliance. Training using samples selected from the aggregated load gave the same accuracy results as training on sub-metered data. Besides, a new conditional hidden Markov model (CHMM) that condition on usage time was proposed for NILM. An online Expectation Maximization algorithm was developed to learn CHMM parameters. Disaggregation based on the proposed CHMM improved accuracy results especially for appliances such as: the lamp, the dishwasher, the stereo and the laptop.
Future work will consider a conditional factorial hidden Markov model (CFHMM) where the time of usage will be used to block sample for a particular hour of the day on a subset of active appliances. The aim of this blocked sampling is to decrease the computational complexity of inference algorithms in factorial hidden Markov models.

References

1. Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., Santini, S.: The eco data set and the performance of non-intrusive load monitoring algorithms. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings. pp. 80–89. ACM (2014)
2. Berges, M., Goldman, E., Matthews, H.S., Soibelman, L., Anderson, K.: User-centered nonintrusive electricity load monitoring for residential buildings. Journal of computing in civil engineering **25**(6), 471–480 (2011)
3. Bonfigli, R., Principi, E., Fagiani, M., Severini, M., Squartini, S., Piazza, F.: Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models. Applied Energy **208**, 1590–1607 (2017)
4. Cappé, O.: Online em algorithm for hidden markov models. Journal of Computational and Graphical Statistics **20**(3), 728–749 (2011)

5. CicchettiMaster, R.: NILM-Eval: Disaggregation of real-world electricity consumption data. Master's thesis, ETH Zurich, Switzerland (2014)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological) pp. 1–38 (1977)
7. Dinesh, C., Makonin, S., Bajic, I.V.: Incorporating time-of-day usage patterns into non-intrusive load monitoring. In: Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on. pp. 1110–1114. IEEE (2017)
8. Egarter, D., Bhuvana, V.P., Elmenreich, W.: Paldi: Online load disaggregation via particle filtering. IEEE Transactions on Instrumentation and Measurement **64**(2), 467–477 (2015)
9. Kelly, J., Knottenbelt, W.: Neural nilm: Deep neural networks applied to energy disaggregation. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments. pp. 55–64. ACM (2015)
10. 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)
11. 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)
12. 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)
13. Lange, H., Bergés, M.: The neural energy decoder: energy disaggregation by combining binary subcomponents. In: NILM2016 3rd international workshop on non-intrusive load monitoring. Retrieved from nilmworkshop. org Google Scholar (2016)
14. Makonin, S., Bajic, I.V., Popowich, F.: Efficient sparse matrix processing for non-intrusive load monitoring (nilm). In: 2nd International Workshop on Non-Intrusive Load Monitoring (2014)
15. Makonin, S., Popowich, F., Bajić, I.V., Gill, B., Bartram, L.: Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring. IEEE Transactions on Smart Grid **7**(6), 2575–2585 (2016)
16. Moulines, E., Bach, F.R.: Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In: Advances in Neural Information Processing Systems. pp. 451–459 (2011)
17. Nalmpantis, C., Vrakas, D.: Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. Artificial Intelligence Review pp. 1–27 (2018)
18. Paradiso, F., Paganelli, F., Giuli, D., Capobianco, S.: Context-based energy disaggregation in smart homes. Future Internet **8**(1), 4 (2016)
19. Parson, O., Ghosh, S., Weal, M., Rogers, A.: Non-intrusive load monitoring using prior models of general appliance types. In: Twenty-Sixth AAAI Conference on Artificial Intelligence (2012)
20. Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., Steinmetz, R.: On the accuracy of appliance identification based on distributed load metering data. In: Sustainable Internet and ICT for Sustainability (SustainIT), 2012. pp. 1–9. IEEE (2012)
21. Salem, H., Sayed-Mouchaweh, M., Hassine, A.B.: A review on machine learning and data mining techniques for residential energy smart management. In: Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on. pp. 1073–1076. IEEE (2016)

22. Weiss, M., Helfenstein, A., Mattern, F., Staake, T.: Leveraging smart meter data to recognize home appliances. In: Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on. pp. 190–197. IEEE (2012)
23. Welikala, S., Dinesh, C., Ekanayake, M.P.B., Godaliyadda, R.I., Ekanayake, J.: Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. IEEE Transactions on Smart Grid **10**(1), 448–461 (2019)
24. Zhang, C., Zhong, M., Wang, Z., Goddard, N., Sutton, C.: Sequence-to-point learning with neural networks for non-intrusive load monitoring. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
25. Zia, T., Bruckner, D., Zaidi, A.: A hidden markov model based procedure for identifying household electric loads. In: IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society. pp. 3218–3223. IEEE (2011)
26. Zoha, A., Gluhak, A., Imran, M.A., Rajasegarar, S.: Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. Sensors **12**(12), 16838–16866 (2012)