A Hierarchical Probabilistic Model for Low Sample Rate Home-Use Energy Disaggregation

Bingsheng Wang * Haili Dong * Arnold P. Boedihardjo † Feng Chen * Chang-Tien Lu *

Abstract

Energy crisis and climate change have caused a global concern and motivated efforts to reduce energy consumption. Studies have shown that providing appliance-level consumption information can help users conserve a significant amount of energy. Existing methods focus on learning parallel signal signatures, but the inherent relationships between the signatures have not been well explored. This paper presents the Hierarchical Probabilistic Model for Energy Disaggregation (HPMED). We derive the discriminative features from low sample rate power readings to characterise device functional modes. The HPMED model bridges the discriminative features, working states, and aggregated consumption. To address the analytical intractable problem, an efficient algorithm is proposed to approximately infer the latent states for disaggregation task. Extensive experiments on a real-world dataset demonstrated the effectiveness of the proposed approach.

1 Introduction

The growing energy consumption issues pose various challenges, for example, the guarantee of sufficient energy supplies, the reduction of greenhouse effect, and "the death of globalization" due to the increase of transportation cost [1]. While the International Energy Agency (IEA) showed that energy demand continues to rise over 2% annually since 1973 [2], the U.S. Energy Information Administration (EIA) predicts that if current energy usage patterns continue, there will be about a 50% increase in world energy consumption by 2030 [3].

Generally, there are two main ways for tackling the energy conservation challenge: using more efficient energy infrastructures, or reducing energy demand by changing consumption habits. The former can incur a high cost due to large capital for developing new technologies, while the latter depends primarily on users' behaviours, which can be achieved in a cost-efficient manner. Previous studies have shown that providing appliance level information to end-users will save a significant amount of energy [5]. This motivates us to employ machine learning techniques to break down aggregated energy consumption and provide users with device level usage information.

Thanks to ubiquitous smart meters, which regularly report the whole-home level power readings, energy disaggregation methods can take the data and use it to predict device level consumption. However, the smart meters can only collect energy readings at a very low resolution (generally 1/900 Hz), which makes it difficult to identify discriminative features from the data. We carefully studied the patterns of low sample rate readings, and learned that most appliances function in multiple states. For example, a microwave can work in "Popcorn", "Defrost", or "Keep Warm" modes. Thus, we introduce the state-based features to characterise the working modes. We then propose a hierarchical probabilistic model to disaggregate the signal by incorporating the state-based features, functional modes, and interval based aggregated consumption.

The major contributions of this paper are summarized as follows:

- **In-depth study on the discriminative features of low sample rate power readings:** We carefully investigate the features for distinguishing the functional modes of devices, and develop an algorithm to extract the state-based features.
- **Design of a Hierarchical Probabilistic Model for Energy Disaggregation (HPMED):** A hierarchical model HPMED is proposed by incorporating the discovered features, working states of devices, and aggregated smart meter readings.
- **Efficient and effective estimation of latent states:** The characteristics of the estimation problem are thoroughly analysed and an efficient algorithm is developed to approximately infer the working states of devices with desirable accuracy.
- **Extensive experiments to validate the effectiveness of HPMED:** We demonstrated that HPMED outperformed other models in both whole-home and device level measures. The extensibility

*Department of Computer Science, Virginia Tech, Falls Church, VA 22043.

†U. S. Army Corps of Engineers, Alexandria, VA 22315.

of HPMED is validated through cross home evaluations.

The remainder of the paper is organized as follows. Section 2 states the problem and related work. The state-based features discovery and extraction are presented in Section 3. Section 4 introduces the hierarchical probabilistic model for energy disaggregation, and presents an algorithm for estimating latent working states. The effectiveness of the proposed approach is illustrated with extensive experiments in Section 5. Section 6 summarizes our work.

2 Problem Statement and Related Work

Energy disaggregation is the task of decomposing a whole-home aggregated energy consumption into its component devices. Assume there are M devices, such as lamp, computer, and dish washer. For each device $j = 1, \ldots, M$, we have an energy consumption matrix $\mathbf{Y}^{(j)} \in R^{N \times L}$, where N is the number of intervals in one day, and L is the number of days for the collected data. The l^{th} column of $\mathbf{Y}^{(j)}$, denoted by $\mathbf{y}^{(j)}_{\cdot,l} \in R^{N \times 1}$, is the energy usage of the l^{th} day for a particular device j, where $l = 1, \ldots, L$. The i^{th} element in vector $\mathbf{y}^{(j)}_{\cdot,l}$, denoted by $y^{(j)}_{i,l}$, is the energy consumption value of device j at interval i in day l, named as interval level consumption. We denote the aggregated energy consumption over all equipments as $\bar{\mathbf{Y}}$, $\bar{\mathbf{Y}} = \sum_{j=1}^{M} \mathbf{Y}^{(j)}$. The l^{th} column of $\bar{\mathbf{Y}}$ holds the aggregated consumption of the l^{th} day for a given household. The i^{th} element of $\bar{\mathbf{y}}_{\cdot,l}$, denoted as $\bar{y}_{i,l} = \sum_{j=1}^{M} y^{(j)}_{i,l}$, is the aggregated consumption at interval i in day l. At training phase, we assume that we can access the individual device consumption matrix $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(M)}$. At testing phase, we can only access the aggregated consumption matrix $\bar{\mathbf{Y}}$, and the goal is to decompose $\bar{\mathbf{Y}}$ into its components $\hat{\mathbf{Y}}^{(1)}, \hat{\mathbf{Y}}^{(2)}, \ldots, \hat{\mathbf{Y}}^{(M)}$.

Early work in power disaggregation focused on **high sample rate** electricity usage readings. A major work called Nonintrusive Appliance Load Monitoring (NALM) was proposed by Hart [6]. Hart identified the different power consumption signatures of various electrical appliances, and suggested using Finite State Machines (FSM) to model the signatures. A large amount of research have been invested to improve NALM. A pattern recognition approach proposed by Farinaccio and Zmeureanu [7] classified aggregated electricity consumption into its end uses. Laughman *et al.* suggested installing higher sample rate meters to collect higher resolution data, and then applying higher harmonics in the aggregated current signal to distinguish loads in the signature space [8]. Additionally, the voltage noise signatures were studied and characterized for classifying the operation of electrical appliances [9].

Recently, due to the broadly deployed smart meters producing aggregated readings at a low resolution, the research focus has shifted to disaggregating **low sample rate** residential power consumption. Kolter *et al.* proposed Discriminative Disaggregation Sparse Coding (DDSC) [10]. They defined the regularized disaggregation error as the objective function aiming at learning better basis functions (signatures) than regular sparse coding for separating aggregated readings. However, it's shown that many devices share similar basis functions especially when the number of parallel devices is large. This leads the failure of DDSC, since it cannot distinguish the set of devices with common basis functions. A Factorial Hidden Markov Model (FHMM) based framework has been presented through considering additional features, such as correlation between activities [11]. However, they only consider two working modes ("On" and "Off") of devices, which is obviously not practicable in real life. For example, portable washer usually works in "Soak", "Wash", "Rinse" and "Spin". Naturally, the FHMM-based method lacks the mechanism to handle complex power signatures.

Existing energy disaggregation techniques are incapable of embodying the intrinsic characters of low sample rate aggregated consumption, which requires the incorporation of certain essential features into the model. DDSC assumes there exist discriminative basis functions for identifying the consumption of devices. In fact, many devices have analogous consumption shapes, thus leading to devices share similar basis functions and causing the failure of disaggregation. FHMM-based methods assume that each appliance has two states ("on" and "off"), which cannot capture all functional modes of devices. The performance of FHMM-based method will be greatly reduced when dealing with complex consumption patterns. This motivates us to identify the inherent features from low sample rate data, and incorporate the discriminative features into a hierarchical model for disaggregation task.

3 Patterns Learning from Smart Meter Readings

In this section, we explore the discriminative features across devices for low sample rate aggregated power readings, and develop an algorithm for learning the features from data.

3.1 Evidence Collection By inspecting the daily energy readings of devices, we found that most devices work in various states, and the corresponding energy consumptions can significantly differ under different states. Take computer as an example, "Standby" state

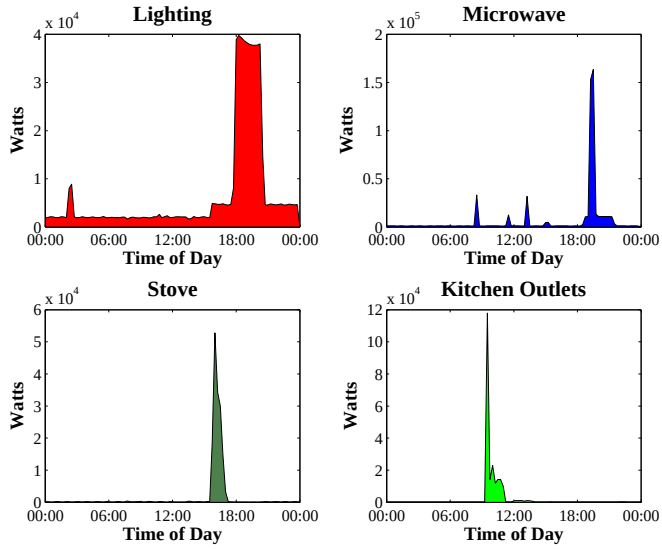

Figure 1: Daily energy consumption of typical equipments.

consumes only 10% of that of "High-performance" state, while "Screen-saver" state consumes about 33.3%. In addition, a computer in "Shut-down" state consumes about 3–5W per hour due to the indicator lights.

In Figure 1, we visualize the daily power consumption of four typical equipments. As shown in the figure, there are at least two states for any equipment in working condition. Specifically, we can use four states to model the functional modes of "Lighting". Most of the time, it consumes a small quantity of power per hour and we use State 1 to describe it. Early in the morning, there is a small peak indicating that the hosts might go to the bathroom and we use State 2 to depict it. Then in the evening, there are two other distinct states. We use State 3 to represent the low consumption rate, and use State 4 to trace the high consumption rate. Similarly, we can model "Microwave" with 4 states, model "Stove" with 2 states, and model "Kitchen Outlets" with 3 states.

In short, we found that most equipments function in multiple states. Since each state contains unique characteristics, it's important to identify these states and select appropriate features to distinguish the functional modes of devices.

3.2 Features Extraction Considering different states, we intend to discover the state-based features. Naturally, the interval level power consumption is one distinct characteristic for differentiating working conditions. However, many devices share similar interval level power consumption under different states. For instance, average power consumption of a one-ton air conditioner is almost the same as that of high power mode of microwave (about 1200 Watts). Fortunately, the start time and duration of microwave is much different from those of air conditioner. This inspires us to extract another set of features to characterize the devices and their working states.

Formally, we use $\mathbf{X}^{(j)}$ to denote the state matrix of device j. Each column is a state vector for one day, i.e., $\mathbf{x}_{:,l}^{(j)}$. The i^{th} element of $\mathbf{x}_{:,l}^{(j)}$, denoted as $x_{i,l}^{(j)}$, indicates the state of device j at interval i in day l. $x_{i,l}^{(j)} = 1, \ldots, K_j$, where K_j is the number of states for a particular device j. We employ the Non-Parametric Bayesian Clustering techniques (**NPBC**) proposed in [12] to discover the number of states K_j, and the state label for each interval consumption. We then extract the state-based start time and duration with consecutive constraints. Suppose vector $\mathbf{a}_k^{(j)}$ contains the start time of device j at state k, and vector $\mathbf{b}_k^{(j)}$ contains the duration of device j at state k, where $k = 1, \ldots, K_j$. The start time is the order number of the interval, at which the state begins (from 1 to N), and the duration is measured with the number of intervals spanned by the corresponding state (from 1 to N). To identify the start time and duration of a specific state for a given device, we break down the time series of device j at state k with consecutive constraints, and then store the start time and duration of state k respectively into $\mathbf{a}_k^{(j)}$ and $\mathbf{b}_k^{(j)}$. The details of the algorithm are shown in Algorithm 1.

Algorithm 1 Learning state-based Features

Input: $\mathbf{Y}^{(j)} \in R^{N \times L}\langle$ power consumption of devices$\rangle$, where $j = 1, \ldots, M$.

Output: $K_j\langle$the number of states for device $j\rangle$; $\mathbf{f}_k^{(j)}\langle$the power consumption of device j at state $k\rangle$; $\mathbf{a}_k^{(j)}\langle$ the start time vector of state k for device $j\rangle$; $\mathbf{b}_k^{(j)}\langle$ the duration vector of state k for device $j\rangle$.

1. **for** $j \leftarrow 1$ to M
2. $\quad [K_j, \text{cluster\_label}] \leftarrow \mathbf{NPBC}\left(\mathbf{Y}^{(j)}\right)$.
3. $\quad$ **for** $k \leftarrow 1$ to K_j
4. $\quad\quad \mathbf{f}_k^{(j)} \leftarrow \mathbf{Y}^{(j)}(\text{cluster\_label} = k)$.
5. $\quad\quad [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_O] \leftarrow$ break down $\mathbf{f}_k^{(j)}$ with consecutive constraints.
6. $\quad\quad \left[\mathbf{a}_k^{(j)}, \mathbf{b}_k^{(j)}\right] \leftarrow$ extract start time and duration from $[\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_O]$.
7. $\quad$ **end for**
8. **end for**

In Line 2, $\mathbf{NPBC}(\mathbf{Y}^{(j)})$ means to apply the Non-Parametric Bayesian Clustering algorithm to $\mathbf{Y}^{(j)}$. In

Line 4, $\mathbf{Y}^{(j)}(\text{cluster\_label} = k)$ means to extract the k^{th} cluster's power consumption from $\mathbf{Y}^{(j)}$. In Line 5, O indicates the number of sub-continuous time series, which are obtained after segmentation. The consecutive constraints are satisfied if and only if the original interval order number are adjacent. For example, if the original interval order number of $c_{1,1}$ is 10, then that of $c_{1,2}$ should be 11 (where $c_{1,1}$ and $c_{1,2}$ are the first two elements of $\mathbf{c}_1$). In Line 6, the start time is the interval order number of the first element respectively in $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_O$, while the duration is the length of $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_O$.

4 A Hierarchical Probabilistic Model

In this section, we propose a hierarchical probabilistic model that incorporates the state-based features with the working state and aggregated consumption.

4.1 Generative Model We begin by studying the distribution of power consumption of devices at a specific working state. Histograms of the consumption of four appliances are shown in Figure 2. Through examining the consumption patterns, we learned that under a particular state, the consumption varies in a small range with a peak in the center, which can be well captured by a Gaussian distribution. Based on the observation, we assume that the consumption of any device at any state follows a Gaussian distribution. Mathematically, for $\forall i = 1, \ldots, N, j = 1, \ldots, M, l = 1, \ldots, L, k = 1, \ldots, K_j$, we have,

$$\Pr\left(y_{i,l}^{(j)} \mid x_{i,l}^{(j)} = k\right) = \mathcal{N}\left(y_{i,l}^{(j)} \mid \mu_k^{(j)}, \nu_k^{(j)}\right), \tag{4.1}$$

where $y_{i,l}^{(j)}$ is the power consumption of device j at interval i in day l, $x_{i,l}^{(j)} = k$ indicates the current working state is k for device j at interval i of day l. $\mu_k^{(j)}$ and $\nu_k^{(j)}$ are the parameters governing the Gaussian distribution, which are respectively the mean and variance. Generally, the mean value of a state represents the average power consumption while the variance captures the transient features between the current state with its adjacent states.

We introduce two new random variables, $s_k^{(j)}$ and $r_k^{(j)}$, which respectively indicate the start time and duration of device j at state k. We set $\mathbf{s}^{(j)} = \left[s_1^{(j)}, s_2^{(j)}, \ldots, s_{K_j}^{(j)}\right]$, and $\mathbf{r}^{(j)} = \left[r_1^{(j)}, r_2^{(j)}, \ldots, r_{K_j}^{(j)}\right]$.

Since we are most concerned with the relationship between aggregated consumption and the state-based features, we develop a three-layer Bayesian probabilistic model to formalize the relation. As shown in Figure 3, the observed variables are represented with shaded nodes, while hidden variables are depicted with white nodes. The only observed variable is the aggregated consumption. The state-based start time and duration lie in Layer 1 (**Feature Layer**), which depicts the functional modes of devices. In Layer 2 (**State Layer**), they are the working states of devices, which depend on the start time and duration. If interval i lies between the start time of state k and the end time of state k (the end time is the start time plus the duration), then the state at interval i is k. The aggregated consumption for a particular interval i is located in Layer 3 (**Consumption Layer**), which is decided by the working states of all devices at interval i.

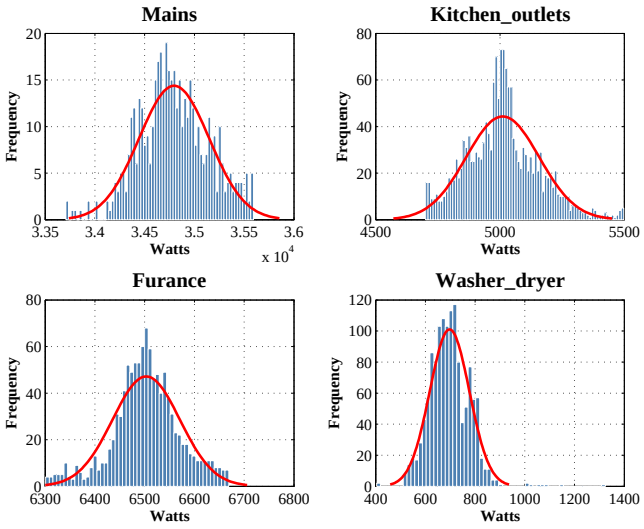

Figure 2: Histograms of the power consumption of appliances under a specific working mode.

We carefully analyze the start time and duration over intervals to identify their corresponding distributions. As shown in Figure 4, we present the start time and duration of "Stove" and "Washer_dryer" under a particular working state. Through inspecting the start time and duration of most devices, we learned that Gamma distribution can capture the frequent variations as it has more freedom in both distribution shape and scale.

Consequently, in Layer 1, the prior probability of the start time and duration respectively follows a Gamma distribution, and for $\forall j = 1, \ldots, M, k = 1, \ldots, K_j$, we have

$$\Pr\left(s_k^{(j)}\right) = \Gamma\left(s_k^{(j)} \mid \hat{\alpha}_k^{(j)}, \hat{\beta}_k^{(j)}\right), \tag{4.2}$$

$$\Pr\left(r_k^{(j)}\right) = \Gamma\left(r_k^{(j)} \mid \bar{\alpha}_k^{(j)}, \bar{\beta}_k^{(j)}\right), \tag{4.3}$$

where $\hat{\alpha}_k^{(j)}$ and $\hat{\beta}_k^{(j)}$ respectively denote the shape and rate parameter of start time for device j at state k;

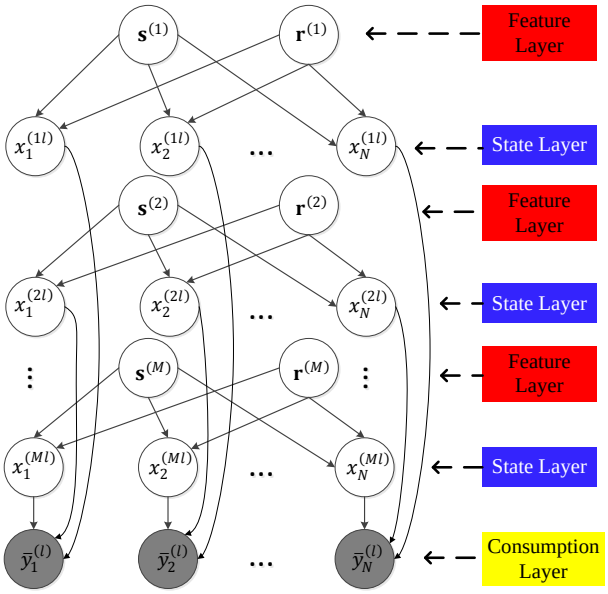

Figure 3: Graphical representation of the three-layer hierarchical probabilistic model.

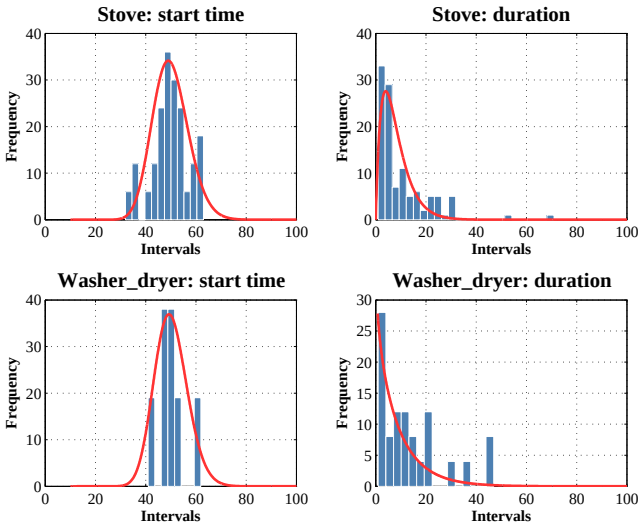

Figure 4: Histograms of the start time and duration of appliances under a specific working mode.

while $\bar{\alpha}_k^{(j)}$ and $\bar{\beta}_k^{(j)}$ respectively denote the shape and rate parameter of duration for device j at state k.

Formally, $\forall i = 1, \ldots, N, j = 1, \ldots, M, l = 1, \ldots, L$, $x_{i,l}^{(j)}$ is dependent on $\mathbf{s}^{(j)}$ and $\mathbf{r}^{(j)}$. For $\forall k = 1, \ldots, K_j$, we have

$$\begin{aligned}
&\Pr\left(x_{i,l}^{(j)} = k\right) = \Pr\left(s_k^{(j)} \leq i \leq s_k^{(j)} + r_k^{(j)}\right) \\
&= \Pr\left(s_k^{(j)} \leq i, r_k^{(j)} \geq i - s_k^{(j)}\right) \\
&= \frac{\gamma\left(\hat{\alpha}_k^{(j)}, \hat{\beta}_k^{(j)} \cdot i\right)}{\Gamma\left(\hat{\alpha}_k^{(j)}\right)} \\
&- \int_0^i \Pr(s_k^{(j)}) \frac{\gamma\left(\bar{\alpha}_k^{(j)}, \bar{\beta}_k^{(j)} \cdot \left(i - s_k^{(j)}\right)\right)}{\Gamma\left(\hat{\alpha}_k^{(j)}\right)} \mathrm{d}s_k^{(j)},
\end{aligned} \tag{4.4}$$

where $\gamma(x, y)$ denotes the lower incomplete gamma function. This equation states that for any i, j, l, if it lies between the start time $s_k^{(j)}$ and end time $s_k^{(j)} + r_k^{(j)}$, then the state of device j at interval i in day l is k. The derivation of Equation (4.4) can be found in [16].

As the values of $s_k^{(j)}$ and $r_k^{(j)}$ is restricted as integers between 1 and N, we employ the following mapping schema for scale purpose.

$$f\left(\delta_k^{(j)} = n\right) = \begin{cases} \int_{-\infty}^{1} \Pr\left(\delta_k^{(j)}\right) \mathrm{d}\delta_k^{(j)} & , n = 1 \\ \int_{n-1}^{n} \Pr\left(\delta_k^{(j)}\right) \mathrm{d}\delta_k^{(j)} & , 1 < n < N \\ \int_{N-1}^{+\infty} \Pr\left(\delta_k^{(j)}\right) \mathrm{d}\delta_k^{(j)} & , n = N \end{cases}, \tag{4.5}$$

where $n \in \mathbf{N}^\star$, $\delta_k^{(j)}$ is $s_k^{(j)}$ or $r_k^{(j)}$. Then, we revise the Equation (4.4) by replacing $\Pr\left(s_k^{(j)}\right)$ with $f\left(s_k^{(j)}\right)$, and substituting the integral operation with summation.

As the aggregated consumption over devices depends on the states of all devices, for $\forall i = 1, \ldots, N, j = 1, \ldots, M, l = 1, \ldots, L$, we have

$$\begin{aligned}
&\Pr\left(\bar{y}_{i,l} \mid x_{i,l}^{(1)} = k_1, x_{i,l}^{(2)} = k_2, \ldots, x_{i,l}^{(M)} = k_M\right) \\
&= \mathcal{N}\left(\bar{y}_{i,l} \mid \sum_{j=1}^{M} \mu_{k_j}^{(j)}, \sum_{j=1}^{M} \nu_{k_j}^{(j)}\right)
\end{aligned} \tag{4.6}$$

The derivation of Equation (4.6) can be found in [16].

4.2 Inference and Learning Based on the generative model, we perform inference to learn the hidden working states of devices, i.e., $\mathbf{X}^{(j)}, j = 1, \ldots, M$. Given any device $j = 1, \ldots, M, k = 1, \ldots, K_j$, parameters $\left\{\hat{\alpha}_k^{(j)}, \hat{\beta}_k^{(j)}\right\}, \left\{\bar{\alpha}_k^{(j)}, \bar{\beta}_k^{(j)}\right\}$ respectively for start time and duration could be derived from the extracted vectors $\mathbf{a}_k^{(j)}$ and $\mathbf{b}_k^{(j)}$. Moreover, the consumption related parameters $\left\{\mu_k^{(j)}, \nu_k^{(j)}\right\}$ can be derived from the consump-

tion records $\mathbf{f}_k^{(j)}$ at state k for device j.

During the course of testing, the only observations are the aggregated consumption over devices, i.e., $\bar{\mathbf{Y}}$. We need to learn the state matrix of all devices (i.e., $\mathbf{X}^{(j)}, j = 1, \ldots, M$), which specifies the interval level working mode of devices. We intend to identify the optimal values of $\mathbf{X}^{(j)}$ using the Maximum a Posterior (MAP) optimal estimator, which maximizes the state posterior $\Pr\left(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(M)} \mid \bar{\mathbf{Y}}\right)$. For a given $\bar{\mathbf{Y}}$, maximizing the posterior is equivalent to maximizing the joint probability $\Pr\left(\bar{\mathbf{Y}}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(M)}\right)$, thus

$$\left\{\mathbf{X}^{\star(1)}, \mathbf{X}^{\star(2)}, \ldots, \mathbf{X}^{\star(M)}\right\} = \underset{\left\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(M)}\right\}}{\operatorname{argmax}} \Pr\left(\bar{\mathbf{Y}}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(M)}\right) \tag{4.7}$$

The optimization problem (4.7) can be decomposed into $N \cdot L$ independent sub-problems, where N is the number of intervals in one day, and L is the number of days to be estimated. The sub-problem associated with interval i of the $l^{\mathtt{th}}$ day is given by

$$\begin{aligned}\left\{x_{i,l}^{\star(1)}, x_{i,l}^{\star(2)}, \ldots, x_{i,l}^{\star(M)}\right\} &= \underset{\left\{x_{i,l}^{(1)}, x_{i,l}^{(2)}, \ldots, x_{i,l}^{(M)}\right\}}{\operatorname{argmax}} \Pr\left(\bar{y}_{i,l}, x_{i,l}^{(1)}, x_{i,l}^{(2)}, \ldots, x_{i,l}^{(M)}\right) \\ &= \underset{\left\{x_{i,l}^{(1)}, x_{i,l}^{(2)}, \ldots, x_{i,l}^{(M)}\right\}}{\operatorname{argmax}} \left[\Pr\left(\bar{y}_{i,l} \mid x_{i,l}^{(1)}, x_{i,l}^{(2)}, \ldots, x_{i,l}^{(M)}\right) \times \prod_{j=1}^{M} \Pr\left(x_{i,l}^{(j)}\right)\right]\end{aligned} \tag{4.8}$$

Considering the fact that device j has K_j states, the theoretical computational cost increases exponentially with the number of devices M (Suppose the number of states for all devices are K, then the complexity is $K^M \cdot N \cdot L$).

This motivates us to propose a heuristic algorithm that can efficiently achieve a local optimum. We first remove the devices which do not satisfy the time and consumption constraints, then the rest of devices will be considered as combinational candidates in the following processes. We define a threshold to preclude searching for global optimum once a satisfactory local optimum is reached. The details of the method are shown in Algorithm 2.

In Line 2, CL represents the candidate list of devices that will be considered for decomposing the consumption of interval i. In **Phase 1: Construct the candidate list**, we use time and consumption constraints to rule out invalid devices. In Line 4, the outliers are detected using the algorithm proposed in [14] with respect to start time and duration constraints. If current i is an outlier under the start time and duration distributions (Gamma distribution) of working states for devices j, then we conclude that the device j is invalid at interval i. Similarly, in Line 5, we detect the invalid devices for interval i using consumption constraints. Based on the fact that $\bar{y}_{i,l} = \sum_{j=1}^{M} y_{i,l}^{(j)}$, $\bar{y}_{i,l}$ should be greater than or equal to any $y_{i,l}^{(j)}$. If for $\forall k = 1, \ldots, K_j$ (except the "off" state, where the consumption is near zero), we have $\bar{y}_{i,l} < \mu_k^{(j)} - 3\nu_k^{(j)}$, then we consider device j as invalid, because there is less than 0.15% probability of occurring in this interval. In **Phase 2: Search optimal solution**, we begin by setting the size of combination to be 1, then evaluate all possible combinations with the current size. During the course of evaluation, suppose the acquired minimum value of the joint probability shown in Equation (4.8) is m_1, and the corresponding maximum value is m_2. Then the convergence criteria in Line 19 is set as $\frac{m_1}{m_2} \leq \lambda$, ($\lambda$ is the predefined threshold). That is if the ratio of the minimum value over the maximum

Algorithm 2 Approximately Infer States

Input: $\bar{y}_{i,l}\langle$the aggregated power consumption over devices $\rangle$.

Output: $x_{i,l}^{\star(j)}\langle$the state for all devices$\rangle$, where $j = 1, \ldots, M\rangle$.

1. **for** $i \leftarrow 1$ to N
2. $\quad CL \leftarrow []$.

{**Phase 1: Construct candidate list**.}

3. $\quad$ **for** $j \leftarrow 1$ to M
4. $\quad\quad \mathcal{O}_1 \leftarrow$ Check if i is abnormal under time constraints of device j.
5. $\quad\quad \mathcal{O}_2 \leftarrow$ Check if $\bar{y}_{i,l}$ is abnormal under consumption constraints of device j.
6. $\quad\quad$ **if** $\mathcal{O}_1 =$ **false and** $\mathcal{O}_2 =$ **false**
7. $\quad\quad\quad CL \leftarrow [CL; j]$.
8. $\quad\quad$ **end if**
9. $\quad$ **end for**

{**Phase 2: Search optimal solution**.}

10. $\quad I_s \leftarrow 1$.
11. $\quad$ **repeat**
12. $\quad\quad$ Evaluate all possible combinations of devices in CL with size I_s.
13. $\quad\quad I_s \leftarrow I_s + 1$
14. $\quad$ **until convergence or** $I_s >$ **size**(CL)
15. **end for**

Table 1: Device level performance, which is reported as Precision Recall F-measure (each occupies one line).

Devices	HPMED	FHMM	DDSC
	92.33%	62.51%	15.42%
1 Mains_1	83.49%	14.89%	0.002%
	87.69%	24.06%	0.004%
	35.41%	45.19%	24.84%
2 Mains_2	91.76%	41.94%	0.01%
	51.10%	43.51%	0.02%
	63.29%	41.24%	100%
3 Kitchen_outlets	81.55%	74.73%	0.10%
	71.27%	53.15%	0.20%
	46.78%	8.76%	48.82%
4 Furnace	96.50%	99.57%	0.43%
	63.01%	16.11%	0.86%
	34.28%	37.92%	48.31%
5 Lighting	70.44%	47.39%	0.03%
	46.12%	42.13%	0.06%
	47.79%	59.96%	35.24%
6 Refrigerator	91.98%	41.06%	0.012%
	62.90%	48.74%	0.025%
	22.17%	10.11%	92.93%
7 Microwave	69.54%	99.59%	0.05%
	33.63%	18.36%	0.10%
	72.21%	7.16%	66.01%
8 Bathroom_gfi[1]	89.58%	100%	0.54%
	79.96%	13.36%	1.07%
	86.57%	96.96%	46.59%
9 Electronics	89.71%	59.49%	0.02%
	88.11%	73.73%	0.04%
	63.16%	4.36%	66.70%
10 Stove	90.77%	100%	9.30%
	74.49%	8.35%	1.83%
	64.42%	5.33%	66.22%
11 Washer_dryer	89.11%	100%	0.76%
	74.78%	10.13%	1.51%
	66.61%	7.30%	59.57%
12 Air_conditioning	89.61%	100%	0.48%
	76.42%	13.61%	0.95%

value is less than λ, then the search is terminated. Accordingly, the combination of the candidates with the maximum joint probability is selected as the optimal value $\left\{x_{i,l}^{\star(1)}, x_{i,l}^{\star(2)}, \ldots, x_{i,l}^{\star(M)}\right\}$.

Since the state matrix $\mathbf{X}^{(j)}$ has been derived, we can estimate the power consumption of individual device using $\Pr\left(y_{i,l}^{(j)} \mid x_{i,l}^{(j)} = k\right)$ as defined by Equation (4.1). The constraint on this estimation is $\sum_{j=1}^{M} y_{i,l}^{(j)} = \bar{y}_{i,l}$. As long as $y_{i,l}^{(j)}$ follows a Gaussian distribution, we apply the method proposed in [13], and perform the disaggregation task through approximating $\sum_{j=1}^{M} y_{i,l}^{(j)} = \bar{y}_{i,l}$ as $\sum_{j=1}^{M} y_{i,l}^{(j)} \leq \bar{y}_{i,l}$. This approach will enforce the sum consumption of all devices to be as close to the aggregated consumption as possible, while preserving the individual working state characteristics.

[1] gfi: ground fault interrupter

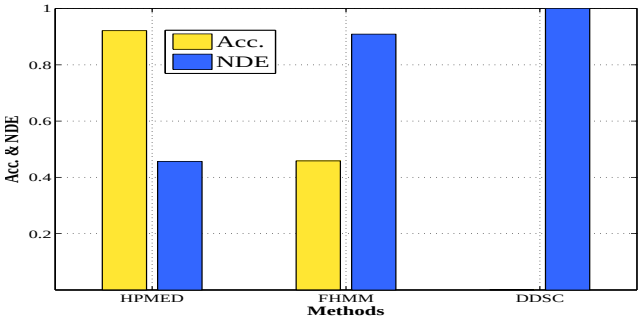

Figure 5: Whole-home level performance of HPMED, FHMM, and DDSC.

5 Experimental Results

In this section, we evaluate the performance of the hierarchical probabilistic model with extensive experiments across several homes.

5.1 Experiment Design

In this section, we describe the data set, and discuss the evaluation metrics.

Data Set: We employed a real-world dataset REDD [15], consisting of 1 Hz power readings for five real homes. Due to cost and privacy concerns, the broadly installed smart meters are not allowed to report readings in such a high granularity (typically as low as 1/900 Hz). We intend to promote the experimental environment towards real-life scenarios. Consequently, for every device j, we aggregate the 1 Hz power readings to be 1/900 Hz.

Evaluation Metrics: For the purpose of examining the performance of models in both global and local scenarios, we inspect the whole-home and device level evaluation metrics. We assessed the whole-home level disaggregation capability with accuracy and normalized disaggregation error. Accuracy measures the average consumption capability, abbreviated as Acc.

$$\text{Acc.} = \frac{\sum_{j,l} \min\left(\sum_i y_{i,l}^{(j)}, \sum_i \hat{y}_{i,l}^{(j)}\right)}{\sum_{i,l} \bar{y}_{i,l}}, \tag{5.9}$$

where $\hat{y}_{i,l}^{(j)}$ is the estimated consumption at the i^{th} interval in l^{th} day for device j. We used the Normalized Disaggregation Error (NDE) to measure the effect of the difference values between true and estimated data

$$\text{NDE} = \sqrt{\frac{\left(\sum_{i,j,l} \left\| y_{i,l}^{(j)} - \hat{y}_{i,l}^{(j)} \right\|\right)^2}{\left(\sum_{i,j,l} y_{i,l}^{(j)}\right)^2}} \tag{5.10}$$

From the above two equations, large accuracy indicates that estimated results almost covers all the true

data, and large normalized disaggregation error hints that the predicted data is significantly different from the true data.

The device level performance was evaluated with precision, recall, and F-measure, where the precision is the fraction of disaggregated consumption that is correctly classified, recall is the fraction of true device level consumption that is successfully separated, and F-measure is $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$.

Table 2: Performance across homes, which is reported as Acc. / NDE.

Home ID	Num. of Devices	HPMED	FHMM
Home 1	8	**0.9215 / 0.4567**	0.4583 / 0.909
Home 2	11	**0.9246 / 0.3311**	0.2773 / 0.9344
Home 3	12	**0.9868 / 0.8103**	0.3583 / 0.9079
Home 4	13	**0.9538 / 0.7081**	0.2658 / 0.9434
Home 5	15	0.9941 / 1.6717	0.3624 / 0.926

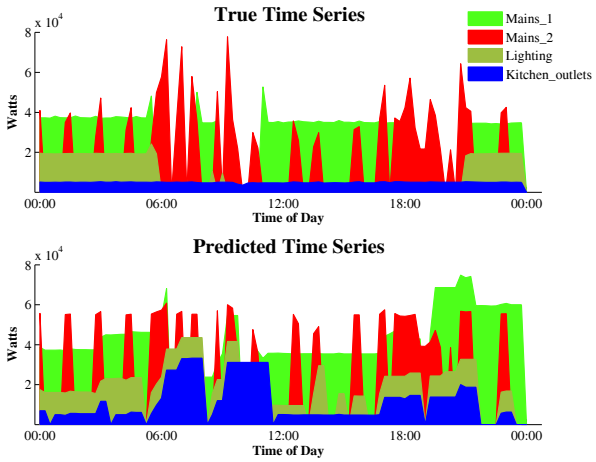

Figure 6: True (top), HPMED predicted (bottom) breakdown in device level for one day.

5.2 Performance Evaluation and Comparison

We compared HPMED with FHMM and DDSC, and applied HPMED to analyse the consumption distribution of households.

Device Level Disaggregation Performance Comparison HPMED, FHMM, and DDSC were assessed with the real data set, where 70% for training and 30% for testing. The household appliances were inspected and the results are shown in Table 1. HPMED outperformed both FHMM and DDSC on disaggregation task. Specifically, HPMED achieved as much as 37.02% higher than FHMM in average F-measure, and achieved 66.82% higher than DDSC. For some equipments, such as "Furnace", "Microwave" and "Bathroom_gfi", FHMM was capable of attaining a near 100% recall, but with low precision. This is caused by the fact that the predicted consumption is much larger than the true consumption induced by the inaccurate power signatures. DDSC failed completely on this task due to the lack of discriminative basis functions. Interestingly, for most devices, DDSC could achieve a fair or even high precision (e.g., more than 90% of precision in "Kitchen_outlets" and "Microwave"). However, the acquired recall is very low. The reason is that the estimated consumption is far smaller than true data leading to a high value in precision and a low value in recall.

We studied the whole-home level performance of HPMED, FHMM and DDSC measured with NDE and Acc. As shown in Figure 5, HPMED is capable of achieving more than 0.9 in Acc. and less than 0.5 in NDE, while FHMM only achieves 0.45 in Acc. with a rather large NDE (more than 0.9). As expected, DDSC has poor whole-home performance. Since DDSC consistently achieved a poor performance in both device and whole-home level measure, we dismiss it in the following discussions.

Overall Evaluation Across Homes To demonstrate the extensibility of the proposed model, we investigated the global performance of HPMED and FHMM across multiple homes, and conducted the comparison in terms of accuracy and NDE. Five homes were considered along with various appliances. As shown in Table 2, HPMED performed much better than FHMM for the first four homes in both accuracy and NDE. However, for the fifth home, FHMM outperformed HPMED in NED. The reason is that the prediction results of HPMED could cover almost all of the true data, but the difference value between the prediction and the ground truth is large. The accuracy of FHMM is much smaller than that of HPMED due to its inaccurate prediction results.

Another pattern shown in Table 2 is that HPMED is more robust than FHMM when the number of involved devices increases. For example, comparing the results of "Home 1" and "Home 2", we found that HPMED remains high performance despite the little increase of both accuracy and NDE, however, FHMM changes greatly with more than 18.00% decrease in accuracy and a small increase in NDE. This is because FHMM is more sensitive to the signatures of devices, which are significantly affected by the number of involved devices.

Consumption Distribution Analysis We analysed the consumption contribution of devices, and intended to provide users with the distribution information for

conservation purpose. We begin by visually inspecting a daily true appliance consumption along with the prediction results by HPMED.

To reveal the main contribution to the total consumption, we show the power usage distribution in Figure 7 (refer [16] to view the distribution analysis for other three homes). The left column shows the true distribution while the right column presents the predicted information by HPMED. For both "home 1" and "home 2", HPMED was capable of labelling the top appliances, {"Mains_1", "Mains_2"}, consuming more than 55% of the total for "home 1" and more than 60% for "home 2", in spite of the misclassification between these two devices. In addition to "Mains_1" and "Mains_2", HPMED properly separated other major classes, such as, "Kitchen_outlets_1", "Kitchen_outlets_2", "Lighting_1" and "Lighting_2" for "home1", and "Refrigerator" and "Microwave" for "home2".

In summary, we compared HPMED with DDSC and FHMM with a large scale of appliances, and showed that HPMED archived the best performance. Besides, we showed that HPMED is more robust than FHMM with respect to the number of involved devices. We also demonstrated that HPMED has competent capability in distribution analysis.

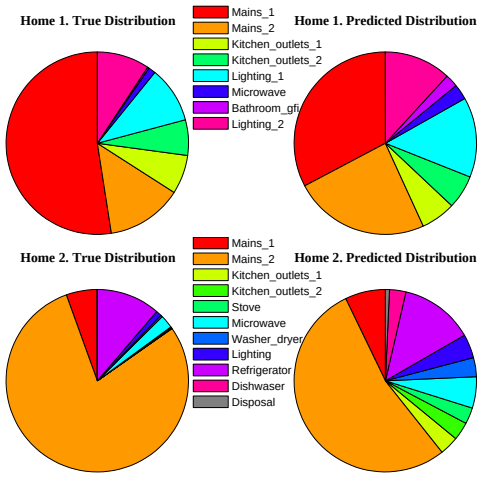

Figure 7: True (left), HPMED predicted (right) energy consumption distribution.

6 Conclusions

In this paper, we present a concrete solution to the disaggregation of low sample rate smart meter readings. We begin by deriving the temporal features start time and duration of functional modes for devices, which are practical for distinguishing devices. HPMED is proposed for the disaggregation task by incorporating the temporal features, functional modes, and the interval level consumption. Using low sample rate consumption data from real homes, we showed that HPMED outperformed other methods in both device and whole-home level performance measures, and was capable of accurately disaggregating aggregated readings into per-appliance usage information.

References

[1] J. Rubin, *Oil and The Death of Globalization*, http://peakoil.com/production/oil-and-the-death-of-globalization/, 2010.

[2] *Power Surge: Energy Use and Emissions Continue to Rise*, http://www.wri.org/publication/content/8601, World Resources, 1998.

[3] T. B. Hurst, *EIA Predicts 50% Increase in World Energy Consumption by 2030*, http://redgreenandblue.org/2008/06/30/eia-predicts-energy-50-increase-in-world-energy-consumption-by-2030/, World Resources, 1998.

[4] *Global Emissions by Source*, http://epa.gov/climatechange/ghgemissions/global.html, 2004.

[5] C. Fischer, *Feedback on Household Electricity Consumption: a Tool for Saving Energy?*, Energy Efficiency, vol. 1, pp. 79–104, 2008.

[6] G. W. Hart, *Nonintrusive Appliance Load Monitoring*, Proceedings of the IEEE, vol. 80, pp. 1870–1891, 1992.

[7] L. Farinaccio and R. Zmeureanu, *Using a Pattern Recognition Approach to Disaggregate the Total Electricity Consumption in a House into the Major End-Uses*, Energy and Buildings, vol. 30, pp. 245–259, 1999.

[8] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford and P. Armstrong, *Power Signature Analysis*, IEEE Power and Energy, vol. 1, pp. 56–63, 2003.

[9] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds and G. D. A. *At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line*, ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 271–288, 2007.

[10] J. Z. Kolter, S. Batra and A. Y. Ng., *Energy Disaggregation via Discriminative Sparse Coding*, Neural Information Processing Systems, pp. 1153–1161, 2010.

[11] H. Kim, M. Marwahy, M. Arlitt, G. Lyon and J. Han, *Unsupervised Disaggregation of Low Frequency Power Measurements*, SIAM Data Mining, pp. 747–758, 2011.

[12] P. Wang, C. Domeniconi and K. B. Laskey, *Nonparametric Bayesian Clustering Ensembles*, ECML PKDD'10, pp. 435–450, 2010.

[13] B. Fortz, M. Labbe, F. V. Louveaux and M. Poss, *The Knapsack Problem with Gaussian Weights*, Technical Report 592, GOM, Universite Libre de Bruxelles, 2009.

[14] M. J. Nooghabia, H. J. Nooghabia and P. Nasirib, *Detecting Outliers in Gamma Distribution*, Communications in Statistics–Theory and Methods, vol. 39, pp. 698–706, 2010.

[15] J. Z. Kolter and M. J. Johnson, *REDD: A Public Data Set for Energy Disaggregation Research*, SustKDD Workshop on Data Mining Applications in Sustainability, 2011.

[16] B. Wang, H. Dong, A. P. Boedihardjo, F. Chen and C.T. Lu, *Appendix: A Hierarchical Probabilistic Model for Low-Sample-Rate Home-Use Energy Disaggregation*, http://filebox.vt.edu/users/claren89/Publications/appendix_siam13.pdf.