Disambiguating Energy Disaggregation: A Collective Probabilistic Approach

Sabina Tomkins and **Jay Pujara** and **Lise Getoor**
University of California, Santa Cruz
{satomkin, jpujara, getoor}@ucsc.edu

Abstract

Reducing household energy usage is a priority for improving the resiliency and stability of the power grid and decreasing the negative impact of energy consumption on the environment and public health. Relevant and timely feedback about the power consumption of specific appliances can help household residents reduce their energy demand. Given only a total energy reading, such as that collected from a residential meter, *energy disaggregation* strives to discover the consumption of individual appliances. Existing disaggregation algorithms are computationally inefficient and rely heavily on high-resolution ground truth data. We introduce a probabilistic framework which infers the energy consumption of individual appliances using a *hinge-loss* Markov random field (HL-MRF), which admits highly scalable inference. To further enhance efficiency, we introduce a temporal representation which leverages state duration. We also explore how contextual information impacts solution quality with low-resolution data. Our framework is flexible in its ability to incorporate additional constraints; by constraining appliance usage with context and duration we can better disambiguate appliances with similar energy consumption profiles. We demonstrate the effectiveness of our framework on two public real-world datasets, reducing the error relative to a previous state-of-the-art method by as much as 50%.

1 Introduction

Households consume over one third of all electricity in the United States [Energy Information Administration, 2016], and opportunities for decreasing this share abound [Dietz *et al.*, 2009]. However, these opportunities are impeded by the lack of information available to residential consumers. Consumers are often uninformed as to which appliances consume the most energy [Gardner and Stern, 2008; Attari *et al.*, 2010], and which actions have the greatest savings potential. Furthermore, there is growing evidence that detailed feedback about energy use can reduce consumption [Brandon and Lewis, 1999; Stern, 1989; Darby, 2006; Van Houwelingen and Van Raaij, 1989]. Currently, this reduction is hampered by the fact that residents receive only aggregate energy information. As a simple analogy, consider receiving a shopping bill with a single figure and being asked to spend less on the next shopping trip. Based on this information alone, it would be difficult to discern how to adjust purchasing habits.

Advanced Metering Infrastructure (AMI), such as smart meters, measure aggregate energy consumption at defined intervals and transmit measurements wirelessly. Smart meters offer a unique opportunity to gather real-time data, learn energy consumption patterns, and ultimately offer actionable insights to consumers. Energy disaggregation (also referred to as non-intrusive load monitoring (NILM)) is the process of determining the energy consumption of individual appliances, given only an aggregated energy reading. A successful disaggregation algorithm can give consumers an itemized energy bill, displaying how much energy is consumed by each appliance, rather than the aggregate monthly reading they currently receive. While there are existing approaches to this problem [Kelly and Knottenbelt, 2015; Lange and Bergés, 2016; Parson *et al.*, 2012], none have been deployed in a real-world setting with low-frequency smart-meter readings.

Here, we propose a probabilistic energy disaggregation framework which determines the most likely collection of appliances that are in use. Our framework uses (soft) constraints and context to disambiguate between similar appliances. By introducing appliance sets we can infer appliance states collectively, allowing us to benefit from the inherent structure in this problem. We also introduce a representation for energy readings which encodes state duration directly. To formulate the inference task we use a *hinge-loss* Markov random field (HL-MRF) [Bach *et al.*, 2015], which allows a flexible probabilistic formulation and admits efficient inference. We evaluate our proposed framework on two real-world data sets. Empirical results demonstrate that our proposed probabilistic model significantly outperforms existing state-of-the-art techniques. In addition to our novel probabilistic formulation, we show the effectiveness of our proposed interval representation and the benefit of jointly modeling appliance states. Finally, we catalog several situations where contextual information is helpful for disambiguating active appliances.

2 Related Work

Hart's seminal paper [Hart, 1992] on non-intrusive load monitoring (NILM) introduced the problem of energy disaggregation. Broadly, there are two classes of solutions, those that require additional hardware and those that do not. Here we review only those approaches which do not require any additional equipment installation and thus are truly non-intrusive.

While alternative approaches exist [Kelly and Knottenbelt, 2015; Lange and Bergés, 2016], factorial hidden Markov models (FHMM)s [Ghahramani and Jordan, 1997] and variants thereof, have been a popular choice for disaggregation algorithms [Kolter and Jaakkola, 2012; Kim *et al.*, 2011; Parson *et al.*, 2012; Johnson and Willsky, 2013]. In this setting each appliance is represented with a single HMM, where the discrete hidden state variables correspond to the state of the appliance, and the observed continuous random variables correspond to the power readings. FHMMs allow multiple HMMs to be joined through a single observed variable in such a way that approximate inference is tractable. However, as the inference is approximate, it is not optimal, a shortcoming addressed by a number of recent papers [Makonin *et al.*, 2015; Shaloudegi *et al.*, 2016].

Our method is similar in spirit to the FHMM line of work as we also employ a structured probabilistic framework. However our HL-MRF structure is more flexible and can model a host of constraints without being restricted by the generative assumptions of an FHMM. Furthermore, we build on the work of Kim et al. [2011] and Li and Zha [2016], by integrating non-traditional context features. Like Shaloudegi et al. [2016], our approach exploits state-of-the-art optimization techniques such as ADMM [Boyd *et al.*, 2011], however our approach ends up being significantly more scalable, taking just a few minutes on homes from the REDD dataset versus over an hour for ADMM-RR.

3 Problem Definition

A disaggregation algorithm, $\mathbf{r} \rightarrow \mathbf{A}$, produces a mapping from a sequence of energy readings, $\mathbf{r}$, to a corresponding sequence of appliance states, $\mathbf{A}$. In the problem setting we consider, we define a sequence of n energy readings, $\mathbf{r} = \langle R_1, \ldots, R_n \rangle$. In addition, we are provided with a set of m appliances, $A = \{a_1, \ldots, a_m\}$. Each appliance a_i is associated with a set of k_i possible states. The number of states varies by appliance, some may have two states (*on*, *off*), while others may have multiple modes (*off*, *low*, *medium*, *high*). For each appliance state, we denote the energy consumed by appliance i in state k as c_i^k. We introduce an indicator $a_{i,j}^s$ that specifies that appliance i is in state s during reading j and define $\mathbf{A}$ as a matrix of these indicators. Each column is a vector of indicators corresponding to a particular reading, such that the j^{th} column has the form $\{a_{1,j}^{s_1}, \ldots, a_{1,j}^{s_{k_1}}, \ldots, a_{m,j}^{s_1}, \ldots, a_{m,j}^{s_{k_m}}\}$.

Using this model formalization, we introduce a probabilistic formulation for the disaggregation problem. We introduce a binary random variable $y_{i,j}^s$ corresponding to each indicator variable, $a_{1,j}^s$. The variable $y_{i,j}^s$ takes value 1 when appliance a_i is in state s during reading j, and 0 otherwise. The goal of disaggregation is to estimate the probability, $P(y_{i,j}^s)$, of each possible appliance state for each reading in the sequence, and determine the most probable state for each appliance, $y_{i,j}^{\hat{s}} = \underset{s \in \{1 \ldots k_i\}}{\operatorname{argmax}} P(y_{i,j}^s)$. This probability estimate should obey the constraint that expected consumption equals the measured usage in the reading, or $R_j = \sum_{i=1}^{m} c_i^s y_{i,j}^{\hat{s}}$.

Frequently, the task of estimating these probabilities is under constrained, such that many potential configurations of appliance states may yield the same observed energy consumption. Thus, disaggregation algorithms can improve identifiability by modeling the joint probability distribution over all appliance states for reading j, $P(Y_j)$. Many possible probabilistic models can be used to characterize this probability distribution. One common example is using previous appliance states to model current appliance states, e.g. estimating the conditional probability $P(Y_j|Y_{j-1})$. In the next section, we discuss how to build a robust probabilistic model of appliance states.

4 Modeling Approach

We propose a flexible framework which can disaggregate individual appliances from aggregate power readings. The framework is designed with real-world applicability as the end goal; it can adapt to multiple categories of information and is able to disaggregate even with coarse power readings. One of the framework's central goals is disambiguating between appliances which have similar power demands, but are used in different contexts for differing lengths of time. We introduce two conceptual representations: appliance sets and an interval temporal formulation.

4.1 Appliance Sets

Rather than predict the states of appliances independently, our model captures a joint configuration of appliances that we refer to as an *appliance set*. The energy consumption of an appliance set can easily be determined by aggregating the energy usage of each appliance. By grouping appliances together we are able to reason jointly about the relative likelihood of sets rather than individual appliances, for example, a heater and an air conditioner is an unlikely pair, while an air conditioner and pool pump is not. Here we only infer the states of those sets which are within a reasonable distance to the observed total power. Thus the input of target appliances to be disaggregated can be split into J unique subsets, corresponding to collections of feasible appliances. Then for any reading R_t there is a single set S_j such that all appliances in S_j are on at time t.

4.2 Interval Representation

In addition to modeling appliance sets, we also explore aggregating several instantaneous readings into an interval representation. If two consecutive readings are similar, it is likely that they correspond to the same appliance set. In this case, inferring the most likely appliance set for an interval rather than for each individual reading in the interval can improve the efficiency as well as the accuracy of the model. To define intervals, we coalesce readings where total power has only minimal fluctuations. When the difference in consumption of two consecutive readings exceeds a threshold δ, our model establishes a new interval. The sequence of power readings is

now indexed by intervals instead of time, so that we have V intervals to disaggregate, rather than N readings, $\langle R_1, \ldots, R_V \rangle$, where $V \ll N$.

5 Probabilistic Disaggregation Framework

Our framework integrates diverse sources of information into a joint probability distribution over active appliance sets. We model this probability distribution as a *hinge-loss* Markov random field (HL-MRF)[Bach *et al.*, 2015]. Finding the most probable appliance set for each reading corresponds to maximum a-posteriori (MAP) inference in the HL-MRF.

To specify an HL-MRF, we use the templating language Probabilistic Soft Logic (PSL). PSL has been successfully deployed in a diverse range of settings, from recommender systems [Kouki *et al.*, 2015] to stance prediction in online forums [Sridhar *et al.*, 2015]. PSL models are specified through weighted logical rules which capture dependencies between variables. Combining the PSL model with input data defines an HL-MRF.

Rules in PSL are expressed in first-order logic and associated with a weight, for example:

$$w_a : \text{ActiveInSet}(A_i, S_j) \wedge \text{AppSet}(R_l, S_j) \rightarrow \text{IsOn}(R_l, A_i).$$

Here the rule has weight w_a and predicates: AppSet, IsOn, and ActiveInSet capture the relationships between the variables: A_i, S_j, and R_l. A predicate and its arguments (either variables or constants), constitute an atom. Atoms in PSL rules have a truth value in the continuous interval $[0, 1]$, allowing a relaxation of Boolean logic.

We outline the rules which define our model, and in Section 5.6 we describe how these rules are used to define a HL-MRF which captures the probabilistic dependencies and constraints in our domain.

5.1 Energy Disaggregation Template

We introduce several logical predicates to capture important elements of our model. Here we consider two appliance states, on and off, however the extension to additional states is straightforward. We capture that appliance i is on during reading l via the atom $\text{IsOn}(R_l, A_i)$. We define a mapping between appliances and appliance sets using $\text{ActiveInSet}(A_i, S_j)$, which is true when appliance i is active in set j. To capture if the appliance set j is active during reading l, we use the atom $\text{AppSet}(R_l, S_j)$, where only one appliance set can be active at a time. In our model we consider all appliance sets, however, in practice we restrict the number of feasible appliance sets considered for any reading as explained in Section 6.1. Accordingly, we introduce the following constraint:

$$\sum_{j=1}^{J} \text{AppSet}(R_l, S_j) = 1.0.$$

5.2 Interval Duration

Appliance usage often follows consistent patterns, and understanding the length of time an appliance spends in each particular state provides a powerful disambiguating signal. In our model, we define the discrete duration classes: *very short*, *short*, *medium*, and *long*. The length of each duration is estimated from data and differs for each dataset, but for example, a very short duration is less than 4 minutes, while a long duration would be more than 19 minutes. Duration classes were found from quartiles of all appliance durations in the training data, for example the threshold under which a duration would be labeled *very short* was the 25% percentile of all appliance on-state durations.

Using the atom $\text{Duration}(R_l, length(R_l))$, we can specify the duration of interval R_l. We can then learn a duration-specific prior for each appliance:

$$w_{dur} : \text{Duration}(R_l, length(R_l)) \rightarrow \neg\text{IsOn}(R_l, A_i).$$

These learned priors can capture patterns such as microwaves rarely being on for long durations and dishwashers rarely being on for very short durations.

5.3 Observed Consumption

The difference between an appliance set's consumption and the observed power reading is modeled with the atom $\text{CloseToConsumption}(R_l, S_j)$. Let the expected consumption of S_j, be $E_j = \sum_{i \in S_j} \mu_i$, $\text{reading}(R_l)$, be the value of reading R_l in watts. Then, $\text{CloseToConsumption}(R_l, S_j)$ equals,

$$1 - \min\left(1, \frac{|\text{reading}(R_l) - E_j|}{\max(E_j, \text{reading}(R_l))}\right).$$

The weighted rule below expresses that the truth value of an appliance set depends on its distance to consumption.

$$w_c : \text{CloseToConsumption}(R_l, S) \rightarrow \text{AppSet}(R_l, S)$$

To express the relationship between appliances and appliance sets we introduce the predicate $\text{ActiveInSet}(A_i, S_j)$, which is 1 if A_i is active in S_j, and 0 otherwise. Thus to propagate information about appliances to appliance sets, and vice versa we use two rules:

$$w_{app} : \text{ActiveInSet}(A_i, S_j) \wedge \text{AppSet}(R_l, S_j) \rightarrow \text{IsOn}(R_l, A_i)$$

$$w_{as} : \text{ActiveInSet}(A_i, S_j) \wedge \neg\text{IsOn}(R_l, A_i) \rightarrow \neg\text{AppSet}(R_l, S_j)$$

5.4 Capturing State Changes

One expectation of energy readings is that changes in observed energy usage correspond to a single appliance changing state, rather than a significant change in the active appliances. Implementing this intuition using appliance sets requires capturing the relationships between appliance sets more directly.

We specify the difference between two consecutive readings using the atom, $\text{Diff}(R_1, R_2, \text{D})$, and apply a threshold $D > \delta$ to generate only meaningful differences. We introduce the predicate Positive for differences greater than 0, and the atom $\text{CloseToDiff}(A, \text{D})$ to capture the distance between the energy consumption of appliance A and the observed difference D. Finally Toggle states that the difference between two appliances sets is exactly appliance A, that is $(S_1 \cup S_2) \setminus (S_1 \cap S_2) = A$. We then put these into two final rules:

$$\begin{aligned} w_{ton} : \quad & \text{Diff}(R_1, R_2, \text{D}) \wedge \text{Precedes}(R_1, R_2) \\ & \wedge \ \text{CloseToDiff}(A, \text{D}) \wedge \neg\text{IsOn}(R_1, A) \wedge \text{Positive}(\text{D}) \\ & \wedge \ \text{Toggle}(S_1, S_2, A) \wedge \text{AppSet}(R_1, S_1) \rightarrow \text{AppSet}(R_2, S_2) \end{aligned}$$

w_{toff} : $\text{Diff}(R_1, R_2, \text{D}) \wedge \text{Precedes}(R_1, R_2)$
$\wedge$ $\text{CloseToDiff}(A, \text{D}) \wedge \text{IsOn}(R_1, A) \wedge \neg\text{Positive}(\text{D})$
$\wedge$ $\text{Toggle}(S_1, S_2, A) \wedge \text{AppSet}(R_1, S_2) \rightarrow \text{AppSet}(R_2, S_1)$

Additionally we create a set of rules to capture the persistence of an appliance being on. To do so we use a predicate $\text{Precedes}(R_i, R_j)$, which is true if index value i, directly precedes index value j, or $i = j - 1$. The following rules allow us to express the probability that an appliance contributes to R_l it will contribute to R_{l+1}.

$w_{stay_{on}} : \text{IsOn}(R_l, X) \wedge \text{Precedes}(R_l, R_{l+1}) \rightarrow \text{IsOn}(R_{l+1}, A_i)$
$w_{turn_{off}} : \text{IsOn}(R_l, X) \wedge \text{Precedes}(R_l, R_{l+1}) \rightarrow \neg\text{IsOn}(R_{l+1}, A_i)$
$w_{turn_{on}} : \neg\text{IsOn}(R_l, X) \wedge \text{Precedes}(R_l, R_{l+1}) \rightarrow \text{IsOn}(R_{l+1}, A_i)1)$
$w_{stay_{off}} : \neg\text{IsOn}(R_l, X) \wedge \text{Precedes}(R_l, R_{l+1}) \rightarrow \neg\text{IsOn}(R_{l+1}, A_i)$

5.5 Contextual Rules

Contextual rules are designed to capture the context in which a resident uses a given appliance. By developing a rich sense of context we reduce reliance on ground truth data, and introduce information which reduces the variance across appliances. Here we introduce two types of contextual information, temporal and temperature.

Temporal Rules

Appliance usage often depends on the time of day, and day of the week. For example, it is more likely that a cooking appliance, such as a microwave, will be used in the evening, than in the middle of the night. Thus we introduce two predicates which state the hour and day of the week at which a reading occurred: $\text{Hour}(R_l, H)$ and $\text{DayOfWeek}(R_l, D)$, where $H \in \{0, 23\}$ and $D \in \{Sunday...Monday\}$. We then learn the relationships between hour, day of the week, and appliance.

$$w_{day} : \text{DayOfWeek}(R_l, D) \rightarrow \text{IsOn}(R_l, A_i)$$
$$w_{hour} : \text{Hour}(R_l, H) \rightarrow \text{IsOn}(R_l, A_i)$$

Temperature Rules

We also model the relationship between temperature and appliance usage. Such a relationship should be particularly strong with heating and cooling appliances such as an air conditioner. To incorporate temperature into the model we introduce the predicate $\text{Temperature}(R_i, \text{Temp})$, where Temp is either *cold*, *mild* or *hot*. Thus we relate appliances to temperature with the following rule:

$$w_{temp} : \text{Temperature}(R_l, \text{Temp}) \rightarrow \text{IsOn}(R_l, A_i)$$

This concludes our overview of the energy disaggregation rules, we now explain how these rules can define a HL-MRF.

5.6 From Disaggregation Templates to HL-MRFs

HL-MRFS are a general class of conditional, continuous probabilistic models, parametrized with a set of weighted *hinge-loss* functions. Hinge-loss functions can model a rich diversity of relationships, and critically, admit highly scalable inference. We now specify how a HL-MRF is defined from a set of weighted logical rules, such as those defined in the probabilistic disaggregation framework. Let us now turn to an example rule from the previous section:

$$\lambda : \text{ActiveInSet}(A_i, S_j) \wedge \text{AppSet}(R_l, S_j) \rightarrow \text{IsOn}(R_l, A_i)$$

where λ is a weight, ActiveInSet, AppSet, and IsOn are predicates, and A_i, S_j, and R_l are all variables. By substituting constants, a_i, s_j, and r_l for the variables, A_i, S_j, and R_l respectively, one obtains three ground atoms: $\text{ActiveInSet}(a_i, s_j)$, $\text{AppSet}(r_l, s_j)$, and $\text{IsOn}(r_l, a_i)$, such that each ground atom takes a value in $[0, 1]$. Suppose that we let the atoms $\text{ActiveInSet}(a_i, s_j)$, $\text{AppSet}(r_l, s_j)$, and $\text{IsOn}(r_l, a_i)$ correspond to three random variables: $y_1, y_2,$ and y_3 respectively.

We would like to assign values to these variables, and to do so we find the values for each which minimize the distance to satisfaction of the participating rule, which is formulated as a hinge-loss. Given two continuous truth values $q, r \in [0, 1]$ a conjunction of q and r is defined as $q \wedge r = max\{q + r - 1, 0\}$. Finally, using the formula $q \rightarrow r = \neg q \vee r$, we arrive at a weighted *hinge-loss* potential. For the above rule this corresponds to:

$$\lambda \cdot max\{y_1 + y_2 - y_3 - 1, 0\}.$$

We can now find the a values for y_1, y_2, and y_3 which minimize this loss.

We have just seen how a rule specified by a PSL model defines a weighted hinge-loss function. To define a HL-MRF we input the set of weighted logical rules defined in PSL, along with our observed variables. The values to the weights can be set from expert knowledge, or learned from data. In this model we learn the values of the weights using training data. We can then jointly infer the maximum a posteriori (MAP) assignments to the unobserved variables. This problem has a tractable convex formulation which is solved directly by the PSL software [1], using the alternating direction method of multipliers ADMM [Boyd *et al.*, 2011].

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors of observed and unobserved continuous random variables in $[0, 1]$, respectively. Formally, a HL-MRF describes the following conditional probability density function over $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$P(\boldsymbol{y}|\boldsymbol{x}) \propto exp\left(-\sum_{j=1}^{m} w_j \phi_j(\boldsymbol{y}, \boldsymbol{x})\right)$$

Where ϕ_j is a *hinge-loss* potential, $\phi_j = \max\{l_j(\boldsymbol{x}, \boldsymbol{y}), 0\}^p, p \in \{1, 2\}$, l_j is a linear function of $\boldsymbol{x}$ and $\boldsymbol{y}$ and w_j is the positive weight associated with ϕ_j.

Let $\mathbf{y}$ be a vector of $IsOn(R_l, A_i)$ variables for all V intervals and all N appliances, and let $\mathbf{s}$ be a vector of $AppSet(R_l, S_j)$ variables for all V intervals and all J appliance sets. Let x denote all atoms which are observed, and for which the truth values are thus known. We then formulate our task as finding the MAP assignments to $\mathbf{y}$ and $\mathbf{s}$ under the probability distribution defined by the HL-MRF formulation of the rules described in 5.1:

$$\underset{\mathbf{y},\mathbf{s}}{\text{argmax}}\, P(\mathbf{y}, \mathbf{s}|x).$$

In the next section we evaluate the performance of this framework on two real-world datasets.

[1] `http://psl.linqs.org`

6 Empirical Evaluation

We evaluate our proposed disaggregation framework[2] on two real-world datasets. We demonstrate the effectiveness of our framework compared to ADMM-RR [Shaloudegi *et al.*, 2016], a recent, state-of-the-art approach. We also explore the situations under which contextual information can help improve model performance.

6.1 Data

We evaluate on two public datasets: The Reference Energy Disaggregation Dataset (REDD) [Kolter and Johnson, 2011] and Pecan Street Inc. (DATAPORT) [2016].

REDD: This is the most widely used dataset for energy disaggregation. The dataset describes six homes, and each home has an average of 21 days. The dataset contains fine-grained meter readings at approximately six second intervals. Following Makonin et al. [2015] and Johnson et al. [2013], we evaluate our model using homes 1, 2, 3, and 6 from the dataset and likewise, we disaggregate the refrigerator, lights, microwave, and dishwasher. Unlike previous work, we omit the furnace as it appears in only a single home, and for home 6 we omit the dishwasher.

DATAPORT: We also evaluate on the Pecan Street dataset which describes meter readings from Austin, Texas. This data set is much larger than REDD (hundreds of homes over several years) and is less well-studied in the context of disaggregation. The readings are coarser, available at either 1 minute or 1 hour intervals. We evaluated on the eight most common appliances: air conditioner, furnace, refrigerator, dishwasher, kitchen outlet, dryer, microwave and clotheswasher. As not all homes have perfect data records; here we choose the five homes (2859,3413,6990,7951,8292) for which there was at least one year of complete data (no records missing for any of the target appliances) and which had completed surveys describing demographics and household features. The models for each home (including the weights) are trained using the first 50% of the data, the next 25% of the data is used as a validation set for model parameters, and the model was evaluated on the final 25% of the data. With DATAPORT this training, validation, and testing is done for each month separately. For both datasets, we estimate the mean and standard deviation of the power consumption for each appliance based on only the training data. Evaluation estimates used both training and validation data. To find the thresholds to partition duration lengths into *very short*, *short*, *medium*, and *long*, we found quartiles for the interval lengths, such that 25% of all duration lengths were assigned to each duration. Similarly, each temperature was uniformly partitioned into one of three categorical labels, corresponding to *cold*, *mild*, and *hot*.

To partition the data into intervals, for each home we let δ be the difference between the average draw of a small appliance less its standard deviation. For REDD homes the small appliance was either the lights or the refrigerator, and for DATAPORT it was always the refrigerator. To assign feasible appliance sets for each interval, we compute the absolute difference between the mean consumption for each appliance set and the observed power, scaled to be in $[0, 1]$ and we then retain only those sets which are within 0.5 of the actual power consumption. If there are no such sets, we select the top three closest sets.

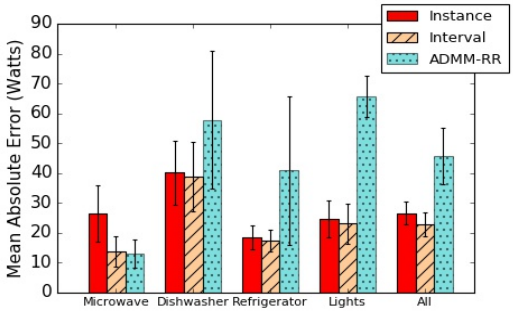

Figure 1: MAE on REDD data.

	Precision	Recall	F-Measure
Dishwasher	0.518	0.598	0.555
Lights	0.708	0.813	0.757
Microwave	0.707	0.712	0.709
Refrigerator	0.851	0.879	0.865
Average	0.696	0.751	0.722

Table 1: Performance of Interval Model on REDD data.

6.2 Results

To estimate appliance consumption, both the appliance state and the power consumption must be inferred. In order to evaluate performance, we measure both at how well an algorithm can predict the appliance states (precision, recall and F-measure), and how well it predicts actual consumption (Mean Absolute Error (MAE)). We evaluate three methods:

Instance: This model treats each instance independently, models the appliance sets jointly and does not use duration information.

Interval: This is the model described in Section 5 where we model both appliance sets and their durations.

ADMM-RR: This is the current state of the art method which uses a factorial HMM model [Shaloudegi *et al.*, 2016]. We use the code they provide online.[3]

Fig. 1 shows the MAE of the three methods on the REDD data set, and Table 1 shows the precision, recall and F-measure, for each appliance using the interval model. Fig. 2 provides a detailed view of the difference between observed and estimated energy usage for one specific home. The interval model performs the best overall, reducing the MAE of ADMM-RR by 50%.

The next set of results are on the DATAPORT data set. This dataset includes contextual features not available in the REDD dataset, so in addition to comparing the instance and interval approaches, we explore a model which captures contextual temporal and temperature features. Fig. 3 compares the MAE of three PSL-based methods with the ADMM-RR base-

[2]Code available: `https://bitbucket.org/linqs/appliance_disambiguation`

[3]To learn the required parameters we used the Matlab HMM toolbox [Murphy, 1998] and the first 75% of the data for each home.

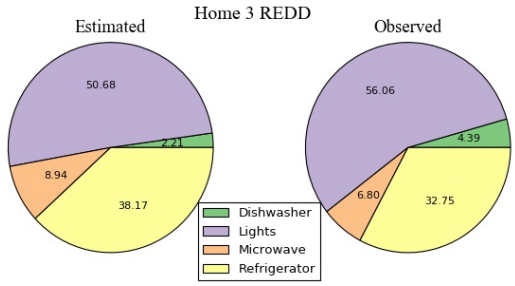

Figure 2: Percentage of total energy consumption of each appliance for a representative REDD home; predictions from the Interval model.

	Precision		Recall		F-Measure	
	Interval	+Context	Interval	+Context	Interval	+Context
Air Conditioner	0.901	0.899	0.815	0.823	0.856	0.859
Clotheswasher	0.226	0.228	0.333	0.274	0.269	0.249
Dishwasher	0.063	0.072	0.360	0.368	0.108	0.121
Dryer	0.591	0.571	0.731	0.749	0.653	0.648
Furnace	0.844	0.829	0.621	0.645	0.716	0.726
Kitchen Appliance	0.045	0.046	0.426	0.358	0.081	0.082
Microwave	0.330	0.354	0.394	0.394	0.359	0.373
Refrigerator	0.675	0.675	0.806	0.828	0.735	0.744
Average	0.459	0.459	0.561	0.555	0.505	0.503

Table 2: Performance of the Interval and Context Models on DATAPORT data. The effect of context varies by home and appliance.

line on the DATAPORT dataset, and Table 2 shows the precision, recall, and F-measure for each appliance for the interval and context-based models. A paired t-test demonstrates the context-based PSL model provides a statistically significant improvement over the interval model on MAE. Table 2 reveals that the effect of context differs by appliance, with the contextual models performing best on the appliances with high energy consumption. Fig. 4 provides a more detailed view of the difference between observed usage and model estimates for one specific home.

To understand the value of different contextual signals, we examine appliances for which temperature and temporal information has the greatest benefit. Contextual rules on temperature improved the predictions for heating and cooling appliances, with statistically significant improvements for the refrigerator and air conditioner. Temporal information improved average predictions for appliances with periodic usage patterns, such as the clotheswasher, dishwasher, dryer, and microwave.

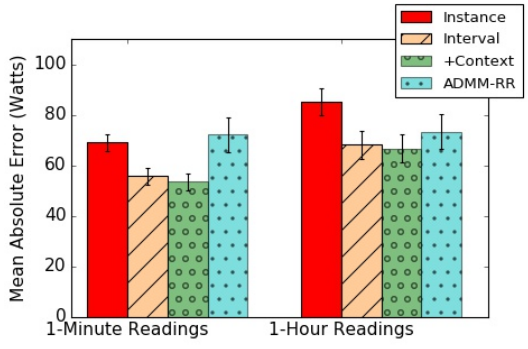

Figure 3: The MAE for Instance, Interval and Context Models on DATAPORT data, with both 1-minute and 1-hour readings. The contextual information provides a statistically significant improvement for readings at both time resolutions.

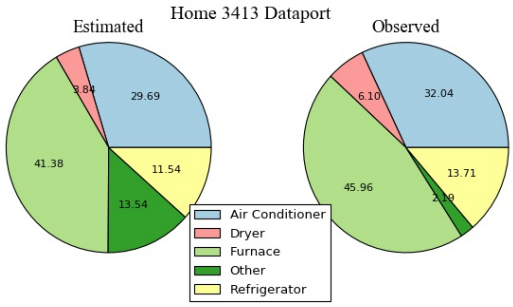

Figure 4: Percentage of total energy consumption of each appliance for a representative REDD home; predictions from the Interval model.

7 Discussion

Our proposed disaggregation framework achieves state-of-the-art performance on two real-world datasets, reducing error by 50% and 25% on REDD and DATAPORT datasets, respectively. Our algorithm adapts to different granularities of data, from the 6 second samples of REDD, to the 1-minute and 1-hour samples from DATAPORT. Not surprisingly, the DATAPORT dataset is more challenging, as we disaggregate more appliances at a rougher resolution, yet the F-measure for the most energy-intensive appliances remains above 0.7. Across datasets, we demonstrate that the interval representation, which aggregates readings into usage events, improves performance over a purely instance-level representation even when readings have low granularity. Beyond representation, one strength of our approach is the ability to incorporate different types of contextual information. Contextual information provides a small reduction in error and targeted improvements in predictive performance, however we had anticipated more pronounced improvements. One potential explanation is the limited range of contextual information due to the small temperature range in the dataset. In future work, we hope to pursue a deeper exploration of contextual information and its efficacy at generalizing predictions across homes.

While smart meters have been installed in homes across the United States, their potential in reducing consumer energy consumption is far from realized. Improved energy disaggregation algorithms can help to reach that potential by discovering appliance-level consumption patterns that consumers need to make informed decisions about their energy usage. In the paper, we propose such an algorithm, which is efficient, scalable, and readily adapted to new sources of information. By performing inference over intervals, and modeling collections of feasible appliance sets, we reduce the complexity of the problem while retaining the advantages of a structured probabilistic formulation. A key advantage of our framework is that additional information can be incorporated easily. For example, we could enrich our models by incorporating richer user and building profiles.

References

[Attari *et al.*, 2010] Shahzeen Z. Attari, Michael L. DeKay, Cliff I. Davidson, and Wndi Bruine de Bruin. Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences (PNAS)*, 2010.

[Bach *et al.*, 2015] Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. Hinge-loss Markov Random fields and probabilistic soft logic. *ArXiv:1505.04406 [cs.LG]*, 2015.

[Boyd *et al.*, 2011] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 2011.

[Brandon and Lewis, 1999] Gwendolyn Brandon and Alan Lewis. Reducing household energy consumption: A qualitative and quantitative field study. *Journal of Environmental Psychology*, 1999.

[Darby, 2006] Sarah Darby. The effectiveness of feedback on energy consumption. A Review for DERFA of the Literature on Metering, Billing, and direct Displays. Technical report, Oxford University, 2006.

[Dietz *et al.*, 2009] Thomas Dietz, Gerald T. Gardner, Jonathan Gilligan, Paul C. Stern, and Michael P. Vandenbergh. Household actions can provide a behavioral wedge to rapidly reduce us carbon emissions. *Proceedings of the National Academy of Sciences (PNAS)*, 2009.

[Energy Information Administration, 2016] Energy Information Administration. Electric power annual, 2016.

[Gardner and Stern, 2008] Gerald T. Gardner and Paul C. Stern. The short list: The most effective actions U.S. households can take to curb climate Change. *Environment: Science and Policy for Sustainable Development*, 2008.

[Ghahramani and Jordan, 1997] Zoubin Ghahramani and Michael I. Jordan. Factorial Hidden Markov Models. *Machine Learning*, 1997.

[Hart, 1992] George W. Hart. Nonintrusive Appliance Load Monitoring. *Proceedings of the IEEE*, 1992.

[Johnson and Willsky, 2013] Matthew J. Johnson and Alan S. Willsky. Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research (JMLR)*, 2013.

[Kelly and Knottenbelt, 2015] Jack Kelly and William Knottenbelt. Neural NILM: Deep neural networks applied to energy Disaggregation. In *ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys)*, 2015.

[Kim *et al.*, 2011] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, and Jiawei Han. Unsupervised disaggregation of low frequency power measurements. In *SIAM International Conference on Data Mining (SDM)*, 2011.

[Kolter and Jaakkola, 2012] J. Zico. Kolter and Tommi Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

[Kolter and Johnson, 2011] J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. *KDD Workshop on Data Mining Applications in Sustainability*, 2011.

[Kouki *et al.*, 2015] Pigi Kouki, Shobeir Fakhraei, James Foulds, Magdalini Eirinaki, and Lise Getoor. HyPER: A flexible and extensible probabilistic framework for hybrid recommender systems. In *ACM Conference on Recommender Systems (RecSys)*, 2015.

[Lange and Bergés, 2016] Henning Lange and Mario Bergés. BOLT: Energy disaggregation by online binary matrix factorization of current waveforms. In *ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys)*, 2016.

[Li and Zha, 2016] Liangda Li and Hongyuan Zha. Household structure analysis via Hawkes processes for enhancing energy disaggregation. In *International Conference on Artificial Intelligence (IJCAI)*, 2016.

[Makonin *et al.*, 2015] Stephen Makonin, Fred Popowich, Ivan V. Bajic, Bob Gill, and Lyn Bartram. Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. *IEEE Transactions on Smart Grid*, 2015.

[Makonin, 2014] Stephen William Makonin. *Real-Time Embedded Low-Frequency Load Disaggregation*. PhD thesis, Simon Fraser Univerity, 2014.

[Murphy, 1998] Kevin Murphy. Hidden Markov Model (HMM) Toolbox for Matlab, 1998.

[Parson *et al.*, 2012] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Artificial Intelligence (AAAI)*, 2012.

[Pecan Street Inc., 2016] Pecan Street Inc. Dataport, 2016.

[Shaloudegi *et al.*, 2016] Kiarash Shaloudegi, András György, Csaba Szepesvari, and Wilsun Xu. Sdp relaxation with randomized rounding for energy disaggregation. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[Sridhar *et al.*, 2015] Dhanya Sridhar, James Foulds, Marilyn Walker, Bert Huang, and Lise Getoor. Joint models of disagreement and stance in online debate. In *Association for Computational Linguistics (ACL)*, 2015.

[Stern, 1989] Paul C. Stern. Contributions of Psychology to Limiting Climate Change. *American Psychologist*, 1989.

[Van Houwelingen and Van Raaij, 1989] Jeannet H Van Houwelingen and W Fred Van Raaij. The effect of goal-setting and daily electronic feedback on in-home energy use. *Journal of Consumer Research*, 1989.

[Zeifman, 2012] Michael Zeifman. Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics*, 2012.