Energy Usage Behavior Modeling in Energy Disaggregation via Marked Hawkes Process

Liangda Li[1,2] and **Hongyuan Zha**[1,2]
[1]Software Engineering Institute, East China Normal University, Shanghai, China
[2]College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Energy disaggregation, the task of taking a whole home electricity signal and decomposing it into its component appliances, has been proved to be essential in energy conservation research. One powerful cue for breaking down the entire household's energy consumption is user's daily energy usage behavior, which has so far received little attention: existing works on energy disaggregation mostly ignored the relationship between the energy usages of various appliances across different time slots. To model such relationship, we combine topic models with Hawkes processes, and propose a novel probabilistic model based on marked Hawkes process that enables the modeling of marked event data. The proposed model seeks to capture the influence from the occurrence and the marks of one usage event to the occurrence and the marks of subsequent usage events in the future. We also develop an inference algorithm based on variational inference for model parameter estimation. Experimental results on both synthetic data and three real world data sets demonstrate the effectiveness of our model, which outperforms state-of-the-art approaches in decomposing the entire consumed energy to each appliance. Analyzing the influence captured by the proposed model provides further insights into numerous interesting energy usage behavior patterns.

Introduction

Energy conservation has become a critical issue in modern society and data analysis methodology has recently been applied to the analysis of energy consumption patterns in households. Several prior studies (Darby 2006; Neenan and Robinson 2009; Wytock and Kolter 2014) have shown that consumers, i.e., household members are more likely to conserve their energy usage when provided with breakdown energy consumption records. However, such fine-grained energy consumption data is not readily available, since it requires numerous additional meters installed on individual appliances. Therefore there has been much interest in the data analysis problem of energy disaggregation — the task of taking a whole-house energy signal and separating it into its component appliances. One powerful cue for breaking down the entire household's energy consumption is user behavior in energy usage(Baptista et al. 2014), which is known to be a major factor in determining the energy consumption in households. Such *energy usage behavior*s can include: how users perform their daily routines, how they share the usage of appliances, and users' habits in using certain types of appliances. Understanding such energy usage behaviors will significantly increase the accuracy of estimating the usage time of each appliance, which consequently benefits the energy disaggregation task.

Despite of the importance of energy usage behaviors, they have not received enough attention in the recent literature, especially how a user's current energy usage behavior influences his/her or other people's future usage behavior. Modeling such influence is important due to the following two reasons: 1) energy usage behaviors rarely depend on the current time slots only. One's energy usage behaviors in the previous time slots also exert a significant impact. For instance, a user's usage time of washing machine can be different from day to day, but his/her sequential behaviors in clothes washing are always similar: first using the washer, and then the dryer. 2) under many circumstances, a user's behavior is not just determined by himself/herself, but influenced by other members in the same household. For example, when parents wake up earlier than usual in the morning, they may also wake up their children earlier than usual. Another instance is that two household members are not able to use the bathroom at the same time, and consequently one member has to postpone his/her usage of the bathroom. Thus, to understand energy usage behaviors, appropriate modeling of influence among the energy usage behaviors of different users in the same household across different time slots is essential.

Unfortunately, the influence between energy usage behaviors is hard to model directly, since the state-of-the-art smart-grid data rarely records the number of household members, and the exact timestamp when a certain member uses a certain appliance. Since the energy consumption of each appliance is relies on the user behavior, we turn to modeling the relationship between the energy usages of different appliances across different time slots, and expect that such relationship will be able to reveal the influence between the energy usage behaviors of different users in the same household. We want to emphasize that such relationship has so far been largely ignored by existing works on energy disaggregation (Kolter, Batra, and Ng 2010; Kolter and Jaakkola 2012; Parson et al. 2012). Those works mostly focus on the distribution of energy consumption of each appliance alone. They either learned the energy usage patterns of each appliance within a certain period (for instance, a week), or studied the influence between energy usage patterns from one time slot to the next. Recent works discussed the dependency between appliances in the same time slot only (Kim et al. 2011). More importantly our method, while modeling the influence between energy usage patterns, also pays attention to the relationship between the energy usages of different appliances across different time slots.

One main challenge in modeling the influence among various appliances across different time slots is how to model the influence between *marked* events, which is defined to be events with marks that contain detailed information of the corresponding event. Under our scenario of energy disaggregation, events are the usage of a certain appliance in a certain time slot, while their marks are the amount of consumed energy. We leverage the general idea of marked point process for our modeling purpose (Rasmussen 2013). Specifically, we propose a novel probabilistic model named marked Hawkes process (M-Hawkes) based on the combination of multivariate Hawkes processes and topic models. This M-Hawkes is designed to model how the occurrence and the mark of an event *together* influence the occurrence and the mark of subsequent events in the near future. In the proposed M-Hawkes model, the topic model part models the distribution of marks of observed events, designed to find user behavior patterns underlying the amount of consumed energy of each appliance in each time slot, while the Hawkes process part models the occurrences of observed events, and captures the influence between different appliances under different energy usage behavior patterns across different time slots.

In a nutshell, our major contributions include: (1) We fully utilize both temporal and energy amount information in addressing the energy disaggregation task, emphasizing the analysis of the underlying energy usage behavior patterns; (2) We consider the relationship among the energy consumption of different appliances across different time slots, which existing works failed to model; (3) We propose a novel probabilistic model that combines Hawkes processes with topic models, which enables the modeling of the influence from the occurrence and the mark of an event to the occurrence and the mark of subsequent events in the future.

Problem Definition

Let us consider a typical scenario in energy disaggregation, where M appliances are used in a sequence of N time slots $T = \{t_n, n = 1, \ldots, N\}$. Multiple appliances can be used simultaneously in one time slot, and certain appliance is not necessarily always in use. Our paper considers the unsupervised setting, i.e., we only observe the total amount of consumed energy X_n in each time slot n, while the amount of consumed energy $x_{m,n}$ of each appliance m used in that time slot is unavailable. The target of energy disaggregation is to predict each $x_{m,n}$ based on the observed T and X.

Instead of straightforwardly predicting $x_{m,n}$ from X_n, we introduce a set of latent variables $\{Y_{m,n}\}$ to denote whether the m-th appliance is used in the n-th time slot, and turn to solving a much easier problem first: which appliances are in use in each of the time slot. The basic intuition is that the usage of one appliance raises the probability of the usage of related appliances (including itself) in the near future. For instance, people are very likely to use dryer after using washing machine. Such self- & mutually exciting nature coincides with the *self- & mutually exciting* property of the multi-dimensional Hawkes process, i.e., the occurrence of one event in the past will trigger events happening in the future.

Multi-dimensional Hawkes Process

The multi-dimensional Hawkes process is a class of self- or mutually-exciting point process models (Hawkes 1971), which are widely used to describe data that are localized at a finite set of time points $\{t_1, \ldots, t_N\}$ (Schoenberg 2010). Formally, the multi-dimensional Hawkes process on an event cascade $\{t_l\}_{l=1}^N$ is defined to be a M-dimensional point process with the intensity of the m-th dimension given by:

$$\lambda_m(t) = \mu_m + \sum_{t_l < t} \alpha_{m_l,m} \kappa(t - t_l)$$

Here μ_m denotes the base intensity of the m-th dimension, $\kappa(t - t_l)$ is a time-decaying kernel, while $\alpha_{m,m'}$ denotes the infectivity from events in the m-th dimension to events in the m'-th dimension. Hawkes process has been widely used in applications, such as earthquake prediction (Ogata 1988), sales modeling (Yan et al. 2015; Errais, Giesecke, and Goldberg 2010), Asset management (Yan et al. 2013), search behavior modeling (Li et al. 2014), crime modeling (Stomakhin, Short, and Bertozzi 2011), and armed conflict analysis (Mangion et al. 2012; Li and Zha 2013).

In our tasks, building a multi-dimensional Hawkes process on Y relates the inference of m-th appliance usage state in the n-th time slot $Y_{m,n}$ with that of other appliances in different time slots, thus can be expected to sharply raise the inference accuracy.

Marked Hawkes Process

Although the (multi-dimensional) Hawkes process has been proved to be effective in modeling the influence between event occurrences in many applications, we find it unable to completely solve our energy disaggregation problem. For one thing, the total amount of consumed energy in each time slot has not been utilized; for another, it only predicts whether an appliance is in use rather than the energy it consumes. A better solution is modeling marked events instead of normal events, where the mark of an event refers to those additional features other than the temporal information that describes the event.[1] In energy disaggregation, *taking the*

[1] A mark can be the casualty of an armed conflict event, the magnitude of an earthquake event, and in our application, the consumed energy of an appliance usage event.

usage of an appliance in a time slot as an event, the corresponding amount of consumed energy is actually the mark of that event. Such marked events are very common in current social networks, as the descriptions of events are usually available.

Since the marks of an event are very likely to be described by numerous features — a vector with each feature represented by continuous or categorical variables, directly modeling the relationship between marks and occurrences of different events is difficult. One widely used effective solution is the topic model, which clusters all observed marks into several topics/categories, with similar marks in the same category.

To enable the modeling of marks of events in Hawkes processes, we further introduce a new set of latent variables $\{Z_{m,n,k}\}$ to denote whether the marks of an event from the m-th dimension, whose occurrence is previously denoted by $Y_{m,n}$, belongs to the k-th category/topic, we propose the following novel multi-dimensional Hawkes process to model the entire event sequence, with the intensity of an event from the m-th dimension occurring in time slot t whose intensity can be written as: whose intensity can be written as:

$$\lambda_m(t) = \mu_m + \sum_{t_l<t} \sum_{m'} Y_{m',l} \sum_{k,k'} Z_{m,n,k} Z_{m',l,k'} \beta_{m,m',k,k'} \kappa(t - t_l). \tag{1}$$

Here the base intensity μ_m captures how often an event from m-th dimension happens spontaneously, while $\beta_{m,m',k,k'}$ models the degree of *influence* between a event from dimension m with marks of category k to a event from dimension m' with marks of category k'. Notice that the proposed new Hawkes process can handle events with multidimensional marks, while in our application, only a single dimension, the amount of consumed energy, is used.

According to the definition of $Y_{m,n}$, we have $Y_{m,n} = \mathrm{HawkesProcess}(\lambda_m(t_n))$. Thus the proposed new Hawkes process straightforwardly models the influence between the occurrence and the mark pattern membership of past events and those of the current event. Assume each appliance has K energy consumption patterns with the k-th pattern denoted as $\theta_{m,k}$, the entire amount of consumed energy in the n-th time slot X_n can be approximated by $\sum_m Y_{m,n} \sum_k \theta_{m,k} Z_{m,n,k}$. The approximation itself does not provide much evidence for the inference of Y and Z, and the learning of θ. However, by constructing a multi-dimensional marked Hawkes process on Y and Z, we relate the inference of $Y_{m,n}$ and $Z_{m,n}$ with that of other appliances in different time slots, thus the inference/learning accuracy can be expected to be increased.

Finally, we present our generative model that produces the entire energy consumption as follows:

- Draw a vector μ of length M that denotes the base intensity of each appliance and a $MK \times MK$ infectivity matrix β that denotes the degree of influence between different appliances under different consumption patterns.
- For each appliance m,
  - draw a K dimensional vector θ_m, where each dimension indicates a single energy consumption pattern of the appliance.
  - draw a K dimensional membership vector $\pi_m \sim \mathrm{Dirichlet}(\alpha)$.
- For the n-th time slot,
  - For the m-th appliance in the n-th time slot,
    - Draw whether it will be used by $Y_{m,n} \sim \mathrm{HawkesProcess}(\lambda_m(\cdot))$, where the intensity λ_m is defined as in Eqn (1);
    - Draw the user energy usage pattern membership $Z_{m,n} \sim \mathrm{Multinomial}(\pi_m)$;
    - Draw the amount of consumed energy of device $x_{m,n} \sim Y_{m,n}\mathrm{Gaussian}(\sum_k \theta_{m,k} Z_{m,n,k}, \sigma)$;[2]
  - Calculate the total amount of consumed energy in the n-th time slot $X_n = \sum_m x_{m,n}$.

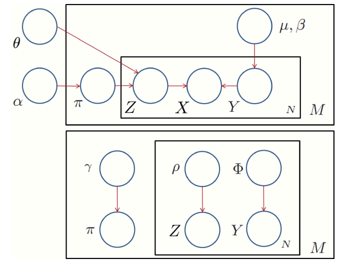


Figure 1: Graphical model representation of M-Hawkes and the variational distribution that approximates the likelihood. The upper figure shows the graphical model representation of M-Hawkes, while the lower figure shows the variational distribution that approximates the likelihood.

Note that in our M-Hawkes model, the number of appliances that can be simultaneously used in the same time slot is constrained by the total amount of consumed energy at that time. Such a constraint not only benefits the inferring of energy usage patterns of each appliance, but also enables the modeling of several events occurring in the same time slot, which existing Hawkes models hardly handled.

Under our M-Hawkes model, the joint probability of data $T = \{N(\cdot)\} = \{\{t_n\}_{n=1}^N\}$, $X = \{\{X_n\}_{n=1}^N\}$ and latent variables π, Y, Z can be written as follows:

$$p(T, X, \pi_{1:M}, Y, Z|\alpha, \theta, \mu, \beta) = P(T, Y|Z, \mu, \beta) \prod_n P(X_n|Y_n, Z_n, \theta)$$
$$\prod_n \prod_m P(Z_{m,n}|\pi_m) \prod_m \prod_n P(Y_{m,n}|\pi_m) \prod_m P(\pi_m|\alpha).$$

Inference

In this section, we derive a mean-field variational Bayesian inference algorithm for our proposed M-Hawkes model.

Variational Inference

Under M-Hawkes model, given observations of both temporal information $T = \{N(\cdot)\} = \{\{t_n\}_{n=1}^N\}$ and consumed energy X of energy consumption event sequences, the log-likelihood for the complete data is given by $\log p(T, X|\mu, \beta, \alpha, \theta)$. Since this true posterior is hard to

[2] In our experiments, we use a constant σ.

infer directly, we turn to variational methods (Blei and Jordan 2005), whose main idea is to posit a distribution over the latent variables with variational parameters, and find the settings of the parameters so as to make the distribution close to the true posterior in Kullback-Leibler (KL) divergence.

$$
\begin{aligned}
& q(\pi_{1:M}, Y, Z|\gamma_{1:M}, \Phi, \rho_{1:N}) \\
=& \prod_m q_1(\pi_m|\gamma_m) \prod_m \prod_n q_2(Y_{m,n}|\phi_{m,n}) q_2(Z_{m,n}|\rho_{m,n})
\end{aligned}
$$

where q_1 is a Dirichlet, q_2 is a multinomial, and $\{\gamma_{1:M}, \Phi, \rho\}$ are the set of variational parameters. We optimize those free parameters to tighten the following lower bound $\mathcal{L}'$ for our likelihood:

$$
\begin{aligned}
\log p(T, X|\mu, \beta, \alpha, \theta) \geq & E_q[\log p(T, X, \pi_{1:M}, Y, Z|\alpha, \theta, \mu, \beta)] \\
& - E_q[\log q(\pi_{1:M}, Y, Z)]. \qquad (2)
\end{aligned}
$$

Isolating terms containing λ in Eqn (2), we have

$$
\mathcal{L}_h = \sum_{m=1}^{M} \sum_n E_q(\log \lambda(Y_{m,n})) - \sum_{m=1}^{M} \int_0^T E_q(\lambda(s))ds, \quad (3)
$$

as the partial likelihood on temporal data assuming consumption pattern distribution is known. On one hand, we have $\sum_{m=1}^{M} \int_0^T E_q(\lambda(s))ds = \sum_{m=1}^{M} b_m \beta_{m,m'k,k'} + T\sum_{m=1}^{M} \mu_m$. Here

$$
b_{m,m',k,k'} = \sum_{n=1}^{N} \sum_{l=1}^{n-1} r_{m,m',lnkk'}(K(t_n - t_l) - K(t_{n-1} - t_l)),
$$

where $K(t) = \int_0^t \kappa(s)ds$, and we define function $r_{m,m',lnkk'} = \phi_{m',l}\rho_{m,n,k}\rho_{m',l,k'}$. On the other hand, in order to update each Hawkes hyper-parameter μ and β independently, we adopt the strategy in (Yang and Zha 2013), and break down the log sum $E_q(\log \lambda(t_n))$ based on Jensen's inequality as:

$$
\begin{aligned}
& \mathbb{E}_q(\log(\lambda_m(t_n))) \geq \eta_{m,nn} \log(\mu_m) - \eta_{m,nn} \log(\eta_{m,nn}) \\
& + \sum_{l=1}^{n-1} \sum_{m',k,k'} \eta_{m,m',lnkk'} \log(r_{m,m',lnkk'} \beta_{m,m',k,k'} \kappa(t_n - t_l)) \\
& - \sum_{l=1}^{n-1} r_{m,m',lnkk'} \eta_{m,m',lnkk'} \log(\eta_{m,m',lnkk'}),
\end{aligned}
$$

where $\{\eta\}$ is a set of branching variables constrained by:

$$
\eta_{m,m',lnkk'} \geq 0,\ \eta_{m,nn} + \sum_{l=1}^{n-1} \sum_{m',k,k'} r_{m,m',lnkk'} \eta_{m,m',lnkk'} = 1.
$$

Under a coordinate descent framework, we optimize the lower bound as in Eqn (2) against each variational latent variable[3] and the model hyper-parameter. For variational latent variables, we have the following process

- update rules for ρ's as:

$$
\begin{aligned}
\rho_{m,n,k} \propto & \exp(\sum_m (\Psi(\gamma_{m,k}) - \Psi(\sum_k \gamma_{m,k})) \\
& + \log([X_n - \sum_{m' \neq m, k' \neq k} \phi_{m',n} \rho_{m',n,k'} \theta_{m',k'}]_+) \\
& - \log(\phi_{m,n} \theta_{m,k}) + \sum_{l=1}^{n-1} f_{l,n} + \sum_{l'=n+1}^{N_m} f_{n,l'}),
\end{aligned}
$$

where we define

$$
\begin{aligned}
f_{l,n} = & \sum_{m',k'} (\eta_{m,m',lnkk'} \phi_{m',l} \log(\frac{\beta_{m,m',k,k'} \kappa(t_n - t_l)}{\eta_{m,m',lnkk'}}) \\
& - \phi_{m',l}(K(t_n - t_l) - K(t_{n-1} - t_l))) \rho_{m',l,k'}
\end{aligned}
$$

- update rules for γ's as:

$$
\gamma_{m,k} = \alpha_k + \sum_n \rho_{m,n,k};
$$

- update rules for ϕ's as:

$$
\begin{aligned}
\phi_{m,n} \propto & \exp(\eta_{m,nn} \log(\mu_m) - \log(\sum_k \rho_{m,n,k} \theta_{m,k}) \\
& + \log([X_n - \sum_{m' \neq m, k} \phi_{m',n} \rho_{m',n,k} \theta_{m,k}]_+) \\
& + \sum_{l=1}^{n-1} \sum_{m',k,k'} \eta_{m,m',lnkk'} \log(b_{mm',lnkk'}) \\
& + \sum_{l=n+1}^{N} \sum_{m',k,k'} \eta_{m',m,nlkk'} \log(b_{mm',nlkk'})).
\end{aligned}
$$

where $b_{mm',lnkk'} = r_{m,m',lnkk'} \beta_{m,m',k,k'} \kappa(t_n - t_l)$.

- and update rules for η as:

$$
\begin{aligned}
\eta_{m,nn} &= \frac{\mu_m}{\mu_m + \sum_{l=1}^{n-1} \sum_{m',k,k'} b_{mm',lnkk'}}, \\
\eta_{m,m',lnkk'} &= \frac{\beta_{m,m',k,k'} \kappa(t_n - t_l)}{\mu_m + \sum_{l=1}^{n-1} \sum_{m',k,k'} b_{mm',lnkk'}}.
\end{aligned}
$$

Learning

We use a variational expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) to compute the parameters in our M-Hawkes model. This variational EM algorithm iteratively approximates the posterior by fitting the variational distribution q and optimizes the corresponding bound against the parameters.

In updating α, we use a Newton-Raphson method, since the approximate maximum likelihood estimate of α doesn't have a closed form solution. The Newton-Raphson method is conducted with a gradient and Hessian as follows:

$$
\begin{aligned}
\frac{\partial \mathcal{L}'}{\partial \alpha_k} &= N(\Psi(\sum_k \alpha_k) - \Psi(\alpha_k)) + \sum_m (\Psi(\gamma_{m,k}) - \Psi(\sum_k \gamma_{m,k})), \\
\frac{\partial \mathcal{L}'}{\partial \alpha_{k_1} \alpha_{k_2}} &= N(\mathbb{I}_{(k_1 = k_2)} \Psi'(\alpha_{k_1}) - \Psi'(\sum_k \alpha_k)).
\end{aligned}
$$

The maximum likelihood estimation of *energy usage pattern* θ can be derived through calculating the first derivative of lower-bound $\mathcal{L}'$ against corresponding parameters. We obtain the update formulas given as follows:

$$
\theta_{m,k} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{x}.
$$

where $\mathbf{A} = [\phi_{m,n} \rho_{m,n,k}]_{n,mk}$ is a matrix of size $n \times mk$, and $\mathbf{x} = [X_n]_n$ is a vector of length n.

To obtain the approximate maximum likelihood estimation of Hawkes hyper-parameters, we optimize the lower bound as in Eqn (2) against each hyper-parameter, and update μ and β independently with closed-form solutions as:

$$
\beta_{m,m',k,k'} = \frac{1}{b_m} \sum_{n,l<n} r_{m,m',lnkk'} \eta_{m,m',lnkk'},\ \mu_m = \frac{1}{T} \sum_{n=1}^{N} \eta_{m,nn}
$$

[3]Here we categorize branching variables η as variational latent variables.

In real world scenario β is usually a sparse matrix, as influence only exist in limited pairs of appliances and patterns. Thus to select effective influence and avoid overfitting, we enforce the sparsity of β by imposing lasso type of regularization as $\|\beta\|_1$, and employ the widely used alternating direction method of multipliers (ADMM) (Boyd 2010; Li and Zha 2014) to address the constraint optimization problem.

Our variation inference algorithm, named Marked-Hawkes (M-Hawkes), can be interpreted intuitively in the following way. The mark pattern distribution γ of each appliance is determined by both the topic/pattern prior and the pattern assignment of each appliance at each time slot. The probability ϕ of an appliance m used in the n-th time slot is jointly determined by: (a) other appliances used in the current time slot; (b) how likely an appliance was used spontaneously; (c) the influence from the occurrence and the mark pattern of past events to the current occurrence; and (d) the influence from the occurrence and the mark pattern of future events to the current occurrence. The energy consumption pattern ρ of an appliance m used in the n-th time slot is jointly determined by: (a) the pattern prior of this appliance; (b) the mark patterns of other appliances; (c) past/future influence to the current mark pattern.

In our mean-field variation inference algorithm, the computational cost of inferring variational variables is $O(NM^2K^2)$. The computational cost of the estimation of topic hyper-parameters is $O(NM^2K^2 + M^3K^3)$. The computational cost of the estimation of Hawkes hyper-parameters is $O(N^2M^2K^2)$, which can be reduced to $O(NM^2K^2)$ by only considering the influence in temporally-close time slots. Thus the total computational cost of our algorithm is $O(NM^2K^2 + M^3K^3)$. Since in real-world scenarios, influence exists only among limited pairs of appliances and patterns, M^2K^2 can be reduced to some much smaller constant, thus the above cost can be viewed as linear in the number of events or time slots.

Experiments

We evaluated our M-Hawkes model on both synthetic and real-world data sets, and compared the performance with the following baselines:

Hawkes: This is a normal multi-dimensional Hawkes process that models the occurrence of events only and no marks of events;

AFAMAP: This method proposed an approximation inference algorithm, named Additive Factorial Approximate MAP, to efficiently solve the additive factorial hidden Markov model by looking at the observed difference in consumed energy, and incorporating a robust mixture component that can account for unmodeled observation (Kolter and Jaakkola 2012).

NIALM: This method, named non-intrusive load monitoring, iteratively separated individual appliances from an aggregate energy consumption record, and updated prior models of general appliance types for each specific appliance instance (Parson et al. 2012).

Synthetic data

Data Generation. Given parameters $(M, N, K, \alpha, \theta, \mu, \beta)$, the synthetic data is sampled according to the proposed generative model. Here each element μ_m and $\beta_{m,m,k,k'}$ are randomly generated in $[0.5\hat{\mu}, 1.5\hat{\mu}]$ and $[0.5\hat{\beta}, 1.5\hat{\beta}]$ respectively before the simulation. In addition, α is a vector of size K, where the element α_k is generated in $[0.5\hat{\alpha}, 1.5\hat{\alpha}]$ before the simulation. Our synthetic data are simulated with two different settings:

- `Small`: $M = 10$, $N = 120$, $K = 3$, $\hat{\mu} = 0.01$, $\hat{\beta} = 0.5$, $\hat{\alpha} = 0.1$, $\hat{\theta} = 10$. Simulations were run 1,000 times using the pre-generated parameters μ, β;
- `Large`: $M = 50$, $N =$ 10,000, $K = 5$, $\hat{\mu} = 0.01$, $\hat{\beta} = 0.5$, $\hat{\alpha} = 0.1$, $\hat{\theta} = 10$. Simulations were run 10 times.

To test the robustness of our method, we add two types of noise to the original synthetic data:

`Event Noisy`: We generate additional 10% of total number of events randomly in the time window of each already sampled event sequence, and add them to the sequence;

`Mark Noisy`: Instead of using the simulated X_n as the consumed energy at the n-th time slot, we use a noisy value X'_n which is obtained by adding Gaussian noise on X_n:

$$X'_n = \max(0.1e + 1, 0)X_n, \ e \sim \mathcal{N}(0, \sigma'). \tag{4}$$

The default value of σ' is set to be 1.

Evaluation metrics. We consider the following evaluation metrics: 1) first, we compare the average log predictive likelihood on events falling in the final 10% of the total time

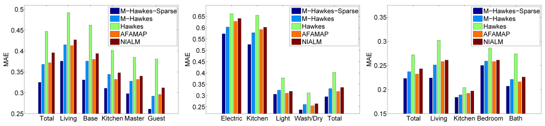
Table 1: Inference and Estimation of M-Hawkes on Synthetic data

Data set	MAE(μ)	MAE(β)	MAE(Y)	MAE(Z)
S-Synthetic	0.065	0.197	0.9251	0.9432
S-E-Noisy	0.077	0.281	0.9042	0.9229
S-M-Noisy	0.092	0.313	0.8847	0.9085
L-Synthetic	0.148	0.346	0.8718	0.8942
L-E-Noisy	0.163	0.353	0.8503	0.8642
L-M-Noisy	0.187	0.386	0.8284	0.8372

"S-" stands for data setting `Small`, "L-" stands for `Large`, "E-" stands for `Event Noisy`, and "M-" stands for `Mark Noisy`.

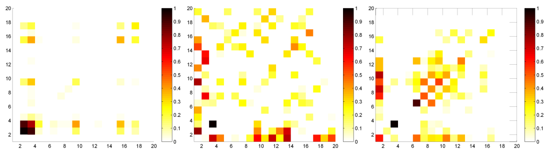
Table 2: Log Predictive Likelihood on Both Synthetic and Real-world Data

Data set	M-Hawkes	Hawkes	AFAMAP	NIALM
S-Synthetic	-96.23	-136.26	-104.28	-108.63
S-E-Noisy	-109.21	-148.32	-120.94	-125.27
S-M-Noisy	-116.93	-161.24	-134.27	-140.05
L-Synthetic	-152.39	-194.38	-168.03	-173.26
L-E-Noisy	-165.82	-208.43	-181.46	-186.85
L-M-Noisy	-171.47	-224.06	-186.94	-191.27
Smart*	-145.39	-182.55	-157.83	-160.35
Pecan	-192.17	-234.88	-209.12	-216.43
REDD	-171.37	-210.26	-182.37	-187.51



(a) `Smart*` (b) `REDD` (c) `Pecan`

Figure 2: Performance Comparison of Energy Disaggregation on Real World Data Sets.



(a) `Smart*` (b) `REDD` (c) `Pecan`

Figure 3: Energy Usage Pattern on Real World Data Sets. Indice of significant appliances: Smart*: 2-lamp, 3-ac, 4-fan, 9-toaster, 15-refrigerator, 17-microwave. REDD: 1-main, 6-dishwaser, 15-kitchen_outlets, 17-light, 19-washer-dryer. Pecan: 1-ac, 2-dishwasher, 13-microwave, 16-refrigerator.

of each event cascade; 2) next we compare the average relative distance between the estimated parameters and ground-truth ones by Mean Average Error (MAE). For instance, the MAE of parameter β and $\frac{1}{M}\sum_m |\frac{\mu_m - \hat{\mu}_m}{\mu_m}|$, which we denote as $\mathrm{MAE}(\beta)$. 3) finally, we measure the performance of energy disaggregation by the MAE between the ground-truth consumed energy of each appliance $x_{m,n}$ and the estimated consumed energy $\hat{x}_{m,n}$, which is calculated based on the inferred $\rho_{m,n}$ and the estimated $\hat{\theta}_m$.

Inference and Estimation. Table 1 evaluates both the accuracy of our proposed variational inference algorithm in parameter estimation and latent variable inference on the synthetic data. We find that, on the small synthetic data, M-Hawkes can recover the Hawkes parameters μ and β very well, and accurately estimate the model's hyper-parameters. On the large synthetic data, M-Hawkes's performance on parameter estimation becomes worse. The shapely increased number of appliances makes the event occurrence prediction more difficult, and further affects the learning of users' energy usage behavior patterns. On both noisy data sets, M-Hawkes's performances in both inference and estimation become worse. We also find that the performance of energy disaggregation become worse with respect to the increase of the number of appliances, which shapely increases the complexity of the problem.

Real-world Data

We also conducted extensive experiments on two real-world data sets. The first data set is Smart* (Barker et al. 2012), which is a high-resolution data set from three homes including over 50 appliances.. The second data set is Reference Energy Disaggregation Dataset (REDD) (Kolter and Johnson 2011). This data set comprises six houses including around 20 appliances. The third data set is Pecan Street [4]. This data set collects one-minute resolution disaggregated data for 450+ homes including around 20 appliances, dating from late 2012 to early 2014.

Model Fitness. Table 2 shows the log predictive likelihood on energy consumption falling in the final 10% of the total time of data. According to Table 2, M-Hawkes fits both synthetic and real-world data better than alternative probabilistic models. The comparison on synthetic data is meaningful since we add noise into it. AFAMAP performs better than the normal multi-dimensional Hawkes process, which shows the importance of modeling marks of events besides the occurrences. On both noisy data sets, the performances of all models become worse. However, the decrease of the performance of M-Hawkes is smaller than baselines, which demonstrates the robustness of our proposed model. Thus when the usage timestamps and the amounts of consumed energy of some appliances are misrecorded, M-Hawkes performs better in energy disaggregation, and learns energy usage behaviors better.

Performance on Energy Disaggregation. To illustrate the effectiveness of the proposed model in energy disaggregation, we compare it with all baselines measured by $\mathrm{MAE}(X)$. Here we use M-Hawkes-NS to denote the M-Hawkes model with no sparsity constraint on Hawkes hyper-parameter β. According to Figure 2, M-Hawkes performs at least 5% better than all compared methods with comparable time costs. Also, M-Hawkes outperforms compared methods on all categorized appliances. Such results demonstrate the importance of modeling the relationship between the consumed energy of different appliances across different time slots. M-Hawkes's advantage over M-Hawkes-NS illustrates that only a limited number of dependencies exist between appliances in real world energy consumption.

Energy Usage Behavior Pattern Analysis. Based on the parameters learned by the proposed M-Hawkes model, we analyze the energy usage behavior patterns detected in real world energy consumption. According to Figure 3, influences exist in only limited pairs of appliances. Moreover, the degrees of those influences are very different. In the Smart* data, the influence between lamp and ac is greater than those between all other pairs of appliances. The influence between refrigerator and microwave is greater than that between refrigerator and toaster, which implies that people are more likely to cook food using microwave than toaster. Notice that Smart* data only recorded significant energy consumptions of refrigerator, which makes its usages easily detectable. In addition, the self-influence on some appliances, such as ac, are also very significant. The interpretation is that those appliances are often used for a long time continuously. The results on REDD also show that rarely used appliances, such as dishwaser and washer-dryer influence much less other appliances than those frequently used appliances, such as light and kitchen outlets. Moreover, the influence between a certain pair of appliances is not always symmetric. In Pecan, the influence from refrigerator to microwave is greater than the influence from microwave to refrigerator. One explanation is that people are used to open refrigerator to fetch food

[4]http://www.pecanstreet.org/

before turn on the microwave to cook them. We also find such phenomenon in Smart* data.

Conclusion and Future Work

In this paper, we formulated the task of energy disaggregation into the modeling of marked event sequences. Our paper presented a probabilistic model that integrates topic models with Hawkes processes to capture the influence from the occurrence and the mark of an event to the occurrences and the marks of future events. In future work, it would be interesting to consider other marks, e.g., the attributes of appliances, into this framework, and investigate the performance of M-Hawkes in other domains. In addition, we'll attempt to directly model the behavior of users instead appliances, and the influence inbetween.

Acknowledgments

This work is supported in part by NSF IIS-1116886 and NIH R01 GM108341.

References

Baptista, M.; Fang, A.; Prendinger, H.; Prada, R.; and Yamaguchi, Y. 2014. Accurate household occupant behavior modeling based on data mining techniques. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, 1164–1170.

Barker, S.; Mishra, A.; Irwin, D.; Cecchet, E.; Shenoy, P.; and Albrecht, J. 2012. Smart*: An open data set and tools for enabling research in sustainable homes. In *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012), Beijing, China*.

Blei, D., and Jordan, M. 2005. Variational inference for dirichlet process mixtures. In *Bayesian Analysis*, volume 1, 121–144.

Boyd, S. 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1):1–122.

Darby, S. 2006. The effectiveness of feedback on energy consumption. *Technical report, Environmental Change In- stitute, University of Oxford.*

Dempster, A. P.; Laird, N. M.; and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 1–38.

Errais, E.; Giesecke, K.; and Goldberg, L. R. 2010. Affine point processes and portfolio credit risk. *SIAM J. Fin. Math.* 1(1):642–665.

Hawkes, A. G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58:83–90.

Kim, H.; Marwah, M.; Arlitt, M. F.; Lyon, G.; and Han, J. 2011. Unsupervised disaggregation of low frequency power measurements. In *SDM*, 747–758. SIAM / Omnipress.

Kolter, J. Z., and Jaakkola, T. 2012. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, April 21-23, 2012*, 1472–1482.

Kolter, J. Z., and Johnson, M. J. 2011. Redd: A public data set for energy disaggregation research. In *Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*.

Kolter, J. Z.; Batra, S.; and Ng, A. Y. 2010. Energy disaggregation via discriminative sparse coding. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada.*, 1153–1161.

Li, L., and Zha, H. 2013. Dyadic event attribution in social networks with mixtures of hawkes processes. In *Proceedings of the 22nd ACM International Conference on Conference on Information; Knowledge Management*, CIKM '13, 1667–1672. New York, NY, USA: ACM.

Li, L., and Zha, H. 2014. Learning parametric models for social infectivity in multi-dimensional hawkes processes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, 101–107.

Li, L.; Deng, H.; Dong, A.; Chang, Y.; and Zha, H. 2014. Identifying and labeling search tasks via query-based hawkes processes. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, 731–740.

Mangion, A. Z.; Dewarc, M.; Kadirkamanathand, V.; and Sanguinetti, G. 2012. Point process modelling of the afghan war diary. *PNAS* 109(31):12414–12419.

Neenan, B., and Robinson, J. 2009. Residential electricity use feedback: A research synthesis and economic frame- work. *Technical report, Electric Power Research Institute.*

Ogata, Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association.* 83(401):9–27.

Parson, O.; Ghosh, S.; Weal, M.; and Rogers, A. 2012. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 22 -26, 2012, Toronto, Canada*, 356–362.

Rasmussen, J. G. 2013. Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability* 15(3):623642.

Schoenberg, F. 2010. Introduction to point processes. *Wiley Encyclopedia of Operations Research and Management Science* 616–617.

Stomakhin, A.; Short, M. B.; and Bertozzi, A. L. 2011. Reconstruction of missing data in social networks based on temporal patterns of interactions. *Inverse Problems.* 27(11).

Wytock, M., and Kolter, J. Z. 2014. Contextually supervised source separation with application to energy disaggregation. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, 486–492.

Yan, J.; Wang, Y.; Zhou, K.; Huang, J.; Tian, C.; Zha, H.; and Dong, W. 2013. Towards effective prioritizing water pipe replacement and rehabilitation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, 2931–2937. AAAI Press.

Yan, J.; Zhang, C.; Zha, H.; Gong, M.; Sun, C.; Huang, J.; Chu, S.; and Yang, X. 2015. On machine learning towards predictive sales pipeline analytics. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.*

Yang, S., and Zha, H. 2013. Mixture of mutually exciting processes for viral diffusion. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, 1–9.