Designing Customized Energy Services Based on Disaggregation of Heating Usage

Dayu Huang, Marina Thottan, and Frank Feather

Abstract—**The deployment of smart meters has made available high-frequency (minutes as opposed to monthly) measurements of electricity usage at individual households. Converting these measurements to knowledge that can improve energy efficiency in the residential sector is critical to attract further smart grid investments and engage electricity consumers in the path towards reducing global carbon footprint. The goal of the reported research is to use smart meter measurement data to identify heating and cooling usage levels for a home. This is important to cost effectively design consumer energy services such as energy audit and demand response targeted towards improving an individual household's heating usage efficiency.**

We present a machine learning approach akin to Non-Intrusive Load Monitoring (NILM) to disaggregate heating usage from measurements of a household's total electricity usage. We use as input 15-minute interval meter data and hourly outdoor temperature measurements. Our approach does not require a manual set-up procedure at each house. The method uses a Hidden Markov Model to capture the dependence of heating usage on outdoor temperature. Compared to existing methods based on linear regression, the proposed method provides details on heating usage patterns and is more flexible to incorporate other system specific information. Preliminary results based on synthetic and real-world usage data demonstrate the feasibility of the proposed approach.

Index Terms—**HVAC, Non-Intrusive Load Monitoring, Hidden Markov Model, disaggregation, energy management service**

I. Introduction

The first step towards smart grid deployment is the installation of smart meters (sensors) at consumer premises. Smart meters provide detailed data on an individual household's electricity usage: Instead of monthly usage data collected via manual reading, it is now possible to collect usage measurements at 15-minutes (or shorter) intervals [1]. To realize the envisioned benefits of smart grid such as energy savings, increased reliability, reduced cost and improved customer satisfaction, the collected data needs to be converted into actionable information for both the consumer and the utility. The benefit to the end-consumer is a crucial element for the wide scale adoption of smart meters.

For a residential consumer the heating and cooling usage is a significant component of the energy consumption (28% of annual usage in U.S. in 2010 [2] and projected to increase worldwide), and is a promising avenue for implementing energy services that optimize the usage and provide benefits to the consumer. However, the challenge in designing these energy services is to extract the heating and cooling usage from the aggregated meter data without compromising the privacy of the consumer. We present a machine learning method that can address this challenge.

The approach presented in this paper provides methodologies to extract heating usage information from aggregate meter readings in order to design customized residential energy services that optimize heating usage. The approach applies to thermostatically controlled heating appliances, however, it can be extended to cooling usage. Optimizing heating usage holds a significant potential for conserving energy and reducing peak usage. For example, an energy audit can improve the thermal efficiency of the building and thus reduce usage. Similarly, demand response programs for the grid [3], [4] can reduce peak usage by temperately adjusting the cycle of the heating appliance without significantly reducing the comfort level. Data analytics plays a critical role to support these energy audit and demand response services by customizing them to the individual household usage patterns. For example, the estimated heating usage of a household could be compared to other similar households. If the usage is significantly higher than average, a recommendation for energy audit can be made to the house owner. In another example, the estimated heating usage characteristics are used to calculate its demand response potential. Such information can be used by the utility in implementing a residential demand response program.

Central to these customized energy services is the estimate of a household's heating usage, which includes not only the average daily energy consumed for heating, but also other characteristics, such as the number of on/off cycles of the heating appliances. The latter is important for estimating a house's potential contribution to a demand response program [3], [4]. Typically the measurement of usage is done at a household level, rather than for individual appliances. This raises the main technical problem addressed in this paper: estimating the heating usage from the aggregate usage data of a household. We address this problem by using Thermostatically Controlled Load (TCL) models to develop a Hidden Markov Model (HMM) for heating usage estimation. Input to this HMM are: hourly (or sub-hourly) outdoor temperature readings [5] and 15-minutes aggregate meter readings for the household.

II. State of the Art

In general the goal of Residential Energy Services, or Home Energy Management Services (HEMS), is to basically save energy at home. There are numerous players in the Residential

Dayu Huang is with Department of Electrical and Computer Engineering, University of Illinois at Urbna-Champaign, Urbana, IL 61801. Email: dhuang8@illinois.edu. Marina Thottan and Frank Feather are with Alcatel-Lucent Bell Labs, Murray Hill, NJ 07974. Emails: {Marina.Thottan, frank.feather}@alcatel-lucent.com. This work was done when Dayu Huang was a summer intern at Alcatel-Lucent Bell Labs.

Energy Services market, including utility companies, government, standards groups, and manufacturers of infrastructure equipment such as the smart meters, thermostats and appliances. The most basic form of Home Energy Management is a home energy audit, which has someone come to the home and catalog and assess its energy efficiency, including the type of heating and cooling appliances, ratings of insulation, window types, and other demographics of the home. These audits are very extensive and therefore expensive. Most other HEMS are centered on the smart meter. Some HEMS are stand-alone solutions that target the consumer directly: the customer uses a smart thermostat and smart appliances that integrate with smart meters to monitor usage, learn a customer's habits, and then control the thermostat in order to optimize heating and cooling usage [6]–[8]. Other HEMS are utility-centric, in that the utility company will use a smart meter infrastructure to monitor individual customer's usage and intelligently present the information and/or provide recommendations. However, a majority of these solutions simply monitor a customer's smart meter and present current and historical usage, along with trend lines and future usage prediction. There are a few companies that have sophisticated analysis to compare a customer's usage to that of his "neighbors", where a neighbor has similar demographics, such as home size [9]. The work that is most closely related to our approach is the method in [10] which disaggregates heating and cooling usage using regression analysis. We use this regression-based method to benchmark and discuss the performance of the proposed HMM-based approach.

This paper is organized as follows: Previous work on heating usage disaggregation is reviewed in Section III. The HMM-based algorithm is introduced in Section IV and evaluated in Section V. Its application to energy audit and demand response is given in Section VI. The paper is concluded in Section VII.

III. Problem Description

Our goal is to disaggregate energy consumption of the heating appliance from total usage for a house. We focus on estimating the usage of heating appliance with on/off states controlled by a thermostat, which is the focus of studies on residential demand response program [3], [4].

Let $y(t)$ denote the aggregate real power consumption of a house at time t, and $w(t)$ denote the power consumption of the heating appliance. Consider meter measurements taken at sampling interval of length Δ. Denote the aggregate energy and the energy consumed by the heating appliance at the kth interval by $y[k] = \int_{(k-1)\Delta}^{k\Delta} y(t)dt$ and $w[k] = \int_{(k-1)\Delta}^{k\Delta} w(t)dt$. Let $\theta_o[k]$ be the outdoor temperature at time $k\Delta$.

In the disaggregation problem, the inputs given to the algorithm are the sequences $\{y[k]\}$ and $\{\theta[k]\}$, and the required output is a sequence of estimates $\{\hat{w}[k]\}$ of the true heating usage $\{w[k]\}$, or estimates of other functionals of $w(t)$. For the energy audit and demand response services, the relevant estimates are: (1) The average heating power over time T: $\frac{1}{T}\int_0^T w(t)dt$; (2) The average energy consumed during each on/off period: $\frac{1}{N(T)}\int_0^T w(t)dt$, where $N(T)$ is the number of times that the appliance is on during $[0, T]$.

The assumptions needed for the proposed method are:

1) The household of interest has one single thermostatically controlled heating appliance that switches between an on and an off state. The heating appliance is turned on when the indoor temperature is lower than a pre-set temperature and turned off when it is higher than another pre-set temperature.
2) The sampling frequency $1/\Delta$ is higher than the frequency that the heating appliance switches between on and off states.
3) The heating appliance consumes the same amount of energy during each period that it is on.

The last assumption is only an approximation: The time that a thermostatically controlled heating appliance stays on and its power consumption depends on the outdoor temperature. One possible way to relax this assumption is to explicitly model this temperature dependence. This will be investigated in the future.

A. Previous Work: Non-Intrusive Load Monitoring

Disaggregating individual appliance usage from the aggregate usage of a house is called Non-Intrusive Load Monitoring (NILM). The main varying characteristics of NILM methods are: (1) the sampling interval Δ (or sampling frequency); (2) the appliance information needed, such as historical usage of individual appliances; (3) the types of appliance whose usage the method disaggregates. The sampling frequency and appliance information are limited by the data collection procedure.

The NILM method reported in [11] uses the real and reactive power as signatures of individual appliances. It first detects the change of aggregate instantaneous power consumption which signifies the on/off state change of an appliance, and then matches how much the power consumption changes to the appliance signatures. The sampling interval in the experiment reported in [11] is $\Delta = 1s$. When the sampling interval increases, the probability that two or more appliances change their power consumption at the same interval increases. For smart meter measurements of energy consumption collected at 15 minutes or less frequently, machine learning methods, such as rule-based algorithm [12] and neural-networks [13] are used. These methods disaggregate the usage of large appliances- such as HVAC, water heaters and pool pumps - by treating the usage of smaller appliances as noise. Other signatures such as the voltage spectrum or transients, which are available from higher frequency measurements, also have been used (see for example [14] [15]). These are not available with measurements of 15-minutes interval.

B. Previous Work: Linear Regression

The state of the art method to disaggregate heating usage from total usage measured at 15-minutes or 1-hour interval and based on linear regression is given in [10]. Define the difference between the outside temperature, $\theta_o[k]$, and a pre-determined baseline temperature, θ_B, as:

$$d[k] = \max\{\theta_B - \theta_o[k], 0\}.$$

One possible choice of θ_B mentioned in [10] is $18.3°C$. Then $y[k]$ is regressed against $d[k]$ to obtain estimates of β_0, β_1:

$$y[k] = \beta_0 + \beta_1 d[k].$$

The estimated heating usage is then given by $\hat{w}[k] = \hat{\beta}_1 d[k]$, where $\hat{\beta}_1$ is the estimate of β_1. The key assumption in linear regression method is that the usage of non-heating appliances is a constant. It neglects the on/off behavior of appliances, and is suitable for estimating average heating power.

C. Our solution

The method we propose builds upon existing methods in NILM. One way to disaggregate heating usage is to first apply methods described in Section III-A and then identify which appliance is a heating appliance. The method proposed here uses the temperature dependence of heating appliance usage directly in the NILM method to help disaggregate heating usage: It uses the change of energy consumption as well as its dependence on the temperature as the appliance signatures.

The proposed method uses HMMs to model energy usage. HMMs are stochastic models that have the ability to capture complex dependences between variables. The reader is referred to [16] for a tutorial. It has been used in some recently proposed NILM methods [17]–[19].

There are two types of variables in an HMM: observed variables and hidden variables where the sequence of hidden variables forms a Markov process. In previous NILM methods, the hidden variables are used to model the on/off state of individual appliances. The observed variables are conditionally independent of each other given the value of the hidden variables, and they are used to model the electric usage. In the proposed method, the hidden variable models the on/off state as well as the indoor temperature seen at the thermostat, and the evolution of the hidden variables depends on the outdoor temperature. This allows us to capture the temperature dependence of the electricity usage. Details of the proposed HMM method are given in Section IV.

One advantage of using an HMM is its flexibility: Additional variables can be added to the HMM. For example, if there is a survey on electric heating appliances in the utility's area, then a variable can be added to incorporate this piece of information. These are important directions to be pursued in future work.

IV. Disaggregation of Heating Usage

We first describe the TCL model for HVAC, which is then used to derive the HMM for heating usage disaggregation.

A. TCL model

The following TCL model and its discrete time version have been used in [3], [4], [20], [21] to describe the evolution of the temperature seen at the thermostat $\theta(t)$:

$$\frac{d}{dt}\theta(t) = -\frac{1}{\tau}(\theta(t) - \theta_f(t) - u(t)\theta_g) \tag{1}$$

where $\theta_f(t)$ is the target temperature that $\theta(t)$ approaches when the heating appliance is off, and θ_g is the additional heat gain when it is on. The binary-valued $u(t)$ is the on/off state of the heating appliance: $u(t) = 1$ when it is on, and $u(t) = 0$ when it is off. The on/off state is determined by how $\theta(t)$ compares to lower and upper set-points of the thermostat θ_l and θ_u: $u(t)$ does not change if $\theta_l < \theta(t_-) < \theta_u$; $u(t) = 1$ if $\theta(t_-) \le \theta_l$, and $u(t) = 0$ if $\theta(t_-) \ge \theta_u$.

This TCL model is the basis for the HMM used in the disaggregation algorithm. The thermal dynamics of a house is more complicated than (1). *Simulating* a house's heating requires detailed models with a large number of variables and parameters. Our approach is *not* to fit a simulation model to the usage data. The heating usage is inferred from the usage data directly and the TCL model is used to distinguish the heating usage from other usages that do not depend on temperature. In statistical inference, it is well-known that a simpler model is preferred when only limited amount of data is available.

B. HMM

The HMM used for the disaggregation algorithm is depicted in Figure 1. The sequences$\{z[k]\}$ and $\{r[k]\}$ model the usage of heating and non-heating appliances, respectively. The total usage $y[k]$ is given by $y[k] = z[k] + r[k]$. Rather than using $z[k]$ and $r[k]$ directly as the variables of HMM, we use the difference between two consecutive values: Define $Z[k] = z[k] - z[k-1], R[k] = r[k] - r[k-1], Y[k] = y[k] - y[k-1]$. The observed variable of the HMM is $Y[k] = Z[k] + R[k]$.

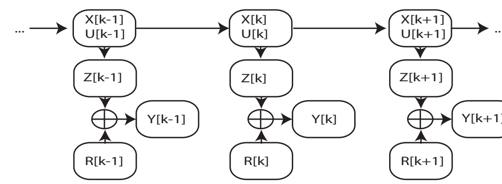

Fig. 1. Hidden Markov Model for heating usage disaggregation.

The change of non-heating usage $\{R[k]\}$ is modeled as an i. i. d. sequence of symmetric Gaussian mixture random variables with $s_0 + 2s_1$ components, where the s_0 components model the change of aggregate usage of small appliances and s_1 components model that of large appliances. The density function is given by

$$f_{R[k]}(r) = \sum_{v=1}^{s_0+2s_1} \alpha_v \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\{-\frac{(r-\mu_v)^2}{2\sigma_v^2}\},$$

where the coefficients $\{\alpha_v\}$ satisfying $\sum_{v=1}^{s_0+2s_1} \alpha_v = 1$ are the probabilities of individual Gaussian components. The parameters μ_v and σ_v are the mean and variance of the vth component. The following constraints are imposed so that the resulting density function is symmetric: $\mu_v = 0$ for $1 \le i \le s_0$, $\mu_{s_0+2i-1} = -\mu_{s_0+2i}$ and $\sigma^2_{s_0+2i-1} = \sigma^2_{s_0+2i}$ for $1 \le i \le s_1$.

We now describe the modeling of heating usage. The change of the heating usage $Z[k]$ depends conditionally on the state variable pair $(X[k], U[k])$ The real-valued variable $X[k]$ models the temperature seen at the thermostat at time $k\Delta$. The two-dimensional vector $U[k] \in \{0, 1\}^2$ models heating gain from the heating appliance at time interval $k-1$ and k. $U[k]_1 = 0$ and $U[k]_2 = 1$ indicates that the heating device is off during $[(k-1)\Delta, (k-2)\Delta)$ and is on during $[k\Delta, (k-1)\Delta)$.[1] The value of $(X[k], U[k])$ depends on the value of $(X[k-1], U[k-1])$ and $\theta_o[k]$. The transition probability is given by:

$$\begin{aligned}&\Pr\{U[k]_2 = 1 | X[k-1] = x\}\\ &= (1-\eta)\chi\{x \le T_l\} + \eta\chi\{x > T_l\},\end{aligned}$$

[1] The HMM has been extended to the case where $U[k]_i$ takes more than 2 values. We only describe the case of binary value due to the limited space.

$$\Pr\{X[k] \in \mathcal{A} | X[k-1] = x, U[k]_2 = u\}$$
$$= \int_{x' \in \mathcal{A}} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\{-\frac{\big(x' - [\lambda x + (1-\lambda)\theta_o[k] + ug]\big)^2}{2\sigma_X^2}\},$$

where χ is the indicator function, η is close to 0.

The dependence between $Z[k]$ and $U[k]$ is given by

$$Z[k] = \check{\mu}(U[k]_2 - U[k]_1) + \epsilon[k]$$

where $\epsilon[k]$ is a Gaussian random variable of mean 0 and variance σ_Z^2.

C. Algorithms

The parameters that appear in the HMM are divided into two groups: The first group includes $T_l, \lambda, \check{\mu}, \{\alpha_v, \mu_v, \sigma_v^2\}$; The second group includes $\sigma_X^2, \sigma_Z^2, g, \eta$. The second group are the same for all houses, and given the availability of training data can be optimized to minimize the estimation error. The first group on the other hand, varies with different houses and are estimated from the usage data.

The disaggregation algorithm computes the maximum-likelihood estimates of $T_l, \lambda, \check{\mu}, \{\alpha_v, \mu_v, \sigma_v^2\}$ and $\{U[k]\}$. The estimate of heating usage is given by $\hat{w}[k] = \check{\mu}U[k]_2$.

The maximum-likelihood estimation of $\lambda, \check{\mu}, \{\alpha_v, \mu_v, \sigma_v^2\}$ and $U[k]$ uses the well-known Baum-Welch algorithm [22],[2] which is guaranteed to converge to a local optimum. The same estimation procedure is repeated with different choices of T_l and the one yielding the maximum likelihood value is used.

V. Evaluations

In this section, the proposed disaggregation algorithm is applied towards two problems: detection of the existence of electric heating and estimation of heating usage per household.

A. Data preparation

Two data sets are used in the experiment: The first data set is real-world data. It is used in the detection problem. It includes two weeks of electricity usage of 76 houses in winter, measured at 15-minutes sampling interval, hourly temperature measurements at a weather station, and a survey filled out by the consumer that describes what type of heating appliance is used. The survey data is converted to a binary label indicating whether the house has electric or non-electric heating. Among the 76 houses, 47 of them were labeled as having electric heating. Hourly temperature is interpolated to obtain out-door temperature at 15-minutes interval.

The second data set is synthetic. It is created for the estimation problem to evaluate the accuracy of the heating usage disaggregation, since the real-world data set does not include the ground truth of heating usage. The synthetic data includes the usage of 100 *virtual* houses generated as follows: The electric heating usage of a virtual house is simulated using HAMbase-S software [23], and the non-heating usage of a virtual house is obtained by taking the real-world usage of houses with non-electric heating.

The simulation of electric heating usage uses the one-zone building model given in [23], in which multiple difference equations are used to compute the thermal dynamics, and the thermal resistance of the walls are chosen so that the lumped thermal resistance is within the range given in [4, Table 1]. The heating appliance has only on and off states controlled by a thermostat. Only one heating usage sequence is simulated and used for all the virtual houses. Modeling the impact of the outdoor temperature on the power and efficiency of the heating appliance, and creating more than one sequence of electricity usage by simulating houses with different thermal parameters will be pursued in future research.

The non-heating usage is obtained from the real-world usage data from 100 real houses labeled as non-electric heating. These houses are evenly divided into two groups according to the size and value of the house and the year when it is built. One group includes larger, newer and more expensive houses and the other group includes smaller, older and less expensive houses. The reason to consider two groups instead of one is based on the observation that on average larger houses have higher non-heating electricity usage than small houses.

B. Detection of electric heating

Currently about a half of the houses in the Unites States built after 2000 are heated using electricity, and newer houses are more likely to have electric heating [24]. The information whether a house uses electric heating is useful, for example, in HEMS in which a household's usage is compared against a selected group of other households to provide feedback on its consumption, since houses heated using electricity generally have higher electricity usage than others.

In this detection problem, the input is the total usage and temperature, and the output is whether the house has electric or non-electric heating. There are two types of errors: type I error in which a house without electrical heated is incorrectly detected as electric heating, and type II error in which a house with electric heating is detected as having no electric heating. The performance of a detection algorithm is characterized by the probabilities of these two errors:

$$\text{Type I error probability} = \frac{\text{Number of Type I errors made}}{\text{Number of houses with no electric heating}},$$
$$\text{Type II error probability} = \frac{\text{Number of Type II errors made}}{\text{Number of houses with electric heating}}.$$

A threshold test is applied to the output of disaggregation algorithms: A house is detected as electrically heated if the estimated average heating power is larger than a threshold. The threshold dictates the trade-off between the two errors.

The error probabilities using HMM and linear regression methods, with varying thresholds, are depicted in Figure 2. In addition the threshold test is applied directly to the raw aggregate usage to obtain a baseline error. In Figure 2, the extreme right is a low threshold, where every house is detected as having electric heating, while the extreme left is a high threshold where no houses are detected as electrically heated.

Three observations are made: First, detection using either disaggregation algorithm performs better than that using raw usage; Second, the probability of error for the linear regression method is almost the same for 15-minutes and 1-hour measurements; Third, the HMM method is worse than linear regression at one end of the curve and better at the other. So from a

[2]We use the standard procedure of the Baum-Welch algorithm, see the tutorial [16] for details.

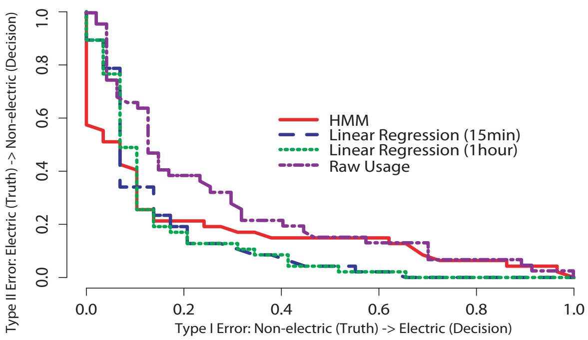

Fig. 2. Probability of detection errors: HMM method with a 15-minute sampling interval (red and solid), linear regression with a 15-minute sampling interval (blue and dashed) or 1-hour sampling interval (green and dotted), threshold test on raw usage (purple and dash-dotted). $\theta_B = 18.3°C$.

purely detection perspective, the linear regression method is favorable, considering its low computational complexity and less stringent requirement on the sampling frequency.

The best method and best threshold value also depends on the cost of a mis-diagnosis. For example, to classify a house as having electric heating when it actually is non-electric (Type I error), might result in a foolish recommendation that the house participate in a demand response program. Similarly, a missed detection (Type II error), i.e., the house is diagnosed as not having electric when it actually does, would result in a lost opportunity. In general we want to balance the two error types and choose a threshold somewhere in the middle, for example, where the curve is closest to the lower left corner of the graph.

It is observed from the usage data that some electrically heated houses consume very little electricity during the night, indicating that the heating might be manually turned off. Future work will exploit the flexibility of the HMM to incorporate such information to improve the detection performance.

C. Estimation of heating usage

The disaggregation algorithm estimates heating usage. Using synthetic data, the accuracy of the algorithm is evaluated by its estimation error, which is defined as:

$$\text{Estimation error} = |1 - (\sum_k \hat{w}[k])/(\sum_k w[k])|.$$

Table I shows estimation errors for HMM and linear regression methods. The results are the mean and 1 standard deviation, and are categorized by the type of house (larger / smaller houses), and the sampling interval (15-minutes / 1-hour). The house type is described in Section V-A. As seen in the table, the HMM method had lower error, especially for tight sampling (15 minute).

Three observations are made: First, the estimation error of the HMM method increases with lower sampling frequency, while the linear regression method is insensitive to the sampling frequency. The reason is that the HMM method uses the difference $Y[k] = y[k] - y[k-1]$ to infer the on/off event of the heating device and thus is affected by how the frequency of the on/off event compares to the sampling frequency. With 1-hour sampling, the second assumption of the HMM method is violated, and the usage in two or more consecutive on-periods sometimes is lumped together.

Second, the error of the linear regression method is larger for larger houses which have higher non-heating usage, compared to those of smaller houses.This is due to the linearity of the regression method. The error of the HMM method does not significantly depend on the amount of non-heating usage since the model explicitly accounts for this. Third, the HMM method with 15-minutes data is more accurate than linear regression.

	Regression	HMM
Larger Houses, Hourly	0.909 +/- 0.477	0.730 +/- 0.100
Larger Houses, 15 Min	0.901 +/- 0.479	0.274 +/- 0.153
Smaller Houses, Hourly	0.669 +/- 0.270	0.669 +/- 0.205
Smaller Houses, 15 Min	0.681 +/- 0.284	0.361 +/- 0.159

TABLE I
ERROR OF ESTIMATION METHODS.

VI. ENERGY SERVICES

We describe applications of the heating usage disaggregation algorithm to demand response and energy audit services. In particular, we show how the algorithm can be used to estimate the demand response potential of a house. We also discuss deployment considerations.

A. Demand Response

Consider the demand response program in which the utility directly changes the on/off state of heating appliances to temporarily change the heating usage, and thus provides demand response to the grid. To maintain consumer's comfort level, the heating appliance only responds to the control signal if the indoor temperature is within the deadband $[\theta_l, \theta_u]$. Therefore, the potential contribution of a household to a demand response program depends on its heating usage characteristics.

The disaggregation algorithm can be used to estimate individual households' potential contribution. The utility could then apply targeted advertisement or give higher incentives to those houses with high potentials to encourage them to participate in demand response.

One important metric of the demand response service is the power capacity over a required duration, i.e., the amount of power that the demand response service can provide consistently over this time interval. Contribution of each household to the power capacity is limited by both the average heating power as well as the average energy consumed during each on/off period. Assuming that the same amount of power capacity must be provided for both increasing and decreasing demand, we use the following formula based on [4] to estimate the potential contribution of a house:

$$\text{potential of a house} = \min\{\text{Average heating power}, \frac{\text{Average energy consumed in each period}}{\text{Duration as a demand response resource}}\}.$$

The required duration as a demand response resource is determined by the type of services it provides. Here we consider a duration of 2 hours for load following. The HMM-based disaggregation algorithm is applied to estimate the average energy consumed in an on-period and the average heating power for the 47 electrically heated houses in the real-world data. A house is recommended to participate in demand response if its estimated contribution is larger than a threshold.

Varying the threshold leads to different number of recommended houses, as well as different total power capacity and

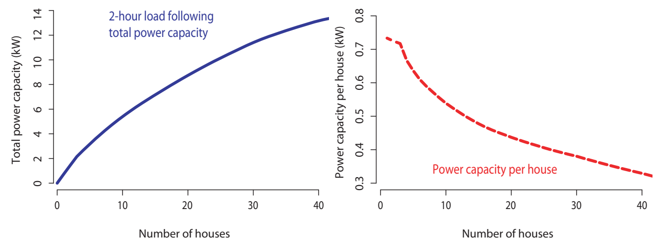

Fig. 3. Total power capacity (left) and average power capacity per house (right), as a function of the number of houses recommended.

average power capacity contributed per house, as depicted in Figure 3. If the cost of demand response program includes both fixed cost per customer and payments according to actual contributions, then both the total capacity and the average capacity per house should be considered in designing and optimizing the demand response program.

B. Energy Audit

Disaggregation algorithms give the estimated heating usage for individual houses. This information can be used to provide feedback to consumers on how they compare to other similar households. Appropriate comparison group for a household may be based on the size or age of the house. Such groupings can be identified using public information. Thus it is possible to provide consumers with information on the amount of savings they may get from an energy audit.

C. Deployment considerations

Data Availability: The disaggregation algorithm requires individual meter usage and outdoor temperature data. We have shown that the performance of the HMM method depends crucially on the sampling frequency. Therefore, it is important to find the optimal sampling frequency prior to deploying this method. The optimal sampling frequency depends on the typical on/off cycle of heating appliances, which can be estimated from a survey, or collecting and analyzing high-frequency usage data from a few representative homes.

Computation: The computational complexity of both the HMM and linear regression methods increases linearly with the number of data samples. The HMM method's computational complexity per data sample is much higher than that of linear regression, and depends on the number of iterations. However this is not a big concern for energy services where the disaggregation algorithm is applied off-line. Moreover, the rich literature on HMM can be leveraged to find approximate methods with smaller computational complexity.

Consumer Population: Consumer characteristics such as people migration, especially those living in rented places could lead to different usage patterns for the same meter. The learning for the disaggregation algorithm is specific to a particular consumer's usage pattern.

VII. Conclusion and Future Work

We have developed an HMM-based algorithm to estimate individual household heating usage from aggregate smart meter data. We have shown how these estimates can be used to design customized energy services that benefit the end-consumer. Even though linear regression based approaches can solve the detection problem it is necessary to have sophisticated methods like HMM to delineate the usage patterns in an intelligent manner. This is essential to create customized energy services for different consumer groups. Similar methods can be extended to extract other information, such as AC or pool pump usage. We are currently pursuing mechanisms to identify appropriate aggregation groups of households for different types of energy services.

References

[1] Pacific Gas and Electric Company, http://www.pge.com/myhome/customerservice/smartmeter/facts/.
[2] U.S. Energy Information Administration, "Annual energy outlook 2012," *DOE/EIA-0383(2012)*, Jun. 2012.
[3] N. Lu, "An evaluation of the HVAC load potential for providing load balancing service," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1263 – 1270, Sep. 2012.
[4] J. Mathieu, D. Callaway, and M. Dyson, "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations," in *2012 ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.
[5] National Climatic Data Center, http://www.ncdc.noaa.gov/cdo-web/.
[6] General Electric Nucleus energy manager, http://www.geappliances.com/home-energy-manager/.
[7] EcoFactor, http://www.ecofactor.com/.
[8] Open Home, http://openhome.org/.
[9] OPOWER, http://opower.com/what-is-opower/reports/.
[10] R. T. Curtis and D. Copeland, "Method and system for disaggregating heating and cooling energy use from other building energy use," Nov. 2010, US Patent App. 12/940,447.
[11] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870 – 1891, December 1992.
[12] J. T. Powers, B. Margossian, and B. A. Smith, "Using a rule-based algorithm to disaggregate end-use load profiles from premise-level data," *IEEE Computer Applications in Power*, vol. 4, no. 2, pp. 42 – 47, 1991.
[13] A. Prudenzi, "A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel," in *IEEE Power Engineering Society Winter Meeting, 2002*, vol. 2, 2002, pp. 941 – 946.
[14] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings*, vol. 24, no. 1, pp. 51 – 64, 1996.
[15] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the flick of a switch: Detecting and classifying unique electrical events on the residential power line," in *Proceedings of the 9th International Conference on Ubiquitous Computing*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 271 – 288.
[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 – 286, 1989.
[17] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements." in *Proceedings of the SIAM Conference on Data Mining*, 2011, pp. 747 – 758.
[18] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, 2011.
[19] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Twenty-Sixth Conference on Artificial Intelligence (AAAI-12)*, Apr. 2012.
[20] S. Ihara and F. C. Schweppe, "Physically based modeling of cold load pickup," *IEEE Transactions on Power Apparatus and Systems*, no. 9, pp. 4142 – 4150, 1981.
[21] R. E. Mortensen and K. P. Haggerty, "A stochastic computer model for heating and cooling loads," *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1213 – 1219, Aug. 1988.
[22] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164 – 171, 1970.
[23] A. W. M. van Schijndel, "Integrated heat air and moisture modeling and simulation," Ph.D. dissertation, Technische Universiteit Eindhoven, 2007.
[24] U.S. Energy Information Administration, "2009 residential energy consumption survey data," 2009.