15 May 2016

An Accurate Method of Energy Use Prediction for Systems with Known Consumption

Jacob A. Mueller

Jonathan W. Kimball
Missouri University of Science and Technology, kimballjw@mst.edu

An Accurate Method of Energy Use Prediction for Systems with Known Composition

Jacob A. Mueller and Jonathan W. Kimball
Department of Electrical and Computer Engineering
Missouri University of Science and Technology
Rolla, MO 65409

***Abstract*—An improved method of nonintrusive load monitoring (NILM) using hidden Markov models is proposed. The proposed method is intended for electrical systems with a closed and known set of devices. Systems that meet this constraint are more commonly found in industrial or commercial contexts than the residential settings in which NILM is traditionally studied. The proposed method is designed to support applications relevant to these contexts, such as fault monitoring and system health assessments. To this end, the energy predictions of the disaggregation algorithm offer accurate descriptions of device behavior in time. The accuracy of the proposed method is validated using standard NILM performance metrics and data from public databases.**

***Index Terms*—Nonintrusive load monitoring, device modeling, hidden Markov models.**

I. Introduction

The value of energy use information provided by an accurate nonintrusive load monitoring system is well-known. Consumers, industry, and utilities all stand to benefit from the continued development of NILM technologies. This breadth of applicability has led to rapid growth in NILM research. Of the many established NILM methods, those constructed around hidden Markov models (HMMs) and their generalizations have enjoyed particular success in recent years [1]–[4].

From the beginning, NILM research been heavily motivated by the goal of providing accurate energy use assessments for appliances in residential settings [5], [6]. Residential electrical systems typically contain a similar set of devices, but are complicated by the fact that the number of devices in the system is continually changing. NILM systems address this challenge using unsupervised training methods or provisions for handling unmodeled loads [7]. The desired output for these systems is a breakdown of the total energy use by device. Energy use predictions are typically calculated from the peak or mean values of the devices' conditional observation distributions given the disaggregated state sequence. This bulk energy use information is useful for many load monitoring applications, but provides limited information regarding device energy use in time.

This study presents a NILM method for characterizing device behavior in electrical systems with known composition. "Known composition" is defined here to mean that the number of loads in the system is known and well-trained models of each load are available. This assumption represents a departure from the more traditional residential context, in which the set of loads is continually changing. The assumption is, however, met by pre-designed electrical systems, such as power systems of satellites, data centers, and electric vehicles of all kinds. For these systems, early warnings of possible device damage are critical, as they affect mission planning, scheduling of maintenance, and other reliability-centered functions. The target application of the proposed method is motivated by these concerns. Namely, the method is intended to aid in detection and identification atypical device behavior, which may be indicative of equipment damage or impending failure. By eliminating the possibility of unmodeled loads, the energy used by each device can be determined with higher precision, and unexpected energy use may be interpreted as the result of changes in device status, rather than contributions of unknown devices.

The proposed method provides energy use profiles in the form of predicted observation sequences that maximize the total likelihood at each sampling instant. A modular training process allows device models to be generated independently, and then combined later as necessary. This results in a system-level description that does not require aggregate training data nor simultaneously collected device data. The method shares the same scalability issues of other HMM-based NILM methods, but is able to use recent sparsity-exploitation approaches to mitigate these problems. In the following sections the proposed method is outlined and results are discussed for a test system of data from public NILM databases.

II. Methodology

The proposed NILM method consists of model training, system-level representation, state sequence disaggregation, and energy use prediction. The observations for the underlying HMMs are instantaneously sampled average active power measurements. That is, the observations are the single-cycle average power during the line frequency period immediately preceding the sampling instant. For dc systems, observations may simply be instantaneous power. States correspond to devices' operational modes. Under these assumptions, devices are accurately modeled by finite state machines with Gaussian conditional observation distributions for each state.

A. Training

As the devices are assumed to be known a priori, a supervised method of training is sufficient to build device models.

From recorded observation data for a single device, the number of states and range of observations for each state may be determined. A corresponding sequence of states can then be found for the sequence of observations [8]. There are several methods that may be used to define range of observations for the states. One such method is given in [1]. The conditional observation distributions for each state are assumed to be Gaussian, and their mean and variance parameters may found using maximum likelihood estimates.

Let device k be (invasively) measured for duration T, where $T \in \mathbb{Z}^+$ is the number of discrete sample instances. Then the resulting observation sequence is $O^{(k)} = \{O_1^{(k)}, O_2^{(k)}, \cdots, Q_T^{(k)}\}$, where each $O_t^{(k)} \in \mathbb{R}^+$. If device k has N_k states, then the sequence of states is $Q^{(k)} = \{Q_1^{(k)}, Q_2^{(k)}, \cdots, Q_T^{(k)}\}$, and $Q_t^{(k)} \in \{1, \cdots, N_k\}$.

Each device's HMM consists of transition matrix, A, observation distributions, ϕ, and initial probability occupation vector, π. Elements of the transition matrix are estimated as

$$A_{ij}^{(k)} = \frac{\sum_{t=1}^{T-1}[I(Q_t = i) \cdot I(Q_{t+1} = j)]}{\sum_{t=1}^{T-1} I(Q_t = i)} \quad (1)$$

where $I(\cdot)$ is the indicator function. Observation distributions for each state of each device are assumed to be Gaussian, i.e. $\phi_i(O_t) \sim \mathcal{N}(\mu_i, \sigma_i)$. For device k, estimates for parameters $\mu_i^{(k)}$ and $\left(\sigma_i^{(k)}\right)^2$ are the sample mean and sample variance, respectively.

$$\mu_i^{(k)} = \frac{\sum_{t=1}^{T}[O_t^{(k)} \cdot I(Q_t^{(k)} = i)]}{\sum_{k=1}^{T} I(Q_t^{(k)} = i)} \quad (2)$$

$$\left(\sigma_i^{(k)}\right)^2 = \frac{\sum_{t=1}^{T}[(O_t^{(k)} - \mu_i^{(k)})^2 \cdot I(Q_t^{(k)} = i)]}{\sum_{t=1}^{T} I(Q_t^{(k)} = i)} \quad (3)$$

B. Model Combination and State Disaggregation

To represent a system of multiple devices, the individual device models are combined into a single composite model. The composite model is itself an HMM, and has the same parameters $\boldsymbol{A}$, $\boldsymbol{\phi}$, and $\boldsymbol{\pi}$ as the device models. Bold type indicates that these parameters pertain to the composite model. In the composite model, observation probabilities are treated as discrete quantities, so $\boldsymbol{\phi}$ takes the form of a matrix.

The composite transition matrix is constructed as the Kronecker product of the device transition matrices. For two matrices X and Y, the Kronecker product $X \otimes Y$ produces a block matrix Z, such that the product of element X_{ij} and matrix Y is the ij^{th} block element of Z. Defined recursively, the composite transition matrix after the inclusion of the k^{th} device model is

$$\boldsymbol{A^{(k)}} = \boldsymbol{A^{(k-1)}} \otimes A^{(k)} \quad (4)$$

where $A^{(k)}$ is the transition matrix of the k^{th} device, and $\boldsymbol{A^{(0)}}$ is 1. The order in which device models are included in the composite model is used to extract device-level state sequences in the disaggregation stage.

To construct the composite observation matrix, the devices' observation distributions must be converted to a discrete representation. A fixed bin size B is chosen and the range of possible observations is partitioned into discrete intervals. For each device state, the probability of an observation in a given bin is the integral of the conditional observation distribution over that bin. Since each device state corresponds to a single Gaussian observation distribution, many bins will have very low observation probabilities. A minimum threshold ϵ may be set such that the bin probabilities are fixed to 0 if the calculated values are less than ϵ. This parameter influences the sparsity of resulting composite observation matrix. For this study, $\epsilon = 10^{-9}$ was used. The choice of B is determined by the resolution of the measurement device to be used in implementation of the system and the characteristics of the observation distributions. This selection is a design tradeoff: the size of B must be small enough that the lowest power state of the lowest power device is not lost, but an overly small B will lead to large memory requirements for the composite observation matrix.

Once the discrete observation probabilities have been calculated, the composite observation matrix is constructed again using the Kronecker product and summing columns that correspond to equal observations. When the composite observation matrix has been determined, the individual devices' observation matrices may be discarded; only the parameters of the continuous distributions need to be retained.

In order to apply the composite model, the aggregate observation is binned using the same B as was used for the model combination process, such that the sequence of observations is integer-valued. Since the composite model is itself an HMM, the Viterbi algorithm may be used to determine the most likely sequence of composite states for the observation sequence. However, the time complexity of the Viterbi algorithm is problematic for systems with large numbers of states, and the number of states in the composite system scales exponentially with the number of devices. To address this issue, the sparse Viterbi algorithm proposed in [9] is used, and composite matrices are stored in a sparse matrix format.

Regardless of whether the standard Viterbi algorithm or sparse Viterbi algorithm is used, the resulting sequence of composite states can then be broken down into individual sequences of device states. The disaggregated state sequences are

$$\hat{Q}_t^{(k)} = \mathrm{mod}\left(\mathrm{ceil}\left(\frac{\hat{Q}_t}{\prod_{i=k+1}^{K} N_i}\right) - 1, N_k\right) + 1, \quad (5)$$

where $\hat{Q}_t$ is the composite Viterbi path at instant t, K is the total number of devices in the system, and mod and ceil are the modulo and ceiling functions, respectively. For device K, the last included in the composite model, the form of (5) reduces to

$$\hat{Q}_t^{(K)} = \mathrm{mod}\left(\hat{Q}_t - 1, N_K\right) + 1 \quad (6)$$

C. Energy Prediction

The approach to energy prediction is the principal difference between the proposed method and existing NILM procedures.

In other methods, the energy use is predicted as the expected value or peak value of the device's conditional observation distribution given the disaggregated device state. While this results in accurate predictions of cumulative energy use, the energy use predictions in time are constant when the state is not changing. For many devices, this offers an inadequate approximation of real behavior.

The assumption of known system composition allows the predicted device observations to be constrained to equal the measured aggregate observation. Let O be the aggregate observation, and let $\hat{O}$ be the K-tuple of device observation predictions. The predictions are subject to

$$g(\hat{O}) = \sum_{i=1}^{K} (\hat{O}^{(k)}) - O = 0. \tag{7}$$

For notational convenience, time indices have been dropped. Because the devices are independent, the total probability of this observation prediction is the product of the conditional observation probabilities for each device. Working in logspace, this probability is described by

$$f(\hat{O}) = \ln\left(P\big[\hat{O}\big]\right) = \sum_{i=1}^{K} \ln\left(\phi^{(i)}\big(\hat{O}^{(i)}|\hat{Q}^{(i)}\big)\right), \tag{8}$$

where $\hat{Q}^{(k)}$ is the disaggregated state of device k.

Using Lagrange multipliers, $\nabla f = \lambda \nabla g$. The i^{th} summation term of (8) depends only on the i^{th} element of $\hat{O}$. For $\hat{Q}^{(i)} = j$, the partial derivative is then

$$\frac{\partial f(\hat{O})}{\partial \hat{O}^{(i)}} = \frac{\partial}{\partial \hat{O}^{(i)}} \ln\left(\phi_j^{(i)}\big(\hat{O}^{(i)}\big)\right) = \frac{\big(\mu_j^{(i)} - \hat{O}^{(i)}\big)}{\left(\sigma_j^{(i)}\right)^2}. \tag{9}$$

The convenient linear form of this equation results from the assumption of Gaussian observation distributions. The i^{th} term of $\lambda \nabla g$ is just λ, so each $\hat{O}^{(i)}$ may be expressed in terms of λ and mean and variance parameters, which are known.

$$\hat{O}^{(i)} = \mu_j^{(i)} - \lambda \left(\sigma_j^{(i)}\right)^2 \tag{10}$$

Substituting these terms into (7), λ may be calculated from the aggregate observation and the sums of the means and variances of the appropriate device states.

$$\lambda = \frac{\left(\sum_{i=1}^{K} \mu_{\hat{Q}^{(i)}}^{(i)}\right) - O}{\sum_{i=1}^{K} \left(\sigma_{\hat{Q}^{(i)}}^{(i)}\right)^2} \tag{11}$$

The individual observation predictions are then calculated by substituting λ back into (10) for each device.

III. Test Results

To verify the performance of the proposed method, a set of experiments were conducted using data from public NILM datasets. Data from Tracebase [10], Smart$*$ [11], and GREEND [12] datasets were used. Observation data were collected for 6 devices over 10 days. Each observation set was downsampled to a period of 5 s. To eliminate bias, a 10-fold cross validation procedure was used; the observations were split into 9 days of training data and 1 day of test data. The aggregate test data was generated as the sum of the test devices' observations, with each device's initial start time randomly generated. For each fold, 10 trials were conducted, corresponding to 10 random device start times. A relatively small set of devices was chosen so that the accuracies resulting from the standard Viterbi algorithm and sparse Viterbi algorithm could be compared. As a further point of comparison, the energy predictions calculated from expected observation values conditional on the disaggregated states were assessed as well. The accuracies of the state and energy predictions were measured using the standard NILM metrics discussed in [13]. Namely, F-Score and modified finite state F-Score were used for state accuracy, and estimation accuracy [14] was used to assess energy predictions.

TABLE I
Accuracy Assessments - Standard Viterbi Algorithm

Device	F-Score/ FS-Fscore	Est. Acc. Proposed Method	Est. Acc. Expect. Method
Stove	0.9511	0.9596	0.9561
Dryer	0.9782	0.9951	0.9919
Dishwasher	0.5007	0.9107	0.9080
Microwave	0.7953	0.7348	0.7323
Refrigerator	0.8208	0.9157	0.8623
LCD TV	0.9309	0.9350	0.9226

TABLE II
Accuracy Assessments - Sparse Viterbi Algorithm

Device	F-Score/ FS-Fscore	Est. Acc. Proposed Method	Est. Acc. Expect. Method
Stove	0.9553	0.9644	0.9608
Dryer	0.9834	0.9953	0.9925
Dishwasher	0.5081	0.9674	0.9651
Microwave	0.7221	0.8743	0.8732
Refrigerator	0.8754	0.9185	0.9080
LCD TV	0.9011	0.9088	0.8943

Accuracies for the standard Viterbi tests and sparse Viterbi tests are shown in Table I and Table II, respectively. The proposed method is more accurate in all cases. On average, the increase in accuracy is +1.29%. The accuracy metric used normalizes predictive error by the device's total energy, so an increase of this magnitude is expected. The improvement provided by the proposed method is its ability to predict transient features. These characteristics are important to the identification of correct device operation, but typically make up a small fraction of the cumulative energy.

The difference in predictive performance is best demonstrated by the results for the refrigerator and LCD TV. The refrigerator's behavior is characterized by transient spikes in active power corresponding to the inrush current of the starting compressor. The transient is consistent enough to require its own dedicated state. The TV is an electronic device, and experiences small variations in power as the image on the screen

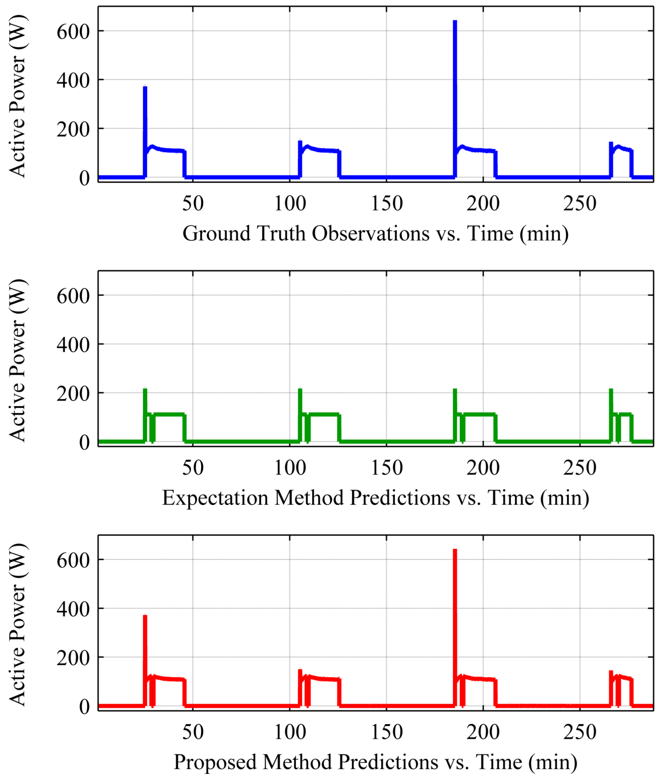

Fig. 1. Comparison of refrigerator ground truth and predicted observations.

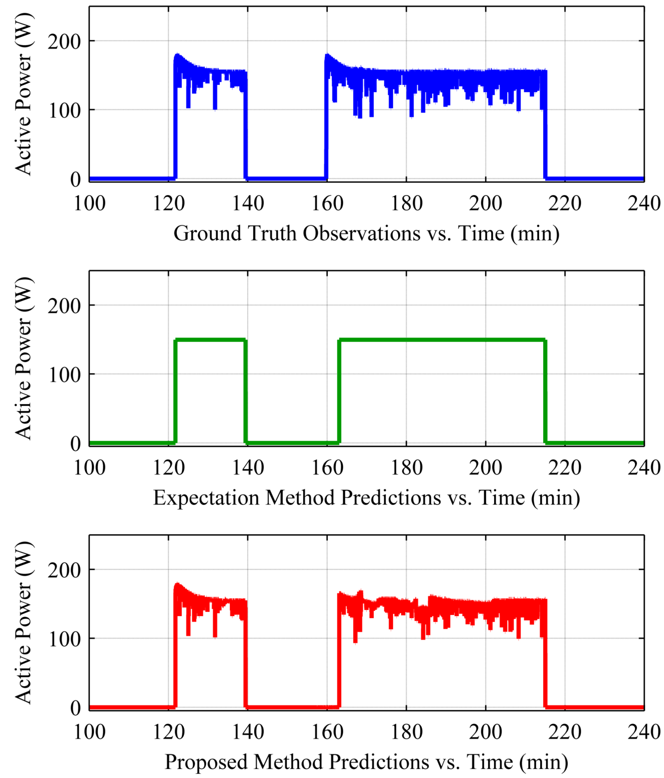

Fig. 2. Comparison of LCD TV ground truth and predicted observations.

changes. These variations are not large enough to necessitate their own states, but are an immediately recognizable feature in the device's energy use profile. A comparison of the ground-truth observations and predictions of the two methods for these devices are shown in Fig. 1 and Fig. 2, respectively. The proposed method is much more capable of predicting these transient effects, while the expectation method provides only a rough approximation. The performance difference shown in these plots is indicative of the overall improvement offered by the proposed method: the refrigerator predictions shown in Fig. 1 represent a increase accuracy of 1.75%, and TV predictions in Fig. 2 represent an increase of 1.36%.

These two devices represent critical areas in the space of devices. The transient behavior of the refrigerator is typical of devices with inductive or capacitive load characteristics, while the TV's power variations due to operational changes are typical of electronic loads. While the total energy use during these transients and variations is small compared to the devices' cumulative energies, their recognition is of critical importance to applications involving the identification of atypical device behaviors.

IV. Conclusion

In this study, an HMM-based NILM method has been extended for systems with known composition. The proposed method provides accurate descriptions of device behavior in time, but requires a static and well-modeled system. Since the system is well-known, the observation predictions for each device may be constrained to equal the measured aggregate observation sequence. The process of determining the device-level observations is then reduced to a constrained optimization problem, which maximizes the likelihood of the predicted observations at each sampling instant. This provides an additional level of accuracy without affecting the functionality of the base NILM method. Moreover, this increased accuracy is possible with only minimal increases to algorithm complexity. The method is most useful for recognizing transient device behavior and energy use variations within a single device state, where the expectation method provides only rough approximations.

In situations where the known system composition assumption is met, the proposed method may be used to support NILM-based system health assessments through the identification of atypical device behaviors. This functionality is particularly relevant to industrial and mobile power systems, which are more likely to contain only known and well-modeled devices. The proposed method of energy estimation is compatible with the sparse Viterbi algorithm, which provides the capability for online operation and allows a larger set of devices to be used before algorithm time-complexity becomes problematic. Nonetheless, the proposed method is still subject to the scalability challenges experienced by all HMM-based NILM approaches. Further work in this study will explore fault monitoring procedures based on this method, and look at new ways of increasing the scalability of the disaggregation algorithm as a whole.

REFERENCES

[1] S. Makonin, F. Popowich, I. Bajic, B. Gill, and L. Bartram, "Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–11, 2015.

[2] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artificial Intelligence and Statistics*, vol. 22, 2012, pp. 1472–1482.

[3] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *26th Conference on Artificial Intelligence*, vol. 22, 2012, pp. 356–362.

[4] T. Zia, D. Bruckner, and A. Zaidi, "A hidden Markov model based procedure for identifying household electric loads," in *37th Conf. Industrial Electronics*, Nov 2011, pp. 3218–3223.

[5] F. Sultanem, "Using appliance signatures for monitoring residential loads at meter panel level," *IEEE Trans. Power Del.*, vol. 6, no. 4, pp. 1380–1385, Oct 1991.

[6] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec 1992.

[7] S. Pattem, "Unsupervised disaggregation for non-intrusive load monitoring," in *11th Int. Conf. Machine Learning and Applications (ICMLA)*, vol. 2, Dec 2012, pp. 515–520.

[8] J. A. Mueller, A. Sankara, J. W. Kimball, and B. McMillin, "Hidden markov models for nonintrusive appliance load monitoring," in *North American Power Symposium (NAPS), 2014*, Sept 2014, pp. 1–6.

[9] S. Makonin, F. Popowich, and I. Bajic, "Efficient matrix processing for nonintrusive load monitoring (NILM)," in *2nd. Int. Workshop on Nonintrusive Load Monitoring*, June 2014.

[10] A. Reinhardt *et al.*, "On the accuracy of appliance identification based on distributed load metering data," in *Proc. 2nd IFIP Conf. Sustainable Internet and ICT for Sustainability (SustainIT)*, 2012, pp. 1–9.

[11] S. Barker *et al.*, "Smart*: An open data set and tools for enabling research in sustainable homes," in *Proc. 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012)*, Aug 2012.

[12] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. Tonello, "GREEND: An energy consumption dataset of households in Italy and Austria," in *IEEE Int. Conf. Smart Grid Communications (SmartGridComm)*, Nov 2014, pp. 511–516.

[13] S. Makonin and F. Popowich, "Nonintrusive load monitoring (NILM) performance evaluation," *Energy Efficiency*, vol. 8, no. 4, pp. 809–814, 2014.

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