How well can HMM model load signals

Lukas Mauch, Karim Said Barsim and Bin Yang
{lukas.mauch,karim.barsim,bin.yang}@iss.uni-stuttgart.de
Institute of Signal Processing and System Theory, University of Stuttgart

***Abstract*—Finding models that can efficiently represent load signals is one key issue in non-intrusive load monitoring (NILM) because they are the foundation of most load disaggregation algorithms. In the past, the factorial hidden Markov model (fHMM) has been proposed as one probabilistic model for the aggregate real power measurement. It is assumed that each load can be represented as one hidden Markov model (HMM) and the HMMs of all loads have been learned successfully before disaggregation. Although fHMM showed some promising results for eventless disaggregation, a detailed investigation on how well HMM is suited to model load signals is still missing till today. In this paper, we study the feasibility of HMM modeling for different categories of loads by using the UK-DALE dataset and propose a method for model adaptation across different houses.**

***Index Terms*—Non-intrusive load monitoring (NILM), hidden Markov model (HMM), model selection, model adaptation**

I. Introduction

Non-Intrusive load monitoring (NILM) is a challenging single channel source separation problem. One of the key issues to solve the separation problem is to find suitable signal models which are able to capture the major properties of different load signals and which can be used to infer the single load components from the aggregate signal.

Finding suitable signal models for NILM is a challenging task not yet solved satisfactorily. Load signals show strong hierarchical time dependencies ranging from seconds, hours to days, weeks and even longer. There are different ways to model load signals. Current event-based disaggregation algorithms divide the measured power signal into segments and model it as a sequence of piecewise constant power levels [1], [2], [3], [4]. For eventless approaches, the factorial hidden Markov model (fHMM) has been proposed as a probabilistic model for the aggregate signal [5], [6], [7], [8]. The key idea of fHMM is that each load contained in the aggregate power signal is modeled by one hidden Markov model (HMM) and the aggregate power signal is the sum of the individual power signals of all loads.

Recently, we proposed a novel combination of HMM with deep neural network (DNN) for load disaggregation [9]. Its basic difference to and its major advantage over fHMM are that not all loads need to be modeled by an HMM (impossible in practice). Instead, our DNN-HMM approach needs only HMM models for the major power loads to be extracted from the aggregate power signal. The remaining loads are considered as noise components and are suppressed by the denoising DNN.

Although fHMM and in particular DNN-HMM showed quite promising results for load disaggregation, there is no investigation yet which loads can be modeled efficiently by HMM. A number of questions remain open: a) Is HMM a suitable model for variable loads? b) Is HMM always a good model for multi-state loads? c) Are the HMM states identical to the load states? d) How to choose the number of HMM states? e) How to adapt a learned HMM model to another similar but not identical load, e.g. in another house [10]?

In this paper, we do not study the disaggregation problem. Instead we go back to the roots and address the above basic questions for the HMM modeling of load signals, supported by experiments on the UK-DALE dataset [11]

II. Hidden Markov Model

A. Basics

An HMM consists of a hidden Markov chain of states $s(1), \dots, s(N)$ of length N and a corresponding sequence of observations vectors $\underline{x}(1), \dots, \underline{x}(N)$. Modelling loads with HMM, the observation could consist of two-dimensional real and reactive power measurements. In this paper, we consider real power measurement $x(n) \in \mathbb{R}$ only. $s(n) \in \{1, \dots, M\}$ is the state at discrete time n and can take M different state values. We assume a first-order Markov chain which is characterized by the initial state probabilities

$$\pi_i = P(s(1) = i),\ 1 \le i \le M,\ \ \underline{\pi} = [\pi_1, \dots, \pi_M]^T \in \mathbb{R}^M \quad (1)$$

and the time-invariant state transition probabilities

$$a_{ij} = P(s(n) = i | s(n-1) = j),\ 1 \le i, j \le M \quad (2)$$

$$\mathbf{A} = [a_{ij}] \in \mathbb{R}^{M \times M} \quad (3)$$

with $\sum_{i=1}^{M} a_{ij} = 1\ \forall 1 \le j \le M$. The continuous-valued observation $x(n) \in \mathbb{R}$ is assumed to have the time-invariant state-conditional probability density function (pdf)

$$\rho_i(x(n)) = p(x(n)|s(n) = i),\ 1 \le i \le M. \quad (4)$$

In this paper, we assume a Gaussian pdf $\rho_i(x(n)) \sim N(\mu_i, \sigma_i^2)$. Let $S = \{s(1), \dots, s(N)\}$ and $X = \{x(1), \dots, x(N)\}$ denote the state and observation sequence, respectively. Let $\underline{\theta} = \{\underline{\pi}, \mathbf{A}, \{\mu_i, \sigma_i\}\}$ be the vector containing all HMM parameters. Then the joint pdf of X and S is [12]

$$p(X, S|\underline{\theta}) = \pi_{s(1)}\rho_{s(1)}(x(1)) \prod_{n=2}^{N} a_{s(n)s(n-1)}\rho_{s(n)}(x(n)). \quad (5)$$

Due to the hidden (missing) states S, the expectation-maximization (EM) algorithm is used to learn the model parameters $\underline{\theta}$ from several independent training samples X_k

$$\hat{\underline{\theta}} = \arg\max_{\underline{\theta}} \prod_{k} p(X_k|\underline{\theta}). \quad (6)$$

For HMM, the EM algorithm is better known as the Baum-Welch algorithm consisting of a forward and backward recursion. During inference, the underlying state sequence S is estimated from a given observation sequence X by using the learned HMM model $\hat{\underline{\theta}}$:

$$\hat{S} = \arg\max_{S} p(X, S|\hat{\underline{\theta}}). \quad (7)$$

B. HMM states vs. load states

Note that the states and state transitions of an HMM model are not necessarily identical to the states and state transitions of a load. The formers are mathematical states and the latters are physical states. Their relationship depends on the type of the load, its transient phase of the power signal and the sampling frequency of the power measurement. In the simplest case, if an M-state load has M perfectly constant power consumption levels with sharp transitions, then the HMM model has exactly M states corresponding to the M power consumption levels. If, however, the load has non-zero transient phase during a load state transition, the power signal $x(n)$ can take more than M different values due to the transient phase of the power signal. The higher the sampling frequency is, the stronger this effect will be. In general, the number of HMM states is always lower bounded by the number of load states.

On the other hand, if one reduces the sampling frequency, or equivalently, if one collects one short block of length $B > 1$ of power measurements $\underline{x}(n) = [x(n), \ldots, x(n-B+1)]^T$ as the observation vector of an HMM state, then such a data block can contain constant power consumptions or different power transient phases. In this case, an HMM state can correspond to a load state with constant power consumption or load state transition. For a variable load with a continuously time-varying power consumption, it is even more hard or impossible to identify load states and HMM states. Hence, the model selection, i.e the choice of a suitable number of HMM states for a given load and a given sampling frequency, plays an important role.

III. Model selection

A. AIC criterion

We use the Akaike information criterion (AIC) to evaluate how well an HMM fits to a specific load and to select the number of HMM states M. Given the training sequence X, the aim is to first estimate the best M-state model parameters $\hat{\underline{\theta}}_M$ from X and then calculate the AIC criterion to be minimized

$$AIC(M) = -2L_M + 2N_M, \;\; L_M = \ln p(X|\hat{\underline{\theta}}_M) \quad (8)$$

L_M is the log-likelihood of $\hat{\underline{\theta}}_M$ for the given observation X. The second penalty term with N_M being the number of estimates parameters is used to avoid order overestimation. The value of M which minimized $AIC(M)$ is the AIC estimate for the number of HMM states.

B. Controlled and uncontrolled multi-state loads

It is easy to imagine that an HMM is a poor model for a variable load like a computer because such a load does not have a finite number of power consumption levels. In this case, a model selection will be difficult because $AIC(M)$ does not show a clear minimum. On the other side, an HMM is even not always a good model for multi-state loads. We distinguish between two types of multi-state loads, the controlled and the uncontrolled ones. The controlled multi-state loads are controlled by a circuit and run different fixed programs. This implies that some state transitions occur more frequently than others. As a result, some elements of the transition matrix $\mathbf{A}$ are large and the remaining ones are close to zero. One example is a dishwasher that typically runs the program heating "h", washing "w" and pumping "p".

Fig. 1a shows the load state transitions of such a dishwasher. This diagram reflects actually the washing program. Fig. 1b shows the corresponding HMM state transitions. Since HMM models the state sequence of a time series, most of the time the HMM state will remain unchanged. This is the reason for the additional "keep in state" self-transitions in Fig. 1b in comparison to Fig. 1a.

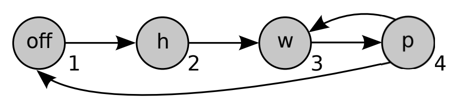

Fig. 1a. Load state transitions

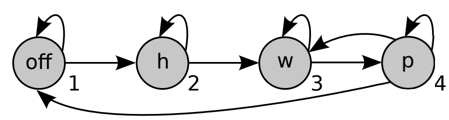

Fig. 1b. HMM state transitions

A controlled multi-state load has a structured transition matrix $\mathbf{A}$ because some elements of $\mathbf{A}$ are always zero. For the dishwasher in Fig. 1b,

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & a_{14} \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{pmatrix}, \;\; a_{ij} \neq 0, \quad (9)$$

when we enumerate the four load states "off", "h", "w", "p" as $s(n) = 1, 2, 3, 4$. If this kind of a priori information (e.g. washing programs, microwave programs) is available, it can be used to constrain some elements of $\mathbf{A}$ to zero. This will reduce the number of parameters to be learned and improve the HMM learning and the inference.

In comparison, uncontrolled multi-state loads do not run fixed programs and the state transitions underly random user behavior. In this case, all or most of the elements of the transition matrix $\mathbf{A}$ are non-zero but small. An example is a lighting circuit with a number of independent on/off lights. The overall lighting circuit is a multi-state load but with random state transitions and random duration at a particular power consumption state. This in turn favors different HMM states with different HMM state transition probabilities with the same observation $x(n)$. In other words, $AIC(M)$ will mostly decrease for increasing M and make the model selection difficult for the uncontrolled multi-state loads.

IV. Model adaptation

Model adaptation is necessary if a learned HMM for a particular load shall be applied across different houses. Assume an HMM with parameter $\hat{\underline{\theta}}_1$ has been learned from the observation sequence X_1 from the distribution $q_1(X)$ in house one and shall be applied to infer the state sequence from the observation sequence X_2 with the distribution $q_2(X)$ of house two.

We assume that both loads in house one and two have the same set of states but differ in $q_1(X)$ and $q_2(X)$ because some states show different power consumptions $\mu_i^1 \neq \mu_i^2$ (e.g two fridges with different compressors) and/or different length of the physical states $N_i^1 \neq N_i^2$ (e.g. good/bad insulated fridge with short/long cooling periods) that effects state transition probabilities of the HMM. Therefore, the model of house one will not be accurate enough for a similar load in house two.

Fig. 2 shows an example of a load with three states from two different houses. We assume loads with constant power consumption in each physical state. We further assume that the signal is periodic. So, we only have to consider one period of the signal to determine the HMM parameters. Each physical state i has a fixed length of N_i^k samples and an amplitude μ_i^k, where $k = 1, 2$ is the index of the house and $i = 1, 2, 3$ are the states. With these assumptions the transition matrix of the HMM has a special form

$$a_{ii}^k = P^k(s(n) = i | s(n-1) = i) = \frac{N_i^k - 1}{N_i^k}, \; 1 \leq i \leq M \quad (10)$$

$$\begin{aligned} a_{ij}^k &= P^k(s(n) = i | s(n-1) = j) \quad (11) \\ &= P^k(s(n) = i | l_j(n)) \underbrace{P^k(s(n) \neq j | s(n-1) = j)}_{1-a_{jj}^k}, \; i \neq j, \end{aligned}$$

where the short notation $l_j(n) = (s(n) \neq j) \cap (s(n-1) = j)$ means that the load leaves state j at time n. $P^k(s(n) = i | l_j(n)) = \frac{|\{n: (s(n)=i) \cap (s(n-1)=j)\}|}{|\{n: (s(n) \neq j) \cap (s(n-1)=j)\}|}$ denotes the probability of entering state i ($\neq j$) conditioned on leaving state j at time n. We propose

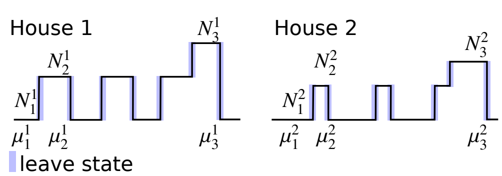

Fig. 2. Two state sequences of a load from different houses

a supervised adaptation method using a short piece $\bar{X}_2$ of sequence X_2 from $g_2(X)$. The basic motivation of model adaptation instead of learning a new HMM model $\hat{\underline{\theta}}$ from X_2 is the dramatic reduction of training data and training time. To adapt the HMM, the parameter vector $\hat{\underline{\theta}}_1$ undergoes a parametric transform

$$\hat{\underline{\theta}}_2 = f(\hat{\underline{\theta}}_1, \underline{\alpha}), \quad (12)$$

resulting in a new parameter vector $\hat{\underline{\theta}}_2$to fit the new distribution $q_2(X)$. $f(\cdot)$ is parametrized by $\underline{\alpha} \in \mathbb{R}^{N_a}$, where N_a is the number of adaptation parameters to be estimated. N_a should be chosen small to allow model adaptation with a short adaptation training sequence $\bar{X}_2$.

Adaptation to different power consumption levels is easily done by

$$\mu_i^2 = \alpha_i \mu_i^1, \;\; 1 \leq i \leq M, \quad (13)$$

where μ_i^k is the mean power consumption of state i for house $k = 1, 2$. Note that Eq.13 models a general state-specific power adaptation. Hence, M scaling parameters $\alpha_1, \ldots, \alpha_M$ are required. In the special case of uniform state-independent power adaptation, only one scaling parameter $\alpha_i = \alpha$ is needed. In contrast, the variances σ_i^2 of $x(n)$ at state i are caused by noise and are assumed to be constant across different houses. Therefore, σ_i^2 is kept constant.

Under the assumption of periodicity, changing the length of the physical states $N_i^1 \rightarrow N_i^2 = c_i N_i^1$ transforms $\mathbf{A}^1 \rightarrow \mathbf{A}^2$ in a special way. Due to the changed length of state i, a_{ii} changes according to

$$a_{ii}^2 = \frac{N_i^2 - 1}{N_i^2} = \underbrace{\frac{c_i N_i^1 - 1}{c_i(N_i^1 - 1)}}_{\lambda_i} a_{ii}^1, \;\; 1 \leq i \leq M. \quad (14)$$

Hence the diagonal elements of $\mathbf{A}$ are scaled by λ_i. Since a scaling of the state length does not affect $P^k(s(n) = i | l_j(n))$,

$$\begin{aligned} a_{ij}^1 &= P^2(s(n) = i | l_j(n))(1 - a_{jj}^2) \quad (15) \\ &= P^1(s(n) = i | l_j(n))(1 - a_{jj}^2) \frac{1 - a_{jj}^2}{1 - a_{jj}^1} \quad (16) \\ &= \frac{1 - \lambda_i a_{11}^1}{1 - a_{ii}^1}, \;\; i \neq j, \quad (17) \end{aligned}$$

i.e. also the adaptation of the off-diagonal elements of $\mathbf{A}$ are parametrized by λ_i, $1 \leq i \leq M$. So the adaptation parameter vector $\underline{\alpha} = [\alpha_1, \ldots \alpha_M \lambda_1, \ldots, \lambda_M]^T$ contains $N_\alpha = 2M$ parameters to be adjusted. The initial state probabilities are assumed to remain the same $\underline{\pi}_2 = \underline{\pi}_1$ and are not adapted.

In order to find the optimal $\underline{\alpha}$, $p(\bar{X}_2 | f(\hat{\underline{\theta}}_1, \underline{\alpha}))$ is computed using the forward algorithm and numerically maximized over $\underline{\alpha}$ using the conjugate gradient (CG) algorithm. This leads to the optimum adaption parameters $\hat{\underline{\alpha}}$. Performance of the adaption is evaluated by comparing the log-likelihoods $\log p(X_2 | \hat{\underline{\theta}}_1)$ and $\log p(X_2 | f(\hat{\underline{\theta}}_1, \hat{\underline{\alpha}}))$ before and after adaptation.

V. Experiments

A. The dataset

We used the UK-DALE dataset [11] to evaluate the HMM modeling of load signals. HMM models of eight different loads are learned from 16h data of real power measurements. The loads can are divided into three categories: controlled multi-state loads, uncontrolled multi-state loads and variable loads, see Table I.

controlled multi-state loads	uncontrolled multi-state loads	variable loads
washing machine dishwasher microwave toaster	fridge lighting ciruit	PC amplifier

TABLE I
LOADS SELECTED FOR EXPERIMENTS

B. Results

Fig. 3 shows the AIC for all loads. In case of variable loads and uncontrolled multi-state loads, AIC(M) keeps decreasing for growing M. Therefore, HMM seems to be not a good model for those loads because it is difficult to find a number of states. On the other hand, HMM is suited for controlled multi-state loads because they can be modelled with only few states.

First we validate the model adaptation with simulated signals X_1 and X_2 of the same form in Fig. 2. The parameters are $\mu_i^1 = 0, 1, 2$ and $N_i^1 = 5, 10, 20$ for $i = 1, 2, 3$ in case of X_1 and $\mu_i^2 = 0, 1.2, 2.3$ and $N_i^2 = 8, 20, 22$ in case of X_2. We trained an HMM on ten periods of X_1 and used one period of X_2 for adaptation. Tab. II compares the true parameters μ_i^2 and a_{ii}^2 to the estimated parameters $\hat{\mu}_i^2$ and $\hat{a}_{ii}^2$ after adaptation. We see that the means μ_i^2 and the diagonal elements a_{ii}^2 of $\mathbf{A}^2$ could be adapted correctly.

	μ_1^2	μ_2^2	μ_3^2	a_{11}^2	a_{22}^2	a_{33}^2
true	0	1.2	2.3	0.875	0.95	0.955
estimated	0	1.19	2.3	0.875	0.94	0.97
	μ_1^1	μ_2^1	μ_3^1	a_{11}^1	a_{22}^1	a_{33}^1
orig.	0	1	2	0.8	0.9	0.95

TABLE II
ADAPTED STATE MEANS AND TRANSITION PROBABILITIES FOR SIMULATED SIGNALS

For three real loads the results of model adaptation are given in Table III. For each load we use an HMM with 5 states. For adaptation we used 1.6h of training data from house 2. The HMM models learned from house 1 can not be shared by house 2 without model adaptation. This can be seen from Table III which compares the log-likelihoods of 3 cases: house 1, data from house 2 by using the HMM model from house 1, data from house 2 by using an adapted HMM model to house 2. A large log-likelihood value indicates a good model fitting and a small value a poor HMM model.

load	$\ln p(X_1\|\underline{\hat{\theta}}_1)$	$\ln p(X_2\|\underline{\hat{\theta}}_1)$	$\ln p(X_2\|f(\underline{\hat{\theta}}_1, \underline{\alpha}))$
fridge	3311	1430	3005
dishwasher	2881	-1614	2453
washing machine	7100	2369	3375

TABLE III
LOG-LIKELIHOOD OF HMM MODELS WITHOUT AND WITH MODEL ADAPTATION

We observe that adaptation is best for the fridge and worst for the washing machine. Comparing the signals of the three loads we noticed that the signals of the fridge and dishwasher from house 1 and 2 differ in amplitude and duration, but seem to have the same states. Because they are periodic, adaptation can be performed easily. In case of the washing machine, the signals look different and we are even not sure whether the loads share the same states or not. Therefore, it is reasonable that the performance of adaptation is worst in this case.

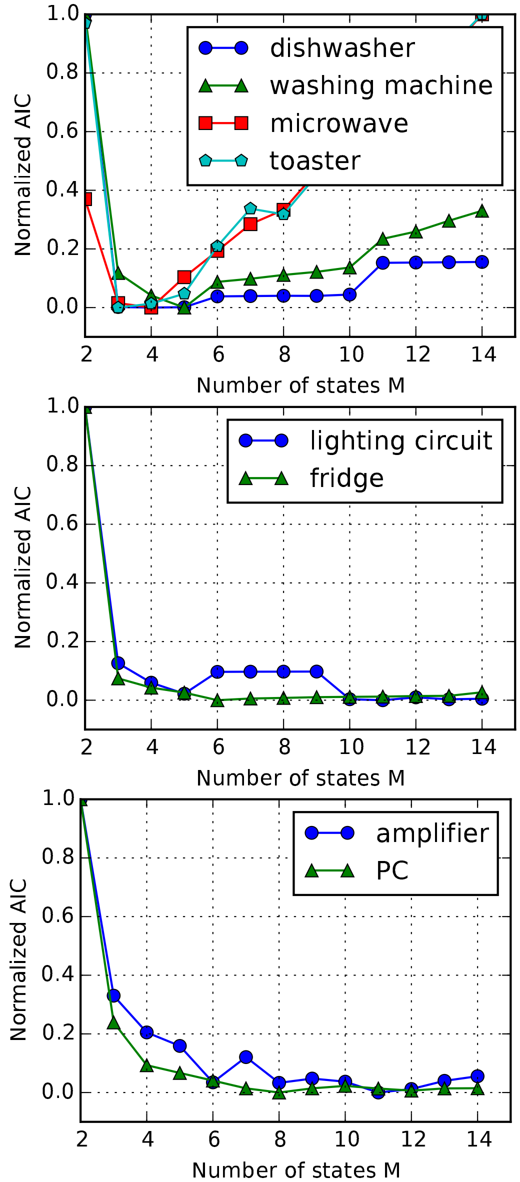

Fig. 3. AIC for loads of different categories

VI. CONCLUSIONS

We investigated how power consumption signals can be efficiently modeled by HMM. We showed that HMM is a suitable model for controlled multi-state loads, but degrades for uncontrolled multi-state and variable loads. Finally, we presented a method for model adaptation between different houses which uses only very little data.

References

[1] Yuanwei Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 4340–4343.

[2] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consumer Electronics*, vol. 57, pp. 76–84, 2011.

[3] J. Froehlich, E. Larson, et al., "Disaggregated end-use energy sensing for the smart grid," *IEEE Trans. Pervasive Computing*, vol. 10, pp. 28–39, 2011.

[4] A. Zoha, A. Gluhak, et al., "Nonintrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, pp. 16838–16866, 2012.

[5] M.J. Reyes-Gomez, B. Raj, and D.R.W. Ellis, "Multi-channel source separation by factorial hmms," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, April 2003, vol. 1, pp. I–664–I–667 vol.1.

[6] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in NIALM systems," in *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, 2004, pp. 3462–3468.

[7] K. Suzuki, S. Inagaki, et al., "Nonintrusive appliance load monitoring based on integer programming," in *Proc. SICE Annual Conf.*, 2008, pp. 2742–2747.

[8] T. Zia, D. Bruckner, and A.Zaidi, "A hidden Markov model based procedure for identifying household electric loads," in *Proc. IECON*, 2011, pp. 3218–3223.

[9] B. Yang L. Mauch, "A novel dnn-hmm-based approach for extracting single loads from aggregate power signals," in *In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

[10] M. Weal O. Parson, S. Ghosh and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proceedings of the 26th Conference on Artificial Intelligence*, 2012.

[11] Jack Kelly and William Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 150007, 2015.

[12] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," *IEEE ASSp Magazine*, 1986.