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 load disaggregation, there is no method to model loads that 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 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.

In this paper, we assume a first-order Markov chain which is characterized by the initial state probabilities

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

and the time-invariant state transition probabilities

$$a_{ij} = P(s(n) = \hat{i} | s(n-1) = j), \ 1 \leq i, j \leq M$$

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

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

$$p(X, S|\theta) = \pi_1 \prod_{n=2}^{N} a_{s(n-1)s(n)} p_i(x(n)).$$

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

$$\hat{\theta} = \arg \max \sum_k p(X_k|\theta).$$
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 \( \hat{\theta} \):

\[
\hat{S} = \arg \max_S p(X, S | \hat{\theta}).
\]  

### 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. The formers are mathematical states and the latters are not necessarily identical to the states and state transitions of a load. The formers are mathematical states and the latters are not necessarily identical to the states and state transitions of a load.

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 \( \hat{\theta} \):

\[
\hat{S} = \arg \max_S p(X, S | \hat{\theta}).
\]

### 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 \( \hat{\theta}_M \) from \( X \) and then calculate the AIC criterion to be minimized

\[
AIC(M) = -2L_M + 2N_M, \quad L_M = \ln p(X | \hat{\theta}_M)
\]

where \( L_M \) is the log-likelihood of \( \hat{\theta}_M \) for the given observation \( X \). The second penalty term with \( N_M \) being the number of observation sequences is used to avoid 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 \( A \) 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](image)

![Fig. 1b. HMM state transitions](image)

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

\[
A = \begin{pmatrix}
  a_{11} & 0 & 0 & a_{14} \\
  a_{21} & a_{22} & 0 & 0 \\
  0 & a_{32} & a_{33} & a_{34} \\
  0 & 0 & a_{43} & a_{44}
\end{pmatrix}, \quad a_{ij} \neq 0,
\]

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 \( A \) 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 \( A \) 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 $\hat{\theta}_1$ 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 $\mu_i \neq \mu_j$ (e.g. two fridges with different compressors) and/or different length of the physical states $N_1 \neq N_2$ (e.g. good/bad insulated fridge with short/long cooling periods) that affects 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 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 $N^i$ samples and an amplitude $\mu^i$, 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

$$a^k_{jj} = P^k(s(n) = j | s(n-1) = j) = \frac{N^k - 1}{N^k}, 1 \leq i \leq M \quad (10)$$

$$a^k_{ij} = P^k(s(n) = j | s(n-1) = i) = P^k(s(n) = j | s(n-1) = i) \cdot P^k(s(n) \neq j | s(n-1) = j), i \neq j, \quad (11)$$

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 $n$. $P^k(s(n) = j | l_j(n)) = \frac{N^k - 1}{N^k}$ denotes the probability of entering state $i \neq j)$ conditioned on leaving state $j$ at time $n$. We propose

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

resulting in a new parameter vector $\hat{\theta}_2$ to fit the new distribution $q_2(X)$. $f(\cdot)$ is parametrized by $\alpha \in \mathbb{R}^{N_a}$, 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 $\hat{X}_2$.

Adaptation to different power consumption levels is easily done by

$$\mu^i_2 = \alpha \mu^i_1, \quad 1 \leq i \leq M, \quad (13)$$

where $\mu^i_2$ 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 $\alpha = \alpha$ is needed. In contrast, the variances $\sigma^2_\alpha$ of $X(n)$ at state $i$ are caused by noise and are assumed to be constant across different houses. Therefore, $\sigma^2_\alpha$ is kept constant.

Under the assumption of periodicity, changing the length of the physical states $N^i_1 \to N^i_2 = c_i N^i_1$ transforms $A^1 \to A^2$ in a special way. Due to the changed length of state $i$, $a^i_{ij}$ changes according to

$$a^2_{ij} = \frac{N^2_i - 1}{N^2_i} c_i (N^1_i - 1) a^1_{ij}, \quad 1 \leq i \leq M. \quad (14)$$

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

$$a^1_{ij} = P^k(s(n) = i | l_j(n))(1 - a^2_{jj}) \quad (15)$$

$$= P^k(s(n) = i | l_j(n))(1 - a^2_{jj}) \frac{1 - a^2_{jj}}{1 - a^1_{jj}} \quad (16)$$

$$= \frac{1 - \lambda_i a^1_{1i}}{1 - a^1_{ii}}, \quad i \neq j. \quad (17)$$

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

In order to find the optimal $\alpha$, $\hat{\alpha} = \arg \max p(\hat{X}_2 | f(\hat{\theta}_1, \alpha))$ is computed using the forward algorithm and numerically maximized over $\alpha$ using the conjugate gradient (CG) algorithm. This leads to the optimum adaption parameters $\hat{\alpha}$. Performance of the adaptation is evaluated by comparing the log-likelihoods $\log p(X_2 | f(\hat{\theta}_1, \alpha))$ and $\log p(X_2 | f(\hat{\theta}_2, \hat{\alpha}))$ 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.
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 X₁ and X₂ of the same form in Fig. 2. The parameters are \( \mu_{1i} = 0, 1, 2 \) and \( N_{1i} = 5, 10, 20 \) for \( i = 1, 2, 3 \) in case of X₁ and \( \mu_{2i} = 0, 1, 2, 3 \) and \( N_{2i} = 8, 20, 22 \) in case of X₂. We trained an HMM on ten periods of X₁ and used one period of X₂ for adaptation. Tab. II compares the true parameters to the estimated parameters after adaptation. We see that the means \( \mu_{2i} \) and the diagonal elements \( a_{ii} \) of \( A^2 \) could be adapted correctly.

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.

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.

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


