Application of Factorial Hidden Markov Model (FHMM) to Non-intrusive Load Disaggregation

Speaker: Wenting Li

Instructor: Qiang Ji

Rensselaer Polytech Institute

Dec. 4, 2018

Background of Non-intrusive Load Disaggregation

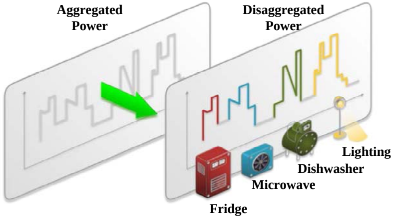


Figure: Non-Intrusive Loads Monitoring: interpret type of each individual loads from the aggregated loads [https://bitbucket.org/aicip/non-intrusive-load-monitoring]

- "Non-intrusive" means **only the aggregated power** consumption of multiple appliances in a household is allowed to measure;
- Consumers concern the **states of each individual loads**, like fridges, lighting;
- These states monitoring are beneficial to detect abnormal conditions and optimize the energy deployment.

Motivations and Contributions

- Motivations
  - If the continuous power observations $X_t^{(m)}$ of the mth individual loads can be modeled by **Gaussian distributions**,
  - the "on and off" states $Z_t^{(m)}$ ("on and off") can be equivalent to be **the hidden state**, and these states are temporal related,
  - Then each **individual load** can be modeled as a HMM with Gaussian observations.
  - Thus the **aggregated power** $\bar{X}_t$, including multiple individual loads, can be modeled by the factorial hidden Markov Model (FHMM), to decode the states of individual loads.

- Main Contributions
  - **Train HMMs with Gaussian observations** to represent each individual loads through the Expectation-Maximization (EM) algorithm;
  - **Construct the FHMM** to characterize the aggregated loads;
  - The power and the states of each individual loads can be **estimated by FHMM decoding** using the aggregated loads.
  - The **parameter sensitivity** of the HMM model is also studied.

Problem Formulation

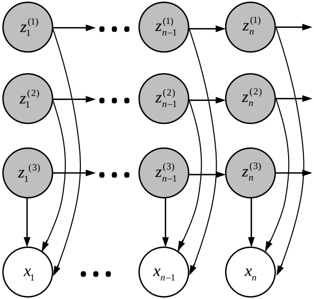


Figure: FHMM

- Given the **aggregated loads** $\bar{X}_t$ of multiple individual,

$$\bar{X}_t = \Sigma_{m=1}^{M} X_t^{(m)} \quad (1)$$

  where the aggregated observation $\bar{X}_t$ is determined by the combination of M possible states $Z_t^{(m)}, m = 1, \cdots, M$ of M individual appliances.
- **Goal:** Predict the states ("on" and "off") $Z_t^{(m)}$ of each individual loads based on FHMM.

Model individual loads with HMM

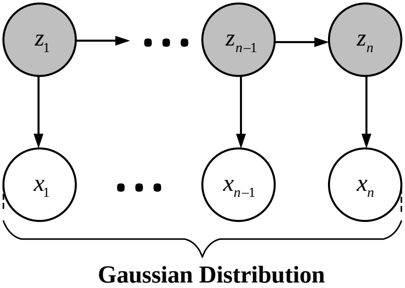


Figure: HMM with Gaussian Observations

- $z_n \in \{1, \cdots, K\}$ denotes the K possilbe switch states at the nth time of the appliance;
- x_n is continuous observations, representing the measured power at the nth time, $x_n \sim \mathcal{N}(\mu, \Sigma)$.
- The parameters for this HMM are $\theta = \{A \in R^{K \times K}, \pi \in R^K, \phi\}, \phi = \{\Sigma \in R, \mu \in R^n\}$,
  - A is the transition matrix
  - π is the prior distribution
  - ϕ include the mean μ and covariance Σ of the Gaussian distribution.

where

Train HMM with EM algorithm (1)

- Following the EM algorithm, the lower bound $Q(\theta, \theta^{old})$ can be maximized

$$Q(\theta,^{old}) = \Sigma_Z p(Z|X,\theta^{old}) \log(p(X,Z|\theta)) \tag{2}$$

$$= \underbrace{\Sigma_{k=1}^{K}\gamma(z_{1k}) \log \pi_k}_{(1)} + \underbrace{\Sigma_{n=2}^{N}\Sigma_{j,k=1}^{K}\xi(z_{n-1,j}, z_{n,k}) \log A_{jk}}_{(2)} \tag{3}$$

$$+ \underbrace{\Sigma_{n=1}^{N}\Sigma_{k=1}^{K}\gamma(z_{nk}) \log(p(x_n|z_n, \phi))}_{(3)} \tag{4}$$

  - where $\gamma(z_{nk}) = \alpha(z_n)\beta(z_n)$
  - $\xi(z_{n-1,j}, z_{n,k}) = \alpha(z_{n-1})p(x_n|z_n)p(z_n|z_{n-1})\beta(z_n)$
  - $\alpha(z_n)\beta(z_n)$ can be **computed recursively**.
- The parameters $\theta = \{A, \pi, \phi\}, \phi = \{\Sigma, \mu\}$ can be obtained by maximizing the (1)$\sim$(3) with probability constraints respectively.

Train HMM with EM algorithm (2)

Algorithm 1 EM algorithm of Gaussian HMM for parameter learning

1: **Given observations x** $\in R^N$
2: **Initialization:** $\theta_0 = \{\pi, A, \mu, \Sigma\}$
3: **while** not converge **do**
4: E-step: Compute $\gamma(z_{nk}), \xi(z_{n-1,j}, z_{n,k})$ by (12) and (22).
5: M-step: Update parameters

$$\pi_k = \frac{\gamma_{1k}}{\sum_{j=1}^{K} \gamma_{1j}} \quad A_{jk} = \frac{\sum_{n=1}^{K} \xi(z_{n-1,j}, z_n, k)}{\sum_{l=1}^{K} \sum_{n=1}^{K} \xi(z_{n-1,j}, z_n, l)}$$

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma_{nk} x_n}{\sum_{n=1}^{N} \gamma_{nk}} \quad \Sigma_k = \frac{\sum_{n=1}^{N} \gamma_{nk}(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma_{nk}}$$

where $k = 1, \cdots, K$.

6: Output: $\theta = \{\pi, A, \mu, \Sigma\}$.

FHMM of the Aggregated Loads

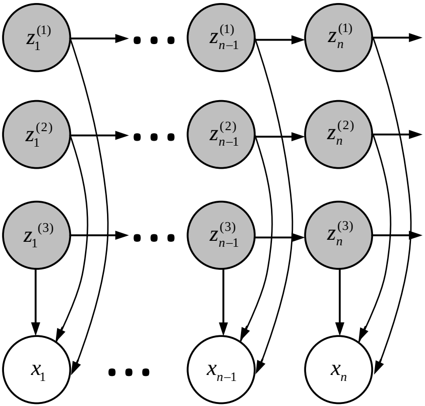


Figure: Factorial Hidden Markov Chain with Gaussian observations

- FHMM utilizes M **HMM chains** to model the observations $x_i, i = 1, \cdots, n$;
- The parameters size of FHMM $\theta^{agg} = \{\pi^{agg} \in R^{K^N}, , A^{agg} \in R^{K^N \times K^N}, \mu^{agg} \in R^{K^N}, \Sigma^{agg} \in R^K\}$ **becomes large if K increases**;
- The parameters of FHMM are defined as the **combination** of the parameters of HMMs $\theta^i = \{A^i, \pi^i, \phi^i\}, i = 1, \cdots, M$.

$$\pi^{agg} = \pi^1 \otimes \pi^2 \otimes, \cdots, \otimes \pi^M \tag{5}$$

$$A^{agg} = A^1 \otimes A^2 \otimes, \cdots, \otimes A^M \tag{6}$$

$$\mu^{agg} = \mu^1 \times \mu^2 \times, \cdots, \times \mu^M \tag{7}$$

where $\otimes$ denotes the Kronecker product, $\times$ denotes the Descartes product.

Decoding the Optimal States with Viterbi Algorithm

- Given the aggregated observations $\mathbf{x}$ and the parameters $\theta^{agg} = \{\pi^{agg}, , A^{agg}, \mu^{agg}, \Sigma^{agg}\}$, the decoding problem is to acquire the best sequence of the hidden variables $\mathbf{z}$ by maximizing the conditional probability $p(Z|X)$

$$\begin{aligned} z_1^*, \cdots, z_N^* &= \arg\max_{z_1, \cdots, z_N} p(z_1, \cdots, z_N | x_1, \cdots, x_N) && (8) \\ &= \arg\max_{z_1, \cdots, z_N} p(z_1, \cdots, z_N, x_1, \cdots, x_N) && (9) \\ &= \arg\max_{z_1, \cdots, z_N} p(z_1) \Pi_{t=2}^N p(z_t | z_{t-1}) \Pi_{t=1}^N p(x_t | z_t) && (10) \end{aligned}$$

- The **best sequence** can be recursively obtained by Viterbi algorithm.
- These $z_1^*, \cdots, z_N^*$ indicate the **states of individual loads**.

Introduction of Datasets

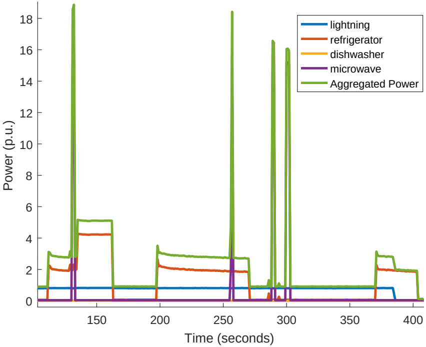
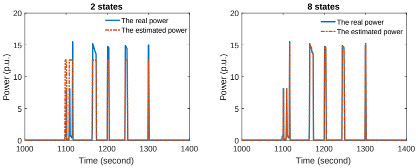


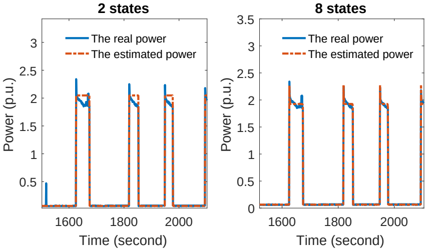
Figure: The typical individual and aggregated power consumption

- The **public dataset** of Reference Engergy Disaggregation Datasets (REDD) [Kolter et al., 2011] validate the method.
- There are six households with 10∼20 various appliances.
- The power consumption of each **individual loads and the aggregated** measurements are also given;
- The main **four type** of individual loads are applied: lighting, refrigerator, dishwasher, and microwave.

HMM with Different Number of Hidden States



(a) The power of microwave



(b) The power of refridgerator

Figure: Modeling individual loads by HMM with different number of states

- The states of **"on and off"** of **individual loads** can be captured by the HMM model;
- For each hidden node $z_n \in \{1, \cdots, K\}$, **the number of states** K will influence the accuracy of modeling the power consumption;
- The figures demonstrate that **a larger** K can reduce the approximation errors, but **more parameters** need to learn;
- As we mainly concern with the "on and off" states, the **HMM with** K**=2** is applied.

State Estimation of the Aggregated Loads with FHMM

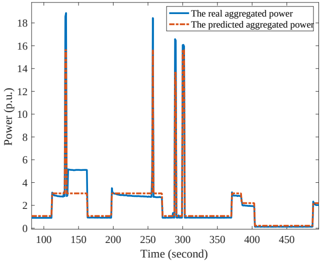
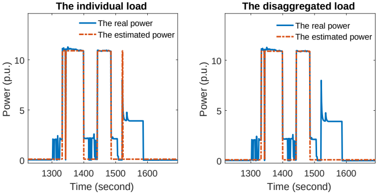


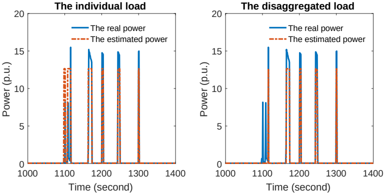
Figure: The real and estimated power of lighting appliance

- The FHMM of **four chains with 16 states** is constructed;
- The various **states can be correctly captured** by the FHMM;
- There are **some difference** between the power consumption estimated by the FHMM and the real one, as only 2 states are used for each loads;
- **A larger number** K of hidden states can reduce the difference.

Disaggregating Different Loads with FHMM (1)



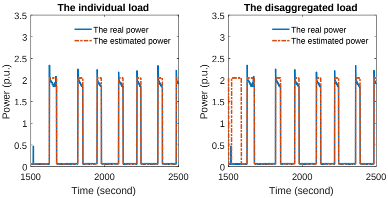
(a) Comparison of the individual and the disaggregated power of refrigerator appliance



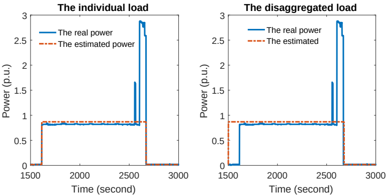
(b) Comparison of the individual and the disaggregated power of lighting appliance

- **"The individual load"** is calculated by the **HMM** from the individual observations;
- **"The disaggregated load"** is decomposed from the aggregated loads by **FHMM**;
- The figures show that the FHMM is able to **correctly estimate the switch on and off** states of individual appliances.

Disaggregating Different Loads with FHMM (2)



(a) Comparison of the individual and the disaggregated power of refrigerator appliance



(b) Comparison of the individual and the disaggregated power of lighting appliance

- Disaggregating different types of individual loads is challenging, if the switch on and off states of various loads are **overlapped**;
- For example, Both the lighting and refridgerator switch their states **during 1500 to 1600 seconds**,
- thus the FHMM **mistakes the states** of these two appliances during this period of time, but this issue can be mitigated if **a higher sampling rate** and more states can be applied.

Percentages of disaggregating loads with FHMM

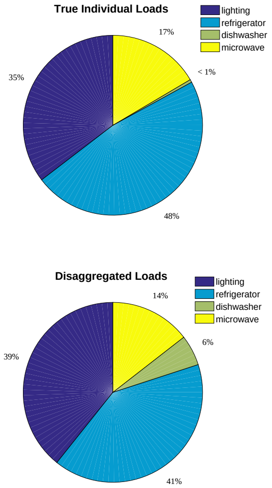


Figure: Comparison of the percentages of true individual loads and the disaggregated loads

- The total power is **decomposed into four types** with the FHMM, and the percentages of each type are shown in the pie diagram;
- For **lighting and microwave**, the difference between the true percentages and the predicted ones is less than 5%;
- For dishwasher and refrigerator, the difference is larger, since these appliances have more **transient states** which are generally ignored by the 2-state HMM.

Conclusions and Discussion

- The on and off states of individual loads can be estimated by a **HMM with Gaussian** observations;
- The **FHMM** constructed with the parameters learned from individual loads is able to estimate the states of the aggregated loads;
- **A larger** K, the number of hidden states, can increase the estimation accuracy but costs more parameters;
- The states of individual loads can be predicted by the **FHMM using the aggregated loads**;
- More number of hidden states, higher sampling rate and more training datasets can **improve the performance** of the FHMM.

References

Kolter, J Zico and Johnson, Matthew J (2011)
REDD: A public data set for energy disaggregation research
Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA,59–62.

Ghahramani, Zoubin and Jordan, Michael I (1996)
Factorial hidden Markov models
Advances in Neural Information Processing Systems, 472–478.

Hart, George William (1992)
Nonintrusive appliance load monitoring
Proc. IEEE, 1870–1891.

Koller, Daphne and Friedman, Nir and Bach, Francis (2009)
Probabilistic graphical models: principles and techniques
MIT press, 612 – 621.

Thank you for your attention!