Energy Disaggregation via Hierarchical Factorial HMM

Yeqing Li, Zhongxing Peng, Junzhou Huang
University of Texas at Arlington,
Arlington, TX 76019, USA
Email: {yeqing.li, zhongxing.peng}@mavs.uta.edu,
jzhuang@uta.edu

Zhilin Zhang, Jae Hyun Son
Samsung Research America-Dallas,
Richardson, TX 75082, USA
Email: zhilin.zhang@samsung.com
jae.son@samsung.com

***Abstract*—This paper considers the problem of energy disaggregation, which aims to decompose a whole home electric consumption into consumptions of individual appliances. Recent studies have shown that the factorial hidden Markov model (FHMM) is a favorable model for this problem. For effectiveness of inference, two key assumptions are often adapted: independence of devices and the one-at-a-time condition (it assumes at most one device changes state at each time step). In this work, we argue that these assumptions in many cases do not hold in practical data. The contradiction of data and assumptions renders disaggregation problem particularly challenging. We attempt to address this problem by introducing a novel inference framework named Hierarchical FHMM that enables effective inference of FHMM when the assumptions are violated. This framework utilizes the relationship between devices to improve the speed and accuracy of inference. Our approach also has the advantage that it can be easily integrated with existing or future inference algorithms of FHMM. Experimental results on two benchmark datasets, REDD and Pecan, demonstrated that our method yields state-of-the-art energy disaggregation results.**

I. Introduction

This paper aims to design a practical FHMM-based energy disaggregation approach. We exploit the structural information such as the correlation between devices and use the information to resolve the problem of violations of key assumptions. Using this kind of information, an inference framework named hierarchical FHMM is developed. This framework consists of two layers: in the first layer, the data is preprocessed to resolve the contradictions by clustering devices into groups; while in the second layer, the signals of the groups are disaggregated via FHMM inference; after the FHMM inference, the inference result of groups are sent back to the first layer and are post-processed to final results of each devices. We demonstrate that our hierarchical approach simplifies the problem so that better performance and better accuracy can be achieved. Furthermore, this approach provides the possibility of solving some difficult problems in disaggregation, such as separate low power level signals from high power level signals. Considering costs in practice, we focus our effort on low-sample-rate power data(generally 1 to 15 minutes).

Our paper has three contributions: **1)** We carefully study the correlations between devices, and develop an algorithm to discover and clustering the signals of correlated devices. **2)** A novel inference framework is proposed, which not only enables FHMM inference methods to work on data that are not satisfied the assumptions but also improve speed and accuracy of inference; **3)** We use two real-life datasets: REDD [1] and Pecan [2] to benchmark our method and demonstrate that its disaggregation performance is better than the existing approaches.

II. Methodology

In this section, we first introduce some key properties of data that conflict with the assumptions in previous method [3]. Then we describe our inference framework to solve this problem.

A. Device Co-relations

In the previous FHMM approaches, devices are represented by independent HMMs. However, in practice, there exists a strong correlation between different devices, which may lead to deterioration of the inference performance. Intuitively, the states of in-home appliances are related to the underneath human activity and systematic connection. For example, when the house owner does the weekly laundry, the washing machine and the dryer are on in order; or if someone is cooking, the lights in the kitchen, the range hood, the oven and other kitchen appliances may be in working condition at the same time period. This kind of correlation is based on some simple intuitions: firstly, some human activities (e.g. laundry, cooking, etc.) involve several electrical devices; secondly, these kind of activities, related to life style, have relatively stable patterns.

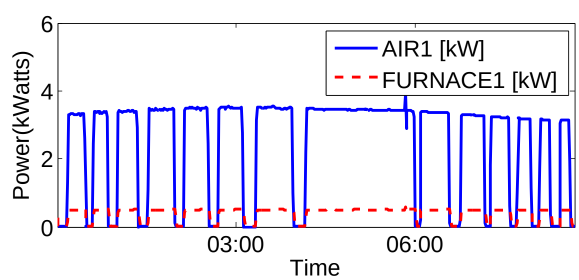


Fig. 1: Correlation between air conditioning(AIR1) and FURANCE1 in Pecan dataset Home 1.

Fig. 1 shows an example of a device correlation in the Pecan [2] dataset. The AIR1 and the FURANCE1 have shown the same rise-and-fall patterns. This situation happens nearly all the time when the two appliances are running. Additionally, since the signal of FURANCE1 is much smaller than that of AIR1, their signals are hard to separate. It will probably lead to the result that all the energy consumption is assigned to AIR1, leaving FURANCE1 deviated far from its ground truth signal.

The existence of correlation between devices contradicts previous assumptions [4] [3] on energy disaggregation problem. The first one is the independence of each appliance. This is one of the the key assumptions that the previous FHMM-based approaches rely on [3]. However, for the correlated devices, the assumption doesn't hold any more. As a result, the model is not accurate to describe the real situation and is error-prone. Another assumption that is problematic is the one-at-a-time condition. Since correlated devices will have the tendency to change state simultaneously and the differences of the output signals have played a critical role in many algorithms, misinterpreting the meaning of the differences will result in mis-assignment of the power.

B. Mining via Hierachical FHMM

At this point, we face the contradiction of the correlation of devices and the independence assumption. To resolve the contradiction and exploit the benefits of the correlation of devices, we develop a Hierarchical FHMM (HieFHMM) approach. The idea is simple. We first transform the data to eliminate their correlation so that they are suitable for the FHMM model to handle. After FHMM model gets the result, we transform the result back to get the disaggregation result of original data. One thing to note is that the transformation is conducted on the training data and will affect the learned model. The transformation does not apply directly on the testing data.

Here, the objective is to resolve the problem of device correlations. We first detect the existence of the correlation via training data and then combine each group of correlated devices into "**super devices**". By doing so, the violation of assumptions are eliminated, which makes data easier for FHMM model to handle. The super devices play a role as an encapsulation of "ugliness" of the reality and let the lower level FHMM model work in a "ideal" situation. In the testing phase, inference are made on these super devices and get a intermediate results. Finally we decompose the intermediate result of each super device into that of each individual devices.

1) Training: In the first step of the training phase, we identify the correlated devices that must be group together. The most widely used similarity measurement for time series signal is Euclidean distance [5]. However, the value of Euclidean distance is hard to interpret, which will be a problem for threshold choosing. Another problem of using Euclidean distance is that it's not a good indicator of synchronization of two signals since it only measure the overall differences of signal magnitude. These motivate us to use normalized cross-correlation (NCC) as our measurement. Let $U = [u_1, u_2, \ldots, u_T]^T$ and $V = [v_1, v_2, \ldots, v_T]^T$ be the two vectors that represent two series respectively. Let $\bar{u}$ be the mean of elements of U, $\bar{v}$ be the mean of elements of V, $\tilde{U} = [u_1 - \bar{u}, u_2 - \bar{u}, \ldots, u_T - \bar{u}]^T$ and $\tilde{V} = [v_1 - \bar{v}, v_2 - \bar{v}, \ldots, v_T - \bar{v}]^T$. Here, $\tilde{U}$ and $\tilde{V}$ represent subtracting U and V by their element means respectly so that they are zero-meaned. The NCC of U and V can be written as:

$$NCC(U, V) = \frac{1}{T} \frac{\langle \tilde{U}, \tilde{V} \rangle}{\|\tilde{U}\|_2 \|\tilde{V}\|_2}, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is inner product of vectors and $\|\cdot\|_2$ is the L^2 norm. NCC can be interpreted a the cosine of the angle between two vectors in hyperspace. Intuitively, it measures how likely each elements of U and V happen together. Its value ranges from 0 to 1, where 1 means the two signals only differ in a constant scale. Observed from our experiments, we find that threshold between 0.8 to 0.9 works well on many datasets.

Clustering. After we compute the NCC of each pair of devices, the relations among all the appliances are modelled as an undirected graph, where each node represents an appliance and each edge represents the strong correlation between nodes. Those pairs of appliances with NCC greater than a threshold (e.g. 0.9) are connected by an edge. The cluster of correlated group of appliances can be solved by computing the connected component of the graph. Fig. 2 shows a correlation graph of six appliances that have been divided into four groups according to the connected components of the graph. We call each group a "**super device**".

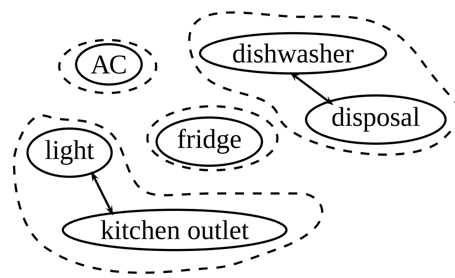


Fig. 2: Correlation graph of appliaces and cluster

Parameter Training. After devices have been grouped into super devices, we can train HMM for each super device. Since the assumption violations are eliminated, the training process is the same as Kolter et al. [3]. It's also required to align the states of each appliance in a group and the states of the group. For example, for appliances $d_1, d_2, \cdots, d_n$ in group g_i, we want to know when group g_i in state s_j, which state is d_k in (where $k = 1..n$). For simplicity, we determine the state of a_k corresponding to the state of g_i by counting scheme. With the scheme, we can choose the state s_h of a_k that occurs most frequently when g_i is in state s_j to correspond with s_j and build a state relation table. Next, we can apply the learned model on testing.

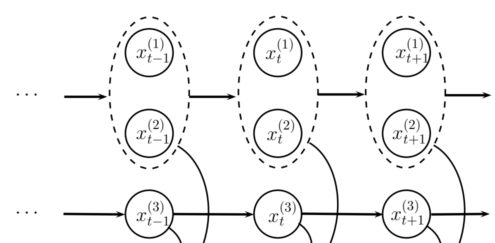


Fig. 3: FHMM with grouped devices. The dashed eclipse means a state of a group, which contains 2 appliances. The rest chain indicate a group with only one appliance.

2) Testing: **Disaggregation**. In the testing phase, AFAMAP [3] algorithm is used to perform inference on the FHMM of super devices. The inference on super devices is faster than than on origin devices because we only need to deal with fewer "devices" that satisfy the independent and one-at-a-time assumptions. Fewer "devices" lead to fewer chains and lower computational complexity and better accuracy. Note that it doesn't reduce the computational complexity by randomly grouping appliances together. The season is, although the number of chains is reduced, the state space of the result chain is the Cartesian product of state spaces of all the appliances in the group. For example, if we randomly combine n appliances where the ith appliance has m_i states, the number of states of the combined chain will be $m_s = \prod_{i=1}^{n} m_i$. In contrast, the number of states of a chain of a correlated group is much smaller than m_s because the two chain are highly correlated, and when the state of one is fixed the state of the other is usually varied between 1 or 2 states. Another benefit is the restoration of the assumptions. After combining those strong correlated appliances, we can make the assumption of the independence of the chains. The one-at-a-time condition can still be true with high probability. Fig. 3 shows an example of FHMM on super devices.

Decomposing. After the signal has been disaggregated according to the groups (super devices), one step remaining is to decompose the signal in each group into signals of each appliance. This is done by looking up the state relation table that we have learned in the training phase. For example, if appliance d_1 is in group g_2 and the state of g_2 in time t is $s_t^g = i$. If we look up the state relation table and find that the corresponding state of d_1 for g_2 at state i is j, then the state of d_1 at time t is $x_t^{(1)} = j$. For those groups with only one appliance, this process is trivial. For those groups with more than one appliance, it is done by using the state relations we have obtained in the training phase. Output of grouped signal disaggregation includes the state sequence of the group. Hence, we can find the corresponding states of each devices by looking up the table. After the states of each chain are decided, the signal of the group can be decomposed by the power level of their states. Note that this decomposing step is not the same as general disaggregation of a synthesis signal with small number of appliances. With the state relation table, the decomposing process is fast and simple. Using this kind of simple lookup table strategy is due to two reasons: first, the computation cost is very low compared to other complex method like FHMM since it uses the learned correlation to make inference; second, and more importantly, it can deal with those devices that differs significant in their power level, which is very difficult for other complex model to handle. The benefits of the proposed method will become clear in the experimental results.

To summarize, the hierarchical FHMM framework is as shown in Fig. 4. We cluster appliances into groups by their correlations. The inference is performed on the group of appliances. By clustering, we restored the independence and one-at-a-time assumptions of chains which are critical for efficient inference. The grouped result is then decomposed into individual results from by their state correlations.

III. Experimental Results

In this section, experiments are conducted on two datasets: REDD [1] and Pecan [2]. We evaluate the Hierarchical FHMM(HieFHMM) approach versus the latest AFAMAP [3] algorithm. All the experiment are performed on a desktop

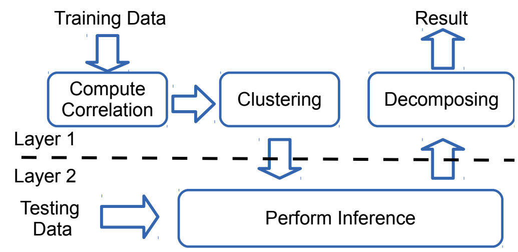


Fig. 4: Hierachical FHMM Framework

computer with an Intel 3.4Ghz i7 CPU and 12.0GB of DDR2 memory. The algorithm is coded in MatLab and we use CPLEX solver for solving quadratic programming. For all experiments, the sampling frequency of power consumption is $1/60$ Hz. To compare the experimental results, three metrics, **precision, recall and F-measure** at the circuit level, are used in our evaluation: **precision** measures what portion of an estimated circuit's energy is truly belong to that circuit, **recall** measures what portion of a given circuits energy is correctly classified, and **F-measure** is the harmonic mean of precision and recall [6]:

$$\text{F-measure} = 2 * \frac{precision * recall}{precision + recall} \tag{2}$$

F-measure can be seen as a combination of precision and recall. For each experiments, we use the first 3 days of individual circuits to train the FHMM model. Note that the amount of training data is not critical as long as we have seen most of target devices has been used. Because we only need a estimation of state transfer probability and power levels. We vary the amount of training data from 1 to 7 days, the results are essentially the same for frequent used devices. And for the case of outliers, it can be handled by the Total Variation term [3].

A. Experiments on REDD dataset

Home	AFAMAP [3]			Proposed method		
	Precision	Recall	F-measure	Precision	Recall	F-measure
1	0.8677	0.6598	0.7496	0.866	0.8439	**0.8548**
3	0.8899	0.7188	0.619	0.6652	0.9536	**0.8342**
4	0.9326	0.2686	0.417	0.8883	0.2791	**0.4248**
5	0.949	0.685	0.7956	0.9455	0.6874	**0.796**
6	0.8874	0.2512	0.3916	0.9662	0.7128	**0.8204**

TABLE I: Comparing proposed method HieFHMM against AFAMAP [3] on the REDD dataset(House-wise result)

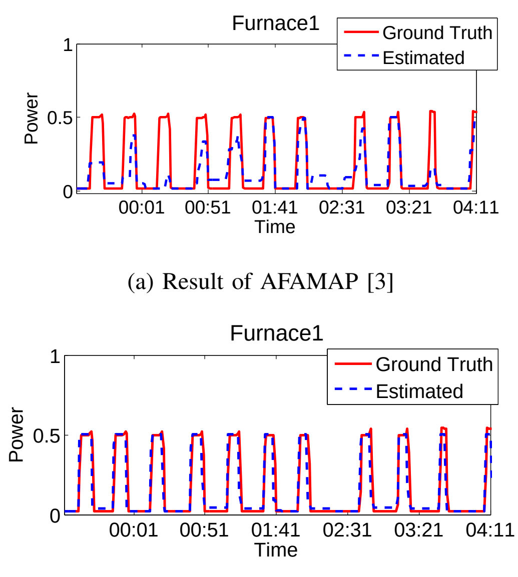
First, REDD dataset [1] is used for our experiments. REDD contains high-frequency(kHZ) current and voltage and low-frequency power measurements (3-4 second intervals) at both household-level and circuit-level from 6 US houses. The data occupy about 2 weeks. We down sample the low-frequency data to one measurement per minute to mimic the practical situation. We choose 5 houses that have the sufficient number of devices and conduct experiments on 7 to 8 frequently used devices at each house. House 2 is eliminated because it has too few appliances to be representative. Our method has discovered the several groups in all houses. House 1 has one group: **[lighting 1, lighting 2]**; House 3 has one group: **[wash dryer 1, wash dryer 2]**; House 4 has one group: **[air conditioning 1, air conditioning 2]**; House 5 has two groups: **[furance, wash dryer]** and **[kitchen outlets 1, kitchen outlets 2, outdoor outlets]**; House 6 has one group: That is the **3 air conditionings**.

The overall experimental results are shown in Table I. It shows that the precision of two algorithms in all houses are close. In the meantime, the recall and F-measure of our method HieFHMM in House 1, 3 and 6 significantly outperform AFAMAP [3] while those two measurements of our method in the House 4 and 5 also better than AFAMAP. Intuitively, the higher precision means higher percentage of estimation is correct (fewer false positive), while the higher recall means higher percentage of truth power level is estimated right (few false negative). The F-measure is a balance of both precision and recall.

B. Experiments on Pecan dataset

Home	AFAMAP [3]			Proposed method		
	Precision	Recall	F-measure	Precision	Recall	F-measure
1	0.9895	0.897	0.941	0.9555	0.9291	**0.9421**
2	0.9827	0.9106	0.9453	0.9832	0.9587	**0.9708**
3	0.9496	0.7442	0.8344	0.957	0.7661	**0.851**
4	0.9571	0.7905	0.8659	0.9841	0.9294	**0.956**

TABLE II: Comparing proposed method HieFHMM against AFAMP [3] on the Pecan dataset(Home-wise result)



(a) Result of AFAMAP [3]

(b) Result of proposed Hierachical FHMM

Fig. 5: Example 2(Small signal): Fragment of results of furance1 in home 4 of Pecan dataset.

To further evaluate our method, experiments are conducted on a commercial Pecan [2] database. The free version of Pecan dataset consists energy use data of household-level and circuit-level of 10 homes in Austin, Texas. The data occupy about 1 week and the sampled in 1 minute. Devices that occupy more than 1% energy of 4 homes are used for the experiment. The groups in these homes are simple: the air conditioner(AIR1) and the FURNACE1 are strongly correlated in all homes.

The overall results are shown in Table II. Precision of method HieFHMM and AFAMAP [3] is very close, while HieFHMM have obtained various levels of improvement in recall and F-measure. For example, in home 4, our method has achieved approximately 20% of improvement in overall recall. The improvement is due to correctly modelling of the correlation between appliances. The result is similar to result on the REDD dataset. Our approach get better results in recall and F-measure while precisions are essentially the same.

The improvement of HieFHMM to AFAMAP [3] is also impressive if look at the visualization of the signal. Fig. 5 shows both the group truth and the estimation of AFAMAP and HieFHMM in a short period of time. The signal is from furance1 in home 4. Although AFAMAP capture some similar forms of the signal, the power level of each event deviates from the ground truth. This is because the furance1 is correlated with AIR1. In contrast, HieFHMM uses correlation between those appliances to deal with the ambiguity so that the result is much more accurate.

IV. Conclusion

In this paper, we have proposed a Hierachical FHMM(HieFHMM) method for the energy disaggregation problem. This method is able to discover the correlation between devices and use the information to improve performance. Because of proper handling the correlations between devices, our method manages to prevent the failure of independence assumption of devices and preserve the one-at-a-time condition in real-world data. Extensive experiments demonstrate that the proposed HieFHMM method can capture the correlation characteristics of the energy signal and lead to better performances of energy disaggregation in terms of disaggregation accuracy and computational cost.

Acknowledgment

This work was funded by the Samsung Research America-Dallas.

References

[1] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011, pp. 1–6.

[2] C. Holcomb, "Pecan street inc.: A test-bed for nilm," *International Workshop on Non-Intrusive Load Monitoring*, pp. 271–288, 2007.

[3] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *International Conference on Artificial Intelligence and Statistics*, 2012, pp. 1472–1482.

[4] H. S. KIM, M. Marwah, M. Arlitt, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *SIAM Conference on Data Mining*, 2011.

[5] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *Proceedings of the VLDB Endowment*, vol. 1, no. 2, pp. 1542–1552, 2008.

[6] H. Shao, M. Marwah, and N. Ramakrishnan, "A temporal motif mining approach to unsupervised energy disaggregation," in *AAAI*, 2013.

[7] H. Wang, H. Dong, A. P. Boedihardjo, F. Chen, and C.-T. Lu, "A hierarchical probabilistic model for low sample rate home-use energy disaggregation," in *SIAM Data Mining*, 2013.