Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation: Applications to Residential and Commercial Buildings

Huijuan Shao[1,2], Manish Marwah[3], Naren Ramakrishnan[1,2]
[1]Discovery Analytics Center, Virginia Tech, Blacksburg, VA 24061
[2]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061
[3]Sustainable Ecosystems Research Group, HP Labs, Palo Alto, CA 94304

Abstract

Non-intrusive appliance load monitoring has emerged as an attractive approach to study energy consumption patterns without instrumenting every device in a building. The ensuing computational problem is to disaggregate total energy usage into usage by specific devices, to gain insight into consumption patterns. We exploit the temporal ordering implicit in on/off events of devices to uncover motifs (episodes) corresponding to the operation of individual devices. Extracted motifs are then subjected to a sequence of constraint checks to ensure that the resulting episodes are interpretable. Our results reveal that motif mining is adept at distinguishing devices with multiple power levels and at disentangling the combinatorial operation of devices. With suitably configured processing steps, we demonstrate the applicability of our method to both residential and commercial buildings.

Introduction

As the saying goes, sustainability begins at home. Greater than ever before, there is now a significant interest in reducing household energy footprints by providing consumers with detailed feedback on their energy consumption patterns. By contrasting such 'drill-down' data with neighborhood profiles, consumers can make better informed decisions about how their daily activities impact the environment as well as their bottom line.

A key step in this endeavor is energy disaggregation. This is the task of, non-intrusively, monitoring aggregate energy usage (electricity, water) at a home/unit and separating it out into individual appliances, subunits, and other spatial dimensions automatically, using machine learning methods. A variety of methods have been proposed, e.g., factorial HMMs (Kim et al. 2010) and sparse coding (Kolter and Jaakkola 2012) but the increasing diversity of appliances to be accommodated and the spatio-temporal coherence properties that must be modeled provides continuing opportunities for algorithm innovation.

Here we propose a temporal motif mining approach (see (Chiu, Keogh, and Lonardi 2003; Yankov et al. 2007) for background) to energy disaggregation. We specifically focus on low-frequency measurements since those can be obtained from smart meters and aim to characterize stable power consumption events, in contrast to transients. The basic idea is to discover the minimal episode which corresponds to a complete state-change cycle by a device or part of a device. Unlike state-of-the-art probabilistic methods that posit detailed temporal relationships and involve complex inference steps, we argue that our method is lightweight and, at the same time, capable of accuracy levels better than or comparable to these more complex methods. Using this approach, we conduct a thorough experimental investigation of our method on a residential dataset (REDD (Kolter and Johnson 2011) as well as a commercial dataset, demonstrating the ability of our approach to disaggregate different classes of electrical loads.

Background

Residential vs commercial buildings. There are significant differences between residential and commercial disaggregation problems. First, the number of devices is one to two orders of magnitude larger in commercial buildings. Although disaggragation of *all* devices is not feasible in commercial buildings, we can disaggregate branches of the electrical infrastructure resulting in a drastic reduction in the number of meters required to monitor loads. The electrical infrastructure in residences and commercial buildings also differs. The former have low voltage levels (e.g., 110V or 220V) and two phase circuits while the latter have three-phase, high voltage lines coming from the utility which feed a hierarchical electrical infrastructure in the building. Heavy duty equipment such as chillers, blowers, pumps, elevators, etc., use three-phase power, which is then split into two phases and stepped down for lighting and plug loads. Residences typically receive two-phase power from the utility, as shown in Fig. 1. Each phase connects to many circuits and in turn each circuit has one or more devices that draw power from it. Devices in residences usually consist of microwaves, refrigerators, ovens, lights, washers/dryers, and air conditioners. Some devices such as washers/dryers typically connect to both phases. Compared to residences, there is more automation in commercial buildings, e.g., blowers, pumps, lights and other devices are controlled by a building management system (BMS) and turn on/off at scheduled times. Most of the past research in disaggregation pertains to residential buildings.

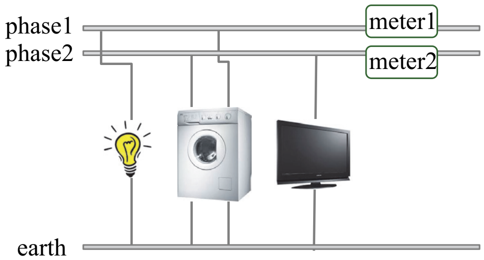


Figure 1: A residential setup for data collection.

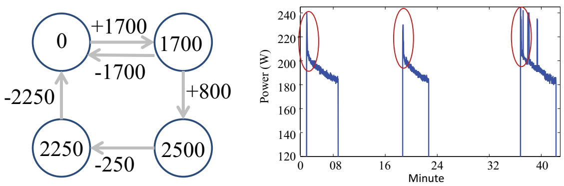


Figure 2: Steady state transitions and transient features at startup.

High frequency vs low frequency sampling. High frequency sampling, typically at the rate of hundreds to thousands of Hz, can reveal transients in the electrical signal which can then be used as features for disaggregation. However, customized HW usually needs to be installed to sample at such high rates. Low frequency sampling, typically at rates of 1Hz or below, can be obtained from smart meters, which are being deployed in increasing numbers by utilities worldwide.

Multiple states and transients. The device to power state mapping is not one-to-one. A given device might involve multiple power states as shown in Fig. 2 (left). For instance, a washer/dryer might function at a fixed power level of 1700W but later change levels based on its workload. Further, as shown in Fig. 2 (right), before the refrigerator reaches a stable state, a transient is observed and, after a period of time, the power consumption stabilizes to a certain level.

Energy disaggregation. Energy disaggregation, initially proposed by (Hart 1992), records only the power at the main entry or several points of a building, and aims to deduce the power consumption of devices in the building over a period of time through analysis of the aggregate. Fig. 3 gives an example of energy disaggregation where a total power time series is disaggregated into fourteen devices over a period of time (here, 8am to 12 noon). For instance, note that it has been deduced that the refrigerator (in purple) is switched on for three periods of time, namely, 8:50am to 9:05am, 10:15am to 10:40am, and 11:50am to 12:05pm.

Challenges. The field of disaggregation has over the last twenty years developed many practical solutions drawing primarily from the field of electrical engineering. However, many challenges remain, including lack of knowledge about the number of power levels of each device, uncertainty about the number of steady states for a given device (e.g., a microwave oven can operate in states of defrost, heat with low power, or with high power), multiple devices exhibiting the same power level (e.g., lights and monitors), concurrent switchings on/off of multiple devices (e.g., printers and PCs), distinguishing start up transients from steady state levels (the former could persist for significant periods in time in commercial buildings), variable speed devices that show continuous power levels, and rare operation of some devices (because they are seldom operated by humans). These challenges are aggravated in commercial buildings (Norford and Leeb 1996) compared to residential buildings.

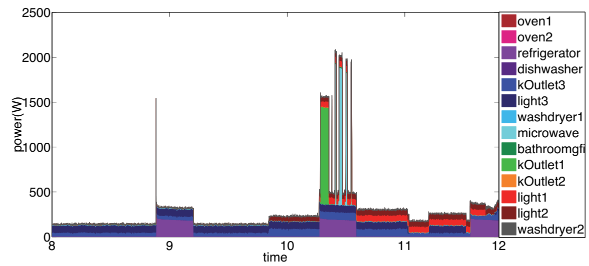


Figure 3: Example of energy disaggregation.

Features from meters. Let us first review the type of features discernible from metered usage data. From low frequency measurements, it is possible to infer features such as steady states, real power, reactive power, low-order harmonics, and the time of day. From high frequency measurements, in addition, we will be able to discern characteristics such as higher-order harmonics and the current or voltage waveform. In addition, from high frequency data, it is possible to discern transient states.

Prior approaches to disaggregation. Initial research focused on using simple device features such as real power and reactive power (Hart 1992). With the development of automated meters, transient states generated when devices turn on have been employed to identify devices (Shaw 2000). Raw current waveforms (Srinivasan, Ng, and Liew 2006), and voltage waveforms (Lam, Fung, and Lee 2007), and transforms of the current waveform (Chan, So, and Lai 2000) have also been adopted as characteristics. In particular, harmonics of non-linear devices have been utilized in prior work (Chan, So, and Lai 2000). Further, non-AC power features such as power line noises (Patel et al. 2007), time of day and device correlations (Kim et al. 2010), can be combined with AC power features to aid disaggregation. The underlying algorithms have been drawn from a variety of domains: supervised learning (Nakano and Murata 2007), data mining, optimization, and signal processing, e.g., kNN (Shaw 2000), SVM (Patel et al. 2007), sparse coding (Kolter, Batra, and Ng 2010). Recent research has placed a great emphasis on building in unsupervised learning features, including hierarchical clustering (Lam, Fung, and Lee 2007), semi-supervised approaches (Parson et al. 2012), factorial HMMs (Kim et al. 2010), and AFAMAP (Kolter and Jaakkola 2012).

Temporal Motif Mining

Early approaches to disaggregation (e.g., Hart(Hart 1992)) assume that only the aggregated current and voltage infor-

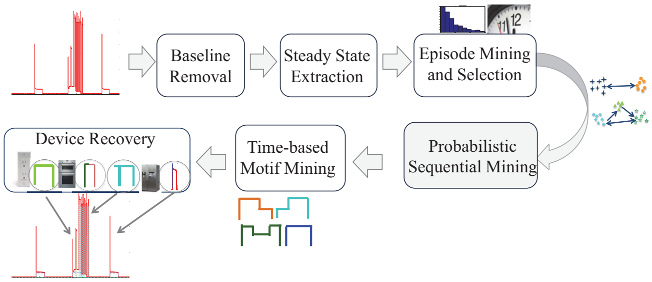


Figure 4: Temporal motif mining framework for disaggregation.

mation is known whereas later work assumes that the number of devices, possible steady states of devices are also known, so that the problem reduces to minimizing the error between the combination of disaggregated devices and the ground truth devices. Here, we assume that the number of devices/number of circuits is known, a reasonable assumption since such information is obtainable from a top-level circuit map of the building.

Our framework (see Fig. 4) unifies clustering and temporal data mining to discover power levels, forms episodes from power levels corresponding to devices, and models the underlying time series as a mixture model whose components correspond to the device episodes. The framework has six key stages, viz. baseline removal, steady states extraction, episode mining and selection, probabilistic sequential mining, motif mining or time-based motif mining, and device recovery. Gray box in Fig. 4 denotes that the step can be neglected (and are typically used when disaggregating for commercial buildings).

Baseline extraction. Baseline removal aims to separate devices that are always on. Given the aggregated (input) power series $P(t)$ over time period T, the baseline power P_{base} is defined such that $P_{\text{base}} \geq \min_t P(t)$ and where $f(P_{\text{base}}) \geq \alpha T$ (a minimum support threshold).

Steady state extraction. Two basic approaches here involve a heuristic method (window-sized filtering) and the more systematic Dirichlet process Gaussian mixture models (DPGMMs) (Görür and Rasmussen 2010). In the former, a mean filter smoothing is typically applied whose window size is adjusted to correspond to the mean or maximal start time duration in the given collection of devices (e.g., this could be just a second in the case of lighting, but higher for say a refrigerator). A DPGMM can be viewed as an infinite-mixture extension of a traditional Gaussian mixture model (GMM). Recall that in a traditional GMM, $\mathbf{y} = \Sigma_{i=1}^{k} \alpha_i N(\mu_i, \Sigma_i)$ where $\Sigma_i \alpha_i = 1$, and each component has a mean μ_i and covariance matrix Σ_i. A DPGMM defines Gaussian priors for all the component means μ_j:

$$p(\mu_j | \lambda, r) \sim N(\lambda, r^{-1})$$

The distribution of λ is set to be a Gaussian prior and the distribution of r is set to have a Gamma prior, so that the number of points in each component i conforms to a multinomial distribution with an unknown number of components. After modeling all the power levels in this manner, we replace all values with their representative (nearest centroid) power levels, record only the differences in successive power levels, and use this 'diffs' time series for further modeling.

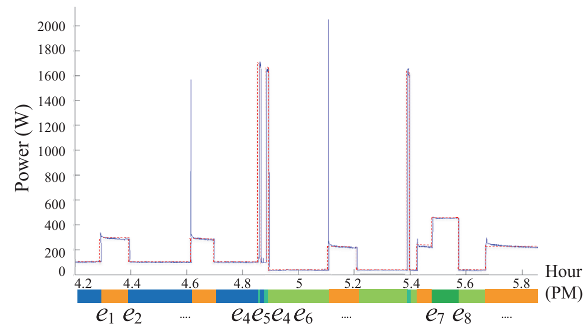


Figure 5: Mining episodes from a symbolic time series.

Episode mining and selection. The goal of episode mining (Patnaik et al. 2009) is to identify repetitive sequences of power level changes and, further, to isolate (select) those episodes that potentially correspond to the operation of a single device. Recall that at this point, we have generated a symbolized time series from the 'diffs' data. Let the set of symbols be S. From the diffs sequence, the transitions between symbols are recorded to help constitute episodes. We set the max episode length to be N, corresponding to the N-1 states of a device. Then all the symbols in the symbol set are permuted with length from 2 to N. As a result, all possible episodes with length from 2 to N are generated. To select valid episodes, some constraints checks are performed.

First, steady state values extracted from the previous step are clustered into a discrete symbol time series and transitions between symbols are recorded to identify episodes. Fig. 5 describes how transition events are generated, resulting in the event series: $(e_1, e_2, e_1, e_2, e_4, e_5, e_4, e_6, e_1, e_2, e_4, e_5, e_1, e_7, e_8, e_1)$. An episode of length N, $E = (e_1 \rightarrow e_2 \rightarrow ...e_N)$, denotes an ordered sequence of (not necessarily consecutive) symbols. To select those episodes that correspond to characteristics of an electrical device, several constraints are introduced:

1. *The sum of the power level changes corresponding to the events of a episode is nearly zero.* Fig. 2 (left) shows an example, where there are two complete episodes for a washer-dryer: (+1700, -1700) and (+1700, +800, -2250, -250).

2. *The sum of the power level changes corresponding to any prefix of a episode is positive.* This constraint is particularly geared toward multiple state devices. Fig 6 shows two examples of episode selection based on this constraint The episode (+100,-100) is retained but the episode (-100,+100) is discarded. As another example, episode (+600, -400, +1000) is chosen and episode (+600, -1000, +400) is discarded. Note that this assumes there are no always on devices.

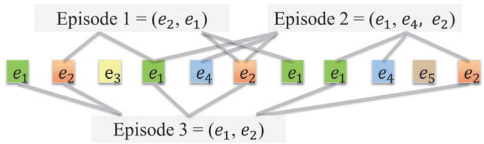


Figure 7: Illustration of motif mining. Note that there are 3 non-overlapped occurrences of Episode 3.

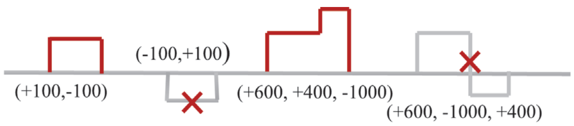


Figure 6: Episode constraints.

3. *The absolute value of the power level of any event in an episode related to a device must be higher than a support threshold over the maximum level in the episode.* In other words, power state changes in a device are assumed to be greater than a support threshold. This condition is intended to exclude cases where the low power consumption of one device inadvertently forms part of the episode of a high power consumption device. For instance, using a support threshold of 0.1, the episode $(1000, -850, -90)$ will get disqualified (because $90 < 100$) since this episode is likely generated by more than one device, rather than a single device.

Probabilistic sequential mining. This step aims to discover devices that exhibit several power levels sequentially and which operate frequently within a very short period of time. We use sequential mining (Agrawal and Srikant 1995), a levelwise framework, with duration constraints to discover such devices. We begin by seeking episodes that satisfy the above three checks and which can be systematically grown into longer chains of power level changes within a user-specified window.

Devices in commercial buildings are often scheduled to turn on/off at fixed time. Therefore, we cluster power levels according to time of day and day of week. We apply hierarchical clustering with Ward Euclidean distance to diffs of power levels. As a result, each set of power level diffs that qualifies the three constraints are chosen. For example, a cluster can identify a power level diff set $S = \{e_1, e_2, ...e_n\}$ belonging to a single device.

Regarding probabilistic sequential mining, a coverage probability θ, say 0.9, is introduced to determine what percent of power levels should be covered for each device. Probabilistic sequential mining only considers the coverage of power levels rather than the sequence of power levels as motif mining.

Motif mining. Motif mining aims to find repetitive episodes in a time series using the technique of non-overlapped occurrences (Laxman, Tankasali, and White 2008). Assume there are five power event change symbols $\{e_1, ..., e_5\}$ and a time series $(e_1, e_2, e_3, e_1, e_4, e_2, e_1, e_1, e_4, e_5, e_2)$ which is produced by these five symbols as shown in Fig. 7. Consider the episode Episode 3, composed of two ordered events (e_1, e_2). In this time series, there are four e_1 and three e_2 occurrences, and three instance of Episode 3. The first e_1 and first e_2 comprise the first instance of Episode 3. The second e_1 and second e_2 make up of the second instance of Episode 3. The third instance is composed of the fourth e_1 and the third e_2. Other possible instances of Episode 3 which may cause overlaps with the above instances are not considered in the non-overlapped count measure, such as (e_1, e_2) which consists of the first e_1 and the second e_2. With this count measure, all episodes that have support greater than the specified threshold are discovered by motif mining. For commercial buildings that have scheduled on/off devices, we adopt a time-constrained version of non-overlapped count, where the episode growth is restricted to events that fall within a specified time window.

Computational complexity Assume m is the number of power levels in the 'diffs' data. Then the computational complexity of DPGMM is $O(mnd^2 + md^3)$, where n is the number of points in diffs data, and d is the number of feature dimensions (e.g., time, date). The computational complexity for the episode generation step is $(p-1)O(m^2)$, where p is the maximal episodes length. Since p, which is 3, and m, which is 14 or 27, are small, we apply a brute force approach. The worst-case time complexity of the motif mining algorithm is $O(msq)$, where q is number of candidate episodes, and s is the size of the episode.

Parameters There are three kind of parameters used: (1) those pertaining to power level generation, (2) threshold for moitif mining, and (3) window size for median filtering. For each of these, a range of values were tried and their values were set based on performance on a test set.

Evaluation

We use precision, recall and F-measures in our evaluation. The standard definition of these metrics are: precision $= \frac{TP}{TP+FP}$, recall $= \frac{TP}{TP+FN}$, F-measure $= \frac{1}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$

We need to define the notions of true/false positives and negatives in the context of disaggregation.

Now suppose there is a ground truth time series X with length T; denote the corresponding disaggregated time series by X^*. For any time $t \in (0, T)$, there are two values: the ground truth value $X_i(t)$ and the disaggregated value $X_i^*(t)$. We define a parameter ρ for the range of true values $X_i(t)$ and another parameter θ as the noise. For any given measurement, there are four total power values at each point: true positive Ψ_{TPi}, false negative Ψ_{FNi}, true negative Ψ_{TNi}, and false positive Ψ_{FPi}.

1. When $X_i(t) > \theta$ and $X_i^*(t) > \theta$, at this point the disaggregation is a true positive. There are three situations in turn:

1.1. When $X_i(t) \times (1-\rho) < X_i^*(t) < X_i(t) \times (1+\rho)$, then

$$\begin{aligned}\Psi_{TPi} &= X_i^*(t) \\ \Psi_{FNi} &= \Psi_{FPi} = \Psi_{TNi} = 0\end{aligned}$$

1.2. When $X_i^*(t) < X_i(t) \times (1-\rho)$, then only the disaggregated power is considered as true positive and the power that is not disaggregated is regarded as a false negative:

$$\begin{aligned}\Psi_{TPi} &= X_i^*(t)\\ \Psi_{FNi} &= X_i(t) - X_i^*(t)\\ \Psi_{FPi} &= \Psi_{TNi} = 0\end{aligned}$$

1.3 When $X_i^*(t) > X_i(t) \times (1+\rho)$, then the disaggregated power is a true positive, and those values which are greater than the truth values are treated as false positive.

$$\begin{aligned}\Psi_{TPi} &= X_i^*(t)\\ \Psi_{FPi} &= X_i^*(t) - X_i(t)\\ \Psi_{FNi} &= \Psi_{TNi} = 0\end{aligned}$$

2. When $X_i(t) > \theta$ and $X_i^*(t) < \theta$, at this point the disaggregation is a false positive. Then,

$$\begin{aligned}\Psi_{FPi} &= X_i(t)\\ \Psi_{TPi} &= \Psi_{FNi} = \Psi_{TNi} = 0\end{aligned}$$

3. When $X_i(t) < \theta$ and $X_i^*(t) > \theta$, at this point the disaggregation is a false negative. Then,

$$\begin{aligned}\Psi_{FNi} &= X_i(t)\\ \Psi_{TPi} &= \Psi_{FPi} = \Psi_{TNi} = 0\end{aligned}$$

4. When $X_i(t) < \theta$ and $X_i^*(t) < \theta$, at this point the disaggregation is a true negative. Then,

$$\Psi_{TPi} = \Psi_{FNi} = \Psi_{FPi} = \Psi_{TNi} = 0$$

For the REDD dataset which features a maximal power level of 4000W, we use $\theta = 30$ and $\rho = 0.2$.

Experiments on REDD dataset

We conduct experiments on the low frequency data from the REDD (Kolter and Johnson 2011) dataset. We focus on 'House 1' since it has the most complete information (for validation purposes) and because it features 18 devices, providing a good test for our algorithm. The sampling frequency of both the mains is 1s and that of each circuit is 3s. The power consumption for devices in this dataset ranges from 50W to 4000W.

Disaggregation experiments

Knowing the ground truth, we synthesize aggregate data with different combinations of devices/circuits and evaluate our algorithm by disaggregating the combined data into the constituent devices. Fig.8 (a),(b),(c) show the plots of precision, recall, and F-measure values for 14 devices. For each device the number of aggregate devices was increased from 2 to 11. Since for k devices, there are ${}^{14}C_{k-1}$ possible combinations for each device, the results show the average over all the combinations. In cases where number of such combinations exceeded 100, 100 combinations were randomly sampled and averaged. Fig. 8(d) plots the power-weighted precision, recall and F-measure for these cases.

From Fig. 8(a), we can see that devices that are used frequently (both consuming low and high power), such as oven2 (4000W), microwave (1527W), kitchen outlet1 (1076W) (kOutlet1), washdryer2 (2712W), refrigerator(193W) and light1(64W) exhibit a stable precision level (above 0.7) even with increase in number of devices.

In contrast, devices such as kOutlet2 (1535W) (kitchen-outlet2), that share similar power levels with microwave (1527W) and bathroomgfi (1605W) show greater precision drops with increase in number of synthesized devices. However, the more frequently such devices are used, the greater the precision level.

As Fig. 8 (c) shows, devices with higher power or frequent use can be disaggregated well by motif mining. If a low power consumption device is prone to be influenced by high power devices, identification depends on the devices masking it; ultimately frequency of use helps disambiguate such situations. Finally, as Fig. 8 (d) shows, precision, recall and F-measure decrease only slightly with increase in the synthesized number of devices. This shows that power levels of devices play a key role in determining accurate disaggregation. When true power levels are supplied, the average precision, recall and F-measure of motif mining fare slightly better than AFAMAP.

Comparison of Motif Mining and AFAMAP

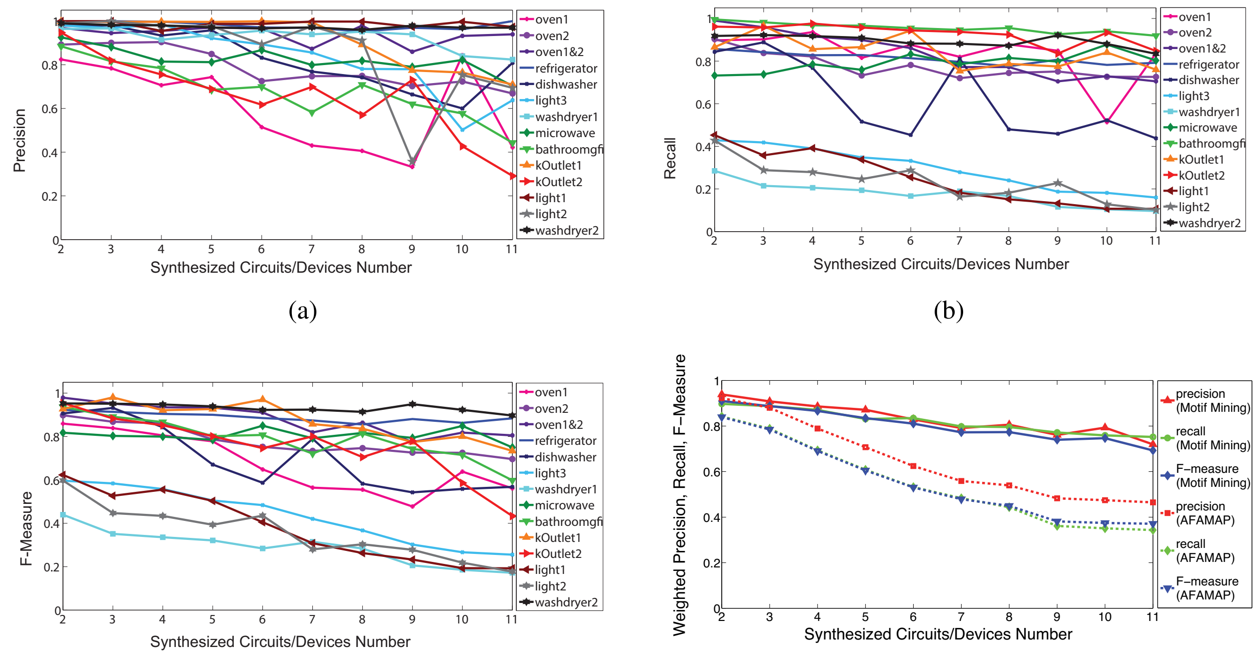
Next, we conduct experiments comparing our approach with the AFAMAP algorithm (Kolter and Jaakkola 2012), and also develop a method that combines motif mining and AFAMAP. Unlike motif mining, AFAMAP requires the power levels of each device; when running AFAMAP separately, we use the ground truth power levels for each device. When using AFAMAP in conjunction with motif mining, we use the power levels from generated episodes as an input to AFAMAP. Table 1 lists the results of the comparison.

In all, there are 18 devices but 4 of them are seldom used; and, thus the remaining 14 devices can be disaggregated by these three methods. For high power consumption devices, such as oven1&2, bathroom_gfi, kitchen_outlet1, kitchen_outlet2 and washdryer2, motif mining performs much better than AFAMAP even when AFAMAP is supplied with the ground truth power levels. For some of the low power consumption devices (such as light1), AFAMAP performs better. For high frequency devices, such as the refrigerator, motif mining performs much better.

Furthermore, by integrating motif mining and AFAMAP, we see the performance is much better than the individual algorithms on multiple state devices such as dishwasher and light3. Since the power level of light3 is low, the performance of the integrated method is better than using only motif mining.

Commercial Building Dataset

We applied our framework to a dataset from a commercial building (from HP Labs' campus in Palo Alto, CA). Data was collected from a branch in the electrical infrastructure of a large building and is composed of a root (aggregate) node and seven child nodes. Although all the nodes are instrumented with meters, we assume only the root and two of the child nodes, a transformer and a sub-panel, are available. The remaining five child nodes are devices that need to



(a) (b) (c) (d)

Figure 8: We increase the number of synthesized circuits from 2 to 11 and calculate performance measures for disaggregation of each device. (a) Precision (b) Recall (c) F-measure (d) The precision, recall and F-measure of all the devices are combined weighed by their average power levels.

Table 1: Comparing Motif mining against AFAMAP on the REDD dataset.

device	True Power (W)	Motif mining			AFAMAP (true power levels supplied)			Motif mining & AFAMAP		
		Precision	Recall	F-Measure	Precision	Recalll	F-Measure	Precision	Recall	FMeasure
oven1&2	4000	0.9297	0.5209	0.6677	0.4902	0.6750	0.5680	0.4008	0.6708	0.5018
refrigerator	193	0.9759	0.7368	0.8396	0.8825	0.3329	0.4834	0.7791	0.5433	0.6402
dishwasher	1113; 900; 400; 200	0.9786	0.2858	0.4423	0.062	0.4104	0.1077	0.5337	0.7431	0.6213
kOutlets3	100; 60	0.1487	0.0318	0.0524	0.6928	0.0439	0.0825	0.467	0.2892	0.3572
light3	282; 90	0.5768	0.1349	0.2187	0.4396	0.023	0.043	0.5519	0.1973	0.2907
washdryer1	466; 50	0.1789	0.1236	0.1462	0.3621	0.5401	0.4336	0.1703	0.6349	0.2686
microwave	1527	0.8035	0.3799	0.5158	0.5909	0.2907	0.3897	0.4512	0.3741	0.4090
bathroomgfi	1605	0.5199	0.6815	0.5898	0.2642	0.7551	0.3915	0.1075	0.406	0.1700
kOutlet1	1076	0.9320	0.6997	0.7993	0.21	0.7313	0.3264	0.2636	0.6394	0.3733
kOutlet 2	1535	0.2233	0.6261	0.3292	0.1153	0.2821	0.1637	0.0234	0.0826	0.0365
light1	64	0.6199	0.1963	0.2981	0.7972	0.0796	0.1447	0.667	0.1759	0.2784
light2	53	0.2603	0.1404	0.1824	0.6658	0.0817	0.1455	0.446	0.2776	0.3422
washdryer2	2711	0.9563	0.8305	0.889	0.7516	0.4237	0.5419	0.6427	0.3301	0.4361

be disaggregated. These are: a pump, a fan, an exhaust fan, a blower, and an elevator. The real power of all nodes are logged at intervals of 10 seconds. Using ground truth data, we combine all five to synthesize the aggregated data.

Table 2: Evaluation measures for commercial building disaggregation.

Device	Precision	Recall	F–measure
Pump and blower	0.99	0.99	0.99
Fan	0.99	0.99	0.99
Elevator	0.75	0.52	0.61

After the processing steps as described in our framework, we find five power levels that often occur in a range of just around 1 minute. Therefore we set the window size to 60 seconds and apply probabilistic sequential mining using a probability of 0.8 (as described earlier). The precision and recall for extracting individual devices is shown in Table 2.

In analyzing these results, we discover that the baseline power is constituted of two devices, namely, the pump and the blower. The elevator shows a sequential episode involving six power levels. The scheduled device is a fan. The only un-disaggregated device in our experiments is the exhaust fan which has very low power consumption compared to others and thus can be disregarded.

Discussion

We have described an intuitive motif-based approach to disaggregation that performs well relative to more complex algorithms that perform detailed modeling of temporal profiles. More importantly, we have demonstrated how our approach is not just an aid to disaggregation but, as a byproduct, also extracts temporal episodic relationships that shed insight into consumption patterns. In this sense, our work goes further than past work into addressing the real goal of disaggregation research, namely, to understand systematic trends in consumption patterns with a view toward identifying opportunities for savings.

References

Agrawal, R., and Srikant, R. 1995. Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, 3–14. IEEE.

Chan, W.; So, A.; and Lai, L. 2000. Harmonics load signature recognition by wavelets transforms. In *Electric Utility Deregulation and Restructuring and Power Technologies, 2000. Proceedings. DRPT 2000. International Conference on*, 666–671. IEEE.

Chiu, B.; Keogh, E.; and Lonardi, S. 2003. Probabilistic Discovery of Time Series Motifs. In *the Ninth ACM International Conference on Knowledge Discovery and Data Mining*, 493–498.

Görür, D., and Rasmussen, C. 2010. Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution. *Journal of Computer Science and Technology* 25(4):653–664.

Hart, G. 1992. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 80(2):1870–1891.

Kim, H.; Marwah, M.; Arlitt, M.; Lyon, G.; and Han, J. 2010. Unsupervised disaggregation of low frequency power measurements. In *11th International Conference on Data Mining*, 747–758.

Kolter, J., and Jaakkola, T. 2012. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Journal of Machine Learning Research - Proceedings Track (JMLR)*, volume 22, 1472–1482.

Kolter, J., and Johnson, M. 2011. Redd: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*.

Kolter, J.; Batra, S.; and Ng, A. 2010. Energy disaggregation via discriminative sparse coding. In *Proc. Neural Information Processing Systems*.

Lam, H.; Fung, G.; and Lee, W. 2007. A novel method to construct taxonomy electrical appliances based on load signatures. *Consumer Electronics, IEEE Transactions on* 53(2):653–660.

Laxman, S.; Tankasali, V.; and White, R. 2008. Stream Prediction using a Generative Model based on Frequent Episodes in Event Sequences. In *the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 453–461.

Nakano, Y., and Murata, H. 2007. Non-intrusive electric appliances load monitoring system using harmonic pattern recognition-trial application to commercial building. In *Int. Conf. Electrical Engineering, Hong Kong, China*.

Norford, L., and Leeb, S. 1996. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Buildings* 24(1):51–64.

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

Patel, S.; Robertson, T.; Kientz, J.; Reynolds, M.; and Abowd, G. 2007. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *Proceedings of the 9th international conference on Ubiquitous computing*, 271–288. Springer-Verlag.

Patnaik, D.; Marwah, M.; Sharma, R.; and Ramakrishnan, N. 2009. Sustainable Operation and Management of Data Center Chillers using Temporal Data Mining. In *the 15th ACM International Conference on Knowledge Discovery and Data Mining*, 1305–1314.

Shaw, S. 2000. *System Identification Techniques and Modeling for Nonintrusive Load Diagnositics*. Ph.D. Dissertation, Massachusetts Institute of Technology, Boston, MA.

Srinivasan, D.; Ng, W.; and Liew, A. 2006. Neural-network-based signature recognition for harmonic source identification. *Power Delivery, IEEE Transactions on* 21(1):398–405.

Yankov, D.; Keogh, E.; Medina, J.; Chiu, B.; and Zordan, V. 2007. Detecting Time Series Motifs under Uniform Scaling. In *the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 844–853.