Recursive Multivariate Piecewise Motif Mining to Disaggregation

Huijuan Shao[1,2], Yaowei Li[3], Fei Li[2], Erin Griffiths [4], Kamin Whitehouse[4], Naren Ramakrishnan[1,2]
[1]Discovery Analytics Center, Virginia Tech, Blacksburg, VA 24061
[2]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061
[3] Game source, McLean, VA 22101
[4]Department of Computer Science, University of Virginia, VA 22904

ABSTRACT

With the advent of modern sensor technologies, significant opportunities have opened up to help conserve energy in residential and commercial buildings. Moreover, the rapid *urbanization* we are witnessing requires optimized energy distribution. This paper focuses on two one problem in improving energy conservation; *energy disaggregation*. Energy disaggregation attempts to separate the energy usage of each circuit or each electric device in a building using only aggregate electricity usage information from the meter for the whole house. We cast this problem as *temporal mining problems*. We exploit motif mining with constraints to distinguish devices with multiple states, which helps tackle the energy disaggregation problem. Our results reveal that motif mining is adept at distinguishing devices with multiple power levels and at disentangling the combinatorial operation of devices.

KEYWORDS

disaggregation, motif mining

ACM Reference format:
Huijuan Shao[1,2], Yaowei Li[3], Fei Li[2], Erin Griffiths [4], Kamin Whitehouse[4], Naren Ramakrishnan[1,2]
[1]Discovery Analytics Center, Virginia Tech, Blacksburg, VA 24061
[2]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061
[3] Game source, McLean, VA 22101
[4]Department of Computer Science, University of Virginia, VA 22904
. 2017. Recursive Multivariate Piecewise Motif Mining to Disaggregation. In *Proceedings of ACM Woodstock conference, Halifax, Nova Scotia, Canada, August 2017 (KDD'17),* 8 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

With the advent of modern sensor technologies, significant opportunities have emerged to help conserve energy in residential and commercial buildings. Moreover, the rapid *urbanization* we are witnessing requires optimized energy distribution. Energy disaggregation attempts to separate the energy usage of each circuit or each electric device in a building using only aggregate electricity usage information from the whole house meter. Usually two-phase or three-phase electric power is connected to residential and commercial buildings. Similarly, water disaggregation aims to discover each water use end by only knowing the hot and cold water usage from the whole house water meter. We generalize these two problems, energy disaggregation and water disaggregation, as a multiple-phase data disaggregation problem. The aim of this paper is to identify electrical devices or water use ends from two phases of aggregated data. Unlike previous work which disaggregate devices from the sum of multiple phases, the time series information from each phase and the correlation of a device between/among phases are fully used. All of this information enables us to characterize more devices. This work makes the following contributions in the field of disaggregation:

(1) It can disaggregate aggregate data from multiple phases.
(2) It can separate the continuously variable loads which are mixed in electricity.
(3) This approach can be used for both electricity disaggregation and water disaggregation.

2 PRIOR WORK

Electricity disaggregation uses the electricity consumption level at the main entry into a building or house to infer whether a device inside the building is on or off. The features used include initial real power and reactive power [6] from a dataset which is recorded in a low-frequency range. With advances in electrical meter technology and the availability of less expensive meters, more and more features are being extracted from the high-frequency data set and used for disaggregation, such as the transient state generated when a device turns on or off [21], the raw current waveform [22], the voltage waveform [10], the transform of the current waveform [2], and harmonics of non-linear devices [2]. Even on-AC power features such as power line noises [17] are exploited jointly with AC power features like time of day, and device correlations [8] in modern systems.

Increasingly, research is being focused on unsupervised learning and semi-supervised learning algorithms because these algorithms do not require the power consumption of each device, and the power or water usages of individual devices are very difficult to obtain. It is only in the last few years that unsupervised learning algorithms have been used, including hierarchical clustering [5], factorial hidden Markov models (FHMMs) [8], additive factorial approximate MAPs (AFAMAP) [9], difference FHMMs [16], and motif mining [20]. Semi-supervised learning algorithms [7, 10] have also been proposed. In this paper, we assume the number of devices and the number of power level states of each device are known. Hence, we formalize the

KDD'17, Halifax, Nova Scotia, Canada

DOI: 10.1145/nnnnnnn.nnnnnnn

disaggregation as a semi-supervised problem and provide solutions to the following three challenging problems.

(1) Several devices may have the same real power, and it is difficult to distinguish these devices using only the recorded aggregated power time stamp.
(2) Many devices may turn on or off at the same time.
(3) Instead of having a discrete range of power levels, there are devices whose power consumption levels vary gradually, e.g., variable speed devices (VSD) and lights with dimmers. Once their power usage is aggregated with that from other devices, disaggregation becomes increasingly difficult.

Since obtaining a low-frequency dataset is more practical in real buildings, we focus mainly on real power, which can be easily extracted from a low-frequency dataset.

Water use disaggregation has emerged in recent years, and so far the applied algorithms are limited to supervised learning algorithms [1]. This paper proposes water disaggregation as a semi-supervised learning algorithm by presuming that we know the number of water use ends and the water usage level of each water user end.

3 DISAGGREGATION FORMALISM

I propose a semi-supervised approach for disaggregation; i.e. I assume that we know the on and off events for a short period of time for all devices or water end uses, and use that information to deduce the power levels or water usage, or to obtain the startup vectors of every device.

For our purposes, we define the disaggregation problem as follows: Given K-phase aggregated power or K aggregated water consumption time series $Y_k = y_1^{(k)}, ..., y_T^{(k)}$, and a set of power or water related and contextual features, $f = f_1, ..., f_T$ over a period of time T, the problem is to estimate the disaggregated power or water consumption of M devices $\hat{X_m} = \hat{x}_1^{(m)}, ...\hat{x}_t^{(m)}, ...\hat{x}_T^{(m)}, m \in [1, M]$, such that a loss function of the sum of the power or water consumption of the M devices and the sum of the K phases of aggregated power or water consumption is minimized.

$$\min_{\hat{x}_t^{(m)}} \{\sum_{t=1}^{T} \mathscr{L}_t(\sum_{m=1}^{M} \hat{x}_t^{(m)}, \sum_{k=1}^{K} y_t^{(k)})\}, \tag{1}$$

where $\mathscr{L}_t$ is the loss function between the sum of M estimated time series at t, and $y_t^{(k)}$ is the ground truth phase k aggregated power or water feature at time t. $\mathscr{L}$ is usually the $\mathscr{L}$1-norm $|\sum_{m=1}^{M} \hat{x}_t^{(m)} - \sum_{k=1}^{K} y_t^{(k)}|$ or the $\mathscr{L}$2-norm $(\sum_{m=1}^{M} \hat{x}_t^{(m)} - \sum_{k=1}^{K} y_t^{(k)})^2$.

4 TEMPORAL DATA MINING AND THE PROBABILISTIC MODEL

Temporal datasets display a character of time-dependency. They are recorded frequently in smart buildings and build scenarios to infer the energy usage of people. Temporal data mining revolves around the techniques (algorithms) that enumerate structures, patterns, and signatures over temporal data (time series, for instance). A survey [12] has investigated several efficient techniques to discover the patterns in ordered data streams. The techniques used to discover significant patterns vary according to the dataset. One of the compelling patterns in temporal data mining is frequent episodes [13].

Frequent Episode Discovery

Frequent episode discovery is proposed in [13]. Given a sequence of events $< (E_1, t_1), ..., (E_n, t_n) >$, where E_i denotes the i^{th} event at the time of t_i, the aim is to find temporal patterns (called *episodes*) that occur frequently in the long sequence. This episode is an ordered event collections. For in-stance an episode $(A \rightarrow B \rightarrow C)$ represents that event type A comes before event type B, which occurs before event type C. The occurring time of these events are unnecessary to be consequent. The frequency threshold is decided by a user. Several data mining algorithms have been researched to discover the frequent episodes [11, 13].

Motif Mining in Multi-variate Time Series Data

Motif mining is a temporal data mining technique that was initially proposed in [3] and [24] and extensively studied in [14, 18, 23]. The fundamental idea behind *motif mining* is that it symbolically encodes the numerical time series data. After this encoding, the symbols combine to form episodes in the data, resulting in patterns that can be mined. Furthermore, by combining domain-specific information and pattern mining techniques, we extract frequent, meaningful episodes from the symbolized time series.

Furthermore, when there are time series that describe the data, we employ *multi-variate motif mining* to find meaningful patterns. The algorithms for multi-variate temporal motif mining are similar to the univariate case, except that the symbolic encoding is represented as a vector. Therefore each time point in the data is represented as a vector of symbols, with each symbol corresponding to one of the several time series that represents the data. Now, the combination of these vector symbols forms episodes that can be mined from the multi-variate time series data. Again, by combining domain specific knowledge, we extract meaningful episodes from the data.

5 RECURSIVE MULTIVARIATE PIECEWISE MOTIF MINING

To solve the problem of separating a multi-dimensional time series into several time series, I propose the approach of recursive multivariate piecewise motif mining. Motif mining has been well studied in previous work [3] and [24]. Multivariate or multidimensional motif mining is further extended in [14] and [23] and [18].

Motif mining is applied to energy disaggregation in [20], in which discrete on/off events are exploited. This research enhances previous work by piecewise motif mining, where the on/off event is comprised of several consecutive data points, i.e. piecewise, other than individual discrete one. Also, I use multivariate motif mining to make full use of two- or three-phase aggregated data.

The framework of recursive multivariate piecewise motif mining to energy disaggregation is illustrated in Figure 1. The input includes multiple-phase aggregated data, such as two-phase data Mains1 and Mains2, and the power levels of each device. During the whole procedure, I recursively apply piecewise motif mining to two-phases and single phase diffs data. The first step is to identify electrical devices which draw power from both phases. Generally these devices consume a large amount of power, such as the water heater indicated by the blue line. These devices draw equal power or disparate power from both phases synchronously. Secondly, I remove the power consumption of the devices which draw power from both phases. This action decreases the noise interference caused by large power consumption and increases the possibility to disaggregate more devices with low-power consumption. Then we apply piecewise motif mining to single-phase data to separate the devices that draw power only from that phase, such as the humidifier indicated by the green line. Generally, multivariate piecewise motif mining is

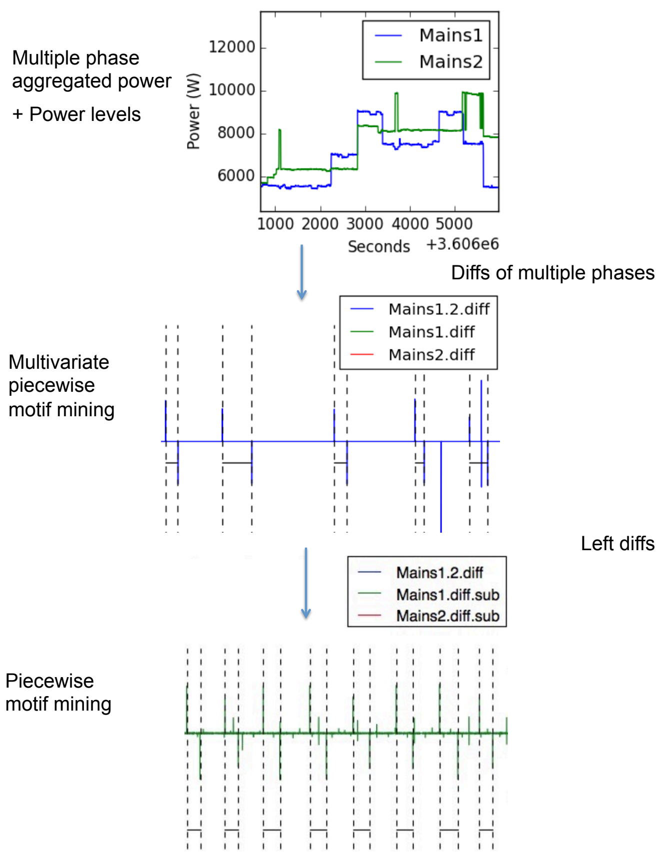


Figure 1: Recursive Multivariate Motif Mining Approach.

divided into four steps, as shown in Figure 2. Step 1 is to search for piecewise events from the two-phase or three-phase data. Step 2 is to encode events from multiple phases. Step 3 aims to mine frequent motifs from the encoded events list. The last step targets to recover devices from mined motifs.

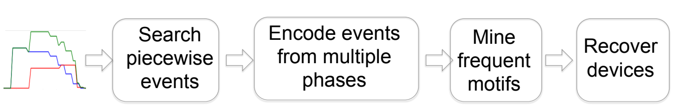


Figure 2: Multivariate Piecewise Motif Mining.

5.1 Piecewise Motif Mining

Motif mining aims to uncover the repetitive patterns in time series data, and works best for discrete events. Piecewise motif mining is proposed for energy disaggregation to detect on/off events.

Definition 5.1. **Piecewise Event** Given a time series diffs data $y_1, ..., y_{n'}$, where $\forall\ |y_i| < \eta$. A piecewise event is the sum of these n' number of diffs data, $e = \sum_{i=1}^{n'} y_i$.

Each piecewise event corresponds to an on/off event of an electric device. The value of η is the noise range of each device, which is usually less than the 10% of $|e|$.

Piecewise Events Search from Multiple Phases

The majority of electrical devices which draw power from multiple phases consume larger amounts of power than electrical devices which connect to single phase. To disaggregate such a device, we need to discover specific on/off events features to separate them. Generally such an electrical device draws power from multiple phases synchronously and constructs a pattern. Some devices may consume equal power from both phases all the time, and so their power consumption patterns from both phases keep the same. Other devices may show different power usage patterns when drawing power from two phases. Algorithm 1 describes how synchronized

Algorithm 1 Search Synchronized Events from Two-phase Aggregated Diffs Data

Require: 2-phases aggregated diffs data $y_k = y_1^{(k)}, ..., y_n^{(k)}$ and $k = 1, 2$, big power consumption threshold θ

```
for i = 0 : n − 1 do
  if |y_i^(1)| > θ then
    for j = i − 5, i + 5 do
      if |y_j^(1)| ∈ [|y_j^(2)| ∗ 0.8, |y_j^(2)| ∗ 1.2] then
        e_i^(1) = e_i^(1) + y_j^(1)
        e_i^(2) = e_i^(2) + y_j^(2)
      end if
    end for
    e_i = e_i^(1) + e_i^(2)
  end if
end for
return e_1, ..., e_i, ..., e_n', ∀e_i > 2 ∗ θ
```

events from two phases are revealed. This input include the two-phase aggregated diffs data and big power consumption threshold θ. This threshold guarantees we only discover big power consumption devices. We review the phase-1 diffs data. If any absolute value $|y_i^{(1)}|$ is greater than θ, both previous and posterior five diffs data points from time i are checked. For these 10 points values, at each time j, if the difference between phase-1 $y_i^{(1)}$ and phase-2 $y_i^{(2)}$ is in the range of $0.2 * |y_i^{(1)}|$, we assert that the diffs data points from these two phases are relatively the same and synchronized. The synchronization implies that these two identical amounts power consumption comes from a single device. Therefore we sum the synchronized power level diffs data and compute the power consumption at time i as e_i. When $e_i > 0$ that denotes an on event, and $e_i < 0$ means an off event for a certain device.

Next we transfer these two-phase diffs data into an ordered on/off event list $e_1, ..., e_{n'}$, then we apply motif mining to this events list. By matching the devices which consume power greater than $2 * \theta$, we can separate all devices which draw equal amount of power from two phases.

5.2 Encoding Events From Multiple Phases

After deleting all the synchronized events from phase-1 and phase-2, we apply multivariate piecewise motif mining to the remaining phase-1 and phase-2 diffs data, to detect devices which consume large amounts of power and draw power from two phases synchronously yet unequally. There are different power drawing patterns from these two phases. We encode these two-phase diffs data, which occur at the same time, as a new event e. Figure 3 gives an example of how the events from two-phase circuits are encoded. We extract an event which consumes power greater than θ, then we check five more data points before and after it. The values of the 11 data points relevant to this event in Main1.diff are [0, 0, -18, 18, 1093, 1830, -196, -68, -37, -36, 0]. The concurrent events listed in Main2.diff are [0, 0, 0, 18, 9, 1946, 440, -51,-36,-36,0]. Since the events at the peak occur in the two phases as $(1830, 1946)$, and the difference of these two powers $1946 - 1830 = 116$ is in the $0.2 * 1830$ range, we consider that these two changes may come from a single device. When looking for insight into these two vectors, we observe that the sum of the changes of phase 1 is 2604W, and the sum of the changes of phase 2 is 2290W. They are in the same range, i.e. $2604 * 0.8 < 2290$. Therefore, we declare that the power changes from these two phases definitely come from a single device. We select two of these values and encode them as $e_{1'} = (1093, 9)$, $e_{2'} = (1830, 1946)$

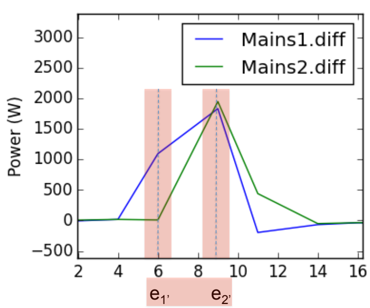


Figure 3: Encoding Events from Multiple Phases.

The piecewise events for this single device are $e = [e_{1'}, e_{2'}]$. Applying frequent motif mining, we separate this large power consumption device which draw power from two phases unequally.

6 EVALUATION

I use precision, recall and F-measure in our evaluation. The standard definitions 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.

Let us suppose there is a ground truth time series x with length T, and denote the corresponding disaggregated time series by $\hat{x}$. For any time $t \in (0, T)$, there are two values: the ground truth value of device m is $x_t^{(m)}$ and the disaggregated value $\hat{x}_t^{(m)}$. We define a parameter ρ for the range of true values $x_t^{(m)}$, and another parameter θ as the noise. For any given measurement, there are four total power values or water usage values of device m at each point: true positive $TP^{(m)}$, false negative $FN^{(m)}$, true negative $TN^{(m)}$, and false positive $FP^{(m)}$.

1. When $x_t^{(m)} > \theta$ and $\hat{x}_t^{(m)} > \theta$, at this point the disaggregation is a true positive. There are three situations in turn:

1.1. When $x_t^{(m)} \times (1-\rho) < \hat{x}_t^{(m)} < x_t^{(m)} \times (1+\rho)$, then

$$\begin{aligned} TP^{(m)} &= \hat{x}_t^{(m)} \\ FN^{(m)} &= FP^{(m)} = TN^{(m)} = 0 \end{aligned}$$

1.2. When $\hat{x}_t^{(m)} < x_t^{(m)} \times (1-\rho)$, then only the disaggregated power or water usage is considered as true positive and the power or water usage that is not disaggregated is regarded as a false negative:

$$\begin{aligned} TP^{(m)} &= \hat{x}_t^{(m)} \\ FN^{(m)} &= x_t^{(m)} - \hat{x}_t^{(m)} \\ FP^{(m)} &= TN^{(m)} = 0 \end{aligned}$$

1.3 When $\hat{x}_t^{(m)} > x_t^{(m)} \times (1+\rho)$, then the disaggregated power or water usage is a true positive, and those values which are greater than the truth values are treated as false positive.

$$\begin{aligned} TP^{(m)} &= \hat{x}_t^{(m)} \\ FP^{(m)} &= \hat{x}_t^{(m)} - x_t^{(m)} \\ FN^{(m)} &= TN^{(m)} = 0 \end{aligned}$$

2. When $x_t^{(m)} > \theta$ and $\hat{x}_t^{(m)} < \theta$, at this point the disaggregation is a false positive. Then,

$$\begin{aligned} FP^{(m)} &= x_t^{(m)} \\ TP^{(m)} &= FN^{(m)} = TN^{(m)} = 0 \end{aligned}$$

3. When $x_t^{(m)} < \theta$ and $\hat{x}_t^{(m)} > \theta$, at this point the disaggregation is a false negative. Then,

$$\begin{aligned} FN^{(m)} &= x_t^{(m)} \\ TP^{(m)} &= FP^{(m)} = TN^{(m)} = 0 \end{aligned}$$

4. When $x_t^{(m)} < \theta$ and $\hat{x}_t^{(m)} < \theta$, at this point the disaggregation is a true negative. Then,

$$TP^{(m)} = FN^{(m)} = FP^{(m)} = TN^{(m)} = 0$$

In our experimental dataset, we set $\theta = 100$ and $\rho = 0.2$. Although the maximal power consumption of all these devices is 11000W, we can still set $\theta < 11000 * 0.1$ because we apply multivariate piecewise motif mining recursively, so the devices which consume a large amount of power are deleted in the first few rounds. Therefore the power noise which is caused by the high-power electronic devices is greatly decreased.

7 EXPERIMENTS

We run experiments on the dataset Study10 from the University of Virginia on electricity disaggregation. This dataset collects data from 02/10/2014 to 02/21/2014 in a residential building. Two individuals were asked to live in an instrumented home for around two weeks. To ensure the data consisted of the personal usage patterns of the participants, they were encouraged to live in and use the home as they normally would use their own. An eMonitor [4] sensor was used to collect both mains data for testing and circuit-level information for ground truth. Additional data, such as the opening of appliance doors and the flicking of light switches, was collected to provide sub-circuit level ground truth information for events such as lights. Both the two-phase aggregated data and each device's data are collected at intervals of 2-3 seconds. In total, 25 devices were connected to two phases at the entry of the house. Five of these devices are seldom operated; less than five times. Fourteen devices consume power less than 100W, and the majority of them are lights. The largest power consumption of these devices is 11000W by indoor heating. The noise caused by the heating device is large; greater than 100W. Therefore we focus on disaggregating the six major electronic devices with power levels greater than 100W.

7.1 Electricity Disaggregation

We assume that we know the power levels of each device. If the power levels of each device are unknown, we can use the sum of two-phase aggregated data and the on/off events of the ground truth to extract them. We set a window size $w = 30s$ ahead and behind of the ground truth events to match the aggregated data. If there is only one power change in the aggregated data during these 60 seconds, this power level change must come from an on/off event of an electrical device. Usually, it takes around 2-5 seconds for an electrical device to reach a steady power level. The on and off events reflect different durations for a device to reach a steady state. Therefore, we measure the minimal duration of the on event and off event of each device. After we go over all the aggregated data and ground truth on/off events, we run a Gaussian mixture model to model the positive power changes and negative power changes independently. The means and standard deviations correspond to the on/off event of each device. The power levels, standard deviation, and on/off duration of each device of dataset study10 are listed in Table 1.

We apply recursive multivariate piecewise motif mining to dataset Study10 and compute the precision, recall and F-measure. Devices which draw power from both phases are separated first. They are heatingIndoor, waterheater and dryer. Figure 4 (a) gives an example of an on event in the two-phase Mains1 and Mains2. Mains1.diff denotes the diffs data from Mains1 and Mains2.diff represents the diffs data from Mains2. Mains1.2.diff shows as blue when Mains1 and Mains2 share the similar power changes. We can see that the power consumption of a specific device jumps twice in two phases simultaneously. The first time, both phases jump 2572W. After nine seconds, the power of both phases increases 2520W. The sum of these four changes is 10184W. Compared with the power levels of all devices, we speculate that these power changes are caused by the device heatingIndoor. Figure 4 (b) shares the same snippet of time series as Figure 4 (a). The red line indicates that the on event of heatingIndoor is recognized. Similarly, the off event plunges twice in two seconds -2877W and -1759W in both phases, as shown in Figure 4 (c). The sum of this off event is -9272W. After matching the power levels, we categorize it as the off event of heatingIndoor as indicated in Figure 4 (d).

The dryer has the same power level as the waterheater at around 4800W. If we disaggregate these two devices from the sum of the two phases, it's difficult to distinguish them, but with multivariate piecewise motif mining, these two devices can be distinguished.

Figure 5 (a) and (b) are the diff data of the dryer and waterheater from the two-phase circuit. We can see that the waterheater draws power from Phase 1 and Phase 2 at the same time, but the dryer shows a different pattern. It draws power from Phase 1 at a lower power of 1093W, then jumps to 1830W; at the same time, it draws power from Phase 2 at the high level of 1946W immediately. We encode the power usage as shown in Figure 3, then apply motif mining to disaggregate them.

After deleting the power consumption from both aggregated phases, we apply piecewise motif mining again to a single phase. We then discover the humidifier from Phase 1 and the microwave from Phase 2. When we only disaggregate the sum of Phase 1 and Phase 2, the precision recall result of the microwave and humidifier is not very accurate because sometimes their power consumptions are similar. However, using multivariate motif mining, we can separate them very clearly with good precision and recall. The precision and recall results for the data set Study10 are listed in Table 1.

Recursive multivariate motif mining is capable of disaggregating continuous variable loads. Figure 5 (c) shows the diff data of heatingOutdoor from the two phases. During this on event, its power levels change nine times, then continue at a relatively stable state. By applying piecewise motif mining, we can successfully identify this as the heatingOutdoor device after matching its power level. If another device D which draws from Phase 1 or Phase 2 is turned on or off during this period, multivariate piece-wise motif mining can still identify this heatingOutdoor device. This is because D only uses one phase's power; hence its power change is not counted in our piecewise event.

7.2 Water Disaggregation and Constraints

Water usage displays different characteristics. The total water consumption is zero most of the time. Whenever a water use end is operated, water is consumed intensively for a period of time. Then it will stay off for a much longer time. We observe that the operations of water use ends reflect a series of user behaviors. For instance, a person may use the toilet in the bathroom first, then wash hands in the sink and finally take a shower afterwards.

Similar to electricity disaggregation, we use a period of aggregated water usage data to extract features and obtain the water flow rate level of each water use end. Table 2 lists the water consumption rate for each device. For instance, taking a shower uses hot water at a flow rate between 0.1822 liter/min and 0.1986 liter/min. Let $\frac{\alpha}{10000}$ denotes this range of water flow rate. The total hot and cold water consumption by shower is 0.1904 liter/min. Therefore, the cold water flow rate caused by shower is $0.1904 - \frac{\alpha}{10000}$ liter/min. Turning on the water for the shower takes around two seconds.

After these calculations, we apply a multivariate piecewise motif mining approach to water disaggregation. For the shower and washing machine, the total flow rate of hot and cold water is high, nearly 0.2 liter/minute. Therefore by only searching the total hot and cold water flow rate, we can identify these two devices. The event of shower usage usually lasts for more than one minute, but

Table 1: Power Levels, Standard Deviation of Power Levels, On/off Duration, Connected Phases and Disaggregation Results of Electricity Devices from Study10.

Device	Power Levels	Standard Deviation	On/off Duration	Phase	Recursive Multivariate Motif Mining			AFAMAP		
					Precision	Recall	F-measure	Precision	Recall	F-measure
HeatingIndoor+ HeatingOutdoor	10590W	1270W	60s	1+2	0.979	0.928	0.953	0.870	0.45	0.598
Waterheater	4450W	350W	2-5s	1+2	0.999	0.997	0.998	0.627	0.882	0.733
Humidifier	1470W	90W	10s	1	0.997	0.992	0.995	0.725	0.858	0.787
Microwave	1850W	200W	10s	2	0.95	0.758	0.843	0.032	0.819	0.06
Dryer	5200W	400W	2-5s	1+2	0.911	0.996	0.952	0.011	0.561	0.021
	875W	225W	2-5s	1						

Table 2: Water Flow Rate Levels of Water End Uses.

Device	Hot water (liter/min*10000)	Cold water (l/min*10000)	Duration (second)
Shower	$\alpha \in (1822, 1986)$	$1904 - \alpha$	on: 2
Washing Machine	$\alpha \in (1988, 2276)$	$2132 - \alpha$	on: 5
DownToilet	0	(1270, 1400)	whole: 50
UpToilet	0	(1480, 1700)	whole: 50

the washing machine uses water for less than one minute, repeating six to nine times. Both the shower and washing machine use hot water and cold water. However, the washing machine uses hot water for only the first one or two times. For the rest of its cycle, only cold water is used. Whenever the washing machines starts, the power consumption starts as well.

Applying piecewise motif mining to the water usage lets us disaggregate the shower and washing machine. The precision, recall and F-measure for the shower disaggregation are 0.999, 0.972, and 0.986, and the precision, recall and F-measure for the washing machine disaggregation are 0.997, 0.969, and 0.983. However, with a variable water flow rate, piecewise motif mining has limitations in handling water use ends such as the toilet. Therefore we use the dynamic time warping subsequence [19] search as a complementary to discover these water use ends. For the two toilets, we apply dynamic time warping to match the time series. The water usage results of one toilet, UpToilet, is shown in Figure 5 (d).

We can see that multivariate piecewise motif mining is capable of disaggregating water use ends which have sharp on/off water flow rates. However, it has limitations in dealing with water use ends with irregular water use patterns, such as toilets and sinks. Since the water usage of toilets is relatively fixed if used alone, some toilet water usages can be disaggregated by using the dynamic time warping subsequence search which was researched in [15].

8 CONCLUSION

This paper proposes a semi-supervised recursive multivariate piecewise motif mining approach to electricity disaggregation. We use data from a period of time to find the features of individual devices, including the power level and standard deviation. Based on these features, we recursively utilize multivariate motif mining to uncover devices that draw power from both phases equally or unequally, then separate the devices from each phase using motif mining. Devices that use a large amount of power are removed from the two phases in the first few rounds, which brings the benefit of decreasing the noise caused by these devices during the process of disaggregating from single phase. Therefore more devices with smaller power consumption are separated using the piecewise motif mining from a single phase. In addition, this piecewise motif mining approach can identify continuously variable loads such as outdoor heating. Furthermore, when we apply motif mining approach to water disaggregation, it can separate water use ends that have steady water usage, such as a shower, but cannot disaggregate those water use ends which consume water variably for the whole cycle, like a toilet.

In the future, for energy disaggregation we will use more features from the aggregated data from the multiple phases, such as the startup shape of individual devices. For water disaggregation, we will explore how to integrate the dynamic time warping with multivariate piecewise motif mining.

REFERENCES

[1] Davide Carboni, Alex Gluhak, Julie A McCann, and Thomas H Beach. 2016. Contextualising Water Use in Residential Settings: A Survey of Non-Intrusive Techniques and Approaches. *Sensors* 16, 5 (2016), 738.

[2] WL Chan, Albert TP So, and LL Lai. 2000. Harmonics load signature recognition by wavelets transforms. In *Proceedings of the IEEE International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*.

[3] Bill Chiu, Eamonn Keogh, and Stefano Lonardi. 2003. Probabilistic discovery of time series motifs. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge discovery and Data Mining*.

[4] EnergyCircle. 2016. eMonitor Sensor. (2016). http://www.energycircle.com/energy-monitors/emonitor-energy-monitor-service-3rd-generation

[5] Hugo Goncalves, Adrian Ocneanu, Mario Berges, and RH Fan. 2011. Unsupervised disaggregation of appliances using aggregated consumption data.

[6] George William Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* (1992).

[7] Matthew J Johnson and Alan S Willsky. 2012. Bayesian nonparametric hidden semi-Markov models. *CoRR* (2012).

[8] Hyungsul Kim, Manish Marwah, Martin F Arlitt, Geoff Lyon, and Jiawei Han. 2011. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the SIAM International Conference on Data Mining*. 747–758.

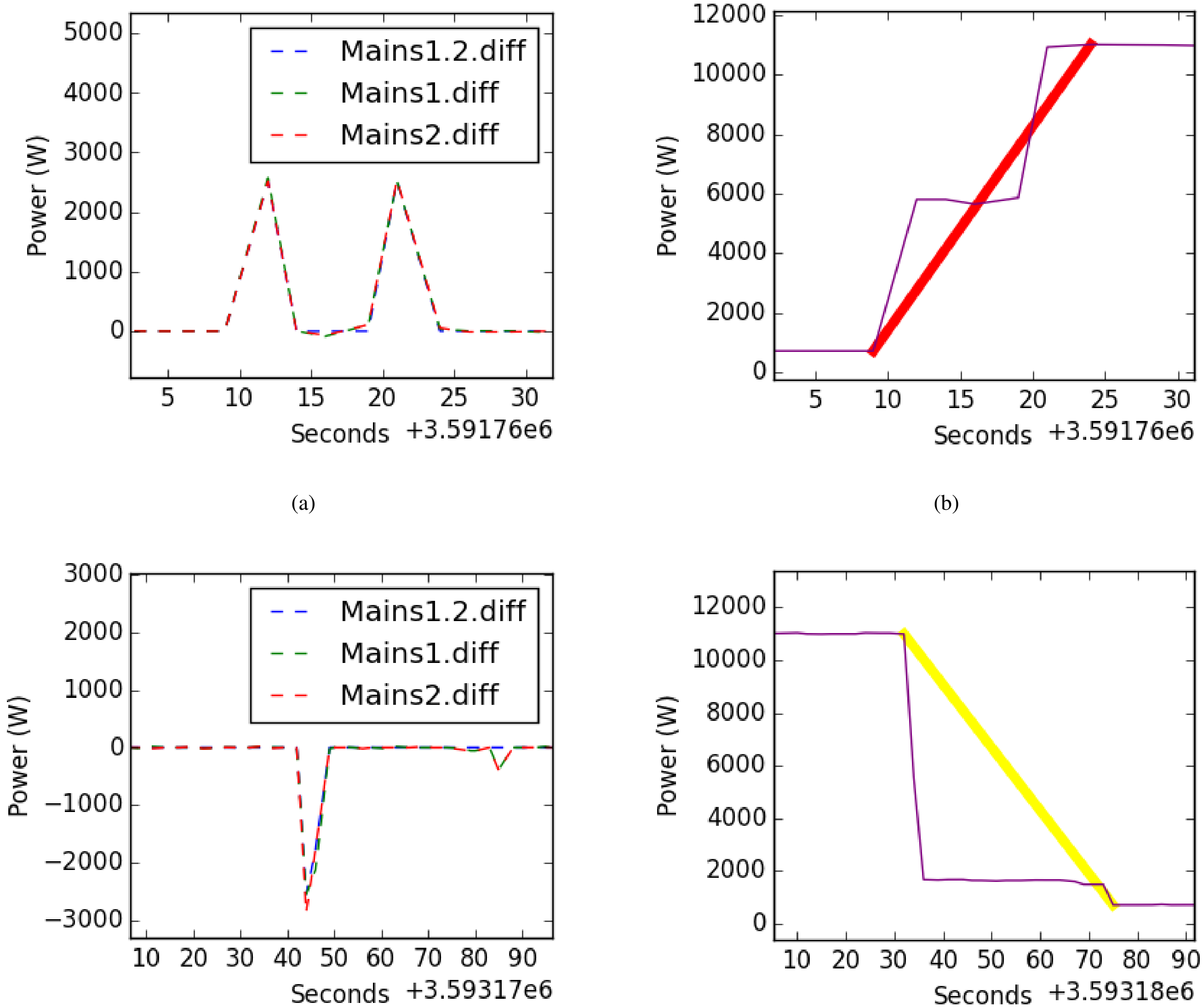
[9] J Zico Kolter and Tommi S Jaakkola. 2012. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *International Conference on Artificial Intelligence and Statistics*.

[10] HY Lam, GSK Fung, and WK Lee. 2007. A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures. *IEEE Transactions on Consumer Electronics* (2007).

[11] Srivatsan Laxman, PS Sastry, and KP Unnikrishnan. 2005. Discovering frequent episodes and learning hidden markov models: A formal connection. *IEEE Transactions on Knowledge and Data Engineering* (2005).

[12] Srivatsan Laxman and P Shanti Sastry. 2006. A survey of temporal data mining. *Sadhana* 31, 2 (2006), 173–198.

[13] Heikki Mannila, Hannu Toivonen, and A Inkeri Verkamo. 1997. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*



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

Figure 4: (a) On piecewise event and (c) Off piecewise event of heatingIndoor. heatingIndoor is disaggregated by motif mining the on event (b) and off event (d).

(1997).

[14] David Minnen, Thad Starner, Irfan A Essa, and Charles Lee Isbell Jr. 2007. Improving Activity Discovery with Automatic Neighborhood Estimation. In *IJCAI*, Vol. 7. 2814–2819.

[15] Khoi Anh Nguyen, Hong Zhang, and Rodney Anthony Stewart. 2013. Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-Environment Research* 7, 3 (2013), 182–201.

[16] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. 2012. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. Toronto, Canada.

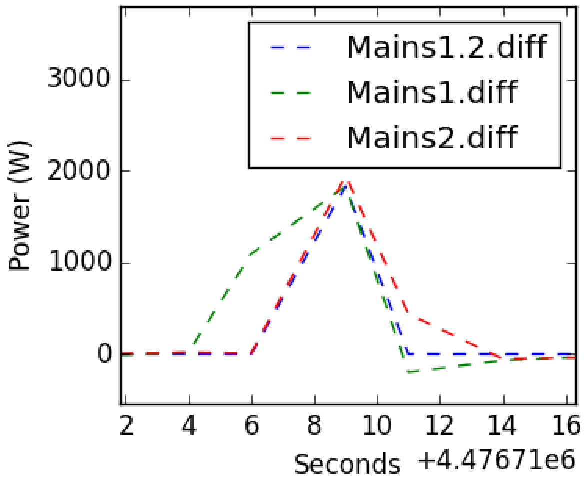
[17] Shwetak N Patel, Thomas Robertson, Julie A Kientz, Matthew S Reynolds, and Gregory D Abowd. 2007. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In *Proceedings of the 9th International Conference on Ubiquitous computing*.

[18] Debprakash Patnaik, Manish Marwah, Ratnesh Sharma, and Naren Ramakrishnan. 2009. Sustainable operation and ,anagement of data center chillers using temporal data mining. In *Proceedings of the 15th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*.

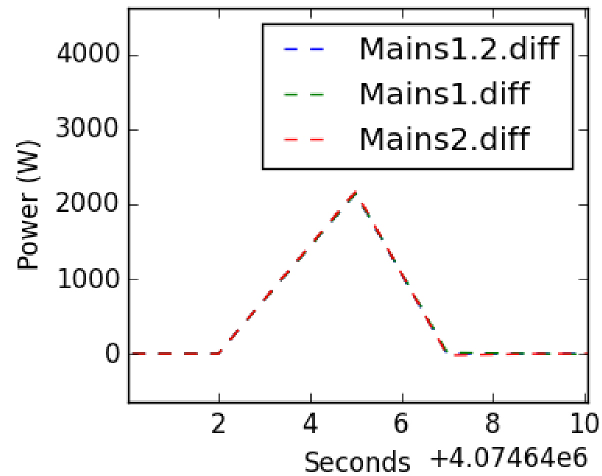
[19] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 262–270.

[20] Huijuan Shao, Manish Marwah, and Naren Ramakrishnan. 2013. A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation: Applications to Residential and Commercial Buildings. Bellevue, U.S.A.

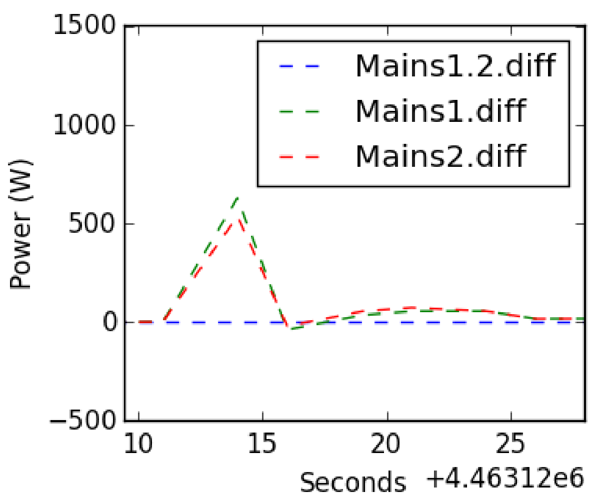
[21] Steven Robert Shaw. 2000. *System identification techniques and modeling for nonintrusive load diagnositics*. Ph.D. Dissertation.



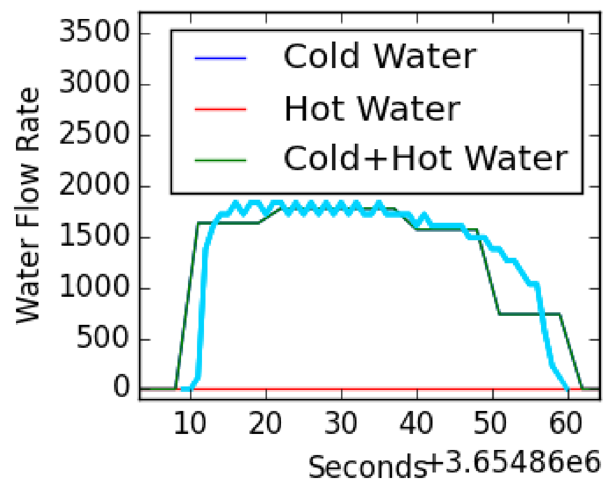
(a)



(b)



(c)



(d)

Figure 5: Disaggregating dryer and continuous variable load heatingOutdoor with multivariate motif mining. The on event of a device and the corresponding diffs in the two phases for (a) dryer, (b) water heater, (c) heatingOutdoor. (d) Disaggregating the toilet water use end with dynamic time-warping subsequence search. This Y-axis is water flow rate in 10000*liter/minute.

[22] D. Srinivasan, WS Ng, and AC Liew. 2006. Neural-network-based signature recognition for harmonic source identification. *IEEE Transactions on Power Delivery* (2006).

[23] Yoshiki Tanaka, Kazuhisa Iwamoto, and Kuniaki Uehara. 2005. Discovery of time-series motif from multi-dimensional data based on MDL principle. *Machine Learning* 58, 2-3 (2005), 269–300.

[24] Dragomir Yankov, Eamonn Keogh, Jose Medina, Bill Chiu, and Victor Zordan. 2007. Detecting time series motifs under uniform scaling. In *Proceedings of the 13th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*.