Energy disaggregation of overlapping home appliances consumptions using a cluster splitting approach

Aiad, Misbah; Lee, Peng Hin

2018

Aiad, M., & Lee, P. H. (2018). Energy disaggregation of overlapping home appliances consumptions using a cluster splitting approach. *Sustainable Cities and Society*, 43, 487-494. doi:10.1016/j.scs.2018.08.020

<https://hdl.handle.net/10356/137056>

<https://doi.org/10.1016/j.scs.2018.08.020>

© 2018 Elsevier Ltd. All rights reserved. This paper was published in *Sustainable Cities and Society* and is made available with permission of Elsevier Ltd.

Downloaded on 26 Apr 2023 19:21:54 SGT

Energy Disaggregation of Overlapping Home Appliances Consumptions Using a Cluster Splitting Approach

Abstract

Non-intrusive load monitoring (NILM) is a set of techniques that aims to decompose the aggregate energy consumptions of a household into the energy consumed by the respective individual appliances. When some of the home appliances have power consumptions levels that overlap with each other, it becomes a challenging problem to disaggregate the energy consumed by each of these appliances. In this work, we present an approach that split the clusters of the overlapping energy consumptions into the respective energy consumed by the individual appliances. The proposed approach involves firstly to analyze the cohesion of devices clusters to determine if a cluster should be split into two clusters. The proposed cluster splitting approach was tested on cases of overlapping devices clusters from six real houses available from the REDD public data sets. The results showed that the performance of the proposed approach depends on the degree of overlapping of the devices clusters, on whether the clusters are tight or loose and on the sizes of the clusters. The proposed approach can be applied to a clustering-based load disaggregation method as a subsequent step to deal with situations of overlapping appliances consumptions, so as to improve the overall energy disaggregation accuracy.

Keywords: non-intrusive load monitoring (NILM), energy disaggregation, energy savings, cluster splitting, normality tests.

1. Introduction

Energy savings and energy management in the residential sector are of increasing interests in the fields of energy systems and smart grids. In the broad scope of energy saving, there is a potential to achieve energy saving by replacing existing appliances with energy-efficient appliances. In addition, it is essential that the energy consumers be knowledgeable to adopt behaviors that lead to reduce their energy consumptions. In general, households' energy smart energy meters provide information about the total aggregate consumption of a household. Studies conducted have shown that energy savings of between 5% to 20% are achievable when the household occupants have information on the detailed breakdown of the energy consumed by each individual device in their homes [1]. This detailed information were found influencing to occupants to use appliances (especially those of high power consumptions) less often [1].

Non-intrusive load monitoring (NILM) is a set of techniques that aims to decompose the total aggregate energy consumptions of a household into the energy consumed by the respective individual appliances that are present in the household [2, 3, 4]. The NILM problem can be defined as follows

Given the total measured power consumption $p(t)$ at any time instant t , we can write

$$p(t) = \sum_{j=1}^N p_j(t) \quad (1)$$

The NILM problem is to estimate each $p_j(t)$, the individual power consumed by the device j at the time instant t , where $j = 1, 2, \dots, N$, $t = 1, 2, \dots, T$ and N is the total number of the operating appliances within the load disaggregation period T [2]. In NILM, the total aggregate measurements from smart meters are used by the disaggregation processes to detect the individual devices and infer their operations. The main objective of an energy disaggregation approach is often to achieve a high disaggregation accuracy for individual appliances. However, there are many challenges in tackling the NILM problem. In this paper, we focus on a specific challenge when two or more energy consumptions clusters overlap and it is necessary to split them into smaller distinct clusters to represent the respective individual appliances. The proposed approach can be applied to a *clustering-based* load disaggregation method as a subsequent step to deal with

situations of overlapping appliances consumptions, so as to improve the overall energy disaggregation accuracy. The proposed clusters splitting method in this paper is not a clustering method nor a NILM approach by itself, but an addition action to be used in *clustering-based* NILM approaches to deal with cases of overlapping clusters and decide if they should be split. The proposed approach refines the outcomes of the home devices clusters by investigating which clusters that may be needed to be split into smaller clusters and then splitting them when necessary.

2. Background and Related Work

The concept of NILM was first proposed by Hart in 1992 [5] where the proposed disaggregation process consisted of two steps. The first step was to capture the significant features whenever a transition in the aggregate signal is detected when a device switches state from the ON state or the OFF state. The primary features that were measured are the change in the real and the reactive power consumptions. The second step is to group these features, also called distinct signatures or attributes into distinct clusters of the different appliances. It was observed that higher power consuming devices are usually easier to be classified since they are placed far from other devices in the features space. There are may be overlapping devices clusters, especially those lower power consuming devices as shown in Figure 1 [5]. The cases of overlapping appliances clusters the basic challenge that we aim to tackle with the proposed clusters cohesion test in this work.

Fig. 1: Features space (real versus reactive power). From [5]. © 1992 IEEE.

Some work on NILM tried to obtain information from the aggregate measurements that may help in the improvement of the disaggregation accuracy (see, for example, [6]). Others have also obtained information, without the need of additional sensors to be installed, such as the time of use, duration of operations of the devices and outdoor weather information to aid in the disaggregation process. In general, extracting useful information and features from a data stream falls under the broad scope of data mining techniques as discussed in [30]. Besides the extracting of additional features, improving the disaggregation and identification algorithms is also essential to enhance the overall disaggregation accuracy. Hidden Markov models (HMMs) and their extensions were found functional in the representations and identifications of household appliances. Kim *et al.* [7] tested four extensions of the HMMs for better representations and

disaggregations. They also used non-electrical features such as the distributions of the ON/OFF durations and the inter-dependency between appliances. Kolter and Jaakkola [8] expressed a new approximate inference method based on the factorial HMM (FHMM). They used two complementary models, namely, the additive and the difference FHMM. The additive FHMM captures the total aggregate output signal while the difference FHMM encodes the signal differences between subsequent levels when a transition is detected in the aggregate total signal. Parson *et al.* [9] proposed an approach in which a one-time supervised learning process with an already labeled data set was used to create general probabilistic models of appliances. Thereafter, the general models are tuned to previously unseen appliances in an unsupervised manner. Gabaldon *et al.* [10] proposed a non-intrusive approach based on the Hilbert transform to identify the most suitable loads that can be used to validate customers' effective responses.

In the domain of machine learning and artificial intelligence, overlapping of clusters has been a challenge in many real-life applications including overlapping appliances clusters in NILM. Wagstaff *et al.* [11] proposed constraint-based clustering which modifies the *K-means* method [12] by considering a set of constraints to be satisfied. The *K-means* clustering is an iterative method of clustering based on achieving the minimum distances of the elements to their cluster center. Lu and Leen [13] used a mean field approximation to produce the *Penalized Probabilistic Clustering (PPC)* to handle increasing complexity in large data sets. In the PPC, a set of constraints are defined based on prior information. These constraints define item-to-item relations, for example, whether they must or must not ultimately belong to a specific cluster. Instead of using item-to-item constraints as in *PPC*, the *Class-Level Penalized Probabilistic Clustering (CPPC)* was proposed by defining cluster-to-cluster constraints [14]. These constraints provide the probabilities of how likely two clusters should be merged, split or kept unchanged. As the number of clusters is often significantly less than the total number of items to be classified, *CPPC* provides a noticeable reduction in complexity over the *PPC* [14]. The hierarchical divisive clustering is a top-to-bottom clustering that repeatedly partition a present cluster into two smaller clusters till reaching some stopping criteria. Common approaches in considering such splitting decisions Some of the common factors used to decide on whether to split or not to split are size priority, cluster cohesion tests and dissimilarities between the possible sub-clusters [15].

Clustering methods used in load disaggregation are usually efficient in identifying non-overlapping and higher power consuming devices [3]. In [16], the standard clustering approach was extended by filtering and smoothing mechanisms to deal with power variations. Nonetheless, this method requires excessive training and it was applicable mainly to high power loads [2].

3. Proposed Approach to Split Devices Clusters

We propose an approach that utilizes some prior information to check if a cluster should be split and perform the splitting, once a decision is made. We are not proposing a new clustering method to tackle the basic NILM problem of energy disaggregation. Such clustering approaches to the NILM problem in energy disaggregation can be found in, for example, [5, 16]. From the outcomes of such clustering approaches to the NILM problem or as Figure 1 has shown, it may result that there are overlapping appliances clusters and the proposed work in this paper seeks to address this. The proposed approach in clusters splitting aims to obtain the individual appliance clusters from the overlapping clusters to enhance the overall disaggregation of the energy consumed by all the respective individual appliances. The ultimate objective is to improve the overall disaggregation accuracy by investigating cases of devices clusters that may have originated from two different home devices. An important prior assumption here is the normality of the data distributions within each cluster. In the context of NILM, each cluster represents the power consumptions of an individual household appliance. This assumption applies mainly to ON/OFF devices with consistent power consumption and to finite state machines (FSMs). In our approach, we have avoided splitting methods that are based on size priority because the devices clusters sizes are usually unpredictable. The sizes of the devices clusters essentially depend on how long the device will be in operation, i.e. the ON

state. For more discussions on devices ON durations, see e.g. [7, 17]. Our proposed approach does not require a field expert to decide on clusters splitting in the context of an efficient NILM algorithm [18].

3.1. Normality of Appliances Power Consumptions

The assumption that a power consumption of an individual home appliance follows a normal distribution has been discussed and adopted in several previous studies. See e.g. [6, 7, 8, 9, 18, 19]. The adoption of a normality fitting of the power consumptions means that the fluctuations or variabilities of the power consumptions of a device can be reasonably modelled using a normal distribution. For cases of non-overlapping devices with consistent consumptions patterns, it is possible to model the devices as a Gaussian mixture model (GMM) [20]. In the proposed approach, we have considered the following:

- a) Only devices with stable single state or multi-state consumptions are considered in the study. Continuously varying consumptions devices such as power tools or light dimmers are excluded.
- b) Some devices may have an overshoot pattern for the consumptions when the states changes. In such cases, we do not include the patterns from the beginning to the end of operation period. Only periods of stable consumptions are considered so as to provide a better normality fitting. Samples in the period of overshoot were ignored and eliminated.
- c) There may be noisy signals or spikes which may alter the normality of a device power consumptions pattern. A 9-sample median filter to remove outliers and noisy signals was applied.

A typical home energy smart meter measures the power consumption of the entire household at a specific frequency (e.g. 1 sample/second). The measured real power values from the smart meter p_j , can be modeled by a one-dimensional normal distribution as follows

$$p_j \sim N(\mu, \sigma) \quad (2)$$

where μ is the mean value of the distribution and indicates the average power consumed by the appliance and σ is the standard deviation, indicating the power fluctuations. Kim *et al.* [7] found that the power consumptions of an appliance often fluctuates within $\pm 20\%$ from its average value. From the normal distribution principles, 99.7% of the data lies in the range between $\mu - 3\sigma$ to $\mu + 3\sigma$. Therefore, the power consumptions may be approximated to fluctuate with standard deviations as

$$\sigma \leq \frac{0.20\mu}{3} = 0.067\mu \quad (3)$$

However, the σ to μ ratio for some appliances may not always be consistent as seen in the results sections. It may be due to the different degrees of normality fitting across appliances and other factors such as noisy signals, effects from neighboring devices, non-linear loads, etc.

3.2. Clusters Cohesion Test to Decide on Splitting

With the assumption of normality of power consumed by the devices, three common normality tests are performed to investigate the cohesion of a cluster. The cohesion tests described here is a general framework that was introduced to specifically suit the clusters of home devices. In cases where a cluster shows a poor fitting to a normal distribution, it could indicate that the cluster is originating from two different sources, i.e. from two distinct devices. The degree of cohesion between cluster items is indicated generally by the degree of how many of these items follow some pattern or are being arranged in a specific manner. The three normality tests used in our approach are

- Anderson-Darling (AD) test [21].
- Kolmogorov-Smirnov (KS) test [22, 23].
- Jarque-Bera (JB) test [24].

These tests are used so as to encompass possible variabilities within the tested samples. For example, while the KS test inspects normality by testing the distance between the cumulative distribution function (CDF) of the reference distribution and the empirical distribution of the sample, the AD test gives more weight to the tails of the distribution. Furthermore, the JB test detects normality by checking the skewness and kurtosis of the tested sample [25]. Though the essence of these tests differs in general, they check the samples under study for normality by testing the following hypothesis:

- *Null hypothesis, H_0 : the sample follows a normal distribution.*
- *Alternate hypothesis, H_1 : the sample does not follow a normal distribution.*

The outcome of each test is usually given in terms of a P_value . Theoretically, a P_value can be in the range

$$0.00 \leq P_value \leq 1.00 \quad (4)$$

P_value is a statistical metric that gives an indication about the tested sample of data, where a larger P_value implies better normality fitting of the tested sample. Therefore, it can be used as a direct indicator of the degree of a cluster cohesion. We may, hence, consider that

$$Cohesion \approx P_value \quad (5)$$

A confidence level, denoted as $(1 - \alpha)$, then should be considered to give final decision about the hypothesis. Typical values for α are 0.01 and 0.05, and they are called the probability of type I error. Type I error indicates how likely we may commit an error by rejecting the null hypothesis when the null hypothesis is true. For more details on confidence levels and hypothesis testing, see e.g. [26, 27]. The conclusion of the hypothesis test is obtained as follows

$$Conclusion = \begin{cases} \text{Fail to reject } H_0, & P_value > \alpha \\ \text{Reject } H_0, & P_value \leq \alpha \end{cases}$$

In our approach, we proposed two steps in order to reach a final decision about splitting. Firstly, we will obtain three distinct P_values from the three normality tests. We then obtain the average of these P_values , denoted as $\overline{PV}$, in order that we could compare it to a specific confidence level while including possible variabilities from the three normality tests as explained earlier. Secondly, since home appliances differ in their consumptions profiles and the degree of fitting to normal distribution, we suggest that the decision on splitting is made by comparing against two confidence levels, α_1 and α_2 , as follows

$$Splitting = \begin{cases} \text{Split}, & \overline{PV} \leq \alpha_1 \\ \text{Test sub_clusters}, & \alpha_1 < \overline{PV} \leq \alpha_2 \\ \text{Do not split}, & \alpha_2 < \overline{PV} \end{cases}$$

The above decision criteria have considered the convention that α_1 is less than α_2 . There is no exact method for choosing the values of α_1 and α_2 but we suggest values that indicate reasonable normality fitting for the applied problem. For example, if we found that the cohesion test shows that the $\overline{PV}$ to be less than the pre-defined α_1 , it would mean that there's a high degree of non-normality within the cluster. Therefore, a decision to split is taken. On the other hand, if the cohesion test shows a $\overline{PV}$ that is larger than the pre-defined α_2 , then no splitting decision is taken as this imply a consistent cluster that most likely has originated from a single source.

In the case when $\alpha_1 < \overline{PV} \leq \alpha_2$, further normality tests on the possible sub-clusters are suggested to be performed before a decision is made. The testing of both the sub-clusters is carried out in three successive steps as follows:

1. Perform the splitting as explained in the subsection 3.3.
2. Reconstruct the sub-clusters.
3. Apply the cohesion tests.

We start with performing the splitting to investigate if any of the smaller clusters can show better normality than the original big cluster. It is important to note that the splitting in step 1 is not final until results from step 3 are obtained. However, the splitting will often make a sharp cut in the assumed normal distributions as shown in Figure 2. Therefore, we suggest to carry out step 2, where the normality tests are carried out on the reconstructed small clusters. A reconstructed cluster consists of the split cluster merged to a portion placed in the cut region that is similar to the other side of the cluster as shown in Figure 2.

For illustrations on how to obtain the reconstructed portions, we consider the clusters X and Y as shown in Figure 2. The reconstructed portion of cluster X may be seen as the items group x obtained from cluster X where each item x_i satisfies

$$\min(X) \leq x_i \leq \mu_X - |\max(X) - \mu_X| \quad (6)$$

where $\min(X)$, μ_X and $\max(X)$ are the minimum item value in cluster X , the mean of cluster X and the maximum item value in cluster X , respectively. Similarly, the reconstructed portion of cluster Y may be seen as the items group y obtained from cluster Y where each item y_i satisfies

$$\mu_Y + |\mu_Y - \min(Y)| \leq y_i \leq \max(Y) \quad (7)$$

where $\min(Y)$, μ_Y and $\max(Y)$ are the minimum item value in cluster Y , the mean of cluster Y and the maximum item value in cluster Y , respectively. These portions of the cluster are then copied in the *inverse order* and appended to the other side of the corresponding cluster so as to obtain the reconstructed version of the clusters.

The notion in this case is that a sub-cluster with a sharp cut in one of its sides is more likely to show poor normality fitting. This poor normality fitting is not due to how the data is naturally distributed but it is actually due to the sharp splitting. That is why we propose to perform normality tests on a reconstructed version of these sub-clusters. The reconstructed portion of these sub-clusters is taken typically to be the same as the corresponding present portion on the other side of the same sub-cluster. In this way, the normality tests results could be more reasonable and thus provide better decision criteria.

Fig. 2: Sharp cuts in sub-clusters and reconstructed sub-clusters.

In step 3, we obtained the $\overline{PV}$'s of both the sub-clusters separately. We can then observe if any of these $\overline{PV}$'s is larger than the $\overline{PV}$ of the original big cluster. In cases that any $\overline{PV}$ of the sub-clusters is larger than that of the original big clusters, it means that the sub-cluster shows a better normality fitting than the big cluster. Hence, each sub-cluster could be originating from a different source and a splitting decision should be taken. If the converse happens, no splitting decision should be taken.

3.3. Performing the Splitting

Once a decision to split the big cluster into two smaller clusters is taken, our approach is to perform an inner *EM* which will assign the big cluster elements into the two small clusters. The *EM* is carried out in two steps. Firstly, we propose to initialize the *EM* with two centroids C_1 and C_2 to be at 2σ distance from the mean value of the big cluster. That is

$$C_1 = \mu + 2\sigma \quad (8)$$

$$C_2 = \mu - 2\sigma \quad (9)$$

where μ and σ are the mean and the standard deviation of the big cluster, respectively. Then, we initialize each small cluster by assigning elements according to which distance from the centroids is the minimum. Secondly, the *EM* is carried out iteratively to obtain the final versions of the two split clusters. The iterations of a typical *EM* consist of two operations: expectation and maximization. In the expectation process, the centers of clusters are updated by using the expectations of all the assigned elements to the corresponding cluster. In the maximization process, the centers of the clusters are kept fixed and the elements are then assigned to clusters based on the maximum likelihood for an element to belong to either clusters.

3.4. Assessment of Splitting Accuracy

The correct assignment of elements to the resulting small clusters is an indication of the accuracy of the splitting methodology. The assignment task is similar to a binary decision as we only need to assign each element to one of the two split clusters. Hence, we can use the *Precision*, *Recall* and *F-measure* [28] to assess the accuracy of elements assignment after the splitting. These accuracy metrics had been used to assess the performance of NILM techniques as in [7, 8, 18]. Table 1 shows the four possible situations that can happen in the assignment of an element to a specific cluster. In Table 1, a True Positive (*TP*) means a correct assignment and a False Positive (*FP*) means an incorrect assignment. A True Negative (*TN*) means a correct non-assignment and a False Negative (*FN*) means an incorrect non-assignment, i.e. a missed item.

Table 1: Actual versus concluded assignments for a specific cluster.

<i>EM</i> conclusion	The element really belongs to the cluster	The element really does not belong to the cluster
Assigned	True Positive (<i>TP</i>)	False Positive (<i>FP</i>)
Not assigned	False Negative (<i>FN</i>)	True Negative (<i>TN</i>)

Precision is a measure that describes the relevance between the assigned elements of a specific cluster (percentage of truly assigned elements from all assigned elements) and it is given by

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Recall is a measure of how many of the elements that actually belong to the cluster were assigned or retrieved correctly and it is given by

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

In the context of accuracy of an energy disaggregation approach, the measure *Precision* can be considered to give a measure of the relevance between the assigned energy. That is, the portion of assigned energy that truly belongs to a device from the whole assigned energy to that device. The measure *Recall* can be considered as to give a measure of energy retrieving accuracy. That is, the portion assigned energy that truly belongs to a device from the whole energy that actually belongs to the device (either assigned or missed) [8]. The *F-measure* is a measure that combines the precision and recall measures as the harmonic mean. It is given by

$$F = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (12)$$

An overall accurate assignment is obtained when both assignments (positive decisions) and non-assignments (negative decisions) are accurate. As an illustration, for a specific cluster X , we need to be accurate when an item is assigned to X and we also need to be accurate when the converse is decided. In cases where an item really belongs to X and it was not assigned to X (false negative decision), it results in a missed item from the cluster X and also in a wrong element that was assigned (wrong positive) to another cluster (say cluster Y). Hence, the overall accuracy of retrieving both the clusters X and Y is decreased. Note that each element must finally be assigned to one cluster only. The accuracy of retrieving each cluster is assessed separately from other clusters. Overall, the F-measure can be used as an indication of the accuracy of retrieving the individual clusters using the tested approach.

4. Results

We applied our proposed approach to overlapping power consumptions clusters of appliances data of six real houses available from the Reference Energy Disaggregation Data Set (REDD) [29]. The REDD provides recorded real data of individual household appliances and the total aggregate power consumptions. The main goal of public data sets is to expedite the research in energy disaggregation and to provide a reference benchmark to compare different disaggregation approaches. In this work, we focused only on cases where there are overlapping of power consumptions between appliances clusters. This usually occurs when two appliances or more have close values of power consumptions at some states of consistent operations. We first searched for possible cases of overlapping between appliances in each house separately. A 9-samples median filter is then applied to filter out the noisy signals and outliers. Table 2 through Table 7 show the possible overlapping appliances power consumptions clusters that are found in house 1 to house 6 from the REDD data. Information on clusters fitting to normal distribution with the respective mean and standard deviation of power consumptions in Table 2 through Table 7 were obtained directly from the sub-metered data of the REDD data set.

Table 2: Overlapping clusters in house 1 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Oven	$N(1657.3, 11.6)$	Cluster 2
Cluster 2	Oven	$N(1687.3, 20.4)$	Cluster 1
Cluster 3	Microwave	$N(1506.9, 43.6)$	Cluster 4, cluster 5
Cluster 4	Bathroom gfi	$N(1597.4, 12.1)$	Cluster 3
Cluster 5	Kitchen outlet	$N(1529.9, 7.7)$	Cluster 3

Table 3: Overlapping clusters in house 2 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Lighting	$N(155.7, 7.5)$	Cluster 2
Cluster 2	Refrigerator	$N(163.0, 6.7)$	Cluster 1

Table 4: Overlapping clusters in house 3 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Electronics	$N(155.7, 7.5)$	Cluster 2, cluster 6
Cluster 2	Refrigerator	$N(98.8, 9.5)$	Cluster 1, cluster 6
Cluster 3	Lighting	$N(181.7, 18.3)$	Cluster 4, cluster 5
Cluster 4	Lighting	$N(189.9, 5.9)$	Cluster 3, cluster 5
Cluster 5	Lighting	$N(196.9, 11.8)$	Cluster 3, cluster 4
Cluster 6	Kitchen outlet	$N(127, 0.7)$	Cluster 1, cluster 2

Table 5: Overlapping clusters in house 4 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Kitchen outlet	$N(137.1, 6.7)$	Cluster 2
Cluster 2	Lighting	$N(101.5, 15.5)$	Cluster 1, cluster 3
Cluster 3	Lighting	$N(93.9, 4.9)$	Cluster 2

Table 6: Overlapping clusters in house 5 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Unknown outlet	$N(169.3, 7.6)$	Cluster 2
Cluster 2	Refrigerator	$N(164.2, 6.2)$	Cluster 1
Cluster 3	Lighting	$N(576.7, 17.1)$	Cluster 4
Cluster 4	Furnace	$N(553.4, 12.5)$	Cluster 3
Cluster 5	Sub-panel	$N(226.7, 4.2)$	Cluster 6
Cluster 6	Sub-panel	$N(216.8, 6.0)$	Cluster 5
Cluster 7	Electric heat	$N(808.7, 19.5)$	Cluster 8
Cluster 8	Electric heat	$N(799.7, 19.0)$	Cluster 7
Cluster 9	Lighting	$N(112.8, 2.7)$	Cluster 10
Cluster 10	Lighting	$N(116.8, 1.3)$	Cluster 9

Table 7: Overlapping clusters in house 6 from the REDD.

Cluster number	Appliance type	Power consumption $\sim N(\mu, \sigma)$ (Watts)	Overlaps with
Cluster 1	Unknown outlet	$N(96.6, 13.0)$	Cluster 2
Cluster 2	Lighting	$N(121.3, 8.8)$	Cluster 1
Cluster 3	Electric heat	$N(463.3, 18.4)$	Cluster 4
Cluster 4	Air conditioner	$N(416.5, 6.8)$	Cluster 3
Cluster 5	Bathroom gfi	$N(945.9, 3.4)$	Cluster 6, cluster 7
Cluster 6	Air conditioner	$N(985.2, 31.0)$	Cluster 5, cluster 7
Cluster 7	Air conditioner	$N(973.5, 31.2)$	Cluster 5, cluster 6

Each cluster in the above tables represents a single stable state for a specific appliance type as shown. The appliance itself may have more than one state of operations, e.g. a finite state machine (FSM), but only the states where overlapping is possible were considered. In addition, each cluster comes from a specific appliance even if they have the same appliance type. For example, cluster 1 and cluster 2 in house 1 come from two independent ovens and so on. Some clusters are defined as outlets or unknown outlets where there is no more information from the REDD about the exact appliance plugged to those outlets. These overlapping clusters may represent typical outcomes of a NILM approach and can be tested by the proposed clusters cohesion test.

The proposed approach is tested on each overlapping cluster. For the proposed splitting decision criteria discussed in subsection 3.2, the two confidence levels $\alpha_1 = 0.01$ and $\alpha_2 = 0.10$ are used to provide reasonable indications about clusters cohesion as clarified in the discussion section. Table 8 through Table 13 show the accuracies of the splitting obtained in terms of the *Precision*, *Recall* and *F-measures* for all the tested overlapping clusters cases in house 1 through house 6, respectively.

Table 8: Accuracy of clusters splitting in house 1 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1	94.5%	79.1%	86.1%
Cluster 2	74.3%	92.9%	82.6%
Cluster 3 (versus cluster 4)	100%	72.2%	83.8%
Cluster 4	82.4%	100%	90.3%
Cluster 3 (versus cluster 5)	100%	51.1%	67.7%
Cluster 5	68.9%	100%	81.6%

Table 9: Accuracy of clusters splitting in house 2 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1	96.2%	59.8%	73.8%
Cluster 2	27.4%	86.4%	41.6%

Table 10: Accuracy of clusters splitting in house 3 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1 (versus cluster 2)	100%	95.7%	97.8%
Cluster 2 (versus cluster 1)	90.4%	100%	95.0%
Cluster 3 (versus cluster 4)	59.9%	66.9%	63.2%
Cluster 4 (versus cluster 3)	75.3%	69.2%	72.1%
Cluster 3 (versus cluster 5)	80.7%	66.9%	73.2%
Cluster 5 (versus cluster 3)	39.4%	57.4%	46.8%
Cluster 4 (versus cluster 5)	76.3%	29.8%	42.9%
Cluster 5 (versus cluster 4)	19.1%	64.2%	29.4%
Cluster 6 (versus cluster 1)	99.3%	100%	99.6%
Cluster 1 (versus cluster 6)	100%	96.8%	98.4%
Cluster 6 (versus cluster 2)	99.3%	100%	99.6%
Cluster 2 (versus cluster 6)	100%	96.8%	98.4%

Table 11: Accuracy of clusters splitting in house 4 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1	75.5%	100%	86.1%
Cluster 2 (versus cluster 1)	100%	75.9%	86.3%
Cluster 3	63.0%	99.6%	77.2%
Cluster 2 (versus cluster 3)	99.3%	45.6%	62.5%

Table 12: Accuracy of clusters splitting in house 5 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1	80.3%	57.5%	67.0%
Cluster 2	54.8%	78.5%	64.5%
Cluster 3	98.6%	23.3%	37.7%
Cluster 4	14.6%	97.6%	25.4%
Cluster 5	99.5%	70.9%	82.8%
Cluster 6	66.4%	99.4%	79.6%
Cluster 7	60.6%	54.7%	57.5%
Cluster 8	58.6%	64.2%	61.3%
Cluster 9	3.4%	89.4%	6.5%
Cluster 10	99.6%	49.4%	66.0%

Table 13: Accuracy of clusters splitting in house 6 from the REDD.

Cluster number	Precision	Recall	F-measure
Cluster 1	27.2%	99.9%	42.7%
Cluster 2	100%	38.4%	55.5%
Cluster 3	89.7%	95.1%	92.3%
Cluster 4	99.3%	98.5%	98.9%
Cluster 5 (versus cluster 6)	8.6%	100%	15.8%
Cluster 6 (versus cluster 5)	100%	48.9%	65.7%
Cluster 5 (versus cluster 7)	8.4%	100%	15.5%
Cluster 7 (versus cluster 5)	100%	47.8%	64.7%
Cluster 6 (versus cluster 7)	59.1%	53.3%	56.0%
Cluster 7 (versus cluster 6)	57.5%	63.1%	60.1%

When a cluster overlaps with more than one cluster, the proposed approach is applied separately on each pair of overlapping clusters. As an illustration on the placement of devices clusters, consider an example of devices clusters in house 1 as shown in Figure 3, which gives indications about the degree of overlapping between clusters. It showed whether a cluster is tight or loose in terms of the spread of the items in the cluster. For example, cluster 4 and cluster 5 are tight but cluster 3 is loose. Cluster 3 and cluster 4 overlap in quite a small degree but cluster 5 totally overlaps with cluster 3.

As an illustration on the cohesion tests and $\overline{PV}$'s of the sub-clusters, consider the case of cluster 3 and cluster 4 in house 1. The normality tests on the big cluster give a $\overline{PV}$ of approximately 0.00, which means that a splitting is required. After performing the splitting, the normality of the sub-cluster 3 was found to be better than the big cluster with $\overline{PV} = 0.033$. The $\overline{PV}$ of sub-cluster 4 has not changed significantly from that of the big cluster. Once either sub-clusters show better fitting for normality than the big cluster, it indicates such sub-clusters are originating from two independent sources as discussed in subsection 3.2.

Fig. 3: Overlapping devices clusters in house 1 from the REDD.

5. Discussions

The results in the previous section showed that the accuracy of retrieving the split devices clusters from the merged one, in general, depends on the degree of overlapping between the clusters, the tightness or looseness of each small cluster and the total number of items in each cluster. The distance between the centers of two overlapping clusters indicates the degree of overlapping of the clusters, while the standard deviation indicates whether a cluster is tight (items are close to each other) or loose (items spread over wide ranges). For the splitting decision proposed in subsection 3.2, we have used the two confidence levels $\alpha_1 = 0.01$ and $\alpha_2 = 0.10$ to provide reasonable indications about clusters cohesion. The value of α_1 was chosen as small as 0.01 so as to decide a splitting once normality or cohesion of a cluster is deduced to be very poor. If $\overline{PV}$ falls in the range between α_1 and α_2 , further normality tests on the possible sub-clusters are suggested to be performed before a decision is made as explained in subsection 3.2. Testing normality fitting of the potential sub-clusters is performed in order to investigate if the sub-clusters may show better normality than the overlapped merged cluster. The value of α_2 should not be chosen too much higher than α_1 as this may more likely leads to errors in the normality test of the cluster.

As seen in Table 8 for house 1, high accuracies were obtained for device cluster 4 and lower accuracies for device cluster 3 (tested versus cluster 5). Cluster 4 showed a high accuracy as its center is quite far from cluster 3 and its items are close to each other (a tight cluster) as seen in Figure 3. Conversely, cluster 3 (versus cluster 5) showed a low accuracy mainly because cluster 5 as a whole is overlapping with it as can be seen in Figure 3. This implies that several items which originally belong to cluster 3 will mostly be assigned to cluster 5. There is an interesting observation that if it happened that there is a 100% in the *Precision* measure for one cluster, the other overlapping cluster should have 100% in the *Recall* measure. This is logical as a cluster retrieved with 100% in the *Precision* measure means that all the items assigned to it do belong to this cluster in reality. In other words, there is a zero element under the category of False Positive (*FP*) decisions. On the other hand, a zero element in the *FP* category for one cluster means a zero element in False Negative (*FN*) category for the other cluster and this imply a 100% in the *Recall* measure. To further illustrate this fact, consider the case of house 1, with the same reasoning applicable to other houses for similar cases. In house 1, cluster 3 has a 100% in the *Precision* measure while clusters 4 and 5 have 100% in the *Recall* measure. From Figure 3, it can be observed that after performing the splitting, all the assigned items to cluster 3 will most likely belong to cluster 3 in reality due to its looseness and the tightness of the other overlapping clusters. However, several items from cluster 3 could be assigned to cluster 4 or cluster 5. This interprets the low *Precision* measure for both cluster 4 and cluster 5. On the other hand, since cluster 4 and cluster 5 are tight, it is more likely that after splitting, all the items that really belong to these clusters would be assigned to the corresponding cluster. This explains the 100% in the *Recall* measure as there no item is missed from the assignment, i.e. zero items in the *FN* category. Such cases may also have happened in the other houses for similar reasons.

When both the overlapping-devices clusters are tight and there is a noticeable distance between their centers, it is more likely to obtain high accuracies in retrieving both of them. Cluster 1 and cluster 2 in house 3 is a good example illustrating this case. A tight cluster is usually known as a cluster with majority of its items are close (to some extent) to its center. That is, items are not spread far from the cluster center as in the opposite case of a loose cluster. The contrast example can be seen from the case of cluster 4 and cluster 5 in house 3, which results in low splitting accuracies. There are some cases when two clusters centers are very close to each other, and the obtained accuracies are very low as seen in the case of cluster 9 in house 5. Besides, clusters with limited number of items may show lower accuracy results as in the case of cluster 4 and cluster 9 in house 5. These clusters have limited number of items since the corresponding appliances were operating for less time durations than other appliances. Therefore, a cluster with fewer items is more likely to be less representative for the actual power consumption model of that specific device. In addition, an error in assigning a small number of items would result in noticeable reductions in the obtained accuracies.

Another interesting case is when two overlapping clusters represent two identical appliances in a given house. From the REDD data set, we found that cluster 6 and cluster 7 in house 6 represent two identical air conditioners. However, each appliance showed a slightly different amount of fluctuation in power consumption, which could be due to uneven noise effects. In such cases, it is reasonable to obtain the *Precisions* and *Recalls* measures at about 50% for both clusters. This can be due to the splitting approach had resulted in somehow two similar small clusters. Each of these small clusters consists of approximately a half of its original items and a half from the other cluster items. Therefore, the number of items under the categories of *TP*, *FP* and *FN* are approximately close to each other. This clarified the obtained accuracies of about 50%.

As far as we know, there is no cohesion testing method that has been applied to the same REDD data set for the purpose of dealing with overlapping consumptions of the devices in an unsupervised fashion. For this reason, we were unable to compare our results to benchmark reference methods. Clusters cohesion tests in the literature were applied in different fields but the nature of the problem (i.e. measurable features, their dimensions, fitting distributions, etc.) are basically dependent on the cases under study.

The approach proposed in this paper is to deal with cases of overlapping appliances clusters and is suggested to work together with NILM cluster-based load disaggregation methods. The NILM cluster-based disaggregation method is first applied for load disaggregation and the proposed approach may then be used to split the overlapped devices clusters, if they exist, contributing to the overall improvement in the non-intrusive loads disaggregation accuracy.

6. Conclusions

An approach for clusters splitting is proposed to be applied by clustering-based NILM approaches. The proposed method is based on the testing of clusters cohesion that is concluded from their degree of normality fitting. The proposed approach was tested on overlapping devices clusters from six houses from the REDD public data set. States where two appliances may overlap in power consumptions were tested. The splitting was carried out by performing an inner expectation maximization (*EM*) for the small clusters. The accuracy of splitting, i.e. assignment of elements to the small clusters, was found to depend on the degree of overlapping, on whether the clusters are tight or loose, and the total number of items in each cluster. We are not proposing a new clustering method to tackle the basic NILM problem of energy disaggregation. Such clustering approaches to the NILM problem in energy disaggregation can be found in, for example, [5, 16]. From the outcomes of such clustering approaches to the NILM problem, it may result that there are overlapping appliances clusters and the proposed work in this paper seeks to address this. The proposed approach in clusters splitting aims to obtain the individual appliance clusters from the overlapping clusters to enhance the overall disaggregation of the energy consumed by all the respective individual home appliances.

References

- [1] Henao N, Agbossou K, Kelouwani S, Dubé Y, Fournier M (2017) Approach in Nonintrusive Type I Load Monitoring Using Subtractive Clustering. *IEEE Transactions on Smart Grid* 8 (2): 812-821.
- [2] Zoha A, Gluhak A, Imran M, Rajasegara S (2012) Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. *Sensors* 12: 16838-16866.
- [3] Zeifman M, Roth K (2011) Nonintrusive appliance load monitoring: review and outlook. *IEEE Transactions on Consumer Electronics* 57(1): 76-84.
- [4] Aiad M, Lee PH (2016) Unsupervised approach for load disaggregation with devices interactions. *Energy and Buildings* 116: 96-103.
- [5] Hart G (1992) Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 12(80): 1870-1891.
- [6] Aiad M, Lee PH (2016) Non-intrusive load disaggregation with adaptive estimations of devices main power effects and two-way interactions. *Energy and Buildings* 130: 131-139.
- [7] Kim H, Marwah M, Arlitt M, Lyon G, Han J (2010) Unsupervised disaggregation of low frequency power measurements. *SIAM International Conference on Data Mining (SDM)*: 747-758.

- [8] Kolter J, Jaakkola T (2012) Approximate inference in additive factorial HMMs with application to energy disaggregation. *International Conference on Artificial Intelligence and Statistics (AISTATS)* 1472-1482.
- [9] Parson O, Ghosh S, Weal M, Roger A (2014) An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence* 217: 1-19.
- [10] Gabaldón A, Molina R, Marín-Parra A, Valero-Verdu S, Alvarez C, (2017) Residential end-uses disaggregation and demand response evaluation using integral transforms. *J. Mod. Power Syst. Clean Energy* 5(1): 91-104.
- [11] Wagstaff K, Cardie C, Schroedl S (2001) Constrained K-means clustering with background knowledge. *Proceedings of the Eighteenth International Conference on Machine Learning*: 577-584.
- [12] Bishop C (2006) *Pattern Recognition and Machine Learning*. Springer-Verlag New York Inc. Secaucus, NJ, USA.
- [13] Lu Z, Leen TK (2007) Penalized probabilistic clustering. *Neural Computation* 19(6): 1528-1567.
- [14] Preston D, Brodley C, Khardon R, Sulla-Menashe D, Friedl M (2010) Redefining class definitions using constraint-based clustering: an application to remote sensing of the earth's surface. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*: 823-832.
- [15] Ding C, He X (2002) Cluster merging and splitting in hierarchical clustering algorithms. *Proceedings of the IEEE International Conference on Data Mining*: 139-146.
- [16] Farinaccio L, Zmeureanu R (1999) Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings* 30: 245-259.
- [17] Johnson MJ, Willsky AS (2013) Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research* 14: 673-701.
- [18] Parson O (2014) Unsupervised training methods for non-intrusive appliance load monitoring from smart meter data. PhD thesis, University of Southampton, UK.
- [19] Xie J, Hong T, Laing T, Kang C (2016) On normality assumption in residual simulation for probabilistic load forecasting. *IEEE Trans. Smart Grid*. DOI: 10.1109/TSG.2015.2447007
- [20] Benaglia T, Chauveau D, Hunter DR, Young DS (2010) mixtools: An R Package for Analyzing Mixture Models. *Journal of Statistical Software* 32(6): 1-29.
- [21] Anderson TW, Darling DA (1954) A test of goodness of fit. *Journal of American Statistical Association* 49(268): 765-769.
- [22] Massey FJ (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* 46(253): 68-78.
- [23] Marsaglia G, Tsang W, Wang J (2003) Evaluating Kolmogorov's Distribution. *Journal of Statistical Software* 8(18): 1-4. DOI: 10.18637/jss.v008.i18.
- [24] Jarque CM, Bera AK (1987) A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 55(2): 163-172.
- [25] Razali NM, Wah YB (2011) Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics* 2(1): 21-33.
- [26] Montgomery D, Runger G (2011) *Applied statistics and probability for engineers*. Fifth edition, John Wiley and Sons Inc.
- [27] Montgomery D (2001) *Design and analysis of experiments*. Fifth edition. John Wiley and Sons Inc.
- [28] Powers DM (2011) Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies* 2(1): 37-63.
- [29] Kolter J, Johnson M (2011) REDD: A public data set for energy disaggregation research. *SustKDD*, San Diego, CA, USA.
- [30] Yu ZJ, Haghighat F and Fung BC (2016) Advances and challenges in building engineering and data mining applications for energy-efficient communities. *Sustainable Cities and Society* (25): 33-38.