
This version is available at https://strathprints.strath.ac.uk/55062/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk
Blind Non-intrusive Appliance Load Monitoring using Graph-based Signal Processing

Bochao Zhao, Lina Stankovic, and Vladimir Stankovic
Department of Electronic and Electrical Engineering University of Strathclyde, Glasgow, G1 1XW, UK
Email: {b.zhao, lina.stankovic, vladimir.stankovic}@strath.ac.uk.

Abstract—With ongoing massive smart energy metering deployments, disaggregation of household’s total energy consumption down to individual appliances using purely software tools, aka. non-intrusive appliance load monitoring (NALM), has generated increased interest. However, despite the fact that NALM was proposed over 30 years ago, there are still many open challenges. Indeed, the majority of approaches require training and are sensitive to appliance changes requiring regular re-training. In this paper, we tackle this challenge by proposing a “blind” NALM approach that does not require any training. The main idea is to build upon an emerging field of graph-based signal processing to perform adaptive thresholding, signal clustering and feature matching. Using two datasets of active power measurements with 1min and 8sec resolution, we demonstrate the effectiveness of the proposed method using a state-of-the-art NALM approaches as benchmarks.

Keywords—Non-intrusive appliance load monitoring, load disaggregation, graph-based signal processing

I. INTRODUCTION

Energy disaggregation, that is, identifying the individual appliance usage from the total, aggregate load, using purely data analytics, referred to as non-intrusive appliance load monitoring (NALM) [1], can significantly enrich energy feedback. Indeed, up to 20% of reduction in energy consumption is expected via appliance-feedback and specific appliance replacement programs [2]. NALM can support home automation, enable activity recognition [3]-[5] and help appliance retrofit decisions. Energy suppliers can better forecast demand, system operators can monitor the effect of smart grid fluctuations on the residential microgrid, and appliance manufacturers can optimise product design to meet customer usage patterns.

NALM can benefit the wider public if it can operate purely on measurements that are already collected for metering and billing purposes. Hence, low-complexity techniques that work with active power only at resolutions of seconds and minutes [6] are of special interest and could be used with a wide range of current off-the-shelf smart metering devices. However, a downside of current NALM approaches for low-rate data (see [7], [8], [9] literature surveys) is their requirement for long training periods and substantial manual effort. To address the challenge of practicality and consumer/industry acceptance, NALM research has focused on unsupervised approaches, that do not require labeled sets for training [10]. These approaches are usually based on hierarchical clustering or different variants of Hidden Markov Models (HMM)s, such as [11], [12], [13], [14].

Although the above state-based approaches can be unsupervised, they use expert knowledge to set a-priori values for each appliance state, and require a training set (usually where appliance operation does not overlap) to build/refine the state models. Since appliance models differ widely between households, using a training set from one house to disaggregate the other has been shown to be ineffective [15], [16]. Unsupervised, time-series approaches, such as [9], either suffer from similar problems or require both active and reactive power and high sampling rates [17].

In this paper, we propose a new, blind, low-rate NALM approach that does not require any training. The approach disaggregates any aggregate active power dataset without any prior knowledge. It relies on graph-based signal processing (GSP) [18]. GSP offers an alternative to conventional signal processing and data mining approaches by embedding the structure of signals onto a graph, leading to a powerful scalable and flexible approach suitable to many data mining and signal processing problems, ranging from image denoising and data compression, to classification, biomedical, and environmental data processing (see [18]-[22] and references therein). We note that recently, the latter two authors of this paper propose a GSP-based NALM approach in [23]. However, the approach of [23] is supervised and employs GSP only for data classification based on [20]. The proposed algorithm operates purely on collected aggregate active power data, and uses GSP three times, first to find an optimal event threshold, then to perform clustering, and finally for feature matching. The approach is deterministic and relies only on time-series data without any manual input, except for the final step of testing when unlabelled disaggregated appliances are labelled through expert knowledge.

II. GRAPH-BASED SIGNAL PROCESSING (GSP)

GSP is based on graph signals obtained by indexing a dataset by nodes of a graph. The basic idea is to represent a dataset using a graph defined by a set of nodes and a weighted adjacency matrix. Each node in the graph corresponds to an element in the dataset while the adjacency matrix defines all edges in the graph and their weights, where assigned weights reflect the degree of similarity, i.e., correlation, between the nodes.
Given an acquired set of measurements \( \mathbf{x} \), we define a graph \( G = \{ \mathbf{V}, \mathbf{A} \} \), where a set of nodes \( \mathbf{V} \) in the graph corresponds to the acquired measurements, and edges are defined by a weighted adjacency matrix \( \mathbf{A} \). The graph signal is then defined as a map from a set of nodes \( \mathbf{V} \) to a set of complex numbers \( \mathbf{s} \), called graph signal, where each element \( s_i \) is indexed by a node \( v_i \in \mathbf{V} \). The adjacency matrix \( \mathbf{A} \) defines all edges in the graph and their weights. The values of \( A_{i,j} \) are often naturally defined by the physical meaning of the collected data. If that is not the case, a Gaussian kernel is often used in the literature \([20],[21],[23]\), where

\[
A_{i,j} = \exp\left(-\frac{\text{dist}(x_i,x_j)}{\sigma^2}\right), \quad (1)
\]

is often used in the literature \([20],[21],[23]\), where \( \text{dist}(x,y) \) can be, for example, Euclidian distance between \( x \) and \( y \). The graphs and signals on graphs defined above can be conveniently used to represent very different data structures, such as time series, images, sensors, tracked objects, social networks, hyperlinked documents etc. \([19],[18]\).

If \( \mathbf{x} \) is piecewise smooth in a defined feature space, then \( \mathbf{s}^T \mathbf{L} \mathbf{s} \) is generally small, where \( \mathbf{L} \) is the graph Laplacian operator \([19]\) given by \( \mathbf{D} - \mathbf{A} \), where \( \mathbf{D} \) is a diagonal matrix with nonzero entries \( D_{i,i} = \sum_j A_{i,j} \). This graph-signal smoothness prior can effectively be used for regularization, since it can be shown that the Laplacian regularizer \( \mathbf{s}^T \mathbf{L} \mathbf{s} \) is a good measure of variation in the signal modulated by weights in \( \mathbf{A} \). Then, we can formulate the smoothness term minimization problem as:

\[
\arg\min_{\mathbf{s}} ||\mathbf{s}^T \mathbf{L} \mathbf{s}||_2^2. \quad (2)
\]

If \( \mathbf{s} \) is an \( N \times 1 \) length discrete signal, then \( \mathbf{L} \) is an \( N \times N \) matrix and \([21],[23]\):

\[
\mathbf{s}^T \mathbf{L} \mathbf{s} = s(1)\mathbf{L}(1,1)s(1) + s(1)\mathbf{L}(1,2) : \mathbf{N})s(2 : \mathbf{N}) + \\
s(2 : \mathbf{N})^T \mathbf{L}(2 : \mathbf{N},1)s(1) + \\
s(2 : \mathbf{N})^T \mathbf{L}(2 : \mathbf{N},2 : \mathbf{N})s(2 : \mathbf{N}), \quad (3)
\]

where \( s(1) \) is the first element in \( \mathbf{s} \), \( s(a : b) = [s(a), s(a + 1), \ldots, s(b)] \), \( a < b \), and similarly for matrix \( \mathbf{L} \). Note that (3) is the same as in \([20],[23]\), except that we replaced a vector of known samples used for training in supervised classification approaches with a randomly picked sample \( s(1) \). Since \( \mathbf{D} \) is a diagonal matrix, \( \mathbf{L} \) is also diagonally symmetric. Thus, since the first term in (3) does not affect minimization, we simplify minimization in (2) to:

\[
\arg\min_{\mathbf{s}} ||\mathbf{s}^T \mathbf{L} \mathbf{s}||_2^2 = \arg\min_{\mathbf{s}} 2s(2 : \mathbf{N})^T \mathbf{L}(2 : \mathbf{N},1)s(1) + \\
s(2 : \mathbf{N})^T \mathbf{L}(2 : \mathbf{N},2 : \mathbf{N})s(2 : \mathbf{N}), \quad (4)
\]

As an unconstrained quadratic programming problem, this minimization has a closed form solution \([21],[24]\):

\[
\mathbf{s}^* = \mathbf{L}(2 : \mathbf{N},2 : \mathbf{N})^\#(s(1))\mathbf{L}(1,2 : \mathbf{N})^T, \quad (5)
\]

where \((\cdot)^\#\) denotes the pseudo-inverse matrix. \( \mathbf{s}^* \) is the smoothness optimization solution, i.e., a solution that minimizes the total graph variation.

### III. Proposed Disaggregation Algorithms

The proposed algorithm consists of filtering, clustering and feature matching steps as shown in the flow chart in Fig. 1.

#### A. Data filtering and clustering

The objective of this step is event detection, i.e., detecting appliance switching on/off as well as all state transitions in multi-state appliances, such as washing machine. Let \( P(t_i) \) be the real power consumption measurement at time instance \( t_i \), for \( i = 1, ..., n \). Then, \( \Delta P(t_i) = P(t_{i+1}) - P(t_i) \), for \( i = 1, ..., n-1 \), denotes the power variation signal between adjacent aggregate power readings. Note that each \( \Delta P(t_i) > 0 \) is defined as an increasing edge and \( \Delta P(t_i) < 0 \) are decreasing edges. As the interval \( t_{i+1} - t_i \) for \( i = 1, ..., \) is constant, we simplify the notation as \( \Delta P(i) = \Delta P(t_i) \). To avoid detecting standby settings, the initial event threshold is set to \( T_0 = 10W \), thus, all \( \Delta P(i) \in (-\infty,-T_0) \cup (T_0, \infty) \) will be regarded as candidates for event occurrences \([23]\).

Next, we design a graph where a sample \( \Delta P(i) \) that is greater than \( T_0 \) (or less than \( -T_0 \)) is associated to a node \( v(i) \) of a graph. \( A(i,j) \) denotes the weight of an edge from node \( v(i) \) to \( v(j) \) that depends on the level of correlation between \( x_i = \Delta P(i) \) and \( x_j = \Delta P(j) \), calculated as in (1) using Euclidian distance measure. We
set $s(1)$ to 1 if $ΔP(i) > T_0$ and -1 otherwise, and initialise all $s(j) = 0, j > 1$. Then, we perform GSP and calculate (5). If $s^*(j) > q$, where $q$ is a constant, then $ΔP(j)$ is assigned to the first cluster of events, and removed. We set then $s(1)$ to the first remaining un-clustered element, i.e., the first $ΔP(j)$ that is less than $q$, and repeat the procedure by adding $ΔP(i)$'s to the next cluster of events. The procedure is iterated until all candidate events are clustered.

Let $μ_i$ and $σ_i$ denote, respectively, the mean value and the standard deviation of Cluster $C_i$. We evaluate the quality of cluster $C_i$ by a ratio $R_i$ as: $R_i = |μ_i| / σ_i$. The mean values of the clusters with the largest $R_i$ will determine $T_N$ and $T_P$, thresholds used for positive and negative edges, respectively. That is, a set of candidate events $Π$ is redefined as: $Π = ΔP ∈ (−∞, T_N) ∪ (T_P, ∞)$.

We continue the same GSP-based clustering process as above for all clusters that have $R_i > K$ by halving $σ$ in (1) in each iteration. After each iteration, all resulting clusters with $R_i ≤ K$, will be saved as final clusters, passed to the following matching cluster in the Feature Matching step, explained in the next section, and removed from $Π$. The clustering will end when there are no remaining elements in $Π$.

B. Feature matching

Since the final clusters compose several “positive” clusters (comprising increasing power edges) and the same number of “negative” clusters (with decreasing power edges), we pair each “positive” cluster with a “negative” cluster into a merged matched cluster, by comparing the absolute mean value of all final clusters. Next, we use GSP to pair each increasing edge in a positive cluster with a decreasing edge in its matched negative cluster, exploiting magnitude differences and time intervals between increasing and decreasing edges as two matching features in the following way.

We start with the magnitude-wise largest matched cluster, that is a matched cluster that has the largest absolute power value. Let $C_P$ and $C_N$ denote, respectively, the set of positive and negative events, namely, increasing and decreasing power edges, in the matched cluster. The task is, for each $C_P(t) ∈ C_P$ to find an optimal match among all candidates $C_N(t) ∈ C_N$. We form a graph by considering only events in $C_N$ that occur after $C_P(t)$ and before $C_P(t + 1)$. This subset of $C_N$ will be regarded as a set of candidates, denoted by $Ω$. Let $Ω_M$ and $Ω_T$ represent, respectively, the set of magnitude differences between $C_P(t)$ and each element in $Ω$ and the corresponding set of time intervals.

We form two graphs: (i) the first one with $A(i, j)$ in (1) calculated using $Ω_M$ and $s(1)$ set to be the average value of the magnitude differences between each element in $C_P$ and the corresponding candidates in $C_N$; (ii) with $A(i, j)$ in (1) calculated using $Ω_T$ and $s(1)$ set to be the median of the time intervals between each element in $C_P$ and the corresponding candidates in $C_N$ (instead of the mean value to reduce the influence the extreme scenarios). In both cases the nodes are indexed by $Ω$. In this way we obtain two solutions from (5), which we denote as vectors: $ω_M^*$ and $ω_T^*$. Then we formulate an optimization problem for feature matching as:

$$\arg\max_i (ω^*(i)) = \arg\max_i (αω_M^*(i) + βω_T^*(i)),$$

where for $i = 1, \ldots, m$, $m$ is number of candidates, that is the length of $ω_M^*$ and $ω_T^*$, and $α$ and $β$ are heuristically chosen, with $(α + β = 1)$ to tradeoff magnitude difference and time. The solution of (6) provides the optimal decreasing edge for a given increasing edge. After pairing all events in a matched cluster, the rejected events will be included into the next matched cluster, i.e., a magnitude-wise smaller cluster. See Fig. 1.

Finally, the above disaggregated appliance clusters are labelled by comparing the disaggregated signature with a database of signatures available for that particular household, which could be done, e.g., via Dynamic Time Warping [25] or cross-correlation. The database of signatures is populated via a time diary, i.e., the signature is extracted at the timestamp the household switches on and off the appliances in their house, preferably with no other appliance operating simultaneously.

IV. RESULTS AND DISCUSSION

We demonstrate the performance of the proposed algorithm using active power readings from two datasets: (1) the REDD public datasets [15] (http://redd.csail.mit.edu) of US houses down-sampled to 1 minute; (2) the UK REFIT dataset (http://www.refitsmarthomes.org) with measurements collected every 8 seconds. We set $K$ for evaluating cluster quality $R_i$ to 10% and the initial Gaussian Kernel weighting function scaling factor $σ_0$ to 20 to avoid over-clustering. $α = β = 0.5$, and $q = 0.98$ to cluster only highly correlated samples.

The evaluation metrics used are adapted from [23], [11], Precision (PR), Recall (RE) and F-Measure ($F_M$). As in [11], we separate true positives (TP) into two cases, accurate true positive (ATP) and inaccurate true positive (ITP). ATP presents the correct claim the detected appliance was running and the corresponding events are correctly named, while, ITP represents the correct claim the detected appliance was running but the corresponding events are wrongly named. False positives (FP) denote a wrong claim that the detected appliance was running, and false negatives (FN) indicate that the appliance operation was not detected. Then:

$$PR = ATP / (ATP + FP)$$

$$RE = ATP / (ATP + ITP + FN)$$

$$F_M = 2 \cdot (PR \cdot RE) / (PR + RE),$$

$PR$ represents the accuracy of event detection, thus a lower $FP$ leads to a higher $PR$. $RE$ denotes the strength of events detection and clustering, thus lower $FN$ and $ITP$ lead to a higher RE is. $F_M$ balances $PR$ and $RE$.
In order to enable like-for-like comparison with [23], we use REDD Houses 2 and 6. The results are shown in Tables I and II. House 2 contains 6 appliances, one appliance is ignored as it was rarely used, i.e., two runs in two weeks. House 6 contains 9 appliances. Approximately two weeks worth consecutive data is used in this experiment. Note that all data is used for testing - there is no training in the proposed method.

<p>| TABLE I. PERFORMANCE OF THE PROPOSED APPROACH FOR HOUSE 2 FROM THE REDD DATASET. |</p>
<table>
<thead>
<tr>
<th>Appliance</th>
<th>ATP</th>
<th>TTP</th>
<th>FP</th>
<th>FN</th>
<th>PR</th>
<th>RE</th>
<th>F_M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microwave</td>
<td>81</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0.98</td>
<td>0.9</td>
<td>0.94</td>
</tr>
<tr>
<td>Toaster</td>
<td>151</td>
<td>19</td>
<td>54</td>
<td>40</td>
<td>0.74</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td>Stove</td>
<td>17</td>
<td>1</td>
<td>84</td>
<td>20</td>
<td>0.17</td>
<td>0.45</td>
<td>0.25</td>
</tr>
<tr>
<td>Fridge</td>
<td>574</td>
<td>23</td>
<td>141</td>
<td>149</td>
<td>0.8</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>DW</td>
<td>18</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>0.67</td>
<td>0.9</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<p>| TABLE II. PERFORMANCE OF THE PROPOSED APPROACH FOR HOUSE 6 FROM THE REDD DATASET. |</p>
<table>
<thead>
<tr>
<th>Appliance</th>
<th>ATP</th>
<th>TTP</th>
<th>FP</th>
<th>FN</th>
<th>PR</th>
<th>RE</th>
<th>F_M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microwave</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>0.77</td>
<td>1</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Toaster</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.57</td>
<td>0.5</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Stove</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>0.7</td>
<td>0.5</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Fridge</td>
<td>439</td>
<td>8</td>
<td>56</td>
<td>132</td>
<td>0.89</td>
<td>0.76</td>
<td>0.82</td>
</tr>
<tr>
<td>DW</td>
<td>26</td>
<td>6</td>
<td>61</td>
<td>5</td>
<td>0.3</td>
<td>0.45</td>
<td>0.42</td>
</tr>
<tr>
<td>Heater</td>
<td>3</td>
<td>0</td>
<td>56</td>
<td>3</td>
<td>0.05</td>
<td>0.5</td>
<td>0.09</td>
</tr>
<tr>
<td>AC</td>
<td>44</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.81</td>
<td>0.86</td>
<td>0.9</td>
</tr>
<tr>
<td>Light</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>12</td>
<td>0.5</td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td>Unknown</td>
<td>146</td>
<td>6</td>
<td>56</td>
<td>65</td>
<td>0.72</td>
<td>0.67</td>
<td>0.69</td>
</tr>
</tbody>
</table>

From Table I, most appliances are disaggregated with accuracy > 75%. The worse performance is achieved for the stove. The poor performance for the heater and light in House 6, shown in Table II, is due to the questionable ground truth data collected using sub-metering for these two appliances.

Next, we present the results for the REFIT dataset. The results from REFIT House 8 for 7 available appliances are shown in Table III. It can be seen that the proposed training-less approach showed very good performance for the kettle and refrigerator, but poor results for the TV, caused by TV being grouped in the same cluster as the freezer, because TV and freezer have close active power range. Note that, our approach shows average performance (across all appliances) of $F_M = 0.49$, which is better than the HMM-based approach in [9] which shows $F_M = 0.46$ for the same house.

Then, we compare our results $F_{M_U}$ to those of the supervised GSP-based approach in [23] $F_{M_S}$ and the results using unsupervised HMM-based method of [12] $F_{M_H}$ as reported in [23]. All three methods are tested using the same data, while the two benchmark methods require additional data for training [23]. The results for the three NALM approaches for both houses are shown in Table IV. The HMM-based approach is the best for the refrigerator disaggregation. In fact, HMMs are known to perform well for events with regular cycles, such as refrigerator that operate usually alone during the night intervals when good statistics can be collected to build a state transition models [12], [25]. Apart from the refrigerator, both GSP-based approaches perform significantly better than the HMM-based approach. The proposed unsupervised GSP-based approach performs, on average, as well as the supervised one of [23], but without the training and supervised labelling overhead.

<p>| TABLE IV. AVERAGED PERFORMANCE OF THREE NALM APPROACHES FOR HOUSES 2 AND 6 FROM THE REDD DATASET. |</p>
<table>
<thead>
<tr>
<th>Appliance</th>
<th>House2</th>
<th>House6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microwave</td>
<td>0.94</td>
<td>0.92</td>
</tr>
<tr>
<td>Toaster</td>
<td>0.73</td>
<td>0.9</td>
</tr>
<tr>
<td>Stove</td>
<td>0.25</td>
<td>0.81</td>
</tr>
<tr>
<td>Fridge</td>
<td>0.78</td>
<td>0.48</td>
</tr>
<tr>
<td>DW</td>
<td>0.77</td>
<td>0.42</td>
</tr>
<tr>
<td>Heater</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AC</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Light</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

V. CONCLUSION AND FUTURE WORK

In this paper, we build on the emerging GSP concepts to develop a novel, blind, unsupervised low-rate NALM approach. Our approach does not require any training, and has lower complexity than traditional approaches, such as FHMM based. We use GSP in the proposed algorithm for solving different optimization problems. Based on the results from two datasets, our unsupervised GSP-based NALM approach performs as well as the supervised GSP-based NALM approach of [23] outperforming an unsupervised HMM-based method.

This paper has further demonstrated the potential of GSP for power disaggregation. Future work will include enhancing the algorithm robustness for dealing with event malposition and appliances’ simultaneous operation, improving performance by averting over-clustering, and developing real-time applications.

ACKNOWLEDGMENT

This work is supported in part by the UK Engineering and Physical Sciences Research Council (EPSRC) projects REFIT EP/K002708, under the Transforming Energy Demand in Buildings through Digital Innovation (BuildTEDDI) funding programme. The authors would like to thank J. Kolter and M. Johnson for making the REDD database available.
REFERENCES


