Unsupervised Clustering of Residential Electricity Consumption Measurements for Facilitated User-Centric Non-Intrusive Load Monitoring

F. Jazizadeh[1], B. Becerik-Gerber[1], M. Berges[2] and L. Soibelman[1]

[1] Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 South Vermont Ave., Los Angeles, CA 90089-2531, E-mail: jazizade@usc.edu, becerik@usc.edu, soibelman@usc.edu
[2] Civil & Environmental Engineering Department, Carnegie Mellon University, Porter Hall 113, Pittsburgh, PA 15213-3890, E-mail: marioberges@cmu.edu

ABSTRACT

Non-intrusive load monitoring (NILM) is a low-cost alternative to appliance level sub-metering, that leverages signal processing and machine learning techniques to estimate the power consumption of individual appliances from whole-home measurements. However, the difficulty associated with obtaining training data sets for the commonly used supervised NILM classification algorithms is a major obstacle in wide commercial adoption of the technology. The diversity of electrical load signatures (patterns of appliances' power draw) demands in-situ training (labeling of the signatures), which often needs to be performed by ordinary users through user-system interaction. To produce the example signatures required for training, continuous interaction with users might be required, which could reduce the success of the training process due to user fatigue. Pre-populating the training data set could help facilitate the process by reducing the number of user-system interactions needed for labeling. Taking into consideration all the issues described above, a study to test the feasibility of autonomous clustering of similar appliances' signatures based on hierarchical clustering was investigated. The information contained in the structure of the binary cluster tree was used for clustering without the need for *a priori* selection of the number of clusters. The assessment, carried out on data collected from a residential setting, showed promising results (with accuracy above 90%, calculated based on the ground truth labels) supporting the feasibility of the approach for unsupervised clustering.

INTRODUCTION

Non-intrusive load monitoring (NILM), as an alternative to appliance level sub-metering, provides the opportunity for low-cost sensing through the application of signal processing and machine learning algorithms. NILM was originally introduced by Hart (Hart 1992) in late 80's. Since the introduction of the NILM concept and during the last two decades, many research efforts have been made to improve NILM for electricity disaggregation. The majority of the research efforts so far have been focused on event-based techniques. These techniques detect significant variations in the sensed signals as events, which are assumed to correspond to operational state transitions of appliances (e.g., a refrigerator compressor going from off to on). Upon event detection, events have to be classified as a specific state transition related to an appliance. For this purpose, signal characteristics extracted from measurements in proximity of the events (i.e., appliance signatures or fingerprints) are used as representative features, and then passed on to a classification algorithm for labeling.

However, as a supervised learning problem, event based NILM systems require the provision of a training data set. A training data set consists of a set of example signatures with labels, which associate the signatures to an appliance state transition. Provision of the training data sets for the commonly used supervised NILM classification algorithms is a major obstacle in wide commercial adoption of the technology. The diversity of electrical load signatures (patterns of power draw) demands in-situ training (labeling), which needs to be performed by ordinary users through user-system interaction. To produce a complete and fairly populated data set of the example signatures, required for the training, continuous interaction with users may be needed, which could increase the need for user-system interaction. For example a user needs to turn a TV on and off for several times and provide the label to NILM system.

Pre-populating the training data set helps facilitate the process by reducing the number of user-system interactions needed for labeling. Pre-populating requires an approach that clusters similar signatures together prior to the user-system interaction step. In a NILM application, the number of clusters depends on the number of appliances state transitions. The appliances could be two-state appliances, such as a light bulb (with on and off states) or multi-state appliances, such as a washing machine (with different states of washing, rinsing, or spinning). Therefore, determining the exact number of states (i.e., number of clusters) is a challenging task that requires accurate monitoring at the appliance level. Accordingly, in this study we investigated the feasibility of autonomous clustering of similar signatures by using hierarchical clustering. This approach was used due to that fact that the information contained in the structure of the hierarchical binary cluster tree could be used for autonomous determination of clustering threshold. The following sections of the paper present a brief review of the NILM research background, the description of the data set used in this study, the proposed methodology for unsupervised clustering, and the findings, followed by the conclusions.

NILM RESEARCH BACKGROUND

Over the course of the last two decades, the majority of the research studies have focused on the event-based NILM, introducing new features and their related data processing and feature extraction processes to improve the performance of the algorithms. Feature extraction analysis, which is impacted by the type of data acquisition technique, has been the focus of many NILM studies (Zeifman and Roth 2011). Variations in steady state power metrics (differences between power metrics before and after an event), such as the real power, reactive power, power factor, RMS current, and RMS voltage, are among the most common steady state features used in the recent research studies (e.g., Marchiori et al. 2011). Other features such as the harmonic content of the current waveform (Akbar and Khan 2007), and information contained in the transients between steady states (Berges et al. 2011) were also integrated in a number of studies to improve the performance of the algorithms. In these studies, the efforts were mainly focused on the NILM event detection and pattern recognition with little attention paid to the challenges associated with training these algorithms. However, in recent years, the training process and its facilitation have been the subject of a number of studies. The introduction of a user-centric NILM system (Berges et al. 2011) and use of non-electricity sensor assisted training (Rajagopal et al. 2013; Schoofs et al. 2010; Taysi et al. 2010) are among these efforts. This study contributes to these efforts by improving upon the user-centric NILM systems.

DATA SET DESCRIPTION

The feasibility assessment was carried out through observations of the feature space (i.e., the collection of extracted signatures). This study was conducted using part of the data (one phase of the two) in the BLUED data set (K. Anderson et al. 2012), collected in a house in Pittsburgh, PA. The data set contains current and voltage raw waveforms sampled at 12KHz for almost a week. These waveforms were processed into power metrics using the spectral envelope coefficient method (Shaw et al. 2008). By using this method, the real and reactive power time series were calculated. The events on the aggregated power time series were labelled for every single load. These ground truth labels were used for performance evaluation in this study. The list of appliances for phase A is presented in Table 1.

Table 1. List of appliances, related numeric label (App ID) and number of events

App. ID	Number of events	Appliance Name	App. ID	Number of events	Appliance Name
108	16	Kitchen Aid Chopper	148	6	Washroom Lights
111	581	Refrigerator	156	94	Bathroom upstairs lights
127	18	Air Compressor	158	19	Bedroom lights
132	8	Hair Dryer	207	36	Circuit 7
147	16	Backyard Lights	1000	65	Unknown Events

Feature Extraction. As noted, the signal characteristics in the proximity of each event could be extracted as a feature vector representing that event. The feature vector used in this study is a vector of real and reactive power (for the fundamental frequency - 60Hz) in a window (with 40 samples (two third of a second) before the event and 60 samples (1 second) after the event) around the appliance state transition event ($\boldsymbol{x_i} = F[n] = \{P_1[n], Q_1[n]\}$). For defining this feature vector, it is assumed that all the appliances are fed on one phase only.

HIERARCHICAL CLUSTERING AND UNSUPERVISED PRUNING

Binary cluster tree generation. A hierarchical clustering approach finds clusters at different levels of granularity in the signature (feature vector) space for detecting the natural partitions in appliances' state transitions. Hierarchical clustering uses distances between the feature vector and connectivity between the clusters to find all possible partitioning in the data set. In this study, an agglomerative hierarchical clustering algorithm was used (D. Manning et al. 2008). This algorithm starts with all data points as singleton clusters; the pairwise proximity matrix of all singleton clusters is generated. Based on the distances between clusters in proximity matrix, clusters are merged to form binary clusters; the proximity matrix is then updated and the iterative merging continues until only one cluster remains. The output of the algorithm is a binary class tree, which represents natural connections between all the vectors in the feature vector space. Figure 1 presents the binary tree generation algorithm.

Input: FV
$C_b \leftarrow \forall \boldsymbol{x} \in FV$
$N_C \leftarrow size(C)$
For each $\boldsymbol{x}_i$
    $D_{pw} \leftarrow dist_x(\boldsymbol{x}_i, \boldsymbol{x}_j)$
End
While $(\exists c \in C \; covering \; all \; N_C \; clusters)$
    $C_b \leftarrow merge\ clusters(D_{pw})$
    $N_C \leftarrow size(C)$
    $D_{pw} \leftarrow dist_c(c_i, c_j)$
End
Output: Binary Cluster Tree	

FV: The set of events' signatures C_b: the binary cluster set	N_C : the number of clusters D_{pw}: the pairwise distance matrix

Figure 1. A high level representation of the agglomerative hierarchical binary cluster tree generation algorithm (D. Manning et al. 2008)

In this study a city block distance metric (distance between feature vectors) and a single linkage metric (distance between clusters) were used:

$$dist_x = \sum_{i=1}^{D} (\boldsymbol{x}_{mi} - \boldsymbol{x}_{ni})$$

$$dist_c(c_i, c_j) = min_{\boldsymbol{x}_m \in c_i, \boldsymbol{x}_n \in c_j} d(\boldsymbol{x}_m, \boldsymbol{x}_n)$$

The binary tree of the clusters (i.e., the dendrogram) generated through the above mentioned algorithm is shown in Figure 3 for the turn-on events. The dendrogram represents the structure of the tree. The inverse *u* shaped links show the relationship between clusters. The height of these connections shows the distance between the clusters. The leaf nodes in the tree represent singleton clusters, which are individual appliance state transitions, and the root node covers all state transitions' feature vectors. In Figure 2, part of the tree has been illustrated and therefore, leaf nodes contain the sub-trees.

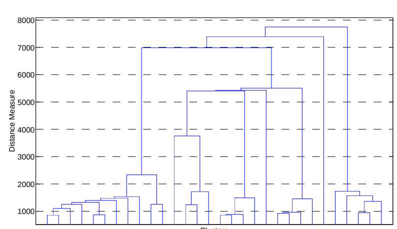

Figure 2. Dendrogram of the "turn-on" events on phase A

Pruning cluster tree into clusters. The cluster tree contains clusters of the data at different levels of granularity. Once the tree is generated, the clusters are obtained by pruning the tree, which could be carried out considering different criteria including the number of clusters, an inconsistency threshold, and a distance threshold. The inconsistency threshold could be used as an approach to find natural segmentation of the data in the feature space. When the heights of two successive links are close together, the links are consistent and there is no clear distinction between the clusters. The inconsistency could be measured by comparing the height of a link with the average height of links below it. A sensitivity analysis for the inconsistency coefficient using data for on and off events showed that very small changes in the inconsistency threshold could result in dramatic changes in the number of clusters, which might not represent the natural partition of the feature space. Consequently, we explored the feasibility of using a global inconsistency measure to determine the distance threshold for pruning the tree in an iterative clustering process as described below and in the evaluation section.

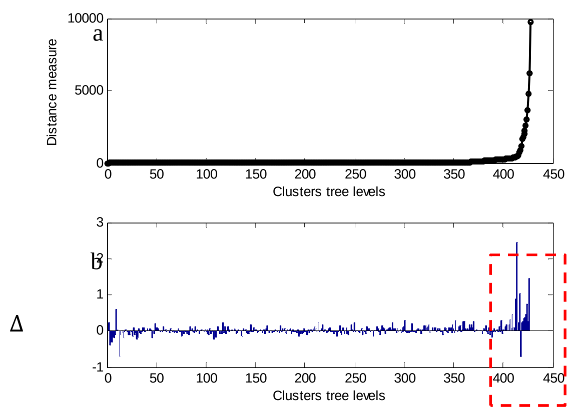

Figure 3.Variation of the tree characteristics from leaf nodes to root; a) distance measure variation; b) variation of rate of growth in distance slope

Figure 3a shows the variation of the distance along the tree. As it can be seen, the distances grow with a very fast rate as we get closer to the root node. It is this characteristic of the tree that we used for finding the distance threshold for pruning the tree. Accordingly, the rate of growth in slope of the distance curve is used to determine the threshold for pruning the tree. The distance threshold is determined as the equivalent point on the tree, where the growth in slope is maximized:

$$\delta_i = dist_{i+1} - dist_i$$

$$\Delta = (\delta_{i+1} - \delta_i)/\delta_i$$

$$\tau_d^* = \underset{tree\ level}{\operatorname{argmax}} \Delta$$

in which, $dist$ is the distance between clusters, δ is the distance slope metric, Δ is the rate of growth in slope, and τ_d^* is the distance threshold for pruning the tree.

EVALUATION AND FINDINGS

Observations of the performance of the proposed heuristic on the data showed that the local inconsistency in distance and slope growth might result in finding a local maximum value for Δ instead of a global value. The local threshold results in unnatural partitioning of the feature space. Accordingly, in this study, the maximization of the Δ value was carried out in 15% of the upper portion of the tree, where the slope growth accelerates – highlighted by dashed line rectangle in Figure 3b. This portion was selected through approximate evaluation of the δ values histogram to find the point of balance between smaller and larger δ values. Moreover, depending on the feature space characteristics, various scales could be observed in the structure of the tree. This holds specifically true for NILM applications, where appliances' signatures with different ranges of variations could be observed. Due to this multi-scale nature, evaluating the distance measure over the entire feature space could result in a tree structure, which in turn brings about a distance threshold that partitions the tree in one scale. Accordingly, the resultant partition will be a partial partition of the feature space, with a residual cluster that contains the feature vectors at a different scale. In this study, to address this challenge, the clustering process was carried out repeatedly on the residual clusters, until no further separation of the feature vectors was possible. In each repetition, the clusters with one or two feature vectors were removed. To provide a sense about the feature space, Figure 4 presents the feature vectors for turn-on events in each class. Titles show the label and number of feature vectors for each class.

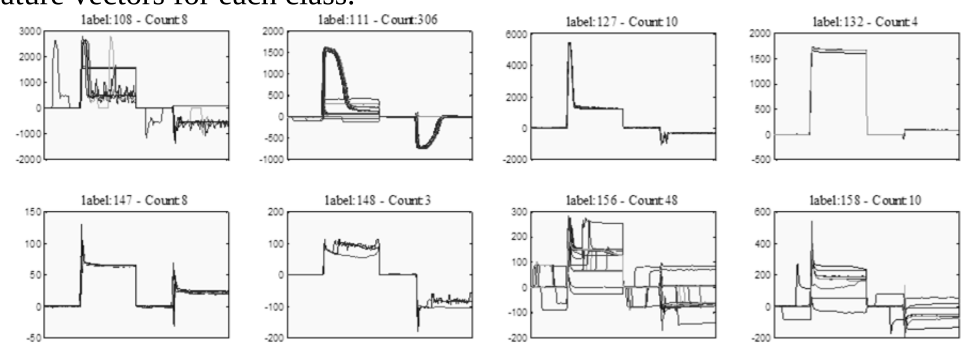

Figure 4. Feature space ground truth distribution with label and number of vectors

Table 2. Cluster-class association matrices for turn-on and turn-off events

	Turn-On Events													Turn-Off Events											
	Autonomously Assigned Cluster Labels													Autonomously Assigned Cluster Labels											
Ground Truth Labels		1	2	3	4	5	6	7	8	9	10	11	12		1	2	3	4	5	6	7	8	9	10	11
	111			125			3	3	23	4	4	4	121	111			3				171		22	100	
	147			8										127				9							
	156	11	3	23		3								147											2
	158				3									148					3						
	207			8										156					14	3	1	29			
	Numbers in columns show the number of feature vectors													158	6	3									
														207				9							1

The matrices, which show the association between ground truth labels and autonomously assigned cluster labels, were presented in Table 2. In fact, Table 2 presents a metric similar to confusion matrix. In this study, an accuracy measure is obtained by matching the cluster labels with the ground truth labels and it is defined as the maximum number of each column divided by total number of feature vectors in that column. Each column represents a cluster. In fact, the ground truth label associated with the maximum number of signatures for each cluster determines the label of the cluster. As Table 2 shows, clusters are comprised of dominant unique feature vectors with accuracy of 0.98 and 0.91 for turn-on and turn-off events, respectively. Although some of the cluster contains feature vectors from different classes, the visual representation of the clusters in Figure 5 shows that similar feature vectors were accurately clustered together. Figure 4 shows that not all feature vectors in the same class are similar. For example in refrigerator class, signatures are associated to compressor and the light. The same pattern was observed in the turn-off events. The clusters for the turn-off events were generated using a kernelized distance through a first degree polynomial kernel function ($k(\boldsymbol{x}_m, \boldsymbol{x}_n) = x_m^T x_n$), which was observed to provide better results. The obtained results show promising performance of the proposed heuristic.

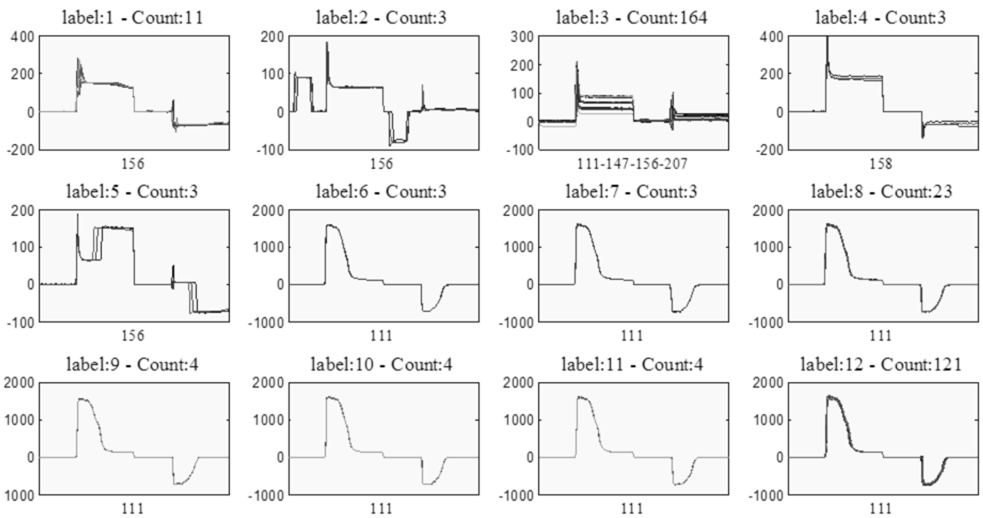

Figure 5. Resultant clusters generated for turn-on events (titles show the cluster label and the number of x_i and horizontal axis label shows the actual labels of x_i

CONCLUSION

In this study, the feasibility of an unsupervised clustering approach with the objective of determining the natural partitioning of appliances' feature space in NILM applications was investigated. The technique could potentially help facilitate the training process of event-based NILM systems (for which supervised learning algorithms are commonly used) by pre-populating the training data set before user-system interaction step. A heuristic approach for unsupervised pruning of a hierarchical binary cluster tree was presented and evaluated. The characteristics of the resulting binary cluster tree, specifically, the fast growth rate of the intra-cluster distance in proximity of the root node, were used to find a distance threshold for pruning the tree. The evaluation of the partitioned feature space on a data set, collected from a residential setting, showed promising results in feasibility of unsupervised partitioning. Developing the proposed heuristic as a fully autonomous heuristic algorithm, further assessment of the approach on different data sets, comparison of the approach with other unsupervised clustering techniques such as spectral clustering and mean shift algorithms, comprehensive evaluation of the performance, and applying unsupervised clustering integrated with a facilitated training framework are among the authors' future research directions.

ACKNOWLEDGEMENT

This work supported by the National Science Foundation under grant #09-30868. The opinions presented here are those of the authors and do not necessarily reflect the views of the National Science Foundation**.**

REFERENCES

Akbar, M., and Khan, Z. A. (2007). "Modified nonintrusive appliance load monitoring for nonlinear devices." *Proc., 11th IEEE International Multitopic Conference, INMIC 2007, December 28, 2007 - December 30,* Inst. of Elec. and Elec. Eng. Computer Society, .

Berges, M., Goldman, E., Matthews, H. S., Soibelman, L., Anderson, K. (2011). "User-Centered Non-Intrusive Electricity Load Monitoring for Residential Buildings." *J. Comput. Civ. Eng.*, posted ahead of print.

Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze,. "Hierarchical Clustering." *In an Introduction to Information Retrieval, Http://Nlp.Stanford.Edu/IR-Book/Pdf/17hier.Pdf,* .

Hart, G. W. (1992). "Nonintrusive Appliance Load Monitoring." *Proc IEEE,* 80(12), 1870-91.

K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Berges. (2012). "BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research." *In Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD), Beijing, China,* .

Marchiori, A., Hakkarinen, D., Qi Han, Earle, L. (2011). "Circuit-Level Load Monitoring for Household Energy Management." *Pervasive Computing, IEEE,* 10(1), 40-48.

Rajagopal, N., Giri, S., Berges, M., Rowe, A. (2013). "Demo Abstract: A Magnetic Field-Based Appliance Metering System." , 307-308.

Schoofs, A., Guerrieri, A., Delaney, D. T., O'Hare, G., Ruzzelli, A. G. (2010). "Annot: Automated electricity data annotation using wireless sensor networks." *Proc., Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference On,* IEEE, , 1-9.

Shaw, S. R., Leeb, S. B., Norford, L. K., Cox, R. W. (2008). "Nonintrusive Load Monitoring and Diagnostics in Power Systems." *IEEE Transactions on Instrumentation and Measurement,* 57(7), 1445-54.

Taysi, Z. C., Guvensan, M. A., Melodia, T. (2010). "Tinyears: Spying on house appliances with audio sensor nodes." *Proc., Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building,* ACM, , 31-36.

Zeifman, M., and Roth, K. (2011). "Nonintrusive appliance load monitoring: Review and outlook." *Proc., 2011 IEEE International Conference on Consumer Electronics, ICCE 2011, January 9, 2011 - January 12,* Institute of Electrical and Electronics Engineers Inc, , 239-240.