Using the Energy Error to Improve the Performance of Unsupervised Approach in Non-Intrusive Load Monitoring

Hader A. D. Azzini, Luiz C. P. da Silva

Department of Electrical Energy Systems
University of Campinas
Campinas, Brazil
hader@dsee.fee.unicamp.br

***Abstract*— this paper proposes the use of the percentage of Energy Error as feedback to improve the performance of an unsupervised approach for Non-Intrusive Load Monitoring. In this approach, different values of thresholds for the composition of clusters in ISODATA algorithm are tested, looking for the value which minimizes the percentage of Energy Error. This approach reduces the percentage of Total Energy Error from around 25% to around 15%, reaching the level of supervised machine learning approaches. The use of unsupervised approach could improve the commercialization of Non-Intrusive Load Monitoring technology since it reduces, or even eliminates, the training phase.**

***Index Terms*— nonintrusive load disaggregation, nonintrusive load monitoring, smart meter, unsupervised machine learning.**

I. Introduction

The basic objective of Nonintrusive Load Monitoring (NILM) is to identify the individual consumption of appliances without having access to them. This is achieved from the detailed analysis of voltage and current signals measured at the house service panel. The waveforms of total current and voltage supplying the monitored load are recorded and analyzed, leading to estimations of individual appliances consumption. Such waveform analysis is performed using appliance-specific electrical characteristics, commonly named signatures, which are capable of identifying which load has been switched ON/OFF on a particular instant. The term "nonintrusive" is related to the absence of physical access to internal points of the electrical installation [1]. There are many kinds of signatures such as changes of active power, changes of reactive power, transient shapes, harmonics, etc. [2].

There are several types of NILM algorithms, and most of them are composed by four basic steps. The first step is the event detection: based on the consumption pattern recorded for pre-determined electrical variables, the NILM algorithm must determine when an appliance switch-ON/OFF event has occurred. The second step consists on to recognize to which appliance the event belongs and it can be done with supervised or unsupervised machine learning techniques. The third step is the calculation of appliance working cycle duration: using events from a specific group, switch-ON and switch-OFF events are combined to determine how long the appliance has been in operation. Finally, the fourth step is the determination of individual appliance consumption, based on the calculated working cycle duration and demanded active power.

For example, considering the power curve of Figure 1, at first the events are detected, which are steps in the power curve. After, the events must be separated for each load, this means that the events of the Electric Shower should be together and so on. In sequence, the ON and OFF events of each load, as similar to the ones show in the two first columns of Table I, should be combined to generate the estimation of operating time. At the end, the consumption of the Electric Shower is estimated.

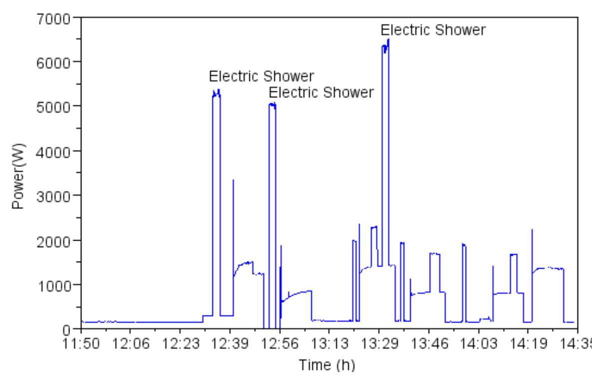

FIGURE 1 CURVE OF ACTIVE POWER OF PHASE 'A' (SAMPLED IN MARCH

The São Paulo Research Foundation – FAPESP gave financial support to this work. Opinions, assumptions, conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of FAPESP.

Part of the results presented in this paper have been obtained through the project "Training in Information Technology," funded by Samsung Electronics Amazonia Ltda, using resources of Computing Law No 8.248/91.

19TH)

TABLE I . IDENTIFIED ELECTRIC SHOWER OPERATIONS

Event	On	off	Operating time	kWh	Total (kWh)
1	12:34:09	12:36:30	0:02:21	0.195	0.543
2	12:53:07	12:55:13	0:02:05	0.175	
3	13:31:08	13:33:12	0:02:04	0.173	

Several NILM algorithms have been developed from late 1980's to the present day [1–5], with special attention to [1], which may be considered a major milestone in this area. In addition, recently, NILM has received great attention under the context of smart grids, due to the development of smart meters [6]. The advancement of smart meter capabilities is highly favorable to nonintrusive load monitoring. However, the nonintrusive monitoring still faces some challenges. As well stated in [2], an algorithm capable of detecting switch-ON/OFF events for all types of loads in a residence was not developed yet. Also, a widely accepted signature, capable of providing a robust, accurate and adaptable identification, has not yet been defined for all appliances.

Another challenge in NILM is the training phase for supervised machine learning approaches [7]. The recognition of which event belongs to which appliance is the most critical step in the NILM and the supervised machine learning can generate good results. However, the supervised machine learning needs a training phase and to obtain the training dataset it is necessary a period of intrusiveness in the home, which goes against the main idea of the NILM. Besides that, obtaining the training dataset could be a laborious work and more susceptible to errors if it is made in a manual way [7]. And the use of training dataset from a house to another is not a very good option. Many factors have influence in the electrical characteristics of an appliance, such as manufacturing design, age, etc., as a consequence the training dataset may not fit so well to all residences.

The unsupervised machine learning approach could reduce or even eliminate the training phase, but it does not generate good results compared with the supervised approach. The use of Iterative Self-Organizing Data Analysis Techniques (ISODATA), an unsupervised machine learning, in [8] did not provide good results compared with the supervised machine learning, using K-Means. In such context, this paper proposes the use of energy error to improve the performance of clustering in the second step of NILM programs.

The paper is organized as follows. Section II presents the ISODATA algorithm and the Adjusted Rand Index (ARI) as way to evaluate the performance of clustering algorithms and classifiers. Section III presents the energy error based approach. Section IV gives a description of the used data set and the method used in the experiments. Finally, Section V outlines the main conclusions and future works.

II. THE RECOGNITION OF EVENTS

The recognition of events can be done with supervised or unsupervised machine learning approaches [9]. In the supervised learning, the recognition is called Classification: based in events already labeled the classifier predicts to which appliance a new event should be labeled. So, the supervised approach needs a training data set, which is a set of events already labeled. In the unsupervised machine learning the recognition is called Clustering: where similar events are grouped together, and each group represents an appliance. The unsupervised approach does not use a training data set, making it easier to be used by costumers and simplifying its acceptance in the market. As stated in [7] the training phase is a complication to commercialize the NILM technology.

In [8] K-Means and ISODATA were used for the recognition of events. However, K-Means was first used to categorize the event samples of training data set and afterwards used in supervised way to recognize the new events of appliances. So, the K-Means was actually used in [8] like a supervised approach. On the other hand, ISODATA was used as unsupervised approach.

A. K-Means and ISODATA

The K-Means is one of the most popular algorithms of machine learning, which relies on the idea that the center of the cluster, named centroid, can be a good representation of the cluster [9]. The algorithm basically has the follow steps [10]:

1. Choose randomly the position of K centroids to K clusters;
2. Associate each one of the data to the nearest centroid;
3. Recalculate the position of each centroid using the current cluster memberships;
4. If a convergence criterion is not met, return to the step 2. Typical convergence criteria are minimum number of transitions of data between clusters or minimal decrease squared error.

The ISODATA algorithm, proposed by Ball e Hall in 1965, is considered a variation of K-Means algorithm[10]. The main difference of ISODATA compared to K-Means is the permission for the clusters to be merged or to be fragmented. Typically a cluster is divided when the variance exceeds a pre-established threshold and two clusters are merged when the distance between the centroids is below a pre-established threshold. Because of those permissions, the ISODATA algorithm does not need the number K of clusters, as in step 1 of K-Means. The main goal of ISODATA is to find the best partition of clusters according to the pre-established thresholds. It is important to make clear that both algorithms need pre-established values, K-Means needs the number of clusters and ISODATA needs the pre-established thresholds to merge and to split the clusters.

The need of pre-established values is a barrier to NILM. Since the NILM aims to identify the individual consumption of appliances without entrance in the house, it is not easy to obtain the values. The number K for K-Means is related to the number of appliances in the house. And for ISODATA, the standard deviation of the events in each cluster and the distances between the centroids are related to characteristics of appliances in the house. In this paper, these values for

ISODATA are obtained testing many values and choosing the one that minimizes the percentage of Energy Error.

B. Adjusted Rand Index

The supervised and unsupervised approaches used in [8] produce different results and the metrics used were not efficient to make the comparison between the two approaches. The Adjusted Rand Index (ARI) is a good option to make that comparison. This Index is one of the most used to make external validation of clusters and determines the similarity of partitions. The ARI lies between -1 and 1, and when the partitions agree perfectly the index is equal to 1. The proximity of the index to 1 means more similarity between the partitions. In (1) it is show how the index is calculated [11]. Consider $\pi^e=\{C_1,\ldots,C_i,\ldots,C_{k^e}\}$ a partition resulting from a clustering algorithm, where Ci is a cluster of the partition, and $\pi^r=\{C_1,\ldots,C_i,\ldots,C_{k^r}\}$ a priori partition, like the ground truth.

$$ARI(\pi^e,\ \pi^r) = \frac{\sum_{i=1}^{k^e}\sum_{j=1}^{k^r}\binom{n_{ij}}{2} - \left[\sum_{i=1}^{k^e}\binom{n_{i.}}{2}\sum_{j=1}^{k^r}\binom{n_{.j}}{2}\right]\Big/\binom{n}{2}}{\left[\sum_{i=1}^{k^e}\binom{n_{i.}}{2}+\sum_{j=1}^{k^r}\binom{n_{j.}}{2}\right]\Big/2 - \left[\sum_{i=1}^{k^e}\binom{n_{i.}}{2}\sum_{j=1}^{k^r}\binom{n_{j.}}{2}\right]\Big/\binom{n}{2}} \quad (1)$$

The n_{ij} is the number of elements in common for the C_i of π^e and C_j of π^r; the $n_{i.}$ is the number of elements in cluster C_i of π^e; the $n_{j.}$ is the number of elements in cluster C_j of π^r; And k^e and k^r are the number of clusters in the partitions π^e and π^r, respectively.

The Adjusted Rand Index is used in this paper to compare the results of K-Means and ISODATA of [8] to the ground truth manually marked.

III. The Percentage of Energy Error

The energy error is an important metric to evaluate the results of load disaggregation in NILM. The percentage of energy error corresponds to the percentage of the total measured energy which was wrongly estimated in NILM. The load curve of the house can be rebuilt by using the information obtained for each load in the disaggregation process. The difference between the estimated and the measured load curves provides the energy error used in this paper as shown in Figure 2 and Figure 3.

The sample time for both curves is the same, so the percentage of energy error related to the measured active power curve could be calculated as show in (2).

$$E_{error}\,(\%) = \frac{\sum |P_m - P_e|}{\sum P_m} \times 100 \quad (2)$$

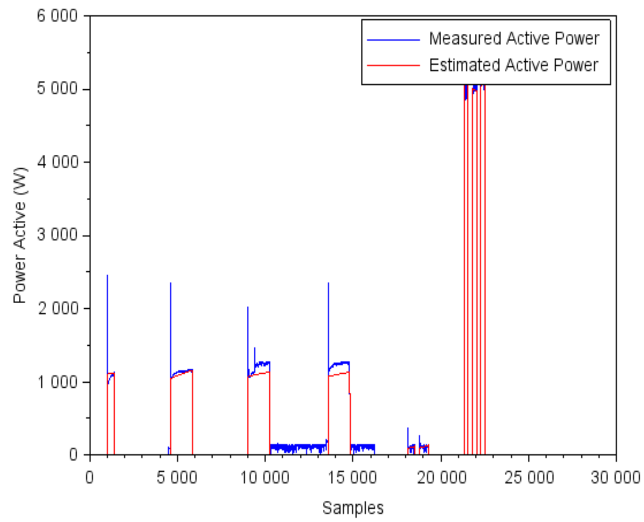

FIGURE 2 ESTIMATED AND MEASURED ACTIVE POWERS

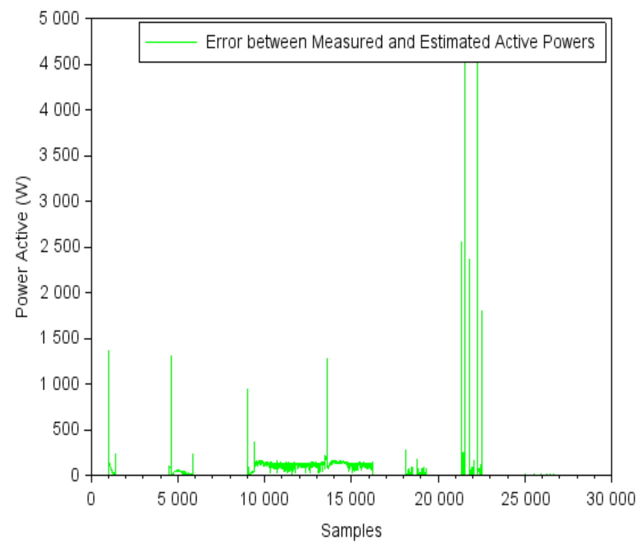

FIGURE 3 ERROR BETWEEN ESTIMATED AND MEASURED ACTIVE POWERS

The P_m is the measured active power and P_e is the estimated active power. The percentage of energy error decreases as the disaggregation becomes near to the ground truth.

IV. Experiments

A. The Data Set

The data set used is the same from [8]. An energy analyzer device was installed on the service panel of an apartment with 3 bedrooms to monitor the two phases supplying the installation. A total of 9 loads have been selected to be switched ON and OFF, generating events of loads. The selected loads are presented in Table II; they are single phase and ON-OFF type (or may be considered as such with good approximation). Measurements were performed on days 19, 20, 22 and 23 of March 2012. Each load was switched ON and OFF manually three times in order to generate the events. The only exception is the Fridge since his thermostat cannot be switched ON and OFF manually. The time of each event was manually marked.

TABLE II. MAIN LOADS IN THE MONITORED APARTMENT

	Loads
1	Air Conditioner in Suite
2	Air Conditioner in Office
3	Air Conditioner in TV Room
4	Electric Shower of Social Bathroom
5	Electric Shower of Suite
6	TV
7	Grill
8	Microwave Oven
9	Fridge

The analyzer recorded many electric variables, but for the NILM program it was chosen the Active Power (P), Reactive Power (Q), and Total Harmonic Distortion of Current (THDI). The energy analyzer was set to generate two samples per second.

B. *The implemented unsupervised approach*

Before the event is recognized it is necessary to detect the event. The event detection method used was The Window with Margins [12]. This method basically uses a scanning window which runs over active power consumption curve. Like other detection methods, this method is not perfect and may generate false positives and false negatives. The parameters used for Window with Margins could lead to more false positives or more false negatives. In general, this implementation is more sensible to false negatives then to false positives.

After the event detection is done, for each event it is calculated the difference between the means of P before and after the event. In order to calculate these means it is used 6 samples for each one. The same calculation of the difference between the means is made also to Q and to THDI. These three values, ΔP, ΔQ, and ΔDHTI, were used as attributes of the event and passed to the clustering algorithm. The detected events were placed into two different three-dimensional spaces, one for ON events and other for OFF events. The analyzer provides the measurements for two phases, so each day of measurement has 4 spaces, one ON space and one OFF space for each phase. Figure 4 shows an example of the spaces of one phase for the day March 20th. The dimensions of the spaces are ΔP, ΔQ, and ΔDHTI.

The ISODATA is applied to the four spaces, containing false positives resulting from the event detection. The clustering is made according to the pre-established thresholds to merge or split clusters. Since in the unsupervised approach it is not used a training phase the clusters are identified by numbers instead of the names of the loads, as in the supervised approach. After that, each event has a label that is a number related to each cluster.

After the clustering step, the centroids of each cluster are calculated. The centroids represent the characteristics of the load. As the ON events space and the OFF events space are different spaces, the way to connect a cluster of ON events to

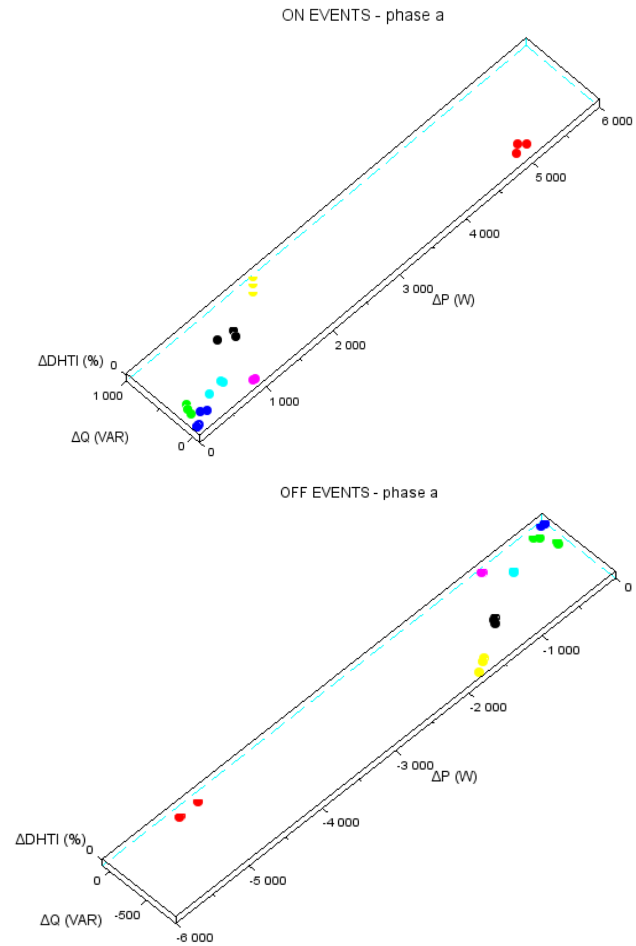

FIGURE 4 ON AND OFF SPACES FOR ONE PHASE

a cluster of OFF events was looking for the centroids which have the three most similar components in absolute values. In this implementation it is possible that the number of clusters in ON space is different than the number of clusters in OFF space. In this case, it is considered the minimum between the numbers of clusters.

After the connection of centroids, each ON event has to be connected to an OFF event to estimate the work periods of each load. In order to avoid overestimation before an ON event is connected to an OFF event it is checked if this work period could actually happen according to the measured power. It is checked if 80% of the samples from measured active power, between the ON and OFF events, have at least 80% of the smallest step of active power from the two events. If the condition is satisfied a new work period is recorded. In order to have an estimative of how well the events were correlated it is calculated the Work Period Success Rate. It consists in dividing the number of well estimated working periods to the real number of measured working periods.

Finally, when all working periods and the steps of power of each event are available, the consumption of each load could be estimated. If the estimated power curve of every appliance is summed up, it is possible to generate figures similar to Figure 2 and Figure 3. After that, the percentage of energy error is calculated.

C. Method Description and Results

The unsupervised approach, using ISODATA, was applied to 3 days of measurements, 19, 20, and 22 of March 2012. The parameters used for Window with Margins were Wd=10 samples, Nm=2 samples, and Wf=12 samples. These values were chosen to avoid any false negative because the interest is to evaluate how the ISODATA provide the clusters with the events, and the absence of some event could lead to a wrong evaluation. These values imply in many false positives in the event detection but it is expected that ISODATA creates a cluster for these unknown events. So the false negatives could generate more error than the false positives.

In [8] two clusters are merged in ISODATA if the Euclidean Distance between the two centroids of the clusters was under 40. If a cluster has less than two events it is merged to another cluster too, the closest one. And a cluster is divided in two if the standard deviation of the Euclidean Distance of the events in a cluster is above 35.

Using the values and the described steps, the NILM software generates the metrics shown in Table III. The supervised approach (SUP) was made using the K-Means with a training phase, using March 23th as training data [8]. The unsupervised approach (UNS) was made using ISODATA.

TABLE III THE METRICS GENERATED BY NILM IN [8]

Day	***March 19th***		***March 20th***		***March 22th***	
Approach	**SUP**	**UNS**	**SUP**	**UNS**	**SUP**	**UNS**
Event Detection Success Rate (%)	79.7	79.7	45.8	45.8	56.6	56.6
ARI ON phase a	1	0.82	0.97	0.97	1	0.77
ARI OFF phase a	0.96	0.71	0.85	0.62	1	0.84
ARI ON phase b	0.64	0.64	0.80	0.65	1	1
ARI OFF phase b	0.70	0.94	1	0.81	0.70	0.76
Work Period Success Rate (%)	85.2	74.1	78.6	60.7	78.8	75.8
Energy Error phase a (%)	20.9	34.6	8.21	19.6	7.98	43.8
Energy Error phase b (%)	4.29	2.71	11.3	41.2	20.2	15.2
Total Energy Error (%)	15.3	23.8	9.36	27.5	12.0	34.3

The ARI was calculated comparing the ground truth to the partitions generated by SUP and UNS. To calculate the ARI the false positives of detection were ignored. The percentage of Total Energy Error was calculated by summing up the energy error of both phases and dividing by the sum of the whole energy measured in both phases.

The UNS approach has the total error higher than 20% for every day. So the main idea is to test different values for the threshold of standard deviation for each phase looking for the one who generate the minimum percentage of energy error.

The tested values for the threshold of standard deviation were the inter numbers between 4 and 80. Table IV shows the best result generated using the percentage of energy error as a feedback. As it can be seen the results for UNS provided a level of total energy error similar to the SUP. The percentage of Total Energy Error decreased from around 25% to around 15% and without the need of a supervised training phase. Figure 5 shows the comparison of these results.

TABLE IV THE METRICS GENERATED USING THE PERCENTAGE OF ENERGY ERROR AS A FEEDBACK

Day	***March 19th***		***March 20th***		***March 22th***	
Approach	**S**	**US**	**S**	**US**	**S**	**US**
Event Detection Success Rate (%)	79.7	79.7	45.8	45.8	56.6	56.6
ARI ON phase a	1	0.55	0.97	1	1	0.65
ARI OFF phase a	0.96	0.54	0.85	0.62	1	0.84
ARI ON phase b	0.64	0.62	0.80	0.70	1	1
ARI OFF phase b	0.70	0.70	1	1	0.70	0.70
Work Period Success Rate (%)	85.2	77.8	78.6	67.9	78.8	72.7
Energy Error phase a (%)	20.9	22.4	8.21	19.6	7.98	14.3
Energy Error phase b (%)	4.29	1.93	11.3	10.5	20.2	11.9
Energy Error Total (%)	15.3	15.5	9.36	16.3	12.0	13.5

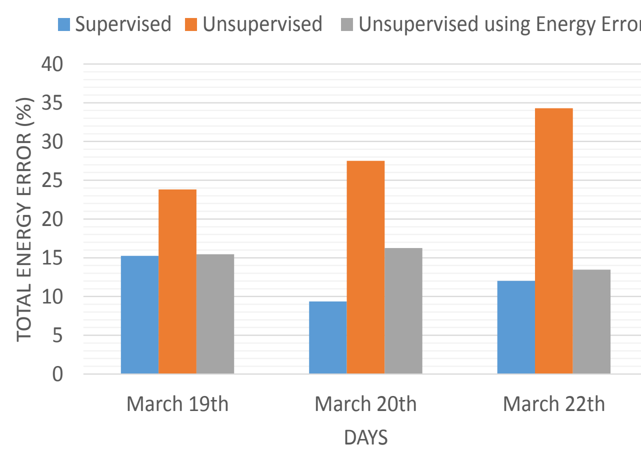

FIGURE 5 TOTAL ENERGY ERROR OF THE APPROACHES

D. Limitations of this approach

Although the results for UNS provided a level of total energy error similar to the SUP, the values of ARI for UNS in Table IV were not improved in general. It means that the proposed approach reduces the energy error but the quality of disaggregation was not improved. It happens because this approach have a tendency to reduce the number of clusters to make it easier to correlate the ON and OFF for the same loads. Since ON and OFF from some loads are correlated only if they are signed for the same cluster this approach tends to merge the clusters to increase the chance of correlating the ON and OFF. Maybe this limitation could be overcome if other event signatures are added to ΔP, ΔQ, and ΔDHTI. As examples it could be added the transient carachteristc of the loads or the amplitudes of the main harmonic components of current.

The clusters in UNS do not have the names of appliances, as in SUP, they are identified by numbers. It makes it very difficult to the owner of the house to recognize which number corresponds to which appliance. However, this problem could be minimized if the NILM have a memory of predefined signatures for many appliances. The NILM could make a search for the name of appliance which have the most similar electrical characteristics. Another possible solution is to have very simple training phase where the owner of the house is asked to switch-ON and switch-OFF the appliances just to correlate the electrical characteristic to the names. These solutions are not so susceptible to errors like the training phase discussed in the Introduction of this paper, since it is used just for the recognition of the names of appliances. The estimation of consumption of loads will not be affected by the training. It is more likely to have an appliance wrongly named than a consumption of this appliance wrongly estimated. But if the name of the appliance is wrong the owner of the house may easily realize that and after some interactions the name could be corrected.

Other limitation of this approach is the fact that all loads in the dataset are considered as ON-OFF. The results could be different if some Finite State Machine (FSM) load was present [2]. In future works, this approach will be adapted in NILMTK for the evaluation of more datasets. The NILMTK is an open source toolkit specifically designed to enable the comparison of energy disaggregation algorithms in a reproducible manner across multiple publicly available data sets [13]. And a network of sensors has been assembled and will be installed in a home to provide better data set from a residence in Brazil. The data set of [8] is composed by few days, the days were not completely measured, and the events were manually generated and marked. Using the sensors network the data set will represent a real profile of a residence, especially because it will not be necessary to manually generate and annotate the events.

V. CONCLUSIONS

This paper has proposed the percentage of Energy Error as a metric to improve the performance of an unsupervised approach for Non-Intrusive Load Monitoring. In this approach different thresholds for the composition of clusters in ISODATA algorithm are tested looking for the one who minimizes the percentage of Energy Error. This approach reduces the percentage of Total Energy Error from around 25% to around 15%, reaching the same error level of supervised machine learning approaches. The use of unsupervised approach could facilitate the commercialization of NILM technology.

This approach has some limitations, as the tendency to reduce the number of clusters and the absence of the name of the appliances. Beyond that, only ON-OFF type loads were tested. In future works it will be tested for different datasets, including loads of the type FSM.

REFERENCES

[1] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[2] M. Zeifman and K. Roth, “Nonintrusive appliance load monitoring: Review and outlook,” *IEEE Transactions on Consumer Electronics*, pp. 76–84, 2011.

[3] M. Dong, P. Meira, W. Xu, and W. Freitas, “An event window based load monitoring technique for smart meters,” *Smart Grid, IEEE Transactions on*, vol. 3, no. 2, pp. 787–796, 2012.

[4] Y. F. Wong, Y. Ahmet Sekercioglu, T. Drummond, and V. S. Wong, “Recent approaches to non-intrusive load monitoring techniques in residential settings,” in *Computational Intelligence Applications In Smart Grid (CIASG), 2013 IEEE Symposium on*, 2013, pp. 73–79.

[5] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, “Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey,” *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[6] A. Ipakchi and F. Albuyeh, “Grid of the future,” *Power and Energy Magazine, IEEE*, vol. 7, no. 2, pp. 52–62, 2009.

[7] A. Rowe, M. Berges, and R. Rajkumar, “Contactless sensing of appliance state transitions through variations in electromagnetic fields,” in *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*, 2010, pp. 19–24.

[8] H. A. D. Azzini, “Evaluation of techniques for nonintrusive monitoring residential loads for the purpose of energy audit,” Dept. Electrical Eng., Federal Univ. of Espírito Santo (In Portuguese), 2012.

[9] O. Z. Maimon, “Data Mining and Knowledge Discovery Handbook.” Springer US, Boston, MA, 2010.

[10] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[11] K. Faceli, *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro, RJ: Livros Técnicos e Científicos, 2011, p. 378.

[12] H. Azzini, R. Torquato, L. C. da Silva, and others, “Event detection methods for nonintrusive load monitoring,” in *PES General Meeting| Conference & Exposition, 2014 IEEE*, 2014, pp. 1–5.

[13] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, “NILMTK: An open source toolkit for non-intrusive load monitoring,” in *Proceedings of the 5th international conference on Future energy systems*, 2014, pp. 265–276.