FACTA UNIVERSITATIS
Series: **Automatic Control and Robotics** Vol. 22, Nº 2, 2023, pp. 103 - 113
https://doi.org/10.22190/FUACR231123008D

Regular Paper

NON-INTRUSIVE LOAD MONITORING USING CURRENT HARMONIC VECTORS AND ADAPTIVE FEATURE SELECTION

UDC (621.3.018.3)

Srđan Đorđević[1], Milan Simić[2]

[1]University of Niš, Faculty of Electronic Engineering, Department of Electronics, Republic of Serbia
[2]University of Niš, Faculty of Electronic Engineering, Department of Measurements, Republic of Serbia

Abstract. *The non-intrusive load monitoring method presented in this paper uses changes in current harmonic vectors to identify the operational state of appliances. The algorithm based on this feature has low complexity, but it may suffer from an information loss caused by a random fluctuation of the current harmonic vectors. In order to deal with this problem, we propose the algorithm which includes a stage which identifies and select a subset of relevant features in the set of available appliance features. The proposed load disaggregation algorithm is demonstrated through experiments on a representative set of household appliances.*

Key words: *Nonintrusive load monitoring, load signature, energy management, power harmonics*

1. INTRODUCTION

The appliance load monitoring has an important role in the electrical energy conservation. According to the study [1], one-third of the electrical energy wastage is a consequence of domestic activities related to electrical consumption. It is clear that consumers can significantly reduce the energy consumption by changing the way of using electrical appliances. This goal can be achieved more efficiently by providing data on the individual appliance consumption to household.

Received November 23, 2023 / Accepted December 12, 2023
Corresponding author: Srđan Đorđević
University of Niš, Faculty of Electronic Engineering, Department of Electronics, Aleksandra Medvedeva 14, 18000 Niš, Republic of Serbia
E-mail: srdjan.djordjevic@elfak.ni.ac.rs

The appliance load monitoring can be realized by using several sensors installed on each target appliance or by using only one sensor attached to the service entry point. The first approach is referred to as Intrusive Load Monitoring (ILM), and the second as the Non-Intrusive Load Monitoring (NILM). In comparison with NILM, ILM is more accurate, but it is also more expensive and has a higher complexity. NILM has received much attention in the last years due to the developments in measurement instrumentation, computer technology, data communications and machine learning algorithms.

The seminal study in the field of NILM suggested the detection of appliance state transitions by examining steady-state changes in aggregate active power [2]. This method is very effective in recognizing high power linear loads but shows poor performance in detecting low-power and nonlinear loads. To overcome limits of NILM based on power measurements, some researchers used appliance features obtained through the analysis of current waveform. The information of the appliance states is present in features extracted from both the time domain and frequency domain of the current waveform. The application of the current harmonic content in the appliance load monitoring is suitable for the recognition of nonlinear loads. These methods are becoming attractive due to the rapid increase of the number of nonlinear devices in the power distribution system.

Srinivasan et al. [3] have shown through experimental evaluation the effectiveness of the NILM method that uses current harmonic vectors in load disaggregation. This method has acceptable load recognition accuracy but its applicability is limited to a small number of household devices. This problem is caused by the fact that the number of appliance signatures depends on the number of target appliances N, as $2^N - 1$. In our previous study [4] we introduced a NILM method based on the harmonic analysis of the current whose applicability is not limited by the number of devices, as opposed to [3]. This method, by analyzing changes in the steady-state current harmonics, extracts features that correspond to each individual appliance. So, the complexity of the method is proportional to the number of target appliances.

This paper extends our previous research on load identification based on current harmonic analysis [4] by addressing the random fluctuation of the current harmonic vectors. The amplitude and phase of the steady-state current harmonics vary not only due to the fluctuations in the power supply voltage but also as a consequence of variations in electrical characteristics of devices. In order to reduce the impact of random signature fluctuations on the accuracy of the NILM system, we propose in this work a load identification algorithm which includes a feature selection stage between the feature extraction and appliance classification. This stage estimates signal-to-noise ratio of each individual appliance feature at the time of prediction. Based on these values, the algorithm identifies features that cannot be reliable detected and marked them as missed. Among the methods for dealing with missed features [5,6], we applied the method that discards missed appliance features from the feature vector.

The rest of the paper is organized as follows. The following section gives a short overview of the non-intrusive load monitoring. Section 3 is devoted to the appliance identification by using current harmonic vectors. The proposed NILM algorithm is explained in Section 4. The experimental procedure and main results are presented in Section 5. Finally, the conclusion is given in Section 6.

2. NON-INTRUSIVE APPLIANCE LOAD MONITORING

The NILM solution for load disaggregation involves the following three stages: data acquisition, feature extraction and appliance classification. The purpose of the data acquisition stage is to collect aggregated load measurements at an appropriate sampling frequency. In the feature extraction stage electrical parameters which uniquely characterize the operational states of appliances are extracted from the raw current and voltage data. A well-known problem with NILM is choosing the appropriate load signature. Namely, high degree of load recognition accuracy can be attained only if the load signatures of the target appliances are well separated in a given signature space. Load signatures can be extracted from the steady-state signal or from the transient signal. The extraction of steady-state load signatures represents a lower-cost solution since it requires low-frequency sampling used in conventional smart-meters as opposed to the transient features which are realizable only with the additional hardware.

The final phase in a NILM system is the classification of the target appliances. A variety of supervised and unsupervised learning approaches have been proposed to solve this issue. The main advantage of the unsupervised learning algorithms is that they do not contain a training phase, which is error-prone and requires user intervention. These methods create a model of appliance behavior which enables the separation of the appliance level data from the aggregated load data. Supervised learning algorithms deal with the task of load desegregation in two different ways: as a pattern recognition problem or as an optimization problem. The pattern recognition approach has attracted more attention of the NILM researchers because optimization methods are vulnerable to errors in the presence of unknown devices.

3. NONINTRUSIVE IDENTIFICATION OF APPLIANCES USING CURRENT HARMONIC VECTORS

The most commonly used steady-state load signature is the active power. In order to improve the accuracy of load disaggregation active power is usually combined with reactive power or with different time and frequency domain characteristics of voltage and current [7-10]. Usually, a load signature, which is suitable for one class of appliances does not perform well in discerning the other categories of loads. The standard solution to this problem is to include more appliance features in the load signature. A comprehensive study based on this concept has been conducted by Liang et al [11,12]. Some authors have suggested the use of non-traditional features to improve the performance of load disaggregation independently of the classification algorithm [13].

The development of a load identification system is nowadays a challenging task because of a rapid increase in the type and number of household appliances. One of the solutions is a load signature composed of current harmonic vectors, which is intended to improve the segregation of the most commonly found class of appliances, small non-linear loads [8]. The main problem with this method is that it requires harmonic signatures with respect to all possible combinations of the target appliances. The amount of computation and the amount of data grow exponentially with the number of devices. Therefore, this algorithm is feasible only for a small set of target appliances.

The NILM method presented in this paper uses changes in current harmonic vectors to identify appliance operations. The application of this load signature in the load monitoring was considered in our previous studies [4, 14]. Its novelty is the use of the current harmonic content in an event-based load disaggregation method. The main advantage of the event based

NILM algorithm is that its complexity is proportional to the number of target devices, N, as opposed to the non-event based one whose complexity is 2^N.

A necessary prerequisite for the implementation of an event-based method is that appliance features meet the feature-additive criterion. The additivity of the current harmonic vectors can be easily proved from the fact that appliances are connected in parallel. According to the Kirchoff's Current Law, harmonic components of the aggregate current can be expressed in terms of the harmonic components of the individual appliances as follows:

$$\bar{I}_h = \sum_{k=1}^{K} \bar{I}_h^k \quad h=1,2,3... \tag{1}$$

where: $\bar{I}_h = |\bar{I}_h| \measuredangle \Theta_h$ are the vectors of the total current, $\bar{I}_h^k = |\bar{I}_h^k| \measuredangle \Theta_h^k$ represent harmonic current vectors of the k-th device, h is the order of the harmonic, K is the number of appliances.

In order to utilize the additivity of the current harmonic phasors for the classification, we use rectangular form of the current vectors, as follows:

$$\mathrm{Re}\{\bar{I}_h\} = \sum_{k=1}^{K} \mathrm{Re}\{I_h^k\} \quad \mathrm{Im}\{\bar{I}_h\} = \sum_{k=1}^{K} \mathrm{Im}\{I_h^k\} \tag{2}$$

A potential challenge of the current harmonics change method is that changes of the current harmonic vectors can be easily masked by the noise. In order to deal with this problem, it is necessary to identify harmonic features in testing instances whose values cannot be reliably determined. The implementation of this task involves the calculation of the signal-to-noise ratio with respect to each current harmonic vector present in the appliance signature.

4. ADAPTIVE FEATURE SELECTION IN THE LOAD DISAGGREGATION

The effectiveness of an NILM algorithm to recognize an appliance state transition is affected by the power network fluctuations and depends on the currently operating devices. This problem is particularly noticeable in the identification of low-power appliances, because their usage signal is easily overwhelmed in the presence of a large aggregated power signal. In this context, it would be of interest to determine a more optimal set of features around the event point. Fig. 1 gives an illustration of this phenomenon. It shows the waveforms of the three high-order current harmonics (3^{rd}, 5^{th} and 7^{th}), recorded during switching-on a

Fig. 1 Current harmonic waveforms related to switching on a fluorescent lamp

fluorescent lamp while the air conditioner was operating in the background. It is noticeable that the seventh current harmonic is irrelevant for the classification as it is overwhelmed with noise.

In an event-based method, the input of the classifier is the difference of an appliance feature around the event point. For this purpose, it is necessary to collect feature samples in separate observation windows, as shown in Fig. 2. Both observation windows are in steady-states, one before and the other after an event. The signal is calculated as the difference between the mean values of an attribute measured after and before an event as:

$$\Delta f_n = \frac{1}{|W_2|}\sum_{s\in W_2} f_n(s) - \frac{1}{|W_1|}\sum_{s\in W_1} f_n(s) \qquad \forall f_n \in F \tag{3}$$

where: n is the index of the appliance feature, F is the set of available features, s is the sample index, $f_n(s)$ is value of the feature f_n measured at the sampling points, Δf_n is the step change of the n-th feature, W_1 and W_2 are windows of samples taken before and after the event, respectively.

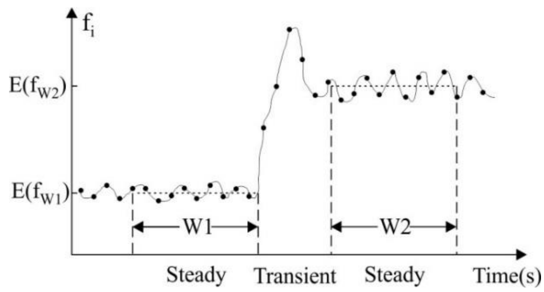

Fig. 2. Edge detection process

Noise is measured as the standard deviation of an attribute's values during a specified time interval, as:

$$\sigma(f_n) = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}\left(f_n(s) - E(f_n)\right)^2} \qquad \forall f_n \in F \tag{4}$$

where $E(f_i)$ is mean value of appliance feature f_n calculated as:

$$E(f_n) = \frac{1}{S}\sum_{i=1}^{S} f_n(s) \tag{5}$$

There are several different metrics used to score features according to their importance. In this study, we perform a feature selection based on the signal-to-noise ratio to eliminate irrelevant features. To access the relevance of the data, we use the signal to noise ratio defined as follows:

$$SNR(f_n) = \frac{\Delta f_n}{\sigma(f_n)} \tag{6}$$

The most common criterion for the step change of the signal that can be reliably detected is that the signal-to-noise ratio is greater than 3 [15]. According to this criterion, the set of relevant appliance features *F'* is:

$$F' = \{ f_n \in F \mid SNR(f_n) > 3 \} \tag{7}$$

where: *F* is the set of available appliance features

The proposed NILM method automatically identifies attributes whose instances do not provide relevant information. Since these features do not contribute to the load recognition they can be classified as missing features. The presence of missing values in the training or testing dataset is a common problem in the field of machine learning. There are two approaches for dealing with missing data at the prediction time: discarding [16] and predicting missed features [17].

In this study we applied the discarding of the missed features. This approach is computationally expensive in the case when the machine learning algorithm requires a separate model for each potential combination of omitted features. As opposed to these machine learning algorithms, the naive Bayes algorithm, which is applied in the proposed NILM method, can be easily modified to work with any subset of features as input data.

The block diagram of the proposed NILM algorithm is shown in Fig. 3. This algorithm is an extension of the conventional NILM algorithm whose data flow phases are: acquisition, event detection, feature extraction and classification. Two additional stages enable dynamic selection of a set of features that is the most appropriate for the load disaggregation. The proposed NILM system continuously extracts features of interest from the composite load signal. The appliance features are calculated from the predefined number of samples in two observation windows during the steady-state operation. After the detection of an on/off appliance event, the average value and variance of the features are calculated from the values collected during two steady-state periods, one that precedes and the other that follows the event point. These data are then used for the selection of useful features from the set of available appliance features.

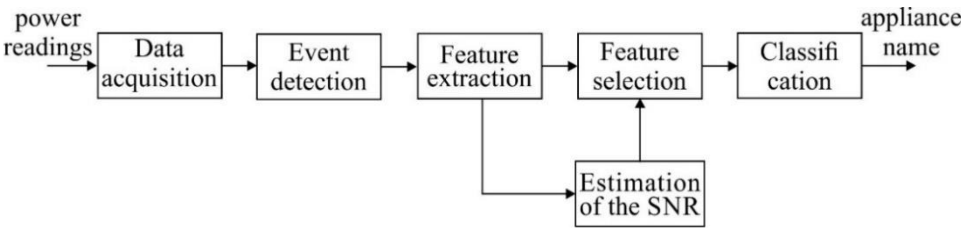

Fig. 3. Block diagram of the proposed NILM process

5. Experimental Procedure

We tested the proposed method on a group of eight widely used single-phase household appliances. The selected group of appliances includes the following electronic appliances: personal computer, TV, laptop, tablet computer, monitor. The remaining considered appliances are mixer, stove and a fluorescent lamp.

Current and voltage measurements were performed using a power analyzer Fluke 435 Series II. Voltage and current waveform are sampled at a frequency of 200 kHz and samples are quantized with a resolution of 16 bits. The schematic and photo of the experimental setup are shown in Fig. 4 and Fig. 5, respectively. As Fig. 4 shows, the analyzer was wired in the configuration for the single-phase measurements. Current measurements were collected by using the current clamp Fluke i2000 Flex put around the connector of a phase. Tested appliances are connected in parallel to the same power outlet. The recorded training and testing data were numerically processed using Matlab software.

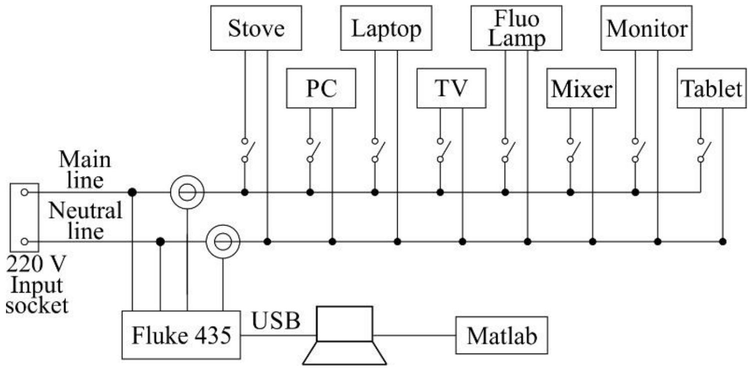

Fig. 4. Schematic of the experimental setup

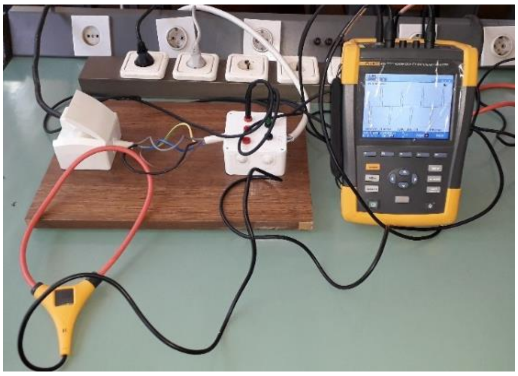

Fig. 5. Photo of the experimental setup

In this study, we use changes in the real and imaginary part of the harmonic current phasors as a load signature. The classification was performed by using naive Bayes classifier because it is a simple technique for classification which can be used very efficiently in the non-intrusive load monitoring. The training data for an appliance was recorded during 100 seconds of its steady-state operation. Each individual appliance's signature is created by measuring the magnitude and phase of the first five odd current harmonics and calculating their real and imaginary part. Table 1 represents a summary of the training data for Bayes classifier, which includes the mean and standard deviation for each appliance feature.

Table 1 Mean and standard deviation of the real and imaginary part of harmonic current vectors

Appliances	Current harmonics (mA)									
	1 st		3rd		5th		7th		9th	
	Re	Im	Re	Im	Re	Im	Re	Im	Re	Im
Fluo. Lamp	322.3 ±5.7	-658 ±27.9	60.5 ±2.2	-46.1 ±2.7	-15.1 ±0.7	-2.0 ±2.4	-1.6 ±2.1	8.8 ±0.5	4.8 ±0.2	0.3 ±1.3
Laptop	115.4 ±18.6	39.3 ±6.0	-157.0 ±13.1	75.3 ±13.6	108.9 ±2.9	99.8 ±16.3	47.6 ±8.8	-103.0 ±12.4	-6.3 ±15.2	84.5 ±3.9
Mixer	358.6 ±7.2	58.2 ±2.6	-6.1 ±1.0	17.8 ±1.4	-10.4 ±0.8	-3.6 ±0.5	6.4 ±0.5	-2.7 ±0.6	-1.3 ±1.1	0.06 ±0.4
PC	82.6 ±10.5	-15.3 ±1.8	-130 ±8.0	20.3 ±4.7	104.3 ±4.5	28.4 ±6.1	69.5 ±1.5	-31.4 ±5.7	35.23 ±2.7	28.3 ±4.0
Tablet	3.5 ±1.8	4.8 ±2.5	-20.5 ±6.2	-3.5 ±1.3	19.9 ±5.3	6.3 ±1.9	18.0 ±5.6	-9.4 ±3.1	15.7 ±2.9	13.2 ±2.8
TV	729.1 ±13.6	71.4 ±3.6	-67.4 ±2.5	41.45 ±5.5	-34.1 ±0.7	2.8 ±2.1	-2.2 ±0.3	4.1 ±0.5	10.7 ±4.1	-0.9 ±0.5

We conducted the experiments by switching ON one of the appliances under testing while a specified combination of the appliances were operating in steady-state. Since the appliances that operate in the background affect the recognition accuracy, events are associated not only to the switching device but also to the group of background appliances. Each of these events was repeated and measured 60 times to determine its recognition accuracy.

Each measurement trial contains two 5-second observation windows, one before and another after the event. In order to avoid transient parts of the current waveform the second analysis window starts ten seconds after the event occurring, which allows the appliance to settle into steady-state. The magnitude and phase angle of the first five odd current harmonics were estimated from a 0.25 second array of current signal samples. In total, there were 40 readings (20 in each observation window) in a measurement trial. The real and imaginary parts of the specified current harmonics were calculated with respect to each reading.

We considered the classification accuracy in two cases. First, we perform classification by using all initial features. In this case, the input variables to the classifier are real and imaginary part of the first five odd current harmonics. The results of these classifications are presented in a confusion matrix in Table 2. The table indicates that PC, mixer and tablet are most frequently misclassified. The mean classification accuracy was 85,9%.

The noise in the features of the classification model were calculated from the energy analyzer measurements during 1 minute period before the switching event. When time series of the current harmonic magnitudes and current harmonic phases have been collected, they are used to calculate time series of the real and imaginary part of the odd harmonic vectors. Then, the noise in each appliance feature is estimated through the standard deviation of its time samples.

Table 3 shows the signal-to-noise ratio related to the devices that operate in the experimental setup. After removing features which are too noisy to provide relevant information, according to the criterion given in (6), we obtain the feature vector used for classification. So created feature vectors associated to each considered case, are listed in Table 4.

Table 2 Classification confusion matrix using the initial set of appliance features

Background appliances	Switching appliance	Predicted load: Laptop	Mixer	Fl. Lamp	Tablet	PC	TV	Accuracy[%]
PC+TV+ Monitor	Laptop	59				1		98.33
	Mixer		57		3			95
	Fl. Lamp			45		15		75
	Tablet				49	11		81.66
TV+Stove+ Fluo. Lamp	Laptop	60						100
	Mixer		60					100
	PC				57	3		5
	Tablet				60			100
PC+Monitor +Fluo. lamp	Laptop	60						100
	Mixer		58		2			96.66
	TV	2			2		56	93.33
	Tablet				52	8		86.66

Table 3 Signal to noise ratio of the appliance features

Background appliances	Switching appliance	Current harmonics: 1 st Re	1 st Im	3rd Re	3rd Im	5th Re	5th Im	7th Re	7th Im	9th Re	9th Im
PC+TV+ Monitor	Laptop	27.99	48.59	29.18	23.19	27.65	20.58	22.98	18.44	1.35	14.19
	Mixer	32.36	43.95	0.12	4.83	2.42	0.67	2.54	0.66	2.94	0.29
	Fl. Lamp	48.84	325.2	0.67	19.59	5.46	7.65	8.77	6.65	0.04	8.21
	Tablet	5.98	4.85	6.69	1.89	8.37	2.07	13.1	2.21	13.81	2.5
TV+Stove+ Fluo. Lamp	Laptop	23.15	157.9	71.75	33.64	43	37.66	22.19	23.68	3.67	79.06
	Mixer	21.51	36.21	3.17	9.05	2.51	0.67	1.32	0.27	0.26	0.28
	PC	5.14	1.01	68.92	3.48	21.68	3.05	7.62	0.74	9.91	0.59
	Tablet	1.07	2.79	21.07	3.98	11.06	2.83	5.63	2.51	14.91	8.95
PC+TV+ Fluo. lamp	Laptop	43.45	25.36	40.52	28.76	40.87	27.75	30.58	24.88	0.82	20.41
	Mixer	45.6	28.89	0.31	5.47	2.3	0.83	2.55	0.63	1.67	0.27
	TV	42.36	18.11	3.67	3.12	3.49	1.11	1.99	0.92	10.18	0.47
	Tablet	8.28	2.16	8.16	2.6	9.11	2.7	9.96	2.77	6.52	2.92

Table 4 Vector of relevant appliance features for individual cases

Background appliances	Switching appliance	Relevant appliance features
PC+TV+ Monitor	Laptop	Re{Ih1} Im{Ih1} Re{Ih3} Im{Ih3} Re{Ih5} Im{Ih5} Re{Ih7} Im{Ih7} Im{Ih9}
	Mixer	Re{Ih1} Im{Ih1} Im{Ih3}
	Fl. Lamp	Re{Ih1} Im{Ih1} Im{Ih3} Re{Ih5} Im{Ih5} Re{Ih7} Im{Ih7} Im{Ih9}
	Tablet	Re{Ih1} Im{Ih1} Re{Ih3} Re{Ih5} Re{Ih7} Re{Ih9}
TV+Stove + Fluo. Lamp	Laptop	Re{Ih1} Im{Ih1} Re{Ih3} Im{Ih3} Re{Ih5} Im{Ih5} Re{Ih7} Im{Ih7} Re{Ih9} Im{Ih9}
	Mixer	Re{Ih1} Im{Ih1} Re{Ih3}Im{Ih3}
	PC	Re{Ih1} Re{Ih3} Re{Ih5} Re{Ih7} Re{Ih9}
	Tablet	Re{Ih3} Re{Ih5} Re{Ih7} Re{Ih9} Im{Ih9}
PC+TV+ Fluo. lamp	Laptop	Re{Ih1} Im{Ih1} Re{Ih3} Im{Ih3} Re{Ih5} Im{Ih5} Re{Ih7} Im{Ih7} Im{Ih9}
	Mixer	Re{Ih1} Im{Ih1} Im{Ih3}
	TV	Re{Ih1} Im{Ih1} Re{Ih3} Im{Ih3} Re{Ih5} Re{Ih9}
	Tablet	Re{Ih1} Re{Ih3} Re{Ih5} Re{Ih7} Re{Ih9}

Table 5 shows the confusion matrix for the classification based on the proposed feature selection method. According to the experimental results, it is apparent that the proposed feature selection method improves the classification accuracy.

Table 5 Classification confusion matrix using the proposed NILM algorithm

Background appliances	Switching appliance	Predicted load						Accuracy[%]
		Laptop	Mixer	Fl. Lamp	Tablet	PC	TV	
PC+TV+ Monitor	Laptop	59				1		98.33
	Mixer		60					100
	Fl. Lamp			45		15		75
	Tablet	9			51			85
TV+Stove+ Fluo. Lamp	Laptop	60						100
	Mixer		60					100
	PC					60		100
	Tablet				60			100
PC+TV+ Fluo. lamp	Laptop	60						100
	Mixer		60					100
	TV					60		100
	Tablet	6			54			90

Discarding of the irrelevant features from the feature vector improves recognition accuracy of the mixer, PC, tablet and TV. The improvement is especially noticeable in the identification of PC. Among the target appliances, laptop and fluorescent lamp were identified with the same accuracy before and after removing of irrelevant features.

6. Conclusion

Fluctuations of the appliance features, which are caused by the variations in the supply voltage and appliances' electrical characteristics, can lead to the reduced classification accuracy of the load disaggregation algorithm. This problem is particularly prominent in the event based NILM methods. In order to deal with noisy features, we proposed a NILM algorithm which estimates the Signal-to-Noise Ratio of each individual appliance feature and remove irrelevant attributes. The experimental results have confirmed that the suppression of the appliance features which do not contain valuable information at the time of prediction increases the classification accuracy of a NILM system.

The proposed solution for the feature selection takes into consideration only the relevancy and ignores the redundancy of the features. However, it provides a simple and computationally effective way to improve the accuracy of the load desegregation.

Acknowledgement: *This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia. This paper was realized as a part of the project TR 32019, funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.*

REFERENCES

[1] P. Shaikh, N. Nor, P. Nallagownden, I. Elamvazuthi, T. Ibrahim, "A review on optimized control systems for building energy and comfort management of smart sustainable buildings," *Renewable Sustainable Energy Reviews*, vol. 34, pp. 409–429, 2014, doi:10.1016/j.rser.2014.03.027.

[2] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992, **doi:** 10.1109/5.192069.

[3] D. Srinivasan, W. S. Ng, A. C Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 780-791, Jan. 2006, doi: 10.1109/TPWRD.2005.852370.

[4] S. Djordjevic, M. Simic, "Nonintrusive identification of residential appliances using harmonic analysis," *Turkish Journal of Electrical Engineering & Computer Sciences,* vol. 26, no. 2, pp. 780-791, 2018, doi:10.3906/elk-1705-262.

[5] M. Saar-Tsechansky, F. Provost, "Handling Missing Values when Applying Classification Models," *Journal of Machine Learning Research*, vol. 8, pp. 1625-1657, 2007, doi:10.5555/1314498.1314553.

[6] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, O. Tabona, "A survey on missing data in machine learning," *Journal of Big Data*, vol. 8, May 2021, doi: 10.21203/rs.3.rs-535520/v1.

[7] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, P. Armstrong, "Power signature analysis," *IEEE Power Energ. Mag.* 2003, vol. 1, no. 2, pp. 56–63, doi: 10.1109/MPAE.2003.1192027.

[8] A. Zoha, A. Gluhak, M. Imran, S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey," *Sensors* 2012, Vol. 12, pp. 16838-16866, Dec. 2012, doi.org/10.3390/s121216838.

[9] C. Agnim, A. Albicki, "Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content." Proceedings of the 17th IEEE Conference on Instrumentation and Measurement Technology IMTC2000, vol. 1, pp. 24-29, 2000.

[10] R. S. Prasad, S. Semwal, "A Simplified New Procedure for Identification of Appliances in Smart Applications," *IEEE International System Conference (SysCon2013)*, Orlando, USA, pp. 339-344, doi: 10.1109/SysCon.2013.6549903.

[11] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load Signature StudyPart I: Basic Concept, Structure, and Methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 551–560, April 2010, doi:10.1109/TPWRD.2009.2033799.

[12] J. Liang, S. K. K. Ng, G. Kendall and J. W. M. Cheng, "Load Signature Study—Part II: Disaggregation Framework, Simulation, and Applications," in *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 561-569, April 2010, doi: 10.1109/TPWRD.2009.2033800.

[13] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," in *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 23-31, February 2012, doi: 10.1109/TCE.2012.6170051.

[14] S. Djordjevic, M. Simic, "Nonintrusive identification and type recognition of household appliances based on the harmonic analysis of the steady-state current," - Electrical Engineering, no. 5, pp. 3319-3328, doi:10.1007/s00202-023-01888-2.

[15] K. Wasielewska, D. Soukup, T. Čejka, J. Camacho, "Evaluation of the Limit of Detection in Network Dataset Quality Assessment with PerQoDA," In: Koprinska, I., *et al.* Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2022. Communications in Computer and Information Science, vol 1753. Springer, Cham., doi:10.1007/978-3-031-23633-4_13.

[16] D. Schuurmans and R. Greiner, "Learning to classify incomplete examples," In Computational Learning Theory and Natural Learning Systems: Addressing Real World Tasks, pages 87–105. MIT Press, 1997.

[17] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," Springer Verlag, New York, August 2009, doi: 10.1007/978-0-387-84858-7.