IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 14, 2019, accepted October 8, 2019, date of publication October 17, 2019, date of current version October 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948079

Non-Intrusive Load Identification Method Based on Improved KM Algorithm

YONG XIAO[1], YUE HU[2], HENGJING HE[1], DONGGUO ZHOU[2], YUN ZHAO[1], AND WENSHAN HU[2]

[1]Electric Power Research Institute, Guangzhou 510663, China
[2]School of Electrical and Automation, Wuhan University, Wuhan 430072, China

Corresponding author: Yong Xiao (xiaoyong@csg.cn)

This work was supported by the Southern Power Grid Corporation under Grant ZBKJXM20170079.

ABSTRACT To achieve load disaggregation in non-intrusive load monitoring (NILM) system, a load event matching method based on graph theory is proposed, which is built on the improved Kuhn-Munkras algorithm. In this method, firstly, an adaptive fitting method using time window is applied to detect the load whether it is switched on and/or off. Particularly, to avoid the fluctuation of load signatures, the kernel density estimation is then built by a number of the independent features of the load switching on, including the active and reactive power signatures. The distribution of load signatures is thereby obtained, allowing the load event to be classified by its features. The load matching, which is based on the improved KM algorithm, is then utilized to resolve the matrix formed by the matching probability of the load event. Similarly, load identification can also be realized by matching the features of events with the signatures in the database. Finally, the experimental results using datasets of our lab and REDD demonstrate that the proposed method can obtain the desirable result for load event matching, and promote the performance in load identification.

INDEX TERMS Non-intrusive, load event, load matching, KM Algorithm, load identification.

I. INTRODUCTION

A non-intrusive load monitoring (NILM) system, of which the measurement device is installed at the main electric power input, has become a mainstream technology to know the energy consumption of individual appliances in a household. Generally speaking, there are several significant advantages of NILM [1] in contrast to the intrusive load monitoring: (1) it's the convenient way to install at main electric power input, instead of the voltage sensor and current sensor in each load, thus reducing the cost [2] and increasing the security of measuring devices; (2) it can help build the ongoing smart grid since the smart meter technology of NILM enables the prediction of the power demand and the decision making for policy makers [3]; (3) it can achieve easy maintenance and extension when the new appliances join, and support the revamping of appliances. Therefore, the study of NILM becomes a hot trend for research institutions.

For NILM, the existing methods can be categorized into three major groups: optimization method, supervised and unsupervised methods. The optimization method [4], [5] is a straightforward way to solve the load disaggregation task, which is to compare the extracted load signatures with the features in database and obtain the minimal error by optimization strategy. In recent research, optimization model is improved by various methods, such as optimized support vector regression [6], optimized bird swarm algorithm [7], genetic algorithm [8], and particle swarm optimization method [9]. The supervised disaggregation methods require existing specific information of devices and need initial training phase. Particularly, it depends on the adequate labelled data for learning the model. Generally, the fundamental of this type of recognition methods is to transform the load signature into a feature space for classification [10], such as Bayesian classifiers [4], Support Vector Machine (SVM) [11], and Artificial Neural Network (ANN) [12] as well as its extensions [13], [14]. The unsupervised methods [15], which can operate without a priori information and reduce the intrusiveness of the training step, take the cluster as its strategy for recognition, including Hidden Markov Models (HMMs) [16] and its extensions [17]. Moreover, some literatures take several algorithms into consideration to promote the performance of NILM, thus obtaining the comprehensive result as discussed in [18].

The associate editor coordinating the review of this manuscript and approving it for publication was Hui Ma.

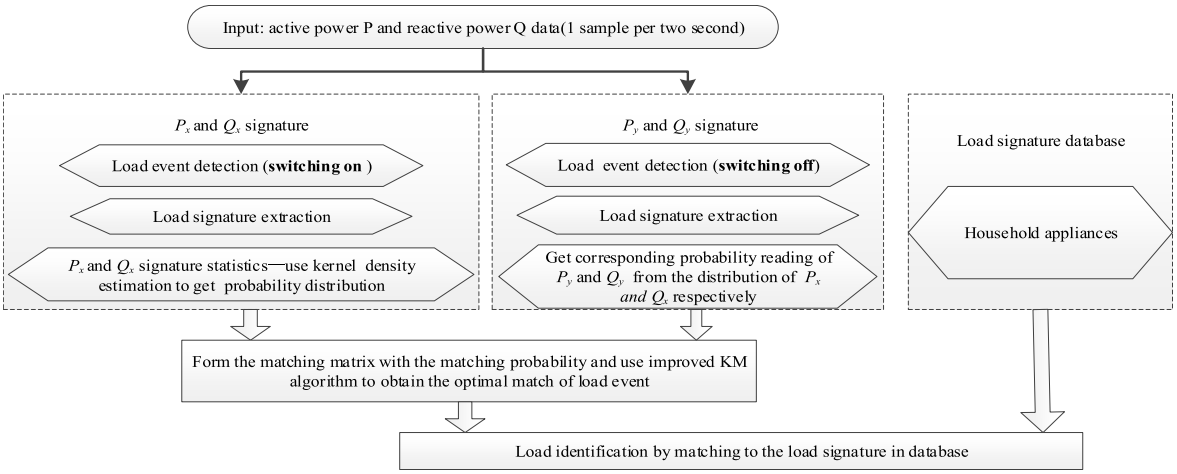


FIGURE 1. Flow chart of the proposed method.

Since there are many methods for NILM, each method has its ability to achieve load identification. However, some modern appliances such as intelligent air-conditions are taken into using in the smart home. This particular type of load may not have the fixed signatures, thus making some of NILM methods fail to identify these loads. Therefore, several methods are proposed to deal with the modern appliances, such as deep learning-based technique [19], particle filtering method [20] and KL-based expansion [21]. These methods in general can also cope with the following appliances [22], [23]: single state, multi-state, and continuously varying appliances [24], among which multi-state appliances are more common and more complicated to be identified. Nevertheless, those above methods may fail to match the load events, especially when multiple appliances work together. The optimization method is not capable to identify the aggregated appliances as the number of load increases, especially for loads with overlapping signatures. Though the supervised methods have high accuracy, it's hard to be implemented for practical household data since they need a lot of training. As for unsupervised methods, they usually have high computational complexity and are prone to fall into local optimum.

In order to alleviate the complexity of load disaggregation, this paper presents a different approach based on graph theory. The graph-based method is potential to solve the problem of NILM, which is used in building mathematic model [25]–[29]. The proposed method transforms the matching problem of load on-off events and load identification process into a bipartite graph model. With the kernel density probability as the weights of the bipartite graph, the improved Kuhn-Munkras algorithm is then used to solve the matching matrix, thus obtaining the optimal matching and finding the type of load in the database.

The rest of the paper is organized as follows. Section II introduces the procedures of the proposed method, including load event detection, load signature representation, statistical model building of load signature, load event matching, and load recognition. The experiments for verification are then tested by using our lab and the REDD datasets in section III. Finally, the conclusions of the work are drawn in the last section.

II. MATERIALS AND METHODS

The load event almost consists of the status of switching on and switching off, i.e., the switching-off event is usually in pair with the switching-on event. In this paper, this match mechanism is utilized to implement the load event matching. Generally, the proposed method consists of load event detection, load signature representation, signature modelling and the procedures using improved KM algorithm to do the matching of load events as well as load recognition, as seen in Fig. 1. It should be noted that in Fig.1, the "P_x and Q_x" denotes the active power and reactive power signature of the load when it is switched on, and the "P_y and Q_y" denotes the power signature when the load is switched off. In the following sections, we will discuss them in detail.

A. LOAD EVENT DETECTION

Load event detection is usually performed to get the status change of appliance whether it's switched on or off. Roughly speaking, it is similar to the edge detection in load power curve, as seen in Fig.2. The active power P undergoes almost a signification change, thus considering as the appliance is switched on or off, as seen the blue curve for example.

In this work, the adaptive method for load event detection is proposed by using the time window. The fundamental idea is to fit the active power data in the window. Assuming that there is a load event happened, the data in two sides of the occurrence point must be quite different, whereas in each side, the power data does not change significantly. Thus, the load detection can be implemented by the fitting method.

Let $\{x_i\}_{i=1\sim N}$ represents the data samples in the time window, which can be divided into two categories. The load

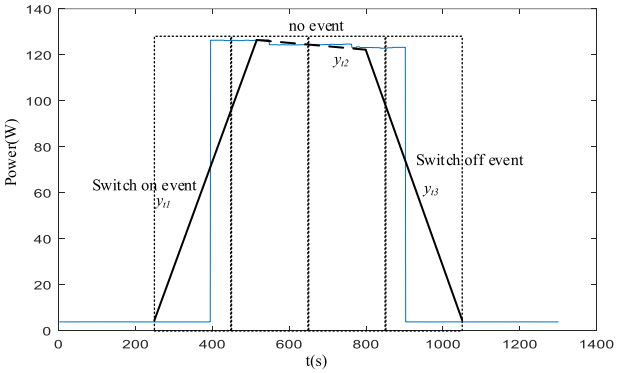


FIGURE 2. Fitting method for load event detection.

events in the window are detected by:

$$y_t = \beta_0 + \beta_1 x_t + e_t, \quad t \text{ is in the time window}, \tag{1}$$

where t is a time point; β_0, β_1, e_t denote a constant term, a slope term and an error term respectively. Obviously, when the appliance works in steady state, its slope in the window is usually considerably stable. On the contrary, when there is an appliance being switched on or off, it will make relatively large error e_t, and the β_1 will vary during the window moves, as seen the black curve in Fig.2.

The occurrence point can be then obtained by the fitting method. Let the data samples in the detection window be divided into two classes by the time k: class $C_0\{x_1, x_2, \ldots, x_k\}$ and class $C_1\{x_{k+1}, x_{k+2}, \ldots, x_L\}$, and the objective function of the method is defined as:

$$min_k \left\{ \sum_{i=1}^{k} (x_i - m(C_0))^2 + \sum_{i=k+1}^{L} (x_i - m(C_1))^2 \right\} \tag{2}$$

where L is the length of the window; $m(C_0)$ and $m(C_1)$ are the centers of the two classes as follows

$$m(C_0) = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{3}$$

$$m(C_1) = \frac{1}{L-k} \sum_{i=k+1}^{L} x_i \tag{4}$$

It's noted that when there is a curve appearing the climbing status, the occurrence point is determined if $\sum_{i=1}^{k} (x_i - m(C_0))^2$ is larger than a threshold.

B. LOAD SIGNATURE REPRESENTATION

Load signatures [30] are inner characteristics of the appliance during its switching on or off and its working. In general, the active power P is considered as the physical quantity of greatest interest [31], since it can directly reflect the energy consumption. The reactive power Q is another important parameter which can determine the type of appliance, such as inductivity, capacitance or resistance of electrical devices.

Let $v(t)$ and $i(t)$ are the transient voltage and current in time t, and the active power and reactive power can be defined as follows:

$$P = \sum_{t=0}^{T} v(t)\, i(t) \tag{5}$$

$$Q = \sum_{t=0}^{T} v\left(t + \frac{1}{4}T\right) i(t) \tag{6}$$

where T is a period obtained from the sampling frequency.

Usually, these two features are the typical steady-state load signatures used in load identification [32], [33]. What's more, the combination of feature P and Q constitutes the P-Q plane [32], which has also become the most commonly used plane in load identification.

C. STATISTICAL MODEL OF LOAD SIGNATURE

In this part, the statistical model is built in order to avoid the obtained signatures deviating from the actual signatures of load event, thus promoting the performance of anti-interference of voltage and current fluctuations.

Let the load events be classified into two types. One is the events of switching on, which is defined as a node set X, and the other is the load events of switching off, which is defined as a node set Y. The matching problem between these two sets can be then solved by graph theory, as seen in the next section.

Let $\{A_i \ (i = 1 \sim n)\}$ represents the appliances needed to be identified in the test scene. The active power of appliances switched on is collected as a set $P_x = \{P_{x1}, P_{x2}, \ldots P_{xn}\}$. Among them, P_{xi} is considered as the independently obtained active power from the appliance i. In general, the obtained value of P_{xi} is followed by Gauss distribution. Similarly, the set of active power changes corresponding to switching off can also be obtained through load event detection, and denoted as $P_y = \{P_{y1}, P_{y2}, \ldots P_{yn}\}$. The reactive power changes of switching-on event is defined as set $Q_x = \{Q_{x1}, Q_{x2}, \ldots Q_{xn}\}$, and the reactive power changes of switching-off event is defined as set $Q_y = \{Q_{y1}, Q_{y2}, \ldots Q_{yn}\}$.

In order to estimate the similarity of signatures P and Q respectively during the matching process, the Parzen window algorithm of kernel density method is used to build the probability density $f(x)$.

Supposing that x_1, x_2, x_3 ... x_n are a set of independent samples, the kernel density of set $\{x\}$ is estimated as:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{7}$$

where n denotes the number of samples, h represents the window width, and $K(\cdot)$ is the kernel function. Generally, kernel function decides the distribution characteristic, and window width can affect the smoothness of the density function. Here, Gaussian function is chosen in this work as follows:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}u^2\right)} \tag{8}$$

It is worth noting that according to Gaussian distribution $N(\mu, \sigma)$, the probability density function $f(x)$ is closely related to the standard deviation σ of the samples. So the optimal width of window h is estimated as:

$$\hat{H} = 1.06\sigma n^{-1/5} \tag{9}$$

D. LOAD EVENT MATCHING

1) KM ALGORITHM

KM algorithm [34], aimed at finding the best matching, is a method based on bipartite graph. Given two acquired set of measurements, $X = \{X_1, X_2, \ldots X_n\}$ and $Y = \{Y_1, Y_2, \ldots, Y_n\}$, a graph G can be defined by the matching nodes, in which each node of X has its corresponding node in Y.

Assuming that M is a match of a graph G, the KM algorithm is employed to get the optimal matching that makes the edge weights $\sum \{\omega\}$ of M maximum. Without loss of generality, the procedure of KM algorithm can be concluded as follows:

1: < initializing the value of feasible vertex labeling >

2: for $i = 1 \rightarrow n$ do

3: $l_x(x)$ and $l_y(y)$ record the vertex label values of nodes in set X and Y respectively

4: $l_x(x_i)$ is set as the maximum weight of all edges $e(x_i, y_i)$ associated with x_i, $l_y(y_i) = 0$

5: end for

6: < find the perfect matching of equal subgraph with Hungarian algorithm >

7: < check: $l_x(x_i) + l_y(y_i) \geq \omega(x_i, y_i)$ >

8: if not, revise the value of the feasible vertex label for accessed vertex x, its feasible label is subtracted from d; for all visited vertex y, the feasible label is added d:

$$\mathrm{d} = min_{x\epsilon X, y\varepsilon Y} \{l(x) + l(y) - \omega(x, y)\} \tag{10}$$

9: end if

10: iterate step6-step10 until the perfect matching of the equal subgraph is found.

2) LOAD EVENT MATCHING BASED ON IMPROVED KM ALGORITHM

Although the KM method is a good solution for matching, it is necessary to do some modification for real problems, especially in the weight of matching.

Assumed that the two bipartite graphs $G(X, Y, W)$ and $H(X, Y, W)$ are constructed to represent matching process of active power and reactive power respectively in this work, where X and Y have been defined as the set of the load switching-on event and load switching-off event respectively; W denotes the weight.

Let $M(x_i, y_j)$ be a matching of X_i and Y_j. The matching probability of load events corresponding to X, Y is indicated by the weight $\omega_G(x_i, y_j)$ and $\omega_H(x_i, y_j)$, which belong to $M_G(x_i, y_j)$ and $M_H(x_i, y_j)$ respectively. In general, the value of the probability can reflect the similarity between the event X_i and Y_j. Therefore, the best solution of matching obtained by the KM algorithm ensures that the sum of the edge weights reaches maximum, that is, the load switching-on signatures are most similar with switching-off signatures.

Let the matching probability of active power and reactive power be a two-dimensional matrix $[f_{ji}]_{n\times n}$ and $[h_{ji}]_{n\times n}$ respectively, where an element in the j-th row of the matrix denotes the signature's probability of a switching-off event, and the values are obtained from the signatures' distribution of the switching-on events. To avoid the fluctuations of voltage and current which can make the matching probability of a single matrix $[f_{ji}]_{n\times n}$ or $[h_{ji}]_{n\times n}$ cause errors, the optimal load matching model can be reformulated as:

$$\{M(i, j)\} = \max(\alpha \sum_{i,j=1}^{n} f_{ji} + \beta \sum_{i,j=1}^{n} h_{ji}), \tag{11}$$

where $\{M(i, j)\}$ is the matching result; i represents the sequence of switching-on events; j represents the sequence of switching-off events; and the index i and j are taken only once during calculation. α and β, which are heuristically chosen, are the weight of $[f_{ji}]_{n\times n}$ and $[h_{ji}]_{n\times n}$ respectively, with $(\alpha + \beta = 1)$ to trade off the factor of active power and reactive power. For convenience, the matrix $[k_{ji}]_{\mathrm{n\times n}}$ is defined as the matching matrix of $\{M(i, j)\}$. Thus, the impact of voltage and current fluctuations can be alleviated during the solution using the improved KM method.

In order to determine the values of α and β, the uncertainty of multiple samples belonging to load signature S is defined as $\mu(S)$

$$\mu(s) = \sqrt{\frac{\sum_{k=1}^{n} (s_k - D(S))^2}{n(n-1)}}, \tag{12}$$

where $D(S)$ is the average value of the signature S from multiple samples; and s_k $(k = 1 \sim n)$ is the independent sample k of signature S.

To distinguish load signatures of different appliances, the uncertainty μ of a signature is used to represent the average distance between the cluster that appliance a belongs to and the clusters of other devices, denoted as μ_a. Thus, the discriminability coefficient g of the signature S between appliance a and appliance b is defined as [35]:

$$g_{ab}(S) = \left\{ \begin{array}{ll} \dfrac{|D_a(S) - D_b(S)| - \mu_a(S)}{\max\{\mu_a(S), |D_a(S) - D_b(S)|\}} & a \neq b \\ 0 & a = b \end{array} \right\}, \tag{13}$$

where D_a and D_b are the average values of load signature belonging to a and b respectively. Notably, $g_{ab}(S)$ is zero when a and b are the same device. Then, the discriminability index G_{dis} of the load signature S can be obtained from "(13)," as:

$$G_{dis}(S) = \frac{\sum_{a=1}^{m} \sum_{b=1}^{m} g_{ab}(S)}{m^2}. \tag{14}$$

Here, m is the number of appliances, and the discriminability of load signatures can be indicated by the difference between discriminability index and one.

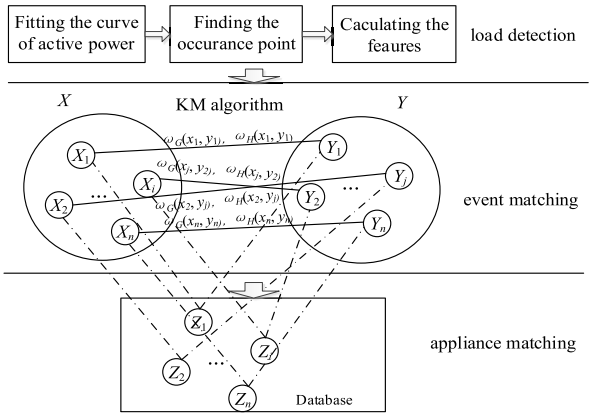


FIGURE 3. Framework of load recognition.

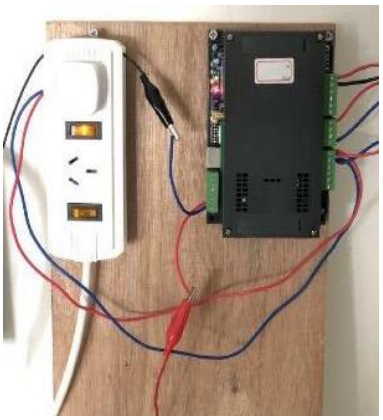


FIGURE 4. Measurement equipment of our lab.

Thus, the values of α and β can be obtained as:

$$\frac{\alpha}{\beta} = \frac{G_{Pdis}}{G_{Qdis}}, \tag{15}$$

where G_{Pdis} and G_{Qdis} are the discriminability index of active power and reactive power respectively. Since $\alpha + \beta = 1$, the corresponding weights of P and Q features can be finally worked out according to "(15),".

E. LOAD RECOGNITION

In our model, the best matching between the switching-on event and the switching-off event is achieved by the improved KM algorithm. Similarly, the load recognition also can be implemented through the matching between signatures of the load event and signatures in the database. Thus, the procedures of the improved KM method can be used to obtain the final identification. For more understanding, the recognition framework is illustrated in Fig. 3.

III. RESULTS AND DISCUSSION

In this section, we shall present the results of the proposed method by using our lab test data and REDD dataset for verification. Notably, α and β are chosen differently according to the different test scenario. The whole algorithm is implemented by the *C* language program in our equipment as seen in Fig. 4, of which the data sampling rate is 8 kHz. And it calculates the signatures per two seconds. The data and results are stored in the MySQL database for the test.

A. TEST SCENARIO

In this test, the data is from our lab, where the devices contain air conditioner, electric heating, electric kettle, TV, microwave oven, hair drier, rice cooker, and induction cooker. The signatures are obtained by switching on and off each appliance for several times, and the P-Q signatures of each appliance are shown in Fig. 5 and Fig. 6. It can be seen that the P-Q signatures have their own distributions. Particularly, some of them have partial overlaps, such as the signatures of induction cooker and electric kettle. The circles drawn in the Fig. 5 and Fig. 6 can help find the areas of the probability

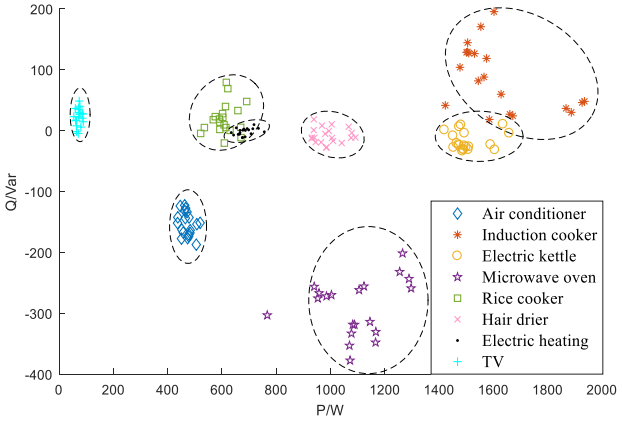


FIGURE 5. P-Q signatures of appliances during on status.

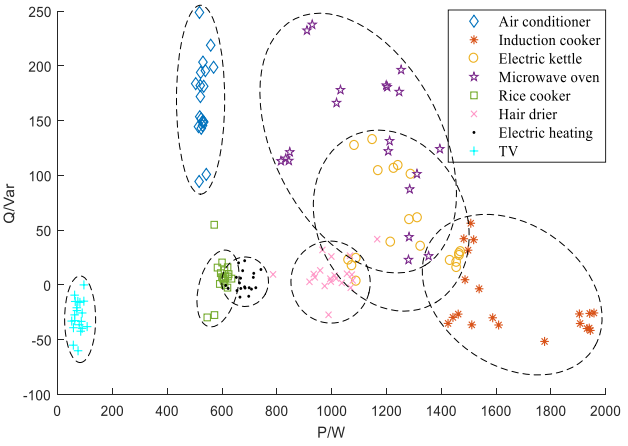


FIGURE 6. P-Q signatures of appliances during off status.

density. It can be seen that the distributions of some appliances are more centralized, whereas some are more dispersed. Here, the range of the active and reactive power information of single appliance is recorded in Table 1, and it's regarded as the load signature database.

Figure 7 illustrates a test by switching on and off the appliances, where the real matching sequence of the appliances

TABLE 1. Household load power information.

Load	Active power P(W)	Reactive power Q(Var)	Average active power(W)	Average reactive power(Var)
Air conditioner	435.9~558.6	-192.8~-115.2	467.3	-173.3
Induction cooker	1442.9~1944.4	33.7~140.5	1497.8	87.1
Electric kettle	1061.9~1636.3	-46.5~11.6	1405.2	-8.6
Microwave oven	832.4~1354.0	-372.9~-232.1	1207.9	-321.6
Rice cooker	535.4~669.2	-8.5~44.1	594.6	-2.8
Hair drier	786.3~1166.5	-21.1~4.0	994.1	-8.5
Electric heating	613.5~741.9	-11.9~22.1	684.7	-2.0
TV	51.4~107.7	8.9~36.6	73.0	22.8

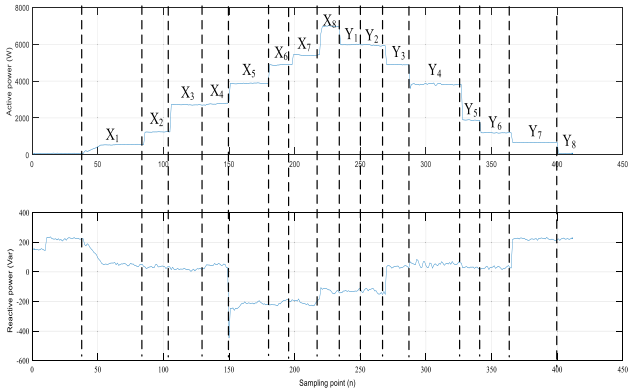


FIGURE 7. Active and reactive power changes of the tested appliances.

switching on and off are: X_1-Y_7, X_2-Y_6, X_3-Y_4, X_4-Y_2, X_5-Y_3, X_6-Y_1, X_7-Y_8, X_8-Y_5. In this sequence, X_1~X_8 represents the air conditioner, electric heating, electric kettle, TV, microwave oven, hair drier, rice cooker, and induction cooker, respectively. Notably, the fluctuation of signature occurs during the appliances work together as seen in Fig.5 and Fig.6. On the basis, α and β are set as 0.60 and 0.40 respectively according to "(15),".

Table 2 shows the load events detected by our method. It's noted that the length of the event detection window in this test is 10 sampling points, and the threshold of the slope term is set to be 20. However, the difference of active power around the occurrence point may deviate from the actual load signature to a certain extent, especially for appliances with climbing power characteristics, such as air conditioners. To our knowledge, it may increase the complexity for load event detection, load event matching, and load identification. Nevertheless, our method can obtain the accurate occurrence point when the climbing shape occurs, as shown in Table 2. This is the key step for the following load signatures extraction.

In addition, to estimate the similarity of load event, a number of tests about each appliance switching on and off are carried out, thus obtaining the sets of load signature data. Here, each test of the appliance is independent from the others. The probability distribution is then drawn by using the kernel density function according to "(7),", as shown

TABLE 2. Load events of the tested appliances.

Load event	Occurrence point	Active power difference around the occurrence point P(W)	Reactive power difference around the occurrence point Q(Var)
X_1	39	119.6	-38.8
X_2	85	679.5	-3.3
X_3	105	1483.7	-17.5
X_4	132	57.2	12.1
X_5	149	127.0	-474.9
X_6	180	1005.7	-6.4
X_7	198	543.9	-3.3
X_8	219	1148.7	75.0
Y_1	234	-979.9	1.1
Y_2	250	-56.7	-21.5
Y_3	269	-1006.0	185.1
Y_4	287	-1282.1	39.6
Y_5	327	-1905.3	-32.1
Y_6	341	-679.9	1.7
Y_7	365	-507.0	193.2
Y_8	400	-590.6	2.6

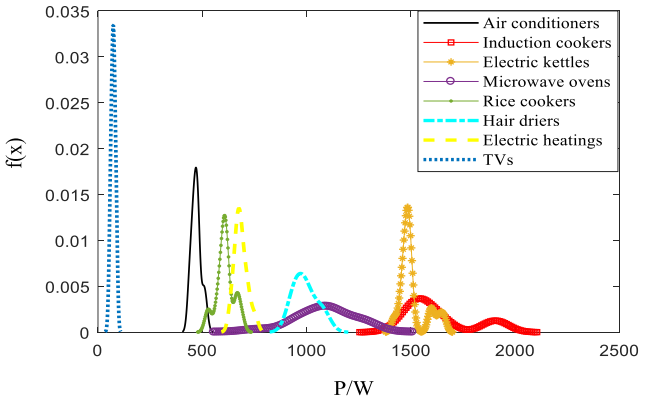


FIGURE 8. Distribution curves of active power.

in Fig. 8 and Fig. 9. It can be seen that the probability distributions of most appliances are followed by Gaussian distribution, although there exist some exceptions, of which the distribution presents several peaks. This can be explained that the appliance may have multiple states during it works.

Table 3 lists the average values from the steady work state of the appliance, which are extracted as the signatures of

TABLE 3. Signatures of load events.

Event	Active power when device on(W)	Reactive power when device on(Var)	Event	Active power when device off(W)	Reactive power when device off(Var)
X_1	476.791	-169.535	Y_1	-1021.024	-9.652
X_2	709.725	-21.742	Y_2	-52.885	10.550
X_3	1487.165	-12.730	Y_3	-1074.068	183.801
X_4	68.931	28.972	Y_4	-1073.918	26.088
X_5	1100.410	-295.364	Y_5	- 1924.100	-27.014
X_6	1009.062	22.907	Y_6	-686.701	1.040
X_7	558.754	23.461	Y_7	-519.064	182.841
X_8	1566.234	79.499	Y_8	-593.005	1.128

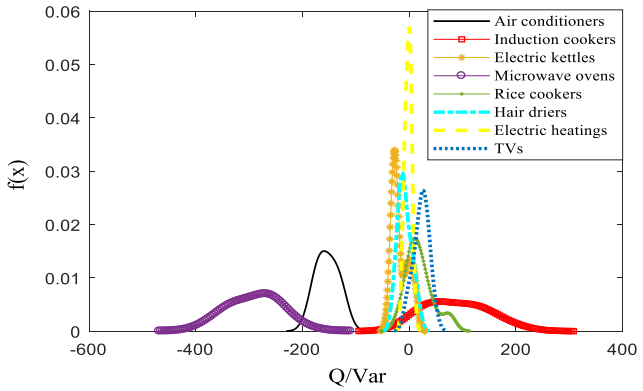


FIGURE 9. Distribution curves of reactive power.

	P_{X1}	P_{X2}	P_{X3}	P_{X4}	P_{X5}	P_{X6}	P_{X7}	P_{X8}
P_{Y1}	0.000	0.000	0.000	0.000	0.296	0.442	0.000	0.000
P_{Y2}	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
P_{Y3}	0.000	0.000	0.000	0.000	0.342	0.279	0.000	0.000
P_{Y4}	0.000	0.000	0.000	0.000	0.342	0.279	0.000	0.000
P_{Y5}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
P_{Y6}	0.000	0.998	0.000	0.000	0.018	0.000	0.190	0.000
P_{Y7}	1.000	0.000	0.000	0.000	0.000	0.000	0.135	0.000
P_{Y8}	0.000	0.002	0.000	0.000	0.002	0.000	0.675	0.000

FIGURE 10. Active power matrix.

load. Notably, the negative sign of active power denotes the switching off event. So, the signatures should be inverted during matching.

Figure.10 and Figure.11 show the intermediate matrixes to match active power and reactive power respectively. It is calculated by substituting the power of switching-off events into the power probability curve of the switching-on events. From the result, it can be seen that the differences between the weights in each column of active power matrix is larger than them of reactive power matrix, which demonstrates that active power is more reliable than reactive power to distinguish different load event.

Figure.12 (a) shows $[k_{ji}]$ constructed by "(11),". The value of each row can reflect the similarity between the

	Q_{X1}	Q_{X2}	Q_{X3}	Q_{X4}	Q_{X5}	Q_{X6}	Q_{X7}	Q_{X8}
Q_{Y1}	0.000	0.081	0.132	0.237	0.000	0.129	0.242	0.208
Q_{Y2}	0.000	0.191	0.128	0.060	0.000	0.297	0.131	0.125
Q_{Y3}	0.754	0.000	0.000	0.000	0.139	0.000	0.000	0.000
Q_{Y4}	0.000	0.000	0.413	0.001	0.000	0.151	0.041	0.072
Q_{Y5}	0.000	0.000	0.002	0.420	0.000	0.015	0.181	0.268
Q_{Y6}	0.000	0.364	0.162	0.141	0.000	0.203	0.203	0.164
Q_{Y7}	0.246	0.000	0.000	0.000	0.861	0.000	0.000	0.000
Q_{Y8}	0.000	0.363	0.162	0.140	0.000	0.204	0.202	0.164

FIGURE 11. Reactive power matrix.

0.000	0.032	0.133	0.095	0.178	**0.317**	0.097	0.083
0.000	0.077	0.128	**0.624**	0.000	0.119	0.052	0.050
0.301	0.000	0.000	0.000	**0.261**	0.167	0.000	0.000
0.000	0.000	**0.413**	0.001	0.205	0.228	0.017	0.029
0.000	0.000	0.002	0.168	0.000	0.007	0.072	**0.707**
0.000	**0.745**	0.162	0.056	0.011	0.081	0.195	0.066
0.699	0.000	0.000	0.000	0.344	0.000	0.081	0.000
0.000	0.146	0.162	0.056	0.001	0.081	**0.486**	0.065

(a)

0	0	0	0	0	1	0	0
0	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0

(b)

FIGURE 12. (a) Sum matrix; (b) matching result of the sum matrix.

switching-on event and the switching-off event. In most cases, it can be considered that the switching-on appliance and the switching-off appliance with the greatest possibility belong to the same appliance. The final result matrix using the improved KM algorithm is shown in Figure 11(b). For more understanding, the load event matching result can be listed as follows: $M(x_1,y_7)$, $M(x_2,y_6)$, $M(x_3,y_4)$, $M(x_4,y_2)$, $M(x_5,y_3)$,

Load event	$\{X_i\}$	6	4	5	3	8	2	1	7
	$\{Y_j\}$	1	2	3	4	5	6	7	8
Air conditioner		0	0	1	0	0	0	**2**	0
Induction cooker		0	0	0	0	1	0	0	0
Electric kettle		1	1	1	**2**	0	1	0	1
Microwave oven		1	0	1	1	0	0	0	0
Rice cooker		0	1	0	0	1	1	0	**2**
Hair drier		**2**	0	1	1	0	1	0	1
Electric heating		1	1	0	0	0	**2**	0	1
TV		0	**2**	0	0	1	0	0	1

FIGURE 13. Load identification matrix.

Load event	$\{X_i\}$	6	4	5	3	8	2	1	7
	$\{Y_j\}$	1	2	3	4	5	6	7	8
Air conditioner		0	0	0	0	0	0	1	0
Induction cooker		0	0	0	0	1	0	0	0
Electric kettle		0	0	0	1	0	0	0	0
Microwave oven		0	0	1	0	0	0	0	0
Rice cooker		0	0	0	0	0	0	0	1
Hair drier		1	0	0	0	0	0	0	0
Electric heating		0	0	0	0	0	1	0	0
TV		0	1	0	0	0	0	0	0

FIGURE 14. Identification result matrix.

$M(x_6,y_1)$, $M(x_7,y_8)$, $M(x_8,y_5)$. In this matching set, the total edge weight of the optimal matching reaches 4.252. The result is almost the sum of the largest value of each column, except the fifth column. This implies that the improved KM algorithm is to obtain the optimal solution within a defined space of all appliances instead of local optimization. So, our method can cope with the situation that the load switched off is wrongly matched by the single matching, thus promoting the performance of load disaggregation. Taking this test as an example, the optimal matching for X_5 in Fig.12(a) corresponds to Y_7, whereas considering all the appliances, the optimal matching for X_5 is Y_3. Eventually, the test result coincided with the actual load events of the tested appliances.

Figure.13 shows the load identification matrix according to the same load recognition method introduced above. It can be seen that, the event Y_1, Y_2, Y_4, Y_6, Y_7, Y_8 can be determined directly, whose matching weight is equal to two. So the rest event Y_3 and Y_5 are easy to be matched by KM algorithm. Finally, Figure.14 shows the identification result obtained by using KM algorithm, and this optimal matching result is consistent with actual test scenarios. This demonstrates that our method is very effective for load recognition.

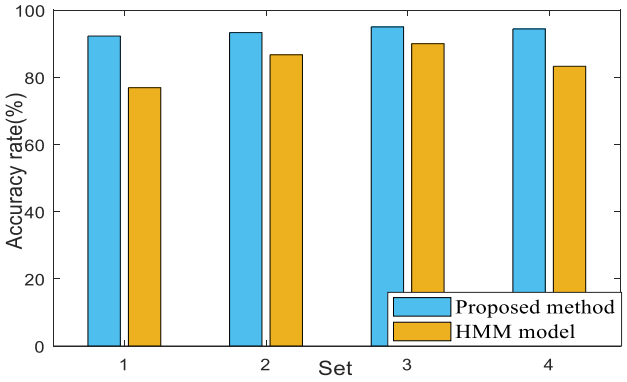


FIGURE 15. Performance of the proposed method and HMM model.

B. REDD DATASET

In order to further verify the proposed method, this paper also use several groups data of house 3 in the REDD dataset for comparison. This test contains twelve different appliances such as furnace, refrigerator, lightings, washer-dryer, microwave, and smoke-alarm. In this dataset, there also exist some multi-state appliances such as the furnace and refrigerator. Here, the value of α and β are set as 0.45 and 0.55 respectively according to "(15),".

For the sake of fairness, our method is compared with the common used HMM model which is to detect the system mode with certain probability by utilizing the relevant information of the system [36]. The results of this method and the HMM model in this test are shown in Fig.15. It can be seen that the proposed method can identify loads effectively compared with the HMM model, and almost has desirable matching and identification results.

IV. CONCLUSION

In this paper, an effective NILM approach built on the emerging graph-based method is proposed, which contains three steps. The first step is to detect load event by using an adaptive load event detection method. Then, the statistical model of load signatures is built, which draws upon the signatures extracted from the load switched on. Particularly, the probability of load event matching is utilized to form the matching matrix, thus improving the KM algorithm to solve the matrix of the load event matching and load identification. Finally, experimental results on the REDD dataset and our lab's data show that the proposed method has a desirable performance in the load event matching and load identification. In the near future, the load identification matrix in our method will be consisted of more appliances and their corresponding load events when the quantity of appliances increases, and then apply our model into wide applications.

REFERENCES

[1] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Improving nonintrusive load monitoring efficiency via a hybrid programing method," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2148–2157, Dec. 2016.

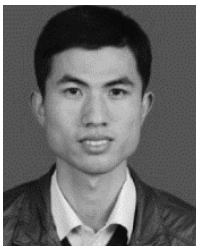
[2] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1739–1747, May 2018.

[3] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, "NILM techniques for intelligent home energy management and ambient assisted living: A review," *Energies*, vol. 12, no. 11, p. 2203, Jun. 2019. doi: 10.3390/en12112203.
[4] B. Qi, L. Liu, and X. Wu, "Low-rate non-intrusive load disaggregation with graph shift quadratic form constraint," *Appl. Sci.*, vol. 8, no. 4, pp. 554–570, Apr. 2018.
[5] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "On the Bayesian optimization and robustness of event detection methods in NILM," *Energy Buildings*, vol. 145, pp. 57–66, Jun. 2017.
[6] Q. Yuan, H. Wang, B. Wu, Y. Song, and H. Wang, "A fusion load disaggregation method based on clustering algorithm and support vector regression optimization for low sampling data," *Future Internet*, vol. 11, no. 2, pp. 51–63, Feb. 2019. doi: 10.3390/fi11020051.
[7] H. Wang, W. Yang, T. Chen, and Q. Yang, "An optimal load disaggregation method based on power consumption pattern for low sampling data," *Sustainability*, vol. 11, no. 1, pp. 251–266, Jan. 2019.
[8] K. T. Chui, M. D. Lytras, and A. Visvizi, "Energy sustainability in smart cities: Artificial intelligence, smart monitoring, and optimization of energy consumption," *Energies*, vol. 11, no. 11, pp. 2869–2888, Oct. 2018.
[9] F. Gong, C. Liu, L. Jiang, H. Li, J. Y. Lin, and B. Yin, "Load disaggregation in non-intrusive load monitoring based on random forest optimized by particle swarm optimization," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Beijing, China, Nov. 2017, pp. 1–6.
[10] 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.
[11] L. Jiang, J. Li, S. Luo, S. West, and G. Platt, "Power load event detection and classification based on edge symbol analysis and support vector machine," *Appl. Comput. Intell. Soft Comput.*, vol. 2012, Nov. 2012, Art. no. 742461.
[12] S. Biansoongnern and B. Plangklang, "Nonintrusive load monitoring (NILM) using an artificial neural network in embedded system with low sampling rate," in *Proc. 13th Int. Conf. Elect. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Chiang Mai, Thailand, Jun./Apr. 2016, pp. 1–4.
[13] Q. Wu and F. Wang, "Concatenate convolutional neural networks for non-intrusive load monitoring across complex background," *Energies*, vol. 12, no. 8, pp. 1572–1588, Apr. 2019. doi: 10.3390/en12081572.
[14] I. H. Cavdar and V. Faryad, "New design of a supervised energy disaggregation model based on the Deep Neural Network for a smart grid," *Energies*, vol. 12, no. 7, pp. 1217–1234, Mar. 2019. doi: 10.3390/en12071217.
[15] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "An unsupervised training method for non-intrusive appliance load monitoring," *Artif. Intell.*, vol. 217, pp. 1–19, Dec. 2014.
[16] T. Zia, D. Bruckner, and A. Zaidi, "A hidden Markov model based procedure for identifying household electric loads," in *Proc. Conf. IEEE Ind. Electron. Soc.*, Melbourne, VIC, Australia, Nov. 2011, pp. 3218–3223.
[17] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," *Energy Buildings*, vol. 116, pp. 96–103, Mar. 2016.
[18] X. Wang, R. Li, D. Zhou, H. Zhou, and W. Hu, "Non-intrusive power load disaggregation method based on decision fusion and its applications," *Power Syst. Protection Control*, vol. 44, no. 7, pp. 115–121, May 2016.
[19] V. Singhal, J. Maggu, and A. Majumdar, "Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2969–2978, May 2019.
[20] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "PALDi: Online load disaggregation via particle filtering," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 467–477, Feb. 2015.
[21] H. G. C. P. Dinesh, P. H. Perera, G. M. R. I. Godaliyadda, M. P. B. Ekanayake, and J. B. Ekanayake, "Residential appliance monitoring based on low frequency smart meter measurements," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Miami, FL, USA, Nov. 2015, pp. 878–884.
[22] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.
[23] M. L. Kulygin, "Development of waveguide semiconductor switches of microwave radiation in the 70- and 260-GHz ranges," *Radiophys. Quantum Electron.*, vol. 57, no. 7, pp. 509–518, Dec. 2014.
[24] M. Dong, P. C. M. Meira, W. Xu, and F. Walmir, "An event window based load monitoring technique for smart meters," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 787–796, Jun. 2012.
[25] H. Lu and S. Wang, "A study on multi-ASC scheduling method of automated container terminals based on graph theory," *Comput. Ind. Eng.*, vol. 129, pp. 404–416, Mar. 2019.
[26] M. Rabbani, R. Yazdanparast, and M. Mobini, "An algorithm for performance evaluation of resilience engineering culture based on graph theory and matrix approach," *Int. J. Syst. Assurance Eng. Manage.*, vol. 10, no. 2, pp. 228–241, Apr. 2019.
[27] B. Zhao, L. Stankovic, and V. Stankovic, "Blind non-intrusive appliance load monitoring using graph-based signal processing," in *Proc. GlobalSIP*, Orlando, FL, USA, Dec. 2015, pp. 68–72.
[28] S. Dipu, P. Konwar, A. De, and S. Goswami, "A graph theory application for fast and efficient search of optimal radialized distribution network topology," *J. King Saud Univ.-Eng. Sci.*, to be published. doi: 10.1016/j.jksues.2019.02.003.
[29] P. N. Yeganeh and M. T. Mostafavi, "Causal disturbance analysis: A novel graph centrality based method for pathway enrichment analysis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published.
[30] Y.-C. Su, K.-L. Lian, and H.-H. Chang, "Feature selection of non-intrusive load monitoring system using STFT and wavelet transform," in *Proc. ICEBE*, Beijing, China, Oct. 2011, pp. 293–298.
[31] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring: A review and outlook," Oct. 2016, *arXiv:1610.01191*. [Online]. Available: https://arxiv.org/abs/1610.01191
[32] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.
[33] A. J. Bijker, X. Xia, and J. Zhang, "Active power residential non-intrusive appliance load monitoring system," in *Proc. AFRICON*, Nairobi, Kenya, Sep. 2009, pp. 1–6.
[34] Y. Zhou and J. Kuang, "A sort method to enhance significant spectral components of test set," in *Proc. 12th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Aug. 2016, pp. 2147–2151.
[35] Q. Li, "Research on harmonic characteristics of electric devices for non-intrusive load monitoring," *Southern Power Syst. Technol.*, vol. 10, no. 10, pp. 73–78, Oct. 2016.
[36] F. Li, S. Song, J. Zhao, S. Xu, and Z. Zhang, "Synchronization control for Markov jump neural networks subject to HMM observation and partially known detection probabilities," *Appl. Math. Comput.*, vol. 360, pp. 1–13, Nov. 2019.

YONG XIAO was born in Jingzhou, Hunan, China, in 1978. He received the Ph.D. degree from Wuhan University. He is currently a Senior Engineer. His research interests include intelligent electric power technology and measurement automation technology.

YUE HU was born in Xianning, Hubei, China, in 1996. She received the B.S. degree in automation from Wuhan University, in 2018, where she is currently pursuing the M.S. degree in control theory and control engineering. She works on signal processing of power information on smart grid.

HENGJING HE was born in Dangyang, Hubei, China, in 1986. He received the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University. He is currently working in research on modern electromagnetic measurement technology and instruments.



DONGGUO ZHOU was born in 1985. He received the B.S. degree in measurement and control technology and instruments and the Ph.D. degree in instrumentation science and technology from Chongqing University, in 2008 and 2013, respectively. His research interests include information signal acquisition and processing.

YUN ZHAO was born in Jingmen, Hubei, China, in 1985. He received the Ph.D. degree in engineering from Wuhan University. His research interest includes electric energy metering.

WENSHAN HU was born in 1980. He received the B.S. and M.S. degrees in automation from Wuhan University and the Ph.D. degree from the University of Glamorgan, in 2008. He is currently a Professor with Wuhan University. His research interests include network control systems and radio power transmission.