

Research Article

Nonintrusive Load Identification for Industrial Users Integrated with LSQR and Sequential Leader Clustering

Shuhui Yi , Yinglong Diao, Junjie Liu, Tian Fang, and Xiaodong Yin 

China Electric Power Research Institute, Wuhan 430070, China

Correspondence should be addressed to Xiaodong Yin; yinxiaodong@epri.sgcc.com.cn

Received 3 April 2022; Revised 4 August 2022; Accepted 24 August 2022; Published 28 September 2022

Academic Editor: Daniel Morinigo-Sotelo

Copyright © 2022 Shuhui Yi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonintrusive load identification for industrial users can accurately acquire the operation of each load. However, it is a major challenge in the demand-side response due to the hardship of collection data for modelling, and high precision measuring equipment is required. Aiming at this situation, a nonintrusive load identification method is proposed, combining the least square QR (LSQR) with the sequential leader clustering algorithm. Firstly, regarding accurate depiction of industrial loads, some appropriate load feature indices of steady-state and transient processes are extracted, respectively. For steady-state processes, the active power, the reactive power, and the root mean square (RMS) current value are selected as the feature indices. In the case of transient processes, ten feature indices of three stages are employed: before, during, and after transient events, consisting of the duration of transient events, the RMS current value before and after transient events, the average value of active power before and after transient events, the average value of reactive power before and after transient events, the maximum RMS current value of transient events, etc. On this base, the LSQR algorithm is proposed to decompose unknown composite power to access the operation of various loads at steady-state. The sequential leader clustering algorithm is propounded to classify transient events of typical industrial loads and further identify which kind of loads had switched. Finally, to validate the effectiveness of the presented model, data of industrial loads from a concrete plant are collected, including blender, cement screw, sewage dump, and inclined belt conveyor, and simulation analysis is fulfilled. The results indicate that the model proposed can effectively achieve the nonintrusive industrial load identification, and least unified residue (LUR) is about 10^{-16} , which is much better than the factorial hidden Markov model (FHMM) and the artificial neural network (ANN) model.

1. Introduction

With the conception of electric-load intelligent (E-LI) being put forward, the number of users increases so rapidly that load types are diversified and users' requirements for electricity service are more and more detailed [1, 2]. On this basis, it is crucial to analyze the electric consumption behavior more precisely and study the key technologies related to load monitoring, e.g., load feature extraction [3, 4], load identification [5], load forecasting [6], and load monitoring [7]. Therefore, research on these key technologies can improve the efficiency of collection-transmission and enhance the accuracy of monitoring and identification for load equipment. Currently, there are mainly two load monitoring approaches: intrusive load monitoring (ILM) and

nonintrusive load monitoring (NILM). ILM is built upon low-end meter devices attached to home appliances in opposition to NILM techniques, and a single sensing point of each electrical device is needed in the ILM [8, 9]. Although ILM can provide higher efficiency and reliability than NILM, it is gradually replaced by the latter. NILM can perform real-time monitoring and analysis of demand-side load. It has many advantages, including low cost, simple communication, easy maintenance, and promotion. Therefore, load identification methods for NILM become a research hotspot of demand-side load monitoring.

Presently, some research results probe into deep learning (DL) intelligent algorithms for NILM in the steady-state process, such as hidden Markov model [10, 11], artificial neural network (ANN) model [12, 13], and graph signal

processing developed approach [14, 15]. In [10, 11], the factorial hidden Markov model (FHMM) is considered, and it is indicated to outperform the other unsupervised disaggregation methods. In [12, 13], the DL model on ANN algorithm is also investigated and has high effectiveness and accuracy in analysis of power consumption behavior for load monitoring. In [14, 15], the graph signal processing developed approach is designed for NILM. The competitive performance of the GSP-based approach is demonstrated with respect to the traditional hidden Markov model and decision tree approaches. Nevertheless, the above research outputs are mainly focused on residential electricity consumption, and there are few achievements on industrial electricity consumption.

Moreover, several methods concentrate on the transient process with a view to further monitor the load switching status, including data acquisition, transient feature extraction, and load identification. Some research evaluates load signatures established by voltage-current (V-I) trajectory [16, 17], and some papers propose signal processing techniques to extract features, such as wavelet transform [18] and Stockwell transform (S-transform) [19, 20]. Some novel algorithms are applied to transient energy analysis to improve recognition accuracy and computational speed. For instance, a stealthy black-box attack construction model is suggested in [21], and ANN model combining with turn-on transient energy analysis is utilized in [22]. Most of load disaggregation methods are designed to focus on the load on/off events of single appliance; however, a current investigation in [23] can effectively deal with the simultaneous on/off events of multiple appliances by applying a cepstrum-smoothing-based load disaggregation method.

The aforesaid research results are mainly aimed at household loads, and the experimental data mostly use open REDD data sets. There are few studies on the load decomposition and load identification for industrial users, and testing data of industrial loads are limited. In [24], it presents a new nonintrusive energy monitoring method for industrial microgrid; however, it is not suitable for demand-side monitoring for industrial users. Moreover, load feature extraction and load identification are only performed for steady-state or transient processes, and there is no overall process to cover steady-state and transient processes.

Given the above, a new method of load identification for industrial users is suggested, integrating with the analysis of steady-state and transient load feature of industrial users. The active power, the reactive power, and the root mean square (RMS) current value are selected as the feature indices of steady-state processes. The least square QR (LSQR) algorithm is utilized to decompose the loads in the steady-state process. In addition, for transient processes, ten feature indices of three stages are employed: before, during, and after transient events, consisting of the duration of transient events, the RMS current value before and after transient events, the average value of active power before and after transient events, the average value of reactive power before and after transient events, the maximum RMS current value of transient events, the maximum active power during transient, and the maximum reactive power during

transient. The sequential leader clustering analysis is applied to classify the transient events of typical industrial loads and further identify transient events.

1.1. The Main Contributions

- (i) According to the characteristics of industrial users, the steady-state and transient feature indices of industrial load are extracted, and the steady-state load feature space and transient load feature space are established.
- (ii) The load decomposition-identification model for industrial scenarios combining the LSQR with the sequential leader clustering algorithm is proposed. The operation of each load is obtained by using the LSQR algorithm model, in combination with the sequential leader clustering algorithm to further identify transient events.
- (iii) Four kinds of loads, including blender, cement screw, sewage pump, and inclined belt conveyor, are selected to test the load decomposition-identification model. The industrial load dataset has collected 30 hours of bus-side three-phase current and voltage signals from a concrete plant by DL850 oscillograph recorder.
- (iv) The results demonstrate that the proposed load decomposition-identification model gives better performance for the nonintrusive industrial load identification compared to the FHMM and ANN model.

The remainder of this paper is organized as follows. Section 2 provides a brief analysis of industrial loads and summarizes the feature indices of steady-state and transient processes. Section 3 describes the framework of the envisaged model, followed by the details of the three components: data processing of raw sampling, load decomposition of the steady-state process rooted on the LSQR algorithm, and the transient process load identification built on the clustering algorithm. Section 4 delineates the data sources and discusses the performance of the proposed model, and a validation of the whole approach is given. Section 5 concludes the paper and envisions the future work.

2. Characteristic Analysis of Industrial Users

The types of industrial users are so diverse that the traits of different types have certain distinct. In accordance with the electricity price classification, the industrial users can be categorized into following, namely, commercial mixed electricity, office-used, agricultural-used, civil-used, non-general industrial users, and large industrial users. The electricity demand of industrial users can be characterized as a sequence fluctuating with time. Modeling and analysis of this sequence can achieve its inherent traits [25, 26]. The equipment-side of a concrete plant is selected as the monitoring target in this paper. And, at 10 kHz sampling frequency, the DL850 oscillograph recorder is operated to record the current wave information of multiple typical

loads on the equipment-side, including blender, cement screw, sewage pump, and inclined belt conveyor. For instance, Figure 1(a) shows the current change after a cement-screw-on event, while Figure 1(b) shows the current change after a cement-screw-off event.

From Figure 1, it shows that the load power consumption of industrial users generally has the following traits:

- (i) Current wave shows obvious fluctuation with time
- (ii) Load on/off events cause significant changes in instantaneous current
- (iii) Current waveform tends to be stable after load on/off events

In the action of load switching, the transient event of an industrial load consists of three stages: before, during, and after transient. The stage before transient refers to the original steady state before a load switches, and the stage after transient stands for another stable state after a load switches. And, some features, such as current and power, will change dramatically during the transient process. Therefore, it is necessary to extract complete features of steady-state and transient processes for load identification.

2.1. Steady-State Features Extraction. The active power and the reactive power are regularly opted as feature indices for the steady-state process:

$$\begin{aligned} P &= VI \cos \phi, \\ Q &= VI \sin \phi, \end{aligned} \quad (1)$$

where P and Q symbolize the active power and the reactive power, respectively. V and I are the RMS values of voltage and current, respectively. ϕ means the phase difference between voltage and current, and $\cos\phi$ is defined as the power factor.

Figure 1 exhibits that the load signals of industrial users are periodic when the frequency of the modulation is a harmonic of the fundamental frequency. In order to characterize the industrial loads more comprehensively, the RMS value of current in the steady-state process is also extracted as a load feature index:

$$I = \sqrt{\frac{\sum_{n=1}^N i(n)^2}{N}}, \quad (2)$$

where N is the number of sampling points in one fundamental period, n is the number of sampling points, and $i(n)$ is the instantaneous current value of each sampling point.

Therefore, as shown in Table 1, the active power, the reactive power, and the RMS current value are extracted to construct the steady-state process feature space. RMS is a feature that is very sensitive to the window length; hence, here, we select the composite-window method based on the cumulative sum (CUSUM) algorithm to calculate three steady-state feature indices. The composite-window method is applicable for the feature indices of both steady-state and transient-state [26].

2.2. Transient Features Extraction. Detection of transient events is a central issue to realize load identification for the transient process. Waveforms between the original steady-state and the new steady-state can be captured to extract the transient features of different loads. In order to outline the transient process more completely, features of three stages delineated above are selected to construct a comprehensive feature space. As can be seen in Table 2, there are altogether 10 feature indices.

In the case of load-on events, the first stage in Table 2 represents the steady-state before the occurrence of load-on events. The second stage represents the process of load-on events. The third stage indicates the new steady-state after load-on events. It is similar in another case, except that event types are load-off events instead of load-on events.

In this paper, transient feature indices are extracted according to the composite-window based on CUSUM [26]. The window length is very important for transient events detection. If the window length is too short, the steady-state fluctuations and glitch of the current easily lead to false detection of the events. However, if the window length is too wide, transient events cannot be accurately identified. Due to the wide duration of industrial equipment start-up transient process, we select the maximum RMS current value during transient as one of the feature indices. This composite-window method can verify the validation of event detection and distinguish transient events due to quality problems, gaps, and impulses, etc., from the connection or disconnection of industrial loads.

The average value of active power before or after transient relies on the active power in each period, the same goes for the average value of reactive power. Take the pretransient process as an example:

$$\begin{aligned} P_f &= \frac{1}{T} (P_1 + P_2 + \dots + P_T), \\ Q_f &= \frac{1}{T} (Q_1 + Q_2 + \dots + Q_T), \end{aligned} \quad (3)$$

where T expresses the numbers of steady-state waveform periods before the transient process. P_n and Q_n ($n = 1, 2, \dots, T$) are active power and reactive power in the n -th period, respectively.

3. Load Identification Method Depending on the Steady-State Process and the Transient-State Process

To obtain an integrated load identification for nonintrusive industrial users, this paper focuses on the load decomposition of the steady-state process combining with identification of the transient process. The designed identification algorithm mainly includes two parts, as can be explicated in Figure 2. Firstly, the steady-state load decomposition can be described by the least squares problem which can be optimized by the LSQR algorithm. Secondly, the sequential leader clustering algorithm is introduced to classify the corresponding load on/off events in the transient process, in such a way as to accurately identify which load or type of loads have a switch action.

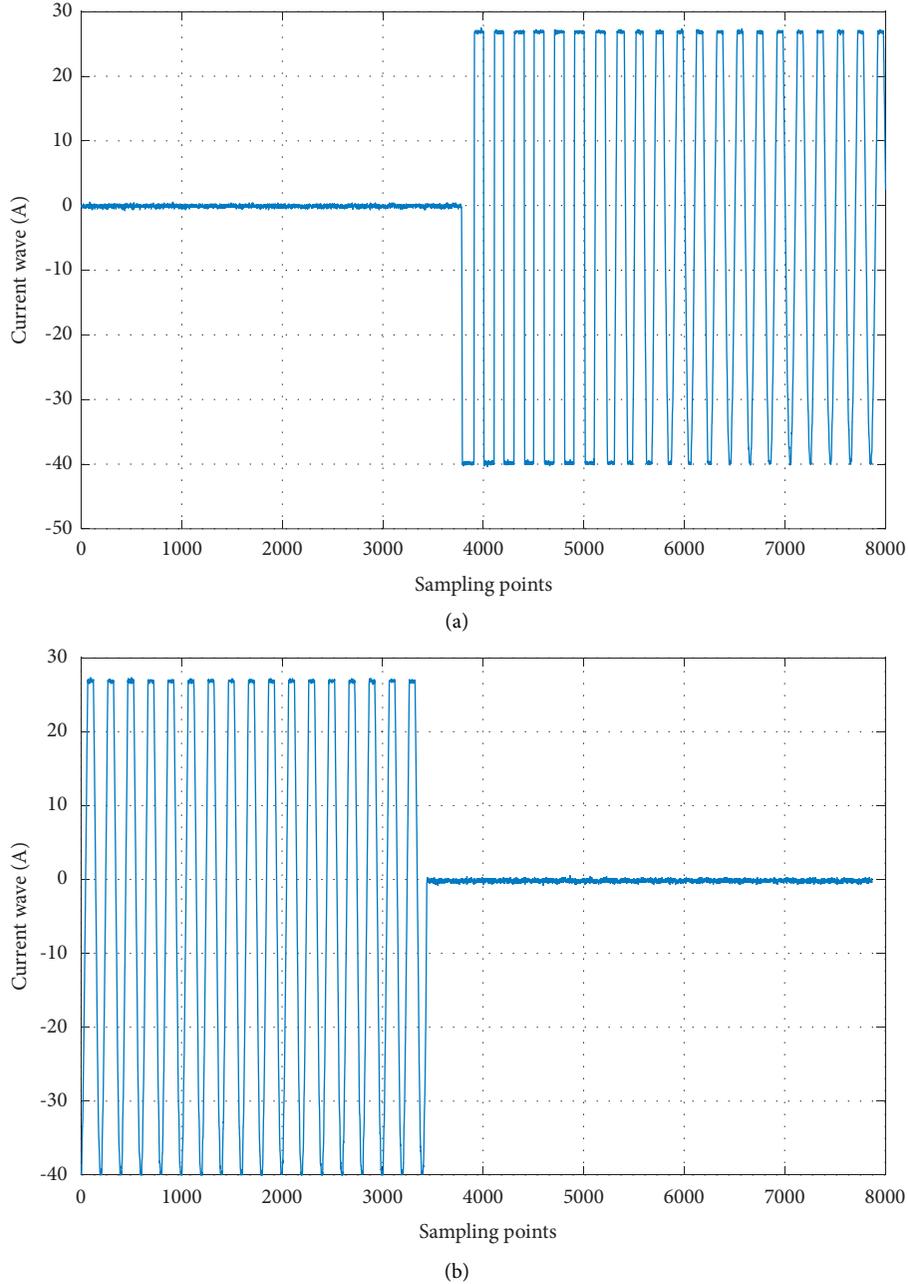


FIGURE 1: Current change of on/off events for cement screw.

TABLE 1: Load steady-state feature indices of industrial users.

Number	Feature indices
1	Active power, P
2	Reactive power, Q
3	RMS current value, I

3.1. Data Processing of Raw Sampling. Random errors are prone to occur when waveforms are recorded due to the complex industrial environment, regardless of the equipment-side or bus-side. Thus, invalid data need to be eliminated and corrected. In order to record the waveforms of collecting data more objectively and comprehensively, the data of voltage and

current are usually sampled at high frequency. However, it will cause a sharp increase in data volume and decrease the data processing efficiency to a certain extent. As well, data of different load features are generally not in a unified dimension; accordingly, it makes normalization more essential. In short, it is necessary to adjust and process the original sampling data, including data cleaning, downsampling, and normalization. Herein, the Irada criterion is applied for data cleaning to eliminate the random error data, and the Nyquist sampling theorem is employed for downsampling to obtain the balance between efficiency and accuracy of sampling. The min-max normalization and Z-score normalization methods are both utilized to normalize parameters.

TABLE 2: Load transient feature indices of industrial users.

Stage	Feature indices
Before transient	RMS current value of the steady state before transient, I_f
	Average active power before transient, P_f
	Average reactive power before transient, Q_f
During transient	Transient duration, T_t
	Maximum RMS current value during transient, I_{\max}
	Maximum active power during transient, P_{\max}
	Maximum reactive power during transient, Q_{\max}
After transient	RMS current value of steady state after transient, I_b
	Average active power after transient, P_b
	Average reactive power after transient, Q_b

3.1.1. Irada Criterion. Assume that s_1, s_2, \dots, s_n are original sampling data, respectively, and $v_i = s_i - \bar{s}$ ($i = 1, 2, \dots, n$) symbolizes residual error, where \bar{s} is the arithmetic mean. σ denotes the standard error in terms of Bessel's formula. v_i satisfies the following formula:

$$|v_i| = |s_i - \bar{s}| > 3\sigma, \quad (4)$$

where s_i is considered as a gross error value that should be corrected. Here, we use the average value of adjacent sampling points to replace the invalid data for correction.

3.1.2. Nyquist Sampling Theorem. Nyquist sampling theorem depicts the relationship between the sampling rate and the signal frequency. It is ruled that the sampling rate f_s must be greater than twice the highest frequency of sampled signal, which is usually called the Nyquist frequency f_N :

$$f_s > 2f_N. \quad (5)$$

3.1.3. Min-Max Normalization. The min-max normalization can transform the original data into the range of [0 1] by the linearization method. Assume that s_{ij} is the value of the j -th feature of i th load. And, $\min(s_j)$ represents the minimum value of the j -th feature among all the power consumption. Similarly, $\max(s_j)$ indicates the maximum value of the j -th feature among all the power consumption.

$$Z(s_{ij}) = \begin{cases} \frac{s_{ij} - \min(s_j)}{\max(s_j) - \min(s_j)}, & \max(s_j) \neq \min(s_j), \\ 1, & \max(s_j) = \min(s_j). \end{cases}, \quad (6)$$

where $Z(s_{ij})$ denotes the normalized value of s_{ij} .

3.1.4. Z-Score Standard Normalization. Z-score standard normalization is usually used to standardize the original data, and it can eliminate the influence of different variable units on the results:

$$Z(s_{ij}) = \frac{s_{ij} - \bar{s}_i}{S(s_{ij})} = \frac{s_{ij} - \bar{s}_i}{\sqrt{(1/n) \sum_{j=1}^n (s_{ij} - \bar{s}_i)^2 / n}}, \quad (7)$$

where \bar{s}_i is the arithmetic mean of all power consumption features of the i th load, $S(s_{ij})$ is the sample standard deviation, and n is the number of the feature indices. $Z(s_{ij})$ has the same meaning as formula (6).

3.2. Load Decomposition of the Steady-State Process Relied on LSQR Algorithm

3.2.1. Problem Description. The LSQR algorithm solution consists of two steps [27, 28], as illustrated in Figure 3. The first is performing Golub–Kahan iterative bidiagonalization to simplify the calculation. The second is to solve the least-squares problem after bidiagonalization.

Given a known load feature database and a composite power feature database with unknown constituent elements in a certain scene, an unknown power can be decomposed into several identifiable loads which belong to the known load feature database. So, the load decomposition of the steady-state process for NILM can be described as

$$\begin{aligned} & \min_{\mathbf{x}} \|\mathbf{H} \cdot \mathbf{x} - \mathbf{y}\|_2, \\ & \text{s.t.} \begin{cases} \mathbf{H} \cdot \mathbf{x} < \mathbf{y} \\ x_i \geq 0 \end{cases}, \end{aligned} \quad (8)$$

where $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_R]$ means the load feature matrix of R kinds of industrial users and $\mathbf{H}_i = [f_{i,1}, f_{i,2}, \dots, f_{i,M}]^T$ expresses the load feature vector of the i th load with M feature factors. $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$ is the feature vector for a composite power expressed by the same M features. $\mathbf{x} = [x_1, x_2, \dots, x_R]^T$ represents the weight vector of the loads, where x_i designates the operation status of i th load. $\|\cdot\|_2$ is the minimum second-order residual norm, which is the fitness function between the estimated composite power and the actual data.

In general, the feature matrix is nonpositive definite. Therefore, the LSQR algorithm is adopted to approximately solve the linear least squares problem as formula (8), which is equivalent to iteratively solving the regular equation $\mathbf{H}^H \mathbf{H} \mathbf{x} = \mathbf{H}^H \mathbf{y}$.

3.2.2. Golub–Kahan Bidiagonalization. By singular value decomposition (SVD) method, the feature matrix H can be constructed as a bidiagonal form:

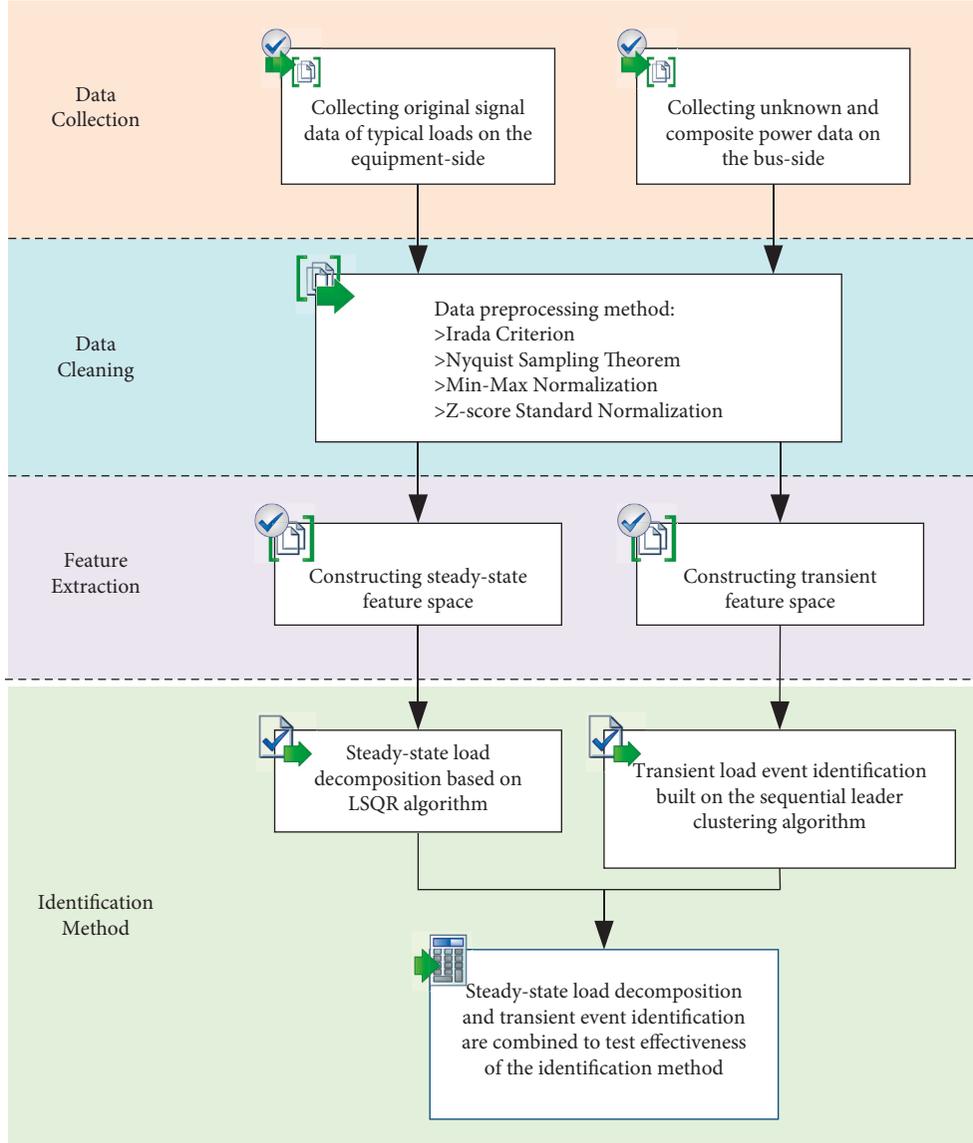


FIGURE 2: Flow chart of the proposed load identification method.

$$\mathbf{U}^H \mathbf{H} \mathbf{V} = \mathbf{B} = \begin{bmatrix} \alpha_1 & & & & \\ \beta_2 & \alpha_2 & & & \\ & \beta_3 & \ddots & & \\ & & \ddots & \alpha_P & \\ & & & & \beta_{P+1} \end{bmatrix}, \quad (9)$$

$$\mathbf{H}[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_P] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P] \begin{bmatrix} \alpha_1 & & & & \\ \beta_2 & \alpha_2 & & & \\ & \beta_3 & \ddots & & \\ & & \ddots & \alpha_P & \\ & & & & \beta_{P+1} \end{bmatrix}. \quad (10)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_P]$ are both unitary matrices, which can be determined by the iterative method. P is the number of iterations. α_k ($k=1, 2, \dots, P$) is the element on the main diagonal of the bidiagonal matrix \mathbf{B} , and β_k is the element on the diagonal below the main diagonal.

In both sides of (9), multiply \mathbf{U} on the left to get

$Hv_k = \alpha_k u_k + \beta_{k+1} u_{k+1}$ or $\beta_k u_k = Hv_{k-1} - \alpha_{k-1} u_{k-1}$ can be deduced by (10). On the other hand, apply the Hermit-transform to both sides of (11) to get $V^H H^H U = B^H$. Then, multiply \mathbf{V} on both sides by the left to get $H^H U = VB^H = VB^T$:

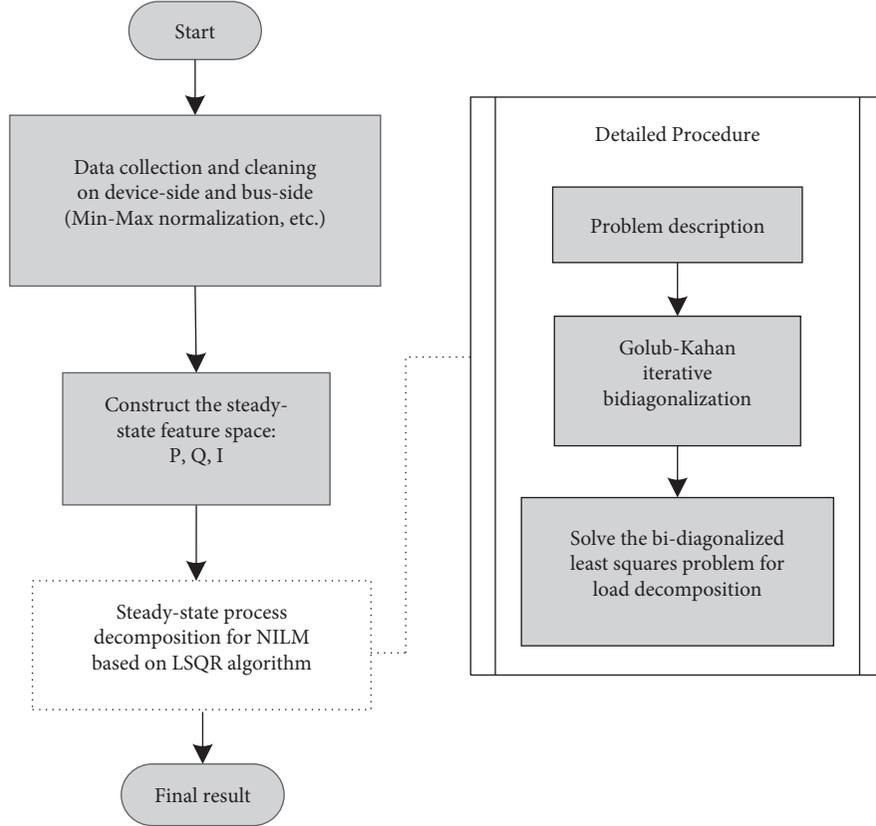


FIGURE 3: Load decomposition of the steady-state process relied on the LSQR algorithm.

$$\mathbf{H}^H [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_P] \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ & \alpha_2 & \beta_3 & & \\ & & \ddots & \ddots & \\ & & & \alpha_p & \beta_{p+1} \end{bmatrix}. \quad (11)$$

Similarly, $\mathbf{H}^H \mathbf{u}_k = \beta_k \mathbf{v}_{k-1} + \alpha_k \mathbf{v}_k$ or $\alpha_k \mathbf{v}_k = \mathbf{H}^H \mathbf{u}_k - \beta_k \mathbf{v}_{k-1}$ ($\beta_1 = 0, k = 1, 2, \dots, P$) can be calculated.

The bidiagonalization depending on the Golub–Kahan method transforms the load feature matrix H into sparse bidiagonal matrix B , and the process can be summarized in Table 3.

3.2.3. Solving the Least Squares Problem. Based on the bidiagonalization of P iterations, the least squares problem of load decomposition designated in formula (8) can be transformed into the following equation:

$$\min_{\mathbf{v}^H \mathbf{x}} \|\mathbf{U}^H \mathbf{H} \mathbf{V} \mathbf{V}^H \mathbf{x} - \mathbf{U}^H \mathbf{y}\|_2 = \min_{\mathbf{c}_p} \|\mathbf{B}_p \mathbf{c}_p - [\beta_1, 0, 0, \dots, 0]^T\|_2, \quad (12)$$

where B_p is a $(P+1) \times P$ bidiagonal matrix. $\mathbf{c}_p = \mathbf{V}^H \mathbf{x}$ denotes the solution of the least squares problem of sparse bidiagonal matrix, which can be attained by QR decomposition of (13). Since V is the unitary matrix, the approximate estimated solution of (8) can be obtained by \mathbf{c}_p in

TABLE 3: Golub–Kahan bidiagonalization of the LSQR algorithm.

Steps	Description
Setting initial values	$\beta_1 = \ \mathbf{y}\ _2$
	$\mathbf{u}_1 = \mathbf{y}/\beta_1$
	$\alpha_1 = \ \mathbf{H}^H \mathbf{y}\ _2$ $\mathbf{v}_1 = \mathbf{H}^H \mathbf{y}/\alpha_1$
For $P = 1, 2, \dots$	
Iteration	$\beta_{p+1} = \ \mathbf{H} \mathbf{v}_p - \alpha_p \mathbf{u}_p\ _2$
	$\mathbf{u}_{p+1} = (\mathbf{H} \mathbf{v}_p - \alpha_p \mathbf{u}_p)/\beta_{p+1}$
	$\alpha_{p+1} = \ \mathbf{H}^H \mathbf{u}_{p+1} - \beta_{p+1} \mathbf{v}_p\ _2$
	$\mathbf{v}_{p+1} = (\mathbf{H}^H \mathbf{u}_{p+1} - \beta_{p+1} \mathbf{v}_p)/\alpha_p$
End	

$$\hat{\mathbf{x}}_p = \mathbf{V} \mathbf{c}_p, \quad (13)$$

where $\hat{\mathbf{x}}_p$ is the estimated result, which signifies the component weight of each known load after load decomposition.

3.3. Load Identification for the Transient Process Underlying on the Sequential Leader Clustering Algorithm. The sequential leader clustering algorithm [29, 30] is approved to detect unknown transient events to realize load identification of the transient process. According to the load feature space established in Table 2, the main process of the load identification based on the clustering algorithm is as follows:

Step 1: when an unknown transient event occurs, the corresponding load feature indices are extracted and the feature vector of the event is constructed as $F_{new} = [T, I_f, I_b, I_{max}, P_f, Q_f, P_b, Q_b, P_{max}, Q_{max}]$. Each component is a dimensionless value normalized by Z-score standard normalization.

Step 2: take the load feature vectors of the transient events corresponding to known load operations as cluster centres. So, the cluster set can be designated as $C = [C_1, C_2, C_3, \dots, C_G]$, where G is the number of known transient events and C_i ($i = 1, 2, \dots, G$) refers to the i th cluster centre. The data dimensions and vector elements of each cluster centre in C are the same as F_{new} . Then, calculate the Manhattan distances between F_{new} and each cluster centre in C , respectively, which can be depicted as $D = [d_1, d_2, d_3, \dots, d_G]$, i.e. $d_i = |F_{new} - C_i|$.

Step 3: find the minimum distance d_{min} in D and select the appropriate threshold β . If $d_{min} < \beta$, then F_{new} belongs to the cluster centre corresponding to d_{min} . Otherwise, F_{new} should be added to set C as a new cluster centre.

Step 4: when a new transient event occurs, repeat Step 1 to Step 3.

The setting of distance threshold β depends on the reliability and accuracy of load monitoring equipment. Herein, the DL850 oscillograph recorder is operated for recording the transient waveform and related load characteristics of common appliances. Then, clustering is conducted to find the β parameter values with the best clustering effect.

4. Experimental Analysis and Results

4.1. Evaluation Index. The least unified residue (LUR) is selected to evaluate the accuracy of load decomposition. The LUR is a measure between an unknown composed load compared with the known database items, and it is defined as follows [3]:

$$E = \frac{\|y - \hat{y}\|_2}{\|y\|_2}, \quad (14)$$

where \hat{y} refers the composite load feature vector estimated by the introduced algorithm and y is the feature vector of an unknown load-on bus-side as defined in formula (8).

4.2. Simulation Results. In this paper, the NILM method is applied in the industrial application scenario [31]. Four kinds of industrial loads are investigated, including blender, cement screw, sewage pump, and inclined belt conveyor, as shown in Table 4. The industrial load dataset has collected 30 hours of bus-side three-phase current and voltage signals from a concrete plant by DL850 oscillograph recorder. Considering the influence of transformers, on the bus-side, the current transformer (CT) ratio is 100:1, the potential transformer (PT) ratio is 1:1, and the sampling frequency is 10 kHz. The proposed load decomposition-identification model is performed to this application scenarios.

TABLE 4: Application scenario.

Load equipment	Number	Rated power (kW)
Blender	2	110
Cement screw	4	18.5
Sewage pump	2	5.5
Inclined belt conveyor	2	55

For three-phase power, the amplitude of voltage or current is the same, but the phase of the waveform differs from each other. The phase of voltage or current is not a determining issue for load decomposition. Hence, data preprocessing is carried out for only one phase.

4.2.1. Data Preprocessing. In this paper, random error data are eliminated by using the Irada criterion to get relatively smooth waveform. Then, according to the Nyquist sampling theorem, the data are sampled at 1 kHz frequency, instead of the original 10 kHz, to balance the efficiency and accuracy of sampling. Figure 4 shows the comparison between the original current waveform and the downsampling of sewage-pump-on event. As shown in Figure 4, the downsampling data do not affect the extraction of load feature, while reducing the data quantity to 1/10 of the original.

4.2.2. Decomposition for the Steady-State Process. In terms of the previous analysis, the RMS value of steady-state current, the active power, and the reactive power are selected as the feature indices. Therefore, the load feature matrix H , which is mentioned in formula (8), is constructed as shown in Table 5. The load feature matrix H here includes four kinds of loads, and each load is described by three feature indices, so matrix H here is a $3 * 4$ matrix.

In order to verify the effectiveness of the proposed algorithm, four groups of unknown power on the bus-side are collected as test data, i.e., y_1, y_2, y_3 , and y_4 . Each test data is the feature vector y in formula (8), which is described with the same features as the load feature matrix H . With reference to the feature indices and extraction methods of the known loads, the corresponding features of each unknown power are extracted, respectively, as shown in Table 6.

Since the units of each feature index are not uniform and the load decomposition results must be nonnegative, the min-max normalization is chosen for dimensionless processing, as formula (6) delineated. Considering the test data y_1 , the proposed load decomposition algorithm is applied to the normalized feature matrix as shown in Table 7.

Figure 5 shows the influence of iterations of the LSQR algorithm on the accuracy of load decomposition. As the number of iterations increases, the accuracy of load decomposition also improves, while when the number of iterations reaches a certain level, the decomposition accuracy does not improve significantly. It can be seen that the optimal iteration number is 3 for y_1 in Figure 5.

Each group of unknown powers are decomposed into a linear combination of four known loads to complete steady-state load decomposition. Based on the optimal iteration of the LSQR algorithm, the decomposition results of each

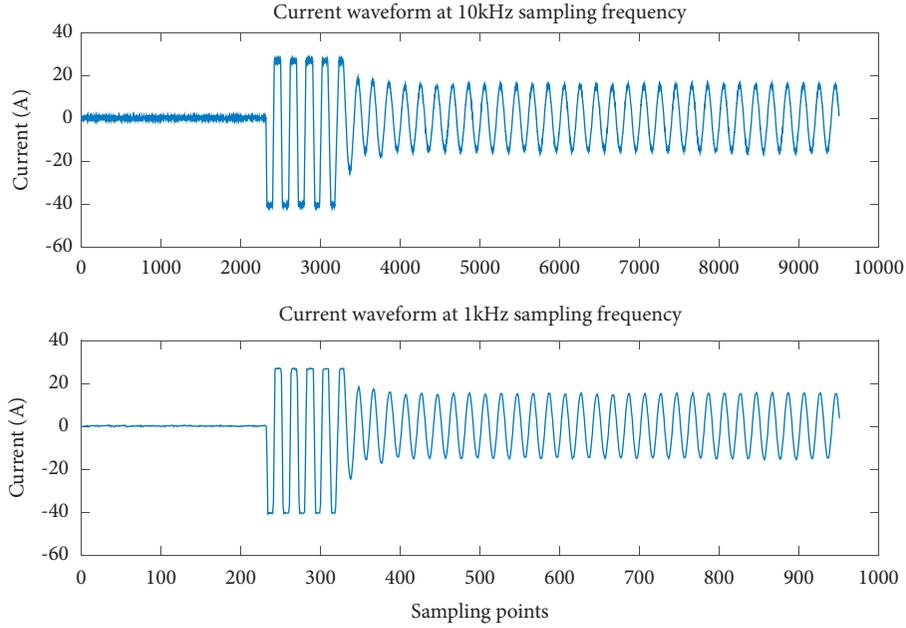


FIGURE 4: Current waveforms before and after data preprocessing for “sewage pump on” event.

TABLE 5: Load steady-state feature matrix table.

Feature indices, unit	Load types			
	Blender	Cement screw	Sewage pump	Inclined belt conveyor
RMS value of steady-state current, A	31.849	25.848	10.806	1.804
Active power, W	5002.000	4725.900	1992.665	312.759
Reactive power, var	5393.400	4821.300	1494.499	275.828

TABLE 6: Features of each unknown power on the bus-side.

Feature indices, unit	Test data			
	y_1	y_2	y_3	y_4
RMS value of steady-state current, A	66.464	251.879	150.383	104.778
Active power, W	12129.499	51307.556	27066.371	18913.008
Reactive power, var	9731.272	41163.100	21714.847	15173.555

TABLE 7: Normalized load steady-state feature matrix table.

Feature indices	Load types				
	Blender	Cement screw	Sewage pump	Inclined belt conveyor	y_1
RMS value of steady-state current	0.0000	0.3718	0.1392	0.0001	1.0000
Active power	0.3968	0.3735	0.1422	0.0001	1.0000
Reactive power	0.5412	0.4807	0.1289	0.0001	1.0000

compound power are shown in Table 8. For test data y_2 , when the number of iterations exceeds one, the decomposition weight appears negative. It does not meet the constraint conditions of (8); hence, the decomposition result of y_2 is obtained by one iteration.

The decomposition weight of each known load is calculated by the proposed identification algorithm, and the weight of each load represents the proportion of the load in the composited power on the bus-side. For instance, the decomposition results of test data y_1 are 0.249 (blender),

0.340 (cement screw), 5.445(cement screw), and 0.003 (inclined belt conveyor), which indicate the proportion of four loads. It is showed that the sewage pump accounts for the largest proportion. According to the decomposition weight of each load, the LUR of the predicted load combination and unknown power on bus-side are calculated to measure the accuracy of load decomposition.

In order to verify the effectiveness of the proposed algorithm, considering a single sample, the load identification models in [11] (FHMM) and in [12] (ANN) are selected to

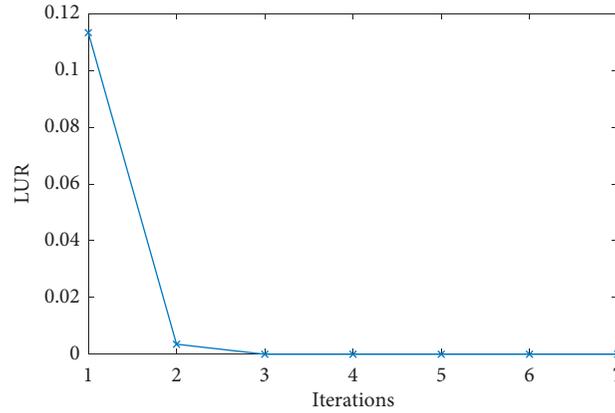


FIGURE 5: Relationship between the iterations of the LSQR algorithm and the LUR.

TABLE 8: Load decomposition results for the steady-state process.

Test data	Load decomposition weights				Optimal iteration number	LUR
	Blender	Cement screw	Sewage pump	Inclined belt conveyor		
y_1	0.249	0.340	5.445	0.003	3	1.433E-16
y_2	4.766	4.152	1.395	0.004	1	0.108
y_3	1.046	0.187	12.514	0.016	4	6.410E-16
y_4	0.656	0.207	8.7	0.008	4	8.988E-11

TABLE 9: Comparison of identification results of different models.

Algorithm	LUR	Calculation time, s
FHMM [11]	1.25E-1	—
ANN [12]	1.04E-2	7.88E-4
The proposed model	1.433E-16	5.31E-2

TABLE 10: Cluster centres for the transient process.

Feature indices	Load types							
	Load on				Load off			
	Class 1 (blender)	Class 2 (cement screw)	Class 3 (sewage pump)	Class 4 (inclined belt conveyor)	Class 5 (blender)	Class 6 (cement screw)	Class 7 (sewage pump)	Class 8 (inclined belt conveyor)
I_f	-0.273	-3.584	-4.028	-11.838	-116.352	-3.643	-12.271	-0.056
I_b	-0.273	-3.546	-4.001	-11.828	-117.160	-3.693	-12.407	-0.056
P_f	-0.211	-3.530	-3.904	-7.277	81.714	5.545	13.691	-0.055
P_b	-0.191	3.258	1.167	-5.587	-47.587	-3.603	-11.525	-0.056
Q_f	-0.207	-3.529	-3.935	-7.819	97.311	5.732	7.165	-0.055
Q_b	-0.185	3.396	-0.132	-6.328	-42.108	-3.602	-11.747	-0.056
I_{\max}	-0.267	-3.535	-3.946	-11.691	-116.358	-3.648	-12.271	-0.056
P_{\max}	0.755	5.376	11.222	18.452	80.803	4.485	13.691	-0.055
Q_{\max}	0.835	5.557	7.409	14.872	96.328	5.732	7.165	-0.055
T	0.017	0.137	0.148	29.044	83.410	-3.306	18.508	0.501

compare with the proposed model in terms of LUR and calculation time. The identification results of different models are shown in Table 9. It can be seen that although the proposed algorithm does not have advantages in calculation time compared with other novel machine learning models, it has obvious advantages in accuracy.

4.2.3. *Identification for the Transient Process.* 10 feature indices are chosen in terms of the transient feature space established in Table 2, and the load types selected are consistent with steady-state load decomposition. The transient load identification underlying on sequential leader clustering algorithm firstly needs to construct the cluster

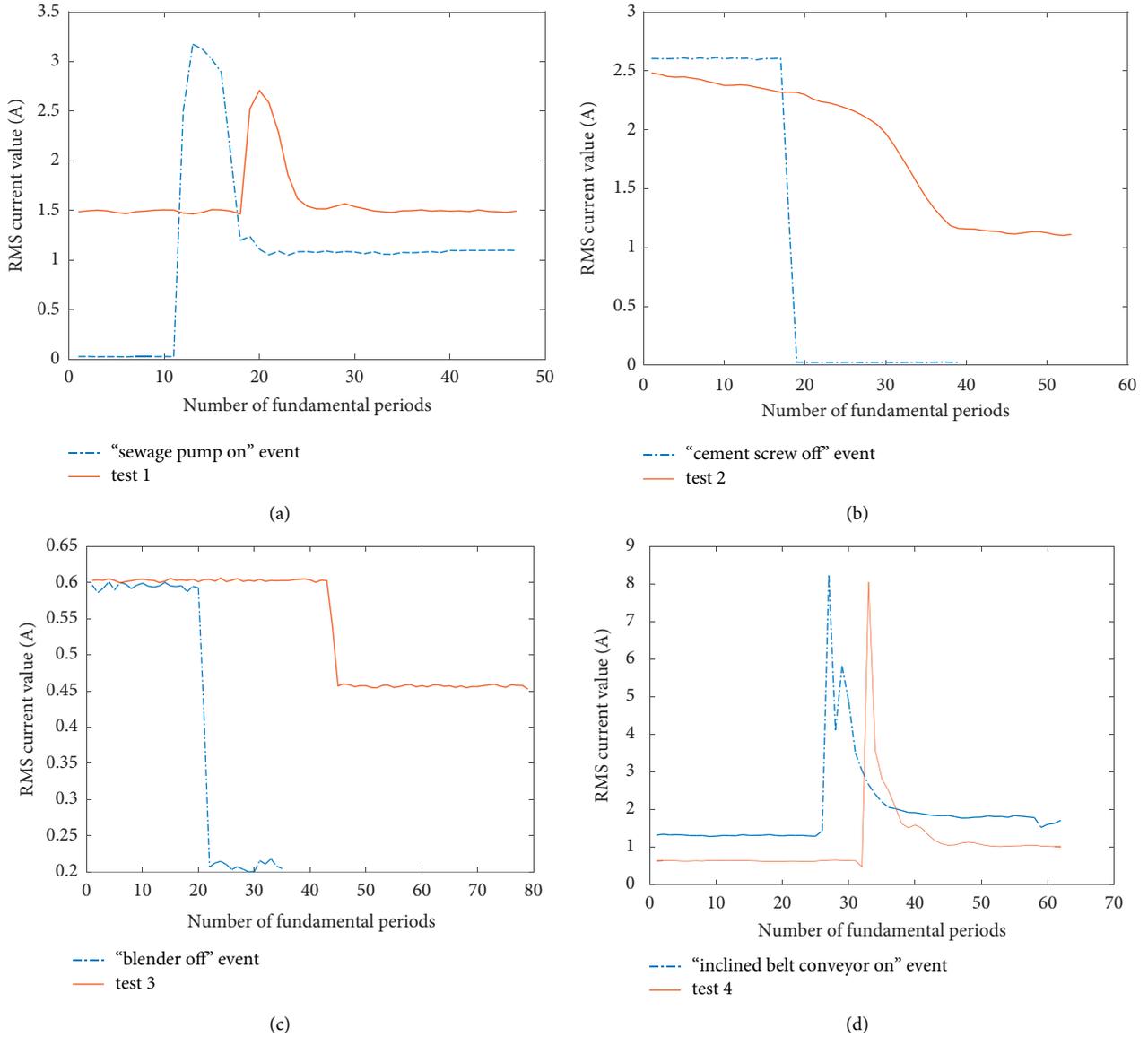


FIGURE 6: Load identification results underlying on clustering.

centres, which signify transient events of typical loads determined by steady-state load decomposition. After a dimensionless processing on the basis of Z-score normalization, as formula (7) defined, the cluster centres can be constructed as Table 10. The eight cluster centres denote eight load on/off events, i.e., blender on/off (class 1/class 5), cement screw on/off (class 2/class 6), sewage pump on/off (class 3/class 7), and inclined belt conveyor on/off (class 4/class 8). According to the experiment, the best clustering effect threshold β is 0.25.

When unknown transient events occur on the bus-side, the related feature indices are extracted and normalized, and then they are classified by the clustering algorithm, so as to identify the specific load switching operation corresponding to the event.

There are four unknown test events (i.e., test 1, test 2, test 3, and test 4), and Figure 6 illustrates the comparison

between unknown test events and cluster centre events on the feature of current RMS value. For example, Figure 6(a) shows the comparison of the RMS current value of unknown transient event test1 with the cluster centre event "sewage pump on." The figure shows that the fitting degree, between the clustered identification results and cluster centre events, is basically consistent. Therefore, the effectiveness of the method presented is further verified.

5. Conclusions

This paper proposes a load decomposition-identification method for industrial users. In order to analyze the transient process caused by load switching operations and the steady-state process after load-on events, the steady-state feature space and the transient feature space of industrial loads are constructed, respectively. LSQR algorithm is proposed to

solve load decomposition least square problem for the steady-state process. The sequential leader clustering algorithm is adopted to classify transient events of the transient process to identify load switching operations. Simulation results prove that the disposed method can effectively decompose the specific loads from unknown composed loads and accurately identify the load switching operation with an error rate of about 10^{-16} . Furthermore, a general process of load identification for industrial users is constructed, including feature extraction, data preprocessing, steady-state, and transient load decomposition-identification. On this basis, the detailed operation information of various loads can be acquired from overall load signals. Nevertheless, it can acquire detailed operations of various loads. The algorithm introduced is based on the industrial users' load signal collected by an oscillograph recorder, and further research on dynamic real-time signal acquisition and online analysis is still needed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the Science and Technology Project of State Grid (Research and Typical Application of Nonintrusive Load Identification and Perception Technology for Special Transformer Users, 5600-202024168A-0-000).

References

- [1] J. W. M. Cheng, G. Kendall, and J. S. K. Leung, *Electric-Load Intelligence (E-LI): Concept and Applications, TENCON 2006-2006 IEEE Region 10 Conference*, pp. 1-4, IEEE, Piscataway, NJ, USA, 2006.
- [2] C. W. Gellings, "Evolving practice of demand-side management," *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 1, pp. 1-9, 2017.
- [3] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load signature Study—Part I: basic concept, structure, and methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 551-560, 2010.
- [4] C. Dinesh, P. Perera, R. Indika Godaliyadda, B. Parakrama, M. Ekanayake, and J. Ekanayake, "Non-intrusive load monitoring based on low frequency active power measurements," *AIMS Energy*, vol. 4, no. 3, pp. 414-443, 2016.
- [5] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 448-461, 2019.
- [6] S. Welikala, N. Thelasingha, M. Akram, P. B. Ekanayake, R. I. Godaliyadda, and J. B. Ekanayake, "Implementation of a robust real-time non-intrusive load monitoring solution," *Applied Energy*, vol. 238, pp. 1519-1529, 2019.
- [7] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle, "Circuit-level load monitoring for household energy management," *IEEE Pervasive Computing*, vol. 10, no. 1, pp. 40-48, 2011.
- [8] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870-1891, 1992.
- [9] P. Franco, J. M. Martinez, Y. C. Kim, and M. A. Ahmed, "IoT based approach for load monitoring and activity recognition in smart homes," *IEEE Access*, vol. 9, pp. 45325-45339, 2021.
- [10] H. Kim, M. Marwah, and M. Arlitt, "Unsupervised disaggregation of low frequency power measurements," 2011, http://hanj.cs.illinois.edu/pdf/sdm11_hkim.pdf.
- [11] B. Yin, Z. Li, J. Xu, and L. Li, "Non-intrusive load monitoring algorithm based on household electricity use habits, neural computing and applications," 2021, <https://link.springer.com/article/10.1007/s00521-021-06088-2>.
- [12] J. Kelly and W. Knottenbelt, "Neural NILM: deep neural networks applied to energy disaggregation," 2015, <https://arxiv.org/abs/1507.06594>.
- [13] H. H. Chang, M. C. Lee, W. J. Lee, C. L. Chien, and N. Chen, "Feature extraction-based hellinger distance algorithm for nonintrusive aging load identification in residential buildings," *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 2031-2039, 2016.
- [14] C. Liu, A. Akintayo, Z. Jiang, G. P. Henze, and S. Sarkar, "Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network," *Applied Energy*, vol. 211, pp. 1106-1122, 2018.
- [15] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1739-1747, 2018.
- [16] Z. Zheng, H. Chen, and X. Luo, "A supervised event-based non-intrusive load monitoring for non-linear appliances," *Sustainability*, vol. 10, no. 4, pp. 1001-2071, 2018.
- [17] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 870-878, 2014.
- [18] H. H. Chang, K. L. Lian, Y. C. Su, and W. J. Lee, "Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification," *IEEE Transactions on Industry Applications*, vol. 50, no. 3, pp. 2081-2089, 2014.
- [19] M. Drouaz, B. Colicchio, A. Moukadem, A. Dieterlen, and D. Ould-Abdeslam, "New time-frequency transient features for nonintrusive load monitoring," *Energies*, vol. 14, no. 5, 2021.
- [20] Y. H. Lin and M. S. Tsai, "Development of an improved time-frequency analysis-based nonintrusive load monitor for load demand identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1470-1483, 2014.
- [21] J. Wang and P. Srikantha, "Stealthy black-box attacks on deep learning non-intrusive load monitoring models," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3479-3492, 2021.
- [22] H. H. Chang, K. L. Chen, Y. P. Tsai, and W. J. Lee, "A new measurement method for power signatures of nonintrusive demand monitoring and load identification," *IEEE Transactions on Industry Applications*, vol. 48, no. 2, pp. 764-771, 2012.
- [23] S. Kong, Y. Kim, R. Ko, and S. K. Joo, "Home appliance load disaggregation using cepstrum-smoothing-based method," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 1, pp. 24-30, 2015.

- [24] Y. Y. Hong and J. H. Chou, "Nonintrusive energy monitoring for microgrids using hybrid self-organizing feature-mapping networks," *Energies*, vol. 5, no. 7, pp. 2578–2593, 2012.
- [25] J. Yu, X. Wu, and W. Liu, "Noninvasive industrial power load monitoring based on collaboration of edge device and edge data center," 2020, <https://ieeexplore.ieee.org/document/9284219>.
- [26] S. Yi, X. Yin, and Y. Diao, "A new event-detection method based on composite windows in nilm for industrial settings," 2019, <https://ieeexplore.ieee.org/document/8975265>.
- [27] C. C. Paige and M. A. Saunders, "LSQR: an algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software*, vol. 8, no. 1, pp. 43–71, 1982.
- [28] Z. Asgari, F. Toutounian, and E. Babolian, "Augmented subspaces in the LSQR Krylov method," *Iranian Journal of Science and Technology Transaction A-Science*, vol. 44, no. 6, pp. 1661–1665, 2020.
- [29] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.
- [30] W. N. Qian and A. Y. Zhou, "Analyzing popular clustering algorithms from different viewpoints," *Journal of Software*, vol. 13, no. 8, pp. 1382–1394, 2002.
- [31] W. Luan, F. Yang, B. Zhao, and B. Liu, "Industrial load disaggregation based on hidden Markov models," *Electric Power Systems Research*, vol. 210, 2022.