Simultaneous Load Identification Method Based on Hybrid Features and Genetic Algorithm for Nonintrusive Load Monitoring

Shu-Hui Yi, Jian Wang, and Jun-Jie Liu

Department of Metrology, China Electric Power Research Institute, Wuhan 430070, China

Correspondence should be addressed to Shu-Hui Yi; 1287460331@qq.com

Received 5 February 2022; Revised 15 April 2022; Accepted 18 April 2022; Published 19 May 2022

Nonintrusive load monitoring (NILM) is a widely accepted technology to conduct load monitoring. Many effective methods have been established to make NILM more practical. However, the focus of current methods is mainly on the identification accuracy and efficiency of single load under the individual appliance operated independently, which have limited support for the identification problem under multiple appliances operated simultaneously. Therefore, a simultaneous identification method is proposed to efficiently identify the total load under multiple appliances operated simultaneously in this paper. The proposed identification method mainly consists of three parts: hybrid features extraction, simultaneous identification optimization model construction, and frequency-weighting-factor-based genetic algorithm (FWF-GA). Firstly, the hybrid feature model, which integrates the features of active power, reactive power, and harmonic magnitude, is constructed by hybrid features extraction. Secondly, the simultaneous identification optimization model is constructed by employing the features of active and reactive power. Thirdly, the developed FWF-GA is used to solve the simultaneous identification optimization problem. In FWF-GA, the relative errors of active power, reactive power, and the frequency-weighting factor of harmonic magnitude are used to evaluate the fitness of an individual. Finally, a NILM practice to identify household appliances is used to demonstrate the validity of the proposed method.

1. Introduction

As a high-grade energy carrier, electricity plays a critical role in economic development [1]. With the development of the economy and society, the demand for electrical power will grow continually. Since electricity production relies on other energies as well as its feature is unsuited to mass storage, energy-saving becomes an urgent need for modern power systems [2]. One best way to optimize energy savings is to conduct load monitoring, which has been widely used in smart grids. Many investigations reveal that both electricity consumers and providers can benefit from load monitoring [3–5]. Customers can smartly use their appliances with the detailed information of load monitoring, such as making a more rational strategy to reduce energy consumption and identifying the failure appliances and inefficient appliances.

For electricity providers, some notable benefits derived from load monitoring include making public policy on electricity consumption, segmenting customers, predicting energy demand, and even detecting illegal electricity behaviors [6].

Traditional load monitoring is conducted by analyzing the collected data that are obtained from the sensors installed on each appliance. This monitoring manner is also called intrusive load monitoring (ILM). The technique advantage of ILM is that it can directly obtain the relatively accurate and reliable monitoring data of the individual appliance, while the key disadvantage lies in the hardware cost, the complicated maintenance, as well as the large number of installation work. In contrast, a cost-effective technique named nonintrusive load monitoring (NILM) has been developed to make load monitoring more practical. NILM only needs a single minimal set of plug-panel sensors.
to acquire the aggregate electric consumption data, and then, the acquired data is treated as a disaggregation problem. NILM, therefore, also refers to load disaggregation [7]. Benefiting from the advantage of technology, NILM has gained tremendous attention and is becoming more and more popular.

NILM aims to identify the operation state of the individual appliance from the information of aggregate load. There are typically four steps summarized in general NILM procedures, including data acquiring, event detecting, feature extracting, and load identifying. Since feature extracting and load identifying dominate more to the efficiency and accuracy, the current researches around NILM mainly focus on these two aspects. The task of feature extraction is to obtain relevant features so as to discriminate different types of appliance loads. It is the inherent requirement to extract features accurately and effectively. Researchers therefore pay especially attention to the feature representation. Many feature representations are developed based on current, voltage, and power. Among them, real power and reactive power are most widely used for their stable performance [8]. Besides, harmonic-based features were proposed to conduct feature extraction as a supplement to real and reactive power features [9–11]. To better detect the transient events, time-frequency tools were introduced to model the nonstationary features. Lin and Tsai [12] proposed a transient feature extraction, where the transient response was modeled based on S-transform. Shaw et al. [13] adopted the short-time Fourier transforms to compute the spectral envelope coefficients during the NILM implementation. Gillis and Morsi [14] applied discrete wavelet transform to represent transient features in event-efficient identification. Another new feature representation method is shape representation, which can directly obtain efficient feature information from the raw current and voltage signals [15–17]. The shape representation was also extended to transient features modeling, such as Stockwell transform-based time-frequency feature [18]. Moreover, He et al. [19] adopted the probability density function to model multiple operating states, where both the power measurements and current signals were considered. As different feature representation indicates the different characteristics of appliances, the synthetic feature is increasingly used in NILM for accurate and efficient identification [20, 21].

The second research focus of the current NILM is load identification. It is a significant goal of NILM that the operating state of individual electrical appliances can be accurately discriminated from the aggregate consumption signal. To achieve this, load identification as one crucial step in NILM is used to distinguish the different types of load. Large quantities of methods were developed in the load identification area, which can be roughly grouped into two categories. The first category is called machine learning-based methods. This type of methods can be implemented without event detection and usually has high efficiency. During its implementation, the discrimination of different types of load is generally treated as a multilabel classification problem, and therefore, the machine learning methods, including unsupervised, semisupervised, and supervised methods, all can be employed to deal with the identification problems. Tabatabaei et al. [22] developed two supervised multilabel classifiers based on a k-nearest neighbor. Lai et al. [23] proposed a hybrid model in which the support vector machine is integrated with the Gaussian mixture model. The hidden Markov model and its improved models were used as the unsupervised classification to implement the load identification, providing a well performance of data clustering [24–26]. Yang et al. [27] proposed a semisupervised multilabel classification method based on deep learning to further improve the efficiency of the deep learning-based identification methods. The second category is optimization-based methods. In the optimization-based method, the load identification is treated as a mathematical optimization problem, where the total consumption signal is considered as the combination of the individual appliance to be identified. For a long time, the optimization-based method was adopted as another identification method, which is always used to identify one load at a time. Therefore, efficiency and accuracy were especially concerned. Some well-known methods, such as mixed-integer linear programming [28] and quadratic programming (QP) [29, 30], were already successfully applied in optimization-based load identification.

Recently, the attention to load identification is also drawn to simultaneous identification problems for multiple electrical loads exist. Hua et al. [31] proposed an event-based NILM method, in which the mixed linear integer programming model was developed to achieve the simultaneous disaggregation of multiple appliances. Liu et al. [32] proposed a deep dictionary learning model and an adaptive-window-based detection approach to conducting the NILM with simultaneous switching operations. Sun et al. [33] proposed the dynamic adaptive particle swarm optimization algorithm (DAPSO) to realize nonintrusive household load identification. Fang et al. [34] proposed a NILM method that leverages advances in statistical learning. Yang et al. [35] proposed an intelligent event-driven nonintrusive load monitoring method based on a convolutional neural network, which extends the NILM application range to the handling process in terms of various and combined characteristics. Furthermore, Zhou et al. [36] proposed a NILM method based on the CNN-LSTM hybrid model to improve the performance of the whole network system. Since different users have different habits, various cases of power consumption are possible. The simultaneous identification therefore is a direction worthy of development and research [37].

The above feature representation and load identification methods can efficiently improve the quality of NILM. There is a clear trend that more and more feature representations use the synthetic feature derived from the different aspects of appliance load. This is because the synthetic feature can better distinguish different appliances compared with a single feature. Furthermore, the good feature representation is always combined with an excellent load identification method to strengthen its practicability. Although many load identification methods have already been developed and well used in practical NILM, their focus is mainly on the single
load identification at a time. The support for the identification problem of the total load under multiple appliances operated simultaneously is still very limited, and both the identification model and the efficient solution algorithm are needed to be further developed. Therefore, to achieve the identification of the total load under multiple appliances operated simultaneously, a simultaneous load identification method based on hybrid features and a genetic algorithm is proposed in this paper, in which the frequency weighting factor-based genetic algorithm (FWF-GA) is developed to efficiently solve the simultaneous identification problem.

The main contributions are highlighted as follows:

(1) This paper proposes for the first time a fitness function based on the frequency weighting factor that is used to evaluate the fitness of individuals powerfully in FWF-GA. This proposed fitness function not only takes the influence of the harmonic feature into account but also considers the relative error of active power and reactive power, and more importantly, during the evaluation of the fitness function, the influence of negligence due to the large gap between active power and reactive power will be avoided.

(2) The test results show that the proposed method has better accuracy and efficiency compared with the method of QP, DAPSO, and CNN. Meanwhile, the proposed method does not need a large amount of data for training like the method based on CNN.

The rest of the paper is organized as follows. Section 2 introduces related preliminaries about NILM. Section 3 proposes the simultaneous identification method, including hybrid features, a simultaneous identification optimization model, and the procedure of FWF-GA. Section 4 mainly uses the load data of five household appliances to test the effectiveness of the proposed method. Finally, Section 5 concludes this work and suggests future work.

2. Preliminaries

2.1. Discrete Fourier Transform. The signal of the time domain can be transformed into the signal of the frequency domain by using discrete Fourier transform (DFT) [38]. The DFT consists in decomposing a signal into a sum of elementary signals, which have the property of being easy to implement and observe [39]. On the basis of DFT, the time signal $A = [a_k] = [a_0, a_1, \ldots, a_{K-1}]$ is transformed into the frequency sequence $B = [b_m] = [b_0, b_1, \ldots, b_{K-1}]$. The DFT is defined by

$$\begin{align*}
B &= [b_m] = DFT[a_k], \\
b_m &= \sum_{k=0}^{K-1} a_k e^{-j2\pi km/K}, 0 \leq m \leq K-1,
\end{align*}$$

where $a_k$ is the $k^{th}$ value of the signal $A$, $b_m$ is the $m^{th}$ DFT value of the signal $A$, $K$ is the size of $A$, and $S$ is the size of $B$.

For each value of $m$ in equation (2), the number of calculations to obtain each value is $K$, and therefore, a total calculative number $K^2$ is required to get the frequency sequence by DFT. When the number $K$ becomes large, the computational cost of common DFT will be extremely large. Consequently, the fast Fourier transform (FFT) is proposed to improve computational efficiency.

According to the symmetry of $e^{-j2\pi km/K}$, the time signal $A$ can be decomposed into the sum of two sequences, which is defined by

$$A = A_1 + A_2,$$

where $A_1$ and $A_2$ are the time sequences of size $K/2$ and $A_1$ is the even sequence, while $A_2$ is the odd sequence. Let $W_K = e^{-j2\pi/K}$, then equation (2) can be transformed into the following form:

$$B = [b_m] = B_1 + W_k^m B_2,$$

where $B_1 = [b_{m_1}]$ and $B_2 = [b_{m_2}]$ are, respectively, the frequency sequence after the DFT of the time signal $A_1$ and $A_2$. $b_m$ can be obtained by

$$b_m = \sum_{k=0}^{K/2-1} a_{2k} W_K^{2km} + \sum_{k=0}^{K/2-1} a_{2k+1} W_K^{2k+1+m} = \sum_{k=0}^{K/2-1} a_{2k} W_K^{2km} + W_K^m b_{m_2},$$

$$0 \leq m_1, m_2 \leq \frac{K}{2} - 1, 0 \leq m \leq K - 1.$$

2.2. Load Feature. The function and internal structure of electrical equipment are different, and the load features are also different. Therefore, different load features can describe the operation of different electrical equipment. In the current study on NILM, the common load features are the current waveform features, current harmonic features, power features, and so on.

2.2.1. Current Waveform Features. From the current waveform, three features can be extracted, which are given by
2.2.2. Harmonic Features. The harmonic features can reflect the structure of electrical equipment. The harmonic features can be calculated by DFT, which can be given by the following equations:

\[
\begin{align*}
M & = |b_m| = \sqrt{\text{Re}(b_m)^2 + \text{Im}(b_m)^2}, \\
P & = \arg(b_m) = \arctan\left(\frac{\text{Im}(b_m)}{\text{Re}(b_m)}\right), \\
k_{T_{\text{THD}}} & = \sqrt{\sum_{k=2}^{\infty} \left(\frac{G_k}{G_1}\right)^2},
\end{align*}
\]

where \(M\), \(P\), and \(k_{T_{\text{THD}}}\) are the magnitude, phase, and total harmonic distortion (THD) of the frequency signal, respectively; \(\text{Im}\) and \(\text{Re}\) represent the imaginary and the real part of the DFT, respectively; \(G_k\) is the harmonic component of the \(k^{th}\) order; \(G_1\) is the fundamental component; and \(L\) is the maximum of effective harmonic order.

2.2.3. Power Features. The power features are the most common indicators to describe the electrical behavior of electrical equipment. The power features contain active power and reactive power, which can be calculated as follows:

\[
\begin{align*}
P & = \sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} U_k I_k \cos(\varphi_k), \\
Q & = \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} U_k I_k \sin(\varphi_k),
\end{align*}
\]

where \(P\) is the active power, \(Q\) is the reactive power, \(U\) is the voltage, \(I\) is the current, \(k\) is the harmonic order, and \(\varphi\) is the phase difference between voltage and current.

3. Simultaneous Identification Method

The proposed simultaneous identification method is mainly used to deal with the problem involving simultaneously identifying multiple loads in NILM. The proposed method is developed on the basis of hybrid features and genetic algorithm, which mainly consists of three parts: hybrid features extraction, simultaneous identification optimization model construction, and optimization model solution by using frequency weighting factor-based genetic algorithm (FWF-GA). The frame of the proposed simultaneous identification method is shown in Figure 1.

3.1. Hybrid Features. In this paper, a hybrid feature model is proposed for NILM prediction. The proposed model is a set of features that is composed of active power, reactive power, and harmonic magnitude. The hybrid features of the appliances are expressed in the following matrix form:

\[
H = \begin{bmatrix}
P_1 & Q_1 & M_{11} & \ldots & M_{1j} & \ldots & M_{1L} \\
P_2 & Q_2 & M_{21} & \ldots & M_{2j} & \ldots & M_{2L} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
P_i & Q_i & M_{i1} & \ldots & M_{ij} & \ldots & M_{iL} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
P_N & Q_N & M_{N1} & \ldots & M_{Nj} & \ldots & M_{NL}
\end{bmatrix},
\]

where \(P_i\) and \(Q_i\) are the active and reactive power of \(i^{th}\) appliance, respectively, \(M_{ij}\) is the \(j^{th}\) harmonic magnitude of \(i^{th}\) appliance, and \(N\) is the total number of appliances. For clarity of presentation, the matrix \(H\) also can be regarded as the combination of matrix \(H = [H_P, H_Q, H_M]\). \(H_P, H_Q,\) and \(H_M\) are given as follows:

\[
H_P = [P_1, P_2, \ldots, P_N]^T, \\
H_Q = [Q_1, Q_2, \ldots, Q_N]^T, \\
H_M = \begin{bmatrix}
M_{11} & \ldots & M_{1j} & \ldots & M_{1L} \\
M_{21} & \ldots & M_{2j} & \ldots & M_{2L} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_{i1} & \ldots & M_{ij} & \ldots & M_{iL} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_{N1} & \ldots & M_{Nj} & \ldots & M_{NL}
\end{bmatrix},\]

Figure 1: Frame of the proposed simultaneous identification method.
where the superscript $T$ denotes the transposition of vectors or matrices. It is noted that when performing the simultaneous identification, both the hybrid feature model of the individual appliance operated independently and multiple appliances operated simultaneously need to be given out.

The single load under the individual appliance operated independently and the total load under multiple appliances operated simultaneously are obtained by real-time monitoring. Through the feature extracting and DFT, the hybrid feature models of single load and total load can be given by

$$
H_S = [H_{S-P}, H_{S-Q}, H_{S-M}],
$$

$$
H_T = [H_{T-P}, H_{T-Q}, H_{T-M}],
$$

where the subscript $S$ denotes the single load that is a load of the individual electrical appliance operated independently and the subscript $T$ denotes the total load that is a load of multiple appliances operated simultaneously. It is noted that $H_{T-P}$ and $H_{T-Q}$ are both constant values, which are defined as $H_{T-P} = P$ and $H_{T-Q} = Q$, respectively, while $H_{T-M}$ is one-dimensional vector, that is, $H_{T-M} = [M_{T1}, M_{T2}, \ldots, M_{Tj}, \ldots, M_{TL}]$, where $M_{Tj}$ denotes the $j$th harmonic magnitude of the total load.

3.2. Simultaneous Identification Optimization Model. The hybrid feature model $H$ is divided into three components due to different uses. The vector $H_P$ and vector $H_Q$ are used to construct the optimization model of simultaneous identification, while the matrix $H_M$ is introduced to the algorithm developed for solving the optimization model. The object of load identification is to find the best combination of the appliance load to approach the total power consumption, and thus, the most possible appliances switched on can be inferred. The objective function therefore is given as follows:

$$
\min f(x) = |H_{T-P} - xH_{S-P}| + |H_{T-Q} - xH_{S-Q}|,
$$

where $x$ is the vector of design variables corresponding to each appliance, that is, $x = [x_1, x_2, \ldots, x_i, \ldots, x_N]$, and $x_i$ is the design variable corresponding to the $i$th appliance.

Considering the limit case of simultaneous appliances, the simultaneous identification problem can be established by an optimization problem, which is given by

$$
\begin{align*}
\min f(x) &= |H_{T-P} - xH_{S-P}| + |H_{T-Q} - xH_{S-Q}| \\
\text{s.t.} & \sum_{i=1}^{N} x_i \leq N, \quad x_i = 0 \text{ or } 1.
\end{align*}
$$

The value of $x_i$ is 0 or 1, and 0 denotes that the $i$th appliance is not running, while 1 denotes that the $i$th appliance is running. Therefore, the optimization problem is a 0-1 integer optimization problem. The genetic algorithm [40, 41] is useful for the integer optimization problem. Therefore, in order to improve the identification accuracy and efficiency, a genetic algorithm based on frequency weighting factors (FWF-GA) is developed to solve the simultaneous identification problem of equation (12).

3.3. FWF-GA. The procedure of simultaneous identification with FWF-GA is shown in Figure 2.

The procedure can be summarized in the following steps:

Step 1: Calculating frequency weighting factors
Since the frequency domain can directly reflect the inherent characteristics of appliance load from a frequency perspective, the frequency weighting factor is defined on the basis of frequency domain information. The frequency weighting factor is defined as follows:

$$
\bar{\xi}_i = \frac{1}{L} \left( \frac{M_{i1}}{M_{T1}} + \frac{M_{i2}}{M_{T2}} + \cdots + \frac{M_{ij}}{M_{Tj}} + \cdots + \frac{M_{il}}{M_{TL}} \right),
$$

where $\bar{\xi}_i$ is the frequency weighting factor of $i$th appliance.

As shown by equation (13), the frequency weighting factor is actually the average value of $L$ terms weight sum. To further clearly illustrate the defined frequency weighting factor, the preceding six terms weight are
take as an example, which is shown in Figure 3. It is worth noting that the value of $L$ is not absolutely fixed, that is to say, the value of $L$ still depends on an individual decision considering which ones play an important role.

Step 2: Constructing fitness function

The fitness evaluation is one crucial step in the GA process, whose fitness function directly determines the accuracy and efficiency of the algorithm. To further enhance the practicability of simultaneous identification problem solving, a new fitness function is developed based on the frequency-weighting factor, which is used to evaluate the fitness of an individual more efficiently.

The new fitness function is defined as follows:

$$ F_t = \left| 1 - \frac{1}{P} \sum_{i=1}^{n} x_i P_i \right| + \left| 1 - \frac{1}{Q} \sum_{i=1}^{n} x_i Q_i \right| + \left| 1 - \frac{\sum_{i=1}^{n} x_i \xi_i}{\sum_{i=1}^{n} \xi_i M T_i} \right| $$  

(14)

Step 3: Generating population

The binary coding is adopted in the proposed genetic algorithm. Each gene stands for the state of one appliance, and therefore, the number of genes that compose an individual equals the number of total appliances. Besides, the orders of gene on an individual correspond to different appliances, for example, four appliances (that are, respectively, marked as 1, 2, 3, and 4) with the state ON, OFF, ON, and ON, are encoded as 1011, in which the value 0 in the second place stands for the state of appliance marked as 2 is switched off. For the first step, the initial population can be obtained randomly.

Step 4: Calculating fitness value

According to equation (14), the fitness value under the current population is calculated. The smaller the value of individual fitness, the better the individual, that is, the combination of the appliance corresponding to the individual is closer to the realistic state.

Step 5: Judging convergence

The convergence criterion can be given by

$$ \frac{F_t^{\text{Cycle}} - F_t^{\text{Cycle}-1}}{F_t^{\text{Cycle}}} \leq \varepsilon \text{ or Cycle} \leq \text{Cycle}_{\text{max}} $$  

(15)

where Cycle denotes iteration time; $F_t^{\text{Cycle}}$ and $F_t^{\text{Cycle}-1}$ are the fitness value corresponding to the $\text{Cycle}^{\text{th}}$ and ($\text{Cycle} - 1)^{\text{th}}$ best individual, respectively; $\varepsilon$ is the convergence accuracy; and $\text{Cycle}_{\text{max}}$ is the maximum number of iterations set. If the convergence is achieved, stop the solution and go to step 9. Otherwise, go to step 6.

Step 6: Selecting and copying

In this step, the superior individuals are selected to generate the next population. The roulette wheel selection approach [42] is adopted to select the superior individuals. The selective probability of an individual is given as follows:

$$ \Pr(s_1) = \frac{F_t(s_1)}{\sum_{j=1}^{n} F_t(s_j)} $$  

(16)

where $\Pr$ stands for the probability, $s_1$ stands for the individual, and $n$ stands for the total number of individuals.

Step 7: Performing crossover and mutation

The crossover operations are first performed for the copied individuals, which are based on the crossover rate and the gene bit to be exchanged. Then, the mutation operations can be performed according to the given mutation rate.

Step 8: Updating population

Through step 7, the new population can be obtained. Then go to step 3 and continue with the next iteration.

Step 9: Obtaining a best individual

When the solution is stopped, give out the best individual, that is, give out the simultaneous identification results.

4. Example Verification

In this section, a load identification problem that contains five household appliances are used to demonstrate the proposed method. To verify the effectiveness of the proposed method, two aspects, efficiency and accuracy are mainly considered.

4.1. Example Description. Five different appliances including a notebook computer (NC), a fluorescent light (FL), a water fountain (WF), a hairdryer (HD), and a television (TV) are used to verify the proposed method. The single load and total load are obtained by real-time monitoring. The single load is the load when five appliances operate independently, but the total load is the mixed load when five appliances operate under different states.

The single load of five appliances is shown in Figure 4. As shown in Figure 4, there is only one operating state for the appliance of NC, FL, and TV, while there are many operating states for the appliance of WF and HD. For the WF, there are three states: heating, cooling, and keeping warm. For the HD, there are four states: 1st hot air, 1st cold air, 2nd hot air, and 2nd cold air. The active power and reactive power of each appliance are listed in Table 1. By using the FFT, the current harmonic characteristics of each appliance are shown in Figure 5.

Table 1. The active power and reactive power of each appliance

<table>
<thead>
<tr>
<th>Appliance</th>
<th>Active Power (kW)</th>
<th>Reactive Power (kvar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>FL</td>
<td>0.05</td>
<td>0.005</td>
</tr>
<tr>
<td>NC</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>FL</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>NC</td>
<td>0.3</td>
<td>0.03</td>
</tr>
<tr>
<td>FL</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>NC</td>
<td>0.4</td>
<td>0.04</td>
</tr>
<tr>
<td>FL</td>
<td>0.3</td>
<td>0.03</td>
</tr>
<tr>
<td>NC</td>
<td>0.5</td>
<td>0.05</td>
</tr>
<tr>
<td>FL</td>
<td>0.4</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Figure 6 shows the total load. It is noted that Figure 6 only gives the partial waveforms of mixed operation states. Eight group data marked in Figure 6 are used to illustrate the effectiveness of the proposed method. The mixed operating states of five appliances corresponding to eight group data are listed in Table 2.

In Table 2, the active power and reactive power are obtained by feature extraction from the total load. The mixed
operating states are the vector of on or off state corresponding to the appliance of NC, FL, heating of WF, cooling of WF and keep warm of WF, 1st hot air of HD, 1st cold air of HD, 2nd hot air of HD, 2nd cold air of HD, and TV. For example, the mixed operating states [1, 1, 1, 0, 0, 0, 1, 0, 0, 1] of fourth group data denote that the current operating state is NC, FL, heating of WF, 1st cold air of HD, and TV.

4.2. Solution Analysis of FWF-GA. The FWF-GA is used to perform the load identification. Considering the different operating states of various electrical appliances, it needs to set 10 design variables, which is \( x = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}] \), where \( x_1, x_2, \ldots, x_{10} \) respectively, stand for the on or off state corresponding to the appliance of NC, FL, heating of WF, cooling of WF, keep warm of WF, 1st hot air of HD, 1st cold air of HD, 2nd hot air of HD, 2nd cold air of HD, and TV. In addition, not only the sum of all the design variables cannot be greater than the total number of appliances, but also the sum of the design variables corresponding to WF and HD cannot be greater than the
maximum number of running states. Consequently, the simultaneous identification optimization model of the example can be established by

\[
\begin{align*}
\min f(x) &= |P - xH_p| + |Q - xH_q| \\
\text{s.t.} & \sum_{i=1}^{10} x_i \leq N; \sum_{i=3}^{5} x_i \leq 1; \sum_{i=6}^{9} x_i \leq 1, \\
x_i &= 0 \text{ or } 1.
\end{align*}
\]  \tag{17}

In the proposed FWF-GA, some system parameters significantly influence the evaluation performance, such as the crossover rate and population size. The sensitivity of these parameters is tested by the data set of mixed load exacted from Figure 6. The length of the data set is 5,000. Table 3 shows the sensitivity of the crossover rate and population size.

In Table 3, SIP is the single identification probability, that is, the probability of identifying each appliance in the total load identification, while AIP is the average identification probability, that is, the probability of identifying all the appliances in the total load identification. As can be seen from Table 3, when the crossover rate is 0.8, the SIP and AIP reach the highest. For the sensitivity of population size, when the population size is 50, the SIP and AIP reach the highest, but the values are very close to those when the population size is 40. Consequently, 0.8 and 50 are chosen as the crossover rate and population size, respectively. Then, the mixed load corresponding to the eight group data can be identified by FWF-GA, and the identification results are shown in Table 4.

From Tables 2 and 4, the design variable values of the other seven groups’ data are the same as the values of the actual mixed operating states, except for the 6th group data. For the 6th group data, the actual mixed operating states is NC, FL, heating of WF, 2nd cold air of HD, and TV, but the identification result is NC, FL, heating of WF, and TV. That is, the running of 2nd cold air of HD is not identified for the 6th group data. This demonstrates that there is one false identification result of eight group data for FWF-GA.

4.3. Different Fitness Function Evaluation. In order to better illustrate the effectiveness of FWF-GA, the algorithms by different fitness functions are used to perform the identification for the eight group data. The algorithms are called Re-GA and Co-GA, and the fitness function is defined by

![Figure 5: Current harmonic characteristics of each appliance.](image1)

![Figure 6: Current and voltage waveforms of multiple appliances operated simultaneously.](image2)
The results of different fitness functions are listed in Tables 5 and 6. In Table 6, the power is the vector of active power and reactive power, and the relative error is the vector of the relative error of active power and reactive power, which can be calculated by

\[
\epsilon_1 = \frac{P_I - P_C}{P_C}, \quad \epsilon_2 = \frac{Q_I - Q_C}{Q_C},
\]

where \(P_I\) and \(Q_I\) are the active power and reactive power by identification, respectively, and \(P_C\) and \(Q_C\) are the active power and reactive power by actual measurement.

As shown in Table 5, the identification of two group data (the 6th and 7th groups) are false for Re-GA, and the identification of three group data (the 2nd, 3rd, and 8th groups) are false for Co-GA, while only the 6th group is false for FWF-GA which is shown in Table 4. These data indicate that the proposed FWF-GA has better accuracy than Re-GA and Co-GA.

From Table 6, comparing the results by Co-GA with that by FWF-GA, the relative error of active power by Co-GA is relatively smaller than that of reactive power in eight group data, which is caused by the numerical gap between active power and reactive power. This is because the fitness function of Co-GA is the absolute error of active and reactive power. The numerical value of active power is relatively larger than reactive power, and the absolute error of active power has a great impact on the optimization results. Consequently, the influence of active power is mainly considered and the influence of reactive power is ignored in order to ensure the minimum target value in the optimization process, which maybe make the identification results incorrect. Additionally, comparing the results by Re-GA with that by FWF-GA, the results of the 6th group are both false by FWF-GA and RE-GA, but the results of the 7th group are also false by RE-GA. FWF-GA and RE-GA both use the relative error of active and reactive power, while FWF-GA also uses the frequency weighting factor of harmonic feature as the adjustment. For further description of the influence of the harmonic feature, the frequency-weighting factor corresponding to the 7th group data is given in Table 7.

From Tables 5 and 6, the mistake of the results of the 7th group data is that the design variable \(x_1\) corresponding to NC is 0, which means that NC is wrongly judged as not running. This is because RE-GA adopts the relative error of
active and reactive power as the optimization objective. For the 7th group data, the relative error of reactive power has a greater impact on the objective value than that of active power. From Table 1, the active power of NC is small, but the reactive power is relatively large. Consequently, RE-GA gives false results. However, for FWF-GA, not only the relative errors of active and reactive power are adopted as the optimization objective but also the magnitude of harmonic feature are used to adjust the influence of active and reactive power. From Table 7, the frequency weighting factor $\xi_i$ corresponding to NC is larger than other factors. Through the adjustment of the harmonic feature, the influence on the optimization result of the open/closed state of NC is changed. Consequently, FWF-GA gives true results.

In order to clearly describe the identification efficiency of different fitness functions, taking the results of the 4th group to draw the iterative history as shown in Figure 7, and the convergence times of 8 group data are shown in Table 8.

From Figure 7, the three algorithms of FWF-GA, RE-GA, and Co-GA tend to be stable after 15 iterations, 10 iterations, and 20 iterations, respectively. As shown in Table 8, the number of convergence times of FWF-GA (which is range from 10 to 16) is less than that of Co-GA (which is range from 12 to 20) while larger than that of RE-GA (which is range from 7 to 15). These show that the FWF-GA has
better computational efficiency than Co-GA. However, due to increasing the adjustment of harmonic features for FWF-GA, the computational efficiency of FWF-GA is not as good as that of RE-GA.

### 4.4. Comparison of Different Methods

In order to verify the effectiveness of the proposed method, the existing method such as DAPSO, CNN, and QP are also used to perform the identification of the data set of mixed load. Table 9 gives the identification accuracy and test time.

<table>
<thead>
<tr>
<th>No.</th>
<th>Convergence times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FWF-GA</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

As shown in Table 9, the minimum of SIP for five appliances (NC is 94.3%, FL is 95.6%, WF is 92.1%, HD is 91.3%, and TV is 94.8%) is obtained by DAPSO and CNN, respectively, while the maximum (NC is 96.7%, FL is 95.0, HD is 95.7%, and TV is 96.4%) is obtained by the proposed method. Furthermore, the minimum of AIP is 93.7% by DAPSO, and the maximum is 96.9 by the proposed method. It can be also derived from Table 9 that DAPSO takes the least time (211.82 s in total), and QP takes the longest time (675.67 s in total). The proposed method takes 260.73 s, which is larger than DAPSO and lower than CNN and QP. These demonstrate that the proposed method has a better accuracy and calculation efficiency.

### 4.5. Influence Analysis of Noise

In load identification, noise inevitably exists. For the purpose of test, the influence of the noise on the identification result of FWF-GA, Gaussian white noise with various signal-to-noise ratios (SNR) is added to the test data. The SNR is designed to be 50 dB, 40 dB, 30 dB, and 20 dB respectively. The identification accuracy with various SNRs is shown in Table 10.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>SIP (%)</th>
<th>AIP (%)</th>
<th>Cost time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NC</td>
<td>FL</td>
<td>WF</td>
</tr>
<tr>
<td>20</td>
<td>94.7</td>
<td>94.6</td>
<td>92.8</td>
</tr>
<tr>
<td>30</td>
<td>95.2</td>
<td>95.3</td>
<td>93.2</td>
</tr>
<tr>
<td>40</td>
<td>95.6</td>
<td>96.1</td>
<td>93.9</td>
</tr>
<tr>
<td>50</td>
<td>96.3</td>
<td>97.5</td>
<td>94.8</td>
</tr>
</tbody>
</table>

As shown in Table 10, the SIP and AIP with different SNR are reduced compared with those without noise, but all the values of identification accuracy are still larger than 90%, which demonstrates that the proposed identification method still maintains a high accuracy under various SNR.

### 5. Conclusions

In this paper, a simultaneous load identification method is proposed to achieve the identification of the total load under multiple appliances operated simultaneously. Three parts including hybrid features extraction, simultaneous identification optimization model, and FWF-GA are mainly involved in the proposed method. The simultaneous identification optimization model is constructed by using the hybrid feature model. FWF-GA is developed based on the GA framework to efficiently solve the identification optimization model. In FWF-GA, the fitness function is constructed on the basis of hybrid features, and the relative error of active power, reactive power, and the frequency weighting factor are used to evaluate the fitness of individuals. The practicability of the developed fitness function is examined by comparing with Re-GA and Co-GA, and the accuracy and efficiency of the proposed FWF-GA are compared with the existing DAPSO, CNN, and QP. The test results all demonstrated that FWF-GA has better accuracy and efficiency than the other methods. In addition, the impact of noise on the identification results is analyzed by adding Gaussian white noise with various SNRs to the test data. The result also demonstrates that FWF-GA still has a high accuracy under various SNRs.

For the shortcomings of the method proposed in this paper, FWF-GA is proposed to identify the stable operation state of all electrical appliances, while instantaneous state on or off with on or off is not considered. In addition, the uncertainty of the appliance operation has an important impact on the identification result. Another important problem is that, in order to get better accuracy, the adjustment of system parameters needs a large calculation cost. Consequently, it would be beneficial to develop an adaptive identification method considering the instantaneous state and the operating uncertainty in future work.

### Data Availability

The labeled data set used to support the findings of this study is 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 supported by the Project of State Grid Corporation of China (Grant no. 5600-202024168A-0-0-0-0).

References


