Research Article
Noninvasive Load Identification Method Based on Feature Similarity

Hongyan Li,1 Xianfeng Ding,1 Dan Qu,2 and Jiang Lin1

1School of Science, Southwest Petroleum University, Sichuan, Chengdu 610500, China
2School of Mathematics and Statistics, Sichuan University of Science & Engineering, Sichuan, Zigong 643000, China

Correspondence should be addressed to Xianfeng Ding; fxxd@163.com

Received 24 June 2019; Accepted 8 January 2020; Published 24 February 2020

Copyright © 2020 Hongyan Li 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.

The traditional power load identification is greatly restricted in application because of its high cost and low efficiency. In this paper, the similarity model is established to realize the noninvasive load identification of power by determining the feature database for the equipment. Firstly, the wavelet decomposition method and the wavelet threshold processing method are used to remove abnormal points and reduce noise of the original data, respectively. Secondly, the transient and steady-state characteristics of electrical equipment (active power and reactive power, harmonic current, and voltage-current trajectory) are extracted, and the feature database for the equipment is established. Thirdly, the feature similarity is defined to describe the similarity degree of any two devices under a certain feature, and the similarity model of automatic recognition of a single device is established. Finally, the device identification and calculation of power consumption are carried out for the part of data in annex 2 of question A in the 6th “teddy cup” data mining challenge competition.

1. Introduction

With the emergence of various new types of power load components in an endless stream, users put forward higher requirements on the reliability, safety, economy, and stability of power system. Smart grid emphasizes bidirectional interaction with users and encourages users to participate in power management through demand response, which is inseparable from detailed control of load operation information. Since the traditional invasive load monitoring system costs a lot in time and investment and has a certain impact on the reliability of the system, it is necessary to develop an economical and effective noninvasive load monitoring and identification system. Hence, strengthening the monitoring of building power consumption is of great practical significance for energy conservation and smart grid.

Noninvasive load monitoring technology has attracted much attention from power companies and scientific research institutions since it was proposed. It is worth noting that Hart [1] established the first noninvasive appliance load monitoring system (NIALM) to develop a monitoring tool that does not affect the target or affect the target as little as possible. It can provide power companies with specific power consumption data of different electrical equipment. Li and Yu [2] further carried out research on noninvasive load monitoring and determined characteristic parameters based on fuzzy clustering results of steady-state load characteristics of electrical appliances, so as to realize noninvasive load monitoring based on differential evolution algorithm. Liang et al. [3, 4] researched on a series of studies in the field of load characteristics and comprehensively introduced the basic concept, system structure, feature method, decomposition framework, system simulation application, and other aspects of noninvasive load monitoring. Cai et al. [5] calculated the similarity between the transient waveform and the fixed characteristic template in the electrical load characteristic database, established the electrical load characteristic membership matrix based on similarity, and determined the characteristic type of electrical load. Zheng et al. [6] studied the microcharacteristics of noninvasive load monitoring, established the household load characteristics database, and
analyzed the load characteristics and extraction methods contained in the fundamental wave and multiple harmonics of current, voltage, and power but lacked of the specific methods to complete the noninvasive identification of electrical load of users. Huang et al. [7] employed instantaneous current and power waveforms to take the decomposed current waveforms as the characteristic values of two similar loads, which could realize the accurate identification of electrical appliances with similar current waveforms. Wu et al. [8] decomposed the sampling current to obtain the independent current generated by the start-up of electrical appliances and established the load identification algorithm of entropy value discrimination to realize the decomposition and recognition of electrical loads. In practice, the research on nonintrusive power load monitoring and decomposition mainly focuses on the optimization and improvement of electrical load feature extraction and load identification algorithm.

Noninvasive power load decomposition and monitoring refers to installing a sensor at the entrance to the grid users, and the device monitors the power consumption and working condition of each or each type of electrical equipment by collecting and analyzing the total power or total current. Hence, power companies can understand the power consumption rules and usage patterns of each or every type of electrical equipment in the user’s home, as shown in Figure 1. The monitoring data of household power load provides a scientific basis for the prediction of load usage in power system and ensures the correctness of decision-making [9]. This paper takes the title A in the 6th “teddy cup” data mining challenge competition as the research background. Firstly, the transient and steady-state characteristics of the electrical equipment are extracted from the original data, the equipment feature database is established, and finally, the similarity model is established to realize the noninvasive load detection of power. The data are available at the teddy cup data mining challenge website.

The data used to support the findings of this study are available at the teddy cup data mining challenge website (http://www.5iai.com/bdrace/tzjingsai/20170921/1253.html#sHref).

2. Data Processing and Establishment of Feature Database

2.1. Data Preprocessing. Table 1 shows the known equipment data and parameters.

2.1.1. Abnormal Points Processing. In this paper, the wavelet decomposition $W_k$ value method is adopted to detect and distinguish abnormal points and mutation points [10]. The specific algorithm is as follows:

**Step 1.** The fitting residuals $e_t$ and $t = 1, 2, \ldots$ were decomposed online based on two wavelet scale.

**Step 2.** The modulus of wavelet decomposition coefficient at two scales was calculated, and the difference value was calculated to obtain $E_{k}$.

**Step 3.** Detection of abnormal points and mutation points.

The active power data of YD1-YD11 were tested by the above outlier test method. Figure 2 shows the abnormal point test results of equipment YD4 in the period from 60 seconds to 290 seconds.


Wavelet noise reduction is to separate signal from noise by using the difference of noise in the time and frequency domain, so as to obtain more ideal noise reduction effect. Let signal $S(t)$ is the polluted noise of $X(t)$, and its basic model can be expressed as

$$S(t) = X(t) + \sigma e(t),$$

where $e(t)$ is noise and $\sigma$ is noise intensity.

After wavelet noise reduction, the processed data is obtained and then the waveform is drawn by MATLAB. Based on length, a sampling period of YD1’s cycle data is taken as an example here to give the signal after noise reduction, which is shown in Figure 3.

2.2. Establishment of Feature Database

2.2.1. Transient Feature Extraction. Transient characteristics refer to the characteristics shown when the working state of electrical appliances changes. As shown in Figure 4, the transient power waveform of electrical appliances’ start-up is a typical load mark.

The following part is the analysis of the implementation methods and load characteristics of transient characteristics, which contains four noninvasive load monitoring: mean current and root-mean-square, transition time of transient and multiple of impulse power (current) [12].

![Diagram](http://www.5iai.com/bdrace/tzjingsai/20170921/1253.html#sHref)
To calculate the mean value of signal $i(t)$, it is necessary to integrate the signal waveform in a period of time:

$$\langle i(t) \rangle = \frac{1}{T} \int_0^T x(t)dt,$$ (2)

where $T$ is the integral time.

(2) Root-Mean-Square. Root-mean-square represents the fluctuation based on mean value of signal. The root-mean-square of signal $i(t)$ is used to represent the voltage of alternating current’s waveform, which is defined as

$$i_{rms} = \sqrt{\frac{1}{T} \int_0^T i(t)dt}.$$ (3)

(3) Transition Time. Set the start time of the transient process as $t_{ton}$ and the end time of the transient process as $t_{toff}$; then the transition time $\Delta t$ can be calculated by the following equation:

$$\Delta t = t_{toff} - t_{ton}.$$ (4)

(4) Multiple of Impulse Power (Current). The formula for calculating the multiple $K_p$ of impulse power (current) is as follows:

Table 1: Electrical equipment and working parameters.

<table>
<thead>
<tr>
<th>Device ID</th>
<th>Device type</th>
<th>Working parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>YD1</td>
<td>Oxon fan</td>
<td>220 V, 60 W</td>
</tr>
<tr>
<td>YD2</td>
<td>Midea microwave</td>
<td>220 V, input: 1150 W, output: 700 W</td>
</tr>
<tr>
<td>YD3</td>
<td>Joyang hot pot</td>
<td>220 V, 1800 W</td>
</tr>
<tr>
<td>YD4</td>
<td>ThinkPad laptop</td>
<td>20 V, 3.25 A/4.5 A</td>
</tr>
<tr>
<td>YD5</td>
<td>Energy saving lamp</td>
<td>220 V, 40 W</td>
</tr>
<tr>
<td>YD6</td>
<td>FUII laser printer</td>
<td>220 V, 5 W</td>
</tr>
<tr>
<td>YD7</td>
<td>Water dispenser</td>
<td>220<del>240 V, 50</del>60 Hz, 4.6 A</td>
</tr>
<tr>
<td>YD8</td>
<td>Wall-hanging air conditioner</td>
<td>220 V, heating: 430 W, refrigeration: 70 W, The total: 500 W</td>
</tr>
<tr>
<td>YD9</td>
<td>Pentium hair dryer</td>
<td>220 V, 50 Hz, 1400 W</td>
</tr>
<tr>
<td>YD10</td>
<td>Skyworth television</td>
<td>220 V, 50 Hz, 150 W</td>
</tr>
</tbody>
</table>

Figure 2: Abnormal points detection map.

Figure 3: Wavelet noise reduction map (the upper and lower are the original signal and the signal is processed using a layered threshold).
where $P_{\text{peak}}$ is the maximum power in the process of transient switching, $P_{s1}$ is the steady-state average power before the input of electrical appliances, and $P_{s2}$ is the steady-state average power after the input of electrical appliances. Applying the above introduction and the single-state data provided in Annex 1, the obtained characteristics database of transient state is as follows.

As can be seen from Table 2, the change form of electricity load from the opening state to the stable state is various. The pure resistive load enters into the steady state directly from the start, while other loads contain pulse current and the starting time and pulse size are different. And the switching transient state of different load is different, so the transient characteristic can be used to distinguish the electrical equipment.

2.2.2. Steady-State Feature Extraction. The steady-state characteristics refer to the characteristics of the electrical appliances in a stable operation state. In other words, the steady-state characteristics are the results of some characteristics analysis differences between the two stable operation states [13]. This paper will use V-I trajectory, power characteristic, and harmonic matrix.

(1) V-I Trajectory. The shape features adopted by V-I trajectory method mainly include the current span, trajectory area, absolute area, standard deviation of instantaneous resistance, curvature, slope, total area, left and right areas, asymmetry, intersection point, etc. [14]. In order to avoid the influence of voltage and current amplitude differences of different loads on the size of V-I trajectory, it is necessary to normalize the two parameters before comparing the shape features. Using the frequency data provided in Annex 1, take the normalized voltage as the abscissa and the normalized current as the ordinate to draw the V-I trajectory curve of some equipment, which is shown in Figures 5 and 6.

As can be seen from the above figure, for resistive loads, such as Joyang hot pot, V-I trajectory is a straight line, while for a load with high harmonic content, such as Midea microwave, V-I trajectory contains at least one intersection point. The two kinds of trajectories differ significantly, so the V-I trajectory can be used as a distinguishing feature of electrical equipment.

$$K_p = \frac{P_{\text{peak}} - P_{s1}}{P_{s2} - P_{s1}}$$  \hspace{1cm} (5)

where $P_{\text{peak}}$ is the maximum power in the process of transient switching, $P_{s1}$ is the steady-state average power before the input of electrical appliances, and $P_{s2}$ is the steady-state average power after the input of electrical appliances.
(3) The absolute area absarea of the normalized V-I trajectory curve, which is defined as

$$\text{absarea} = \sum_{m=1}^{v_{\min} - 1} \frac{1}{2} (V_{m+1}' - V_m') (I_{m+1}' - I_m'),$$  \hspace{1cm} \text{(9)}$$

(4) The standard deviation of instantaneous resistance $D[15]$, which is defined as

$$D = \sqrt{\frac{\sum_{m=1}^{NT} (R_n - \bar{R})^2}{NT}},$$  \hspace{1cm} \text{(11)}$$

where $R_n = (V_n'/I_n')$ is the instantaneous resistance of the $n$-th sampling point, $V_m'$ is the $n$-th sampling point and represents the normalized voltage value, $I_n'$ is the $n$-th sampling point and represents the normalized current value, $m \in [1, NT + ip]$, NT are the number of sampling points in a period, and $ip$ is the number of preset interpolation points. $\bar{R}$ is the average value of $R_n$.

According to the size of the power, the working state of the equipment is divided into several gears; the greater the power, the higher the gear. From device 1 to device 11, there are at most five working states, so the working state of the device is divided into five levels. The device data of one-second period is randomly selected from each running state to draw the V-I trajectory. Based on the above steps and the single-state data provided in Annex 1, the V-I trajectory feature database is obtained, and the V-I trajectory feature of gear 1 of each device is obtained (the default line represents that the device does not have this gear).

As can be seen from Table 3, the V-I trajectory characteristics of gear 1 of each equipment, especially the difference between the current span and the standard deviation of instantaneous resistance are relatively large, and the differences of the obtained track are very obvious, so the V-I trajectory characteristics can be used to distinguish electrical equipment.

(2) Power Characteristics. Active power is the total power consumed by the load during operation. If the load is pure resistance, the voltage-current waveform will always be in phase, so there is no reactive component. However, due to the presence of inductive or capacitive elements, there is always a phase shift between the current and voltage waveforms, which produces or consumes reactive power. Active power and reactive power are calculated as follows [16]:

\begin{table}[h]  
\centering  
\caption{Transient eigenvalues.}  
\begin{tabular}{|c|c|c|c|c|}  
\hline  
 & Mean current & Root-mean-square value & Transition time & Multiple of impulse power \\
\hline  
YD1 & 110.021 & 71.860 & 4 & 1.003 \\
YD2 & 3229.075 & 2635.514 & 3 & 3.376 \\
YD3 & 4362.044 & 3874.353 & 1 & 1.014 \\
YD4 & 213.690 & 74.269 & 4 & 1.612 \\
YD5 & 98.903 & 87.131 & 1 & 1.005 \\
YD6 & 25.685 & 16.077 & 2 & 1.030 \\
YD7 & 435.202 & 1079.552 & 6 & 7.127 \\
YD8 & 1242.107 & 863.184 & 2 & 1.005 \\
YD9 & 100.004 & 54.712 & 7 & 1.551 \\
YD10 & 2153.324 & 2129.396 & 3 & 1.037 \\
YD11 & 289.769 & 239.148 & 6 & 1.011 \\
\hline  
\end{tabular}  
\end{table}
In this paper, the current variance of harmonic content rate of each device under different working conditions is calculated to describe the variation trend of harmonic content rate of each device under different working conditions. The default value indicates that the gear does not exist in the device. For example, device 1 cannot be switched 4th to 5th gear. The result is shown as Table 5.

As can be seen from Table 5, under the closed state, the variance of harmonic content rate of YD1, YD2, YD3, YD5, and YD6 is greater than other equipment. For one device, such as YD4, the variance of harmonic content rate is firstly small under the closed state, and then the harmonic content rate increases rapidly when switching to the first gear. In addition, the higher the gear shift is, the lower the variance harmonic content rate is, and the harmonic content rate is almost constant. Therefore, the variance of harmonic content rate can be used as the identification basis.


3.1. Similarity and Weight Coefficient. To automatically identify an unknown single device, the characteristic similarity of load mark can be analyzed [14]. Domain feature similarity $S$ is defined as

\[
S = \frac{1}{\|\mathbf{Y}_x - \mathbf{Y}_i\|^2}
\]

where $\mathbf{Y}_x$ represents the eigenvector of the unknown device $x$, $\mathbf{Y}_i$ is the eigenvector of device $i$. The larger the value of $S$, the higher the similarity between the unknown device $x$ and the known device $i$.

The similarity of load mark extracted in this paper is divided into four types of calculation, where $(1/\|\mathbf{Z}_1 - \mathbf{Z}_i\|^2)$ represents the similarity of transient characteristic of device YD1, and device YD5. Similarly, $(1/\|\mathbf{V}_1 - \mathbf{V}_i\|^2)$ and $(1/\|\mathbf{X}_1 - \mathbf{X}_i\|^2)$ represent the similarity of V-I trajectory characteristic and harmonic characteristic of device YD1 and device YD9.

$H_{\mathbf{X}}$ represents the contrast similarity of the active power reactive power, defined as the image similarity between the active power and the reactive power contrast figure of the two devices. The specific similarity calculation employs the histogram method [18]. Firstly, calculate the histogram of

\[
P = \sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} U_k I_k \cos(\phi_k),
\]

\[
Q = \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} U_k I_k \sin(\phi_k),
\]
the two images, respectively, and then calculate the distance measure of the two images, the Pap distance is chosen as a measure, which is defined as

\[
d(H_1, H_2) = \sqrt{1 - \frac{1}{H_1 H_2 N^2} \sum_{i=1}^{N^2} \sqrt{H_1(I)H_2(I)}}. \quad (14)
\]

Finally, we calculate the contrast similarity between the active power and the reactive power of the device YD_i and the device YD_x. The total similarity is calculated by weight, and the weight is determined by entropy method [19].

\[
w_i = \frac{1 - H_i}{m - \sum_{i=1}^{m} H_i}, \quad i = 1. \quad (15)
\]

Through the entropy value method, the weight of each feature similarity is \( w = (0.151, 0.342, 0.375, 0.132) \).

3.2 Establishment of the Similarity Model. Load identification model based on similarity is a weighted sum of all kinds
of feature similarities to obtain a total similarity of load feature similarity. The specific model is as follows:

\[
S_{tx} = \frac{w_1}{\|Z_i - Z_x\|^2} + \frac{w_2}{\|V_i - V_x\|^2} + \frac{w_3}{\|X_i - X_x\|^2} + w_4 H_{tx}^2, \tag{16}
\]

where \(Z_i\) is the eigenvector of transient of YD_i, \(Z_x\) is the eigenvector of transient of YD_x, \(V_i\) is the eigenvector of V-I trajectory of YD_i, \(V_x\) is the eigenvector of V-I trajectory of YD_x, \(X_i\) is the eigenvector of harmonic of YD_i, \(X_x\) is the eigenvector of harmonic of YD_x, and \(H_{tx}\) is the comparison similarity of active power and reactive power between the device to be tested and the known device.

It can be seen from the characteristic data of Table 6 that when the devices \(X_1\) and \(X_2\) to be tested are in the first gear position, the V-I trajectory curve caused by the standard deviation characteristic of instantaneous resistance is relatively large.

As can be seen from Table 7, the unknown equipment \(X_1\) produces few harmonics, and the unknown equipment \(X_2\) produces abundant harmonics. It can be seen that the third and fifth harmonics content rate of \(X_2\) nearly 50%, but the harmonic content of \(X_1\) is less than 1%.

As can be seen from Table 8, in the closed state, the variance of harmonic content rate under each operating state has little difference. For equipment \(X_1\), firstly, the variance of harmonic content rate is small in the closed state, then gradually decreases with the increase of gear switch, and finally remains almost constant.

Through the established model and relevant data, the calculation results of the similarity between the unknown device \(X_1, X_2, \) and YD1 to YD11 are as follows.

### Table 6: The V-I trajectory characteristics of gear 1 of the device.

<table>
<thead>
<tr>
<th></th>
<th>The current span</th>
<th>Area</th>
<th>Absolute area</th>
<th>Standard deviation of instantaneous resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>1.642</td>
<td>0.022</td>
<td>0.054</td>
<td>64.249</td>
</tr>
<tr>
<td>X2</td>
<td>1.277</td>
<td>0.049</td>
<td>0.079</td>
<td>13.390</td>
</tr>
</tbody>
</table>

### Table 7: Amplitude of kth harmonic content rate of device.

<table>
<thead>
<tr>
<th>Amplitude</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>2-norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.98</td>
<td>0.88</td>
<td>0.95</td>
<td>0.98</td>
<td>0.88</td>
<td>0.99</td>
<td>0.94</td>
<td>3.01947</td>
</tr>
<tr>
<td>X2</td>
<td>15.7</td>
<td>54.12</td>
<td>21.21</td>
<td>49.32</td>
<td>15.3</td>
<td>34.93</td>
<td>17.81</td>
<td>29.42</td>
<td>15.72</td>
<td>29.94</td>
<td>99.16741</td>
</tr>
</tbody>
</table>

### Table 8: Variance of harmonic content rate under each operating state.

<table>
<thead>
<tr>
<th>Variance</th>
<th>Closed</th>
<th>Gear 1</th>
<th>Gear 2</th>
<th>Gear 3</th>
<th>Gear 4</th>
<th>Gear 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>0.047832</td>
<td>0.02007</td>
<td>7.51E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>0.047832</td>
<td>0.080229</td>
<td>0.046945</td>
<td>0.037575</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 9: Similarity results of device 1.

<table>
<thead>
<tr>
<th>Similarity</th>
<th>Similarity of transient characteristic</th>
<th>Similarity of V-I trajectory</th>
<th>Comparison similarity of active power and reactive power</th>
<th>Similarity of harmonic</th>
<th>The total similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device 1—YD1</td>
<td>0.321</td>
<td>0.302</td>
<td>0.945</td>
<td>1.000708</td>
<td>0.63868</td>
</tr>
<tr>
<td>Device 1—YD2</td>
<td>1.228</td>
<td>1.134</td>
<td>0.675</td>
<td>1.000417</td>
<td>0.98983</td>
</tr>
<tr>
<td>Device 1—YD3</td>
<td>1.147</td>
<td>0.920</td>
<td>0.961</td>
<td>1.000407</td>
<td>0.99420</td>
</tr>
<tr>
<td>Device 1—YD4</td>
<td>0.434</td>
<td>0.973</td>
<td>0.782</td>
<td>1.000839</td>
<td>0.81413</td>
</tr>
<tr>
<td>Device 1—YD5</td>
<td>0.321</td>
<td>0.305</td>
<td>0.956</td>
<td>1.0011</td>
<td>0.64316</td>
</tr>
<tr>
<td>Device 1—YD6</td>
<td>0.148</td>
<td>1.265</td>
<td>0.943</td>
<td>1.000979</td>
<td>0.89269</td>
</tr>
<tr>
<td>Device 1—YD7</td>
<td>1.320</td>
<td>1.176</td>
<td>0.910</td>
<td>1.001123</td>
<td>1.09019</td>
</tr>
<tr>
<td>Device 1—YD8</td>
<td>3.001</td>
<td>0.664</td>
<td>0.898</td>
<td>1.002281</td>
<td>1.26949</td>
</tr>
<tr>
<td>Device 1—YD9</td>
<td>0.297</td>
<td>0.346</td>
<td>0.928</td>
<td>1.010204</td>
<td>0.64403</td>
</tr>
<tr>
<td>Device 1—YD10</td>
<td>1.361</td>
<td>0.786</td>
<td>0.926</td>
<td>1.006951</td>
<td>0.98783</td>
</tr>
<tr>
<td>Device 1—YD11</td>
<td>0.604</td>
<td>0.515</td>
<td>0.955</td>
<td>1.034707</td>
<td>0.76915</td>
</tr>
</tbody>
</table>

### 3.3. Application of Model

#### 3.3.1. Feature Extraction and Recognition of Unknown Devices.

By the method of V-I trajectory and harmonic matrix, the feature matching data of unknown device \(X_1\) and \(X_2\) are extracted as follows.
value of the original data. Secondly, the data is transformed by wavelet noise reduction, and pretreatment of the sampled data points of each device is completed. Finally, the abnormal point detection results of a certain device are obtained, and the waveform diagram after noise reduction is drawn.

In the process of feature extraction, firstly, the transient characteristic of a single device are extracted by analyzing the preprocessed data, which includes active power, reactive power, harmonic current, and voltage-current trajectory (V-I trajectory). Secondly, the computation and extraction methods of the characteristic values of each load characteristic are given. Finally, the transient characteristic values of the equipment are obtained, containing the V-I trajectory characteristics of gears 1, 2, 3, 4 and 5, the comparison diagram of active power and reactive power of each equipment, the amplitude of \( \frac{k}{3} \) \( k = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 \) harmonic content rate of the equipment, and the variance of harmonic content rate of each operating state of the equipment.

In the automatic identification of a single device, this paper identifies any single device by establishing a similarity model. Based on the load characteristics of four types extracted, a similarity-based load identification model is established. Firstly, the feature similarity is defined to denote the similarity degree of any two devices, and the weight coefficient of similarity of each feature is determined by the entropy value method. Secondly, the weighted sum of feature similarity is used to determine the total feature similarity, and the device with the highest similarity is selected to match with the unknown device. Finally, the similarity feature data between the unknown device and devices 1-11 are obtained. According to the calculation results, the unknown device \( X_1 \) is determined as device 8, and the unknown device \( X_2 \) is determined as device 9.

### 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 they have no conflicts of interest.
Acknowledgments

Hongyan Li expresses sincere thanks to all those who have helped in the course of writing this paper, would like to show sincere gratitude to Mr Xianfeng Ding, who has given so much useful advices on writing and has tried his best to improve the paper, and also would like to express gratitude to classmates who offered references and information on time. Without their help, it would be much harder to finish this paper.

References