Research Article

Noninvasive Load Identification Method Based on Feature Similarity

Hongyan Li [ID],[1] **Xianfeng Ding** [ID],[1] **Dan Qu** [ID],[2] **and Jiang Lin**[1]

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

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

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 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.
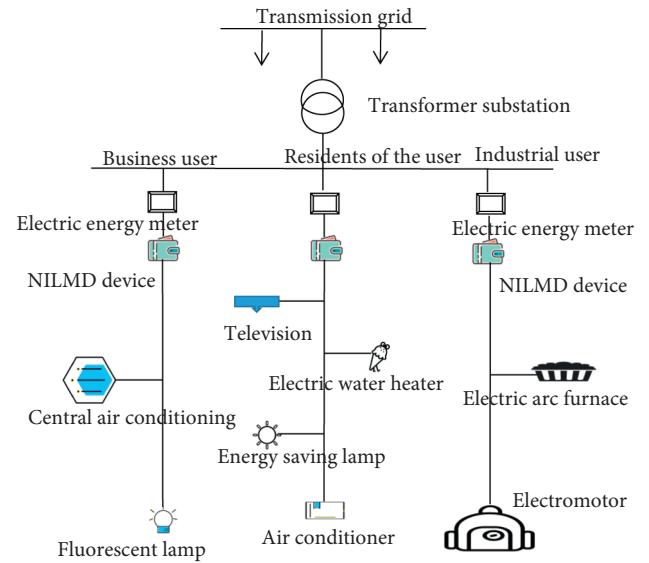

FIGURE 1: Schematic diagram of noninvasive power load monitoring and decomposition system.

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.

2.1.2. Noise Reduction Processing.
We perform data noise reduction through wavelet threshold process [11].

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 $S(n)$ is the polluted noise of $X(t)$, and its basic model can be expressed as

$$S(t) = X(t) + \sigma e(t), \tag{1}$$

where $e(t)$ is noise and σ 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].

TABLE 1: Electrical equipment and working parameters.

| Device ID	Device type	Working parameter
YD1	Oxon fan	220 V, 60 W
YD2	Midea microwave	220 V, input: 1150 W, output: 700 W
YD3	Joyang hot pot	220 V, 1800 W
YD4	ThinkPad laptop	20 V, 3.25 A/4.5 A
YD5	Incandescent lamp	220 V, 40 W
YD6	Energy saving lamp	220 V, 5 W
YD7	FUJI laser printer	220~240 V, 50~60 Hz, 4.6 A
YD8	Water dispenser	220 V, heating: 430 W, refrigeration: 70 W, The total: 500 W
YD9	Wall-hanging air conditioner	220 V, 2600 W
YD10	Pentium hair dryer	220 V, 50 Hz, 1400 W
YD11	Skyworth television	220 V, 50 Hz, 150 W
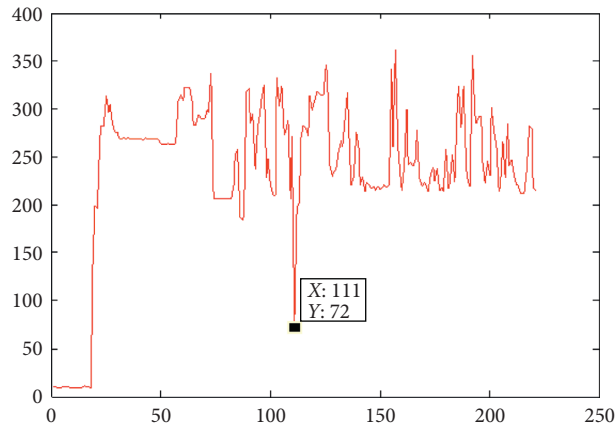
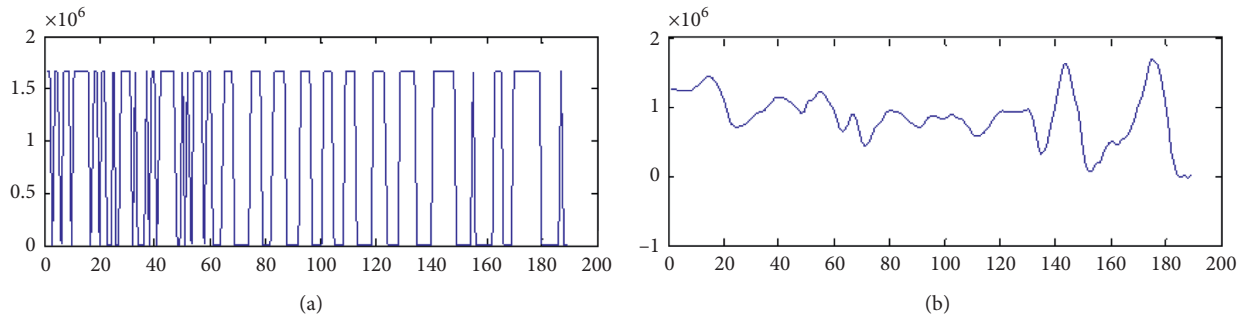FIGURE 2: Abnormal points detection map.

(a)

(b)

FIGURE 3: Wavelet noise reduction map (the upper and lower are the original signal and the signal is processed using a layered threshold).

(1) Mean and Root-Mean-Square. To calculate the mean value of signal $i(t)$, it is necessary to integrate the signal waveform in a period of time:

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

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

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

(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 Δt can be calculated by the following equation:

$$\Delta t = t_{\mathrm{toff}} - t_{\mathrm{ton}}. \tag{4}$$

(4) Multiple of Impulse Power (Current). The formula for calculating the multiple K_P of impulse power (current) is as follows:
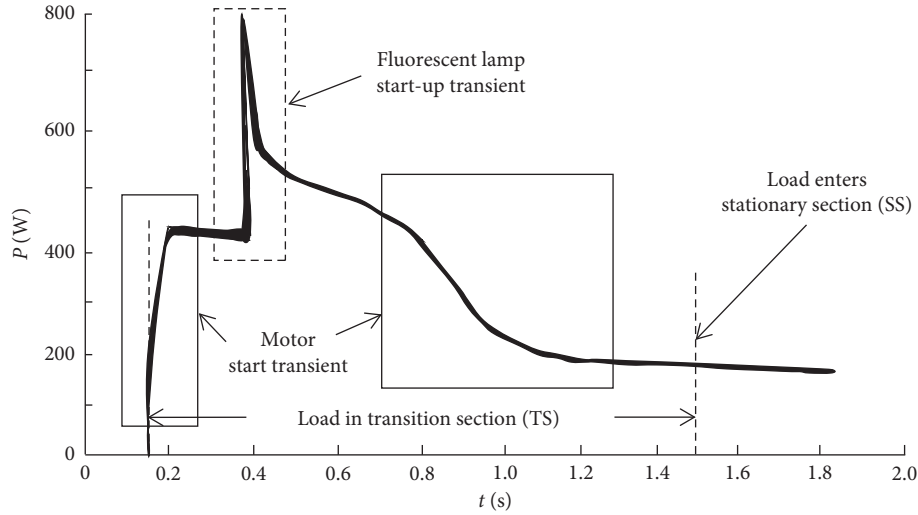
FIGURE 4: Electrical start-up transient power waveform.

$$K_p = \frac{P_{\text{peak}} - P_{s1}}{P_{s2} - P_{s1}}, \qquad (5)$$

where P_{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.

(1) Current span itc, which is defined as

$$\text{itc} = \max(I) - \min(I), \qquad (6)$$

where I is the current sequence and $\max(I)$ and $\min(I)$ represent the maximum and minimum values of the current sequence.

(2) The trajectory area of the normalized V-I trajectory curve

The normalized sequence value V'_m is obtained from voltage sequence V_m, which is defined as

$$V'_m = \frac{V_m}{\max(V)}, \qquad (7)$$

where $\max(V)$ is the maximum value of the voltage sequence and $m \in [1, \text{NT} + \text{ip}]$, NT are the number of sampling points in a period, and ip is the number of preset interpolation points.

The normalized value I'_m is obtained from current sequence I_m, which is defined as

$$I'_m = \frac{I_m}{\max(I)}, \qquad (8)$$

where $\max(I)$ is the maximum value of the current sequence and I is the current sequence, the maximum voltage point is $v\max(V'_{v\max}, I'_{v\max})$, the minimum voltage point is $v\min(V'_{v\min}, I'_{v\min})$, and the trajectory area is area, which is defined as

TABLE 2: Transient eigenvalues.

	Mean current	Root-mean-square value	Transition time	Multiple of impulse power
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	2153.312	2129.346	3	1.037
YD11	289.769	239.148	6	1.011

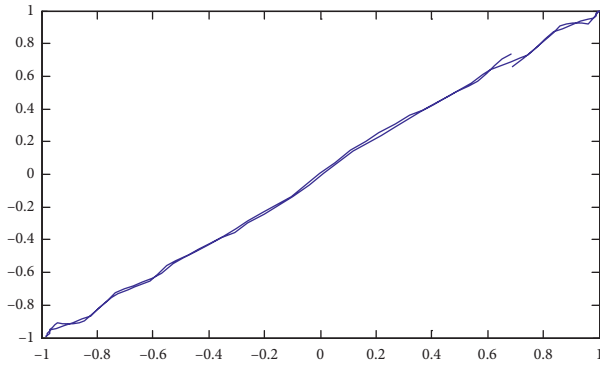FIGURE 5: YD3 (Joyang hot pot).
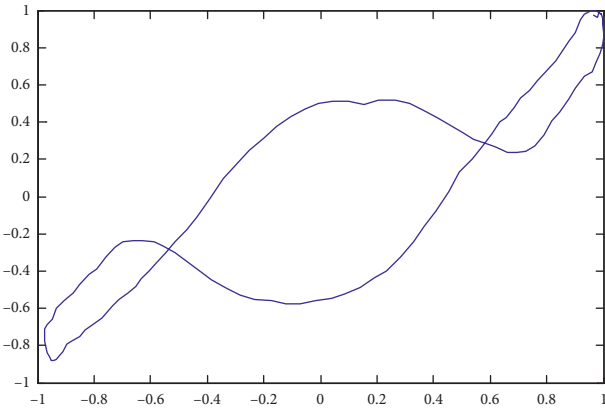
FIGURE 6: YD2 (Midea microwave).

$$\text{area} = \sum_{m=1}^{NT+ip-1} \frac{1}{2} \left(V'_{m+1} - V'_m \right) \left(I'_{m+1} - I'_m \right), \qquad (9)$$

(3) The absolute area absarea of the normalized V-I trajectory curve, which is defined as

$$\text{absarea} = \sum_{m=1}^{v\min-1} \frac{1}{2} \left| V'_{m+1} - V'_m \right| \left(\left| I'_{m1} - I'_m \right| + \left| I'_{m2} - I'_{m+1} \right| \right), \qquad (10)$$

where $m1$ satisfies $\min(|V'_{m1} - V'_m|)$, $m1 \in [v\min + 1, NT + ip]$, $m2$ satisfies $\min(|V'_{m2} - V'_m|)$, $m2 \in [v\min + 1, NT + ip]$.

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

$$D = \sqrt{\frac{\sum_{n=1}^{NT} \left(R_n - \overline{R} \right)^2}{NT}}, \qquad (11)$$

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. $\overline{R}$ 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]:

TABLE 3: The V-I trajectory characteristics of gear 1 of each device.

V-I trajectory characteristics	The current span	Area	Absolute area	Standard deviation of instantaneous resistance
YD1	3.253	0.0723	0.085	4.410
YD2	77.582	0.012	0.072	59.810
YD3	95.075	0.077	0.077	2.120
YD4	11.086	0.014	0.022	30.925
YD5	4.100	0.072	0.079	3.874
YD6	2.336	0.017	0.060	170.892
YD7	16.185	0.016	0.024	38.097
YD8	31.091	0.069	0.084	0.191
YD9	0.881	0.062	0.110	6.843
YD10	36.897	0.027	0.051	12.263
YD11	11.780	0.059	0.062	9.119

$$P = \sum_{k=0}^{\infty} P_k = \sum_{k=0}^{\infty} U_k I_k \cos(\varphi_k),$$

$$Q = \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} U_k I_k \sin(\varphi_k), \tag{12}$$

where U is the effective value of voltage when the power load is running, I is the effective value of current when the power load is running, ϕ is the power factor angle when the power load is running, and k is the number of harmonics.

We draw the images of active power and reactive power of each device on the same coordinate axis and obtain the comparison diagram of active power and reactive power of each device. The comparison diagram of YD1 device and YD9 device is shown in Figure 7.

As can be seen from Figure 7, the active power of YD1 equipment is greater than the reactive power, while the active power of YD9 equipment is not always greater than the reactive power, among which the active power is less than the reactive power during a sampling period, so YD9 equipment is obviously different from other equipment in the comparison of active power and reactive power.

(3) Harmonic Matrix. The harmonic data contains the unique characteristics of different electrical appliances. The harmonic of load voltage or current can be extracted by Fourier transform or wavelet transform and further identified the load. It should be noted that most loads produce even harmonic with small amplitude and odd harmonic with large amplitude. Low harmonic contains a large amount of information [17]. Therefore, this paper selects the 2nd to 11th harmonic data to study. Calculate the amplitude of each harmonic content rate of each device, and obtain the following harmonic feature database. The data in each row is the amplitude of the kth $(k = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)$ harmonic content rate of each device.

It can be seen from Table 4 that resistive loads, such as incandescent lamps and kettles, produce few harmonic. While nonresistive loads, such as induction cooker, electric fan, produce rich harmonic. It can be seen that the second and third harmonic contents of YD1, YD2, YD3, YD5, YD6, and YD8 are above 90%, but the harmonic contents of YD9, YD10, and YD11 are significantly lower than 90%, which can distinguish these loads.

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. Mathematical Model and Application of Automatic Identification of Single Device

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

$$S = \frac{1}{\left\| \|Y_i - Y_x\| / \|Y_i\| \right\|^2}. \tag{13}$$

where Y_x represents the eigenvector of the unknown device x. 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 $(1/\| \|Z_i - Z_x\| / \|Z_i\| \|^2)$ represents the similarity of transient characteristic of device YD_i and device YD_x. Similarly, $(1/\| \|V_i - V_x\| / \|V_i\| \|^2)$ and $(1/\| \|X_i - X_x\| / \|X_i\| \|^2)$ represent the similarity of V-I trajectory characteristic and harmonic characteristic of device YD_i and device YD_x.

H_{ix} 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
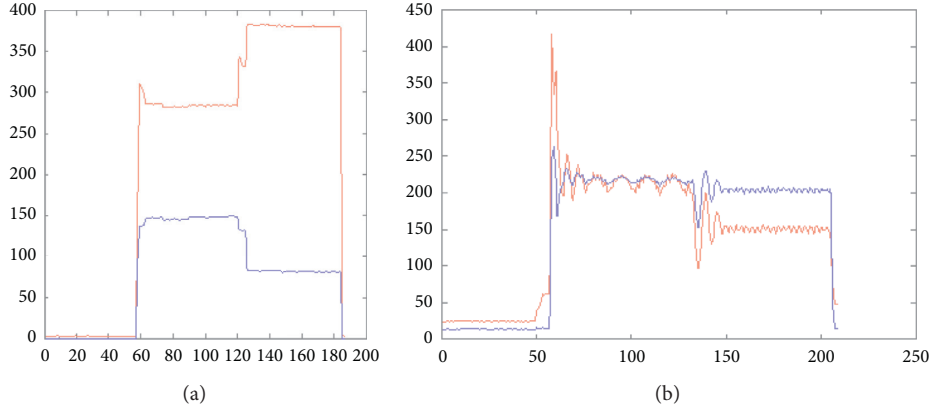
(a)

(b)

FIGURE 7: Comparison diagrams of active power and reactive power (left and right are device YD1 and device YD9, respectively).

TABLE 4: Device amplitude of kth harmonic content rate.

Amplitude	2	3	4	5	6	7	8	9	10	11	Two norms
YD1	99.68	99.71	96.9	98.28	94.51	89.62	98.51	98.19	95.49	96.21	305.958
YD2	99.2	99.44	95.8	96.9	99.32	98.6	98.04	97.98	97.89	99.33	310.714
YD3	99.9	94.5	96.48	99.01	93.5	97.43	97.54	93.74	97.31	96.84	305.622
YD4	21.06	96.19	19.6	88.75	18.47	82.88	17.43	76.13	15.01	62.38	188.104
YD5	97.15	99.85	97.84	99.45	95.64	99.68	99.54	96.9	93.93	94.94	308.364
YD6	97.83	92.98	97.89	94.48	96.08	96.67	84.73	94.65	98.6	97.28	301.039
YD7	79.93	94.24	36.77	83.92	19.74	78.01	19.49	73.88	13.26	58.86	199.043
YD8	99.51	98.09	98.93	92.82	98.65	97.63	92.07	98.63	99.54	98.93	308.366
YD9	17.75	53.85	34.85	62.77	25.88	31.83	32.73	37.85	42.27	46.98	128.647
YD10	66.59	38.27	38.37	21.88	14.4	20.07	8.67	22.51	10.96	13.5	96.665
YD11	6.14	31.41	6.05	10.67	4.64	15.31	4.53	17.42	4.01	8.97	43.025

TABLE 5: Variance of harmonic content rate under each operating state of the device.

Variance	Closed	Gear 1	Gear 2	Gear 3	Gear 4	Gear 5
YD1	763.2436	1.21656	0.97821	0.24295		
YD2	702.8478	768.2751	0.582456	0.58845	90.05154	42.24794
YD3	777.4941	703.4177				
YD4	17.23165	555.0503	480.2903	474.9479		
YD5	582.9355	0.088918				
YD6	583.3648	220.6958				
YD7	8.813617	480.7053	473.4962	75.56688	32.17358	1.453375
YD8	166.9674	290.4106	0.128187	175.1147	274.4583	
YD9	36.81396	38.92979	51.2206	38.28987		
YD10	16.59171	107.8209	0.063964	12.29556	8.93058	
YD11	9.557222	5.628287				

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

$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\overline{H_1}\overline{H_2}N^2}} \sum_I \sqrt{H_1(I)H_2(I)}}. \quad (14)$$

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].

The entropy weight coefficient [20] of each target is expressed as follows:

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

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

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

	The current span	Area	Absolute area	Standard deviation of instantaneous resistance
X1	1.642	0.022	0.054	64.249
X2	1.277	0.049	0.079	13.390

TABLE 7: Amplitude of kth harmonic content rate of device.

Amplitude	2	3	4	5	6	7	8	9	10	11	2-norms
X1	0.98	0.99	0.97	0.98	0.88	0.95	0.98	0.88	0.99	0.94	3.01947
X2	15.7	54.12	21.21	49.32	15.3	34.93	17.81	29.42	15.72	29.94	99.16741

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

Variance	Closed	Gear 1	Gear 2	Gear 3	Gear 4	Gear 5
X1	0.047832	0.02007	7.51E-06			
X2	0.047832	0.080229	0.046945	0.037575		

TABLE 9: Similarity results of device 1.

Similarity	Similarity of transient characteristic	Similarity of V-I trajectory	Comparison similarity of active power and reactive power	Similarity of harmonic	The total similarity
Device 1—YD1	0.321	0.302	0.945	1.000708	0.63868
Device 1—YD2	1.228	1.134	0.675	1.000417	0.98893
Device 1—YD3	1.147	0.920	0.961	1.000407	0.99420
Device 1—YD4	0.434	0.973	0.782	1.000839	0.81413
Device 1—YD5	0.321	0.305	0.956	1.0011	0.64316
Device 1—YD6	0.148	1.265	0.943	1.000979	0.89269
Device 1—YD7	1.320	1.176	0.910	1.001123	1.09019
Device 1—YD8	3.001	0.664	0.898	1.002281	1.26949
Device 1—YD9	0.297	0.346	0.928	1.010204	0.64403
Device 1—YD10	1.361	0.786	0.926	1.006951	0.98783
Device 1—YD11	0.604	0.515	0.955	1.034707	0.76915

of feature similarities to obtain a total similarity of load feature similarity. The specific model is as follows:

$$
S_{ix} = \frac{w_1}{\left\|\|Z_i - Z_x\|/\|Z_i\|\right\|^2} + \frac{w_2}{\left\|\|V_i - V_x\|/\|V_i\|\right\|^2}
$$
$$
+ \frac{w_3}{\left\|\|X_i - X_x\|/\|X_i\|\right\|^2} + w_4 H_{ix}^2, \tag{16}
$$

where Z_i is the eigenvector of transient of YD_i, Z_x is the eigenvector of transient of YD_x, V_i—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_{ix} is the comparison similarity of active power and reactive power between the device to be tested and the known device.

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.

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 of X_1 and X_2 has little difference. For equipment X_1, firstly, the variance of harmonic content rate is small in the closed state, and then the harmonic content rate slowly decreases when switching to the 1st level, and finally the harmonic content rate continues to decrease when switching to the 2nd level. For the equipment X_2, 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 10: Similarity results of device 2.

Similarity	Similarity of transient characteristic	Similarity of V-I trajectory	Comparison similarity of active power and reactive power	Similarity of harmonic	The total similarity
Device2—YD1	2.488	0.772	0.9585	1.01977	1.22088
Device2—YD2	1.013	1.047	0.7076	1.011319	0.931428
Device2—YD3	1.009	1.003	0.9296	1.010978	0.983971
Device2—YD4	1.386	1.278	0.8192	1.021115	1.110883
Device2—YD5	2.021	0.753	0.9456	1.031931	1.120526
Device2—YD6	0.609	1.041	0.9594	1.027882	0.927835
Device2—YD7	1.038	1.197	0.9267	1.029494	1.050958
Device2—YD8	1.039	0.976	0.9022	1.073814	0.986183
Device2—YD9	4.198	1.025	0.9735	1.510558	1.741483
Device2—YD10	1.018	1.044	0.9365	1.156208	1.029209
Device2—YD11	1.187	1.146	0.9492	0.881266	1.042491

TABLE 11: Real-time power consumption of device 1.

ID	Time	Real-time power consumption (W)
Device 1	2018/1/16 16:04	0.1130
Device 1	2018/1/16 16:04	0.1117
Device 1	2018/1/16 16:04	0.1117
Device 1	2018/1/16 16:04	0.1121
Device 1	2018/1/16 16:04	0.1115
Device 1	2018/1/16 16:04	0.1115
Device 1	2018/1/16 16:04	0.1121
Device 1	2018/1/16 16:04	0.0605
Device 1	2018/1/16 16:04	0.0001
Device 1	2018/1/16 16:04	0.0001

Based on the result shown in Tables 9 and 10, the similarity between the unknown device X_1 and device 8 is the highest; that is, the unknown device X_1 is device 8. The similarity between the unknown device X_2 and device 9 is the highest; that is, the unknown device X_2 is device 9.

3.3.2. Calculation of Real-Time Power Consumption of Unknown Equipment. In the equipment data given in Annex 2, U, I, and PFC are the measured voltage, current, and power factor, respectively. The specific calculation formula of real-time power consumption is as follows:

$$P = \frac{U \times 10^{-1} \times I \times 10^{-3} \times PFC \times 10^{-3}}{3600}, \quad (17)$$

where U represents voltage, I represents current, and PFC represents power factor.

According to the above calculation formula and the data given in Annex 2, the real-time power consumption of the unknown device is obtained. There are some data of the real-time power consumption of the unknown device 1.

Table 11 shows partial data of calculation results of real-time power consumption of unknown equipment X_1.

4. Conclusion

Based on the data analysis, this paper firstly uses MATLAB to detect and distinguish the abnormal points and mutation points by using the method of wavelet decomposition W_k

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 kth ($k = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$) 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

References

[1] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[2] P. Li and Y. X. Yu, "Online decomposition of non-invasive power load," *Journal of Tianjin University (Natural Science and Engineering Technology Edition)*, vol. 42, no. 4, pp. 303–308, 2009.

[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] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load signature study—part II: disaggregation framework, simulation, and applications," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 561–569, 2010.

[5] W. Cai, H. G. Yang, and H. M. He, "Household load identification method and system based on proximity matching of transient characteristics: China," CN103439573A, 2013.

[6] X. X. Zheng, Q. Q. Liu, and S. F. Lin, "Research on the microscopic characteristics of residential load for non-interference load monitoring," *Power System Protection and Control*, vol. 42, no. 10, pp. 62–71, 2014.

[7] T. D. Huang, W. S. Wang, and K. L. Lian, "A new power signature for non-intrusive appliance load monitoring," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1994-1995, 2015.

[8] X. Wu, B. Qi, and L. Han, "Non-invasive fast identification algorithm for residential load based on template filtering," *Automation of Power System*, vol. 41, no. 2, pp. 135–141, 2017.

[9] R. X. Liang, M. Q. Yang, J. X. Dong, and W. Wang, "Research on intelligent identification and monitoring technology of non-invasive home appliances," *Science and Technology Innovation and Productivity*, vol. 25, no. 3, pp. 56–58, 2019.

[10] W. X. Su, Y. L. Zhu, F. Liu, and K. Y. Hu, "Detection algorithm for anomaly and mutation points in time series," *Computer Research and Development*, vol. 51, no. 4, pp. 781–788, 2014.

[11] Y. Feng and X. H. Wang, "Wavelet transform noise reduction processing and its Matlab implementation," *Data Acquisition and Processing*, vol. 6, no. 21, pp. 37–39, 2006.

[12] C. Yao, *Research and Design of Non-Invasive Low-Voltage Fault Monitoring System for User Internal Loop*, North China Electric Power University, Beijing, China, 2017.

[13] O. D. Luo, *Non-Invasive Load Feature Extraction and Recognition Research*, Southeast University, Nanjing, China, 2018.

[14] D. Hua, X. M. Chen, and L. J. Wang, "Non-invasive load decomposition method and system based on V-I trajectory: China," CN107302217A, 2017.

[15] 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.

[16] B. Jiang, *Deep Learning Based Non-invasive Residential Load Decomposition method*, Hefei University of Technology, Anhui, China, 2017.

[17] B. S. Li, H. J. Qiu, and X. Zhou, "Design of power load automatic monitoring system based on big data," *Electronic Design Engineering*, vol. 27, no. 8, pp. 95–98, 2019.

[18] G. Bradski and A. Kaehler, *Learning OpenCV*, O'Reilly Media, Inc., Newton, MA, USA, 2008.

[19] W. X. Zhu, Z. B. Jiang, G. L. Zhao, and J. C. Yang, "Residual resource evaluation of CBM mining based on principal component analysis and entropy method," *Coal Mine Safety*, vol. 50, no. 5, pp. 162–167, 2019.

[20] H. Wang and C. Y. Guo, "Study on the influence of linear dimensionless method on the index weight of entropy method," *China Population Resources and Environment*, vol. 27, no. 11, pp. 95–98, 2017.