Study of steady-state feature extraction algorithm based on EMD

Xianqing Huang¹, Bo Yin¹, Rui Zhang², Zhiqiang Wei¹,*

¹College of information science and engineering, Ocean University of China, Qingdao, China
²Qingdao Haier Smart Technology R&D Co., Ltd., Qingdao, China

*Corresponding author e-mail: seiouc16@163.com

Abstract  The accuracy of appliance identification is mainly determined by the selection of features in Non-Intrusive Load Monitoring (NILM). The time domain features extracted based on the appliance steady-state depends highly on the stable voltage. This paper proposes a frequency domain features extraction method, which applies the empirical mode decomposition (EMD) principle. The combination of frequency domain features and time domain features can effectively reduce the impact of voltage fluctuations on the accuracy of electrical identification. Through experimental comparison, it is proved that this method can solve time domain features overlapping, with higher anti-interference ability and robustness, especially for the non-resistive electronic loads.

1. Introduction

The implementation of smart grids, which is the development trend of the power system in the future, will fundamentally change the evaluation method of the power system, and then solve the problem of insufficient voltage in the power distribution system [1]. The intelligentization of load monitoring in power systems is a prominent feature of smart grids. With the loading monitoring, the precise information of various types of loads can be detected strictly, such as the status parameters, power consumption, run-time and so on. Thus, the loading monitoring can help users to allocate and save the electronic energy reasonably. It also improves the public awareness of eco-civilization and contributes to restrain the greenhouse effect and the climate change.

The Non-Intrusive Load Monitoring (NILM) [2], whose feature extraction technology and load recognition technology have become the research focuses in recent decades, is the main monitoring approach of smart grids. There are three steps in the NILM, which are event detection, feature extraction and electrical identification, respectively. The accuracy of load identification is determined by the selection of features, which are generally divided into the transient features and steady-state features. The load signature (LS) reflecting the transient process of electronic device is closely related to the function of the device. Considering the relatively few feature overlapping, the load monitoring based on the transient features can achieve satisfactory efficiency.
However, there are few features in the steady-state, and most of the steady-state feature extraction methods are developed to identify the current, power and other information in the time domain in previous studies. Besides, the voltage fluctuations often occur in the real-time electric environment, with the difference between the maximum and minimum voltage about 24V, as shown in Fig. 1. The voltage fluctuation has an obvious effect on the current and power. Hence, it is not reliable to identify the load status only based on the time domain features in electronic steady-state.

![Real-time voltage data](image)

**Figure 1.** The real-time voltage monitoring records

Therefore, this paper intends to analyze the steady-state feature extraction in the NILM using real-time datasets. In particular, a steady-state frequency domain feature extraction method using the empirical mode decomposition (EMD) principle is proposed. The combination of frequency domain features and time domain features can effectively reduce the impact of variable voltage on the accuracy of appliance identification. The structure of the paper is as follows. The related work is shown in section 2. Section 3 describes the EMD principle and three kinds of frequency domain feature extraction algorithms based on EMD presented in this paper. Section 4 details experimental verification and results analysis. The summary is presented in section 5.

2. Related Work

Steady-state refers to the status which is no failure, switching and gear changes in all kinds of loads, and the electronic parameters of the whole system are relative steady. Many other methods for the steady-state extraction have been proposed in previous work during the few years. The variable-speed drive (VSD) power estimation method proposed by Liang et al. [3], which can extract the steady-state information of current, voltage and power, and take the corresponding peaks, mean values and root mean square (RMS) values as the features, can distinguish the load types easily. However, it is difficult to solve the feature overlapping. Besides, the power estimation methods based on relationships between fundamental and higher harmonics of the measured current [4] and a switching-function [5], which could be used for NILM, have been developed to estimate the power consumption of VSDs. However, both methods can’t solve the local feature overlapping and crossover distortion of the harmonic component and make further distinction for different SVDs. Moreover, combing the classification algorithms, the steady-state V-I trajectory based on the load signatures for the NILM can achieve a better efficiency than that for the power feature [6]. However, this method cannot work in the small power loads and the loads with continuous changes.

In addition, ElectriSense, as a single point sensing for NILM, has been presented in Gupta et al. [7].
This method can be used to identify the electronic devices with switch mode power supplies (SMPS). Different from the monitoring of transient noise, ElectriSense performs event detection and load identification by applying electromagnetic interference (EMI). This family-friendly device is able to distinguish similar devices in the home. However, this method is difficult to be generalized, because of the poor capacity of resisting disturbance and generality. Patel et al. proposed a method of monitoring the operation of a household load using a single sensor [8]. This is a plug-in sensor that monitors the noise generated on the power line during home load state switching and load operation. This method uses machine learning techniques for noise analysis, and the accuracy of electrical identification is 85%-90%. The advantage of this approach is that it has the capability of resisting disturbance and its testing process is simpler. However, there are also some disadvantages to the multi-mode loads.

Therefore, we put forward the frequency domain feature extraction method based on EMD in this paper. Three types of steady-state features are selected and the time domain feature are combined with frequency domain feature to improve the capability of identification and anti-interference.

3. Methods

3.1 EMD

The EMD, a method that has obvious advantages in dealing with nonlinear non-stationary signals, can adaptively decompose the original signal into a set of better-performing intrinsic mode functions (IMFs) according to the time scale features of the data itself [9]. This method does not require any basis function to be preset, and can be applied to the decomposition of any type of signal in theory, so it has been applied quickly and effectively in the engineering field. Each eigenmode component contains local features of different time scales of the original signal and satisfies two conditions. Firstly, the number of extreme points of the signal and the number of zero crossings are equal or at most one different. Secondly, the average value of the upper envelope and the lower envelope is zero, and the upper and lower envelopes are respectively composed of local maximum values and local minimum values.

The EMD, which can realize the transition of the multi-signal to the single signal, is an adaptive and temporally local algorithm. As originally proposed, EMD method is implemented through a shifting process:

$$s(t) = \sum_{i=1}^{N} c_i(t) + r_n(t)$$

where, $s(t)$ is any given data; $c_i(t)$ are the IMF components; $r_n(t)$ is the residual, which is a monotonic function with no more than one extremum, showing the trend of $s(t)$. The detailed description of the EMD method can be found in the studies of Huang et al. [9, 10].

3.2 Feature selection algorithms

The original data $s(n)$ can be decomposed into M IMFs by employing the EMD method. The number of the samples including in the M-th mode is N, in which the IF of the nth sample is $f_{mn}$. Based on this method, we can extract the features of all IMFs.

3.2.1 Contribution Rate of IMFs (IMFCR). The contribution rate is the ratio of each IMF in the total IMFs. The energy of all components, calculated by the time domain energy of M IMFs, that is,
\[ E(x_i(n)) = \sum_{n=1}^{N} |x_i(n)|^2 \quad 1 \leq i \leq M \]  

Total energy of \( s(n) \) is the sum time domain energy of all components. The feature coefficient is defined as the contribution rate \( k_i \) of \( i \)-th IMF.

\[ E(s(n)) = \sum_{i=1}^{N} E(x_i(n)) \quad 1 \leq i \leq M \]  

\[ k_i = \frac{E_i}{E} \quad 1 \leq i \leq M \]  

3.2.2 Average IF of IMFs (IMFAIF). The mean value of the IF \( (f_{mn}) \) during some time is the IF of the \( m \)-th component, which is defined as \( \overline{f_m} \), be

\[ \overline{f_m} = \frac{1}{N} \sum_{n=1}^{N} f_{mn} \]  

The mean IF can exhibit the feature of the frequency distribution of the original data.

3.2.3 Center frequency of the strongest IMF (SIMFCF). The IF of each IMF always fluctuates around a center frequency, which is defined as the center frequency of the \( m \)-th IMF. Assuming the IF of the \( n \)-th samples in the \( m \)-th IMF is \( b_{mn} \), thus the instantaneous strength is

\[ Q_{mn} = b_{mn}^2 \]  

The center frequency of the \( m \)-th IMF is defined as \( \tilde{f}_m \), be

\[ \tilde{f}_m = \frac{\sum_{n=1}^{N} Q_{mn} f_{mn}}{\sum_{n=1}^{N} Q_{mn}} \]  

The mean strength of the \( m \)-th IMF is

\[ \tilde{B}_m = \frac{\sum_{n=1}^{N} Q_{mn}}{N} \]  

The center frequency of the IMF with the strongest mean strength is defined as the strongest center frequency of strongest IMF, \( \tilde{f} \), be

\[ \tilde{f} = \tilde{f}_{\text{max}}(\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_M) \]  

The center frequency of the strongest IMF reflects the magnitude of the frequency of the strongest component.

The frequency domain features, which can be extracted by above three algorithms, and the time domain features are taken as the criterion of classification. The experimental analysis will be given in the following sections.
4. Experimental Results and Discussion

4.1 Data preparation
The STMicroelectronics 32-bit chip (STM32), which is used as the core processor, the current sensors and voltage sensors are attached to create a NILM in this paper. The sample frequency is about 4KHz. To verify the effect of voltage fluctuation on the accuracy of load identification, the experiment is divided into two parts. One part is to collect the state of load devices during three periods of the wide voltage fluctuations, which are 8:30am, 11:30am and 5:30pm. There are 50 groups of samples were collected for each load devices during each period. The other part is to collect 150 groups of samples for each load devices under the stable voltage conditions. The data collected from seven load devices usually used in family, including hair dryer, microwave oven, vacuum cleaner, refrigerator, kettle, desktop PC and TV. The load types and data acquisition records are shown in Table 1.

Table 1. Load types and data acquisition records

| NO. | Load type       | 8:30am | 11:30am | 17:30pm | Stable Voltage |
|-----|-----------------|--------|---------|---------|----------------|
| 1   | Hair dryer      | 50     | 50      | 50      | 150            |
| 2   | Microwave oven  | 50     | 50      | 50      | 150            |
| 3   | Vacuum cleaner  | 50     | 50      | 50      | 150            |
| 4   | Refrigerator    | 50     | 50      | 50      | 150            |
| 5   | Kettle          | 50     | 50      | 50      | 150            |
| 6   | Desktop PC      | 50     | 50      | 50      | 150            |
| 7   | TV              | 50     | 50      | 50      | 150            |

4.2 Performance of time-domain features
Taking the opening point of the load devices as the starting, the data within 2 seconds, that is, 100 periods of steady-state, are extracted from each group. Six features of time domain, including the current peak, current mean value, current RMS, current peak-to-peak value and current energy, and the apparent power during this period, are extracted as the criteria of classification.

![Figure 2](image-url)  
**Figure 2**. The actual classification and predictive classification of time domain features using SVM under the condition of: (a) variable voltage; (b) stable voltage.

For the 50 groups of data sampled during different periods from each load devices, the 30 groups are the training samples, and the remaining 20 groups are the test samples. Similarly, for the 150
groups of data, the 90 groups are taken as the training samples, and the remaining 60 groups are the test samples. The support vector machine (SVM) is used to predict and classify load devices in this paper. The result, as shown in Fig. 2, indicates that the accuracy of load identification is seriously influenced by the voltage fluctuations. For the classification of loads, it doesn’t satisfy the actual application requirements only employing the six features of time domain.

Table 2. The result of prediction based on time domain features

| NO. | Load Type      | Variable Voltage | Stable Voltage |
|-----|----------------|------------------|----------------|
|     |                | Correct/Total    | Accuracy       | Correct/Total | Accuracy       |
| 1   | Hair dryer     | 43/60            | 71.7%          | 52/60         | 86.7%          |
| 2   | Microwave oven | 41/60            | 68.3%          | 51/60         | 85.0%          |
| 3   | Vacuum cleaner | 39/60            | 65.0%          | 45/60         | 75.0%          |
| 4   | Refrigerator   | 35/60            | 58.3%          | 42/60         | 70.0%          |
| 5   | Kettle         | 47/60            | 78.3%          | 55/60         | 91.7%          |
| 6   | Desktop PC     | 38/60            | 63.3%          | 44/60         | 73.3%          |
| 7   | TV             | 40/60            | 66.7%          | 47/60         | 78.3%          |

As shown in Table 2, the maximum identification rate of selected loads under the variable voltage achieves 78.3%, and the maximum identification rate under the stable voltage is as high as 91.7%. The effect of voltage fluctuations on the identification rate is about 12%. With taking the time domain features as the criterion of classification, the identification rate of resistive loads is higher than that of non-resistive loads.

4.3 Performance of proposed algorithm

Figure 3. The actual classification and predictive classification of frequency domain features combined with time domain features using SVM under the condition of: (a) variable voltage; (b) stable voltage.

The steady-state data during the 100 periods are decomposed by EMD method into 9 IMFs and one trend in this paper. We define the maximum and minimum values of IMFCR and IMF AIF as the features. Additionally, we also take the SIMFCF as the features. These five frequency domain features and the time domain features are taken as the criterion of classification. The result, as shown in Fig. 3, indicates that the accuracy of load identification is obvious higher than that in section 4.2, especially
for non-resistive devices, no matter whether the voltage changes or not.

Table 3. The result of prediction using proposed algorithm

| NO. | Load Type      | Varying Voltage | Stable Voltage |
|-----|----------------|-----------------|----------------|
|     | Correct/Total  | Accuracy        | Correct/Total  | Accuracy        |
| 1   | Hair dryer     | 57/60           | 95.0%          | 57/60           | 95.0%          |
| 2   | Microwave oven | 56/60           | 93.3%          | 55/60           | 91.7%          |
| 3   | Vacuum cleaner | 55/60           | 91.7%          | 56/60           | 93.3%          |
| 4   | Refrigerator   | 52/60           | 86.7%          | 53/60           | 88.3%          |
| 5   | Kettle         | 59/60           | 98.3%          | 60/60           | 100.0%         |
| 6   | Desktop PC     | 54/60           | 90.0%          | 54/60           | 90.0%          |
| 7   | TV             | 55/60           | 91.7%          | 56/60           | 93.3%          |

As shown in Table 3, the frequency domain features extracted based on EMD can reflect the character of load devices. This character is hardly affected by voltage changes, which is an inherent property of electronic equipment. This performance is very apparently for the non-resistive load devices. The accuracy of load identification will increase continuously with the rising number of samples.

5. Conclusion

The frequency domain feature extraction method based on EMD is proposed and the concepts of IMFCR, IMFAIF, and SIMFCF are presented in this paper. Besides, five frequency domain steady-state features are extracted based on these three types of concepts. Combined with time domain features, these five features can be used in the identification of NILM. This method reduces the dependency of load identification on the stable voltage and improves the accuracy of load identification under the variable voltage. Furthermore, the test results show that the good capability of this method to solve the feature overlapping in time domain and improve the anti-interference capability, especially for the non-resistive load devices.

Acknowledgments

This work was supported in part by China International Scientific and Technological Cooperation Special (2015DFR10490), Qingdao national laboratory for marine science and technology Aoshan science and technology innovation project (2016ASKJ07-4) and Qingdao innovation and entrepreneurship leading talent project (13-cx-2).

References

[1] Kerry D. McBee, Marcelo G., Simões. Utilizing a Smart Grid Monitoring System to Improve Voltage Quality of Customers, IEEE Transactions on Smart Grid, 3(2012) 2: 738-743.
[2] Hart G W, Nonintrusive appliance load monitoring, Proceedings of the IEEE, 80(1992) 12: 1870-1891.
[3] Liang J, Ng S K K, Kendall G, and Cheng J W M, Load signature study, part I: basic concept, structure, and methodology, IEEE Transactions on Power Delivery, 25(2010) 2: 551-560.
[4] Lee K D, Lee S B, Norford L K, Armstrong P R, Holloway J, and Shaw S R, Estimation of variable-speed-drive power consumption from harmonic content, IEEE Transactions on Energy Conversion, 20(2005) 3: 566-574.
[5] Wichakool W, Avestruz A, Cox R W, and Leeb S B, Modeling and estimating current harmonics of variable electronic loads, IEEE Transactions on Power Electronics, 24(2009) 12:2803-2811.

[6] Hassan T, Javed F, and Arshad N, An empirical investigation of VI trajectory based load signatures for non-intrusive load monitoring, IEEE Transactions on Smart Grid, 5(2014) 2: 870-878.

[7] Gupta S, Reynolds M S, and Patel S N, ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home, Proceedings of the 12th ACM international conference on Ubiquitous computing, Copenhagen, Denmark, (2010): 139-148.

[8] Patel S N, Robertson T, Kientz J A, Reynolds M S, and Abowd G D, At the flick of a switch: Detecting and classifying unique electrical events on the residential power line, Lecture Notes in Computer Science, 4717(2007): 271-288.

[9] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N-C, Tung C C, and Liu H H, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. Roy. Soc. Lon., 454A(1998): 903-995.

[10] Huang N E, Shen Z, and Long R S, A new view of nonlinear water waves-the Hilbert spectrum, Ann. Rev. Fluid Mech., 31(1999): 417-457.