Non-intrusive Load Monitoring Method Based on SAGA-FCM Algorithm

Zhou-da HU, Yang FANG, Wen-quan YE and Yuan-xi YIN
Yudean Group Guizhou Co., Ltd. Guiyang 550081, Guizhou, China

Keywords: Non-intrusive load monitoring, Delta feature extraction, Genetic simulated annealing algorithm, The fuzzy C-means algorithm.

Abstract. Existing non-intrusive algorithms have some deficiencies such as low recognition accuracy and the accuracy decreases significantly when the amount of monitoring data is too large. An improved fuzzy C-means algorithm (FCM) is proposed in this paper which will make up the shortcomings of existing algorithms. In order to optimize the clustering process, the simulated annealing algorithm (SA) and genetic algorithm (GA) are introduced to fasten the process. The experimental and comparison results fully demonstrate the superiority of the new FCM in terms of recognition, robustness, stability and reliability. Furthermore, the new FCM is applied to practical household appliances. All the comparison results consistently indicate that the new FCM is highly competitive and can be used in big data monitoring environment.

Introduction

It is an important goal of the smart grid to build an interactive service system for power grids and power users to realize smart power usage [1]. Load monitoring is the basis for “intelligent interaction” between grid and power users. Non-Intrusive Load Monitoring (NILM) was first proposed by Hart in the 1980s. Its aim is to decompose the load, that is, to install sensors at the entrance of the power user to obtain load information, and identify it through complex algorithms. User's power usage information. Compared with intrusive load monitoring, it has the advantages of economy, simplicity, privacy protection, practicability and easy promotion. It is especially suitable for ordinary residents [2, 3]. With the rise of data science and smart home, NILM has become a research focus in academic community. The core content of NILM is load identification. In [4], the heuristic algorithm is used to solve the steady-state current decomposition model, but the load type is single and the decomposition error is large. In [5], the transient characteristics of the start and stop of the household load are proposed as the load imprint, which is compared with the standard library for closeness, but the effect of superimposing multiple loads is not good. In [6], clustering is used to cluster reactive power and active power, but the recognition rate of devices with similar power characteristics is limited. Reference [7] uses fuzzy recognition method to convert reactive power and harmonic amplitude into Row fuzzy C-means clustering, but not suitable for large-data monitoring scenarios. In [8], the load current is extracted as the load template filter, and the total current is filtered. However, with the increase of the monitoring data, the recognition accuracy of the method will be significantly reduced, and the application in practical engineering is limited.

In response to the above problems, this paper proposes a new NILM method. After sampling the load data of the power user, the isolated noise point is removed by median filtering, and the difference feature of the active power and the harmonic amplitude is extracted by the difference extraction method to obtain a reliable cluster sample. The simulated annealing algorithm is combined with the genetic algorithm for FCM cluster analysis to realize the identification of different devices. Finally, the actual data is tested to verify the accuracy and stability of the method identification in different environments.

Analysis of Household Load Characteristics

The NILM system collects data and analyzes and identifies the data through the NILM system by installing an information collecting device such as a smart meter at the residential line of the resident user, and then uploads the identification result and various types of electrical parameter
data to the NILM server, and simultaneously presents the data. To the user terminal, it is convenient for the user to operate feedback and form an interactive mode [9].

The unique characteristics of the electrical equipment that can be reflected in the operation are called the load fingerprint [10]. Active power and reactive power are the most commonly used load marks. However, the power characteristics of low-power appliances are relatively close. It is difficult to distinguish them accurately only by active power and reactive power. Therefore, it is necessary to add a new load mark for more accurate discrimination. In [11], through the load monitoring of multiple single devices, the load characteristic database of 32 kinds of electrical equipment was established, and it was found that the current amplitude of the third harmonic has a greater discrimination for different low-power electrical equipment. Therefore, this paper selects the three-current harmonic amplitude with the largest difference as the incremental load imprint to solve the problem of low recognition of low-power electrical equipment.

Data Preprocessing and Difference Feature Extraction

Data Preprocessing

In order to reduce random changes in the monitoring data, this paper uses a filter to process the data. Median filtering is a nonlinear smoothing technique based on sorting statistics theory, which has a very effective effect on eliminating noise points and abnormal data [12]. The main steps are as follows:

Step 1. Let the data sequence be \( y_1, y_2, y_3, \ldots, y_n \). After the data is sorted in ascending order, it is \( y_{i1}, y_{i2}, y_{i3}, \ldots, y_{in} \).

Step 2. When \( n \) is an odd number, the median is represented by (1):

\[
\text{median}_i = \frac{y_{(n+1)/2}}{2}
\]

When \( n \) is an even number, the median is represented by (2):

\[
\text{median}_i = \frac{1}{2} \left( y_{(i+1)/2} + y_{i+1/2} \right)
\]

Differential Feature Extraction

For any electrical equipment \( m \), define the load stamp at any time \( t \) as follows:

\[
y_m(t) = [\beta_{m,1}, \beta_{m,2}, \cdots, \beta_{m,N}]
\]

Where \( \beta_{m,n} \) is the \( n \)-th load imprint of device \( m \) at time \( t \), and \( N \) is the total number of load imprints.

When multiple devices are running at the same time, let \( J \) be the number of devices, then the total load stamp is defined as follows:

\[
\varphi(t) = \sum_{m=1}^{J} y_m(t)
\]

A variety of devices operate at the same time. The load characteristics of low-power devices are easily obscured by the load characteristics of high-power devices. To increase the recognition rate of low-power devices, the load characteristic difference is extracted as a load imprint, and the load difference at any sampling time \( t \) is defined. feature:

\[
\varphi(\Delta t) = \varphi(t + \Delta t) - \varphi(t)
\]

\[
= \left\{ \sum_{i=1}^{N} y_i(\Delta t) \mid \Delta t = T \right\}
\]

Where \( T \) is the sampling period, which is the sampling interval, so that the sampling period is sufficiently small, and every two sampling moments can be equated with a change of at most one electrical load state, ie
\[ \varphi(\Delta t) = \begin{cases} y_m(t) & t \not\in s \\ 0 & t \in s \end{cases} \]  

Through (5), the difference characteristics of the same time series of each load characteristic are obtained. Because the equipment is in normal operation, it is affected by the fluctuation of the power grid, and the power has slight fluctuation. This part of data needs to be eliminated.

The time series corresponding to the difference data of 0.5% of the active power is sorted in ascending order, defined as a normally open time series \( s \), and the sequence needs to be eliminated from the total time series \( t \), namely:

\[ \Delta \equiv \begin{cases} y_m(t) & t \not\in s \\ 0 & t \in s \end{cases} \]

Through the equations (5) and (6), the decomposition problem of various equipment load characteristics is transformed into a single electrical load characteristic problem at any time, and then the NILM method is used for cluster analysis to realize the monitoring and decomposition of various equipment loads.

**Principle and Implementation of the NILM Method in This Paper**

**FCM Algorithm in the Implementation of NILM**

Fuzzy C-means clustering (FCM) is an algorithm for determining the automatic classification of data by assigning each data point to a certain class according to membership degree [13]. Therefore, this paper chooses FCM algorithm as the main algorithm for classification and identification of load imprint.

The load imprint obtained by the acquisition and NILM system is made into the data set \( X=\{x_{1,m}, x_{2,m}, \ldots, x_{n,m}\} \), the number of samples is \( n \), each sample is an \( m \)-dimensional vector, and the power is used in this paper. And two-dimensional vectors of harmonics. Define \( u_{ik} \) as the membership of the sample \( x_i \) for the class \( D_k \). The objective function \( J_g \) is as follows:

\[ J_g(U,v) = \sum_{k=1}^{cn} \sum_{i=1}^{n} (u_{ik})^g (d_{ik})^2 \]  

In the function, \( cn \) is the number of clusters, which means that \( X \) is divided into subsets of \( cn \) categories \( \{D_1, D_2, \ldots, D_{cn}\} \). \( U \) is a similar classification matrix, and the cluster centers of each type are \( \{v_1, v_2, \ldots, v_{cn}\} \). \( g \) is the weighting coefficient, and this paper takes 3. In addition, \( d_{ik} \) is the Euclidean distance, which is used to measure the distance between the sample \( i \) and the center point of the category \( D_k \). The function is as follows:

\[ d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - v_{ij})^2} \]  

The essence of the FCM algorithm is to find the optimal classification by iteratively, so that the function value \( J_g \) generated by the classification is minimized. During the period, the membership function sum of each cluster is required to be 1, and the membership degree \( u_{ik} \) and \( cn \) cluster center \( \{v_i\} \) of the sample \( x_i \) for the class \( D_k \) are calculated as follows:

\[ u_{ik} = \frac{1}{\sum_{j=1}^{cn} \left( \frac{d_{ij}}{d_{ik}} \right)^{2g}} \]  

Assume \( I_k=\{i|cn\in[2,n];d_{ik}=0\} \), for, \( i \in I_k, u_{ik}=0 \).

\[ v_{ik} = \frac{\sum_{i=1}^{n} (u_{ik})^2}{\sum_{k=1}^{n} (u_{ik})^g} \]
Through (10) and (11), the load imprint membership degree, cluster center and reclassification are modified several times. When the algorithm converges, the data of cluster center and membership degree are theoretically determined.

After obtaining the cluster center, the load attribution is confirmed by comparing the actual test value with the cluster center. Given the subset $D$ of the domain $U$ and the subset $V$ of the cluster center, the maximum and minimum closeness is calculated as follows:

$$N(D, V) = \frac{\sum_{i=1}^{n} u_D(x) \wedge u_V(x)}{\sum_{i=1}^{n} u_D(x) \vee u_V(x)}$$

The above formula: For taking the big operator, for the small operator, $u_D(x)$, $u_V(x)$ is the membership of the element $x$ to $D$, $V$ in the domain $U$. When the following formula is satisfied, it can be considered that $D$ and $V_j$ are closest to each other, and $D$ is attributed to $V_j$.

$$N(D, V) = \max_{i \in [1..n]} N(D, V_j)$$

### SAGA Optimization of FCM Algorithm

The FCM algorithm has a high search speed, but it is easy to converge to local minimum points. In this paper, the simulated annealing algorithm (SA) and genetic algorithm (GA) are combined to use fuzzy C-means clustering for load imprinting.

According to the advantages and disadvantages of the two, the two algorithms are organically combined and applied to FCM cluster analysis (SAGA-FCM), which can effectively overcome the phenomenon of GA early premature and late evolutionary stagnation, and design according to FCM clustering problem in NILM method. The genetic coding method and optimization model can make the algorithm get rid of the local optimum and converge to the global optimal solution more efficiently, so as to improve the accuracy of NILM recognition. The specific implementation steps are as follows:

- **Step 1.** Initialization control parameters: annealing initial temperature $T_0$, termination temperature $T_g$, temperature cooling coefficient $k$, population individual size $S$, maximum evolution number $M$.

- **Step 2.** Generate an initial population $Chrom$, randomly initialize $cn$ cluster centers, and each cluster center membership degree is calculated by (10). The individual fitness function uses a sorted fitness allocation function, and the objective function $J_g$ is calculated by equation (8). Individual chromosome coding: using binary coding, each chromosome consists of $cn$ cluster centers, $m$-dimensional sample vectors, and the number of sample variables to be optimized is.

- **Step3.** Set the loop count variable $g=0$.

- **Step 4.** Genetic manipulation of selection, crossover, and mutation of the population $Chrom$. The new individual is generated by using equations (10) and (11) to calculate the membership of the cluster center each sample, and the individual fitness. If so, replace the old individual with a new individual; otherwise, accept the new individual with probability and discard the old individual.

- **Step 5.** If yes, go to step $d$, otherwise go to step $f$.

- **Step 6.** If the process ends, return the global optimal solution; otherwise, perform the cooling operation and go to step $c$.

### Experimental Results and Analysis

In order to verify the recognition ability of the algorithm under complex load conditions, 10 kinds of household appliances such as electric fan, microwave oven, hot water bottle, notebook computer, incandescent lamp, printer, water dispenser, air conditioner, hair dryer and TV were selected. These devices are randomly started and stopped 200 times.
Fig. 1 and Fig. 2 are obtained, and the number set is divided into class A, and then calculate the closeness of each set of numbers in the standard load library and the set of Ai class numbers, and identify the device type of the experimental scene. Due to the variety of start-stop difference feature is more complicated, and it is impossible to visually observe, judge and distinguish the normalized current harmonic amplitude difference feature as in scene 1, but after introducing current harmonic information, the recognition rate of the low-power electrical equipment is increased by the algorithm, and the overall load recognition rate has also been improved accordingly.

Fig. 3 shows the variation of the algorithm and the traditional FCM algorithm objective function $J_g$ with the number of iterations.
The convergence speed of the algorithm is fast, and the value of the final objective function is always lower than the traditional FCM algorithm, which means that the load recognition ability of the algorithm is always higher than the traditional FCM algorithm. This means that with the increase of the monitoring load characteristic data, the accuracy rate of the algorithm in this paper is significantly smaller than that of the traditional FCM algorithm, and it has better robustness. Therefore, in the face of complex monitoring environment, the stability of this algorithm is higher, and it is more suitable for NILM large data volume monitoring environment.

Summary

The recognition accuracy of the traditional NILM method will be greatly reduced. In view of the problem, a non-intrusive load monitoring method based on the SAGA-FCM algorithm was proposed in this paper. The experimental and comparison results fully demonstrate the superiority of the new FCM in terms of recognition, stability and reliability. However, the calculation structure of this method is relatively complicated, the recognition speed is slow, and the recognition ability of electrical equipment with similar power characteristics and harmonic characteristics is limited.

Acknowledgement

This research was financially supported by the Science and Technology Foundation of GuiZhou Province (Grant No.[2018]2151).

References

[1] Cheng Xiang, Li Linzhi. A Survey of the Research on Non-intrusive Load Monitoring and Disaggregation[J]. Power System Technology, 2016, 40(10): 3108-3117.

[2] Wu Xin, Han Xiao. Non-Intrusive Residential Load Decomposition Algorithm Solving Underdetermined Equations Based on Signal Smartification[J]. Grid Technology, 2017, 41(9):3033-3040.

[3] Qi Bing, Han Lu. A non-intrusive residential load identification algorithm based on genetic optimization [J]. Electrical Measurement and Instrumentation, 2017, 54(17):11-17.

[4] LI Peng. Non-Intrusive Power Load Monitoring and Disaggregation Technique[D]. Tianjin University, 2009.

[5] Gao Yun, Yang Hong geng. Household Load Identification Based on Closeness Matching of Transient Characteristics[J]. Automation of Electric Power Systems, 2013, 37(09):54-59.

[6] LIN Y H, TSAI M S. Non-intrusive load monitoring by novel neuro-fuzzy classification considering uncertainties [J].IEEE Trans on Smart Grid, 2014, 5(5):2376-2384.

[7] Sun Yi, Cui Can, Lu Jun, etc. Non-intrusive Load Monitoring Method Based on Delta Feature Extraction and Fuzzy Clustering [J].Automation of Electric Power Systems, 2017, 41(04):86-91.

[8] Wu Xi, Qi Bing,etc. Fast Non-intrusive Load Identification Algorithm for Resident Load Based on Template Filtering. Automation of Electric Power Systems, 2017, 41(02):135-141.

[9] Sun Yi, Zhang Lu,Zhang Honglei,etc. A Non-Intrusive Household Load Monitoring Method Based on Dynamic Adaptive Particle Swarm Optimization Algorithm[J]. Grid Technology, 2018, 42(06):1819-1826.

[10] Liang J, Ngskk, Kendal G, et al. Load signature study: Part I basic concept, structure, and methodology [J]. IEEE Trants. on Power Delivery, 2010, 25(2):551-560.
[11] Yang Aibing, Dong Tianqiang, Chen Bo, Zhou Ke. Measurement and Analysis of Channel Transmission Characteristics for Low-voltage Power Networks [J]. Laboratory research and exploration, 2017, 36(01):136-139+143.

[12] Kong Weigong, Li Lirong. Simulation Research on the Grid-connected System Inverters Based on Active Power Filter [J]. Laboratory research and exploration, 2017, 36(07):103-107.

[13] Xu Xueliang, Yi Pengfei, Wang Haibin, ZHANG. Study on breath sound recognition based on FCM clustering algorithm [J]. Modern Electronic Technology, 2015, 38(12):16-19.