Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/55/20241663

Research on the design of non-invasive load monitoring system

Taizhong Chen

Nanning No.2 High School, Nanning, Guangxi, 530000, China

2825353423@qq.com

Abstract. In industrial production, machine failure will cause production efficiency decline, maintenance costs increase and other problems, while the traditional manual maintenance can only be carried out after the failure, cannot play an early warning effect, and has low efficiency. Non-invasive load identification and condition estimation can predict upcoming failures, remind maintenance in time, reduce failures, and improve production efficiency. In this paper, a non-intrusive load monitoring system is designed based on a non-contact high-frequency synchronous power data acquisition device. A non-intrusive on-line load monitoring system is constructed by building the load feature database offline and training the model. After event detection and feature extraction of the data sent by the power data acquisition device to the server, these data are input into the model for load identification. The experimental results show that the system has high precision of load identification and can meet the needs of practical life and engineering.

Keywords: aviation, Object recognition, Industrial production defect detection, Troubleshooting, State estimation

1. Introduction

Non-intrusive Load Monitoring, or NILM, is a technology designed to obtain information about a user's energy consumption at minimal cost. The technique was proposed by Professor Hart in the 1990s, and has both academic and practical characteristics. NILM can monitor the household electricity load by installing equipment at the user's power entrance (i.e., the household bus) [1]. Compared with the traditional method, this method reduces the installation cost and is highly accepted by users. The non-invasive load monitoring system can decompose the operation of each power load from the whole load. By analyzing the usage of each load, we can deeply understand the power consumption behavior of users and form an optimal power supply strategy. This helps to improve the power quality and reliability of the power supply side. At the same time, using fine-grained electricity consumption information as the basis, the user can optimize their own electricity consumption behavior and reduce unnecessary energy consumption.

In summary, non-invasive load monitoring technology uses low-cost equipment to obtain information about household energy consumption. Through comprehensive analysis of each power load and corresponding strategies, the quality and reliability of power supply can be improved, and users can be encouraged to optimize power consumption behavior and reduce energy waste. The development of this technology is of great significance in both academia and practice.

2. Design of a non-invasive load identification scheme

2.1. Event detection and current separation

2.1.1. Event detection method based on nonparametric CUSUM. A load event is usually defined as a load switch or a change in the operating state of the load. When an event occurs, the current and power sequence will change significantly, so the event detection problem can be regarded as a change point detection problem. Cumulative Sum (CUSUM) is a commonly used algorithm for change point detection. The CUSUM algorithm can be divided into parametric CUSUM and non-parametric CUSUM, depending on whether it is necessary to know the probability distribution of data in advance [2].

In view of the constant change in power load, it is more meaningful to adopt the non-parametric method in the NILM field for random variables such as power and current. This is because we cannot obtain the probability distribution of these random variables in advance.

By using non-parametric CUSUM algorithm, the load events existing in the power system can be effectively detected in NILM research without relying on any assumed knowledge about the probability distribution of the data [3]. This method can be adapted to load variations of different types and sizes and provides accurate and reliable results.

Therefore, when dealing with NILM related problems, the study of non-parametric methods can better cope with the challenges brought by complex diversity, wide range fluctuation and unknown probability distribution, so as to provide a feasible and effective method for the detection and analysis of power system load events.

The main idea of non-parametric CUSUM is that when the statistics of concern are higher or lower than the average under normal operating conditions, an event may occur. The difference between the monitoring amount and the average is accumulated, and when the accumulation exceeds a certain threshold value, an event is determined to occur, and then the moment of the event can be pushed backward. The algorithm flow is as follows: Let a time sequence be $\mathrm{T} = \{\mathrm{t}(\mathrm{k})\}(\mathrm{k} = 1,2,\cdots)$, t is the moment, and the statistical function s is defined as:

$$\mathrm{So} = 0 \tag{1}$$

$$\mathrm{Sk} = \max(0, \mathrm{Sk} - 1 + \mathrm{tk} - \mu\mathrm{o} - \beta) \tag{2}$$

Where: μo is the mean value of the sequence before the change point; Beta is the noise level. When Sk > 0, the sampling sequence is considered to change at this time. If the threshold of event occurrence h is not reached, the initialization delay time m = 0. Let m = m+1 and continue the cumulative sum until sk > h, the moment when the event is considered to be detected can be backward deduced to t = k - m.

2.1.2. Current separation method based on parallel circuit. After the event is detected, we need to isolate the current sequence for each individual load from the total load current data. Since each load current is independent of each other and conforms to the basic principle of parallel circuit, the overall total household current can be regarded as the superposition of each load current [4].

Considering that the voltage signal is always relatively stable and the phase is unchanged, and the phase of the current is determined by the phase of the voltage, it is only necessary to ensure that the current signal at the steady-state moment is extracted under the same initial phase. By using the principle of parallel circuit, two sets of current sequences at steady state time are extracted before and after the event, and then they are subtracted to get the only current sequence in and out of the load part generated during the single load event.

This paper will choose the moment when the voltage crosses zero and begins to rise as the benchmark to extract the correct sequence under these critical moments. The specific steps are as follows:

(1)The steady-state current sequence after the event is extracted. When the steady-state voltage crosses zero and rises, the steady-state current sequence is collected.

(2)The steady-state current sequence before the event is extracted. The steady-state current sequence of the same length is collected using the time when the steady-state voltage crosses zero and rises before the event occurs as the reference.

(3)Single load current separation. Make a difference between the current sequence in Step 1 and the current sequence in Step 2, as shown in the equation.

$$\mathrm{Is(k) = Im1(k) - Im2(k)(k = 1,2,\cdots,N)} \tag{3}$$

Where: Is (k), Im1 (k), Im2 (k) are the KTH sampling points of single load, steady-state current before and after the event respectively; N is the length of the current sequence.

2.2. Feature extraction

Load characteristics are the key to load identification. After event detection and current separation, we can obtain the current sequence of a single load during steady-state operation [5]. Figure 1 shows the steady-state current waveforms for six common household loads. Although using the complete current sequence directly as the load feature can achieve a high level of recognition, it will increase the computational complexity because the sequence is too long. Therefore, it is necessary to extract effective load characteristics from the current sequence.

Nonlinear devices (such as various types of electronic devices) have different degrees of harmonic components, while common linear devices (such as incandescent lamps, electric heating hair dryers, etc.) have power level differences. Therefore, we can use the fundamental and odd harmonics (3rd, 5th, 15th) in the steady-state current as load characteristics to distinguish them. In this paper, fast Fourier transform is used to extract fundamental and odd harmonics in steady-state current, which are further analyzed and processed as load characteristics [6].

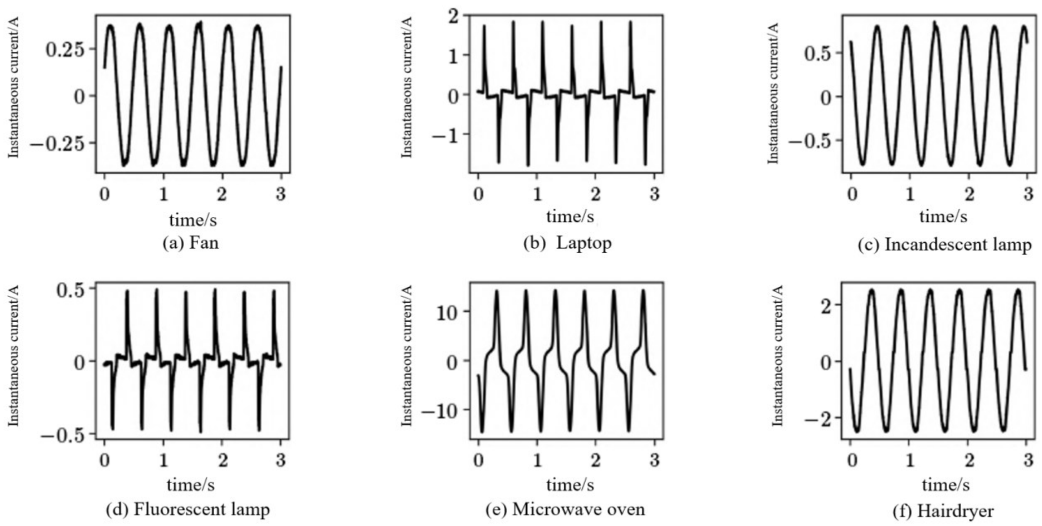

Figure 1. 6 types of household load steady-state current waveforms

2.3. Random forest load identification

Random forest is a set method composed of multiple decision trees. Decision tree is constructed with the principle of layer by layer analysis and logical reasoning, and it presents a kind of tree-like structure. In the decision tree, the sample feature is used as the root node, and the internal node determines the classification result by judging the feature. The leaf nodes represent the final classification result. The whole classification process is shown in Figure 2.

Each decision tree can be regarded as an independent classifier, and a random forest is a collection of these decision trees. The core idea is to comprehensively consider the classification results of all decision trees by means of ensemble learning, and select the category with the highest frequency as the final output.

In a random forest algorithm, we first train multiple different and independent decision trees. Each decision tree is trained using a different subset of samples and features to ensure diversity and generalization of the model. Then, in the test stage, for the samples to be classified, each decision tree will give its own classification judgment, and count the occurrence times of each category. Finally, the category that appears most frequently is assigned as the final output category for the entire random forest.

In this way, random forests combine the advantages of multiple decision trees with better robustness and accuracy. At the same time, since each decision tree is constructed independently, and samples and features are randomly selected for training, random forest also has the ability to prevent overfitting to a certain extent.

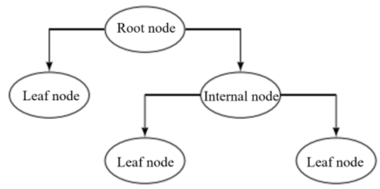

Figure 2. Flowchart of decision tree

3. Experimental analysis

3.1. Evaluation index

The evaluation index should focus on the overall classification accuracy and the classification effect of each type of equipment. In this paper, accuracy (A), precision (P), recall (R) and F1 score (F1-score) F1 were used to evaluate the model. Accuracy rate A represents the proportion of correctly classified samples in the number of samples in the test set; accuracy rate P represents the probability that when the label predicted by the algorithm is an electrical appliance, its actual label is the electrical appliance; recall rate R represents the probability that an electrical appliance sample is correctly identified by the algorithm as the electrical appliance; F1 score F1 is a comprehensive index combining accuracy rate P and recall rate R, and their specific calculation method is as follows:

$$A = (T + R)/N1 \tag{4}$$

$$P = T/(T + F) \tag{5}$$

$$R = T/(T + Q) \tag{6}$$

$$F1 = 2PR/(P + R) \tag{7}$$

Where: N1 represents the total number of test set samples; T represents the number of samples that are consistent between the electrical label obtained after model identification and the actual sample label; R indicates that the model is not identified as an electrical label, and the actual label is not the number of samples of the electrical appliance; F represents the number of samples whose identification label is an appliance but whose actual label is not. Q represents the number of samples where the identification label is not for an appliance but the actual label is for that appliance.

3.2. System overall scheme design

The non-invasive load monitoring system is divided into two parts: offline modeling and online monitoring.

3.2.1. Off-line modeling. A non-contact high frequency power data acquisition device is used to collect the steady-state current data of a single load during normal operation. Then the fundamental and odd harmonics of steady state current are obtained by fast Fourier transform and taken as load characteristics. The same equipment extracts multiple groups of samples to ensure the diversity of samples and improve the generalization degree of the model. All the samples are fed into the machine learning model and the model automatically adjusts and optimizes the parameters.

3.2.2. On-line monitoring. The non-contact high frequency power data acquisition device is installed at the bus. Event monitoring, current separation and load feature extraction are carried out on the data received by the server, and the extracted feature is input into the off-line model to identify the type of equipment.

4. Result analysis

In this experiment, high-frequency current data of PLAID dataset was used to verify the algorithm. After event detection, current separation and feature extraction, 816 sets of valid samples from 6 kinds of devices were obtained. 80% of the randomly scrambled samples are used as model training and 20% as model performance testing, and the sample-label pairs are input into different machine learning classification algorithms. The integrated development environment used in this experiment is PyCharm, and the machine learning algorithm in SciKIT-LEARN library is used for the experiment. The overall classification accuracy of K-nearest neighbor algorithm is 99.40%, the overall classification accuracy of multi-layer perceptron is 95.70%, and the overall classification accuracy of random forest algorithm is 98.80%. The classification effect of single device in the model is shown in Table 1.

Table 1. Classification effect of single equipment

Equipment	Algorithm	P	R	F1
Fluorescent lamp	KNN	1.00	1.00	1.00
	MLP	0.97	1.00	0.98
	RF	0.97	1.00	0.98
Fan	KNN	0.95	1.00	0.97
	MLP	0.80	0.89	0.84
	RF	0.95	1.00	0.97
Hairdryer	KNN	1.00	1.00	1.00
	MLP	1.00	1.00	1.00
	RF	1.00	1.00	1.00
Laptop	KNN	1.00	0.97	0.98
	MLP	1.00	0.94	0.97
	RF	1.00	0.97	0.98
	KNN	1.00	1.00	1.00
Microwave oven	MLP	1.00	1.00	1.00
	RF	1.00	1.00	1.00
Incandescent lamp	KNN	1.00	1.00	1.00
	MLP	0.93	0.90	0.92
	RF	1.00	0.97	0.98

It can be seen from Table 1 that F1 scores of single equipment are all high, which proves the effectiveness of taking harmonics as load characteristics in this experiment.

5. Conclusion

A non-intrusive load monitoring system based on a non-contact high-frequency synchronous power data acquisition device is designed in this paper. The hardware part uses non-contact high-frequency power data acquisition devices, including a non-contact current measurement module, an analog-to-digital

DOI: 10.54254/2755-2721/55/20241663

conversion module, a synchronous sampling signal generation module, a data processing module, and a communication module. In the software part, the non-parametric cumulative sum method is used for event detection, single load current extraction is carried out by parallel circuit theory, steady-state current fundamental wave and singular harmonic wave are used as load characteristics, and machine learning method is used to identify load. The validity of current fundamental wave and harmonic wave as load characteristics is proved by open data set. Finally, the whole scheme design of non-invasive load monitoring system is introduced, including off-line modeling and on-line monitoring.

The disadvantages of this example are: the classification effect is good on this data set, but the generalization ability on other data sets is unknown; the number of data samples is not enough. In the subsequent research, we can try to increase the sample data set by synthesizing a few classes or data enhancement methods to improve the generalization ability of the model.

References

[1] Zhao A J, Cui P F, Yu J Q, et al. Non-invasive household power load decomposition model based on self-attention [J]. Building science, 2023, 39 (04): 113-121+212.

[2] Zhang S Q, Wang W, Qian Y G, et al. Research progress of non-invasive load monitoring for deep learning [J]. Computer system application, 2023, 32 (03): 25-47.

[3] Liu H, Liu C Y , Zhao H R, et al. Non-invasive decomposition of electric and gas load of public buildings based on deep learning [J]. Power grid technology, 2023, 47 (03): 1188-1197.

[4] Luan W P, Wei Z, Liu B, et al. Test and evaluation method of non-invasive load monitoring algorithm [J]. Power grid technology, 2022, 46 (11): 4568-4579.

[5] Peng B G, Pan Z N, Yu T, et al. Graph data modeling and graph representation learning methods and their application to non-invasive load monitoring problems [J]. Proceedings of the CSEE, 2022, 42 (17): 6260-6274.

[6] Wang Y, Yang W, Xiao X Y, et al. A non-invasive residential load monitoring method based on fine identification of U-I trajectory curve [J]. Power grid technology, 2021, 45 (10): 4104-4113.