Research on non-intrusive load event detection algorithm

Xiaoli Huang [1] ,Dan Wang[1]
1. College of Electronic and Information, Xihua University
6607784@qq.com,1943443347@qq.com

***Abstract*—Event detection is an important step of non-intrusive load disaggregation. The accuracy of event detection directly affects the correct rate of load classification identification. Aiming at the shortcomings of existing event detection and feature extraction algorithms, using the steady-state active power as the feature, three improved event detection algorithms are proposed: improved CUSUM bilateral accumulation algorithm, improved sliding chi-square-GOF algorithm, and improved sliding Cepstrum analysis. The algorithm compares and analyzes the event detection accuracy of the three event detection methods. The experimental results show that the accuracy is better than most of the literature results, and the improved sliding chi-square-GOF algorithm can effectively detect large reference power events, avoiding missed detection of this situation.**

Keywords—Event detection;improved CUSUM bilateral accumulation algorithm;improved sliding chi-square-GOF event detection algorithm;improved sliding cepstrum event detection algorithm

I. INTRODUCTION

Non-intrusive load disaggregation and monitoring is an important technology of demand side management[1]. It directly collects the total electricity meter data to obtain the detailed electricity consumption of residents, and provides users with detailed electricity consumption, which is convenient for users to self-check and manage their own electrical behavior, thereby effectively reducing energy consumption. The basic framework includes data measurement[2], data processing, event detection[3],feature extraction[4],and load identification[4].Non-intrusive load disaggregation and monitoring are mainly divided into two categories: event-based non-intrusive load monitoring and non-intrusive load monitoring that is not event-based. The event-based non-intrusive load monitoring algorithm needs to identify the event occurrence time from the real-time change of the total power consumption data, and use the load feature extracted from the event occurrence time to match the classified characteristic load characteristic database to identify it. Non-event based load identification[5] first uses the marked data for disaggregation model training, and uses the trained model for load classification and recognition [6, 8].

Event detection is the first step of event-based non-intrusive load disaggregation, which is divided into difference threshold judgment and statistics-based methods and template matching methods [7, 9]. The first method is simple and efficient, but it is affected by noise and abnormal fluctuation points.

The statistical method parameters are assumed to be arbitrarily large and the accuracy is low. The third method needs to know the duration of each state. The accuracy of EMI based event detection is as high as 93.26%. The Hilbert transform[7] is used to extract the transient feature envelope for event detection and matching. Event detection can be attributed to the time-series change point detection problem. The change point detection problem belongs to the sequential detection problem in mathematical statistics. It belongs to the category of data analysis, and it plays an increasingly important role in many fields. The change point is the data point where the statistical law changes significantly. In this paper, the time corresponding to the feature change point is the time when the load state changes. Change point detection is divided into parametric and non-parametric ways, non-parametric methods mainly depends on the characteristics of the time series itself to determine whether the event occurs, such as CUSUM algorithm, GLRT, Cepstrum and chi-square test, but parameter settings have a greater impact on event detection accuracy. It can detect whether an event occurs by assuming that both sides belong to the same probability distribution. The Cepstrum[10] method uses homomorphic analysis to detect event occurrence. The parameter method needs to know a probability distribution in advance, so its application is limited.

For event-based non-intrusive load disaggregation and monitoring, accurate event detection is particularly important. This paper mainly studies how to accurately detect the event occurrence time from the total power consumption data. This paper selects the simple active power as the feature, and uses the UK-DALE[11] steady-state data as the data source to perform event detection. The improved sliding chi-square-GOF event detection method and the improved sliding cepstrum analysis event detection algorithm and the improved CUSUM bilateral accumulation algorithm are used to analyze the application of these three algorithms in event detection.

II. THE THREE IMPROVED EVENT DETECTION ALGORITHMS

The event detection algorithm needs to detect a situation in which the load power characteristics change significantly. In order to obtain the feature sequence, the feature sequence is obtained according to the sliding

978-1-7281-1907-6/19/$31.00 ©2019 IEEE

window model, and the basic schematic diagram is as follows:

$p_1, p_2, \cdots, p_n, p_{n+1}, p_{n+2}, \cdots, p_{2n}, \cdots, p_{k-1}, p_k$

$w(t_1)$ $\cdots$ $w(t_s)$

$w(t_1+1)$

Fig 1 Sliding window model

$w(t_1)$ indicates the power sequence of the sliding window at the t_1 moment, each sliding window extending a sequence value backward from the previous window. An event detection algorithm is then used to determine if a load change has occurred within each sliding window. The three event detection algorithms designed in this section are based on this sliding window model, but the improved sliding CUSUM bilateral accumulation algorithm and the improved sliding chi-square-GOF algorithm are divided into bilateral windows, and the remaining algorithms are based on single window.

A. Improved sliding CUSUM bilateral accumulation algorithm

The CUSUM algorithm belongs to the sequential analysis in the statistical algorithm. It accumulates the feature information in the sample data, accumulates the small deviation, and enlarges the variation characteristics of the sequence, so as to achieve the purpose of abnormal point detection. It divides each sliding window into two parts, the transient window and the mean window, the transient window is the front part, and the mean window is the back part. For power time series $P = \{p(1),...,p(k)\}_{k=1}^{k=\infty}$,define the mean calculation window W_m And transient calculation window W_d ,The length is m and n ,generally take $m=n$, M_m is the sequence mean within the mean window, M_d is the mean value in the transient window.

$$M_m = \frac{\sum_{k}^{k+m-1} p(k)}{m} \tag{1}$$

$$M_d = \frac{\sum_{k+m}^{k+m+n-1} p(k)}{n} \tag{2}$$

Separately defined g^+_k , g^-_k , they are used to determine if an event has occurred:

$$g^+_k = \max\{0, g^+_{k-1} + M_d - (M_m + \beta)\}, g^+_0 = 0 \tag{3}$$

$$g^-_k = \max\{0, g^-_{k-1} - M_d + (M_m - \beta)\}, g^-_0 = 0 \tag{4}$$

β indicates the level of fluctuation when the power time series is stable, M_d is the mean value of the power characteristics in the window for the transient state, M_m is the mean value of the power feature in the mean calculation window. If the load is switched on, $M_d > M_m + \beta$, g^+_k Start greater than zero, when the range of change is more than H , it is regarded as a positive event,also,use g^-_k to judge whether power drop is a negative event. Literature[12] uses the traditional CUSUM bilateral sliding detection algorithm,when $0 < g^+_k < H$,it is possible that the event will occur at this time , By introducing a delay factor $d = d+1$,slide backwards until $g^+_{k+d} > H$,An event was detected at this time,Put back d moment to get the moment of the event,.if g^+_{k+d} is not monotonous increase, it will be regarded as characteristic fluctuation. It will not continue to judge, but will be discarded directly. The process of the algorithm is complex, and for the data flow which fluctuates frequently, each time series point needs to be judged by delay, and the computational complexity is too large. The normal power fluctuation has been taken into account in the design of the β fluctuation factor. Therefore, an improved CUSUM bilateral cumulative sum algorithm for eliminating delay link is designed in this section, which not only reduces the data overlap with other devices, but also reduces the complexity of the algorithm.

B. Improved sliding chi-square-GOF algorithm

Chi-square test is a hypothesis test used to solve the statistical analysis of counting data. It does not make any assumptions about the overall distribution, so it belongs to non-parametric test. Chi-square test is a hypothesis test method. The basic principle of chi-square test is to compare the difference between two groups of samples to verify whether two groups of samples belong to the same distribution. Like bilateral CUSUM algorithm, it is based on sliding window:Front window: $W_1 = \{p_j \mid i \le j \le i+n\}$, Back window: $W_2 = \{q_k \mid i+n+1 \le k \le i+2n+1\}$, Assuming Front Window Power Data Sequence p_j ($i \le j \le i+n$) satisfies the condition of independent identical distribution and obeys the probability distribution G(p) ,Assuming Back Window Power Data Sequence q_k ($i+n+1 \le k \le i+2n+1$) satisfies the condition of independent identical distribution and obeys the probability distribution F(q).The distribution of power data sequence in front and back windows is unknown. Chi-square fitting test is to test the following two hypotheses:

$$\begin{aligned} H_1 &: G(p) \ne F(q) \\ H_0 &: G(p) = F(q) \end{aligned} \tag{5}$$

Chi-square hypothesis test statistics：

$$l_{GOF} = \sum_{i=1}^{n} \frac{(q_i - p_i)^2}{p_i} \tag{6}$$

If $l_{GOF} > \chi^2_{\alpha,n-1}$,The hypothesis H_0 is rejected at the 100(1-α)% level, At this point load events occur. However, this method has serious defects in the process of event detection, that is, when the reference load is too large compared with the load power change value, the event is likely not to be monitored, because the statistics become too small at this time, resulting in the hypothesis test quantity falling into the rejection domain, and the event cannot be detected. To solve this problem, the paper [8] proposes a chi-square fitness test based on voting mechanism. Its basic principle is to find the maximum value in each window by sliding GOF sequence. If the maximum value is obtained, a vote will be obtained. Finally, the window exceeding the threshold number of votes will be regarded as an event in the window. The parameters training and selection of this method are complex, and the test results depend on the parameters. The essential reason why events cannot be detected is that the reference load changes too much compared with the load power. Since the denominator reference power is too large, we can effectively solve the problem by improving the statistical formula as follows:

$$l_{\mathrm{GOF}} = \sum_{i=1}^{n} \frac{(q_i - p_i)^2}{\alpha \times \mathrm{abs}(q_i - p_i)} \tag{7}$$

In order to amplify multiple, the original statistics use the reference load as denominator. When the reference load is too large, the statistics will become very small. Therefore, through the above improvement, setting the appropriate amplification multiple can make the statistics meet the needs of event detection. Experiments with MATLAB show that the improved algorithm can detect high reference power event detection problems, and can be enlarged or reduced appropriately according to the change range of event detection power, so the method is simpler and more efficient.

C. *Improved sliding Cepstrum algorithm*

This algorithm is cepstrum analysis in signal processing. It is widely used in the separation of input characteristics and system characteristics. It is typically used in speech signal processing, etc. Compared with the previous two detection algorithms based on bilateral sliding windows, Cepstrum algorithm uses single-sided window detection.For sliding window signal sequence: $X=(p_{i+1},\dots p_{i+n})$，In order to detect events from each sliding window, the time domain signal is first transformed into the frequency domain signal：

$X[k]=\sum_{\mathrm{j}=1}^{n} x[\mathrm{j}]\mathrm{e}^{-\frac{2\pi\ \mathrm{i}k}{n}\mathrm{j}}, 0 \le k < \mathrm{N}$，Then the inverse Fourier transform of logarithmic spectrum is used to transform the signal into homomorphic domain：

$c(n)=\frac{1}{\mathrm{N}}\sum_{\mathrm{k=0}}^{\mathrm{N-1}}\log_{10}(|X[k]|)\mathrm{e}^{\frac{2\pi\ \mathrm{i}k}{\mathrm{N}}n}, 0 \le n < \mathrm{N}$，then transform again: $\hat{X}[k]=\sum_{\mathrm{j}=1}^{n}\mathrm{c}(\mathrm{j})\mathrm{e}^{-\frac{2\pi\ \mathrm{i}k}{n}\mathrm{j}}, 0 \le k < n$,Convert to decibel level: $\hat{X}_{\mathrm{dB}}[k]=20\log_{10}(\hat{X}[k])$.

$$\min_{0\le k<\mathrm{N}}(\hat{X}_{\mathrm{dB}}[k]) > \tau \tag{8}$$

Formula (8) has a negative decibel after calculation. Therefore, it is equivalent to directly process the original homomorphic signal and compare the absolute value of the decibel signal with the threshold value. Formula (8) is changed as follows:

$$\min_{0\le k<\mathrm{N}}(\mathrm{abs}(\hat{X}_{\mathrm{dB}}[k])) > \tau \tag{9}$$

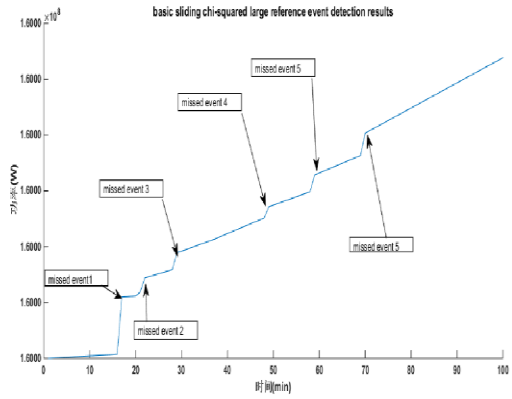
III. THE EXPERIMENT RESULTS

Because the sampling rate of power sequence data in this paper is low and the sliding window length is set to 3, most of the sliding window length can be fully included in the load state change process, so the window length of the event detection algorithm in this section is chosen as $m=n=3$, $\beta=15\mathrm{W}$, $H=20\mathrm{W}$; For improved sliding chi-square-GOF algorithm, $\alpha=200$,the significant level $\alpha=0.05$; For improved sliding Cepstrum algorithm, threshold $\tau=2.5$.

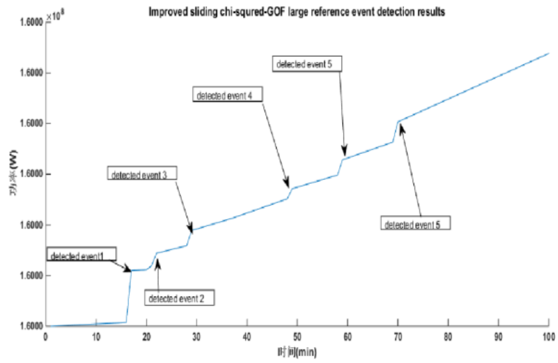
Table. 1 Improved Sliding CUSUM Bilateral Cumulative and Algorithmic Event Detection results

appliances	the accuracy of Improved CUSUM	the accuracy of Chi-squared GOF	the accuracy of Improved Cepstrum
Heater	0.89	0.93	0.83
solar pump	0.89	0.94	0.89
PC	0.73	0.6	0.87
Washing machine	0.95	1	1
Dishwasher	0.76	0.92	0.92
TV	0.94	0.94	0.94
Kitchen lamp	0.97	1	1
Home Cinema Computer	0.90	0.95	1
Kettle	1	0.96	0.96
Oven	0.97	0.97	1
Refrigerator	0.96	1	1
Microwave Oven	1	1	0.95
LCD	0.86	0.96	0.96
Toaster	0.63	0.68	0.95
Living room lamps	0.78	1	1
Total accuracy rate	0.88	0.92	0.95

Further design experiments verify the improved detection effect of the sliding chi-square-GOF algorithm on large reference power events. The comparison of the test results of high-power events is as follows:



(a)



(b)

Fig. 2 Detection results of high reference power events for each algorithm

As shown in Figure 2, a is the basic sliding chi-square-GOF high-reference power event detection result, b is the improved sliding chi-square-GOF high-reference power event detection result, which proves that the improved sliding chi-square-GOF algorithm can effectively detect high-reference power events.

IV. CONCLUSIONS

In summary, all three event detection algorithms can detect load events better. The event detection accuracy of various devices of various algorithms is higher than 0.60, and the total correct rate of the three improved algorithms is above 0.88. Among them, the improved sliding Cepstrum algorithm has the highest event accuracy, and the improved sliding CUSUM accumulation algorithm accuracy. The rate is the lowest. And the improved sliding chi-square-GOF algorithm can effectively detect large reference power events. Compared with the literature [7], the improved motion detection accuracy of the sliding chi-square-GOF algorithm is higher than the adaptive GOF algorithm proposed in the literature about 0.1, but it only performs algorithm verification on 10 single-load events. The UK-DALE contain tens of thousands of the total power characteristics of experimental data with simultaneous load events. At the same time, the accuracy of the three event detection algorithms is superior to the literature [12].

ACKNOWLEDGMENT

This work was supported by the Xihua University Graduate Innovation Fund (Project No.: ycjj2017061); Ministry of Education "Chunhui Project" (Project No.: Z2011089)

REFERENCES

[1] S.Darby, "The effectiveness of feedback on energy consumption",A Review for DEFRA of the Literature on Metering, Billing and direct Displays, vol.486,2006.

[2] Hart GW. Non-intrusive appliance load monitoring. Proc IEEE 1992;80:1870–91.

[3] Xiaofan J,Dawson-Haggerty S,Dutta P,Culler D.Design and implementation of a high-fidelity AC metering network.In:Proceedings of the international conference on information processing in sensor networks IPSN;2009.p.253–64.

[4] Hosseini S S, Agbossou K, Kelouwani S, et al. Non-intrusive load monitoring through home energy management systems: A comprehensive review[J].Renewable & Sustainable Energy Reviews, 2017, 79:1266-1274.

[5] Jin Y,Tebekaemi E,Berges M, et al.Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities[C],IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2011:4340-4343.

[6] Cominola A,Giuliani M,Piga D,et al.A Hybrid Signature-based Iterative Disaggregation algorithm for non-intrusive load monitoring[J]. Applied Energy, 2017, 185:331-344.

[7] Alcalá J, Ureña J, Álvaro Hernández, et al. Event-Based Energy Disaggregation Algorithm for Activity Monitoring From a Single-Point Sensor[J]. IEEE Transactions on Instrumentation & Measurement, 2017, PP(99):1-12.

[8] Yang C C, Soh C S, Yap V V. A non-intrusive appliance load monitoring for efficient energy consumption based on Naive Bayes classifier[J]. Sustainable Computing Informatics & Systems, 2017, 14.

[9] Jin Y. A time-frequency approach for event detection in non-intrusive load monitoring[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2011, 8050(4):80501U-80501U-13.

[10] Kong S, Kim Y, Ko R, et al. Home appliance load disaggregation using cepstrum-smoothing-based method[J]. IEEE Transactions on Consumer Electronics,2015, 61(1):24-30.

[11] Kelly J, Knottenbelt W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes[J]. 2015, 2.

[12] Yuan Ling. Real-time power load statistics and event detection based on spark [D].Huazhong University of Science and Technology, 2015.