International Journal of Recent Advances in Multidisciplinary Topics
Volume 6, Issue 10, October 2025
www.ijramt.com | E-ISSN: 2582-7839 | RESAIM Publishing (www.resaim.com)



# Robust Random Cut Forest Based Event Detection Method for Non-Intrusive Load Monitoring

Il Myong To[1], Chol Jun Han[2*]
*[1,2]Robotics Institute, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea*

***Abstract*: Non-Intrusive Load Monitoring (NILM) can provide the appliance-level power consumption information without deploying submeters in each load, in which load event detection is one of the crucial steps. However, the existing event detection methods are not efficient to detect both starting time of event (STE) and ending time of event (ETE), and the scalability is limited. To tackle these problems, this paper has proposed an event detection method based on robust random cut forest (RRCF) that is an unsupervised learning algorithm for detecting anomalous data points within a dataset. It has been validated on REDD dataset. The experimental results show that the proposed event detection method performs better than state-of-the-art.**

***Keywords*: Non-intrusive load monitoring, Event detection, Robust random cut forest, Adaptive threshold.**

## 1. Introduction

The feedback on energy usage information can provide electricity customers a basis to better control their electricity utilization behaviors and ultimately save energy [1].

Non-intrusive load monitoring (NILM), which was initially proposed by Hart [2], can provide the individual electrical load usage information without deploying submeters in each load. According to the acquired information from NILM, the electricity consumers can know and adjust their electricity utilization behavior to achieve energy saving.

NILM is mainly divided into two categories, namely, non-event-based NILM and event-based NILM. In NILM study, a switch action or change in the working state of a load is called an event. The event-based method could have a better performance than non-event-based NILM method because of the acquired rich load features that can greatly enhance the identification accuracy [3].

The main purpose of event detection in NILM is to detect the starting time of event (STE) and ending time of event (ETE) when state transition occurs from the aggregated measurements [4]. Accurate event detection is the prerequisite for precise NILM [5]. For this reason, many scientists have carried out in-depth investigations in this field. According to the implementation principles, the existing event detection methods can be mainly divided into three categories: expert heuristics, probabilistic model, and pattern matching.

Expert heuristic methods mainly utilize the common-sense knowledge and propose a set of decision rules for event detection.

Probabilistic methods are based on the statistical probability distribution of aggregated load data to detect the change after an event occurs. Typical probabilistic methods include likelihood ratio and cumulative sum (CUSUM).

The pattern matching methods detect events by matching the sequence fragments corresponding to the event transient process with the known feature library.

Although the existing approaches have obtained a series of achievements, they are still bearing from some inadequacies. Firstly, most of the existing methods only focused on detecting the STE while ignoring the ETE [3], [5]. Though some methods have considered detecting both STE and ETE [11], they are only applicable to the specific long-transient events and cannot be generalized to other types of events. Yet the STE and ETE are very crucial for subsequent load identification, as most of the event-based NILM methods rely on extracting load features.

To cope with these problems, this paper has proposed an event detection method based on robust random cut forest (RRCF) algorithm which can improve the accuracy and scalability of state-of-the-art.

The main contributions of this work are as follows.

- The proposed method can detect both STE and ETE with high accuracy. Thus, it can provide a good foundation for the subsequent load identification.
- The proposed method has high sensitivity in the scenario with high sampling rate, which means it can detect events that occur within a short period of time.
- The proposed method has high scalability and practicality. In both scenarios with high and low sampling rates, it performs well and can be implemented in real time because of its low computational complexity.

## 2. Introduction of Robust Random Cut Forest Algorithm

This section will introduce the principle of RRCF outlier detection algorithm that lays the foundation for event detection.

### *A. Principle of RRCF Algorithm*

RRCF [4] is an outlier detection algorithm for dynamic data streams. It has been applied in various scenarios.

The first step of the RRCF algorithm is to create a random

*Corresponding author: hcj901128@star-co.net.kp

forest of trees where each tree is obtained by using a partition of the sample dataThis section will introduce the principle of RRCF outlier detection algorithm that lays the foundation for event detection. In which the anomaly score is defined as the expected change in complexity of the tree as a result of adding or removing that data point to the tree. The random cut forest assigns an anomaly score by computing the average score from each constituent tree and scaling the result with respect to the sample size.

Anomaly score can manifest during unexpected spikes in time series data, breaks in periodicity, or unclassifiable data points.

### *B. Calculation of the Anomaly Socre*

The procedure for calculating an anomaly score is as follows. Given a set of points and a point . Let be the depth of in tree. Consider now the tree produced by deleting as. Let the depth of y in be. Fig. 1 shows the tree before and after deleting one data point, Fig. 1 (a) is the tree before deleting, and Fig. 1 (b) is the tree after deleting. The anomaly score for data point is calculated as,

$$Score(x,Z) = \sum_{y \in (Z-\{x\})} (f(y,Z,T) - f(y,Z-\{x\},T)) \quad (1)$$

In the tree, the depth of an anomalous data point is usually much shallower than the depth of a normal data point. Thus, when the anomalous data points are added or deleted, the anomaly score will be larger. Therefore, a low anomaly score means that the corresponding data point is “normal” and a high anomaly score means that the corresponding data point is “anomalous”.

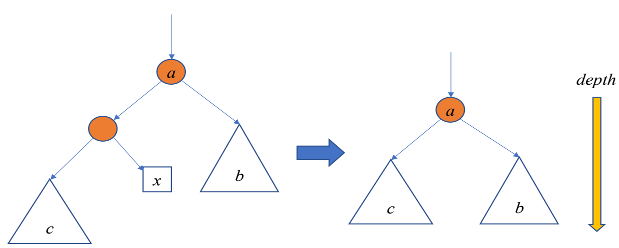


Fig. 1. Deleting one data point from tree

## 3. Principle of Detection Method Based on Robust Random Cut Forest

This section will introduce the procedure of the proposed event detection method, the principle of the RRCF based preliminary detection, details of post processing, and finally the complete algorithm.

### *A. Calculation of the Anomaly Score*

The flow chart of the proposed method is shown in Fig. 2. The functions of each module are as follows.

*1) Data Preprocessing*

The mean-pooling preprocess is applied for the aggregated power time series with high sampling rate to eliminate the fluctuation, while unnecessary for those with low sampling rate. Then, the power differential series are obtained by calculation.

*2) RRCF Based Preliminary Detection*

The anomaly score of each data point in the power differential series is calculated and the possible event is preliminarily detected.

*3) Post-Processing*

When possible event is detected, the misidentification event is further inhibited by power difference threshold. Then the STE and ETE are finely adjusted by linear fitting.

*4) Adaptive Power Difference Threshold Updating*

The standard deviation of the aggregated power data is calculated at each moment to update the adaptive power difference threshold value.

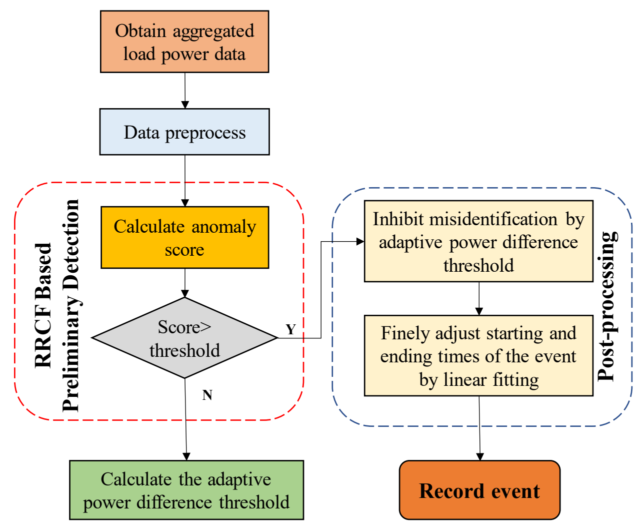


Fig. 2. Flow chart of the proposed method

### *B. Preliminary Detection Based on RRCF Algorithm*

Firstly, for data with high sampling rate, the mean pooling process is applied to eliminate the fluctuation and to reduce frequency; but for data with low sampling rate, the mean pooling is not necessary.

Fig. 3 (a) shows the aggregated power time series consisting of two events. When the event occurs, the power value becomes outlier relative to the power values in the previous steady state. Hence, these data points can be detected as outlier by the RRCF algorithm.

We use the power differential series as shown in Fig. 3 (b) to calculate the anomaly score. Based on the aggregated power time series with the length of, the power differential series is obtained as,

$$\Delta P(t) = P(t+1) - P(t) \quad (2)$$

As can be seen from Fig. 3 (b), during the steady state, the power differential series are very close to 0. When an event occurs, the power difference value has changed suddenly and becomes an outlier. Thus, we can use RRCF algorithm to detect event based on the power differential series.

First, the aggregated power time series are obtained in real time and the anomaly score is calculated for each point in the power differential series. The threshold of anomaly score is set in advance and compare with the anomaly scores. When the current anomaly score is greater than the threshold, it means that the power difference value has changed, indicating that an event may occur.

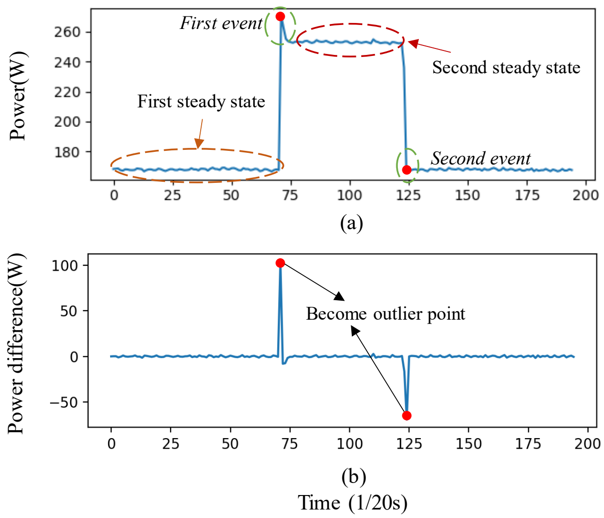


Fig. 3. Data series acquisition
(a) The aggregated load power series, (b) Power differential series

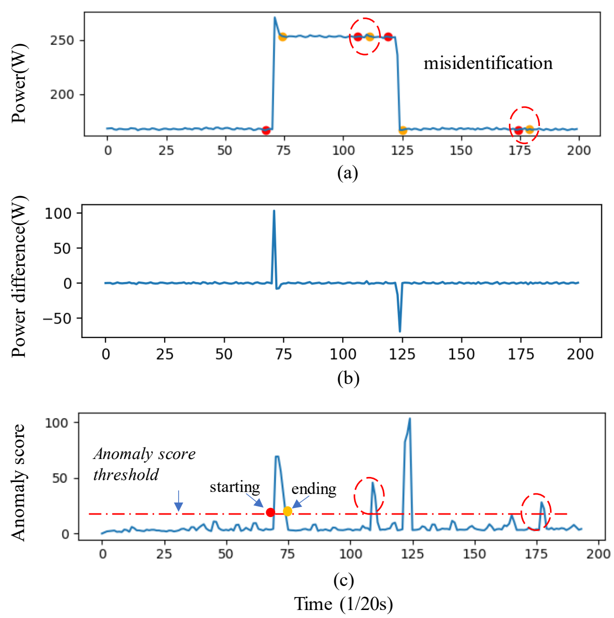


Fig. 4. The event detection result for Fig.3.
(a) The aggregated load power series, (b) Anomaly score, (c) Anomaly score

When the anomaly score converges to threshold, it can be preliminarily deduced that the event is finished.

When the threshold of the anomaly score is set to 20, the event detection results for Fig. 3 are shown in Fig. 4. The red circle indicates STE, and the orange circle indicates the ETE. As can be seen from Fig. 4, in addition to the two real events, there are also two falsely-detected events caused by power fluctuation. Therefore, it is necessary to further determine whether the event is reasonable or unreasonable by post-processing.

### *C. Post -Processing*

The main purpose of post-processing is to eliminate misidentifications caused by power fluctuation and accurately locate the STE and ETE.

#### *1) Inhibition of the Misidentification Based on Adaptive Power Difference Threshold*

In this paper, we use the standard deviation of the aggregated power signal in steady state to adjust the threshold, as shown below,

$$\Delta P_{thr} = max\ (\Delta P_0, sd + \Delta P_0 \cdot arctan\ (\frac{sd}{\Delta P_0}) \cdot \frac{4}{\pi}) \tag{3}$$

Where $\Delta P_{thr}$ is the adaptive power difference threshold, is the preset threshold in the zero standard deviation and is the standard deviation of the steady state before the latest event.

The value of is defined by the electric customer. When the absolute value of the power difference is greater than the threshold, the event is considered to be true; otherwise, it is considered to be false and can be eliminated.

#### *2) Adjustment of the STE and ETE*

The purpose of event detection is to extract the features of the load causing the event, thus it is very important to accurately locate the STE and ETE. However, the STE and ETE found by RRCF algorithm are only the starting and ending times of outlier in the aggregated power time series, as shown in Fig. 6, which are not accurate enough. We still need to find the accurate STE and ETE in the steady state.

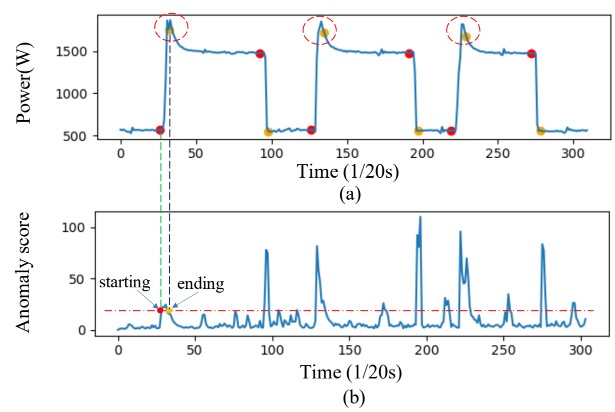


Fig. 5. The inaccurate ETE

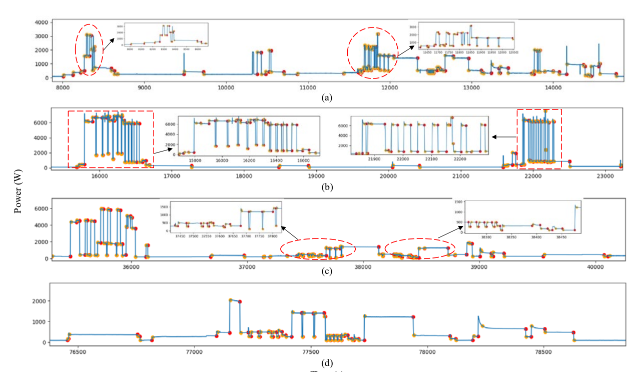

Fig. 6. Event detection results on REDD dataset

Table 1
Compare with other methods

| Literatures | Aggregated signal | Sampling rate | Phase | Precision (%) | Recall (%) | F1 score (%) |
|---|---|---|---|---|---|---|
| Proposed method | Active power | 20 Hz | Phase A | 99.40 | 100.00 | 99.70 |
| | | | Phase B | 92.60 | 95.50 | 94.03 |
| | | 1 Hz | Phase A | 99.60 | 99.40 | 99.50 |
| | | | Phase B | 94.60 | 74.20 | 83.17 |
| [12] | Active power | 60 Hz | Phase A | 99.20 | 99.20 | 99.20 |
| [14] | Active power | 1 Hz | Phase A | 99.30 | 81.01 | 89.22 |
| | | | Phase B | 77.72 | 57.04 | 65.79 |
| [26] | Active power | 1 Hz | Phase A | 98.87 | 100.00 | 99.43 |
| | | | Phase B | 79.98 | 92.55 | 85.81 |

As can be seen from Fig. 5, the ETE detected by RRCT algorithm is slightly earlier, that is, the event is considered to have ended before the steady state is completely reached. In order to find the accurate STE and ETE, we select several data points near the starting and ending points detected by the RRCF, and then carry out linear fitting to calculate the slope and goodness of fit, as shown in Fig. 7.

## 4. Case Study

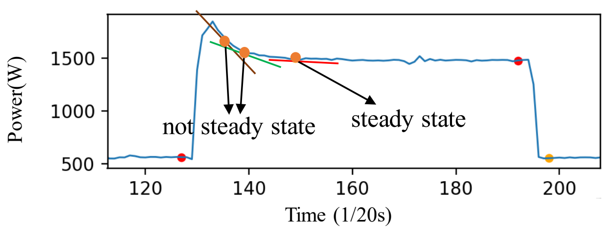


Fig. 7. The principle of the accurate locating of the event

### *A. Validation on the REDD Dataset*

REDD dataset provides the aggregated load power series recorded at 1 Hz and collected from six real houses. Considering the complexity of the appliance composition, we select the aggregated load power series for about one day from House 1 to validate the proposed method.

The event detection results on REDD dataset are shown in Fig. 7. As can be seen from Fig. 14, the proposed method performs well in the scenario with low sampling rate. REDD dataset also has a lot of repetitive events, as shown in Fig. 6 (b)-(d).

The validation result in REDD dataset shows that the proposed event detection method can performance well in different household, which means it has high practicality.

### *B. Comparison with Other Methods*

Event detection results of the proposed method are compared with other methods in the literatures, as shown in Table I.

From the comparison results, it can be seen that the proposed method performs better than state-of-the-art algorithms. The event detection methods proposed in [3] and [5] only focus on the STE, but not on the ETE. The event detection methods proposed in [11] try to detect whole transition process of the event, but in those methods another event is not allowed to occur during long-transient event detection, which is not controllable in reality.

## 5. Conclusion

This paper has proposed an event detection method based on RRCF algorithm. The power differential series is input into the random forest and the anomaly score is calculated for each data point to roughly detect the STE and ETE. Then the post processing is carried out to inhibit the misidentifications by using adaptive power difference threshold and accurately locate the STE and ETE.

The proposed method is validated on the REDD dataset.

This study provides a practical foundation for NILM problems. In the next research, we will combine the proposed event detection method to explore the NILM system, which plays a significant role in improving the performance of energy-efficiency systems and promoting sustainability in buildings.

## References

[1] D. Hua, F. Huang, L. Wang, and W. Chen, “Simultaneous disaggregation of multiple appliances based on non-intrusive load monitoring,” Electr. Power Syst. Res., vol. 193, p. 106887, Apr. 2021.

[2] G. W. Hart, “Nonintrusive appliance load monitoring,” Proc. IEEE, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[3] L. Yan, W. Tian, H. Wang, X. Hao, and Z. Li, “Robust event detection for residential load disaggregation,” Appl. Energy, vol. 331, p. 120339, Feb. 2023.

[4] A. U. Rehman, T. T. Lie, B. Vallès, and S. R. Tito, “Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low Complexity Statistical Features,” IEEE Trans. Instrum. Meas., vol. 69, no. 3, pp. 751–759, Mar. 2020.

[5] M. Lu and Z. Li, “A Hybrid Event Detection Approach for Non-Intrusive Load Monitoring,” IEEE Trans. Smart Grid, vol. 11, no. 1, pp. 528–540, Jan. 2020.

[6] A. Yasin and S. A. Khan, “Unsupervised Event Detection and On-Off Pairing Approach Applied to NILM,” in 2018 International Conference on Frontiers of Information Technology (FIT), Dec. 2018, pp. 123–128.

[7] B. Völker, P. M. Scholl, and B. Becker, “Semi-Automatic Generation and Labeling of Training Data for Non-intrusive Load Monitoring,” in Proceedings of the Tenth ACM International Conference on Future Energy Systems, New York, NY, USA, Jun. 2019, pp. 17–23.

[8] F. Ciancetta, G. Bucci, E. Fiorucci, S. Mari, and A. Fioravanti, “A New Convolutional Neural Network-Based System for NILM Applications,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–12, 2021.

[9] A. E. Lazzaretti et al., “A Multi-Agent NILM Architecture for Event Detection and Load Classification,” Energies, vol. 13, no. 17, Art. no. 17, Jan. 2020.

[10] S. Houidi, F. Auger, H. B. A. Sethom, D. Fourer, and L. Miegeville, “Multivariate event detection methods for non-intrusive load monitoring in smart homes and residential buildings,” Energy Build., vol. 208, p. 109624, Feb. 2020.

[11] W. Luan, Z. Liu, B. Liu, Y. Yu, and Y. Hou, “An Adaptive Two-Stage Load Event Detection Method for Nonintrusive Load Monitoring,” IEEE Trans. Instrum. Meas., vol. 71, pp. 1–14, 2022.

[12] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research,” in Proc. Workshop Data Min. Appl. Sustain. (SIGKDD), San Diego, CA, USA, 2011, pp. 1–6.

[13] S. Guha, N. Mishra, G. Roy, and O. Schrijvers, “Robust Random Cut Forest Based Anomaly Detection On Streams,” in International Conference on Machine Learning, Vol 48, San Diego, 2016, vol. 48.