Front. Energy 2015, 9(2): 231–237
DOI 10.1007/s11708-015-0358-6

RESEARCH ARTICLE

Chuan Choong YANG, Chit Siang SOH, Vooi Voon YAP

A systematic approach to ON-OFF event detection and clustering analysis of non-intrusive appliance load monitoring

Abstract The aim of non-intrusive appliance load monitoring (NIALM) is to disaggregate the energy consumption of individual electrical appliances from total power consumption utilizing non-intrusive methods. In this paper, a systematic approach to ON-OFF event detection and clustering analysis for NIALM were presented. From the aggregate power consumption data set, the data are passed through median filtering to reduce noise and prepared for the event detection algorithm. The event detection algorithm is to determine the switching of ON and OFF status of electrical appliances. The goodness-of-fit (GOF) methodology is the event detection algorithm implemented. After event detection, the events detected were paired into ON-OFF pairing appliances. The results from the ON-OFF pairing algorithm were further clustered in groups utilizing the *K*-means clustering analysis. The *K*-means clustering were implemented as an unsupervised learning methodology for the clustering analysis. The novelty of this paper is the determination of the time duration an electrical appliance is turned ON through combination of event detection, ON-OFF pairing and *K*-means clustering. The results of the algorithm implementation were discussed and ideas on future work were also proposed.

Keywords non-intrusive appliance load monitoring, event detection, goodness-of-fit (GOF), *K*-means clustering, ON-OFF pairing

1 Introduction

Non-intrusive appliance load monitoring (NIALM) disaggregates the total power consumption into individual electrical appliances, which is switched ON and OFF [1]. From the disaggregated data, the energy usage of individual electrical appliance can be determined [1,2].

The NIALM methodology is implemented according to the operational model states of the electrical appliances [1,3,4]. The appliances model states are two states of operation — either switched ON or OFF; finite state machine (FSM) — repeatable switching pattern with finite number of operating states, for example refrigerator, hair dryer and cooker; continuously variable — infinite number of operating states, for example induction motor, fluorescent lamp and electric vehicle battery; and permanent consumer devices — another category of appliances which remain in operation throughout the day or weeks and consumed energy constantly, for example home alarm system and smoke detector [3,4].

To analyze the signature from the NIALM methodology, the appliance signatures need to be known. An appliance signature is the quantifiable value of the total load that provides information on the operating state of the individual appliance in the load [1]. The appliance signatures are categorized into steady-state signatures, transient signatures and others type of signatures [1,3–5].

The NIALM methodology utilizes steady-state signature analysis by obtaining steady-state features during the operation of the electrical appliances [1]. To study NIALM, various steady-state methods have been proposed including power change [1,4,6], time and frequency domain characteristics of *V-I* waveforms [4,7,8], *V-I* trajectory [4,9,10], and steady-state voltage noise [4,11,12].

In the power change method, the features are the real power and the reactive power in NIALM for tracking the switching ON or OFF operation of the electrical appliances [1]. The advantage of this method is that it can identify high power load appliances easily. Moreover, it also requires low sampling rate. However, the low power appliances will be difficult to differentiate due to the overlapping of real and reactive power in the plot of reactive power versus the real power feature space [4,6].

To overcome the challenges for the low power

Received August 25, 2014; accepted November 22, 2014

Chuan Choong YANG (✉), Chit Siang SOH, Vooi Voon YAP
Faculty of Engineering and Green Technology, University Tunku Abdul Rahman, Jalan University, Bandar Barat, Kampar 31900, Perak, Malaysia
E-mail: yangcc@utar.edu.my

appliances in the power change method, the time and frequency domain characteristics of voltage (V), current (I) and waveforms have been proposed [4,7,8]. With this analysis, special appliance features, namely, the peak and root mean square current and voltage values, the phase differences and power factor information are utilized to determine the status of an appliance. Therefore, the resistive or inductive load of an appliance can be easily identified. However, this method requires high sampling rate and is not able to separate overlapping of switch ON events.

Another method is to perform a V-I trajectory to group electrical appliances [4,9,10]. Each of the electrical appliance will have its unique V-I trajectory plotted by utilizing the normalized current and voltage values. The shape features of a V-I trajectory and the distinctive V-I curves allow a detailed classification of electrical appliances. However, this method requires intensive computational power and is unable to identify the distinctive V-I curve for smaller electrical loads.

Another approach has been proposed in the steady-state method which is the steady-state voltage noise [4,11,12]. This method analyzes the steady-state voltage noise produced once the electrical appliance starts operation. In this method, motor-based appliances are easier to be identified because these appliances produce synchronous voltage noise. Besides, the identification of simultaneous switching events can be detected with higher accuracy. However, this method requires additional hardware for measurement and it is sensitive to the wiring structure of the appliance and environment.

In transient signature, a small variation, either in power or current, is linked with the switching ON of appliances that provide signature for researchers to study [5]. Researcher discovered that the transient signature is linked to the nature of the electrical appliances and this can be utilized as the signature of the appliance [13–15]. The transient signature analysis can be categorized into transient power [4,7,14,16,17], start-up current transients [4,6,16] and high frequency sampling of voltage noise [4,11,18].

The novelty of this paper is the determination of the time duration an electrical appliance is turned ON through combination of event detection, ON-OFF pairing and K-means clustering. This contribution is simple and intuitive and yet extremely useful for user.

2 Methodology

The objective of this paper is to describe the approach of ON-OFF event detection and clustering analysis in a systematic manner for non-intrusive appliance load monitoring (NIALM). To monitor the efficient usage of electricity, the NIALM method was utilized.

In the algorithm, the event-based methodology was used [19]. The first step was the power measurement. The power measurement data were obtained from the REDD data set [20]. The REDD data set contains two columns of data type, the coordinated universal time (UTC) timestamps as integers and power readings in apparent power (VA) of the appliances on the circuit. This was followed by median filtering. This step was conducted to remove noise from the data. The outcome of filtering was a smoother data with less impulsive noise, which allowed easier identification of events in the subsequent stage of processing. The median filtering was a two stage filtering with different window sizes. The first stage was to improve the signal to noise ratio (SNR). In the second stage, the filter exploited the nonlinearity of the median filter to remove impulsive noise [21].

The next task in the NIALM methodology was the event detection, whose purpose was to determine the time when an appliance changed state (for example switched ON or OFF). In this paper, the ON-OFF based approach [1,22] and the goodness-of-fit (GOF) methodology [19,23] was combined to detect events in the total power consumption data. The ON-OFF pairing approach was based on the grouping of ON-OFF pairs of each of the appliances on the circuit. The GOF methodology was based on the χ^2 statistical test. It was used to compare two populations of data samples. In event detection, if the first data sample came from pre-event and the second data sample originated from post event, then there should be a quantifiable difference between the two samples. This difference could be evaluated using the χ^2 statistical test. The reason for adopting this method is its simplicity and improved performance as reported in other studies [2,23].

After the completion of the event detection, cluster analysis was performed on the detected events to group similar events into the same cluster. Detected events consisted of pairs of power consumption when an appliance was turned ON (positive value) and turned OFF (negative value). Detected events from the same appliance shared similar positive and negative power consumptions. To group these events, K-means clustering was performed on the pairs of positives and negatives power consumption values. Figure 1 shows the flowchart of the methodology implementation.

3 Algorithm

3.1 Appliance signature data

In this paper, the REDD data set was utilized and the data used were labeled as house 1 which have 20 monitoring points consisting of whole home monitoring (2 mains) and various other electrical appliances. The electrical appliances consist of electronics, lighting, refrigerator, disposal, dishwasher, furnace, washer dryer, smoke alarms, bathroom, kitchen outlets and microwave. The data set consists

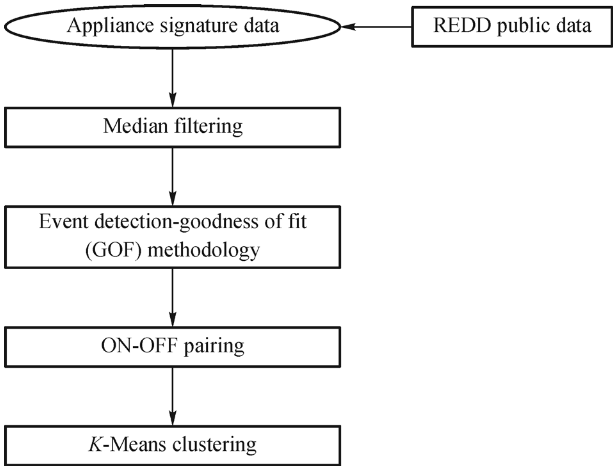

Fig. 1 Flowchart of methodology implementation

of the total apparent power consumption in unit Volt-amp (VA) recorded every 1 s.

3.2 Median filtering

After the data set from house 1 was loaded into the MATLAB folder, median filtering was performed twice in the algorithm. The first median filtering algorithm was used to reduce the noise, while the second one was used to smoothen the signal for processing in the subsequent stages. The data used in the filtering process are up to 100000 data points.

3.3 Event detection

The event detection methodology used in the algorithm was the GOF methodology [19,23]. As the data set used was from the continuous power data, the data was separated into n-samples of frames. The frames could be identified as pre- and post- event frames. In this paper, the event detection algorithm was to detect the switching "ON" or "OFF" of electrical appliances. Besides, the algorithm would determine the time this switching "ON" or "OFF" occurred.

The GOF method aimed to establish that for some available probability distribution, a set of data could be derived from it. The GOF test utilized the chi-square test or χ^2 test [23]. Equation (1) was implemented using MATLAB.

$$l_{\mathrm{GOF}} = \sum_{i=1}^{n} \frac{(y_i - x_i)^2}{x_i}. \tag{1}$$

For the calculation, $l_{\mathrm{GOF}} = \chi^2$ was equated. The formula of l_{GOF} used was in Eq. (1). The data sample x_i was the sample from the pre-event window and y_i was the data sample from the post-event window. A total number of n data samples in the pre-event window were compared with a total number of n data samples in post-event window to determine the amount of deviation of both population of data sample. Equation (1) quantified this deviation. If an event took place, the value of l_{GOF} was expected to be high.

For comparison with the calculated value l_{GOF}, the chi-square distribution table, $\chi^2 = \chi^2_{\alpha,\mathrm{d}f}$, value for alpha, α = 0.01 and 0.05 were used. For every calculated chi-square test, if $l_{\mathrm{GOF}} = \chi^2_{\alpha,\mathrm{d}f}$, an event was detected. From the event detected, the total sum of false positive was calculated from Eq. (2) [19].

$$\Delta P_e = \frac{1}{w_3} \sum_{i=e+w_2+1}^{e+w_2+w_3} P(i) - \frac{1}{w_1} \sum_{i=e-w_1}^{e-1} P(i). \tag{2}$$

where the term w_1, w_2 and w_3 are the window lengths of the signal, w_1 and w_3 are used to determine the pre- and post-events means, and w_2 is used to allow a delay for the transient before it reaches steady-state.

The total power for the false positives ΔP_{FPS} is shown in Eq. (3) [19]. To determine the total false positive, the power which was less than 50 W was considered noise in the signal analysis [24]. The total sum of false positive was the value that the algorithm detected as an event, even though through manual inspection of the signal, that portion of signal was not an event but signal noise.

$$\Delta P_{\mathrm{FPS}} = \sum_{f \in F} |\Delta P_f|, \tag{3}$$

where F is the set of all false positives.

3.4 ON-OFF pairing

In the algorithm, the ON/OFF appliance model was utilized [1,22]. The ON-OFF pairing methodology used in the algorithm was performed after the GOF. In the algorithm, the cluster positive members were separated from the cluster negative ones. Next, the algorithm scanned each of the cluster positive members and calculated the current cluster positive member and the next cluster negative one. If the calculated values of both cluster were less than 10% difference, therefore both the cluster positive member and cluster negative one is a switching ON and OFF event [22]. Cluster positive members were the detected events with positive power consumptions corresponded to the switching ON of the appliances. On the other hand, cluster negative one were the detected events with negative power consumptions corresponded to the switching OFF of the appliances.

3.5 *K*-means clustering

The *K*-means clustering methodology, one of the popular methods of unsupervised learning [25,26], was used in the algorithm. The number of clusters, K, was initialized to 3. The number of clusters was the number of centroid. K-means iteratively assigned a pair of ON-OFF events to the cluster with its centroid closer to the power consumption values of the ON-OFF event. The centroids of all clusters were calculated. The process repeated itself until no change to the cluster memberships appeared. The purpose of performing K-means clustering was to group similar ON-OFF pairs (i.e. the events with similar poser consumption values) into the same cluster, since similar ON-OFF events have greater likelihood of originating from the same appliance or the same state.

4 Results

4.1 Event detection results

Figure 2 demonstrates the aggregated power consumption versus time with window size n used in GOF event detection when K was set to 3. The blue square boxes indicate that an event is detected.

There are several false detections in the detected events. Therefore, all noises for power consumption of less than 50VA were filtered out. The result obtained from the algorithm, that is, apparent power versus time is depicted in Fig. 3.

4.2 ON-OFF pairing result

From the ON-OFF pairing algorithm, the pairs of switching ON and OFF event of the appliances were identified, as displayed in Fig. 4. The ON-OFF pairing algorithm was performed on the clusters of the apparent power consumption detected in the preceding step, that is, the GOF event detection algorithm.

4.3 K-means clustering result

From the ON-OFF pairing results, the data were set to three clusters. In the K-means algorithm implementation, three clusters have been formed, as exhibited in the result of the cluster in Fig. 5.

Based on the result in Fig. 5, the combination of the GOF event detection, ON-OFF pairing and K-means cluster analysis methodology made it possible to identify the switching ON and OFF events based on the cluster of the electrical appliances. The time difference between a pair of ON-OFF event was the time duration that the appliance was turned on. This information is useful to inform the user the amount of electricity consumed by each appliance and time duration of the use. It is also useful for utilities to create an itemized billing of electricity consumption that lists down all appliances and the time durations of the use of each appliance.

5 Conclusions

In this paper, a systematic approach toward ON-OFF event detection and K-means cluster analysis were performed for NIALM. The result for the GOF event detection indicated that the algorithm was able to filter out

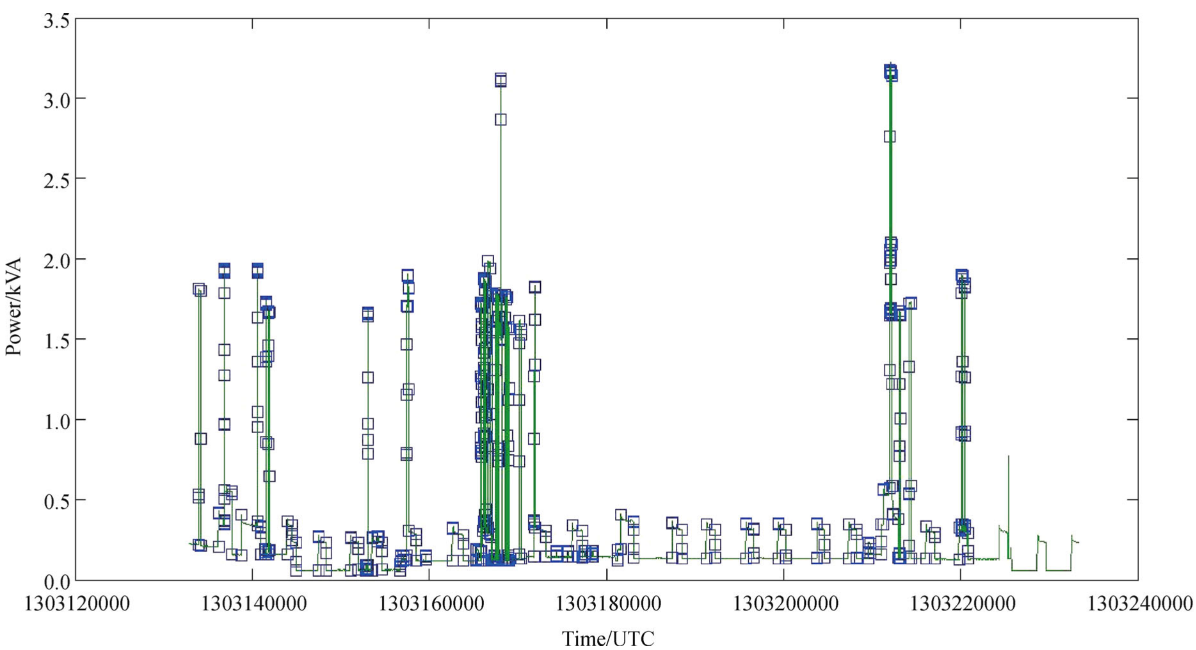

Fig. 2 Aggregated power consumption versus time with window size n used in GOF event detection set to 3

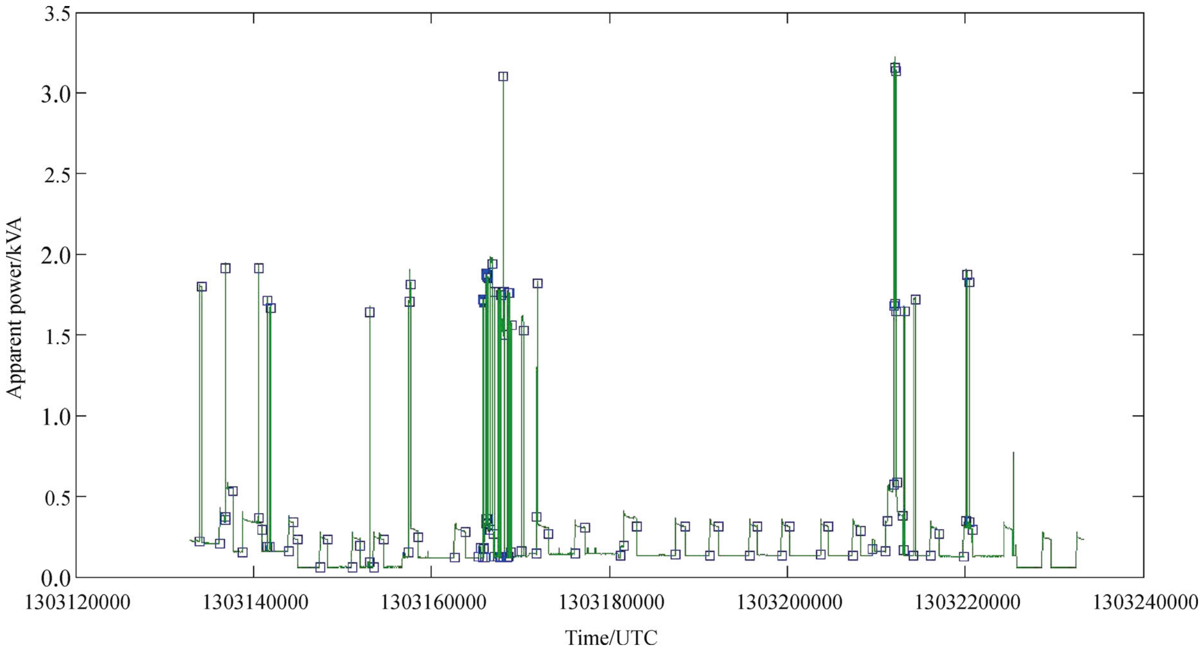

Fig. 3 Apparent power versus time

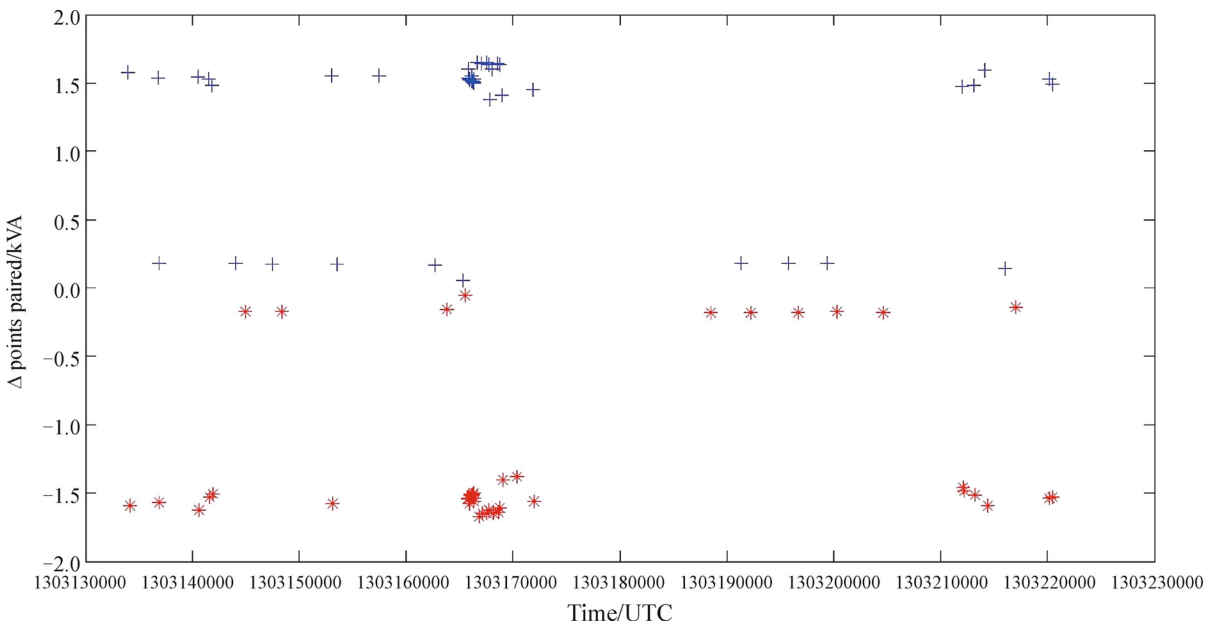

Fig. 4 Pairs of switching ON and OFF event of appliances

the noise and detect the edge of the events. Subsequently, from the GOF event detection results, an ON-OFF algorithm was performed to obtain the ON-OFF pairs of the events detected. The *K*-means algorithm was able to identify the ON-OFF paired events into pre-set clusters accordingly. With this proposed combination of methodology, the future work will focus on the classification of detected events into appliance categories utilizing classification methods such as *k*-nearest neighbor or naïve Bayes methods.

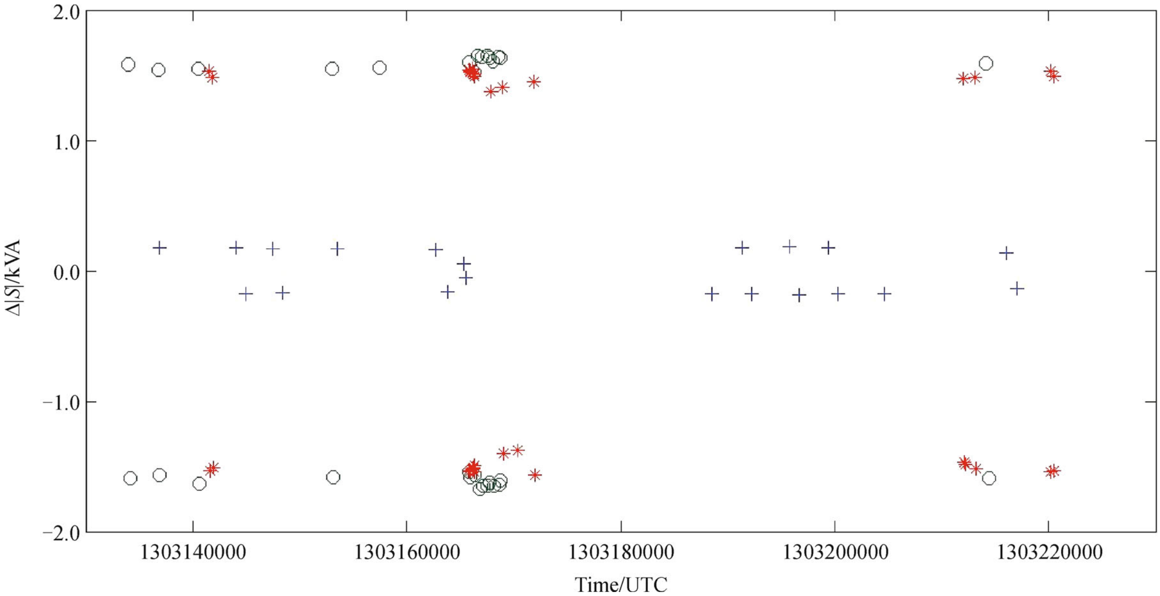

Fig. 5 Three clusters formed

Acknowledgements The work is funded by Ministry of Science, Technology and Innovation (MOSTI) Malaysia under the MOSTI Science fund project (No. 06-02-11-SF0162).

References

1. Hart G W. Nonintrusive appliance load monitoring. Proceedings of the IEEE, 1992, 80(12): 1870–1891
2. Yang C C, Soh C S, Yap V V. Comparative study of event detection methods for non-intrusive appliance load monitoring. Energy Procedia, 2014, 61, 1840–1843
3. Zeifman M, Roth K. Nonintrusive appliance load monitoring: review and outlook. IEEE Transactions on Consumer Electronics, 2011, 57(1): 76–84
4. Zoha A, Gluhak A, Imran M A, Rajasegarar S. Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. Sensors (Basel), 2012, 12(12): 16838–16866
5. Wong Y F, Sekercioglu Y A, Drummond T, Wong V S. Recent approaches to non-intrusive load monitoring techniques in residential settings. In: IEEE Symposium on Computational Intelligence Applications in Smart Grid. Singapore, 2013, 73–79
6. Norford L K, Leeb S B. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. Energy and Building, 1996, 24(1): 51–64
7. Laughman C, Lee K, Cox R, Shaw S, Leeb S, Norford L, Armstrong P. Power signature analysis. IEEE Power and Energy Magazine, 2003, 1(2): 56–63
8. Liang J, Ng S K K, Kendall G, Cheng J W M. Ng S K K, Kendall G, Cheng J W M. Load signature study Part I: basic concept, structure, and methodology. IEEE Transactions on Power Delivery, 2010, 25 (2): 551–560
9. Lee W K, Fung G S K, Lam H Y, Chan F H Y, Lucente M. Exploration on load signatures. In Proceedings of International Conference on Electrical Engineering (ICEE), Sapporo, Japan, 2004, 1–5
10. Lam H Y, Fung G S K, Lee W K. A novel method to construct taxonomy electrical appliances based on load signatures. IEEE Transactions on Consumer Electronics, 2007, 53(2): 653–660
11. Patel S N, Robertson T, Kientz J A, Reynolds M S, Abowd G D. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In: Proceedings of the 9th International Conference on Ubiquitous Computing. Innsbruck, Austria, 2007, 271–288
12. Gupta S, Reynolds M S, Patel S N. ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing. Copenhagen, Denmark, 2010, 139–148
13. Leeb S B, Shaw S R, Kirtley J L. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. IEEE Transactions on Power Delivery, 1995, 10(3): 1200–1210
14. Chang H H, Yang H T, Lin C L. Load identification in neural networks for a non-intrusive monitoring of industrial electrical loads. Lecture Notes in Computer Science, 2008, 5236: 664–674
15. Figueiredo M B, Almeida A D, Ribeiro B. An experimental study on electrical signature identification of non-intrusive load monitoring (NILM) systems. Lecture Notes in Computer Science, 2011, 6594: 31–40
16. Chang H H. Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. Energies, 2012, 5(12): 4569–4589
17. Shaw S R, Leeb S B, Norford L K, Cox R W. Nonintrusive load

monitoring and diagnostics in power systems. IEEE Transactions on Instrumentation and Measurement, 2008, 57(7): 1445–1454

18. Hazas M, Friday A, Scott J. Look back before leaping forward: four decades of domestic energy inquiry. IEEE Pervasive Computing, 2011, 10(1): 13–19
19. Anderson K, Berges M, Ocneanu A, Benitez D, Moura J M F. Event detection for non- intrusive load monitoring. In: Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society (IECON). Montreal, Canada, 2012, 3312–3317
20. Zico Kolter J, Johnson M J. REDD: a public data set for energy disaggregation research. In: Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability. San Diego, USA, 2011, 1–6
21. Arias-Castro E, Donoho D L. Does median filtering truly preserve edges better than linear filtering? Annals of Statistics, 2009, 37(3): 1172–1206
22. Giri S, Lai P, Berges M. Novel techniques for ON and OFF states detection of appliances for power estimation in non-intrusive load monitoring. In: Proceedings of the 30th International Symposium on Automation and Robotics in Construction and Mining (ISARC). Montreal, Canada, 2013, 522–530
23. Jin Y, Tebekaemi E, Berges M, Soibelman L. Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities. In: Proceedings of the 2011 International Conference on Acoustics, Speech and Signal Processing (ICASSP'11). Prague, Czech Republic, 2011, 4340–4343
24. U.S. Department of Energy. Estimating Appliances and Home Electronic Energy Use. 2014–11–05, http://energy.gov/energysaver/articles/estimating-appliance-and-home-electronic-energy-use
25. Goncalves H, Ocneanu A, Berges M. Unsupervised disaggregation of appliances using aggregated consumption data. In: Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, USA, 2011, 1–6
26. Wang Z, Zheng G. Residential appliances identification and monitoring by a nonintrusive method. IEEE Transactions on Smart Grid, 2012, 3: 80–92