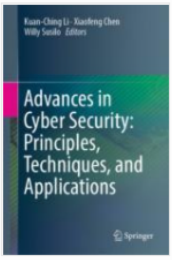

Advances in Cyber Security: Principles, Techniques, and Applications pp 23-48 | Cite as

Non-intrusive Load Monitoring Algorithms for Privacy Mining in Smart Grid

Authors Authors and affiliations

Zijian Zhang, Jialing He , Liehuang Zhu, Kui Ren

Chapter
First Online: 07 December 2018

Non-Intrusive Load Monitoring Algorithms for Privacy Mining in Smart Grid

Zijian Zhang, Jialing He*, Liehuang Zhu, Kui.Ren

Abstract Non-intrusive load monitoring (NILM) method is essentially artificial intelligence algorithms for energy conservation and privacy mining. It obtains consumers' privacy data by decomposing aggregated meter readings of consumer energy consumption into the individual devices energy consumption. In this chapter, we first introduce the background and the advantage of the NILM method, and the classification of NILM method. Secondly, we demonstrate the general process of NILM method. The specific process contains data preprocess, event detection and feature extraction, and energy consumption learning and appliance inference. Furthermore, we introduce a supervised NILM example and an unsupervised example. We describe their processes, and discuss their characteristics and performances. In addition, the applications of NILM method are depicted. Lastly, we conclude this chapter and give the future work.

Zijian Zhang
School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 P.R. China. e-mail: zhangzijian@bit.edu.cn

Jialing He (Corresponding author)
School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 P.R. China. e-mail: hejialing@bit.edu.cn

Liehuang Zhu
School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 P.R. China. e-mail: liehuangz@bit.edu.cn

Kui.Ren
Institute of Cyber Security Research and School of Computer Science and Engineering, Zhejiang University, Hangzhou 310058, China.

1 Introduction

Nowadays, Smart grid [62] is usually equipped with smart meters to allow utility companies to monitor the grid more granularly, which allows them to predict changes in demand more accurately, detect failures more quickly and adapt pricing and electricity generation more dynamically. It is common knowledge that utility companies all over the world are very willing to be bringing in time-of-day usage data. This data tries to discourage electricity consumption among the peak hours and defers these consumptions to off-peak time (a.k.a. peak shaving). When the tariffs are flexibly adjusted due to peak shaving, smart meters are crucial to monitor customers' electricity consumption data at a high time resolution, in order to remind customers' uneconomical usages of electricity appliances in real time. Smart meters also benefits customers as they can monitor and adapt their energy consumption in real time to reduce costs.

Apart from the benefits to the utility companies and customers, smart meters are also favored by the electricity manufacturers, advertisement companies and insurance companies because meter readings can uncover the customers' activities in houses by Appliance Load Monitoring (ALM) methods. Generally, the usage of a certain electrical device, e.g. a computer or an air conditioner etc., can be identified via monitoring of the meter readings. In a special case, even the program on TV can be identified based on its corresponding electricity usage. The in-house activities and when they occur have highly commercial value for habit and preference perception, personalized advertisement push, insurance and risk evaluation.

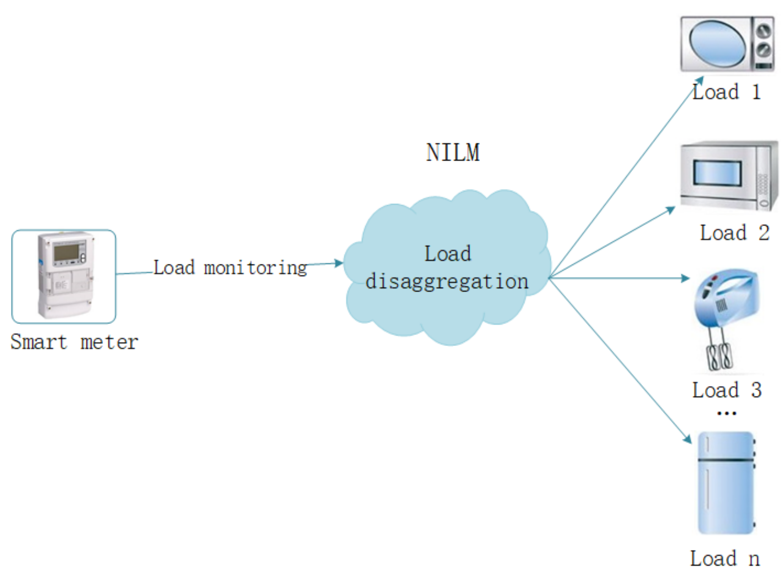

Fig. 1 The non-intrusive load monitoring system diagram.

There are two major approaches to monitor the energy consumption of appliances, normally are called Intrusive Load Monitoring (ILM) and Non-Intrusive Load Monitoring (NILM) [1, 2, 3]. The ILM approach requires to install multiple meters in houses. These meters take responsibility for recording the energy consumption of all the appliances incessantly. Although the results of ILM approach are usually accurate, it is impractical because of its high demanding costs, multiple meter configurations and the process for installation is very complicated as well [4, 5, 6]. On the contrary, the NILM approach is convenient to be implemented as it merely needs to install a single meter outside houses [7, 8, 9, 10]. The non-intrusive load monitoring system diagram is depicted in Fig. 1. However, since no special hardware or software is assumed to be installed or run in the customer residences, the working status of each appliance is a completely black box. The only way to analyze the customers' activities has to disaggregate the high resolution meter readings directly. Therefore, the precision of the detection and the largest number of the electricity appliances are both limited. Table 1 compares the advantage and disadvantage of the ILM and NILM methods. From the viewpoint of the practicability and cost, the NILM methods outperform ILM methods because the NILM methods do not intrude the customers' residences and they require only one smart meter. Therefore, this chapter mainly focuses on introducing the NILM approach, and the effort to increase the number of the appliances and to improve the accuracy of the result.

Table 1 The Advantage and Disadvantage of the ILM and NILM Approaches

Parameters	ILM	NILM
Indoors Installation	Necessary	Not Necessary
Numbers of Smart Meter	Large	Only One
Numbers of Electricity Appliances	Small	Large
Accuracy	Relatively High	Relatively Low

2 General Process of the NILM Approach

The task of the NILM approach aims to disaggregate a aggregated electricity signal that often means the power consumption into the involved individual devices' contributions [3]. Formally, for the total aggregation consumption $Y(Y \in Y_1, Y_2, ..., Y_T)$ of all the appliances in the house at the time $t(t \in 1, 2, ..., T)$, the purpose of the NILM approach is to recover the consumption y_t^i of the $i-th$ device at time t. If there are N appliances in total, and $i \in 1, 2, ..., N$, we have $Y_t = \sum_1^N y_t^i$.

The general process of the NILM approach often requires three steps, as shown in Fig. 2. First is data acquisition. This step processes the raw smart meter readings, mostly for the missing consumption or outliers. Next is event detection. In this step, the task is to capture the major change of the power consumption through the smart readings. These major changes of the power consumption often can identify the related appliance because the changes usually happen when the appliance turned on or off. Finally, the specific NILM algorithms are designed to infer the usage of the electric appliances due to the feature of their energy consumption.

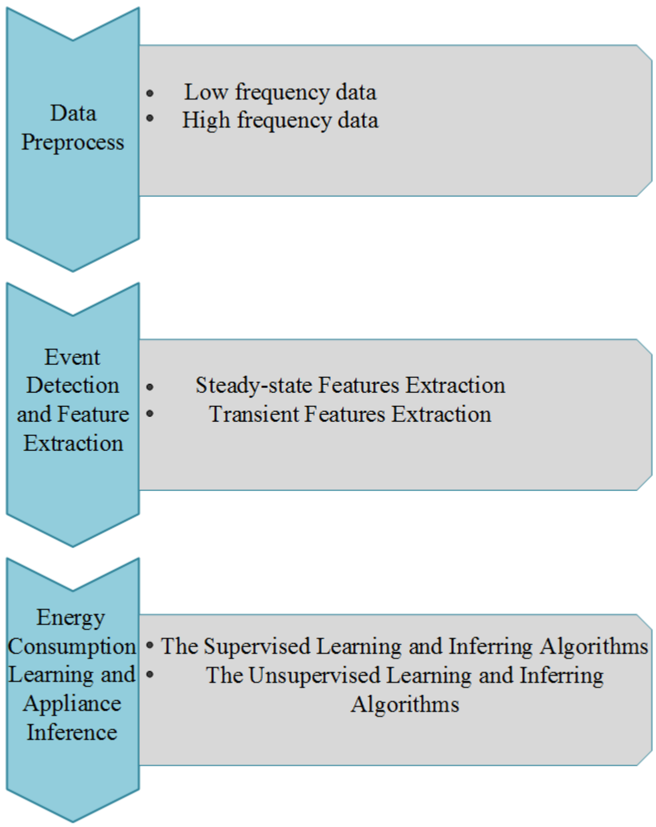

Fig. 2 The general process of the NILM approach

According to the operation mode of the appliance, all the appliances can be divided into four types [13].

Type 1: ON/OFF Appliances belong to this category only has two states of operations (ON/OFF). Most of the appliances like TV and toaster are belonged to this category.

Type 2: Finite State machines (FSM) These appliances own two or more operating states (the number is finite). And whole switching cycle is repeated frequently

among these states. Examples of such devices include refrigerator (OFF/ON/Defrost) and dryer (OFF/Heater + Motor/Motor only).
Type 3: Continuously Variable Consumer Device Appliances belong to this category with variable power draw but without fixed number of states. The regular rules of these appliances' power consumption is hard to capture because there is no repeatability. It causes the difficulty for NILM algorithm to disaggregate them from the aggregated load. Devices like power drill and dimmer lights are in this category.
Type 4: Permanent Consumer Device Devices like smoke detector belong to this category because the rate they consuming energy remains active and constant through days, weeks or even months.

2.1 Data Preprocess

In the NILM approach, data preprocess is the first step to deal with the raw aggregated consumption that is generated by a smart meter at a regular rate. According to the size of the sampling frequency, the aggregated consumption data can be divided into two types. One is low frequency data and the other is high frequency data [11, 12, 61].

From the viewpoint of low frequency, it is usually no more than 1 Hz. The sampling time interval is about 1/60 second. The cost of low frequency meters is inexpensive, but only major power changes can be captured from low frequency data [1]. From the viewpoint of high frequency, it varies from 10 KHz to 100 MHz. The cost of high frequency meters is expensive, but more information like outliers can be obtained from high frequency data [13].

Several benchmark datasets, such as BLUED [14] and REDD [15] are currently available to be downloaded. These datasets contain both the low-frequency data and the high-frequency data. For instance, the REDD dataset provides the low-frequency data with 1 Hz and high-frequency data with 15 KHz. This dataset collects energy consumption data in 6 houses for more than one month. As a benchmark dataset, it not only has the aggregated meter readings but also has the energy consumption of each individual appliance. If a dataset misses part of energy consumption data, the average of the neighboring data could be filled to achieve data integrity.

2.2 Event Detection and Feature Extraction

In the NILM approach, the event detection is critical for disaggregation as it helps to recognize the working states of the appliances. Here we define an event as a major change of the meter readings. The event detection has a complex process because of the multiple kinds and working states of the electricity appliances. In fact, two states, finite state, continuously variable states and constant states are four kinds of states according to the pattern of their energy consumption [16, 60].

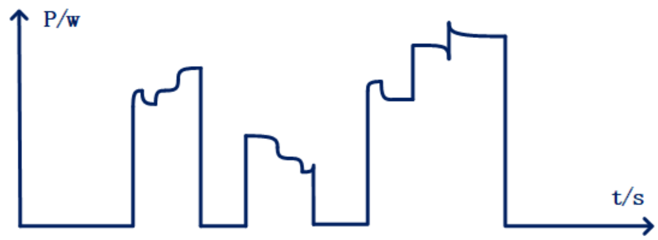

Fig. 3 The power monitored of appliances

Event detection algorithms concentrate on analyzing all the events in the consumption data. Roughly, when an event occurs, some operations must be run for some electricity appliances. For instance, for a two-state appliance, the operations could be switching the appliance on or off. Almost all the event detection algorithms rely on the fact that the meter readings are fluctuate from time to time, as shown in Fig. 3. Hence, the events can be found in terms of transient changes and steady-state changes, accordingly.

There exists some feature extraction algorithms for characterizing the detected events of the appliances. Steady-state event detection algorithms mainly watch the power change of appliances. For example, when the energy consumption changes from a higher value to a lower value, some corresponding appliance is identified to be turned off. The steady-state event detection algorithms are mainly applied to analyze the low frequency data. The transient event detection algorithms view the state transitions as a signature to identify the appliances because the transitions are unique for the appliances. The features of the state transitions include but not limited to the duration, shape, size and harmonics of the waveforms [2]. However, these signatures cannot be well extracted unless the sampling frequency is higher than 1 KHz. The transient event detection algorithms are common used for analyzing high frequency data.

After the steady-state and transient events are detected, the NILM approach starts to extract the features of the electricity appliances. The electricity appliances have steady-state features and transient features in accordance with the two kinds of events.

Steady-state Features Extraction

We also take a two-state appliance as the example. The steady-state features [16] consist of the real power and the reactive power. The real power represents the total amount of the energy consumption of each electricity appliance. For the pure resistor circuit, since the current and voltage waveforms have the same phase, there is no reactive energy. For inductor and capacitor circuit, the current and voltage, the current and voltage waveforms have different phases. Thus, the real power and the reactive power both exists. The real power can distinguish the higher-power appliances like water pumps because their power draw characteristics are distinct [17, 18, 19, 59]. But multiple states transform simultaneously will lead to erroneous result. Furthermore, the steady-state features are not proper for the appliances when the feature space of the real power and reactive power exist a considerable overlap, typically for the appliances with low-power as depicted in the Fig. 4.

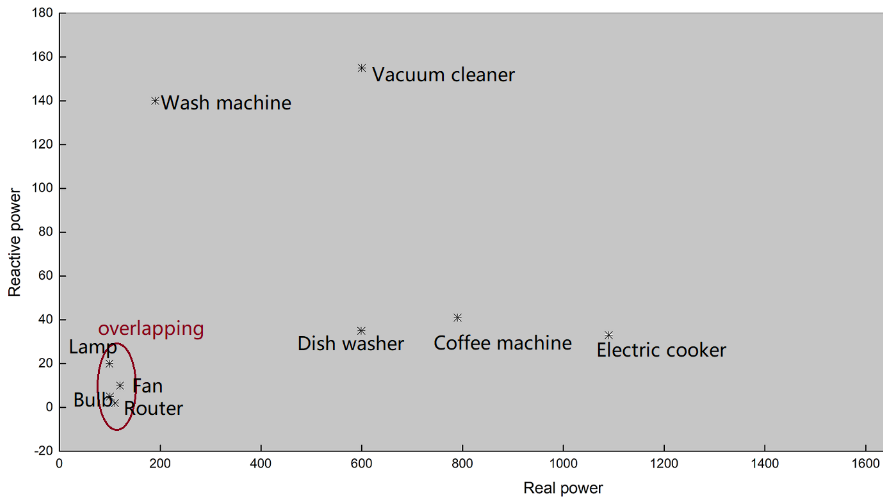

Fig. 4 The power monitored of appliances

Apart from the real power and the reactive power, the peak power, the root mean square of the voltage and current values, the phase difference and the information of power factor are also common features to disaggregate the meter readings. For example, the power factor information is a ratio between the real power and the apparent power. This feature has good performance for disaggregating the two-state electricity appliances in the kitchen [20].

Besides the traditional features, a novel disaggregation algorithm that uses voltage and current trajectory to differentiate a group of appliances has been proposed [21, 22]. The voltage and current trajectory differenciate the class of appliances into different groups with high accuracy, giving further sub-division with each individual group. The result shows that this algorithm is more effective than the existing algorithms because of a taxonomy of the electrical appliances according to the dedicated voltage and current curves. Gupta et al. [23] also explored that the steady-state voltage noise can characterize the appliances if they equip with a Switch Mode Power Supply.

Although several efforts are made in the steady-state feature extraction algorithms, these algorithms have often two disadvantages. One is to require additional hardware for measurement. The other is that the algorithms are quite sensitive when monitoring to the wiring architecture.

We provide a introduction of the existing steady-state feature extraction algorithms in Table 2.

Table 2 Summary of steady-state methods.

Features type	Extraction methods	Advantages	Disadvantages
Real and Reactive Power [7, 17, 18, 19, 24]	Calculation of change	Simple Intuitive	Overlap (P-Q plane); Poor performance for Type 2, 3 and Type 4 loads.
V-I waveforms [20, 25, 26, 27, 28, 29, 30, 31]	Calculation of a series of index $I_{\{}RMS\}, I_{\{}avg\}, I_{\{}peak\}, V_{\{}RMS\}$	Devices can easily be categorized into resistive, inductive and electronic loads	High sampling Low accuracy for Type 3 loads , overlapping features for Type 1 and 2 category, unable to distinguish between overlapping activation events
V-I Trajectory [32, 33]	Shape features of V-I trajectory	Detail taxonomy of electrical appliances can be formed due to distinctive V-I curves	Difficult to distinguish smaller loads
Voltage Noise [34, 23]	Fast Fourier Transform (FFT)	Those with SMPS can be recognized with high accuracy	The ability of anti-interference is very poor; Poor generality

Transient Features Extraction The Transient features extraction algorithms reduce the overlaps of the steady-state features, thereby improving the accuracy of disaggregation. Correspondingly, these algorithms need higher frequency data.

One classic transient features extraction algorithm analyzes the spectral envelope of the waveform based on Short-Time Fourier Transform (STFT) which has been proven to be useful in detecting the appliances with variable-state. Unfortunately, this algorithm can just detect whether the appliances are turned on or not in a certain time period, but not determine the specific time point when the appliances are turned on. To solve this problem, the revised algorithm that applies the correlate spectral envelopes with real power and reactive power components is proposed [45, 46]. However, the revised algorithm often suffers from the excessive training.

Comparing with Short-Time Fourier Transform, the wavelet transform has also been used to identify all the transient features of appliances. The transient response features and transient ftime of a transient features extraction algorithm are proven to be less than those of a steady-state features extraction algorithm [47, 57].

Several recent works [11, 13, 58] show good results via sampling the voltage noise that happens with the transient events frequently(like turning from off to on). The idea is based on the fact that all of the appliances would emit voltage noises. Specifically, all of the noises are normally divided into three categories: steady-state continuous noise, on-off transient noise and steady-state line voltage noise. Table 3 demonstrates some typical transient-based NILM approaches.

Table 3 Summary of transient-based methods

Features type	Extraction methods	Advantages	Disadvantages
Transient power [35, 30, 36, 37]	FFT; Power spectrum envelope estimation; Calculation of waveform vector.	Appliances with same power draw characteristics can be easily differentiated. Recognition of Type 1,2,3 loads	Continuous monitoring, high sampling rate not suitable for Type 4 loads
Start up Current Transients [17, 38, 36]	Calculation of a series of index current spikes, size, duration...	Works well for Type 1,2 loads, distinct transient behavior in multiple load operation scenario	Poor detection of Simultaneous activation deactivation of sequences, unable to characterize Type 3 and 4 loads, sensitive to wiring architecture, appliance specific
Voltage Noise [34, 39]	Spectrum analysis	Able to distinguish similiar loads	There is a certain error in identifying simultaneous activation

2.3 Energy Consumption Learning and Appliance Inference

In this process, the specific states of all the appliances at each time point are classified or clustered via the features detected before. The whole process contains two types of algorithms. One is energy consumption learning algorithms and the other is appliance inference algorithms. The energy consumption learning algorithms are used to train the appropriate parameters, while the appliance inference algorithms are used to infer the specific states of appliances at all the time points. These algorithms can be categorized into two kinds, the supervised learning and inferring algorithms and unsupervised learning and inferring algorithms [54, 55, 56].

The Supervised Learning and Inferring Algorithms

The supervised learning and inferring algorithms first train parts of the meter readings due to the energy consumption of each appliance, in order to infer the states of the appliances. Then they test the other of the meter readings to evaluate the accuracy of the inference. Since installing the meter readings to monitor all the appliances is an expensive and time-consuming process, the scalability of these algorithms are limited.

Some may argue that these algorithms belong to ILM approach, because they need to install smart meters in houses. But they are traditionally viewed as NILM approach as the test process does not require to intrude the houses and intruding behavior is only once during in the training phase.

The Unsupervised Learning and Inferring Algorithms

Unlike the supervised learning and inferring algorithms, the unsupervised learning and inferring algorithms neither require to train data for each appliance, nor need to monitor the energy consumption of each appliance. The parameters are adjusted only through the meter readings themselves [40].

The unsupervised learning and inferring algorithms can further be divided into three subgroups as suggested by [41]. First is to require unlabeled the training data to build the appliance model or to populate appliances database. The appliance model is either generated manually [42] or produced automatically [43] during training the meter readings. The appliances in houses are assumed to be known before.

The second uses the labelled data from an unknown house to build appliances models This algorithm is then used to disaggregate a brand new house. Most deep learning based NILM methods fall in this category.

The third neither requires to train before disaggregation, nor requires to monitor the energy consumption of each appliance. In other words, this algorithm has not any prior knowledge [44, 45].

Generally speaking, NILM algorithms based on supervised learning emerge in an endless stream, but there are not many kinds of load involved, and the scene of processing is relatively simple. Compared with supervised algorithms, the unsupervised ones have low accuracy, but they have the advantage of reducing manual intervention and have good prospects for development. The improved and integrated algorithms can cope with complex scene, which are worth continuing to study. Table 4 depicts the comparison of some of the supervised and unsupervised algorithms.

Table 4 Summary of learning and inferring algorithms

Methods type	Algorithm types	Features type	Application Scenarios	Accuracy
Supervised NILM	SVM	Steady Transient	Common appliances; The accuracy is low when Type 4 appliance is not considered; Not much type of load;	66%-100%
	Nearest Neighbour (KNN)	Transient	Non concurrent events; Linear load; Single load and multi load operation scenes;	78%-100%
	Adaboost	Steady	Some common appliances; Consideration of complex conditions such as noise interference;	98%-99%
Unsupervised NILM	HMM	Steady	Not contain much appliances; No Type 4 Appliances;	52%-98%
	Motif mining	Steady	Appliances' number is known; Low accuracy for Type 2 appliances;	15%-98%
	Deep learning	Steady Transient	Concurrent events; Similiar load features;	50%-100%

3 Examples

In this section, we take two examples to further explain the two kinds of learning and inferring Algorithms.

3.1 A Supervised Learning and Inferring Example

In this section, we will introduce Wang's scheme [52] which is based on sparse coding. The scheme proposed a new clustering algorithm–Probability Based Double Clustering(PDBC), which can promote the efficiency of clustering the device consumption features. Real experiments based on data set–REDD showed the average disaggregation accuracy of this scheme could reach 77%.

There are total n appliances in a house. Each device is modeled based on sparse coding. As to each device $i = 1, 2, ..., n$, the feature matrix of the device is learned at the frequency of q HZ. Once obtaining enough features for the $i-th$ involved appliance, the following approximation would be utilized:

$$\begin{gathered} T_w\left(x_i\left(t\right)\right) \approx X_i a_i\left(t\right), T_w\left(x_i\left(t\right)\right) = \\ \left(x_i\left(t\right), x_i\left(t+\frac{1}{\mathrm{q}}\right), x_i\left(t+2\times\frac{1}{\mathrm{q}}\right), \ldots, x_i\left(t+T_w\right)\right) \end{gathered} \tag{1}$$

In Eqs1, T_w was the time of a sliding window, which denoted a lasting period of time during the training process and $T_w \ll T_{tr}$ where T_{tr} means the whole training time. m_i denoted the features that captured from the $i-th$ device. $X_i \in R^{m_i \times T_w}$. $a_i\left(t\right)$ demenstrated the activations of feature matrix X_i. $a_i\left(t\right)$ was sparse that contained mostly zero atoms and only one none-zero element that is 1. Then the scheme utilized a matching algorithm to get the $a_i\left(t\right)$ so that they could calculate $X_i a_i\left(t\right)$ to obtain an approximate solution of all involved individual device. There were mainly two stages in this scheme, which are learning the feature and power consumption matching.

3.1.1 Learning the feature

In this stage, the PBSC algorithm was proposed to learn each device feature matrix. In the scheme, it assumed that the training process consumes T_{tr} time and the sampling frequency is q Hz, so there were $(T_{tr} - T_w + 1) \times \mathrm{q}$ sliding windows,as to a sliding window, $T_w \times \mathrm{q}$ elements are included.

Assumed that each sliding window was regarded as a data point in the algorithm. The data point was defined as $\mathrm{P}_i \in (1, (T_{tr} - T_w + 1) \times \mathrm{q})$, therefore there were total k unique points P_{uniq} after removing the repeated points. The vector $\mathrm{R} = (r_1, r_2, r_3, \ldots, r_k)$ denoted the repeat times for each data point. $d_{ij} = \left\|P_i - P_j\right\|_2$ was defined as the distance of two data points, the distance matrix D was a $\mathrm{k} \times \mathrm{k}$ scale symmetric matrix with a diagonal of 0, $\mathrm{D} = \left\{\mathrm{d}_{ij}, \mathrm{i}, \mathrm{j} \in \{1, 2, 3, \ldots, k\}\right\}$.

Then the PBDC algorithm is utilized to perform the cluster task, specifically, m is a upper limit that could represent the number of clustering centers. Then in the clustering process, it just repeatedly compare the number of the actual clustering centers and m. Therefore, they could make the clustering efficient with the number of clustering centers stabilized at a certain controllable value.

After this first clustering process, the center point represented the point with the larger distance from the other data points, here the result was just evaluated by the difference of the data values. The result is not fair enough because some of the data points that with small value have a high probability for being buried in large. So a second clustering was further performed to promote the efficiency.

In the second clustering, all the clusters got from the first clustering were further clustered. The number of clusters was set through the probability set C. The C_i in the set means the probability of each individual cluster.As to all repeated points, C_i was computed through Eqs2 and 3.

$$\mathrm{C}_i = \frac{num\,(Clusteri)}{\sum_{i=1}^{k} r_i} \tag{2}$$

$$num\,(Clusteri) = \sum r_i, r_i \in \{Clusteri\} \tag{3}$$

The algorithm 1 in [52] described the details.

Algorithm 1 Learning Feature Matrix [52]

Input: Device training data set D_{tr},
 Sliding window size w,
 The number of feature vector m.
Output: Device feature matrix X.
1. Compute P_{uniq} and R from D_{tr}
2. Compute distance matrix D:
3. **for** $i = 1, \ldots, size(R)$ **do**
4. **for** $j = i, \ldots, size(R)$ **do**
5. $d_{ij} = \|P_i - P_j\|_2$
6. **end for**
7. **end for**
8. First Clustering: improved FSFDP(D,$m/10$)
9. Second Clustering:
10. **for** $k = 1, \ldots, size(clusters)$ **do**
11. $C_i = num(cluster_i)/\|\mathrm{R}\|_1$
12. improved FSFDP(D_i,$m * C_i$)
13. clustering result is collected to X
14.**end for**
15.**retrun** X

3.1.2 Power consumption matching

In the leaning feature matrix stage, the feature matrix for each individual device X_i has been obtained. In the power consumption matching stage, the main task was to getting $a_i(t)$. The Max-Min pruning Matching (MMPM) algorithm was proposed to commit the task. The main contribution of MMPM was to performing the pruning optimization for shorting the decomposing time while obtaining the global optimum solution. The optimization process was consisted of two processes which were minimum pruning process and maximum pruning. It assumed arg_{max} denoted the maximum pruning parameters, μ represented the threshold of pruning.

The contribution of maximum pruning was to obtain the global optimal result and cease the matching loop when they got the global optimal result. To perform this matching process, getting the order j that represents the maximum element of $T_w(y(t))$. Then sorted each feature matrix $X_i, i \in \{1,2,3\ldots,n\}$ according to the $j-th$ element as the descending order, they regarded the element in the X_i in the $j-th$ column as the maximum. Then the maximum pruning parameters were calculated:

$$arg_{max} = y(t+(j-1)q) - \sum_{i=n-i}^{n} max_i \tag{4}$$

When $arg_{max} > \mu$, it depicted that the arg_{max} in the remaining loops would be larger than the given pruning threshold, so all the remaining loops would be cut off.

The contribution of minimum pruning process was to cut a invalid loop once the value of the vectors is found too small. For each loop, they obtained a remainder power $T_w(r(t))$ which denoted the difference between the aggregated power $T_w(\mathrm{y}(t))$ and the value of upper loop, the minimum pruning judgement regulation was defined:

$$\min(T_w(r(t))) + \mu < 0 \tag{5}$$

If this judgement regular was set up, then the remaining loop would make the $\min(T_w(r(t)))$ smaller, so cutted off all the remaining invalid loops. All the details were described in algorithm 2 in [52].

Algorithm 2 MMPM algorithm [52]

Input: Device feature matrixs $X_1 X_2 \dots X_n$
Test data Y.
Pruning threshold u
Output: Disaggregation result $\tilde{Y}_1 \tilde{Y}_2 \dots \tilde{Y}_n$
get the windows size w and each device feature numbers m_i
while $X_1 \neq NULL$
  get one feature vector form X_1
  compute the remainder energy $T_w(\mathrm{y}(t))$ and arg_{max}
  if($arg_{max} > \mu || min(T_w(\mathrm{y}(t))) + \mu < 0$)
    break;
  end if
  overlay record the feature vector into $\tilde{Y}_1$
  while $X_2 \neq NULL$
    get one feature vector form X_2
    compute the remainder energy $T_w(\mathrm{y}(t))$ and arg_{max}
    if($arg_{max} > \mu || min(T_w(\mathrm{y}(t))) + \mu < 0$)
      break;
    end if
    overlay record the feature vector into $\tilde{Y}_2$
    ...
      while $X_n \neq NULL$
        get one feature vector form X_n
        compute the remainder energy $T_w(\mathrm{y}(t))$ and arg_{max}
        if($arg_{max} > \mu || min(T_w(\mathrm{y}(t))) + \mu < 0$)
          break;
        end if
          overlay record the feature vector into $\tilde{Y}_n$
      end while
    ...
  end while
end while

3.1.3 Experiment Analysis

All the experiments in the paper [52] is executed through the REDD data set [15] which is a public data set constructed for electricity disaggregation. REDD contained six different houses' power consumption signals. For each individual house, about 20 kinds of devices were measured and the aggregated power was recorded in the same time. The measuring time lasted about two weeks and the sampling rate is 1/3 HZ which is a low frequency. The data of the House 5 was excluded because the

Table 5 Energy disaggregation accuracies (%) [52]

	House 1	House 2	House 3	House 4	House 6	Average
Simple	41.4%	39.0%	46.7%	52.7%	33.7%	42.7%
FHMM	71.5%	59.6%	59.6%	69.0%	62.9%	64.5%
PED	81.6%	79.0%	61.8%	58.5%	79.1%	72.0%
EPSC	84.3%	82.7%	70.2%	71.0%	78.9%	77.4%

data in this house is lack of enough fluctuations, which were not proper for feature extractions.

In the experiment, Wang's scheme utilized a month of recorded data with 5 usual household appliances. They used one week data for learning the features and all the rest of the data for power consumption matching. The size of learned feature matrix was set as $20 \times m$ which represented that there are m feature vectors and the length of the vector is 20.

In the last, they got five feature matrixes, the time of the power consumption matching process is about 10 seconds for a sliding window's power readings. The disaggregation accuracy was denoted as follows:

$$acc_{\text{energy matching}} = 1 - \frac{\sum_{t \in \psi} \sum_{i=1}^{M} \left\| \mathrm{T}_w(\mathrm{x}_i(\mathrm{t})) - \tilde{\mathrm{T}}_w(\mathrm{x}_i(\mathrm{t})) \right\|_1}{2 \sum_{t \in \psi} \left\| \mathrm{T}_w(\mathrm{x}_i(\mathrm{t})) \right\|_1} \tag{6}$$

In Eqs 6, $\psi = \{1, T_w + 1, 2T_w + 1, \ldots\}$. They compared their scheme with the PED algorithm [53], supervised FHMM algorithm [8] and the simple mean prediction method.

Fig. 5 showed the time with different feature vector numbers for the algorithm. Table 5 showed the disaggregation accuracy of all the five houses that the REDD data set recorded. Their algorithm performed better than all the other schemes, promoted the accuracy about 5.4% higher.

3.2 An Unsupervised Learning and Inferring Example

An unsupervised learning and inferring example based on deep learning algorithms is illustrated here. We first recall the usage of the deep learning algorithms [46, 47, 48] for energy disaggregation. Next, we introduce a deep neural network ar-

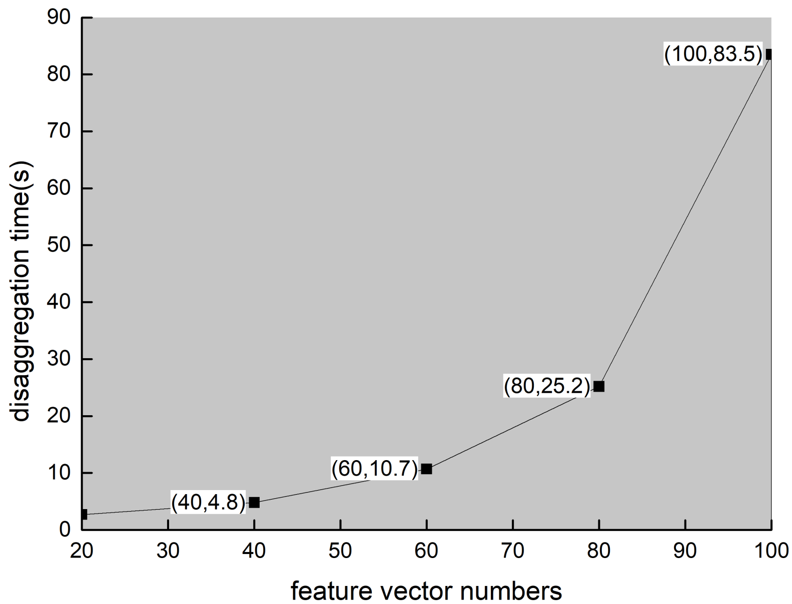

Fig. 5 The time with different feature vector numbers for the algorithm [52]

chitecture that is a form of recurrent neural network referred to as Long Short-Term Memory (LSTM) through the data training, neural network architecture, disaggregation, and results analysis procedures [63].

3.2.1 Data training

During data training process, deep neural networks take a tremendous number of training data just because they own a lot of trainable parameters. So it is critical to acquire large training datasets. For energy disaggregation, tremendous amounts of aggregated data can be effectively created by combing real appliances' activations randomly (the activation denotes the power of an appliance in a complete operating cycle).

The data set of UK-DALE [49] was used as source data in this example. UK-DALE contains the load data of 6 houses, and the data of house 1, 2, 3, 5, 6 is used in the experiment. The experiment contains five appliances which are washing machine, kettle, microwave, dish washer and fridge.

In this experiment, they trained the networks on both synthetic aggregated data and the real aggregated data in 50:50 ratios. Each individual appliance was trained

Table 6 Number of training activations per house [63]

Appliance	House 1	House 2	House 3	House 5	House 6
Kettle	2836	543	44	716	176
Fridge	16336	3526	0	4681	1488
Washing machine	530	53	0	0	51
Microwave	3266	387	0	0	28
Dish washer	197	98	0	23	0

Table 7 Number of testing activations per house [63]

Appliance	House 1	House 2	House 3	House 6	House 6
Kettle	54	29	40	50	18
Fridge	168	277	0	145	140
Washing machine	10	4	0	0	2
Microwave	90	9	0	0	4
Dish washer	3	7	0	3	0

by one network. The input for each network is a window of aggregated power consumption, the desired output of the network is the target appliance's power consumption. In their experiment, the size of the window varied for different appliances, like 1536 samples for the dish washer and 128 samples for the kettle. They reserved the last week of data for training and the rest of the data for testing. Table 6 depicted the number if appliance training activations. Table 9 depicted the number of testing activations. Table 8 depicted the specific houses used for training and testing.

Appliances activations can be extracted by $Electric.get_activations()$ function In the NILMTK tool [50].

The arguments passed to this method are shown in Table 9. For simple applications like toasters, they extracted activations through analysing consecutive samples that were above a predefined power threshold. If the activations were shorter than the threshold duration, then they were thrown. As to some more complicated appliances like washing machine whose power consumption might drop below for short periods of time during a operating cycle, NILMTK ignored these sub-threshold power consumption.

Table 8 Houses used for training and testing [63]

Appliance	Training	Testing
Kettle	1,2,3,4	5
Fridge	1,2,4	5
Washing machine	1,5	2
Microwave	1,2	5
Dish washer	1,2	5

Table 9 Arguments passed to $get_{activations}()$ [63]

Appliance	Max power (watts)	On power threshold (watts)	Min. on duration (secs)	Min. off duration (secs)
Kettle	3100	2000	12	0
Fridge	300	50	60	12
Washing machine	2500	20	12	30
Microwave	3000	200	12	30
Dish washer	2500	10	1800	1800

Locate the activations in the house's meter readings for the target appliance. The code could decide to whether to include the target appliance's activations with the probability of 50% for each training example. If the code decided not to include the target appliance, it would choose a random window of aggregated data without any activations of the target appliance. Otherwise, activations of a target appliance would be included and randomly positioned in a window. In the same time, the time window of the real aggregated data was loaded and loaded together as the input of the network.

Synthetic aggregate data: In order to create the synthetic aggregated data, they extracted the five target appliances' activations from all of the training data. Firstly, two vectors with zero elements were defined, one was the input of the network and the other was the output. The length of the vector was the size of the window related to the network. All the five appliances' activations were scanned through and were decided whether to be included in the training sequence. The probability of appearing in the sequence is 50%, and the probability of being 'distractor' of other appliance was 25%. Specifically, for a target appliance, the activation of the appliance was randomly chose to added in a random position of the input vector. Distractor appliances' activations could appear any positions of the sequence.

3.2.2 Neural Network Architecture

The LSTM was utilized in this example. Specifically, the feed forward neural network that mapped from a solo input vector to a solo output vector was utilized. If a new input vector was cast into the network, then the net would lost the memory about the previous input. In their experiment, RNNS was trained by backpropagation through time (BPTT) [51]. Bidirectional layers were added to enhance the effect of RNNS. Bidirectional RNN, in which there were two parallel RNNS. One of the RNN was utilized to read the input sequence forwards and the other RNN was used to read the sequence backwards. The output of the network was constituted by concatenating the outputs of the above two parallel RNNS.

The architecture is depicted as follows:

1. Input (The length of the input vector is related with the duration of an appliance)
2. 1D conv (filter stride is set as 1, size is 4, number of filters is set as 16, border mode is same and activation function is set as linear)
3. Fully connected (N is set as 1, activation function is set as linear)
4. Fully connected (N is set as 128, activation functionis set as TanH)
5. bidirectional LSTM (N is set as 256 with peepholes)
6. bidirectional LSTM (N is set as 128 with peepholes)

For each time window, the network dealt with a sample of aggregated data and obtained a solo sample of the related target appliance.

3.2.3 Disaggregation

They would preprocess the input data with zeros element in the beginning and the end of the input vector. The net was slide along the input sequence to cope with the complicated situation in which the input time window of a net was at most up to few hours.

3.2.4 Results

The disaggregation result on an unseen house of this experiment is shown in Fig. 6. We use the following metrics in this experiment:

TP means how many true positives exist

FN means how many false negatives exist

FP means how many false positives exist

N means how many negatives in ground truth exist

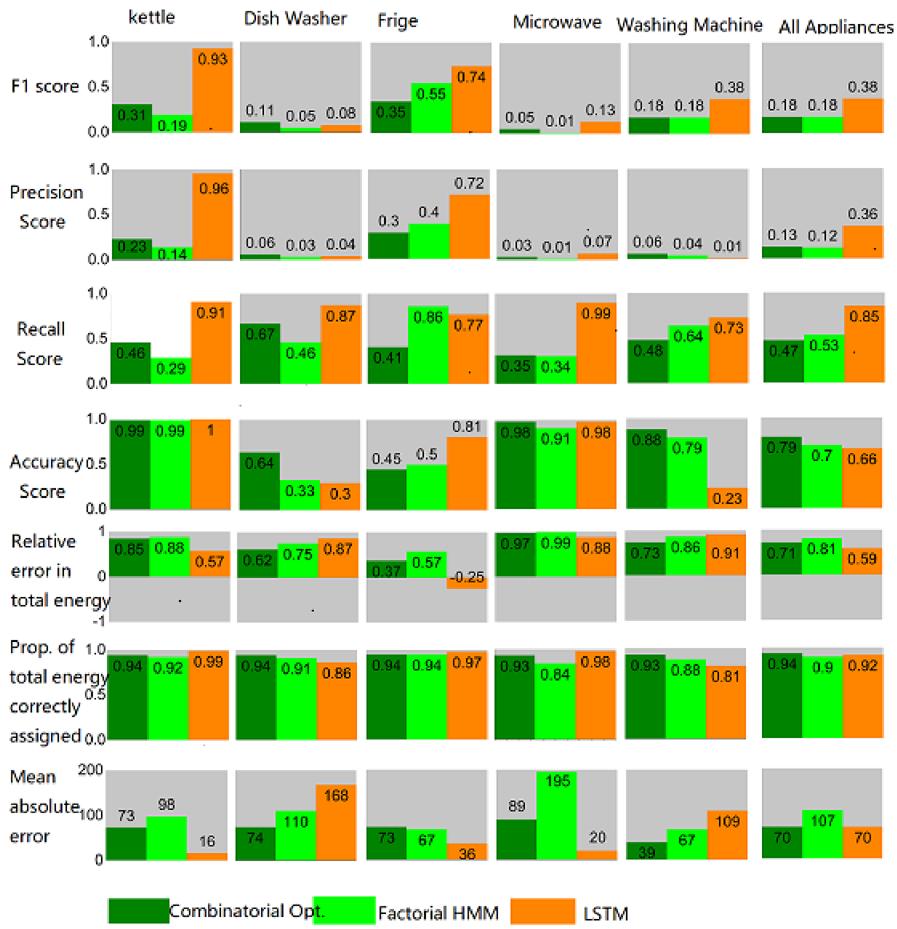

Fig. 6 The general process of the NILM approach

P means how many positives in ground truth exist
E means aggregated actual energy
$\hat{E}$ means aggregated predicted energy
$y_t^{(i)}$ means the $i-th$ appliance's actual power consumption at time t
$\hat{y}_t^{(i)}$ means the $i-th$ appliance's estimated power at time t
$\bar{y}_t^{(i)}$ means the aggregate real power at time t
$precision$ is set as $\frac{TP}{TP+FP}$
$recall$ is set as $\frac{TP}{TP+FN}$
$F1$ is set as $2 \times \frac{precision \times recall}{precision + recall}$
$accuracy$ is set as $\frac{TP+TN}{P+N}$

mean absolute error is set as $\frac{1}{T}\sum_{1}^{T}|\hat{y}_t - y_t|$
relative error of whole energy is set as $\frac{|\hat{E}-E|}{max(E,\hat{E})}$
proportion of total energy correctly assigned is set as $1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}|\hat{y}_t^i - y_t^i|}{\sum_{t=1}^{T}\bar{y}_t}$

4 Applications of NILM

NILM is a promising technology in practical application. It can bring various benefits to users and power companies. For example, it can help users to save electricity power and help power companies to arrange power transmission rationally [64]. In this section we will depict some practical applications of NILM.

4.1 Optimization of power use strategy

Based on the detection data of users' appliances through NILM system, we can analyze users' electricity power usage habits, energy consumption of appliances and other information. If these information can be fed back to users, so that users can take targeted energy-saving measures. If the electricity power usage information of other users is combined, it can provide effective energy saving suggestions for the users. At present, there have been NILM applications like automatic management system for commercial users [65].

4.2 Fault detection and diagnosis

Through the data collected by the NILM system, we can get the current, power and other waveform data of the appliances. For many appliances, when some of their components' physical characteristics have been changed, the current and power waveform would appear distorted, and the transient, steady-state features extracted and some other estimation parameters got from the power waveform may change. So, we can carry out fault diagnosis according to the situation. For example, [66]shows that those refrigeration controllers whose regulating performance are bad may appear big swings for their power signals. Based on the envelop analysis of power signals, the corresponding relationship between diagnostic parameters and equipment health is further established, and a non-intrusive fault diagnosis for motor load is achieved [67, 68, 69].

4.3 Demand response

Load disaggregation enables electric power companies to understand the working characteristics, power consumption rules of different load classes and even single electric equipment and current electric power and controllable capacity. Those information can help power companies to formulate dynamic price and demand response incentive policies more scientifically. In addition, the analysis and processing of the power consumption detail monitoring results can be used to adjust, improve and scientifically evaluate the energy efficiency projects in power companies.

Under normal circumstances, the dynamic price or the demand response can motivate the user's friendly interaction with power companies, which can achieve load peak shifting and shaving, improve the load coefficient, thereby improving the economic efficiency of the power system operation, (such as energy saving, prolonging the service life and reducing the operating and maintenance costs of power equipment.) improving the utilization rate of short-term power assets, and postponing long-term infrastructure investment. In case of emergency, if the power grid is under overload or low voltage, emergency load rejection can be realized through demand response protocol, thereby improving system stability, preventing system collapse, and ensuring the safe operation of the power system.

And like electric water heater, air conditioner, heater, refrigerator, washing machine, vacuum cleaner and electric kettle etc. with load can be shifted in the power system, which not only can quickly respond to the peak load demand of the power grid, but also have little influence on the users with power supply interruption in a short time. So they are the main force when the users interacting with the power grid and responding the demand of power grid, and the power system implementing demand management.

5 Conclusions and future work

Smart meters today can recorded the power consumption of the related house at a high time resolution and delivers these recorded data to a utility company or institution. This is very good for monitoring the smart grid and then control the power grid better. It also helpful for users electricity conservation by adjusting electricity usage time.

Not only are smart meters conducive for the management of the power grid, but also the high resolution of meter readings can be utilized for exposing customers' activities in house. Therefore, this chapter mainly introduced a NILM approach for mining customers' privacy without intruding into customer residence, i.e. install no device or embed no Trojan virus in the house. Specifically, the core idea was to disaggregate the incessant readings for indicating the consumption of every electricity appliance by designing the energy consumption learning and appliance inference algorithms, thereby exposing customers' daily activities. The existing energy consumption learning and appliance inference algorithms were often divided into two categories which are the supervised and unsupervised algorithms. These algorithms aimed to classify or cluster the working states of the electricity appliances, in order to predict customers' activities.

As discussed before, since no special hardware or software is assumed to be installed or run in the customer residences, the working state of each appliance at each time point is a completely black box. So the only way to analyze the customers' activities has to disaggregate the meter readings directly. This makes the result of the NILM approach not accurate when the number of the appliances are large. Hence, how to improve the precision of the inference and enlarge the number of the electricity appliances are two research directions for the NILM apporach from the viewpoint of privacy mining.

From the viewpoint of privacy preserving, consumption patterns can be identified through the high-frequency data, creating considerable risks in customers' privacy, especially in the United States and European countries. Several cities like Capitola and Ross in California, have already begun to refuse the installation of smart meters, and the European Commission's Article 29 Working Party also strongly advocates privacy enhancing technologies in smart grid. The need and urgency for such technologies is exacerbated by the fact that more than 8 million smart meters have already been installed in the United States. Similarly, European Parliament's directive requires 80% smart meter adoption in all European households by 2020, and 100% by 2022. Worse, the electricity energy is not abstract data. Therefore, adversaries could secretly install their own smart meters outside the residences to record the energy consumption. This type of attacks cannot be resisted against by the traditional cryptographic primitives. Hence, developing new technologies for preserving customers' privacy without having bad impact on the management of the smart grid is a potential research direction.

References

1. Esa N F, Abdullah M P, Hassan M Y, A review disaggregation method in Non-intrusive Appliance Load Monitoring, *Renewable & Sustainable Energy Reviews*, 2016, 66:163–173.
2. Faustine A, Mvungi N H, Kaijage S, et al. A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem[J]. 2017.
3. Zoha A, Gluhak A, Imran M A, et al. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey[J]. *Sensors*, 2012, 12(12):16838.
4. Jiang X, Dawson-Haggerty S, Dutta P, et al. Design and implementation of a high-fidelity AC metering network, *International Conference on Information Processing in Sensor Networks. IEEE*, 2009:253–264.
5. Suzuki K, Inagaki S, Suzuki T, et al. Nonintrusive appliance load monitoring based on integer programming, *Sice Conference. IEEE*, 2008:2742–2747.
6. Ridi A, Gisler C, Hennebert J, A Survey on Intrusive Load Monitoring for Appliance Recognition, *International Conference on Pattern Recognition. IEEE Computer Society*, 2014:3702–3707.
7. Hart G W, Nonintrusive appliance load monitoring, *Proceedings of the IEEE*, 1992, 80(12):1870–1891.
8. Kolter J Z, Recent advances in algorithms for energy disaggregation, In : *BECC Conference*, 2011.

9. Tsai M S, Lin Y H, Development of a non-intrusive monitoring technique for appliance' identification in electricity energy management, *International Conference on Advanced Power System Automation and Protection. IEEE*, 2012:108–113.
10. Adabi A, Mantey P, Holmegaard E, et al. Status and challenges of residential and industrial non-intrusive load monitoring, *Technologies for Sustainability. IEEE*, 2015:181–188.
11. Zeifman M, Roth K, Nonintrusive appliance load monitoring: Review and outlook, *IEEE Transactions on Consumer Electronics*, 2011, 57(1):76–84.
12. Belley C, Gaboury S, Bouchard B, et al. An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances, *Pervasive & Mobile Computing*, 2014, 12(3):58–78.
13. Zoha A, Gluhak A, Imran M A, et al. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey, *Sensors*, 2012, 12(12):16838–16866.
14. Anderson K, Ocneanu A, Benitez D, et al. BLUED : A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research. In *Proceeding of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, p:1–5, 2012.
15. Kolter J Z and Johnson M J, REDD: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA, volume 25, pages 59–62. Citeseer, 2011.
16. Hart G W, Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
17. Norford L K, Leeb S B, Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms, *Energy & Buildings*, 1995, 24(1):51–64.
18. Farinaccio L, Zmeureanu R, Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy & Buildings*, 1999, 30, 245–259.
19. Marceau M L, Zmeureanu R, Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. *Energy Conversion & Management*, 2000, 41, 1389–1403.
20. 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*, 4–6 July 2004; pp. 1–5.
21. Lam H, Fung G, Lee W, A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signaturesof. *IEEE Transactions on Consumer Electronics*, 2007, 53(2):653–660.
22. Madden S, Franklin M J, Hellerstein J M, et al. TAG:a Tiny AGgregation service for ad-hoc sensor networks. *Acm Sigops Operating Systems Review*, 2002, 36(SI):131–146.
23. 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*, 26–29 September 2010; pp. 139–148.
24. Marchiori A, Hakkarinen D, Han Q, et al. Circuit-Level Load Monitoring for Household Energy Management, *IEEE Pervasive Computing*, 2011, 10(1):40–48.
25. Liang J, Ng S K K, Kendall G, et al. Load Signature Study–Part I: Basic Concept, Structure, and Methodology. *IEEE Transactions on Power Delivery*, 2010, 25(2):551–560.
26. Najmeddine H, Drissi K E K, Pasquier C, Faure C, Kerroum K, Diop A, Jouannet T, Michou M, State of Art on Load Monitoring Methods. *In Proceedings of the 2nd IEEE International Conference on Power and Energy Conference*, Johor Bahru, Malaysia, 1–3 December 2008; pp. 1256–1258.
27. Kato T, Cho H S, Lee D, Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments. *In Proceedings of the 7th International Conference on Smart Homes and Health Telematics*, Tours, France, 1–3 July 2009; Volume 5597, pp. 150–157.
28. Cole A, Albicki A, Nonintrusive Identification of Electrical Loads in a Three-Phase Environment Based on Harmonic Content. *In Proceedings of Instrumentation and Measurement Technology Conference*, Baltimore, MD, USA, 1–4 May 2000; Volume 716, pp. 24–29.

29. Suzuki K, Inagaki S, Suzuki T, Nakamura H, Ito K, Nonintrusive Appliance Load Monitoring Based on Integer Programming. *In Proceedings of SICE Annual Conference*, Tokyo, Japan, 20–22 August 2008; Volume 174, pp. 2742–2747.
30. Laughman C, Lee K, Cox R, Shaw S, Leeb S, Norford L, Armstrong P, Power signature analysis. IEEE Power&Energy Magazine. 2003, 1, 56–63.
31. Li J, West S, Platt G, Power Decomposition Based on SVM Regression. *In Proceedings of International Conference on Modelling, Identification Control*, Wuhan, China, 24–26 June 2012; pp. 1195–1199.
32. Schoofs A, Guerrieri A, Delaney D, O'Hare G, Ruzzelli A, ANNOT: Automated Electricity Data Annotation Using Wireless Sensor Networks. *In Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks*, Boston, MA, USA, 21–25 June 2010; pp. 1–9.
33. Rowe A, Berges M, Rajkumar R, Contactless Sensing of Appliance State Transitions through Variations in Electromagnetic Fields. *In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, Zurich, Switzerland, 3–5 November 2010; pp. 19–24.
34. 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, 16–19 September 2007; pp. 271–288.
35. Zeifman M, Roth K. Nonintrusive appliance load monitoring: Review and outlook, *IEEE Transactions on Consumer Electronics*, 2011, 57(1):76–84.
36. 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*, 2007, 5236:664–674.
37. Shaw S R, Leeb S B, Norford L K, et al. Nonintrusive Load Monitoring and Diagnostics in Power Systems. *IEEE Transactions on Instrumentation & Measurement*, 2008, 57(7):1445–1454.
38. Cole A I, Albicki A, Data Extraction for Effective Non-Intrusive Identification of Residential Power Loads, *In Proceedings of Instrumentation and Measurement Technology Conference*, St. Paul, MN, USA, 18–21 May 1998; Volume 2, pp. 812–815.
39. Hazas M, Friday A, Scott J, Look back before leaping forward: Four decades of domestic energy inquiry. *IEEE Pervas. Comput.* 2011, 10, 13–19.
40. Bonfigli R, Squartini S, Fagiani M, et al. Unsupervised algorithms for non-intrusive load monitoring: An up–to–date overview, *International Conference on Environment and Electrical Engineering. IEEE*, 2015:1175–1180.
41. Anderson K D, Berges M E, Ocneanu A, et al. Event detection for Non Intrusive load monitoring, *Conference on IEEE Industrial Electronics Society. IEEE*, 2012:3312–3317.
42. Makonin S, Popowich F, Bajic I V, et al. Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring[J]. IEEE Transactions on Smart Grid, 2016, 7(6):2575–2585.
43. Parson O, Ghosh S, Weal M, et al. Non–intrusive load monitoring using prior models of general appliance types, *Twenty-Sixth AAAI Conference on Artificial Intelligence. AAAI Press*, 2012:356–362.
44. Zhao B, Stankovic L, Stankovic V. On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing. *IEEE Access*, 2017, 4:1784–1799.
45. Jia R, Gao Y, Spanos C J, A fully unsupervised non-intrusive load monitoring framework, *IEEE International Conference on Smart Grid Communications. IEEE*, 2015:872–878.
46. Fukushima K, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 1980, 36(4):193–202.
47. Atlas L E, Homma T, Ii R J M, An Artificial Neural Network for Spatio-Temporal Bipolar Patterns: Application to Phoneme Classification. *Neural Information Processing Systems*, 1987:31–40.

48. LeCun Y, Bottou L, Bengio Y, and Haffner P, Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
49. Kelly J and Knottenbelt W, The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from uk homes. *Scientific Data*, 2(150007), 2015. doi:10.1038/sdata.2015.7.
50. Batra N, Kelly J, Parson O, Dutta H, Knottenbelt W, Rogers A, Singh A, and Srivastava M, NILMTK: An open source toolkit for non-intrusive load monitoring. *In Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014. doi:10.1145/2602044.2602051.
51. Werbos P J, Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.
52. Dongshu W, Jialing H, Mussadiq A R, Zijian Z, and Liehuang Z, An Efficient Sparse Coding-based Data-mining Scheme in Smart Grid. MSN 2017 accepted.
53. Elhamifar E, Sastry S, Energy Disaggregation via Learning Powerlets and Sparse Coding. *AAAI*. 2015: 629–635.
54. Lai Y X, Lai C F, Huang Y M, Chao H C, Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. *Information and Sciences*. 2012, doi:10.1016/ j.ins.2012.10.002.
55. Bao C, Ji H, Quan Y, et al. Dictionary learning for sparse coding: Algorithms and convergence analysis. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(7): 1356–1369.
56. Guvensan M A, Taysi Z C, Melodia T, Energy monitoring in residential spaces with audio sensor nodes: TinyEARS. *Ad Hoc Networks 2012*, doi: 10.1016/j.adhoc.2012.10.002.
57. Yoo J, Park B, Hur K, Context Awareness-Based Disaggregation of Residential Load Consumption. *In Proceedings of the 18th International Federation of Automatic Control (IFAC) World Congress*, Milano, Italy, 28 August–2 September 2011; pp. 13691–13695.
58. Berges M, Rowe A, Poster Abstract: Appliance Classification and Energy Management Using Multi-Modal Sensing. *In Proceedings of the 3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, Seattle, WA, USA, 1 November 2011.
59. Anderson K, Ocneanu A, Benitez D, Carlson D, Rowe A, Berges M, BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research. *In Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability*, Beijing, China, 12–16 August 2012.
60. Saitoh T, Osaki T, Konishi R, et al. Current Sensor Based Home Appliance and State of Appliance Recognition. *Sice Journal of Control Measurement & System Integration*, 2010, 3(2):86–93.
61. Lin G Y, Lee S C, Hsu Y J, et al. Applying power meters for appliance recognition on the electric panel, *Industrial Electronics and Applications. IEEE*, 2010:2254–2259.
62. Shao H, Marwah M, Ramakrishnan N, A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation. *In Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring*, Pittsburgh, PA, USA, 7 May 2012.
63. Kelly J, Knottenbelt W, Neural NILM: Deep neural networks applied to energy disaggregation, *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015: 55–64.
64. Cheng X, Li L, Wu H, Ding Y, Song Y, Sun W (2016), A survey of the research on non-intrusive load monitoring and disaggregation, 40. 3108-3117. 10.13335/j.1000-3673.pst.2016.10.026.
65. Batra N, Parson O, Berges M, et al. A comparison of non-intrusive load monitoring methods for commercial and residential buildings [J/OL]. 2014-08-27[2015]. http://arxiv.org/abs/1408.6595.
66. 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 Buildings*, 1996, 24(1):51–64.

67. Shaw S R, Leeb S B, Norford L K, et al. Nonintrusive load monitoring and diagnostics in power systems, *IEEE Transactions on Instrumentation and Measurement*, 2008, 57(7):1445–1454.
68. Orji U, Remscrim Z, Laughman C, et al. Fault detection and diagnostics for non-intrusive monitoring using motor harmonics, *Applied Power Electronics Conference and Exposition*. Palm Springs, CA:IEEE, 2010:1547–1554.
69. Yan R, Gao R X, Energy-based feature extraction for defect diagnosis in rotary machines, *IEEE Transactions on Instrumentation and Measurement*, 2009, 58(9):3130–3139.