Non-intrusive load monitoring algorithm based on household electricity use habits

Bo Yin^{1,2} · Zhenhuan Li¹ · Jiali Xu² · Lin Li¹ · Xinghai Yang³ · Zehua Du¹

Received: 18 October 2020 / Accepted: 24 April 2021
© The Author(s) 2021

Abstract

The construction of smart grid is an important part of improving the utilization rate of electric energy. As an important way for the construction of smart grid, non-intrusive load decomposition methods have been extensively studied. In this type of method, limited by transmission cost and network bandwidth, low-frequency data has been widely used in practical applications. However, the accuracy of device identification in this case faces challenges. Due to the relatively single characteristics of low-frequency data, it is difficult to express the operating status of complex electrical appliances, resulting in low decomposition performance. In this paper, a non-intrusive load is proposed based on household electrical habits by studying the relationship between household electricity consumption habits and load status decomposition method. The Gaussian mixture model and time information are used to model the probability distribution of the electrical appliance state. This probability distribution is then used as the observation probability distribution of the factor hidden Markov model. In such a way, the BH-FHMM model is proposed. Finally, load decomposition is carried out through the load decomposition process of the FHMM model. In order to verify the performance of the proposed method, an experimental comparison is conducted based on the REDD data set. According to the results, a significant improvement in equipment recognition accuracy is obtained.

Keywords Factorial hidden Markov model · Gaussian mixture model · Non-intrusive load monitoring · Household electricity habits

1 Introduction

Energy is essential for human society while the global economy actually run on energy. Every major progress of human civilization is accompanied by improvement and replacement of energy. The development of energy has greatly promoted the development of the world economy and human society. However, with the rapid development of the economy, energy problems are imminent. The

excessive energy consumption and the large-scale use of electrical appliances call for smart grid [1].

To cope with the above problems, the power industry, which mainly uses fossil fuels, not only needs to replace traditional power generation methods with more environmentally friendly technologies (such as geothermal energy, hydropower, etc.), but also needs to use electricity more efficiently. Therefore, improving consumption management on the supply side and increasing energy conservation awareness and capabilities on the demand side have become important measures to deal with the energy crisis and global warming. A large part of electricity consumption comes from household electricity and other building electricity. According to statistics from relevant US departments, almost 70% of electricity consumption comes from building electricity [1]. Therefore, reducing household electricity consumption is of great significance for energy saving and emission reduction.

✉ Zehua Du
duzehua@stu.ouc.edu.cn

¹ College of Information Science and Engineering, Ocean University of China, Qingdao, China

² Pilot National Laboratory for Marine Science and Technology, Qingdao, China

³ School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, China

With the development of intelligent technologies (big data, intelligent optimization control, intelligent decision support, etc.), the development of the smart grid has been accelerated to improve the utilization rate of electric energy and to provide high-quality electric services. At present, with the development of informatization, networking, and virtualization technologies, the coverage of smart meters is getting higher and higher, providing abundant electricity consumption data for the grid. How to effectively use these data has become key challenge in smart grid research.

Load monitoring, also known as non-intrusive load monitoring, was first proposed to monitor and decompose internal electricity usage information based on load information at power access points [2]. In this method, the electricity data of home bus is used to obtain the electricity consumption data of various appliances in the home. Accurate and efficient load monitoring can provide strong data support for power grid decision-making, proper power usage guidance, and implement accurate load forecasts to improve the understanding of power consumption. An improved power plan can then be formulated to reduce the waste of power resources [3]. Besides, load monitoring technology is also of great significance for fault detection and user behavior monitoring.

Non-intrusive load monitoring methods can be briefly divided into event-based and non-event based types. In event-based non-intrusive load monitoring methods, KNN [4], Naive Bayes [5], neural networks [6], and state vector machines (SVMs) [5] are normally used to process high-frequency power data and to identify electrical switches based on event detection. The accuracy of the load monitoring method is very high, and the recognition accuracy of a single electrical appliance can reach more than 90%. However, this method requires training on various events in the circuit, which leads to the poor classification effect when multiple electrical appliances change at the same time. Meanwhile, these load identification methods are usually based on transient characteristics, which require high collection equipment and a large amount of high-frequency power data. Besides, the classification effect of multi-state electrical appliances is poor. Each state must be treated as a separate category, and the classification effect of resistance-type electrical appliances with similar power is not good. Therefore, research on non-event-based load monitoring algorithms has gradually developed. In non-event based non-intrusive load monitoring methods, the properties of the Markov chains in hidden Markov models are determined to effectively use low-frequency power data. The HMM model of each electrical appliance is obtained through the data of each labeled electrical appliance and then is expanded into an FHMM model. After that, the convex optimization method is introduced to

obtain the optimal solution. This type of method has fast training speed and improves the performance of the load identification for overlapping events. Because of the advantages related to hardware requirements, non-event based methods are widely used to construct smart grids. However, compared to that for event-based methods, the identification accuracy provided by non-event based methods is lower because of the poor usage of low-frequency features, especially when there are many types of appliances. In this paper, an improved method called BH-FHMM is proposed to optimize the traditional factorial hidden Markov model (FHMM) by introducing household electricity habits. In this method, a probability distribution of time series information for electric appliances is used to replace the Gaussian distribution in the FHMM, and a Gaussian mixed model (GMM) is applied. Moreover, several mean values are introduced as load distribution values to reduce the error between the predicted power and actual power.

The main contributions of this paper are as follows. (1) A non-intrusive load decomposition method called humans-factorial hidden Markov model (BH-FHMM) is proposed based on users' electricity consumption habits to improve the identification accuracy of the non-event based methods. (2) Household electricity consumption habits are introduced as additional features to extend the traditional low-frequency feature set. Then, the resulting data set is fused with time series information, and a GMM is used to model the probability distribution of time series information for electric appliances. (3) The mean values of different Gaussian distributions are used as the load distribution values for different time periods to reduce the error between the predicted power and actual power.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related work. Section 3 discusses the FHMM model based on household electricity habit improvements. Section 4 presents our results based on power load data from actual homes. Section 5 summarizes the study.

2 Related work

Load monitoring was first proposed by Hart [7]. The non-intrusive load monitoring method, which uses a finite state approach to describe the state transitions of electric appliances, was then proposed based on steady-state power variation research at power access points. Non-intrusive load decomposition technology, also known as load decomposition technology, refers to the technology that decomposes the data of various electrical appliances in the home through the data on the bus. Compared with similar methods, the NILM method only needs to detect the total

circuit, and has a relatively lower cost. Moreover, some of them can directly use the data of existing smart meters to perform load decomposition, which greatly reduces the implementation difficulty of load decomposition and increases the monitoring range. Therefore, with the development of simple, economical and easy-to-use methods of data collection, non-intrusive load monitoring (NILM) has quickly become a research hotspot. This method can be used for conventional embedded system development and SoC embedded systems with hardware acceleration functions. However, the power consumption of certain lighting types, informatic devices, and electronic equipment is ineffective.

In these methods, load monitoring performance is improved by using features extracted from a large amount of high-frequency data [8]. The advantage is that it can be applied in MEL monitoring and provide a viable non-intrusive alternative to the conventional methods of monitoring MEL power consumption and usage. Laughman et al. treated harmonics as a third dimension and added them to a classification model as extended features [9]. NILM is an ideal platform for extracting useful information about a system that includes electromechanical equipment. Few sensors are used, so the installation cost is low, and reliability is high. Leeb et al. used the first coefficient of short-term FFT as an extension of the harmonics to distinguish among variable-load devices [10]. The advantage of commercial NILM methods based on event detection algorithms is that they can be built with parallel processing architectures composed of inexpensive microprocessors and microcontrollers. Zoha adopted wavelet transform to represent the characteristics of load [11]. The above methods extract features through a large amount of high-frequency data, which effectively improves the accuracy of load decomposition. However, due to the needs for various special equipment and the large amounts of data, this method cannot be widely promoted.

Hence, the focus of research has shifted to low-frequency data with active power and reactive power components [12]. Data for these components are relatively easy to collect, especially for active power, for which direct collection can occur with electricity meters. The NILM method has many advantages over intrusive methods; however, the disadvantage of this approach is that it cannot be used for the high-precision monitoring of all equipment types.

Kim et al. [13] considered an existing FHMM and three new models (FHSM, CFHMM, and CFHSM), but these methods only work well for appliances with simple or modestly complex power signatures and perform poorly for devices with complex signatures. Bonfigli et al. [14] used active and reactive power information to establish a two-dimensional HMM. Then, the K-means method was used to

cluster the working status of each appliance. The drawback of this method is that it is difficult to obtain reactive power data in a practical environment.

Biansoongnern et al. [15] proposed a measurement system that can determine the power consumption of split-type air conditioning units and refrigerators in households and does not require the installation of instruments directly on appliances. A method based on segmented integer quadratic constraint programming (SIQCP) was proposed to disaggregate a household power profile at the appliance level. This method involves an iterative process; therefore, errors can be amplified as iterations proceed. Zeifman et al. [16] used mutually independent features to increase accuracy of model. However, the high false positive rate is also introduced. Overall, most researchers have focused on improving the decomposition accuracy by adding load data characteristics and using electrical characteristics to improve non-intrusive load monitoring algorithms.

Since deep learning has gradually developed, load decomposition techniques based on deep learning have become widely popular. Kelly et al. [17] first introduced deep learning into the field of load decomposition, using long-term memory recurrent neural networks, Autoencoders and rectangular networks that regress the start time, end time and average power demand of each appliance to perform load decomposition. The results show that this method is better than the traditional method for load decomposition and has good generalization performance. Since then, Bonfigli et al. [18] improved the Autoencoder model proposed by Kelly, using a brand-new network, and optimized for the noise environment, so that the network still has better performance under noise conditions. Krystalakos et al. [19] modified the recurrent neural network in Kelly's paper and proposed an online load decomposition method based on sliding windows. This method uses sliding windows to achieve online decomposition, but due to the use of recurrent neural networks in the process, therefore, the complexity of training is high. Based on previous methods and experience, Kaselimi et al. [20] proposed a multi-channel CNN load decomposition model, which uses active power, reactive power, apparent power and current as input, and considers previous electrical appliances. The state of it has reached a higher level.

Load decomposition can also provide consumers with accurate energy-saving recommendations. Fischer et al. [21] proposed an energy-related recommendation system based on household usage. The system compares the user's current energy prices with the prices in the market and provides recommendations. Experiments have shown that users who use the system have reduced their energy tariffs by 10%.

Load decomposition can improve the accuracy of energy demand forecasting. The load decomposition algorithm

allows grid operators to predict the corresponding energy demand better. Basu et al. [22] predict the use of electrical appliances in the future based on load decomposition and future weather information. Rao et al. [23] proposed a new method that can identify electrical activities and predict the future work status of equipment. This method uses a support vector machine based on edge analysis as a device identification model, and at the same time uses an autoregressive moving average model as a prediction model. As a result, the recognition rate and the estimated accuracy of future power consumption of electrical appliances reached 75% and 90%, respectively. NILM technology can also be used to formulate smart heating strategies. Spiegel et al. [24] used load decomposition to obtain the electric furnace usage of households and optimize the heating plan accordingly.

Load decomposition can be used to detect malfunctioning equipment. The refined power consumption of equipment can also be used to detect faulty electrical appliances in several households. Through load decomposition technology, information such as the operating cycle, power and current of electrical equipment can be obtained. Using this information, faulty electrical equipment can be detected.

Load decomposition can monitor households. The load decomposition technology can be used to monitor the occupants. The application in this area may involve personal privacy issues, but this technology can also have a very positive impact, especially in monitoring the health of the elderly. Belley et al. [25] proposed an activity recognition method based on NILM technology and applied it to a smart home system to simulate the daily situation of patients with Alzheimer's disease for testing. Using this NILM-based behavior monitoring method, the system can identify daily activities effectively with less investment and relatively limited data.

Load decomposition can provide information for the macro-control of the power sector. Detailed power consumption information can also help the power sector conduct macro-control better. Kong et al. [26] upgraded the AMI architecture based on a large amount of smart meter data obtained from the Australian AMI architecture. The power sector uses the rich data provided by smart meters to enhance interaction with customers and demand-side management.

Hui Liu et al. [27] proposed a method to estimate the energy consumption of appliances in a building. In the study, weighted current harmonic vectors were proposed to increase the weights of useful harmonics. However, the designed deep LSTM framework only considered unidirectional dependence, and real load disaggregation may involve bidirectional dependence. Min Xia et al. [28] proposed a composite deep LSTM for load disaggregation.

The proposed method reduced the disaggregation complexity and improved the efficiency of disaggregation. However, in experiments, some electrical appliances were in the “off” state for a long time, and accurate disaggregation was difficult when they entered the “on” state of power consumption.

3 Proposed methods

3.1 HMM and FHMM algorithms

Hidden Markov model is a statistical model about time series. The model is a double random process. It describes a random sequence of unobservable states generated randomly by Hidden Markov chain, and then an observation value is generated by each state to generate a random sequence of the process. Among them, the Markov chain describes a sequence of states, and the value of each state depends on the previous finite states. The sequence of states randomly generated by the Hidden Markov chain is called the state sequence, and the state is not observed. Each unobservable state generates an observation, and the resulting random sequence of observations is called the observation sequence [29]. Each position in the sequence can be regarded as a moment. Hidden Markov chain can be indirectly described through a probabilistic model of a time sequence, such as an HMM, as shown in Fig. 1.

The formal definition of Hidden Markov can be defined as:

$$\lambda = (Q, V, \pi, A, B) \quad (1)$$

It includes two state sets and three probability distributions. The specific definition of each parameter is as follows:

- (1) Q is the set of all possible unobservable hidden states, it can be defined as $Q = \{q_1, q_2, q_3 \dots q_I\}$, where I is the number of all possible states.
- (2) V is the set of all possible observations, it can be defined as $V = \{v_1, v_2, v_3 \dots v_M\}$, where M is the number of all possible observations.

The state sequence S is a set with a value space of Q , and the observation sequence O is a set with a value space of V . Figure 3-1 shows a hidden Markov

Fig. 1 Diagram of a hidden Markov model

model, where $S = \{s_1, s_2, s_3, \dots, s_T\}$ represents state sequence, $O = \{o_1, o_2, o_3, \dots, o_T\}$ represents the observation sequence, T is sequence length.

- (3) π is the initial probability distribution, define $\pi = \{\pi_1, \pi_2, \pi_3, \dots, \pi_I\}$, it indicates that at the initial time point $t = 1$. Probability of appearance of all elements in Q , which is $\pi_i = P(s_1 = q_i), i = (1, 2, 3, \dots, I)$. Obviously $\sum_{i=1}^I \pi_i = 1$. In many load decomposition problems, a uniform distribution is usually used to represent the initial probability distribution.
- (4) A is the state transition distribution, the general form is a state transition matrix $A = [a_{ij}]_{I \times I}$, where a_{ij} refers to the probability that the state is q_j at time $t + 1$ under the condition that the state at time t is q_i , it's expression is $a_{ij} = P(s_{t+1} = q_j | s_t = q_i), 1 \leq i, j \leq I$.
- (5) B is the observed probability distribution. The general form is a probability matrix $B = [b_j(k)]_{I \times M}$, where $b(k)$ refers to the probability that the observed state is v_k when the state is q_j at time t , it's formula is $b(k) = P(o_t = v_k | s_t = q_j), k = 1, 2, \dots, M; j = 1, 2, \dots, I$. In practical problems, when the observation probability is continuously different values, Gaussian distribution is usually used as the observation probability distribution. When the observed value is a few finite values, it is expressed in matrix form.

It can be seen from the above analysis that the initial probability distribution π and state transition distribution A determine the random process of the hidden Markov chain, and the observation probability distribution B determines the random process of generating observations from the state. The three parameters jointly determine the hidden Markov model. Therefore, the general form of the hidden Markov model can also be represented by $\lambda = (\pi, A, B)$.

It can be seen from the definition of the model that the hidden Markov model relies on two basic assumptions:

- (1) Homogeneous Markov assumes that at a certain time t , the state s_t only depends on its state at the previous moment s_{t-1} . It has nothing to do with any other moments in the sequence, and it has nothing to do with the moment t itself.

$$P(s_t | s_{t-1}, o_{t-1}, \dots, s_1, o_1) = P(s_t | s_{t-1}) \quad (2)$$

- (2) Observation independence hypothesis, that is, the observation o_t of the model output at time t depends only on the current state s_t , has nothing to do with other states, and does not depend on the time t itself.

$$P(o_t | s_t, s_{t-1}, o_{t-1}, \dots, s_1, o_1) = P(o_t | s_t) \quad (3)$$

Hidden Markov model, as a probability model, is mainly used to solve three basic problems.

- (3) Probability calculation problem, that is, the model $\lambda = (\pi, A, B)$ and observation sequence $O = \{o_1, o_2, o_3, \dots, o_T\}$ are known, and the probability $P(O | \lambda)$ of the observation sequence O appearing under the model is calculated.
- (4) Learning problems, that is, given enough observation data $O = \{o_1, o_2, o_3, \dots, o_T\}$, use the observation data to estimate the parameters of the model $\lambda = (\pi, A, B)$, so that the probability $P(O | \lambda)$ of using the model to calculate under these parameters is maximized.
- (5) Decoding problem, that is, given the observation data $O = \{o_1, o_2, o_3, \dots, o_T\}$ and model parameters $\lambda = (\pi, A, B)$, finding the corresponding state sequence $S = \{s_1, s_2, s_3, \dots, s_T\}$ maximizes the conditional probability $P(O | S)$.

Each basic problem has its own specific solution method, the following this thesis will introduce these solutions in turn:

- (1) For probability calculation problems, forward-backward algorithms are usually used.

Forward algorithm is a dynamic programming algorithm that defines the local state of dynamic programming by defining forward probability. The forward probability is defined as: the probability that the state q_i at time t is and the observation sequence $o_1, o_2, o_3, \dots, o_t$ is, that is

$$\alpha(i) = P(o_1, o_2, o_3, \dots, o_t, s_t = q_i | \lambda) \quad (4)$$

The steps of the forward algorithm are as follows:

Step 1 Calculate the initial value

$$\alpha_1(i) = \pi_i b_i(o_1), 1 \leq i \leq I \quad (5)$$

Step 2 Recursive calculation, for $t = 2, 3, \dots, T$,

$$\alpha_t(i) = \left[\sum_{j=1}^I \alpha_{t-1}(j) a_{ji} \right] b_i(o_t) \quad (6)$$

Step 3 Get the final result

$$P(O | \lambda) = \sum_{i=1}^I \alpha_T(i) \quad (7)$$

The backward algorithm is similar to the forward algorithm. In the forward algorithm, the first two steps are mainly to calculate the forward probability, and the final probability $P(O | \lambda)$ is the sum of the final forward variables. The backward probability can also be defined as:

$$\beta(i) = P(o_{t+1}, o_{t+2}, o_{t+3}, \dots, o_T, s_t = q_i | \lambda) \quad (8)$$

The steps of the backward algorithm are as

follows:

Step 1 Calculate the initial value

$$\beta_T(i) = 1, 1 \leq i \leq I \quad (9)$$

Step 2 Recursively calculate from back to front, for $t = T - 1, T - 1, \dots, 1$

$$\beta_t(i) = \sum_{j=1}^I a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), i = 1, 2, \dots, I \quad (10)$$

Step 3 Get the final result

$$P(O|\lambda) = \sum_{i=1}^I \pi_i b_i(o_1) \beta_1(i) \quad (11)$$

- (2) Learning problems, also known as parameter training problems, are divided into supervised and unsupervised learning algorithms. The data of the supervised algorithm needs to have the observation sequence and the corresponding state sequence, while the unsupervised algorithm only needs the observation sequence.

The training process of the supervised algorithm is relatively simple, which can be obtained by simple statistical methods. The initial state probability π_i is the frequency of each state in the training data in the first position.

Suppose the frequency of state i transition to state j in the training data is A_{ij} , then the state transition probability is

$$a_{ij} = \frac{A_{ij}}{\sum_{j=1}^I A_{ij}}, i = 1, 2, \dots, I; j = 1, 2, \dots, I \quad (12)$$

Suppose the frequency of state i in the training data and observation k is B_{jk} , then the observation probability

$$b_i(k) = \frac{B_{jk}}{\sum_{k=1}^M B_{jk}}, i = 1, 2, \dots, I; k = 1, 2, \dots, M \quad (13)$$

The cost of supervised training is too expensive, so the more common way to train Hidden Markov models is the unsupervised method, which only uses observation data to estimate model parameters, and the cost is relatively low. The main algorithm used is Baumwell Odd algorithm. The algorithm is essentially an expectation maximization (expectation maximization, EM) algorithm. The EM algorithm starts from the initial parameter $\lambda = (\pi^{(0)} A^{(0)} B^{(0)})$, iterates continuously, and uses the forward-backward algorithm to evaluate until the probability is no longer significantly improved. The iterative process includes the following steps:

Step 1 Calculate the EM algorithm Step E: Calculate the expectation of the log likelihood function

$$Q(\lambda, \lambda^k) = \sum_I \log P(O, I|\lambda) P(O, I|\lambda^k) \quad (14)$$

Among them, $P(O, I|\lambda)$ is the log-likelihood function of the model, λ^k is the estimated value in the first k round, and λ is the model parameter to be maximized. According to the definition of Hidden Markov model, formula (14) can be written as:

$$\begin{aligned} Q(\lambda, \lambda^k) = & \sum_I \log \pi_i P(O, I|\lambda^k) \\ & + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t, i_{t+1}} \right) P(O, I|\lambda^k) \\ & + \sum_I \left(\sum_{t=1}^{T-1} \log b_{i_t}(o_t) \right) P(O, I|\lambda^k) \end{aligned} \quad (15)$$

Step 2 Calculate the EM algorithm Step M: Maximize the function $Q(\lambda, \lambda^k)$, find the model parameters π, A, B , and finally get:

$$\pi_i = \gamma_1(i) \quad (16)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (17)$$

$$b_j(k) = \frac{\sum_{t=1, o_t=v_k}^T \gamma_t(j)}{\sum_t \gamma_t(j)} \quad (18)$$

Among them, $\gamma_t(i)$ is the probability that the state of the model is q_i at time t , $\xi_t(i, j)$ is the probability that the state is q_i at time t and the state is q_j at time $t + 1$, and the algorithm is from front to back:

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^I \alpha_t(j) \beta_t(j)} \quad (19)$$

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^I \sum_{j=1}^I \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \quad (20)$$

Step 3, keep iterating until $P(O|\lambda)$ is no more obvious improvement.

- (3) The decoding problem can be solved by the Viterbi algorithm [30]. The Viterbi algorithm uses the idea of dynamic programming to reduce the computational cost of the decoding problem. The specific steps are:

Step 1 Initialize $t = 1$,

$$\delta(i) = \pi_i b_i(o_1), \varphi_1(i) = 0 \quad (21)$$

Step 2 Recursion for $t = 2, 3 \dots T$

$$\delta_t(i) = \max_{1 \leq j \leq N} \delta_{t-1}(j) a_{ji} b_i(o_t) \quad (22)$$

$$\varphi_t(i) = \operatorname{argmax}_{1 \leq j \leq N} \delta_{t-1}(j) a_{ji} \quad (23)$$

Step 3 Termination

$$P^* = \max_{1 \leq i \leq N} \delta_T(i) \quad (24)$$

$$s^* = \operatorname{argmax}_{1 \leq i \leq N} \delta_T(i) \quad (25)$$

Step 4 The optimal path backtracking, for $t = T-1, T-2 \dots 0.1$

$$s_t^* = \varphi_{t+1}(s_{t+1}^*) \quad (26)$$

Find the optimal path $S^* = (s_1^*, s_2^* \dots s_T^*)$.

Among them, $\delta_t(i)$ represents the maximum probability of all single paths whose state is q_i at time t , and $\varphi_t(i)$ represents the $t-1$ th node of the path with the highest probability among all single paths whose state is q_i at time t .

3.2 Factorial hidden Markov model

For smart grids with a single appliance, an HMM could be used to precisely build a corresponding model. The actual state of an electric appliance $S\{S_1, S_2, \dots, S_T\}$ can be represented by an implicit sequence of states. Additionally, the actual observed power $O\{O_1, O_2, \dots, O_T\}$ is an observation sequence. However, in a multi-appliance models built based on HMMs, the state o_t^i of a single appliance is unobservable. Moreover, the complexity of a total power model may increase exponentially. In order to solve these problems, FHMM is introduced here. The factorial hidden Markov model extends the structure of the hidden Markov model. There are multiple Markov chains in this model,

Fig. 2 Diagram of a factorial hidden Markov model

and each layer is an independent random process, but the observation state at each time point is the aggregate function value of the observations produced by each hidden state. The model is shown in Fig. 2. The commonly used aggregate functions in NILM problems include additive functions and different functions. Here, the additive functions are used. The observation value is then the sum of the observations of each Markov chain.

The factor hidden Markov model can also be expressed in the form of $\lambda = (\pi, A, B)$, but the three parameters all depend on multiple Markov chains. Assuming that N Markov chains are included, then:

π represents the probability distribution of the initial state: $P\{s_1^1, s_1^2, \dots, s_1^N\}$.

A is the state transition probability distribution: $P(s_t^i | s_{t-1}^i) = P(s_t^1, s_t^2, \dots, s_t^N | s_{t-1}^1, s_{t-1}^2, \dots, s_{t-1}^N)$.

B represents the output probability distribution from state to observation: $P(o_t | s_t^1, s_t^2, \dots, s_t^N)$.

Since each Markov chain is independent of each other, the initial state probability distribution and state transition probability distribution of the model can be expressed as:

$$\pi = P(s_1^1, s_1^2, \dots, s_1^N) = \prod_{i=1}^N \pi^i \quad (27)$$

$$P(s_t^i | s_{t-1}^i) = \prod_{i=1}^N P(s_t^i | s_{t-1}^i) \quad (28)$$

For the additive factor hidden Markov model, the observation vector O can be defined as:

$$O = \sum_{i=1}^N o^i \quad (29)$$

The observation probability distribution can be defined as:

$$P(s_t | O_t) = \prod_{i=1}^N P(o_t^i | s_t^i) \quad (30)$$

It can be seen from the above formula that the factorial hidden Markov model uses multiple HMM models to represent the probability distribution. If only the HMM model is used for description, assuming that each layer has I states and there are N layers in total, then one $I^N \times I^N$ state transition matrix is needed. The factorial hidden Markov model only needs $N I \times I$ state transition matrices, which greatly saves the calculation cost.

Fig. 3 Load disaggregation framework

The load disaggregation framework of the FHMM is shown in Fig. 3. In addition, the disaggregation steps are as follows.

(1) Model parameter training.

This problem can be described as knowing the power data of each electrical appliance and obtaining the HMM model of each electrical appliance, which can be solved by using the learning problem of the HMM model introduced above. This process is unsupervised. The supervision mentioned in this article is for the FHMM model. This article uses the data of different appliances with appliance tags to train the HMM model. This process is unsupervised because the training data does not mark the status of the appliance. Due to the volatility of circuit load, the traditional FHMM load decomposition model usually uses Gaussian distribution to represent the observation probability distribution of electrical appliances. For a single appliance i , the observation probability distribution at time t can be expressed as

$$p(o_t^i | s_t^i) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(o_t^i - \mu^i)^2}{2\sigma^2}} \quad (31)$$

where σ^i is the variance of appliance i , and μ^i is the mean value of appliance i .

(2) Load state estimation

Based on the model established in the first step, the problem can be described as, in the T time period, for the total load containing N appliances, the model parameter λ and the total power $O(O_1, O_2, \dots, O_T)$ are known to find each the optimal state sequence $S(S_1, S_2, \dots, S_T)$ of the electrical appliance. This problem can be regarded as an optimization problem with the largest posterior probability, which is

$$\max_S P(S|O) \quad (32)$$

Using the independence between the Markov chains in the factorial hidden Markov model can be obtained:

$$P(S|O) = \prod_{i=1}^N P(s_1^i) \prod_{t=2}^T \prod_{i=1}^N P(s_t^i | s_{t-1}^i) \cdot \prod_{t=1}^T P(o_t | s_t) \quad (33)$$

After logarithmic simplification, it can be solved through the following iterative process:

$$Q_1(s_1) = \sum_{i=1}^N \lg P(s_1^i) + \lg P(o_1 | s_1^1, s_1^2, \dots, s_1^N) \quad (34)$$

$$Q_t(s_t) = \max_{s_{t-1}} \{Q_{t-1}(s_{t-1}) + \sum_{i=1}^N \lg P(s_t^i | s_{t-1}^i)\} + \lg P(o_t | s_t) \quad (35)$$

Compared with the ordinary HMM model, the iterative process of the FHMM model must calculate the state transition process of all Markov chains at each step. The solution method is still a dynamic programming process, and the best path calculated in the previous step will be retained until the last moment T . At this time, according to the optimal state combination at time T , we can get all the states on the Markov chain.

(3) Load distribution

The load distribution problem can be described as the problem of solving the observation output o with a known state s , which can be constructed as an optimization function with the following objectives:

$$\max \prod_{t=1}^T \prod_{i=1}^N p(o_t^i | s_t^i) \quad (36)$$

Meet the following constraints:

$$o_t = \sum_{i=1}^N o_t^i, o_t^i \geq 0 \quad (37)$$

When the observation probability is represented by Gaussian distribution, the objective optimization function is equivalent to:

$$\min \sum_{t=1}^T \sum_{i=1}^N \frac{1}{2\sigma_i^2} (o_t^i - \mu_{s_t^i}^i)^2 \quad (38)$$

In practical applications, the average value of Gaussian distribution is usually used to represent the predicted power. Corresponding to time t , the power o_t^i of electrical appliance i is represented by the mean value of Gaussian distribution $\mu_{s_t^i}^i$ corresponding to s_t^i .

3.3 BH-FHMM algorithm

According to the analysis of household electricity habits, there is a strong relationship between human dietary habits and the operation status of electric appliances. Hence, the BH-FHMM algorithm is proposed in this paper. In this method, a GMM is used to obtain the observation probability distribution of the operation status of electric appliances; then, the operation status can be predicted.

3.3.1 Analysis of household electricity habits

In real life, the operating status of an appliance is related to household electricity habits. For example, the operation of dishwashers mainly occurs in certain time periods, as

Fig. 4 Proportion of dishwasher use in each hour

shown in Figs. 4 and 5. These time periods are related to human dietary habits.

Considering this phenomenon, 5 types of electric appliances with high power consumption in the REDD data set [31] were analyzed based on the following equation.

$$per = \frac{\text{num}(o_h^i)}{\text{num}(o^i)} (o > \delta) \quad (39)$$

where $\text{num}(o_h^i)$ represents the number of power values larger than a threshold δ during period h for electrical appliance i and $\text{num}(o^i)$ represents the number of power values larger than threshold δ during all periods for electrical appliance i .

The analysis results are shown in “Appendixes 1 and 2.” “Appendix 1” shows the proportion of the operating time for electrical appliances in different periods. It is obvious that electricity habits are different among households. “Appendix 2” shows the average power use of appliances in different periods. The power distribution corresponds to the proportion of the operation time. Thus, electricity habits can be used as inputs to improve the performance of the load monitoring model.

3.3.2 Time information model based on a GMM

The electrical HMM model consists of a state transition matrix A , an observation probability distribution matrix B , and an initial probability distribution π . Since the device always tends to maintain the current state, the state transition matrix A is usually close to the identity matrix [32]. Therefore, the focus of load decomposition research is the observation probability matrix B . Because appliances generally maintain a consistent state, the state transition matrix is normally similar to a unit matrix [33]. Thus, for load composition determination, the key process is to form an observation probability matrix. In this paper, the observation probabilities and the mean values at different times over 24 h are obtained by the GMM, as shown in Fig. 6. Thus, the electrical characteristics in different periods can be effectively fitted. The FHMM is composed of HMMs for individual electrical appliances. Hence, the

Fig. 5 Power distribution of dishwasher use over 24 h

Fig. 6 Diagram of a GMM with time information for a single appliance

framework of the BH-FHMM can be established by adding a GMM, as shown in Fig. 7.

On this basis, this paper adds the time information of household electricity consumption habits and improves the performance of the load monitoring model. In addition, to solve the model complexity problem caused by too many types of equipment, this paper improves the K-means algorithm to aggregate the data collected within 24 h. This method effectively reduces the number of unnecessary states generated after clustering the total power.

Earlier load decomposition algorithms usually use two states, on and off. And the decomposition result is the typical value of the two states. Although this method is effective for switch-type appliances such as electric kettles, it has great limitations for decomposing multi-state appliances (divided into finite-state appliances and continuous-change appliances). Due to the large variance of the operating power set of the appliances, the error between the decomposed operating power and the actual power is relatively large.

Kong proposed a k-means method based on an iterative determination of k value to cluster the state of electrical appliances. Here, the method is introduced to cluster the states of electrical appliances and to train the state transition matrix. The average power at different times is then calculated according to the time and power information obtained above. Hence, the number of states of the segment can be simplified. The specific steps are:

Step 1 Set the maximum number of states $K_{\max}$, the initial number of states K_0 and the state threshold δ for a single electrical appliance.

Step 2 Use the initial state number for clustering.

Step 3 Calculate the difference value of the clustering center, if the difference value is less than δ , the clustering ends, otherwise go to 4.

Step 4 The number of states $K = K + 1$, execute 1.

Clustering each electrical appliance using the above method can obtain the state number n of the electrical appliance, and then use it to train to obtain the state transition matrix A , and then compare the maximum power in different time periods with different clustering centers. Next, find the clustering center k_{tar}^i with the smallest difference between the two, and use the number of clustering centers less than or equal to the cluster center value as the number of states in the current time period to determine the number of Gaussian distributions that need to be trained, which is

$$n_h^i = \text{num}(k_j^i \leq k_{tar}^i) \quad (40)$$

where n_h^i represents the state number of the i th electrical appliance at time h , and k_j^i represents the j th cluster center of the i th electrical appliance.

For a certain electrical appliance i , the probability that the state is s_t^i at time t can be expressed as:

$$p(s_t^i) = p(s_t^i | s_{t-1}^i) p(o_t^i | s_t^i) \quad (41)$$

Because of the introduction of different Gaussian models in different periods, the GMM could be used to determine the observation probability associated with each state. The corresponding equation is as follows:

Fig. 7 Diagram of the BH-FHMM

$$p(o_t^i | s_t^i) = \sum_k \varphi_{hk} \frac{1}{\sqrt{2\sigma_k^2\pi}} e^{-\frac{(o_t^i - \mu_k)^2}{2\sigma_k^2}} \quad (h \leq 24, k \leq 24) \quad (42)$$

where σ is the Gaussian distribution variance, μ is the mean of the Gaussian distribution, φ is a 24×24 weight matrix, and h is the hour in which the timestamp occurs.

In addition, an improved K-means method is proposed based on reference [34]. This approach is used to determine the number of states of electric appliances, and the number of Gaussian distributions used for training can be verified, as shown in the equation below.

$$n_h^i = \text{num}(k_j^i \leq k_{tar}^i) \quad (43)$$

where n_h^i represents the number of states for the i th appliance at time h and k_j^i represents the j th clustering centre for the i th appliance.

4 Experiment

4.1 Experimental data set

The load decomposition algorithm proposed in this paper based on household electricity consumption habits uses low-frequency power data and adds non-traditional characteristics of household electricity consumption habits. This allows the algorithm to perform better when the amount of data is low. To verify the performance of this algorithm, the REDD data set is used as the data set of this experiment. The REDD data set is one of the most common used data sets in the field of load decomposition. The data set includes the electricity consumption data of six households in more than one month, as well as high-frequency data and low-frequency data. The sampling rate of high-frequency data is 16.5 kHz, which includes the sum of the all voltage of two households (house3 and house5). The

Fig. 8 nilm_metadata class diagram

waveform of the total current, and the low-frequency data includes the electricity consumption information of all households. The sampling rate of the data of each appliance is 1/3 Hz at the sampling rate of the total circuit of 1 Hz.

4.2 Data pre-processing

With the rapid development of load decomposition technology, different data sets have been introduced one after another. However, due to the lack of a unified structure, a cumbersome data pre-processing process is required when using these data sets for load decomposition. To solve this problem, NILM researchers introduced the publicly available data description `nilm_metadata`, including the data set description mode and the core metadata part. The class diagram is shown in Fig. 8.

Since the release of the REDD data set is earlier than the release of `nilm_metadata`, it needs to be converted. This article uses NILMTK [35] as a conversion tool. NILMTK is an open source tool in the field of load decomposition, which can organize data sets into the form of `nilm_metadata`, and provides simple evaluation indicators and the realization of load decomposition algorithms. Use NILMTK's `conver_redd` function to convert the REDD data set in `dat` format into structured data in `h5` format that conforms to `nilm_metadata`.

The algorithm in this paper is aimed at the situation when the amount of data is low, so the sampling rate uses a lower 1/60 Hz. Because the FHMM model does not perform well when the number of electrical appliances is large, and the actual application value is low when the number of electrical appliances is too small. Therefore, this article performs load decomposition for all households using the top five electrical appliances that consume power. This operation can be achieved by using the `select_top_k` function of NILMTK.

In the process of data collection, there will be some abnormal values due to sensors or data transmission. In

order to eliminate the influence of outliers, this paper uses median filtering to filter the original data. The value of each point in the time series data in the data set will be replaced by the median value of its neighboring points. The data before and after house1 filtering is shown in Fig. 9, where the left picture is before filtering, and the right picture is after filtering. It can be seen that the median filter has basically filtered out most of the singular values, and the filtered data waveform is smoother.

4.3 Evaluation index

The evaluation of the load decomposition problem is a comprehensive problem. It is necessary to evaluate the accuracy of the prediction of the state of the electrical appliance and the gap between the final power obtained by the decomposition and the actual power. When a single indicator evaluates the load decomposition problem, there will always be the problem of insufficient evaluation. Therefore, this article uses threshold-based state estimation accuracy, *F1*-measure, and root mean square error as the evaluation indicators of load decomposition from both accuracy and power difference. The specific introduction of each indicator is as follows:

- (1) Accuracy is the most intuitive indicator for judging the decomposition performance. Only using the accuracy rate has a good evaluation ability for the closed state of the electrical appliance, but it is difficult to reflect the error between the predicted value and the actual value of the electrical open state. Therefore, this article uses the state estimation accuracy rate with a threshold. For the predicted value greater than 0, the difference between the predicted value and the actual value will be calculated and judge whether the state prediction is accurate according to whether it is greater than the threshold. The specific definition of the accuracy of each appliance is as follows:

Fig. 9 Data comparison before and after filtering

To evaluate the performance of the BH-FHMM, three indices are introduced. The first index is called state estimation accuracy and is expressed as

$$\text{Acc} = \frac{\sum_t^T \sum_i^N F(s_t^i, \hat{s}_t^i + \delta)}{TN} \quad (44)$$

where T is the length of the time series, N is the number of appliances, $s_t^i, \hat{s}_t^i$ represent actual power and predicted power, respectively, and δ represents the threshold, which is defined as 20% of the maximum power in this article, and F is the judging function, which is defined as:

$$F(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases} \quad (45)$$

- (2) The accuracy rate does not fully reflect the decomposition performance of the model because most electrical appliances have been turned off for a long time, and there will still be a high accuracy rate when the predicted values are all 0. This problem can be solved by using F1-measure. F1-measure is a commonly used evaluation standard in binary classification, and it is also a classic evaluation index in the field of load decomposition. In this paper, the improved index proposed in [13] is used as the second index, and the equation is as follows:

$$F\text{-measure} = 2 \frac{PR}{P + R} \quad (46)$$

where $P = \frac{1}{N} \sum_{i=1}^N \frac{ATP_i}{ATP_i + ITP_i + FP_i}$ and $R = \frac{1}{N} \sum_{i=1}^N \frac{ATP_i}{ATP_i + ITP_i + FN_i}$.

- (3) Accuracy and F1-measure mainly evaluate the performance of the model from the aspect of state estimation accuracy. In order to evaluate the gap between the load decomposition result and the actual value, this paper uses RMSE as the load evaluation index, which is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (gt - \text{pred})^2} \quad (47)$$

where gt is the actual power value and pred is the predicted power value.

Table 1 Comparison of state prediction accuracy between the two methods

	house1	house2	house3	house4	house5	house6
FHMM	0.941	0.923	0.931	0.926	0.916	0.940
BH-FHMM	0.959	0.960	0.943	0.951	0.903	0.934

4.4 Results

4.4.1 Experiment based on a household

In this section, firstly, the load decomposition is carried out with the household as the unit. Each household takes the top five electrical appliances in the training set and takes the average of the index results to represent the decomposition index of the family.

Table 1 shows the estimated accuracy rates of electrical appliances in different households. It can be seen that the accuracy rates of the two models are both above 90% and have a relatively high level. The reason is that although the threshold is added to the accuracy index, most electrical appliances are in the off state most of the time. The two probability models based on the Markov chain structure can have a better prediction of the off state. Nevertheless, in most families, the performance of the BH-FHMM model proposed in this paper is 1–3% higher than that of the traditional FHMM model.

Table 2 shows the comparison of the mean values of F1-measure of different household appliances. Compared with the accuracy rate, F1-measure combines the accuracy rate and the recall rate to evaluate the model more comprehensively. In all households, the performance of the BH-FHMM model proposed in this paper has a certain degree of improvement compared with the traditional FHMM model. Except for house2, which has increased by 2.7%, other families have increased by more than 10%.

The above two tables are to take the average of all electrical appliances results to evaluate the performance of the model at the household level. Each household will have its own household electricity consumption habits that can be used as non-traditional characteristics of load decomposition. The decomposition results show that the BH-

Table 2 Comparison of the F-measure between the two methods

	house1	house2	house3	house4	house5	house6
FHMM	0.700	0.811	0.567	0.567	0.712	0.651
BH-FHMM	0.835	0.838	0.722	0.626	0.858	0.762

Fig. 10 Two index values (RMSE and F-measure) for typical appliances based on two methods

FHMM model proposed in this paper effectively extracts this non-traditional feature and improves the performance of model load decomposition.

4.4.2 Experiment based on typical appliances

The ultimate goal of load decomposition is to get the operating conditions of a single appliance. Next, this section will analyze performance of the BH-FHMM model in combination with specific appliances. Several typical appliances, such as dishwashers, refrigerators, lights, sockets, CE appliances, and electric stoves, are selected for analysis in the entire data set. The selected appliances are the top five appliances in the household that consume electricity and contain different operating states. Figure 10 shows the RMSE and F measurement tables for household appliances. The RMSE reflects the error magnitude of power decomposition. Hence, the larger the RMSE is, the

greater the deviation from the actual value, as shown in Fig. 10a. The power decomposition errors of the method proposed in this paper for seven typical electrical appliances are all less than those obtained with the FHMM method. The F-measure values of the typical appliances are shown in Fig. 10b for the two methods. Notably, compared with the FHMM method, the BH-FHMM can improve the power decomposition accuracy of the appliances, except for the refrigerator (fridge). Notably, a fridge is an appliance that is randomly opened, so the corresponding household electricity use habits are difficult to predict.

Figure 10a shows the comparison of the RMSE results of these electrical appliances. The smaller the value, the smaller the difference between the decomposed power and the actual power. It can be seen from the table that the RMSE values of all appliances in the BH-FHMM model are improved by about 10% compared with the FHMM model, which shows that the predicted values obtained by

Fig. 11 Comparison of the results of two methods of dishwasher decomposition

Fig. 12 Comparison of the decomposition results of the two methods of CE appliances

Fig. 13 Comparison of the decomposition results of the two methods of socket

using multiple Gaussian models to fit the status of appliances are closer to the true value.

Figure 10b is a comparison of F1-measure of several typical electrical appliances. The F1-measure values of all appliances in BH-FHMM are greater than FHMM. Among them, the improvement of electric lamps is the most prominent, which is an increase of 40%. This is due to the strong correlation between the on-off state of electric lamps and household electricity consumption habits. The performance of refrigerators is basically the same. This is because refrigerators are automatically controlled and regularly cycled electrical appliances, which have a low correlation with household electricity consumption habits.

The actual decomposition results of three typical electrical appliances are given below for comparative analysis. Figures 11, 12 and 13 are the load decomposition results of

dishwashers, CE appliances and sockets, respectively. The left figure shows the results of BH-FHMM, and the right figure shows the results of the FHMM model. The horizontal axis is time and the vertical axis is power. The orange part is the actual power of the appliance, and the blue part is the predicted power of the model. It can be seen from Fig. 11 that for a typical FSM appliance such as a dishwasher, the BH-FHMM model reduces the misjudgment of low-power on events by introducing electricity usage habits, and makes the curve fit the actual curve more closely. It can be seen from Fig. 10 that the RMSE is reduced by about 11 and the F1-measure is increased by about 9%. For CE appliances, such as CVD appliances, the operating power is related to time, so different power values are used for different time periods to match. Hence, the curve is better fitted to the real data, and its RMSE is

reduced from 58 to 32. The socket is a relatively special device. The device itself produces almost no power consumption. Its power consumption depends on the connected appliances. The REDD data set does not specify the device connected to the socket, so it can be seen as a black box. It can be seen from the results in Fig. 10 that the RMSE of the socket has been reduced by about 8 and the F1-measure has been increased by about 23%. From Fig. 13, it can be seen that for this black box device, the decomposition result of the BH-FHMM model proposed in this paper is even better. Because of close to the actual results, so the BH-FHMM model proposed in this paper has better decomposition performance.

5 Conclusions

A novel NILM method called the BH-FHMM is proposed, and household electricity habits are used as new model features. This is the first study to introduce human habits to improve the performance of NILM-based model. In this method, an improved K-means algorithm is first used to obtain the state number for each appliance. According to the relationship between human habits and the operating status of electric appliances, a household electricity habit model is built based on a GMM. The GMM is used to obtain the observation probability distribution of the

operating status for electrical appliances and then to predict the operating status. Finally, the BH-FHMM is constructed by combining the FHMM with the household electricity habit model.

The performance of the BH-FHMM is verified through an experimental comparison of the traditional FHMM method and the BH-FHMM method based on the REDD data set. According to the results, the improved K-means algorithm can eliminate the infrequent states of electrical appliances and reduce false predictions. In the experiment based on households, the accuracy of the results per household increased by 1% to 2%, and the F-measure increased by an average of more than 10%. In the experiment based on typical electrical appliances, the RMSE for most appliances was reduced by 10%, and the F-measure was also improved, especially for lights, which closely reflect household electrical habits. Hence, the excellent performance of load power decomposition with the BH-FHMM was verified.

Appendix 1: Proportion of on-time major electric appliances in households at different times

Appendix 2: The average power use of major electrical appliances in households at different times

Funding This study was funded by National Natural Science Foundation of China (61972367), and Key R & D projects of Shandong Province (2019JMRH0109).

Declarations

Conflict of interest The author declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Zeifman M, Roth K (2011) Nonintrusive appliance load monitoring: Review and outlook. *IEEE Trans Consum Electron* 57(1):76–84
2. Hoyo-Montano JA et al (2016) Overview of non-intrusive load monitoring: a way to energy wise consumption. In: *International Conference on Power Electronics IEEE*, 2016
3. Menke JH, Bornhorst N, Braun M (2019) Distribution system monitoring for smart power grids with distributed generation using artificial neural networks. *Int J Electr Power Energy Syst* 113:472–480
4. Gupta S, Reynolds MS, Patel SN (2010) ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home
5. Marchiori A et al (2011) Circuit-level load monitoring for household energy management. *Pervasive Computing*, IEEE
6. Esa NF, Abdullah MP, Hassan MY (2016) A review disaggregation method in non-intrusive appliance load monitoring. *Renew Sustain Energy Rev* 66:163–173
7. Nguyen TK, Dekneuveld E, Jacquemod G (2017) Development of a real-time non-intrusive appliance load monitoring system: an application level model. *Int J Electr Power Energy Syst* 90:168–180
8. Zeifman M, Akers C, Roth K (2015) Non-intrusive monitoring of miscellaneous and electronic loads. In: *IEEE international conference on consumer electronics*, 2015.
9. Laughman C, Lee K, Cox R, Shaw S, Armstrong P (2003) Power signature analysis. *IEEE Power Energy Mag* 1(2):56–63
10. Leeb S, Shaw S, Kirtley J (1995) Transient event detection in spectral envelope estimates for non-intrusive load monitoring. *IEEE Trans Power Deliv* 10(3):1200–1210
11. Zoha A, Gluhak A, Imran MA, Rajasegarar S (2012) Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. *Sensors* 12(12):16838–16866
12. Najafi B, Moaveninejad S, Rinaldi F (2018) Data analytics for energy disaggregation: methods and applications. In: *Big data application in power systems*
13. Kim H, Marwah M, Arlitt M, Lyon G, Han J (2011) Unsupervised disaggregation of low frequency power measurements. In: *Proceedings of the 2011 SIAM international conference on data mining*, 2011, pp 747–758
14. Bonfigli R, Principi E, Fagiani M, Severini M, Piazza F (2017) Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Appl Energy* 208:1590–1607
15. Biansoongnern S, Plungklang B (2016) Non-intrusive appliances load monitoring (NILM) for energy conservation in household with low sampling rate. *Proc Comput Sci* 86:172–175
16. Zeifman M, Roth K (2011) Non-intrusive appliance load monitoring: Review and outlook. *IEEE Trans Consum Electron* 57(1):76–84
17. Kelly J, Knottenbelt W (2015) Neural NILM: deep neural networks applied to energy disaggregation. *Computer Science*, 2015.
18. Bonfigli R, Felicetti A, Principi E, et al (2017) Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. In: *Energy and Buildings*, 2017, p 158
19. Krystalakos O, Nalmpantis C, Vrakas D (2018) Sliding window approach for online energy disaggregation using artificial neural networks, pp 1–6
20. Kaselimi M, Protopapadakis E, Voulodimos A, et al (2019) Multi-channel recurrent convolutional neural networks for energy disaggregation. In: *IEEE Access*, 2019
21. Fischer J, Ramchurn S, Osborne M, et al (2013) Recommending energy tariffs and load shifting based on smart household usage profiling, 2013
22. Basu K, Debusschere V, Bacha S (2014) Residential appliance identification and future usage prediction from smart meter. In: *Conference of the IEEE industrial electronics society*. 2014.
23. Rao K, Ravichandran D, Mahesh K (2016) Non-intrusive load monitoring and analytics for device prediction
24. Spiegel S, Albayrak S (2014) Energy disaggregation meets heating control. In: *Acm symposium on applied computing*. 2014
25. Belley C, Gaboury S, Bouchard B et al (2014) An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances. *Pervasive Mob Comput* 12:58–78
26. Kong W, Yan X, Dong ZY, et al (2015) An extended prototypical smart meter architecture for demand side management. In: *2015 IEEE 13th international conference on industrial informatics (INDIN)*. 2015
27. Liu H, Wu H, Chen C, Wang Z (2020) An improved non-intrusive load disaggregation algorithm and its application. In: *Sustainable cities and society*, vol 53, 2020
28. Xia M, Liu W, Wang K, Song W, Chen C, Li Y (2020) Non-intrusive load disaggregation based on composite deep long short-term memory network. *Expert Syst Appl* 160
29. Kauw M, Benders RMJ, Visser C (2015) Green methanol from hydrogen and carbon dioxide using geothermal energy and/or hydropower in Iceland or excess renewable electricity in Germany. *Energy* 90(P1):208–217
30. Viterbi AJ (1967) Error bound for convolutional codes and asymptotically optimum decoding algorithms. *IEEE Trans Inform Theory* 13
31. Kolter JZ, Johnson MJ (2011) REDD: a public data set for energy disaggregation research. *Sustkdd* 25

32. Kolter Z, Jaakkola T, Kolter JZ (2012) Approximate inference in additive factorial HMMs with application to energy disaggregation
33. Zico KTJJ (2012) Approximate inference in additive factorial HMMs with application to energy disaggregation
34. Kong W, Dong ZY, Ma J, Hill D, Zhao J, Luo F (2016) An extensible approach for non-intrusive load disaggregation with smart meter data. *IEEE Trans Smart Grid* 2016:1
35. Batra N et al (2014) NILMTK: an open source toolkit for non-intrusive load monitoring. *Statistics* 2014:265–276

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.