Non-intrusive load disaggregation based on composite deep long short-term memory network ☆

Min Xia [a,b,*,1], Wan'an Liu [a,b,1], Ke Wang [c], Wenzhu Song [a,b], Chunling Chen [a,b], Yaping Li [c]

[a] *Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[b] *Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[c] *China Electric Power Research Institute, Nanjing 210003, China*

ARTICLE INFO

ABSTRACT

Non-invasive load monitoring (NILM) is a vital step to realize the smart grid. Although the existing various NILM algorithms have made significant progress in energy consumption feedback, there are still some problems need to further addressed, such as the exponential growth of state space with the increase of the number of multi-state devices, which leads to the dimension disaster; and it is difficult to capture the power fluctuation information effectively because of the neglect of time-dependency problem load disaggregation; traditional disaggregation involves a process of one sequence to one sequence optimization, which is inefficient. In our study, a composite deep LSTM is proposed for load disaggregation. The proposed algorithm considers the process of load disaggregation as a signal separation process and establishes regression learning from a single sequence to multiple sequences to avoid dimension disaster. In addition, an encoder-separation-decoder structure is introduced for load disaggregation. Encoder completes the effective encoding of the mains power and differential power information, the time-dependency of the encoding process implemented by a deep LSTM, separation realizes the disaggregation process by separating the encoded information, and decoder decode the separated signal into the sequences of corresponding electrical appliances. Compared with the one sequence to one sequence disaggregation method, the proposed method simplified disaggregation complexity and improves the efficiency of disaggregation. The experimental results on WikiEnergy and REDD datasets show that the proposed method can reduce the disaggregation error and improve the comprehensive performance of event detection. Besides, our study can provide conditions for the realization of the bidirectional interaction of the smart grid and the improvement of the smart grid scheduling.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Load monitoring technology, as one of the key technologies of advanced metering infrastructure (AMI), is a key step to realize smart grids. Traditional load monitoring adopts an intrusive way, that is, sensors are installed on the users' electrical appliances to record the usage. The advantage of the intrusive method is that the monitoring data is accurate and reliable, and the disadvantages are poor operability, high implementation cost and low user

acceptance (Ridi, Gisler, & Hennebert, 2014; Froehlich et al., 2010). Therefore, non-intrusive load monitoring (NILM) technology is urgently needed to achieve load monitoring. NILM proposed by Hart (1992) is a feasible system to implement appliance-level load monitoring. That is, only a single monitoring device is installed to obtain the energy consumption data of the whole household without increasing the marginal cost, the running state of each electrical equipment in the household can be accurately identified. The key to the implementation of NILM is to estimate the appliance-level power consumption by utilizing the energy consumption data obtained from the single monitoring device, and non-intrusive load disaggregation (NILD) is the main technical means of the implementation of the estimation process. NILD disaggregates the mains power information into the information of each electrical appliance, and then the appliance-level power consumption can be obtained to track the power consumption behavior with the aim of providing accurate power consumption feedback to consumers

---

(Evora, Hernandez, & Hernandez, 2015). The realization of NILD has high application value, which can bring benefits to the power company and users (Paterakis, Erdinc, Bakirtzis, & Catalao, 2015). Compared with intrusive load monitoring, the remarkable advantages of NILD mainly manifested in its economic investment is small and its practicability is strong (Belley, Gaboury, Bouchard, & Bouzouane, 2015). Therefore, since the NILD was put forward, it has attracted wide attention.

At present, mathematical optimization and pattern recognition are the two major methods of addressing NILD problems, and these two methods must be analyzed based on load signatures (LS) (Hassan, Javed, & Arshad, 2013). LS is defined as unique information that reflects the state of power consumption in the operation of an electrical appliance. LS includes the operating characteristics of the load, which are determined by the working conditions of the electrical appliance. LS is divided into three types of steady state, transient and running mode, in which the steady-state and transient depend on the characteristics of the components inside the equipment, and the running mode is determined by the operation control strategy of the equipment (Lee, Fung, Lam, Chan, & Lucente, 2004). The characteristics of LS are repeatable, so the load types and their usage status can be identified based on LS characteristics, which are the realization principle of NILD (Rahimpour, Qi, Fugate, & Kuruganti, 2017).

Combinatorial optimization (CO) transforms the NILD problem into 0–1 planning problem (Zoha, Gluhak, Imran, & Rajasegarar, 2012). Chang, Lin, Chen, and Lee (2013) used a heuristic algorithm to realize load disaggregation based on the steady current. However, in the case study of Chang's work, only two induction motors with different power and one load with switch state are tested, the multi-state devices and continuously variable electrical appliances are not involved, which is not representative. Besides, there is a large room for improvement of the disaggregation error in Chang's work. The model used by Lin and Tsai (2013) is similar to that of the literature (Chang et al., 2013), which could handle the simultaneous events and considered the use of electrical appliance to be close to the active power consumption, but the recognition accuracy was not high. Piga, Cominola, Giuliani, Castelletti, and Rizzoli (2015) proposed a sparse optimization method to solve the NILD problem and improved the disaggregation accuracy. Ahmadi and Marti (2015) proposed a load disaggregation method based on feature matching (also called LS matching), the objective of which is to find and detect the electrical equipment with the highest similarity of load characteristics in the feature library. Because of the simplicity and easy to implement of CO algorithm, it has been widely applied for load disaggregation since it was put forward, however, CO methods mainly have the following problems (Batra, Dutta, & Singh, 2013): (1) The model is essentially a complete problem of NP-hard, the efficiency of disaggregation is a challenge; (2) Accurate identification needs a complete load feature library, which is often difficult to meet in practice; (3) The method based on the optimization theory is a discrete method, which can only be used to analyze the discrete state of the electrical appliances, so the load with very strong load fluctuation is difficult to be modeled.

The essence of load disaggregation based on pattern recognition is to identify the electrical appliances by studying the LS characteristic mode of each electrical equipment. Unsupervised learning based load disaggregation does not require labeled data for electrical appliance, thereby reducing manual intervention (Munshi & Mohamed, 2017). Shao, Marwah, and Ramakrishnan (2013) proposed a pattern mining method to realize unsupervised energy disaggregation. Through the steady state characteristics of the mains power, the scene sequence was generated, and the scene of the applied electrical appliance was screened out, then the usage of each equipment was obtained by using the pattern mining

technology. Martins, Pinto, and Bittencourt (2018) disaggregated six different industrial machines of a Brazilian factory with factorial hidden Markov model (FHMM). The disaggregation results are compared on normalized disaggregation error (NDE), normalized signal aggregated error (SAE), and F1 score. The results show that FHMM can realize the prediction of on/off of electrical appliances. However, when the number of electrical appliances increases, the complexity of FHMM for load disaggregation increases exponentially, resulting in the final disaggregation performance is far from good. Parson, Ghosh, Weal, and Rogers (2014) proposed an approach which combines a one-off supervised learning process over existing labeled appliance data sets, with an unsupervised learning method over unlabeled household aggregate data. Kolter, Batra, and Ng (2010) used a sparse coding algorithm to learn the power consumption models of each device and used these models to predict the power consumption of different devices. Johnson and Willsky (2013) used the hierarchical Dirichlet hidden semi-Markov model to achieve unsupervised power disaggregation. The algorithm was fast and flexible, but the accuracy was low. Compared with the supervised learning algorithm, the load disaggregation algorithm based on unsupervised learning is with lower accuracy and lacks the ability to identify complex continuous state loads because of lack of prior supervision information (Esa, Abdullah, & Hassan, 2016). Zaen, Achkar, Carrillo, and Hutter (2018) compared the performance of unsupervised clustering and neural network on detection and classification of refrigeration units of display cabinet, under counter fridge and right fridge. Although unsupervised clustering algorithm can effectively detect the power cycle, it cannot precisely estimate the power consumption of refrigeration units. In comparison, the precision rate and recall rate of a supervised neural network algorithm in electrical appliances are all over 90% which indicates neural network can achieve accurate power disaggregation.

Due to the reliability of supervised learning method, there are many researches in this field such as K-nearest neighbor algorithm (Tsai & Lin, 2012), neural network (Chang, Chen, Tsai, & Lee, 2011; Chang, Lian, Su, & Lee, 2013; Park, Baker, & Franzon, 2019; Martinez, Pacheco, & Robergs, 2018), support vector machine (Saitoh, Osaki, Konishi, & Sugahara, 2010), Adaboost algorithm (Hassan, 2012), fuzzy system (Lin & Tsai, 2014; Welikala, Dinesh, Ekanayake, Godaliyadda, & Ekanayake, 2019; He, Du, Yang, Harley, & Habetler, 2012), etc. Cominola, Giuliani, Piga, Castelletti, and Rizzoli (2017) combines FHMM and iterative subsequence dynamic time warping (ISDTW) to conduct supervised NILM and semi-supervised NILM to track the power consumption of multiple devices running at the same time. Their work solved the problems of intrusive data sampling, the inconsistency of NILM metrics and the difficulty of identifying states of devices operating at the same time. The performance on AMPds dataset (Makonin, Popowich, Bartram, Gill, & Baji, 2013) evaluated by power contribution error (PCE), F1 score and R2 score show the effectiveness of supervised and semi-supervised algorithms. Compared with Kolter's work (Kolter & Johnson, 2011), reduced the average PCE from over 50% to 4%. Du, Restrepo, Yang, Harley, and Habetler (2013) integrated supervised self-organizing mapping and Bayesian hybrid identifier to make a probabilistic judgment on the category of a device, but this method fails to consider the problem of load classification for multi-state transfer. Lin and Tsai (2014) combined Fuzzy C-Means based on particle swarm optimization (PSO) with neural fuzzy classification was used to identify household appliances and could effectively deal with the LS similarity of household appliances. Pereira and Nunes (2018) makes a comprehensive comparison of supervised event detection algorithms in the NILM on 23 evaluation metrics, which includes True positive rate (TPR), False positive rate (FPR), Area under roc curve (AUC), accuracy, precision, recall rate et al. The results demonstrate that

under the circumstance of data imbalance, the selection of performance metrics is vital. Guo, Wang, and Kashani (2014) used the modified implicit Markov model based on differential observation to solve the overlap of active signals between target electrical appliances and other electrical equipment. Non-parametric classification was firstly done by Wang and Zheng (2011) based on the number of events and occurrence time, and then linear discriminant analysis was used to identify the categories of electrical appliance. Park et al. (2019) proposed an appliance identification algorithm based on a cogent confabulation neural network, which did not require multiplications during the identification phase and improved the speed of load disaggregation. Welikala et al. (2019) invested a NILD method which incorporates appliance usage patterns (AUPs) to improve performance of active load identification and forecasting.

Although supervised NILM generally achieved better load disaggregation performance than that of unsupervised NILM, however, there are still some deficiencies in the above supervised NILM methods: (1) Most of the above supervised methods require handcraft feature extraction (Basu, Debusschere, Bacha, Maulik, & Bondyopadhyay, 2014), which is a time-consuming and laborious work. Besides, in the presence of multi-state devices, the manually extracted features cannot guarantee the disaggregation performance of devices under various running states. (2) Traditional methods usually modeled NILM task as a classification problem to detect the running states of individual appliances. However, such methods cannot capture the fluctuation information of power consumption, fail to provide some valuable information such as how the working patterns and operating efficiency of a device impacts the power consumption. (3) Some methods such as denoising AutoEncoder (DAE) (Kelly & Knottenbelt, 2015), CNN (Martinez et al., 2018; Xia, Liu, Wang, Zhang, & Xu, 2019) modeled the load disaggregation problem as a regression problem, their work are based on the assumption that the power consumption information is time-independency. However, the current state of the a load may have a certain dependency on the previous operation state (for example, the washing cycle of the washing machine may occur after the pre-washing cycle), time dependency usually exists especially in the load disaggregation of multi-state equipments, which the power consumption curve often presents periodic operation pattern. (4) Generally, the current methods (such as hidden Markov model (HMM) based methods (Kim, Marwah, Arlitt, Lyon, & Han, 2011)) do energy disaggregation by establishing one-to-one (one input to multiple output) mapping from the mains power sequence to the target sequence. When the number of electrical equipment is large (especially for multi-state equipment), the time complexity of model training increases as well.

To address the above limitations in current supervised NILD, we propose a composite long short-term memory network (CD-LSTM) for load disaggregation. The proposed CD-LSTM satisfies some outstanding characteristics, making it superior to the previously supervised NILD algorithms. The arising limitations and drawbacks motivated CD-LSTM equipped with the following features: (1) Automatic load feature extraction ability which benefits from the deep learning framework (Singh & Majumdar, 2017; Singhal, Maggu, & Majumdar, 2018) allows CD-LSTM to conduct end-to-end load disaggregation to simplify the feature extraction process. (2) CD-LSTM modeled load disaggregation problem as a regression learning process, which not only distinguish the state of electrical appliances but also realized precise appliance-level load disaggregation. (3) The time-dependency problem in multi-state appliances is solved by the design of deep LSTM framework (Kim, Le, & Kim, 2017; Mauch & Yang, 2015). The power disaggregation of each time step takes the power fluctuation information of the previous time steps into account to improve the disaggregation performance. (4) Load disaggregation process is regarded as a signal separation process, driving CD-LSTM designed as an encoder-separation-decoder and one-to-many (one input, multiple outputs) structure. Compared with one-to-one structure, CD-LSTM avoided repeated model training and improved the disaggregation efficiency.

Experimental results on two open NILD datasets show CD-LSTM can reduce the disaggregation error and improve the disaggregation efficiency over existing benchmark models, providing more robust bidirectional feedbacks to both smart grid and power consumers.

2. Proposed composite deep long short-term memory network

2.1. Review of long short-term memory network

Long short-term memory network (LSTM) is a special recurrent neural network (RNN), which can learn long-term dependence. LSTM was first proposed by Hochreiter and Schmidhuber (1997), and then improved by Graves (2012), which has been widely applied in time series correlation analysis. LSTM can overcome the long-term dependencies problem in RNN (Williams & Zipser, 1989; Huang, Xu, & Yu, 2015; Bao, Yue, & Rao, 2017; Zhu et al., 2016). Although RNN (You, Park, & Oh, 2017) is designed to be able to process the whole time sequence information, the final input data can be well memorized, the memory for the earlier input data is insufficient. LSTM solves the problem of long term dependence, which makes it possible to memorize long term data without complex parameter adjustment.

In all recurrent neural network, there is a chain form of duplicate neural network modules. In RNN, the structure of duplicate modules is very simple, which contains only one *tanh* layer (Chemali, Kollmeyer, Preindl, Ahmed, & Emadi, 2017). Compared with RNN, the duplicate module structure of LSTM is more complicated. The schematic diagram of the LSTM structure is shown in Fig. 3, in which the duplication modules are represented by the small rectangles. Each small rectangle represents a neural network, and LSTM contains 4 layers of neural networks. $\oplus$ and $\otimes$ represent element-wise addition and multiplication. σ is the *sigmoid* activation function, and *tanh* represents the *tanh* activation function.

2.1.1. The cell state of LSTM

As showed in Fig. 1a, there is a straight line running through all the elements in LSTM, which represents the state of LSTM. In the process of information transfer, each unit can increase or decrease the information of a state. These operations allow information to selectively affect the state of each moment through the "gate" structure. The gate structure contains an operation of *sigmoid* function σ and a vector dot product $\otimes$. The output of the *sigmoid* activation function is a value between $(0, 1)$, which directly controls the proportion of the information transfer. If the output value C is 0, the neuron does not allow the message to be transmitted, while the output value C of 1 means all the information can pass through the neuron (Kong et al., 2017).

2.1.2. Forget gate

In the LSTM structure, the first step is to decide which information needs to be discarded, which is determined through the "Forget gate". The "Forget gate" will decide which part of the state unit C_{t-1} should be forgotten based on the current input value x_t and the output value h_{t-1} at the last time. The structure of the "Forget gate" is shown in Fig. 1b, and its formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{1}$$

where W_f, b_f are the connection weights and bias values of *sigmoid* layer in "Forget gate" respectively.

(a) The cell state of LSTM

(b) The "Forget gate" structure

(c) The "Input gate" structure
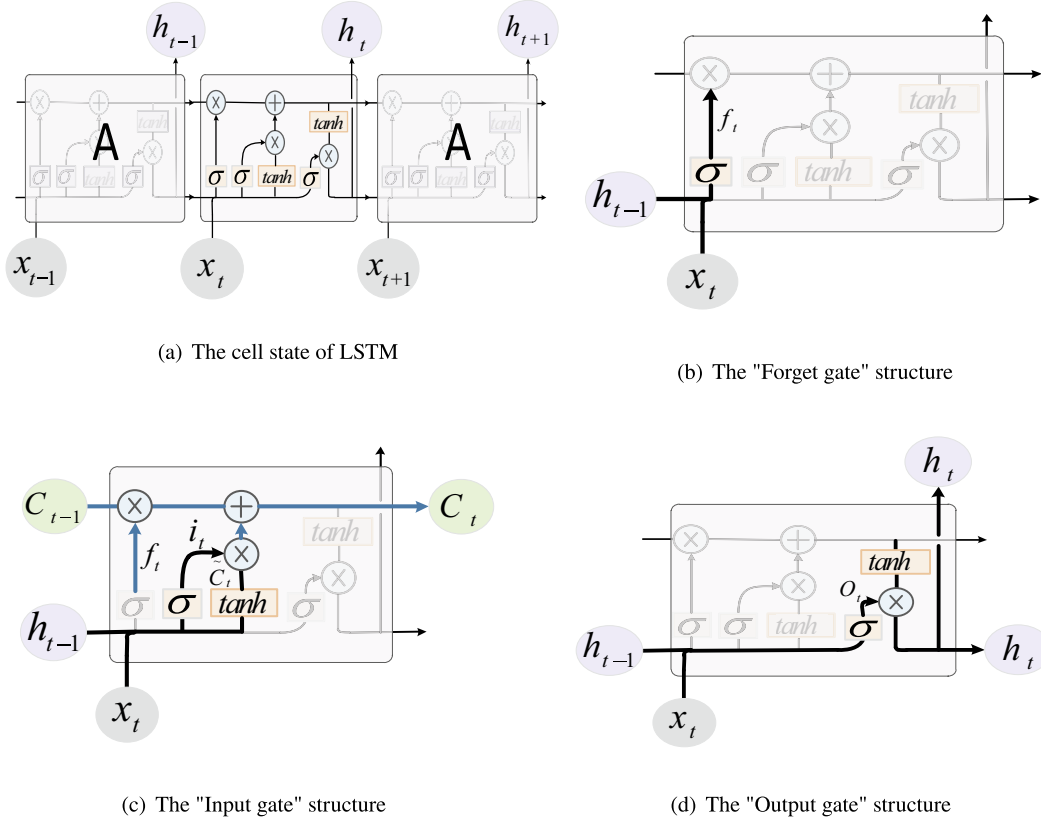
(d) The "Output gate" structure

Fig. 1. Schematic diagram of LSTM structure.

2.1.3. Input gate

After LSTM "forgot" part of the previous status information, the current state also learns new information from the current input, which is achieved through the "Input gate". As showed in Fig. 1c, the "Input gate" contains two parts: the *tanh* layer and the *sigmoid* layer. The *tanh* layer is used to create a new candidate value vector. The *sigmoid* layer is used to control which information of the *tanh* layer is added to the current state. The *tanh* layer and the *sigmoid* layer are described as:

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$

$$\tilde{C}_t = tanh(\mathbf{W}_C \cdot [h_{t-1}, x_t] + b_C), \tag{3}$$

where $\mathbf{W}_i, b_i$ are the connection weights and bias values of *sigmoid* layer in "Input gate" respectively, $\mathbf{W}_C, b_C$ are the connection weights and bias values of *tanh* layer in "Input gate" respectively.

Through the operation of "Forget gate" and "Input gate", LSTM can effectively filter out which information should be forgotten and which new information is worth learning (Cortez, Carrera, Kim, & Jung, 2018). Therfore, current cell state C_t can be updated as:

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \tag{4}$$

the updating process is shown in the blue thick lines in Fig. 1c.

2.1.4. Output gate

After getting the updated state, LSTM needs to determine the output value of the current time, and the output value is determined through the "Output gate". As showed in Fig. 1d, the "Output gate" controls the output of the current state through a *sigmoid* function, which is described as follows.

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O), \tag{5}$$

$$h_t = O_t \otimes tanh(C_t), \tag{6}$$

where W_O, b_O are the connection weights and bias values of *sigmoid* layer in "Output gate" respectively. In Eq. (6), "$\otimes$" is the dot product of vectors.

2.2. Deep LSTM network

Deep LSTM is a variant of LSTM, which can enhance the expressive ability of the model. Compared with LSTM, deep LSTM duplicates the structural morphology many times. The parameters in the same layer are consistent, and the parameters in different layers are different. Fig. 2 shows the time step expansion structure of deep LSTM networks, each of which represents a structural morphology. In designed deep LSTM, dropout operation (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) is added to LSTM, which can make the network more robust. When training neural network models, the use of dropout can prevent over-fitting (Dahl, Sainath, & Hinton, 2013). In the deep LSTM of this work, dropout operations are used only between different levels of modules, rather than in the same level of modules. As showed in Fig. 2, the solid line part indicates that no dropout is used, and the dashed line indicates the use of dropout.

2.3. Design of composite deep LSTM network

Researches of load disaggregation based on deep learning methods often trains a network model for every target appliance, and the model inputs the mains power data and outputs the power of a target electrical appliance. This leads to a lot of repetitive model training, which is cumbersome and more time-consuming. For load disaggregation, CD-LSTM is proposed in this study, the input mains power data can be simultaneously decomposed into the power of multiple target electrical appliances, thus avoiding
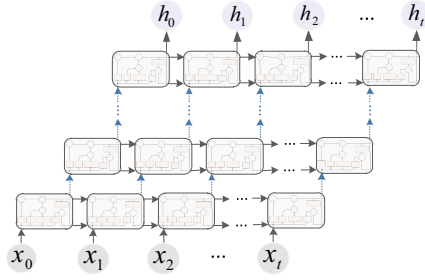
Fig. 2. Structure of deep LSTM network.

repeated training of models. The network structure is shown in Fig. 3. Since the input data integrates the original mains power data with the differential data which obtained through the mains power differential operation (see differential processing in Section 3.1.2 for details), a convolution layer with filter size 1×2 is firstly used to extract the mains power and differential information (Ting, Tan, & Sim, 2019), and the output of the convolution layer can be denoted as $\mathbf{x}^{conv} \in R^{1 \times L}$, where L is the length of a sample. In this way, we not only extract the integrated mains power and differential information through a convolution operation, but also transfer the disaggregation target to the single-channel signal separation, then the NILM task can be described as estimating the power consumption information $Y_1(t), Y_2(t), \ldots, Y_M(t)$ of M target electrical appliances by using a given discrete total power data $X(t)$:

$$X^{conv}(t) = \sum_{i=1}^{M} Y_i(t), \tag{7}$$

In the window sliding phase, we describe how to divide the mains power sequence into K segments of length L. We can get $\mathbf{x}^{conv} \in R^{1 \times L}$ after performing convolution, where:

$$\begin{cases} \mathbf{x}_a^{conv} = X(t), \\ \mathbf{y}_{i,a} = Y_i(t), \end{cases} \tag{8}$$

where $t \in [aL, (a+1)L], a = 1, 2, \ldots, K, a$ denotes ath sequence segment. Each mains power segment and the power segment of the individual appliance can be represented as a weighted summation of the basis signals $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N]$ while $\mathbf{B} \in R^{N \times L}$, and $\mathbf{b} \in R^{1 \times L}$:

$$\begin{cases} \mathbf{x}_a^{conv} = \mathbf{c}_a \mathbf{B}, \\ \mathbf{y}_{i,a} = \mathbf{d}_{i,a} \mathbf{B}, \end{cases} \tag{9}$$

where $\mathbf{c} \in R^{1 \times N}$ is the weight vector of mains power segment, $\mathbf{d}_{i,a} \in R^{1 \times N}$ represents the weight vector of the ath power segment of the ith electrical appliance, and $\mathbf{c}_a$ satisfies the following conditions:

$$\mathbf{c}_a = \sum_{i=1}^{M} \mathbf{d}_{i,a} \tag{10}$$

from Eq. (10), we can further know:

$$\mathbf{c}_a = \sum_{i=1}^{M} \mathbf{d}_{i,a} = \sum_{i=1}^{M} \mathbf{c}_a \odot (\mathbf{d}_{i,a} \oslash \mathbf{c}_a) = \mathbf{c}_a \odot \sum_{i=1}^{M} \mathbf{m}_{i,a} \tag{11}$$

Among them, it can be seen from Eq. (11) that $\mathbf{m}_{i,a} = \mathbf{d}_{i,a} \oslash \mathbf{c}_a$, obviously, $\mathbf{m}_{i,a}$ represents the relative proportion of the ith power segment of target electrical appliance to the mains power segment weight vector $\mathbf{c}_a$. $\odot$ and $\oslash$ represent element-wise multiplication and division respectively. Consequently, a load disaggregation problem can be converted into utilizing $\mathbf{x}_c^{conv}$ to estimate the weight vector $\mathbf{d}_{i,a}$ task for each appliance.

Fig. 3 shows the structure of the CD-LSTM network. CD-LSTM consists of three main parts: encoder, separation, decoder. The encoder is used to get the weight vector $\mathbf{c}_a$ of the ath mains power segment, separation part is designed for obtaining the ath mask vector $\mathbf{m}_{i,a}$ of ith individual appliances, finally, the decoder can decode the mask vector into its corresponding weight vector $\mathbf{d}_{i,a}$ of power segment of individual appliances.

In the encoder structure, it can be seen from Eq. (9) that $\mathbf{x}_a^{conv} = \mathbf{c}_a \mathbf{B}$, where $\mathbf{x}_a^{conv} \in R^{1 \times L}$, $\mathbf{c}_a \in R^{1 \times N}$, $\mathbf{B} \in R^{N \times L}$, encoder vector $\mathbf{c}_a$ can be implemented by a fully connected layer, which can be obtained by a fully connected layer with N hidden nodes, thus, we can get $\mathbf{c}_a \in R^{1 \times N}$. Combined with the structure of LSTM in
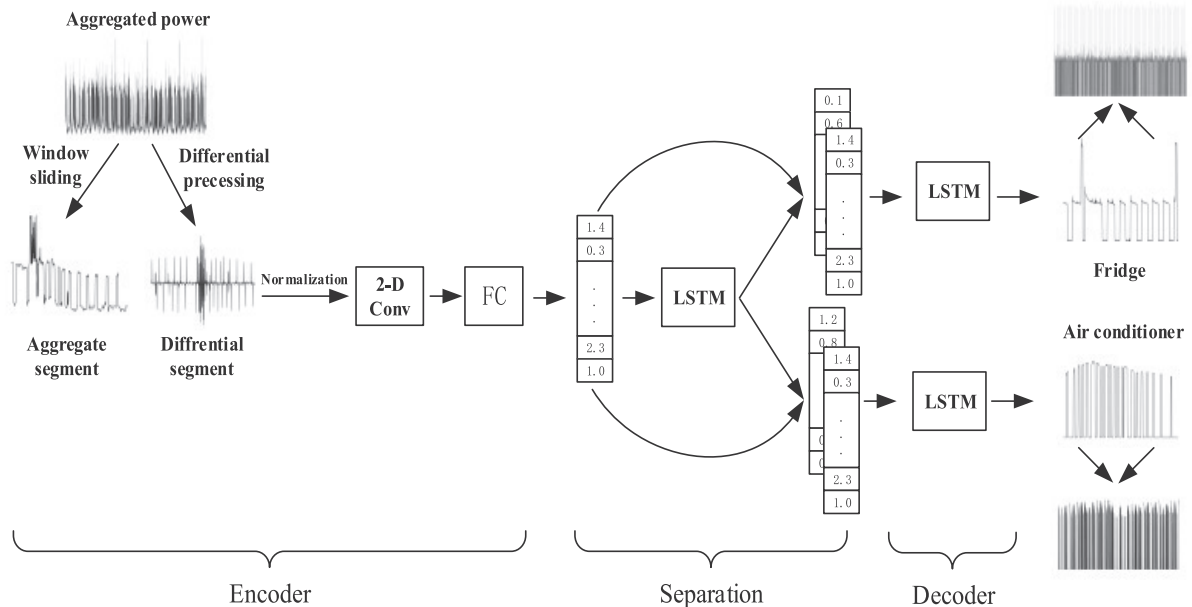

Fig. 3. Structure of composite deep LSTM network.

Fig. 1, the structure of deep LSTM in Fig. 2 and the mechanism analysis of "Forget gate", "Input gate" and "Output gate" in LSTM structure, it can be seen that in the encoding process of load disaggregation at each time step, deep LSTM-based NILD task effectively utilizes the information of previous time steps to adequately solve the long-dependency problem in multi-state load disaggregation.

The separation network produces for each target source $M = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_K] \in R^{K \times N}$. Such a mask matrix or vector is widely used in deep learning applications. In the image field, it can be regarded as high-level feature maps, and in the analysis of sequence data, the dimension vector with size $1 \times N$ is known as mask vector. Let $\mathbf{m}_{i,a} = [m_i(j, 1), m_i(j, 2), \ldots m_i(j, N)] \in R^{1 \times N}$ be the mask vector for ath sequence of target appliance s, thus the mask matrix for all K segments of target appliances s can be defined as $\mathbf{M} = [\mathbf{m}_{i,1}, \mathbf{m}_{i,2}, \ldots, \mathbf{m}_{i,K}] \in R^{K \times N}$. The goal of our study is to find an estimate $\mathbf{M}$ using a deep learning model $f(\mathbf{C}; \Phi) = \mathbf{M}$. Because LSTM has obvious advantages in time series analysis, we take advantage of a LSTM with N hidden nodes to separate the encoded vectors in our separation network. The separation network takes encoded sequential vector $\mathbf{c}_a \in R^{1 \times N}$ as input, where $\mathbf{c}_a$ can be specifically expressed as $\mathbf{c}_a = [c_a(1), c_a(2), \ldots, c_a(N)]$. Therefore, the function between $\mathbf{c}$ and $\mathbf{m}$ can be modeled by Eqs. (1)–(6).

In addition, because of the deep LSTM structure used in the separation stage, the problem of vanishing gradient is inevitable. In order to slightly alleviate the effect of vanishing gradient, skip connection is introduced in disaggregation network, which borrows the idea of ResNet (He, Zhang, Ren, & Sun, 2016) and DenseNet (Huang, Liu, Van Der Maaten, & Weinberger, 2017) to operate information flow and solve the problem of vanishing gradient. ResNet skip connection, the implementation can be simply expressed as $skip(\mathbf{c}_a, \mathbf{m}_{i,a}) = \mathbf{c}_a + \mathbf{m}_{i,a}$, and skip connection in DenseNet can be easily realized as channel fusion: $skip(\mathbf{c}_a, \mathbf{m}_{i,a}) = concat(\mathbf{c}_a, \mathbf{m}_{i,a})$. In our experiment, DenseNet-type skip connection operation is adopted to transfer information flow. Another advantage of such skip connection is that it combines encoding vector with mask $\mathbf{m}_{i,a}$, which is helpful for reconstructing electrical power segments. Besides, dropout (Srivastava et al., 2014; Dahl et al., 2013) skill is applied in LSTM which helps to avoid overfitting.

In the decoding part, the mask vectors we separate from encoded mains vector can be regarded as the feature vector of target sequence, thus, separation network can be regarded as the encoder network of decoder network. In other words, separation network as both separation function and encoding ability. If we consider the separation network as a encoder network of decoder network, in the separation network, we adopt LSTM to finish the separation and encoding process, therefore, in order to accurately reconstruct the power consumption sequence of each electrical appliance, M LSTM with L hidden node is combined together to jointly decode the corresponding mask. The whole architecture of CD-LSTM is designed as one-to-many (one input and multiple outputs). Compared with the traditional NILM algorithm, the advantage of such design is that it can simultaneously disaggregate the power consumption of multiple running devices, and avoid redundant repeated inference processes.

3. Expermental settings

3.1. NILD datasets and preprocessing

3.1.1. NILD dataset

According to the sampling frequency, NILD can classified into high-frequency NILD and low-frequency NILD. The low-frequency indicate the sampling rate is lower than 1 Hz (sampling period higher than 1 second), high-frequency sampling refers to the sampling rate is higher than 50 Hz (sampling period lower than 0.02 s), although high-frequency NILD can extract richer features from high-order harmonics to improve the performance of NILD, maintaining high frequency sampling often requires the installation of custom expensive monitoring devices. In contrast, low-frequency NILD is much more economical. However, the experimental data sampled with a low-frequency contains less information, which makes low-frequency NILD a challenge. Therefore, in this study, we focus on low-frequency NILD.

The experimental data in this study are derived from open datasets WikiEnergy and REDD (Kolter & Johnson, 2011). WikiEnergy dataset is the power dataset released by the Pecan Street Inc company through the WikiEnergy project (Pecan Street Inc, 2013), which already contain more than 600 household electrical data, including household mains power and data from each independent household electrical source monitored at intervals of one minute. The WikiEnergy began in January 2011 and is still collecting data from most buildings. REDD dataset is a public dataset created by Kolter and Johnson (2011) for energy disaggregation research, which contains low-frequency with sampling period 3 s of power data from 6 different households and high-frequency current and voltage data of two households.

According to different working conditions, electrical equipment is mainly divided into three categories:

(1) electrical appliance with a single state. In the course of operation, the service rating and power rating of these appliances are basically the same.
(2) electrical appliance with limited multi-state. This type of electrical appliances usually has a limited number of discrete working states, the electrical power of each state is discrete, and the different power values are different working states.
(3) electrical appliance with continuous change state. The electrical power of such appliances has no constant mean value, but fluctuates continuously within a certain range.

In the data selection of this experiment, above three kinds of electrical appliances should be selected. Therefore, 5 electrical appliances are selected from household No. 18, No. 20, and No. 25 in WikiEnergy dataset, the No. 3 household with low-frequency data in REDD is selected for our study.

In WikiEnergy dataset, air conditioner, refrigerator, washing machine, microwave oven, and dish washer are selected for experiment. Four electrical appliances of the third household in the REDD data set are used for testing: LITE, MICR, BATH, and DIFF. The specific reasons for selecting the above electrical appliances for load disaggregation experiment are:

(1) The cost of supervised learning experiments on every electrical appliance in each household is tremendous and impractical, while the experiments on some common and representative electrical appliances are necessary.
(2) Our research does not involve the power consumption disaggregation of some low-power electrical appliances, because of such appliances in the data set are easy to be interfered by noise, and the noise information sometimes triggers the event detection of low-power electrical appliances, resulting in the wrong classification. The research of load disaggregation on low power consumption devices is still one of the challenges in NILD, we leave this for future exploration.
(3) The power consumption of above selected electrical appliances accounts for over 70% of the total consumption of the entire household, we conduct load disaggregation experiments on commonly used electrical appliances.

(4) The historical power consumption of the selected electrical devices includes mode disaggregation from simple to complex.

3.1.2. Differential processing

In load disaggregation, mains power of the instantaneous load is used as the sequence of characteristic observation values. The differential signal is a numerical representation of the difference between two physical quantities. Mains power values of adjacent time points after filtering are subtracted, and the differential waveforms of mains power are obtained. The consumption of mains power and its differential waveform are shown in Fig. 4. After differential processing, each non-zero value indicates that the state of the electrical appliance has changed. The formula for calculating the difference waveform is as follows:

$$\Delta X_t = X_t - X_{t-1}, \tag{12}$$

where X_t represents the mains power consumption at time point t, and X_{t-1} is the mains power consumption at time point $t - 1$, ΔX_t is the results of differential processing.

3.1.3. Data normalization

Different time series data often have different basic physical properties, which will affect the result of data analysis. In order to eliminate such influence between characteristics of the data, data normalization is performed on original power data. After normalizing, all the raw data can be mapped into the same order of magnitude, which is suitable for the comprehensive comparison and evaluation (Affonso, Rossi, Vieira, & de Leon Ferreira, 2017). Normalization methods usually include min–max standardization, Z-score standardization, decimal scaling standardization and so on. The min–max is used for normalization in our study. Min–max standardization, also called deviation normalization, is a linear change to the original data, and the result is mapped to $[0, 1]$. The conversion formula is as follows:

$$X_t^* = (X_t - X_{min})/(X_{max} - X_{min}), \tag{13}$$

where X_{max} is the maximum value of mains power sequence or power consumption sequence of target appliances, and X_{min} is the minimum value sequence, X_t denote unnormalized power value consumed at time point t, and X_t^* is the normalized results.

3.1.4. Input data with sliding window

In this work, the original mains power data and the differential processing data are used as the input data of the network, and the real power of each electrical appliance is used as the target sequences. Deep learning relies on large amounts of training data, thus, in our study, first 80% of the mains power sequence and target sequence are used as a training sequence, where the training
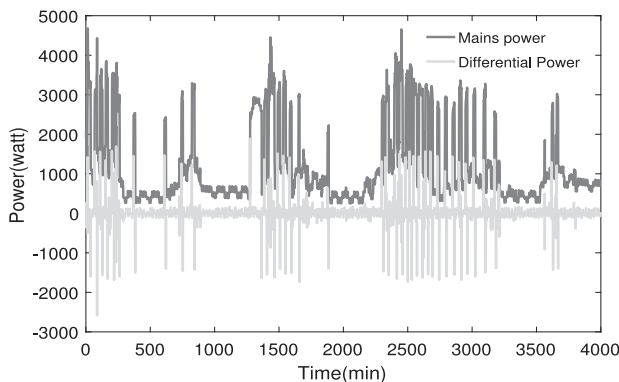
Fig. 4. Mains power consumption and differential power of mains power.

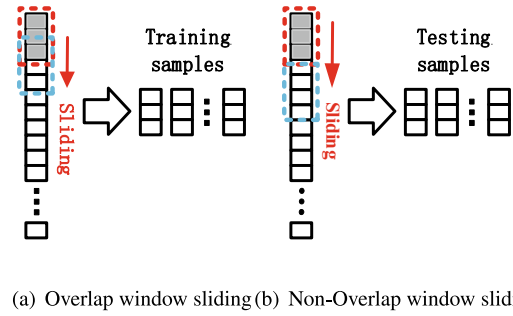(a) Overlap window sliding (b) Non-Overlap window sliding

Fig. 5. Different sliding pattern on training sequence and testing sequence.

sequence contains the mains power sequence and the power sequence of individual appliances, the remaining 20% of mains sequence and target sequence are regarded as the test sequence. The input of CD-LSTM requires to be a fixed size, which a sliding window is performed on training sequence and testing sequence. Fig. 5a shows the different window sliding strategy applied on the training sequence and testing sequence, as showed in Fig. 5a, overlapping sliding process of the training sequence is adopted to increase the training samples, such data augmentation skill has been widely used in image processing and sequence domain to improve the performance of deep learning algorithms. Assume that the length of the training sequence is z, a sliding window with window size n and stride μ performed sliding operation on training sequence, $(z - n)/\mu + 1$ can be obtained. Similarly, as showed in Fig. 5b, h/n test samples can be obtained by non-overlapping sliding on test sequence with length h and both window size and strides are n. CD-LSTM maps the mains power samples $X_{m:m+n-1}$ into the power samples $Y_{m:m+n-1}$ of the respective individual appliances. Where m represents the time at which the data begins to slide, n represents the size of the sliding window, and $m : m + n - 1$ represents the time segment from m to $m + n - 1$. Mains power samples and the samples of individual appliances are paired into (x, y) for training, where x, y represents the samples sliding from mains power sequence X and target power sequence Y.

3.2. Benchmark models

In this work, other four methods are compared with the proposed method. FHMM, CO, K-nearest neighbors (KNN) are classic methods for load disaggregation. Kelly and Knottenbelt (2015) indicated that the performance of DAE is better than traditional LSTM. Thus, the proposed model will be compared with FHMM, CO, KNN, and DAE. In the above method, FHMM and CO generate discrete data, which can only decompose electrical appliances into discrete values but cannot achieve continuous power value representation.

The FHMM model clusters the historical power data of each electrical appliance to get the cluster center for each electrical appliance, which can indicate the power values of each electrical appliance in different states. Then the HMM of a single appliance is obtained based on the number of clusters (i.e. the number of states). The parameters of a hidden Markov model include mean value, covariance value, initial probability, and state transition probability. The parameters of each hidden Markov chain for different appliances are different. Then the Kronecker product is used to obtain the initial probability, state transition probability, mean value and covariance of combined hidden Markov model. Each state of the combined hidden Markov model represents the state combination of various electrical appliances. With each input of a mains power data, the combined hidden Markov model will get a state, which represents the state combination of the electrical

appliances, and decodes the combined state to get the state of each electrical appliance.

Consistent with FHMM's model assumptions, the CO model first realizes the discrete state representation of each electrical appliance by clustering. Given a mains power data, the mains power of all possible combinations of electrical appliances is calculated. Then the optimal combination of load disaggregation can be obtained by finding the minimum residual error between the power of all possible appliances combination and real mains power.

KNN is a classical machine learning algorithm, which can not only applied for event detection in NILM but also predict the power consumption of individual electrical appliances by establishing regression learning, and carry out load disaggregation task. The principle of KNN algorithm is to use the target values of K nearest training samples to determine the value of testing samples. It predicting target load power consumption according to the similarity of the trained mains power sequence and target sequence.

DAE is an encode-decode structure algorithm based on neural network technology. Kelly and Knottenbelt (2015) modeled the process of load disaggregation as the process of denoising. DAE seeks to recover the clean power consumption information of target appliance from the mains power signal with noise. The classic DAE is an unsupervised algorithm which can reconstruct the input signal. Kelly and Knottenbelt (2015) trains the DAE with the mains power and the power consumption information of the individual appliance as the target, trying to find the mapping pattern between the mains power with noise and the power consumption information of the target appliances.

3.3. Evaluation measures

In order to compare the proposed method with other methods on all the data series, adequate evaluation indexes should be selected to identify the performance. In this work, two kinds of load disaggregation evaluation indexes are used: mean absolute error and signal aggregate error. The mean absolute error (MAE) is used to evaluate the error of forecast value and real value:

$$MAE = 1/T \sum_{i=1}^{T} |p_t - g_t|, \qquad (14)$$

where g_t is a true power of an electrical appliance at time t, p_t is the forecast value at time t, T is the number of predicted points. In this study, SAE is used to evaluate the error in a period of time:

$$SAE = |\hat{e} - e|/e, \qquad (15)$$

where $\hat{e}$ and e represent the predicted total energy consumption and real total energy consumption respectively. This measurement is very meaningful because it is helpful to the daily power report.

When the load power is disaggregated, the on/off state of the electrical appliance can be distinguished by thresholding, which can also reflect the accuracy of the disaggregation problem. The thresholds of Air conditioner, Fridge, Washing machine, Microwave, Dish washer are 100 W, 50 W, 20 W, 200 W and 100 W respectively. In terms of judging the on/off state of electrical appliances, this work uses four indexes: recall rate, precision, accuracy and F1 measure.

Recall is the ratio of correctly predicted positive observations to all observations in the actual class.

$$Recall = TP/(TP + FN), \qquad (16)$$

where True Positives (*TP*) – The number of both predicted electrical state and the real state are "on", False Negatives (*FN*) – The number of predicted state is "on" when the real state is "off". There are two possibilities: one is to predict the original positive class into a pos-

itive class (*TP*), and the other is to predict the original positive class as a negative class (*FN*).

Precision is the proportion of the number that is correctly predicted to be "on" state to the state that is predicted as "on" state.

$$Precision = TP/(TP + FP), \qquad (17)$$

where False Positives (FP)-Real state is "off" when the predicted state is "on". Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

$$Accuracy = (TP + TN)/(P + N), \qquad (18)$$

where P is the number of positive samples in real samples, N is the number of negative samples in real samples.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

$$F1 = 2 \times (Precision \times Recall)/(Precision + Recall). \qquad (19)$$

Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially the data set has imbalance class distribution. The on/off classification of electrical appliances in this work has imbalance class distribution.

4. Result

Load disaggregation as well as to the electrical on/off state judgment are two important aspects of load identification. To accomplish both aspects, we firstly use CD-LSTM to achieve load disaggregation, and then utilize the disaggregation results to identify the state of an electrical appliance.

Fig. 6 shows the disaggregation results of No. 18 household of WikiEnergy. As can be seen from Fig. 6, FHMM and CO have bad disaggregation results. The load disaggregation in FHMM and CO can only be roughly expressed by the combination of discrete states. Therefore, Fig. 6 shows that the predicted power is discrete for FHMM and CO. Therefore, optimization technologies such as FHMM and CO can provide consumers with operable power consumption feedback, showing which electrical appliances are on and which of them are off, but these methods cannot accurately offer the detailed power consumption information of each appliance. By contrast, KNN, DAE, and CD-LSTM can achieve good results in appliances with high frequency of use such as Air conditioner and refrigerator. However, the effect of KNN in Washing machines, Microwave, and Dish washer with low usage appliances is not as good as DAE and CD-LSTM. The effect of DAE depends in part on the adjustment of parameters, and each electrical appliance has to train a separate model. Therefore, in order to achieve good results for each appliance, it will take more experience in adjusting parameters for DAE. In this study, only a simple parameter adjustment is needed for CD-LSTM, and CD-LSTM can output several decomposed values of electrical appliances at the same time with good performance.

Table 1 shows the disaggregation performance on No. 18 household of WikiEnergy. As can be seen from Table 1, KNN has the minimum error in the Air conditioner in the MAE index. The proposed CD-LSTM method is better than other methods in the MAE index of four kinds of electrical appliances such as Fridge, Washing machine, Microwave, and Dish washer. In the SAE index, the DAE is with the smallest error in the dishwasher, but CD-LSTM method has made the smallest error in the other four electrical appliances. It can be seen from Table 1 that the results of method CO and FHMM are not good, while the KNN, DAE, and CD-LSTM methods have better results. Therefore, we focus on comparing the three methods of KNN, DAE, and CD-LSTM. In addition, for the Air condi-
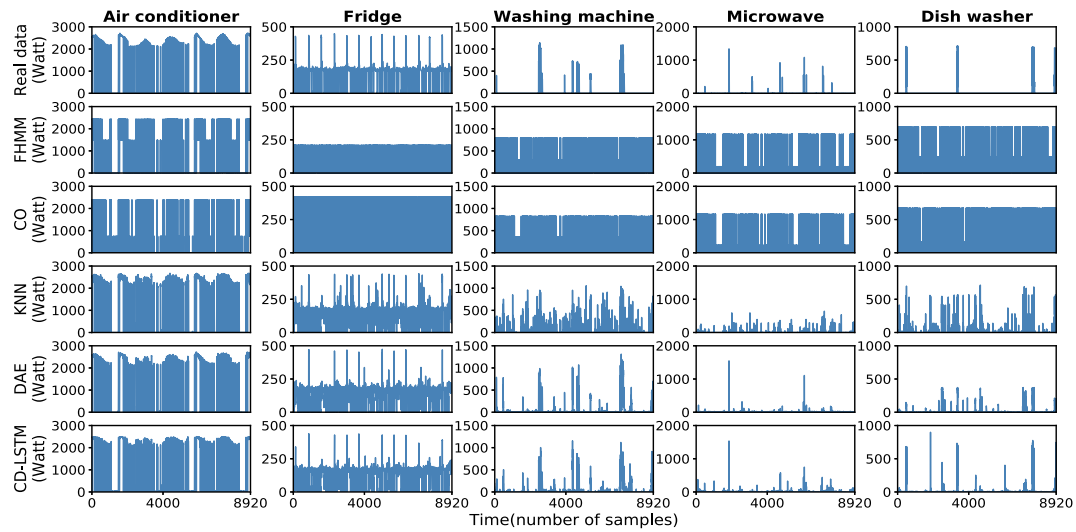
Fig. 6. Load disaggregation comparison of household 18 in WikiEnergy data set.

Table 1
Comparison of load disaggregation evaluation indexes of household 18.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
MAE	FHMM	153.594	96.120	267.563	144.714	232.219
	CO	330.280	161.521	225.114	243.283	213.127
	KNN	**38.232**	24.678	34.331	7.745	20.380
	DAE	83.287	18.823	**18.058**	5.226	26.618
	CD-LSTM	59.115	**18.122**	21.145	**5.076**	**10.562**
SAE	FHMM	0.140	0.022	7.881	60.335	14.001
	CO	0.276	0.664	6.352	107.287	12.600
	KNN	0.010	0.019	0.185	1.925	0.301
	DAE	0.013	0.036	0.187	1.285	**0.008**
	CD-LSTM	**0.003**	**0.016**	**0.161**	**1.224**	0.262

Bold item represents the best value for each column.

tioner, the above three methods are very effective, so here we only consider the comparison of the other four kinds of appliances.

Fig. 7 shows the finer grained load disaggregation comparison of three methods of DAE, KNN, and CD-LSTM for No. 18 household. The comparison results in Fig. 7 are time segments from Fig. 6. It can be seen from Fig. 7 that the proposed CD-LSTM can achieve the good fitting of various electrical appliances, which is superior to the DAE in the disaggregation of Fridge, Microwave, and dish washer. For the disaggregation of Washing machine, these two methods have different performances for different data points, but the difference of MAE is very small. The Fig. 7 also shows that the prediction results of KNN fluctuate greatly, and the results are worse than that of DAE and CD-LSTM.

Table 2 shows the comparison of evaluation indexes of electrical appliances on/off of No. 18 household. As can be seen from Table 2, the recall rate of the CO method for the Microwave is the highest, and the recall rate of the FHMM method is the highest for the dish washer. However, the precision of these two methods is very low, which indicates that these two methods classify too many negative samples into positive samples, and they prefer to divide the samples into positive samples. From the simulation results, it can be seen that the appliance with the greater the frequency of use or the longer the state of use, the more accurate the classification.

Fig. 8 shows the comparison of load disaggregation between five kinds of electrical appliances of No. 20 household. It can be seen from the simulations that the disaggregation results of FHMM and CO are very poor, and KNN is not ideal for sparse and abrupt

data (Washing machine, Microwave, and Dish washer). The good results of Air conditioner and Fridge can be obtained by DAE and the proposed model. For the appliance with sparse and abrupt data, the fitting effect of DAE and CD-LSTM is not good enough, but it is better than the methods of FHMM, CO, and KNN.

Fig. 9 shows the load disaggregation comparison of three methods of DAE, KNN, and CD-LSTM for No. 20 household. The comparison results in Fig. 9 are from time segments of Fig. 8. Fig. 9 shows that the performances of CD-LSTM and DAE are better than that of KNN, and the results of CD-LSTM are superior to DAE. As seen in Fig. 9a, Fridge is a multi-state device whose state is on in a long run, the disaggregation effect of DAE on Fridge is better in the area with higher power consumption, but the disaggregation error in the area with lower power consumption is significantly greater than that of CD-LSTM. Same conclusion can be drawn from the disaggregation results of Washing machine and Microwave. The above four disaggregation results show that although KNN has the ability of load disaggregation, the disaggregation error is large. Generally, the disaggregation curve of CD-LSTM is the closest to the real power consumption curve. In conclusion, the simulations demonstrate the effectivity of the proposed method.

Table 3 is the comparison of disaggregation error on No. 18 household of WikiEnergy dataset. Table 3 shows that the MAE of DAE is superior to CD-LSTM in the disaggregation of Air conditioner, but is inferior to CD-LSTM in the disaggregation of Fridge, Washing machine, Microwave, and Dish washer. For the SAE index, KNN has the best results in Air conditioner and Fridge. This is because KNN is not very good at fitting for the high frequency used
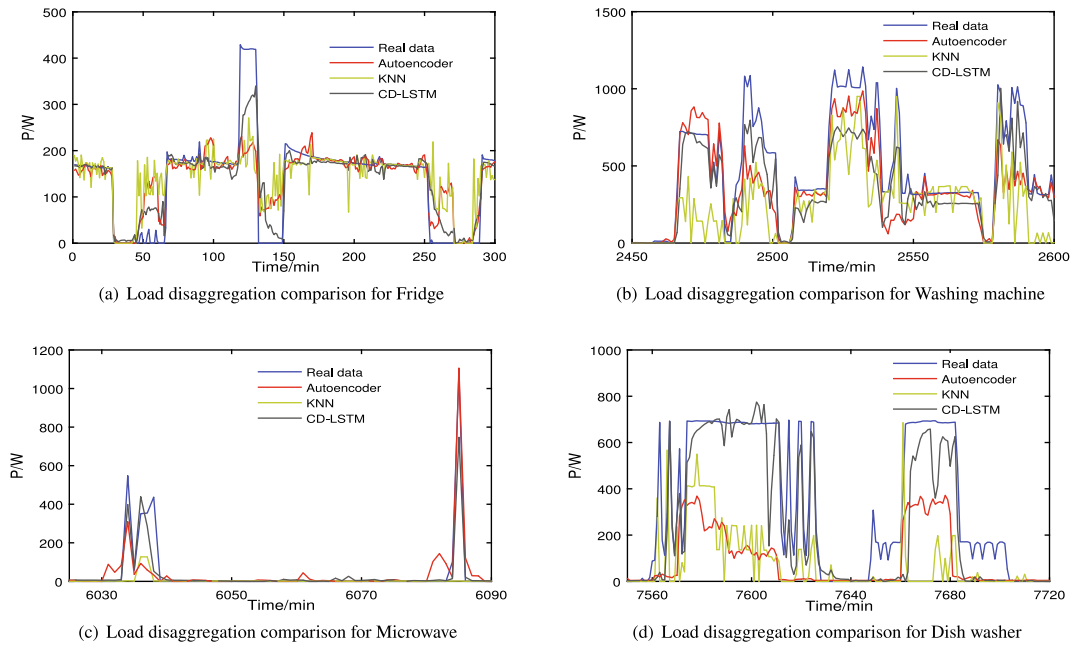
(a) Load disaggregation comparison for Fridge

(b) Load disaggregation comparison for Washing machine

(c) Load disaggregation comparison for Microwave

(d) Load disaggregation comparison for Dish washer

Fig. 7. Load disaggregation comparison of three methods of DAE, KNN, and CD-LSTM for No. 18 household.

Table 2
Comparison of evaluation indexes of electrical appliances on/off of No. 18 household.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
Rec	FHMM	0.944	0.645	0.664	0.216	**0.775**
	CO	0.902	0.683	0.580	**0.690**	0.676
	KNN	0.987	0.989	0.856	0.111	0.544
	DAE	0.997	0.994	0.969	0.296	0.767
	CD-LSTM	**0.998**	**0.994**	**0.976**	0.279	0.631
Prec	FHMM	0.977	0.661	0.067	0.015	0.042
	CO	0.829	0.689	0.075	0.004	0.040
	KNN	0.945	0.886	0.382	0.039	0.303
	DAE	0.827	0.914	**0.550**	0.423	0.255
	CD-LSTM	**0.979**	**0.922**	0.464	**0.530**	**0.412**
Acc	FHMM	0.974	0.540	0.433	0.947	0.375
	CO	0.904	0.580	0.551	0.462	0.423
	KNN	0.976	0.906	0.899	0.989	0.932
	DAE	0.930	0.934	**0.945**	0.995	0.903
	CD-LSTM	**0.992**	**0.939**	0.924	**0.996**	**0.949**
F1	FHMM	0.960	0.653	0.122	0.029	0.080
	CO	0.864	0.686	0.133	0.007	0.076
	KNN	0.965	0.935	0.529	0.058	0.390
	DAE	0.906	0.953	**0.702**	0.348	0.382
	CD-LSTM	**0.989**	**0.957**	0.630	**0.367**	**0.498**

Bold item represents the best value for each column.

electrical appliances, but it has good performs in overall trend tracking.

Table 4 shows the comparison of the evaluation indexes of on/off judgement of No. 20 household. It can be seen from Table 4 that the CD-LSTM method proposed in this study has obvious advantages over other methods, especially in the disaggregation of the following four kinds of electrical appliances: Fridge, Washing machine, Microwave, and Dish washer. The CO method has a good recall rate in the disaggregation of Washing machine, but the precision, accuracy, and F1 values are very low, which indicates that the CO method misclassifies of too many negative samples into positive samples.

Fig. 10 shows the comparison results of the five electrical load disaggregation of No. 25 household. Similar to the results of No. 18 and No. 20 household, the proposed method and DAE method

are better than FHMM, CO, and KNN. The comparison results in Fig. 11 are time segments from Fig. 10. Fig. 11 shows that the CD-LSTM has the best curve approximation ability. As viewed from Fig. 11a that the disaggregation results of Fridge presents three stages of change in this time period, the power fluctuates back and forth around 0 watts, 180 watts, 700 watts. The disaggregation results of KNN are comparatively unstable and the curve fluctuates sharply. Although KNN can approximately draw the trend of power consumption of Fridge, it has not yet achieved accurate disaggregation results. In contrast, the disaggregation curves of DAE and CD-LSTM are closer to the real power consumption curves. However, by further comparison, compared with CD-LSTM, DAE seems difficult to capture the time points when the power fluctuate greatly, similar phenomena are also reflected in the power disaggregation of the Washing machine in No. 25 household of WikiEnergy. This
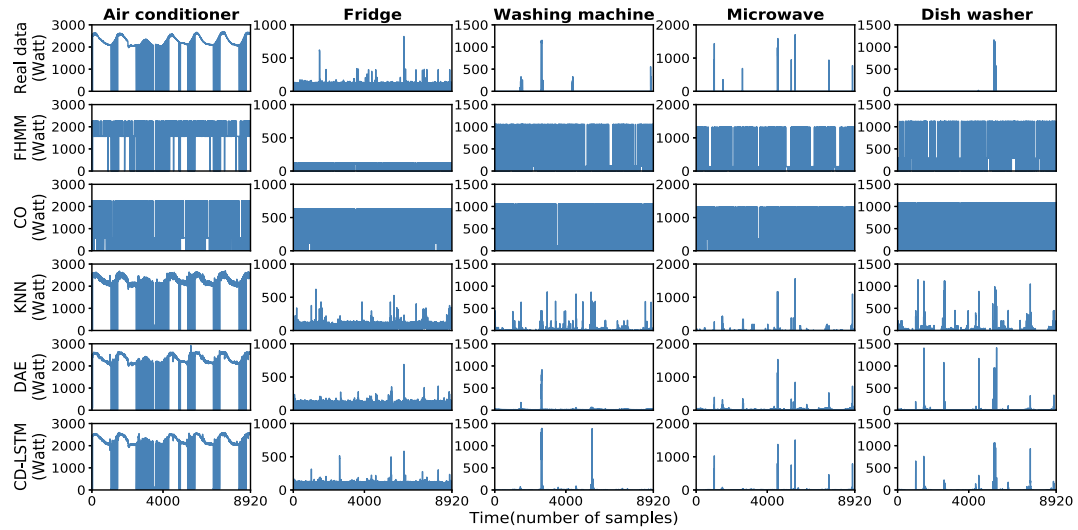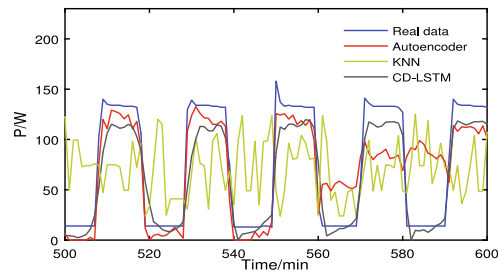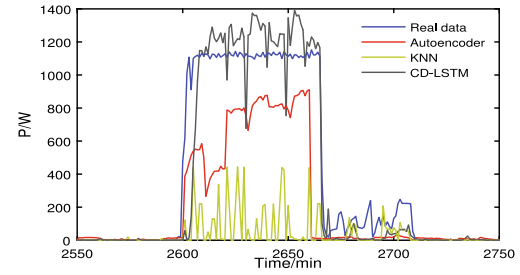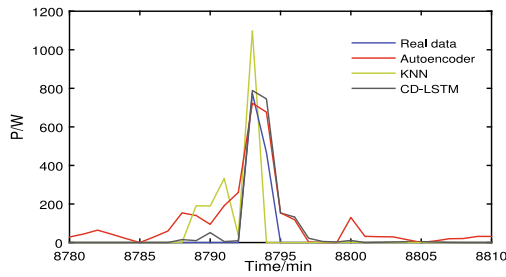
Fig. 8. Load disaggregation comparison for No. 20 household.

(a) Load disaggregation comparison for Fridge

(b) Load disaggregation comparison for Washing machine

(c) Load disaggregation comparison for Microwave

(d) Load disaggregation comparison for Dish washer

Fig. 9. Load disaggregation comparison of three methods of DAE, KNN, and CD-LSTM for No. 20 household.

Table 3
Comparison of load disaggregation evaluation index of number No. 20 household.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
MAE	FHMM	426.230	59.695	333.501	405.658	388.749
	CO	865.723	246.275	405.168	525.447	390.098
	KNN	54.491	45.156	14.096	4.223	17.217
	DAE	**38.547**	21.608	14.773	12.306	9.860
	CD-LSTM	45.433	**20.588**	**6.532**	**4.061**	**6.803**
SAE	FHMM	0.210	0.034	32.004	125.261	35.516
	CO	0.433	2.984	39.709	162.512	35.758
	KNN	**0.0007**	**0.017**	0.490	0.292	1.103
	DAE	0.010	0.029	0.156	2.341	0.580
	CD-LSTM	0.007	0.071	**0.055**	**0.028**	**0.046**

Bold item represents the best value for each column.

Table 4
Comparison of evaluation indexes of electrical appliances on/off of No. 20 household.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
Rec	FHMM	0.973	0.676	0.623	0.500	0.615
	CO	0.850	0.739	**0.740**	0.618	0.650
	KNN	0.997	0.803	0.290	0.324	0.333
	DAE	0.9990	0.949	0.333	**0.735**	0.608
	CD-LSTM	**0.9997**	**0.953**	0.293	0.588	**0.683**
Prec	FHMM	**0.998**	0.571	0.030	0.006	0.020
	CO	0.942	0.564	0.034	0.003	0.023
	KNN	0.985	0.676	0.230	0.275	0.046
	DAE	0.992	0.907	0.437	0.556	0.130
	CD-LSTM	0.994	**0.926**	**0.481**	**0.998**	**0.175**
Acc	FHMM	0.977	0.545	0.386	0.733	0.296
	CO	0.838	0.546	0.364	0.378	0.356
	KNN	0.986	0.679	0.943	0.994	0.899
	DAE	0.993	0.918	0.963	0.996	0.940
	CD-LSTM	**0.995**	**0.932**	**0.966**	**0.998**	**0.952**
F1	FHMM	0.985	0.619	0.057	0.012	0.039
	CO	0.894	0.640	0.065	0.007	0.045
	KNN	0.991	0.734	0.256	0.297	0.082
	DAE	0.996	0.927	**0.378**	0.633	0.214
	CD-LSTM	**0.997**	**0.939**	0.364	**0.741**	**0.278**

Bold item represents the best value for each column.

Fig. 10. Load disaggregation comparison for No. 25 household.

is mainly due to DAE are designed based on the assumption of the time-independency of power consumption data. On the contrary, because CD-LSTM combines the power fluctuation information of the previous time points in the disaggregation process of each time step, its ability to capture the time point when the electrical state transforms is stronger, and the disaggregation error is lower.

Table 5 shows the MAE and SAE index of different methods in most cases. It can be seen that the MAE of KNN and DAE is smaller than that of CD-LSTM in the disaggregation of Fridge. This is because in 1520–1580 min and 1880–1920 min, the CD-LSTM method failed to respond well to the closure of an electrical appliance. From the simulation results of above three households, it can be seen that the proposed method CD-LSTM is better than the other methods on the MAE index, and the performance on the SAE index is better than the other methods in most cases. If we pay attention to the fitting of electrical power curves, the CD-LSTM method is the best choice. If we only focus on the overall power consumption of an electrical appliance in a certain period of time, KNN and CD-LSTM method are both good choices.

Table 6 shows the comparison of appliances' state prediction of No. 25 household of WikiEnergy. It can be seen from Table 6 that CD-LSTM performs best in Air conditioner, Washing machine and Microwave, DAE performs best in Fridge and Dish washer. Although the recall rate of FHMM is very high, its comprehensive performance is very poor.

To verify the reliability and stability of the proposed method, REDD dataset is used for comparative experiments. Because of the poor effect of FHMM and CO in the experimental comparison of Wikienergy data set, the FHMM method and CO method are removed in the REDD data set experiment. The sampling period of REDD low-frequency data is 3 s. Four appliances were selected from No. 3 household for further experiment: LITE, MICR, BATH, and DIFF.

Fig. 12 shows the load disaggregation result for REDD data set with the method of DAE and CD-LSTM. It can be seen from the Fig. 12 these two methods have good disaggregation results on four kinds of electrical appliances. Due to the lower sampling period of REDD data set, the data of the time series are larger, and more
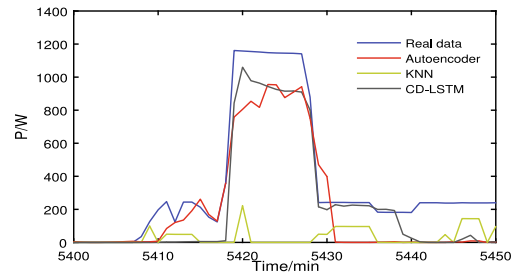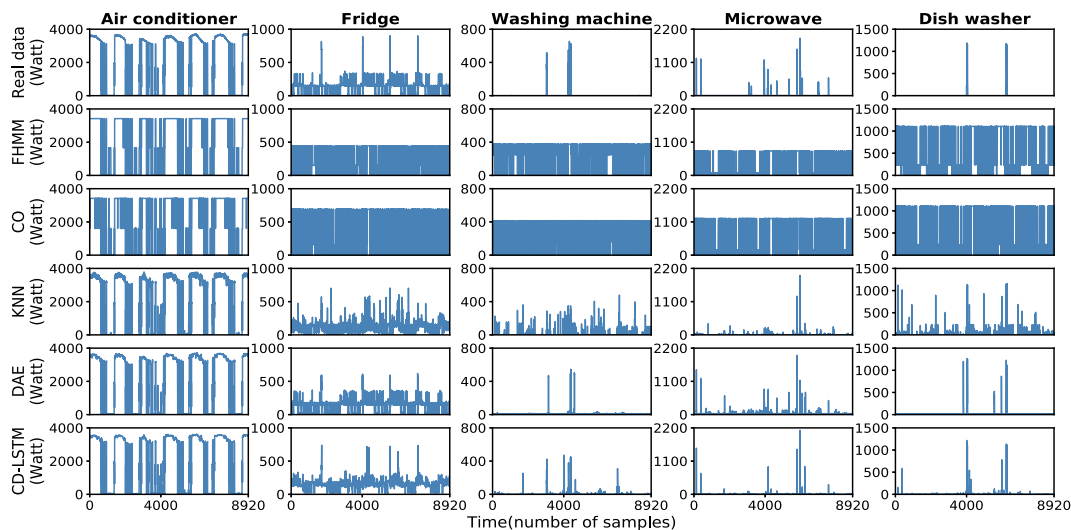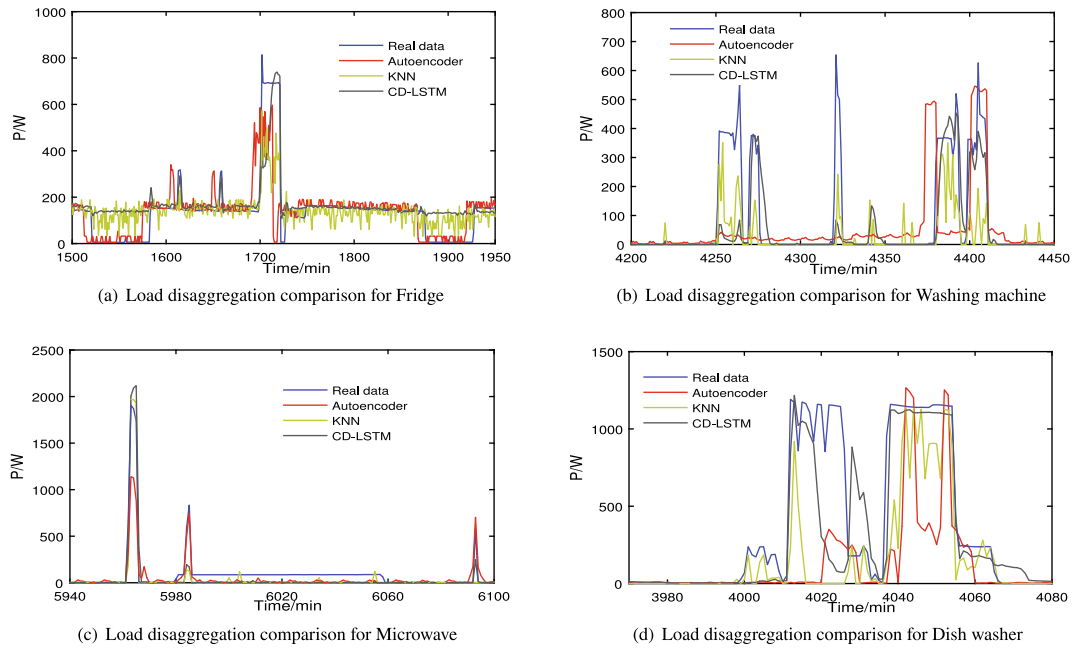
(a) Load disaggregation comparison for Fridge

(b) Load disaggregation comparison for Washing machine

(c) Load disaggregation comparison for Microwave

(d) Load disaggregation comparison for Dish washer

Fig. 11. Load disaggregation comparison of three methods of DAE, KNN, and CD-LSTM for No. 25 household.

Table 5
Comparison of load disaggregation evaluation index of number No. 25 household.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
MAE	FHMM	255.611	178.454	244.388	283.516	503.941
	CO	325.053	270.758	228.088	467.880	422.668
	KNN	42.366	**36.859**	5.846	6.366	9.873
	DAE	53.802	37.105	9.548	17.015	12.172
	CD-LSTM	**35.687**	37.854	**3.232**	**5.217**	**6.468**
SAE	FHMM	0.0321	0.967	40.811	50.039	61.031
	CO	0.0430	1.769	38.028	82.509	51.004
	KNN	**0.0002**	**0.0224**	0.277	**0.048**	0.159
	DAE	0.001	0.0792	1.359	1.979	**0.037**
	CD-LSTM	0.006	0.168	**0.111**	0.435	0.270

Bold item represents the best value for each column.

Table 6
Comparison of evaluation indexes of electrical appliances on/off of number No. 25 household.

Index	Method	Air-C	Fridge	Wash	Micr	Dish
Rec	FHMM	0.983	0.743	**0.949**	**0.806**	0.814
	CO	0.961	0.747	0.917	0.613	0.748
	KNN	0.998	0.983	0.595	0.194	0.595
	DAE	0.995	0.945	0.632	0.354	0.435
	CD-LSTM	**0.999**	**0.997**	0.861	0.387	**0.870**
Prec	FHMM	0.895	0.716	0.019	0.008	0.016
	CO	0.918	0.709	0.019	0.006	0.015
	KNN	0.990	0.864	0.141	0.429	0.143
	DAE	0.991	**0.942**	0.207	0.275	**0.606**
	CD-LSTM	**0.993**	0.853	**0.372**	**0.800**	0.295
Acc	FHMM	0.918	0.612	0.238	0.683	0.277
	CO	0.923	0.606	0.244	0.664	0.368
	KNN	0.992	0.877	0.964	0.996	0.942
	DAE	0.991	**0.920**	0.975	0.995	**0.987**
	CD-LSTM	**0.995**	0.876	**0.986**	**0.998**	0.968
F1	FHMM	0.937	0.729	0.037	0.015	0.031
	CO	0.939	0.727	0.036	0.011	0.030
	KNN	0.994	0.919	0.228	0.267	0.231
	DAE	0.993	**0.944**	0.312	0.309	**0.507**
	CD-LSTM	**0.996**	0.920	**0.519**	**0.522**	0.440

Bold item represents the best value for each column.
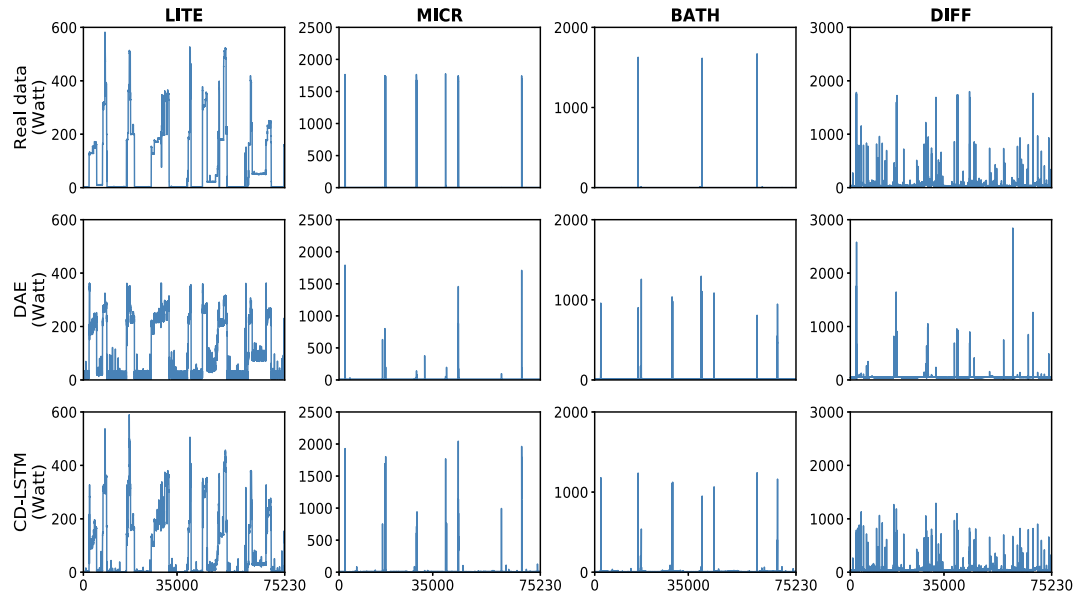
Fig. 12. Load disaggregation comparison for REDD data set.

(a) Predicted versus actual energy signal for LITE

(b) Predicted versus actual energy signal for DIFF
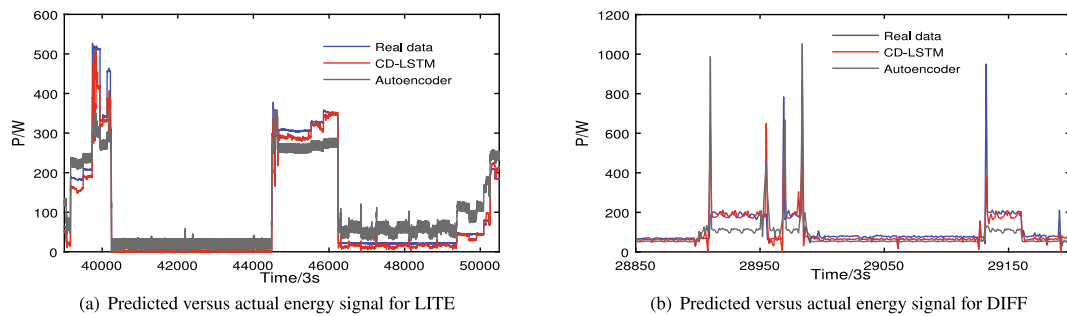
Fig. 13. Predicted versus actual energy signal for two appliances with two different methods of DAE and CD-LSTM.

training samples can be provided to improve the training effect. It can be seen from Fig. 12 that the disaggregation ability of the CD-LSTM method on DIFF is better than that of DAE. The comparison results in Fig. 13 are time segments of Fig. 12. Fig. 13 shows that CD-LSTM method can perfectly match the actual power of the appliances, it has better performance than DAE method, which indicates that the learning ability and generalization ability of CD-LSTM model is better than that of DAE.

Table 7 shows the comparison of load disaggregation evaluation indexes for REDD data sets. From the simulation results, in terms of the two indicators, CD-LSTM model is superior to DAE in four electrical appliances. From the practical experience, with the increase of samples, the results obtained by CD-LSTM model will be better. Table 8 shows the comparison of on/off indices of REDD data sets. As can be seen from Table 8, except the recall rate of LITE, CD-LSTM is superior to DAE. Table 8 indicates that when there are enough

Table 8
Comparison of evaluation indexes of electrical appliances on/off in REDD data set.

Index	Method	LITE	MICR	BATH	DIFF
Rec	DAE	**0.9995**	0.292	0.977	0.220
	CD-LSTM	0.972	**0.921**	**0.999**	**0.713**
Prec	DAE	0.939	0.831	0.079	0.258
	CD-LSTM	**0.999**	**0.857**	**0.091**	**0.478**
Acc	DAE	0.979	0.995	0.993	0.997
	CD-LSTM	**0.991**	**0.998**	**0.996**	**0.998**
F1	DAE	0.968	0.432	0.146	0.237
	CD-LSTM	**0.985**	**0.887**	**0.168**	**0.572**

Bold item represents the best value for each column.

samples to train the network, CD-LSTM can also get good results for sparse power data.

5. Conclusion

Load monitoring is an important part of household intelligent power consumption. In this study, NILD under low-frequency data is studied, which avoids the high cost of hardware and a large amount of high-frequency sampling data in the study of intrusive load disaggregation. Firstly, the mains active power is processed by the differential processing, then the original active power data

Table 7
Comparison of load disaggregation evaluation index in REDD data set.

Index	Method	LITE	MICR	BATH	DIFF
MAE	DAE	34.951	19.318	9.357	18.174
	CD-LSTM	**11.035**	**7.439**	**7.422**	**10.025**
SAE	DAE	0.191	0.389	8.247	0.311
	CD-LSTM	**0.087**	**0.209**	**6.616**	**0.121**

Bold item represents the best value for each column.

and the differential data are used simultaneously as the input of the network, thus enriching the fluctuation information of NILD. The proposed method is designed as an encoder-separation-decoder structure, while the encoder and the decoder are consists of deep LSTMs for solving the problem of time-dependency in multi-state devices; and the disaggregation the process is simulated as a signal separation process which implemented by a separation structure. The multi-branch output network model in this study can simultaneously disaggregate the active power value into multiple target appliances and avoid the problem that each electrical appliance should train a single model separately, thus reduces the complexity of the model. Finally, the cross-layer connection is introduced in the network to enhance the transmission of network information, thus improving the accuracy of model training. In the simulation experiments of two public data sets in Wikienergy and REDD, the method of this paper is compared with several load disaggregation methods studied previously, and the method of this method has a better disaggregation effect, and the accuracy and other indexes are obviously improved. We applied the proposed model on two data sets such as Wikienergy and REDD. Results on two benchmark data sets show that our method outperforms the advanced model, and the proposed CD-LSTM model can be a robust and efficient way to deal with NILD's issues, providing favorable technical support for realizing the bidirectional interaction of the smart grid and power consumers, and improving the scheduling level of the smart grid.

Although our model improves the performance of NILD, there are still some problems that need to be further discussed: (1) In our study, the design of deep LSTM only considers uni-directional dependency, while the real load disaggregation may have a bidirectional dependency. For example, there are multiple operation cycles (pre-washing, washing, drying) in the operation state of the Washing machine. Each cycle may interrelate in a long time. The current power consumption reasoning not only benefits from the previous power consumption information but also has a certain correlation with future power consumption. Therefore, in future work, we will consider the bidirectional dependency in load disaggregation, and some necessary improvements can be introduced to solve this problem, such as designing a deep bidirectional LSTM (Kiperwasser & Goldberg, 2016; Zeyer, Doetsch, Voigtlaender, Schlöter, & Ney, 2017) to capture the bidirectional dependency relationship (2) It is noted that in our experiment, some electrical appliances are in the "off" state in a long time, and the power consumption of such appliances in the "on" state is difficult to get accurate disaggregation. The reason for this phenomenon is that if NILM is regarded as a classification process based on event detection, the number of "on" of some electrical appliances in a period is far less than the number of "off", which makes the final prediction results tend to "off". In other words, there is a high degree of imbalance in the disaggregation data. Based on the above limitation, we will introduce some imbalance learning methods into our future work. Feasible data re-sampling methods such as synthetic minority oversampling technique (SMOTE) (Fernandez, Garcia, Herrera, & Chawla, 2018), a generative model such as a generative adversarial network (GAN) (Goodfellow, 2016; Antoniou, Storkey, & Edwards, 2017) to generate some "on" power consumption data would be a good choice. (3) In this study, a deep learning algorithm is proposed for NILD, but its black-box attribute is widely criticized. In our experiment, although the performance and efficiency of CD-LSTM have been improved, how the mains power fluctuation affects the power output of electrical appliances remains unknown. Thus, in future work, we will gradually tend to open the black box. Some interesting skills will be considered, such as introducing attention mechanism (Vaswani et al., 2017; Xia, Liu, Xu, Wang, & Zhang, 2019) to capture the time point of state change, and some model-agnostic methods would also be helpful

for us to open the black-box model, such as local interpretable model-agnostic explanations (LIME) (Ribeiro, Singh, & Guestrin, 2016) and SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Min Xia: Conceptualization, Methodology. **Wan'an Liu:** Writing - original draft, Formal analysis. **Ke Wang:** Data curation, Investigation. **Wenzhu Song:** Visualization, Investigation. **Chunling Chen:** Validation, Writing - review & editing. **Yaping Li:** Data curation.

References

Affonso, C., Rossi, A. L. D., Vieira, F. H. A., & de Leon Ferreira, A. C. P. (2017). Deep learning for biological image classification. *Expert Systems with Applications, 85*, 114–122.

Ahmadi, H., & Marti, J. R. (2015). Load decomposition at smart meters level using eigenloads approach. *IEEE transactions on Power Systems, 30*(6), 3425–3436.

Antoniou, A., Storkey, A., & Edwards, H. (2017). Data augmentation generative adversarial networks. arXiv preprint arXiv: 1711.04340.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one, 12*(7).

Basu, K., Debusschere, V., Bacha, S., Maulik, U., & Bondyopadhyay, S. (2014). Nonintrusive load monitoring: A temporal multilabel classification approach. *IEEE Transactions on Industrial Informatics, 11*(1), 262–270.

Batra, N., Dutta, H., & Singh, A. (2013). Indic: improved non-intrusive load monitoring using load division and calibration. *2013 12th international conference on machine learning and applications* (Vol. 1, pp. 79–84). IEEE.

Belley, C., Gaboury, S., Bouchard, B., & Bouzouane, A. (2015). Nonintrusive system for assistance and guidance in smart homes based on electrical devices identification. *Expert Systems with Applications, 42*(19), 6552–6577.

Chang, H. H., Chen, K. L., Tsai, Y. P., & Lee, W. J. (2011). A new measurement method for power signatures of nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications, 48*(2), 764–771.

Chang, H. H., Lin, L. S., Chen, N., & Lee, W. J. (2013). Particle-swarm-optimization-based nonintrusive demand monitoring and load identification in smart meters. *IEEE Transactions on Industry Applications, 49*(5), 2229–2236.

Chang, H. H., Lian, K. L., Su, Y. C., & Lee, W. J. (2013). Energy spectrum-based wavelet transform for non-intrusive demand monitoring and load identification. In *2013 IEEE Industry Applications Society Annual Meeting* (pp. 1–9). IEEE.

Chemali, E., Kollmeyer, P. J., Preindl, M., Ahmed, R., & Emadi, A. (2017). Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Transactions on Industrial Electronics, 65*(8), 6730–6739.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & Rizzoli, A. E. (2017). A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring. *Applied Energy, 185*, 331–344.

Cortez, B., Carrera, B., Kim, Y. J., & Jung, J. Y. (2018). An architecture for emergency event prediction using LSTM recurrent neural networks. *Expert Systems with Applications, 97*, 315–324.

Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8609–8613). IEEE.

Du, L., Restrepo, J. A., Yang, Y., Harley, R. G., & Habetler, T. G. (2013). Nonintrusive, self-organizing, and probabilistic classification and identification of plugged-in electric loads. *IEEE Transactions on Smart Grid, 4*(3), 1371–1380.

Esa, N. F., Abdullah, M. P., & Hassan, M. Y. (2016). A review disaggregation method in non-intrusive appliance load monitoring. *Renewable and Sustainable Energy Reviews, 66*, 163–173.

Evora, J., Hernandez, J. J., & Hernandez, M. (2015). A MOPSO method for direct load control in smart grid. *Expert Systems with Applications, 42*(21), 7456–7465.

Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research, 61*, 863–905.

Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., & Patel, S. (2010). Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing, 10*(1), 28–39.

Goodfellow, I. (2016). NIPS 2016 tutorial: generative adversarial networks. arXiv preprint arXiv: 1701.00160.

Graves, A. (2012). Supervised sequence labelling. In *Supervised sequence labelling with recurrent neural networks* (pp. 5–13). Berlin, Heidelberg: Springer.

Guo, Z., Wang, Z. J., & Kashani, A. (2014). Home appliance load modeling from aggregated smart meter data. *IEEE Transactions on Power Systems, 30*(1), 254–262.

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE, 80*(12), 1870–1891.

Hassan, T. (2012). Bi-level characterization of manual setup residential non-intrusive demand disaggregation using enhanced differential evolution. In *Proc. 1st Int. Workshop Non-Intrusive Load Monitoring.*

Hassan, T., Javed, F., & Arshad, N. (2013). An empirical investigation of VI trajectory based load signatures for non-intrusive load monitoring. *IEEE Transactions on Smart Grid, 5*(2), 870–878.

He, D., Du, L., Yang, Y., Harley, R., & Habetler, T. (2012). Front-end electronic circuit topology analysis for model-driven classification and monitoring of appliance loads in smart buildings. *IEEE Transactions on Smart Grid, 3*(4), 2286–2293.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv: 1508.01991.

Johnson, M. J., & Willsky, A. S. (2013). Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research, 14*(Feb), 673–701.

Kelly, J., & Knottenbelt, W. (2015). Neural NILM: deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments* (pp. 55–64). ACM.

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* (pp. 747–758). Society for Industrial and Applied Mathematics.

Kim, J., Le, T. T. H., & Kim, H. (2017). Nonintrusive load monitoring based on advanced deep learning and novel signature. *Computational Intelligence and Neuroscience, 2017.*

Kiperwasser, E., & Goldberg, Y. (2016). Simple and accurate dependency parsing using bidirectional LSTM feature representations. *Transactions of the Association for Computational Linguistics, 4*, 313–327.

Kolter, J. Z., Batra, S., & Ng, A. Y. (2010). Energy disaggregation via discriminative sparse coding. In *Advances in neural information processing systems* (pp. 1153–1161).

Kolter, J. Z., & Johnson, M. J. (2011). REDD: A public data set for energy disaggregation research. *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA* (vol. 25, pp. 59–62).

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid.*

Lee, W. K., Fung, G. S. K., Lam, H. Y., Chan, F. H. Y., & Lucente, M. (2004). Exploration on load signatures. *International conference on electrical Engineering (ICEE)* (Vol. 152).

Lin, Y. H., & Tsai, M. S. (2013). Development of an improved time-frequency analysis-based nonintrusive load monitor for load demand identification. *IEEE Transactions on Instrumentation and Measurement, 63*(6), 1470–1483.

Lin, Y. H., & Tsai, M. S. (2014). Non-intrusive load monitoring by novel neuro-fuzzy classification considering uncertainties. *IEEE Transactions on Smart Grid, 5*(5), 2376–2384.

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 2017-Decem (Section 2), 4766–4775.

Makonin, S., Popowich, F., Bartram, L., Gill, B., & Baji, I. V. (2013). AMPds: a public dataset for load disaggregation and eco-feedback research. In *2013 IEEE electrical power & energy conference* (pp. 1–6). IEEE.

Mauch, L., & Yang, B. (2015). A new approach for supervised power disaggregation by using a deep recurrent LSTM network. In *2015 IEEE global conference on signal and information processing (GlobalSIP)* (pp. 63–67). IEEE.

Martinez, R., Pacheco, D., & Robergs, D. (2018). NILM power disaggregation via artificial neural networks. In *Proceeding of the 4th international workshop on non-intrusive load monitoring* (pp. 1–4).

Martins, P. B. de M., Pinto, R. G. D., & Bittencourt, S. P. (2018). Load disaggregation of industrial machinery power consumption monitoring using factorial hidden markov models. The international workshop on non-intrusive load monitoring (NILM), 6.

Munshi, A. A., & Mohamed, Y. A. R. I. (2017). Extracting and defining flexibility of residential electrical vehicle charging loads. *IEEE Transactions on Industrial Informatics, 14*(2), 448–461.

Pecan Street Inc, (2013).http://DataPort.Cloud.

Park, S. W., Baker, L. B., & Franzon, P. D. (2019). Appliance identification algorithm for a non-intrusive home energy monitor using cogent confabulation. *IEEE Transactions on Smart Grid, 10*(1), 714–721.

Parson, O., Ghosh, S., Weal, M., & Rogers, A. (2014). An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence, 217*, 1–19.

Paterakis, N. G., Erdinc, O., Bakirtzis, A. G., & Catalao, J. P. (2015). Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. *IEEE Transactions on Industrial Informatics, 11*(6), 1509–1519.

Pereira, L., & Nunes, N. J. (2018). An experimental comparison of performance metrics for event detection Algorithms in NILM. In *Proceedings of the 2016 workshop on smart grids at the 2016 IEEE international conference on big data (BigData '16).* https://doi.org/10.1145/2674061.2675037.

Piga, D., Cominola, A., Giuliani, M., Castelletti, A., & Rizzoli, A. E. (2015). Sparse optimization for automated energy end use disaggregation. *IEEE Transactions on Control Systems Technology, 24*(3), 1044–1051.

Shao, H., Marwah, M., & Ramakrishnan, N. (2013). A temporal motif mining approach to unsupervised energy disaggregation: Applications to residential and commercial buildings. In *Twenty-seventh AAAI conference on artificial intelligence.*

Tsai, M. S., & Lin, Y. H. (2012). Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation. *Applied Energy, 96*, 55–73.

Rahimpour, A., Qi, H., Fugate, D., & Kuruganti, T. (2017). Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint. *IEEE Transactions on Power Systems, 32*(6), 4430–4441.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144).

Ridi, A., Gisler, C., & Hennebert, J. (2014). In *A survey on intrusive load monitoring for appliance recognition In 2014 22nd International Conference on Pattern Recognition* (pp. 3702–3707). IEEE.

Saitoh, T., Osaki, T., Konishi, R., & Sugahara, K. (2010). Current sensor based home appliance and state of appliance recognition. *SICE Journal of Control, Measurement, and System Integration, 3*(2), 86–93.

Singh, S., & Majumdar, A. (2017). Deep sparse coding for non-intrusive load monitoring. *IEEE Transactions on Smart Grid, 9*(5), 4669–4678.

Singhal, V., Maggu, J., & Majumdar, A. (2018). Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning. *IEEE Transactions on Smart Grid.*

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958.

Ting, F. F., Tan, Y. J., & Sim, K. S. (2019). Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications, 120*, 103–115.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Wang, Z., & Zheng, G. (2011). Residential appliances identification and monitoring by a nonintrusive method. *IEEE Transactions on Smart Grid, 3*(1), 80–92.

Welikala, S., Dinesh, C., Ekanayake, M. P. B., Godaliyadda, R. I., & Ekanayake, J. (2019). Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. *IEEE Transactions on Smart Grid, 10*(1), 448–461.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation, 1*(2), 270–280.

Xia, M., Liu, W., Wang, K., Zhang, X., & Xu, Y. (2019). Non-intrusive load disaggregation based on deep dilated residual network. *Electric Power Systems Research, 170*, 277–285.

Xia, M., Liu, W., Xu, Y., Wang, K., & Zhang, X. (2019). Dilated residual attention network for load disaggregation. *Neural Computing and Applications, 31*(12), 8931–8953.

You, G. W., Park, S., & Oh, D. (2017). Diagnosis of electric vehicle batteries using recurrent neural networks. *IEEE Transactions on Industrial Electronics, 64*(6), 4885–4893.

Zaen, J. Van, Achkar, C. El, Carrillo, R., & Hutter, A. (2018). Detection and classification of refrigeration units in a commercial environment: comparing neural networks to unsupervised clustering.Nilmworkshop.Org.

Zeyer, A., Doetsch, P., Voigtlaender, P., Schlöter, R., & Ney, H. (2017). A comprehensive study of deep bidirectional LSTM RNNs for acoustic modeling in speech recognition. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2462–2466). IEEE.

Zoha, A., Gluhak, A., Imran, M., & Rajasegarar, S. (2012). Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. *Sensors, 12*(12), 16838–16866.

Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., & Xie, X. (2016). Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. In *Thirtieth AAAI conference on artificial intelligence.*