Socio-demographic Information Extraction from Load Profile Using Convolutional Neural Network

Yibo Wang[1], Qian Wang[1], Zhengrun Wu[1(✉)], and Bing Zhu[2]

[1] Software College, Northeastern University, Shenyang, China
{20185315,20185224,20185178}@stu.neu.edu.cn
[2] Shenyang Electric Power Survey, Design Institute Co., LTD, Shenyang, China
zhubing@syepdi.com

Abstract. Reasonable estimation of socio-demographic information by using smart meter data is the application direction of future load profile user behavior analysis. The full mining of socio-demographic information has attracted more and more attention because the socio-demographic characteristics of consumers can help energy suppliers provide consumers with personalized services, thereby gaining an advantage in business competition. Nowadays, the simplicity of the current feature extraction methods has the information content of smart meter data not fully excavated, which leads to the low accuracy of the training model. This paper uses deep learning methods to infer the possibility of household socio-demographic characteristics from consumers' electricity smart meter data. A deep convolutional neural network (CNN) uses different feature extraction methods of one-dimensional convolution and two-dimensional convolution respectively, and some measures are used to prevent the model from overfitting. After a lot of repeated experiments, our model has a stronger identification ability than other previous models. That's because different feature extraction methods can better decompose consumers' heterogeneous electricity consumption behavior. Finally, we compare and discuss the results, thus supporting the modeling of users' electricity consumption behavior and the design of a customized demand management strategy.

Keywords: Convolutional Neural Network (CNN) · Deep Learning · One-Dimensional Convolution · Two-Dimensional Convolution · Socio-Demographic Information · Smart Meter

1 Introduction

Smart meter is the research foundation in the fields of nonintrusive load monitoring and load profile behavior analysis, which is used to frequently measure and observe household electricity consumption to obtain a large amount of fine-grained data [6]. At present, one of the main applications of smart meter data is the analysis of consumer behavior services, including demand-side management improvements and supply-demand linkage

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 703–715, 2023.
https://doi.org/10.2991/978-94-6463-010-7_72

incentive policies. The ultimate aim is to gain insight into consumer socio-demographic characteristics by extracting relevant demand patterns from consumers' electricity consumption habits and household characteristics and then to offer new income opportunities and improve demand-side management. For this goal, energy suppliers need to understand their consumers in an agile and efficient way [3]. The socio-demographic characteristics of consumer households can help energy suppliers provide consumers with personalized services, improve various aspects of their decision platforms, and thus gain an advantage in business competition. What's more, the socio-demographic characteristics of social families are also of vital importance to local government decision-makers. In particular, it is subject to the objective conditions of the census programs, such as the difficulty of obtaining the actual situation of the family and the uncertainty of frequent population movements. It is the application direction of future load profile behavior analysis to estimate household socio-demographic characteristics by using electricity meter data.

There are also many authors engaged in research in the field of identifying household socio-demographic characteristics from smart meter data. Beckel et al. [2] constructed a household characteristic estimation system called CLASS, where feature selection is implemented manually. Yi Wang et al. [7] used a deep CNN to extract data features of smart meters automatically. They solved the problem that the features extracted manually cannot effectively simulate the high variability and nonlinearity of simple load distribution, and then compared their model with existing advanced machine learning methods comprehensively.

Deep learning has great advantages in load forecasting [8], electricity price forecasting [10], etc. In previous works, most of the feature extraction methods are manual or automatic for the aggregated meter data, which is relatively simple. The information content of the smart meter data cannot be fully explored. Moreover, the accuracy of the training model is also affected by the overlapping degree of different dimensional features of the smart data.

This paper uses the same deep learning method to infer the possibility of a family's socio-demographic characteristics from consumers' smart meter data. We use one-dimensional convolution and two-dimensional convolution to decompose consumers' heterogeneous electricity consumption behavior, and ultimately support the analysis of consumers' electricity load profile user behavior and the design of a customized demand management strategy.

In summary, this paper makes the following three contributions:

1) It uses deep learning CNN to extract the features of smart meter data and adopts different feature extraction methods of one-dimensional convolution and two-dimensional convolution combined with a classifier to identify the socio-demographic information of consumers.
2) A large number of parameters involved in the CNN are prone to overfitting of model training, which leads to the weakening of model generalization ability. In order to alleviate the problem of overfitting, this paper puts forward the effective principle of parameter training and uses a variety of evaluation methods to measure the accuracy of the model. In order to alleviate the problem of overfitting, this paper puts forward

the effective principle of parameter training and uses a variety of evaluation methods to measure the quality of the model.
3) It shows the results of the training model with different feature extraction methods and draws a conclusion by comparing the results.

2 Problem Definition

In this paper, we obtain socio-demographic information from consumer questionnaires, including gender, age, social class, and living conditions of households.

Let $i \in I$ denote the individual family; let $i \in J$ denote the label of the survey question. Then $y_{i,j}$ denote the $j - th$ characteristic of the $i - th$ family. s_i is defined as the meter readings of the $i - th$ household over a period of time. In order to achieve better learning results for the deep learning model, we process the original aggregated electrical energy data through a feature extractor $E_{j,k}$ and then send it to the model for training:

$$\alpha_{i,j} = E_i(s_i, w_{1,j})$$

Then for the $j - th$ survey question label, the classifier $C_j(a_{i,j}, w_{2,j})$ needs to be trained. Among them, C_j represents the non-linear mapping relationship between the information of household electricity load profile after feature extraction and the label of the $j - th$ survey question. In the end, we can estimate the $j - th$ characteristics of the $i - th$ family through the model we trained:

$$\hat{y}_{i,j} = C_j(a_{i,j}, w_{2,j})$$

In the process of training the feature extractor $E_{j,k}$ and the classifier $C_j(a_{i,j}, w_{2,j})$, when the number of training samples is D_j, the cross-entropy loss function is used in it:

$$L_j(w_{1,j}, w_{2,j}) = \frac{1}{D} \sum_{i=1}^{D_j} [y_{i,j} \log \hat{y}_{i,j} - (1 - y_{i,j}) \log(1 - \hat{y}_{i,j})]$$

In summary, the model in this paper mainly solves the following three problems in the study of social demographic information:

1. Selection and training of feature extractors E_j to obtain feature extraction data $\alpha_{i,j}$.
2. Selection and training of classifiers $C_j(a_{i,j}, w_{2,j})$ to obtain the predicted socio-demographic characteristic $\hat{y}_{i,j}$.
3. Use appropriate model training methods to obtain model parameters $w_{1,j}$ and $w_{2,j}$.

3 Method

3.1 The Advantages of CNN

Automatic Feature Extraction: There is some current work using load profile to predict socio-demographic characteristics, which used some manually extracted features

and combined with some classification algorithms [6]. For example, Beckel, C. et al. [2] extracted the maximum and average features manually and applied SVM for classification. In fact, this method is unsatisfactory, because the features manually extracted by humans are subjective. It is difficult to predict whether these features are useful for classification. If these features are noisy for classification, they can even decrease the performance of the model. CNN can automatically change the parameters of the convolution kernel according to the gradient descent through the loss function of the classification error, and then automatically extract the features. So the advantages of CNN make the classification perform better and increase the accuracy of our model.

Non-linear Relationship: IN fact, the power load in an area usually has certain stability and periodicity related to the weather and time. As for the electrical load of a family, there are many influencing factors, including socio-demographic characteristics such as family income, family population, and housing area. In this way, the relationship between load profile and these characteristics is highly non-linear. Some traditional feature extraction methods are linear, such as the Principal component analysis (PCA) used by Yi Wang et al. [5]. A neural network with two or more layers can simulate the non-linear relationship in any situation by using a non-linear function as the activation function. Therefore, multi-layer convolutions and fully connected layers can be used to learn the nonlinear relationship between household load profile and socio-demographic characteristics.

Shared Weight: For the initial input of 7×48, if a fully connected network is used as the first layer to extract its features, a large number of training parameters are required. For CNN, it uses convolution to scan the original data, and every part of the data will be scanned by the same convolution kernel to share the parameters in the convolution kernel. Compared with the general fully connected network, it can greatly reduce the number of parameters of the network and reduce the risk of overfitting while making the model training faster.

Convolution Dimension: The use of one-dimensional convolution to automatically extract features pays more attention to the feature correlation between hours of household smart meter data in a day. The two-dimensional convolution is often used in the field of image processing, which can automatically extract local associated feature information on multiple channels of an image. Compared with one-dimensional convolution, the two-dimensional convolution used for household smart meter data focuses more on the characteristic correlation between the day and another day. Different feature extraction methods enable the information content of smart meter data to be fully excavated, and further improve the accuracy of our model.

3.2 Our Proposed Network

Tables 1 and 2 show the network structure of one-dimensional convolution design and two-dimensional convolution design. For such a network structure, the reasons are as follows.

Table 1. Hyperparameter And Parameters of One-Dimensional Convolution Model

Lay Type	Hyperparameters	Parameters
Conv1D	Input size: 48 × 7	176
	Kernel size: 3	
	Kernel number: 8	
Conv1D	Input size: 46 × 8	2064
	Kernel size: 16	
	Kernel number: 16	
MaxPooling1D	Input size: 23 × 16	0
Dropout	None	0
Dense	Input size:16	51

Table 2. Hyperparameter And Parameters of Two-Dimensional Convolution Model

Lay Type	Hyperparameters	Parameters
Conv2D	Input size: 48 × 7 × 1	56
	Kernel size: 2 × 3	
	Kernel number: 8	
Conv2D	Input size: 46 × 6 × 8	1168
	Kernel size: 3 × 3	
	Kernel number: 16	
MaxPooling2D	Input size: 44 × 4 × 16	0
Dropout	None	0
Flatten	None	0
Dense	Input size:704	1410

For a family, the smart meter data of a week is more cyclical than a day, because the former takes the difference between weekdays and weekends into account. And this periodicity reflects the electricity consumption behavior of this family. The data of higher time span has no significant additional power consumption characteristics compared with the data of one week and will increase the data dimension and slow down the training speed. In order to avoid the above situation and increase the number of samples to prevent overfitting, this paper selects 7 × 48 m data of one week as the input of our network.

The deeper the layers of the convolution network, the better the nonlinear expression ability. The deeper network can fit more complex feature input and is more conducive to the classification of population information. At the same time, our training data is

limited, so the network structure cannot be too complex to prevent overfitting. In this paper, we apply the two convolution layers.

The use of pooling layer can retain the main features and reduce the amounts of parameters. The input dimension of train data is 7×48, which is far smaller than the general application scenario (such as convolution image), so it is more reasonable to use only one layer of pooling.

It is also one of the ways to prevent overfitting by applying a dropout layer after the maxpool layer. It makes the activation of a neuron invalidate with a certain probability when propagating forward, and prevents the model from relying on a local feature of data to enhance the generalization ability.

There are 2291 hyperparameters in the one-dimensional convolution model and 2634 hyperparameters in the two-dimensional convolution model. Cross-validation and grid search are used to get their initial values.

3.3 Description of the Layers

Activation Function: Each neuron receives an output from the former node and transmits an input to the later node in the neural network. There is an activation function between the output of the former node and the input of the later node, such as $f_{\tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. These typical nonlinear functions make the expression ability of neural network optimized. But they have a certain degree of soft saturation with the disadvantage of the gradient disappearance. In this paper, we choose $f_{relu}(x) = \max(0, x)$ as our activation function, because there is no saturation problem when $x > 0$, keeping the gradient unchanged to alleviate the problem of gradient disappearance. Particularly, we select cross entropy as loss function and choose as an activation function to speed up weight updating.

Convolutional Layers: Convolution layer has the function of extracting data features. Each convolution layer has convolution kernels, and the weight of convolution kernel can be learned. In high-layer neural network, convolution operation can extract the desired features according to the objective function, and then increase the performance of the model [1].

The $j - th$ convolution kernel have learnable weight matrix $W_{i,j}$ and bias matrix $b_{i,j}$. The convolution result can be expressed as:

$$f_{conv}(X_{i,j}) = \sum_{j=1}^{M} (X_{i,j} * W_{i,j} + b_{i,j})$$

Full Connection Layer: Each neuron in this layer receives input from all the neurons in the former layer. Suppose Y_i represents the input vector before the fully connected layer i, and the neuron in the fully connected layer j has learnable weight vector $W_{i,j}$ and bias vector $b_{i,j}$. For the current input vector, the calculation result can be expressed as follows.

$$f_{dense}(Y_i) = Y_i * W_{i,j} + b_{i,j}$$

Pooling Layer: The pooling layer is applied between successive convolution layers, which is used to prohibit over fitting and keep the features unchanged. In this paper, we

choose the 1D and 2D maximum pooling layer for separate models. The large value in the matrix means that some specific features may be detected, which can better retain the features. Supposing the output of pool layer is a^l, and m, n is the area a_{ij}^{l-1} covered by the pool core, the forward propagation process of the maximum pooling layer is as follows.

$$a_{ij}^{l} = \max(a_{mn}^{l-1}),\ i \leq m,\ n \leq i+2$$

Dropout Layer: The direct function of dropout layer is to increase the orthogonality between features in each layer. Our model can produce the best result after cross validation when the dropout rate of hidden nodes is equal to 0.5. By using a random variable r which obeys Bernoulli binomial distribution with parameter p and adding this random variable to the input of standard neural network, the neural network with dropout can be described as:

$$r_j^{(l)} \sim Bernoulli(p)$$

Classification: CNN often uses softmax as a classifier. The output of the convolution layer and the input of softmax is z. C is the dimension and y_i is the probability that the prediction object belongs to the class.

$$y_1 = S(z)_i = \frac{e^{zi}}{\sum_{j=1}^{C} e^{zi}},\ i = 1, \ldots, C$$

Loss Function: The objective is to minimize the classification error, which is evaluated by cross entropy.

3.4 Prohibit Over Fitting

Data enhancement: Data enhancement is to enlarge the numbers of the sample according to the prior knowledge under the condition of keeping the label of the sample unchanged, so that the new samples also conform to or approximate to the real distribution of the data.

In the research of socio-demographic characteristics, we use low-frequency smart meters and socio-demographic characteristic questionnaire information for three months to get the corresponding relationship between electricity consumption behavior and socio-demographic characteristics. Even if every household's electricity consumption behavior may be affected by weather, emergencies and irregular behavior, it will not interfere with the performance of the model too much.

Previous studies have shown that weekly load support documents can reveal the social demographic information of consumers more or less. If the data set contains r-week smart meter data, the training data set can be expanded r times.

Dropout: Dropout makes the neural network independent of some features when training for many times. To enhance the generalization performance of neural network, we adopt the average result of several times of training, so that the predicted result of neural network will be stable after some neurons discarded.

4 Results and Discussion

4.1 Dataset Description

The data set we use in this section is provided by Commission for Energy Regulation (CER). This data set contains 4232 households with more than 536 days of smart meter data, which is recorded every 30 min. We selected 929 households participating in the power demand response plan. We deleted 147 households whose smart meter data was less than 536 days, leaving 782 households. For each family, we used 76 weeks of data to train our proposed CNN. We rank the IDs of all consumers in ascending order, and then use the first 80% as the training set and the last 20% as the test set.

The data set also provides two data sets containing questionnaire information, which contains socio-demographic information about consumers. We selected 13 survey questions from "Smart meters Residential pre-trial survey data" and used them as labels for our supervised learning. The specific question content and answer classification can be seen in Table 3. These questions include information about the occupants of the house, such as gender, age, and income, as well as information about the house, such as the age of the house and the number of bedrooms.

Table 3. Socio–demographic Information

No.	Question	Socio–demographic Information Question	Answers	Number
1	*200*	*Sex from voice?*	*Male*	*24928*
			Female	*24168*
2	*300*	*The age of the chief income earner?*	*18–35*	*76*
			35–65	*13148*
			65 +	*11248*
3	*310*	*The employment status of the chief income earner?*	*Employed*	*25916*
			Unemployed	*23180*
4	*401*	*Occupation of chief income earner is?*	*Less*	*5700*
			Medium	*20520*
			More	*21052*
5	*405*	*Have internet access in the home or not?*	*Yes*	*31768*
			No	*17328*
6	*406*	*Have broadband in the home or not?*	*Yes*	*27892*
			No	*3876*
7	*410*	*Best describes the people that live with?*	*Alone*	*11020*
			Over 15	*25916*
			With children	*12160*

(*continued*)

Table 3. (*continued*)

No.	Question	Socio–demographic Information Question	Answers	Number
8	*430*	*How many hours people stay at home during the day?*	*0–1*	*20672*
			2–4	*17024*
			5–6	*380*
9	*450*	*Best describes your home?*	*Apartment*	*1140*
			Bungalow/Detached house	*21280*
			House	*26524*
10	*4531*	*Approximately how old is your home?*	*0–10*	*152*
			30–75	*2584*
			75+	*2584*
11	*460*	*How many bedrooms in the home?*	*1–3*	*27664*
			4+	*21280*
12	*4704*	*Best describes how to cook in the home?*	*Electric cooker*	*34884*
			Non-electric cooker	*14212*
			None	*4408*
13	*4906*	*The proportion of double-glazed windows?*	*Low proportion*	*2736*
			High proportion	*41952*

4.2 Metrics

Generally speaking, for binary classification problems, we can get the values of TP, FN, FP, TN according to the confusion matrix. In the past, the evaluation standard of many works was to use the F1 score, which is an indicator of the accuracy of the binary classification model, to achieve the trade-off between the precision rate of the training classification model and the recall rate. It is defined as follows:

$$F_1 = 2\frac{Pr \times Re}{Pr + Re}$$

For multi-class classification problems, we can use the sum of TP, FP, and FN of each class to calculate the Micro-F1 score, which is defined as follows:

$$Pr = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} TP_i + \sum_{i=1}^{N} FP_i}$$

$$Re = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i}^{N} TP_i + \sum_{i}^{N} FN_i}$$

N is the total number of class of the current classification problem. In addition, we also used another metric, accuracy. Construct a $M \times M$ square matrix S, where $S_{a,b}$ represents the number of samples that class a is divided into class b. Only when $a = b$, $S_{a,b}$ represents the number of samples correctly classified by the model. Therefore, for multi-classification problems, the accuracy metric is defined as follows:

$$\sum A = \frac{\sum_{a=1}^{M} S_{a,a}}{\sum_{a=1}^{M} \sum_{b=1}^{M} S_{a,b}}$$

4.3 Results

Based on the description of Part III, we used two different convolution methods for extracting features in the same CNN architecture. One-dimensional convolution regards the user's one-day data as a sequence and seven-day data as 7 channels. It focuses on extracting the relationship of the user's electricity consumption behavior between hours in the same day. While two-dimensional convolution tends to extract the electricity consumption behavior characteristics of the same electricity usage time period in adjacent days.

Figures 1 and 2 show the result of accuracy and F1-Score obtained by using one-dimensional convolution and two-dimensional convolution. It can be seen that using two-dimensional convolution to extract features can increase the classification accuracy compared with one-dimensional convolution on average 3.18% and F1-Score on average 3.48%. We believe that this may be a manifestation of the stronger feature extraction capabilities of two-dimensional convolution compared to one-dimensional convolution. In view of the specific classification results of different socio-demographic information, we get the following conclusions.

First of all, regardless of whether one-dimensional convolution or two-dimensional convolution we used, the accuracy of prediction for (6) whether there is broadband, (12) cooking method and (13) double-glazed ratio is above 75%. This shows that the three socio-demographic information will have a significant impact on consumers' electricity consumption behavior. Even if different feature extraction methods are used, we have reason to believe that it is very easy to identify these three socio-demographic characteristics from smart meters. The presence of broadband directly affects users' online behavior and indirectly affects the power consumption of a series of electronic devices; whether to use electricity for cooking will directly affect the total power consumption; Glass is related to heat preservation. We guess that this will affect the use of household heating equipment, such as air conditioning and floor heating.

At the same time, the accuracy of (7) family member profile and (4) family classes using one-dimensional convolution is between 50% and 60%. We believe that these socio-demographic characteristics have weakened relationship with electricity consumption behavior. However, in the case of using two-dimensional convolution to enhance feature extraction, their accuracy is still improved by 3.22% on average. This gives

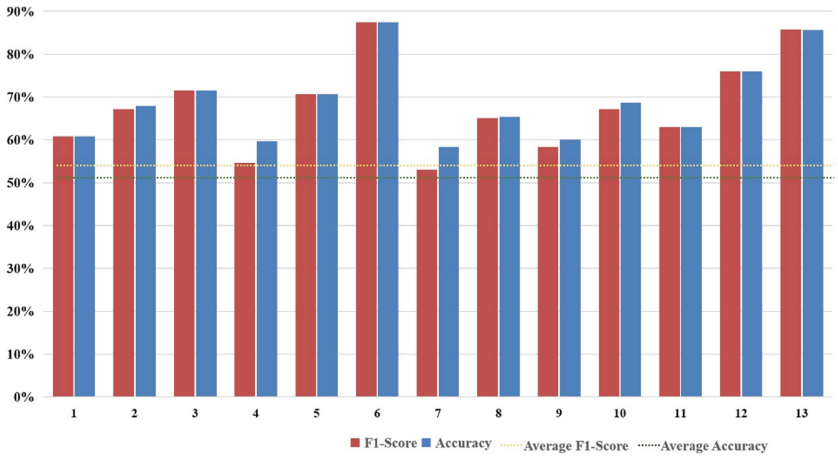

Fig. 1. Population information recognition using one-dimensional Con-volution

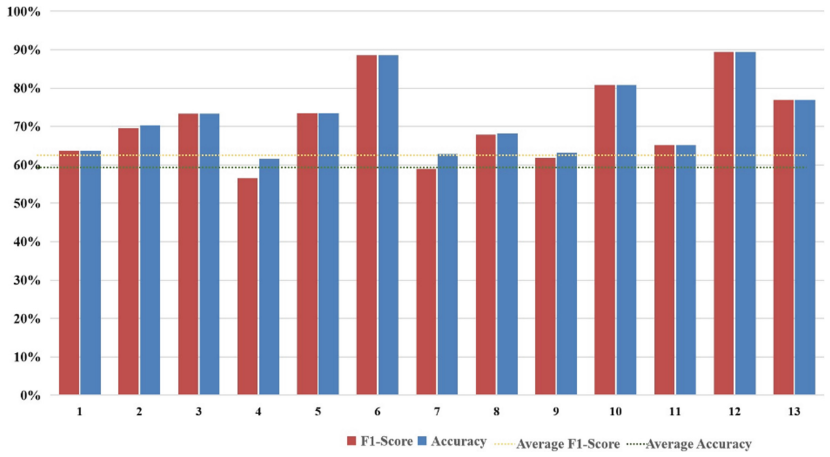

Fig. 2. Population information recognition using two-dimensional Con-volution

us inspiration that effective feature extraction methods can enhance the classification performance.

In addition, it can be seen that although there are 12 classification results indicating that the use of two-dimensional convolution will improve the accuracy, there is another result that shows the two-dimensional convolution reduces the accuracy, which may be due to overfitting. It inspires us that it is not always better to use higher dimensional convolution to extract features. The specific results depend on the real experiment rather than empirical judgment.

5 Conclusion

In this paper, we use a CNN-based method to identify consumers' socio-demographic information through smart meter data. We use one-dimensional convolution and two-dimensional convolution respectively, which can take into account the correlation between different times of the day and between different days. We use a variety of methods such as dropout, data enhancement, and weight sharing to prevent overfitting and promote the generalization ability of our model. However, there is still much room for improvement in our model. Among the classification results of our 13 socio-demographic survey questions, 5 of them have an accuracy of about 60% or less, which is unsatisfactory. Our CNN architecture is relatively simple. We may try to improve our network architecture in the future to strengthen our ability to extract features from smart meter data, thereby improving our classification accuracy. In addition, we will also try to compare or combine with other traditional feature extraction methods to improve the accuracy of model classification.

References

1. Pw A, Mwa B, Re A (2020) Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones - sciencedirect. Appl Energy 278:115563
2. Beckel C, Sadamori L, Staake T, Santini S (2014) Revealing household characteristics from smart meter data. Energy 78:397–410
3. Customer insights and collections strategies: a utility provider guide to optimisation, technical report (2013). Experian
4. Sadamori L, Beckel C, Santini S (2013) Automatic socio-economic classification of households using electricity consumption data. Futur Energy 75–86
5. Qiang Y, Yang L, Tianjian C, Yongxin T (2019) Federated machine learning: concept and applications. ACM Trans Intell Syst Technol 10:1–19. 01
6. Rouzbeh R, Amin G (2018) Rethinking the privacy of the smart grid: what your smart meter data can reveal about your household in Ireland. Energy Res Soc Sci 44:312–323
7. Wang Y, Chen Q, Gan D, Yang J, Kirschen DS, Kang C (2018) Deep learning-based socio-demographic information identification from smart meter data. IEEE Trans Smart Grid 1
8. Deng Z, Wang B, Xu Y, Xu T, Zhu Z (2019) Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting. IEEE Access 7(99):88058–88071
9. Deng Z, Liu C, Zhu Z (2021) Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network. Int J Electr Power Energy Syst 125:106499
10. Deng Z, Zhu Z, Wang Y, Guo H, Chai C, Wang B (2020) Unified quantile regression deep neural network with time-cognition for probabilistic residential load forecasting. Complexity 2020