

Study of Residential Power Load Patterns Based on Clustering and Deep Belief Network

Wang Baoyi^{a,*}, Lv Jin^b, Zhang Shaomin^c

School of Control and Computer Engineering, North China Electric Power University,
Baoding, 071003, China

a.email: wangbaoyiqj@126.com,

b.email:lvjin1992@foxmail.com, c.email:zhangshaomin@126.com

Keywords: Residential Power; Load Patterns; Clustering; Classification

Abstract: The study of power load patterns is the premise and basis of power distribution network maintenance. In view of the shortage of the existing power load model focusing on industry, agriculture, commerce and other large users, not on residents, In this paper, a method of residential power load patterns based on clustering and deep belief network is proposed. Firstly, use the improved k-means clustering algorithm for the residential electricity load clustering analysis to extract the typical load curve of each cell; and then a depth belief network classifier is constructed to classify the typical load curves of each cell, to identify the residential power load patterns and provide reliable support for distribution network maintenance. The effectiveness of the method is demonstrated by experiments on power data.

1. Introduction

Residents as a small part of the electricity users, because of their small amount, scattered characteristics, previous studies often pay attention to industrial, [1], agriculture, commercial [2] and other large users, and ignore residents, however, with the continuous improvement of the residents' electricity consumption and the popularity of smart meters, the study of the residential power load patterns has gradually become a hot spot in the field of intelligent electricity consumption.

In recent years, some scholars have carried out research on the power load patterns. The literature[3] through the parallel K-means cluster analysis divided intelligent residential users into the elderly, workers, the elderly workers, and then according to different types of residents' needs made the different scheme; In literature [4], a distributed clustering algorithm based on adaptive K-means and CFSFDP is proposed, and extracted 10 typical power load patterns. With the arrival of the big data age, the machine learning [5] method based on deep learning [6] technology can effectively analyze and process these data [7][8]. The survey found that some types of community similar, the electricity habits vary little, so the power load patterns in these areas are similar. If the power load patterns of each household in a residential district are classified one by one, it will increase the computational complexity of data processing. Therefore, only one kind of typical power load pattern is classified, which can represent the load patterns of all the residents, thus

reducing the complexity of data processing. In addition, most of the current clustering algorithms are used to classify the load patterns, whose result is in a variety of clutter, the purpose is not clear enough to increase the operating costs of the power company.

Based on the above analysis, this paper proposes a classification model based on clustering and deep belief network (IKDBNC). Firstly, the improved k-means clustering algorithm is used to cluster the power load of each district, and the typical load curve of each cell is extracted; Then a deep belief network classifier is constructed to classify the typical load curves of each cell and to identify the load patterns of the residents. The algorithm takes full advantage of K-means clustering analysis to reduce the amount of data and deep belief network classifier to classify clearly.

2. Design Idea of Residential Power Load Patterns Classification

The general idea of the study on the residential power load patterns is as follows: First of all, a sample of sufficient and comprehensive power load samples was obtained; Secondly, the improved k-means clustering is used to cluster the feature vectors, and the typical load curve of the cell is extracted; Then, the deep belief network classifier is used to classify the typical load curve of the district according to the classification rule of the power load pattern; Finally, an example is given to verify the accuracy of the method. The design idea of this method is shown in figure 1.

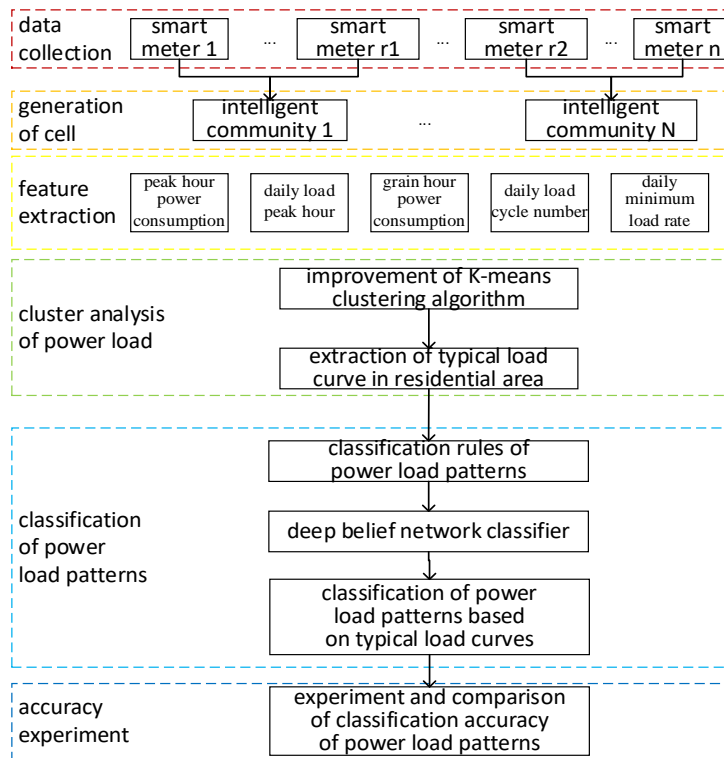


Figure 1 Design idea of residential power load patterns classification.

3. Classification Method of Residential Power Load Pattern

3.2 Improvement of K-means Clustering Algorithm for Residential Power Load

For a fixed feature weight is not flexible enough, according to the results of each iteration, dynamically different weights to attributes are assigned, with the iteration, the algorithm can automatically identify the different attributes of the importance weights, the direction of adjusting

the weights reflect the importance of different attributes of clustering[9].In addition, this section also aims at the problem that K-means clustering needs to determine the number of clusters in advance, and makes corresponding improvements. According to whether the clustering meets the tolerable error threshold, the number of clusters K is adjusted dynamically.

Taking cluster analysis of a cell load as an example, the main steps are described as follows.

Step 1: The daily load curve is expressed as a vector, and the vector normalization is processed, the n daily load curve can be represented as a $n \times 5$ row vector matrix X.

The Euclidean distance $d(b,c)$ between the weighted vector b and the vector c is obtained:

$$d(b,c) = \sqrt{\sum_{j=1}^5 w_j (x_{bj} - x_{cj})^2} \quad (1)$$

x_{bj} represents the data of row b and column j and x_{cj} represents the data of rows c and column j in vector rows X.

Step 2: all normalized feature data obtained by this cell are analyzed by K-means clustering which the value of K is K_{\min} , and K_{\min} clusters are obtained, and these clusters are labeled as untested clusters.

Step 3: check the untested cluster of the area to determine whether it meets the clustering error index, and judge whether the r cluster obtained by step 2 satisfies the clustering error index, or whether the r clustering meets the following conditions:

$$E_r = \sum_{j=1}^m w_j \sum_{l=1}^{n_r} (p_{rl}^j - C_r^j)^2 \leq B \quad r = 1, 2, \dots, K_{\min} \quad (2)$$

B is the tolerable error threshold, that is the overall maximum error that can be accepted by the clustering results; E_r is the total representation error of the rth cluster; n_r is the number of daily load curves that belong to the rth cluster; C_r^j is the jth feature of the cluster center of the r cluster; p_{rl}^j is the jth characteristic value of the lth daily load curve that belongs to the rth cluster.

Step 4: when the judgment result of testing an unchecked cluster in step 3 is correct, the checked cluster mark that satisfies the clustering error index B is the checked cluster; When the judgment result of the test of an unchecked cluster in step 3 is not, the K-means clustering analysis is carried out for the untested cluster that does not satisfy the clustering error index B, and the K value is 2, the clustering of the untested clusters that do not satisfy the clustering error index B is divided into two new clusters, and the two new clusters are labeled as untested clusters.

Step 5: repeat steps 3 and 4 until all of the unchecked clusters in the cell have been labeled as checked clusters or the total number of clusters containing the unchecked clustering and the checked cluster has reached the maximum number of clusters, and all the clustering centers at this time are labeled as typical load curves.

3.2 Deep Belief Network Classifier for Load Curve Classification

The bottom of the deep belief network classifier (DBNC)is stacked by several layer Restricted Boltzmann Machine(RBM), and the top layer adds the last layer representing the expected output variable, that is the classification layer. Here, the classifier usually used Softmax, which is a

generalization of the Logistic classifier in classification problems, it also gives the probability of classification results with the results, and is suitable for nonlinear multi classification problems, if it combines with the deep belief network, it will get better performance of discriminant[10][11].

Step 1: Select sample data

The data collected from the local data center are used as the sample data, which are unlabeled data and can be used as pre training samples and test samples. The small number of tag samples used in the tuning phase can be obtained by collecting the load pattern recognition experimental data.

Step 2: Select characteristic variables

The peak power consumption, daily load peak time, power consumption at the valley time, daily load cycle number and daily minimum load rate are selected as inputs to the global deep belief network classifier.

Step 3: Coding for Power load patterns

The residential power load patterns can be divided into seven types: stationary type, midday load type, night load type, one peak type, two peak type, three peak type and other type (not belonging to the first six types), as shown in figure 2.

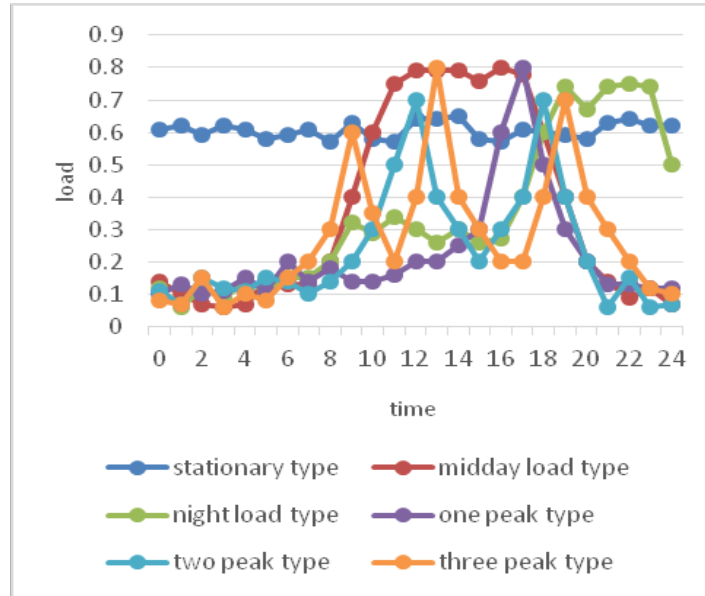


Figure 2 Residential power load patterns.

Step 4: Classification model of residential power load patterns

The input of the model is the five eigenvalue of the load curve, finally, the output of the top Softmax classifier is the probability value of the corresponding samples belonging to different states, and the state with the maximum probability is the final result of classification.

Step 5: Set the parameters $\theta = \{W_{ij}, a_i, b_j\}$ of RBM, the number of network RBM layer n and the initial learning rate ρ of global deep belief network classifier model. Where W_{ij} represents the connection weight between the visible unit i and the hidden unit j, a_i represents the offset of the visible unit i, and b_j denotes the offset of the hidden unit j.

Step 6: The unlabeled samples are trained by CD algorithm to preprocess the RBM layer at the bottom of the model.

Step 7: The whole network is optimized by BP algorithm by using the labeled samples of

training set.

Step 8: Save the trained network and test it.

4. Example Analysis

The data set used in this example analysis is from the Irish Commission for Energy Regulation (CER). In this paper, 3000 households were selected from the data, including 20 plots, the data are collected each 30min, 48 data points are collected for each user every day, and the acquisition lasted for nearly 1 years, so there are a total of 1 million load curves.

In the improved K-means clustering stage, through changing the tolerable error threshold B , it calculated α which represents the ratio of the number of cluster centers obtained by the improved k-means clustering analysis of each local data center to the total number of local data. As shown in Figure 3, as the B increases, the value decreases gradually, meaning that the larger the error of the clustering, the smaller the size of the local model. In particular, an inflection point can be found, and when the tolerance error is 0.4, it only requires the 1/5 local data to well represent all the load curves of the local data center, equivalent to the global clustering, the load curves are decreased by about 4/5.

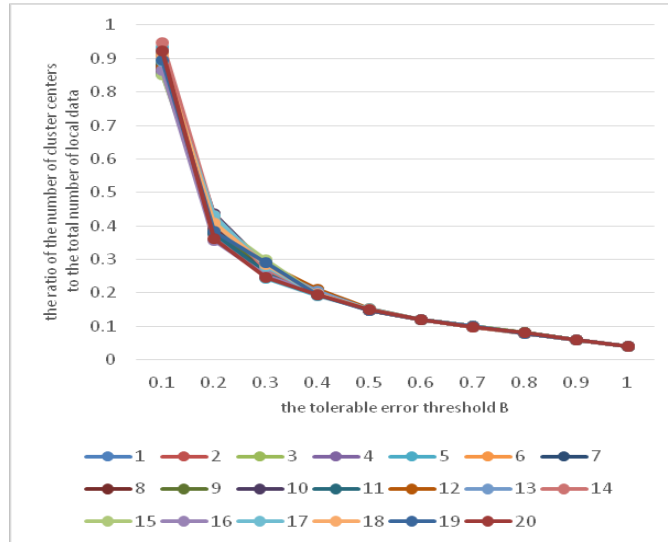


Figure 3 Relationship between the value of α and the threshold B .

So there are about 200 thousand load curves aggregated into the global data center. The characteristic values corresponding to the 200 thousand load curves are taken as the sample data, which are unlabeled data as pre training sets. In addition, 1000 groups of load pattern recognition experiment data were used as tuning set and test set according to 4:1.

The parameters of the IKDBNC model are set up, which the tolerance error threshold B is 0.4, the number of the network RBM layer is 6 and the initial learning rate of IKDBNC is 0.1.

For the IKDBNC method, the following tests are carried out from the measured data.

(1) The classification of power load patterns for pre training set of 10000, 50000, 100000 and 200000 was tested, as shown in table 2. It can be seen from the table that with the increase of the pre training set, the average correct rate of power load classification is increasing.

Table 2 Residential power load patterns classification case on the different pre-training sets.

serial number	pretraining set	debug set	test set	accuracy rate
1	10000	800	200	80.35%
2	50000	800	200	88.67%
3	100000	800	200	92.31%
4	200000	800	200	95.72%

(2) For the load pattern classification method based on SVM and BPNN, epochs and LR represent the BPNN maximum training iterations and the learning rate, they are 1500 and 0.01 respectively, SVM regularization coefficient and kernel parameter is 2048 and 0.03 respectively, the initial values of these parameters were determined from the experimental data by a large number of experiments to determine. Tag samples are used to train them separately, the load pattern classification for the test set is shown in table 4.

Table 4 Residential power load patterns classification case based on IKDBNC, BPNN and SVM.

method	accuracy rate
IKDBNC	94.41%
SVM	88.33%
BPNN	83.79%

By Table 4, we can see that the IKDBNC model proposed in this paper is more accurate than the classification methods based on BPNN and SVM.

5. Conclusions

This paper constructs a classification model based on improved k-means and deep belief network. It only needs to classify a certain class of typical load pattern in the area through this model which can get all the residential power load pattern classification, thereby greatly reducing the time complexity of data processing.

This paper abandons the traditional classification method of power load curve, namely clustering algorithm, and puts forward the IKDBNC model to classify the power load pattern, this classification is carried out according to the regulations in advance, which can meet the power company's needs better, and make promotion service to provide the basis for its value.

References

- [1] Yundong Gu, Sujie Zhang, Junshu Feng. Multi-model fuzzy synthesis forecasting of electric power loads for larger consumers[J]. Transactions of China Electrotechnical Society, 2015, 30(23):110-115.
- [2] Na Yu, Lezheng Yu, Guoqing Li. Controllable load management strategy for commercial users based on multi-agent in smart grid environment[J]. Automation of Electric Power Systems, 2015, 39(17): 89- 96.
- [3] Suxiang Zhang, Jianming Liu, Bingzhen Zhao, et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology, 2013, 37(6): 1542-1546
- [4] Wenjun Zhu, Yi Wang, Min Luo, et al. Distributed Clustering Algorithm for Awareness of Electricity Consumption Characteristics of Massive Consumers[J]. Automation of Electric Power Systems, 2016, 40(12):21-27.
- [5] Karun P Warrior, Shrenik M, Nimish Soni. Short-term electrical load forecasting using predictive machine learning models[C]// India Conference. IEEE, 2017.
- [6] Bin Yu, Shaozi Li, Suxia Xu, et al. Deep learning: a key of stepping into the era of big data[J]. Journal of Engineering Studies, 2014, 6(3): 233-243
- [7] Haben Stephen, Singleton Colin, Grindrod Peter. Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data[J]. IEEE Transactions on Smart Grid, 2016, 7(1):136-144.

- [8] Fahiman Fateme, M.Erfani Sarah, Rajasegarar Sutharshan, et al. Improving load forecasting based on deep learning and K-shape clustering[C]// International Joint Conference on Neural Networks. IEEE, 2017:4134-4141.
- [9] Sihai Li. Zibin Man. K-means Clustering Algorithm Based on Adaptive Feature Weighted[J]. Computer Technology and Development, 2013, 23(6):98-101.
- [10]Yong Ma, Changchun Bao, Bingyin Xia. Speaker segmentation based on discriminative deep belief networks[J]. Journal of Tsinghua University (Science and Technology), 2013, 53(6): 804-807.
- [11]Xin Shi, Yongli Zhu, Churila Sa, et al. Power transformer fault classifying model based on deep belief network[J]. Power System Protection and Control, 2016, 44(1): 71-76.