IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 7, 2020, accepted January 29, 2020, date of publication February 3, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971033

Incorporating Load Fluctuation in Feature Importance Profile Clustering for Day-Ahead Aggregated Residential Load Forecasting

NANTIAN HUANG[1], WENTING WANG[1], SINING WANG[2], JUN WANG[2], GUOWEI CAI[1], AND LIANG ZHANG[1]

[1]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China
[2]Beijing Zhongdian Puhua Information Technology Company, Ltd., Beijing 100192, China

Corresponding author: Nantian Huang (huangnantian@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0900104, and in part by the Research and Development of Industrial Technology in Jilin Province under Grant 2019C058-8.

ABSTRACT Residents clustering in different periods of load fluctuation and aggregated forecasting can increase the load prediction accuracy. But the strength of load fluctuation reflects the difference in the electricity consumption behavior of residents and affects the cluster results of residents. This paper presents a new day-ahead aggregated load-forecasting method for distribution networks based on the load fluctuation and feature importance (FI) profile clustering of residents. First, the input features are determined, the FI profile of residents is determined, and residents are clustered according to the FI profile. Then, the crow search algorithm is used to optimize the initial cluster centers for preventing the clustering results from falling into a local optimum. And the cluster verification index S_Dbw, the sum of the average scattering for the clusters and the inter-cluster density, is used to evaluate the cluster quality. The optimal clustering results of the aggregated load for different fluctuation periods are determined via statistical experiments. Finally, a random forest predictor based on ensemble learning is selected. According to the optimal clustering results in different fluctuation periods, a rolling forecasting model is constructed to realize day-ahead aggregated load forecasting in a residential distribution network.

INDEX TERMS Aggregated load forecasting, feature importance, load fluctuation, crow search algorithm, S-Dbw, random forest.

I. INTRODUCTION

Load forecasting plays a vital role in power system, including safe operation, economic optimization scheduling, and clean energy consumption [1]–[3]. With the large-scale popularization of smart meters [4], the high-resolution electricity load data recorded by smart meters can break the restrictions of the physical structure of traditional power system measurement and improve the accuracy of load forecasting [5]. Accurate aggregated day-ahead load forecasting for a distribution network is important for the economic and safe operation [6] of the distribution system.

Unlike traditional load forecasting, aggregated load forecasting is a bottom-up forecasting method based on smart meters, and the scale of the aggregated load is generally small. For residential distribution networks, the difference in residents' electricity consumption behavior will increase the fluctuation and complexity of the aggregated load, which will have a negative impact on power quality [7]. What's more, it will also make the aggregate residential load forecasting more difficult [8], [9].

The commonly used loading forecasting methods include the time-series method and the machine-learning method. The time-series method (such as the autoregressive moving average) only considers the impact of the time factor on prediction. They often exhibit great deviation when significant changes occur in other features [8]. Machine-learning is widely used in load forecasting and has achieved great results [10], [11], including artificial neural networks (ANNs), support vector regression (SVR), decision trees (DT),

The associate editor coordinating the review of this manuscript and approving it for publication was Dwarkadas Pralhaddas Kothari.

random forests (RF), deep neural networks (DNNs), etc. ANNs exhibit good performance for nonlinear loads. However, when the input dimension is high, constructing a suitable ANN is difficult. In SVR, the structure and parameters are adjusted according to the different inputs, and the optimization process is complicated. Ensemble learning methods [12] and deep learning methods [13], [14] have achieved satisfactory performance in load forecasting problems. For ensemble learning methods, Ref. [15] proposed an ensemble framework to predict the day-ahead average household energy consumption. And it confirms that the ensemble learning can bring solutions to the challenges that forecasting energy consumption at smaller aggregation level. Ref. [16] proves that the ensemble learning method can significantly improve the prediction accuracy of electricity load with high fluctuation. For deep learning methods, Ref. [5] proposes a model fusion method for different DNNs based on the deep learning. Moreover, the fusion of different DNNs and group prediction can cooperate with each other, which can effectively improve the prediction accuracy. In [17], an innovative neural network architecture consisting of a radial basis function, a convolution, a pooling, and two fully connected layers is proposed and used in load forecasting.

The traditional load-forecasting methods generally only build a single model for predicting the load in the area of interest, and they are not suitable for aggregated load forecasting. For aggregated load forecasting, related research shows that the application cluster identified customer groups with similar load consumption patterns from smart meters can improve the forecasting accuracy [6]. Load data collected by a smart meter (SM) provides a basis for analyzing the load characteristics of a large number of users [18]. The prediction accuracy can be increased by reducing intra-class differences [19]. Ref. [20] clusters high-dimensional data according to indicators describing the characteristics of electricity consumption. In [9], an aggregate load forecasting method based on SM data was proposed, which has the advantages of small computational burden and few requirements for historical data. In [12], the ensemble method was proposed for forecasting the aggregated load with sub-profiles. In [21], this was achieved by using K-means to cluster users according to the load curve. Then, ANN models were constructed for each class. The final load forecasting results are obtained by summarizing the forecasting results. In [22], a shape-based approach that classifies and predicts consumer energy consumption behavior at the household level was proposed. The method is based on the dynamic time warping. In [23], a finite mixture model based on clustering was presented. Data from four key time periods were used to form relevant attributes for clustering. In [24], an M-shaped model was used to cluster daily electricity load curves.

Although high-resolution smart meter data can improve load forecasting accuracy, the high input dimensionality of the prediction model caused by it cannot be ignored. Ref. [5] proposed an aggregated load forecasting method based on deep neural network and two-terminal sparse coding. And two-terminal sparse coding can achieve feature extraction and dimensionality reduction, which overcomes the challenges brought by fine-grained smart meter data. In addition, it can effectively improve the accuracy of day-ahead aggregated load forecasting. Ref. [25] proposed a smart meter data compression method based on stacked convolutional sparse auto-encoder. And this method can achieve significant enhancements in model size, computational efficiency, and reduction in reconstruction errors while maintaining the maximum number of details. It is important to effectively avoid the influence of high feature dimensions on the forecasting results of the aggregated residential load. RF is a common ensemble learning algorithm, and it can reduce the variance of the prediction results to some extent. Additionally, it screens the important features when DT generation occurs and its nodes split, reducing the effect of redundant features on the prediction accuracy [26]. Additionally, RFs have significant advantages to load forecasting with high-dimensional feature sets [27]–[29].

Existing research has achieved breakthrough results for aggregated load forecasting, but most of them model the each cluster of overall data. The impact of load fluctuation on forecast results is rarely considered. Moreover, the problem of whether the prediction model can adapt to high-dimensional feature inputs remains to be solved. The contributions of the present study are fourfold.

1) Load-fluctuation analysis: The load fluctuation of residents is strongly related to load forecasting. The proposed method is time-segmented in the sense that the load fluctuation varies in different periods. The method aims at clustering the aggregated load and establishing forecasting models for different fluctuation periods, as described in Section II.

2) Clustering based on feature importance (FI) profile: We use FI for clustering SM users, rather than the feature value, as described in Section III A. Clustering using load curves or statistical load features has drawbacks. It is difficult to ensure that users in the same cluster have similar responses to inputs of the predictor. Moreover, clustering via FI allows analysis of the effects of non-load characteristics on the future electricity load.

3) Optimizing initial centers and quality of clustering: The proposed clustering method, which is described in Section III B, is optimized by crow search algorithm (CSA) and S_Dbw. The CSA is applied to search the initial cluster center for preventing the clustering results from falling into a local optimum. And S_Dbw is used as the evaluation index of the clustering quality, which can improve intra-cluster similarity and inter-cluster difference.

4) Efficient forecasting approach: The proposed RF-based approach can reduce the impact of redundant features on the prediction result during the forecasting process, and no feature selection is required. And it also shows positive performance in reducing the accumulative errors of the rolling forecast. The forecasting results are presented in Section V, and Section VI concludes the paper.

II. DATA ANALYSIS AND TIME SEGMENTATION

The load characteristics of different distribution networks vary greatly. Therefore, it is important to study the residential load pattern of the target distribution network before establishing the prediction model.

A. RESIDENTIAL LOAD DATASET OF SM

In this study, SM data for residential customers from the Commission for Energy Regulation (CER) in Ireland are used, comprising the electricity consumption recorded at intervals of 30 min. In the case of missing values in the load profiles, the mean of the points in the vicinity of the missing ones was employed [21], [30]. If an SM had serious data loss, the corresponding load was omitted. Finally, 3790 residents were selected for the experimental analysis. We used the SM data from August 2009 to December 2010. The training set and test set were constructed as described in [21].

B. RESIDENTIAL LOAD FLUCTUATION IN TIME DOMAIN

Because the residential load consumption of different load customers differs in both magnitude and time of use, clustering load customers can improve the current practices for load forecasting [21]. Load fluctuations affect the user clustering and load forecasting, as demonstrated in Section V. In this section, the load fluctuation of the SM datasets for different periods is analyzed, using the standard deviation $\sigma(t)$, which is defined as:

$$\sigma(t) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(L_n(t) - \frac{\sum_{n=1}^{N} L_n(t)}{N})} \quad (1)$$

where $n(n = 1, 2, \ldots, N)$ is the number of SMs, and $L_n(t)$ is the load value of n^{th} SM in t.

Owing to differences in the geographical location and social environment, the electricity-consumption behavior differs for residents in different regions, and the results of load-fluctuation analysis vary for different datasets. Taking the CER dataset as an example, instead of finding the optimal number of load-fluctuation periods, it provides a new idea of dividing the load-fluctuation period. Fig. 1 shows the analysis and fluctuation of the total residential load to be forecasted for one year. Fig. 1(a) shows the standard deviation of the annual electricity load of the SM users. It embodies the load fluctuations for this year at t. A larger $\sigma(t)$ indicates greater fluctuation at t. Fig. 1(b) shows a box plot of whole power-consumption distribution of all the users at t. The black boxes represent the fluctuation range of the total residential load at t. The black lines in the boxes represent the median. The red circles represent outliers in the data. Longer boxes indicate a wider range of load fluctuations.

As shown in Fig. 1(a), the power-consumption fluctuations at night exhibit a decreasing trend. The first valley appears at 06:30 in the morning, and the fluctuation increases thereafter. The time of day with the greatest fluctuation is 19:30, and the fluctuation declines thereafter. As shown in Fig. 1(b), from 01:00 to 06:30, the fluctuation range of the load is small, but the outliers are numerous and concentrated. From 06:30 to 19:30, the fluctuation range is relatively stable, and the outliers are numerous and dispersed. There are no outliers from 19:30 to 00:30.

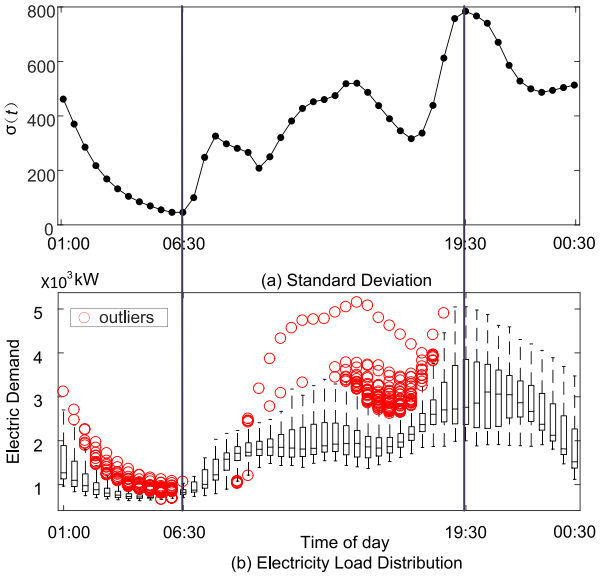


FIGURE 1. Standard deviation and electricity load distribution at every time point from August 2009 to August 2010.

The daily load from the CER dataset can be divided into three periods according to the standard deviation and load fluctuations.

a) Period 1: 01:00-06:30, low fluctuation; outliers are numerous and concentrated.
b) Period 2: 07:00-19:30, moderate fluctuation; outliers are numerous but scattered.
c) Period 3: 20:00-00:30, high fluctuation; no outliers.

Clearly, the load fluctuation differs for different periods. Thus, to improve the load-forecasting accuracy, we seek to cluster the customers according to the different periods of load fluctuation.

C. INFLUENCE OF LOAD FLUCTUATION ON CLUSTERING RESULTS

In this section, it is proven that the load fluctuation is related to the clustering results. Usually, the consumption behavior patterns (CBPs) of consumers are used as the K-means clustering basis. In this method, the data points can be arranged in a 3790 $\times$ 35 data matrix, where 3790 is the number of SMs, and 35 (five segments per day for the seven days of the week) is the number of features [21]. CBP-K-Means is the baseline clustering method used in this study.

Because we intend to utilize the SM data for load forecasting, the mean absolute percentage error (MAPE) of the predictor is regarded as an index for determining k_{opt}. A smaller MAPE indicates a more accurate forecast result. The MAPE

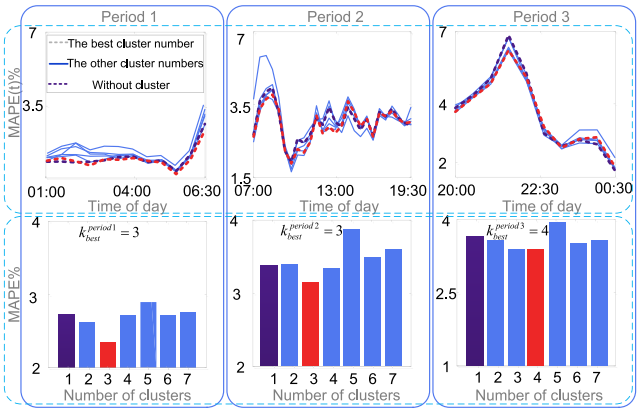


FIGURE 2. MAPE and MAPE(t) for three periods at different k values obtained using the CBP-K-means-predictor.

is defined as follows:

$$\text{MAPE} = \frac{1}{N_t}\sum_{n_t=1}^{N_t}\frac{1}{L_r}\left|L_r - L_p\right| \times 100\% \tag{2}$$

where L_r is the real load, L_p is the forecasting load, and n_t ($n_t = 1, 2, \ldots, N_t$) is the number of L_p values. We take the sum of the MAPEs at the same t for each day in the test set, and the average of these sums represents the error at this time point, i.e., MAPE(t). The MAPE used to select k_{opt} is obtained using the RF predictor. A detailed description of this method is presented in Section IV.

Fig. 2 shows the $k_{opt}^{period1}$, $k_{opt}^{period2}$, and $k_{opt}^{period3}$ obtained using the CBP-K-Means-Predictor. In the histogram, the red represents k_{opt}, and the black represents non-clustering. The fluctuation of the distribution-network aggregated load varies during different periods. k_{opt} must be selected specifically at various load fluctuations. The k_{opt} values corresponding to different periods are not equal. Clearly, the load fluctuation is related to the clustering results, and the clustering has positive effect on load forecasting.

III. OPTIMAL FI-CLUSTERING

The proposed method involves clustering SM users based on the FI rather than the feature value. Because different users have different FI sets, the FI can reflect different responses of different user loads to the predictor input features. Compared with CBP-clustering, FI-clustering is not limited by the type of features and can be used to analyze the relationship between multiple types of features and the predicted objects. The K-means clustering method is optimized in this section.

A. FI ANALYSIS BASED ON RReliefF

Numerous FI algorithms have been developed. After careful consideration, RReliefF (Algorithm 1)—a robust and widely used feature-weight calculation method—was selected for analyzing the FI of the SM users.

1) RReliefF ALGORITHM [31]

Each feature is given a final weight value $W(F)$ by calculating the weight $W_{dF}(F)$ between the features F, the weight

Algorithm 1 RReliefF

```
function RReliefF (W_dL, W_dF(F), W_dL&Df(F), W(F), A, b, f)
for a from 1 to A
    randomly select instance R_a
    select G instances Ig nearest to R_a
    for g from 1 to b
        W_dL = W_dL + diff(L, R_a, I_g) · d(a, g)
        for F from 1 to f
            W_dF(F) = W_dF(F) + diff(F, R_a, I_g) · d(a, g)
            W_dL&dF(F) = W_dL·dF(F) + diff(L, R_a, I_g)·
                         diff(F, R_a, I_g) · d(a, g)
for F from 1 to f
    W(F) = (W_dL·dF(F)/W_dL)
           − ((W_dF(F) − W_dL·dF(F)) /(f − W_dL))
```

W_{dL} between the electricity load values L, and the weight $W_{dL\&dF}(F)$ between the electricity load value L and feature F. Then, a random instance $R_a(a = 1, 2, \ldots, A)$ and its G nearest instances $I_g(g = 1, \ldots, G)$ are selected. The number of iterations is b, and W_{dL}, $W_{dF}(F)$, and $W_{dL\&dF}(F)$ are calculated. Finally, the importance of each feature is obtained via f iterations. The term $d(a, g)$ takes into account the distance between the two instances R_a and I_g. The function $diff(\,)$ [31] is used for calculating the distance between instances to find the nearest neighbors. Compared with other feature-weight analysis methods, RReliefF can address incomplete data and noisy data.

2) EFFECTIVENESS OF FI PROFILE CLUSTERING

RReliefF is applied to calculate the FI of each SM user. The SM users are clustered according to the FI, avoiding the effects of differences in the data types and power consumption on the clustering results.

To verify the effectiveness of the FI profile clustering, clustering based on CBP is employed as a baseline for comparative experiments. Four users were randomly selected for analysis. (The FI dimension is the same as the construction feature dimension of Section IV, i.e., 345. The CBP dimension for each user is 35. See Section II C.)

If the number of clusters is 2, the CBP clustering results show that user1 and user 2 are the same class and that user3 and user4 are in another class. However, in the FI profile clustering results, user1, user2, user3 and user4 are in the same class.

Fig. 3 shows the CBP and FI profiles for user1-user4. For the CBP clustering, the load curves for user1 and user2 are similar, and the load values are large. The load curves for user3 and user4 are similar, and the load values are small. For the FI profile clustering, although the four users have differences in the load values, the FI profiles are similar in shape. Additionally, the users are in the same cluster.

As shown in Fig. 3, the FI method allows analysis of different types of features during the clustering, avoiding the

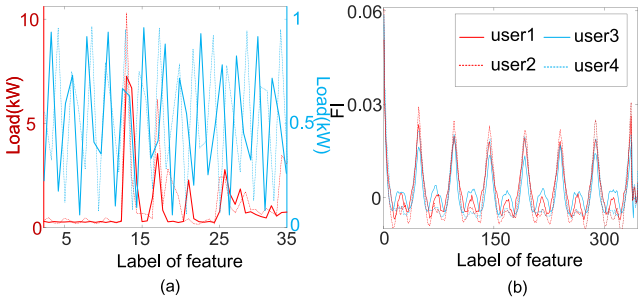


FIGURE 3. CBP and FI profiles for user1, user2, user3, and user4.

effects of user power-consumption differences on the clustering results. Compared with the CBP method, this technique improves the similarity of the responsiveness of intra-cluster users to the predictive load and enhances the prediction accuracy.

B. CLUSTERING IMPLEMENTATION FOR FI GROUPING

Clustering algorithms analyze the similarities and differences between different data by analyzing and mining the entire dataset. The K-means method is widely used in clustering. However, it has two shortcomings related to: 1) the criteria for measuring similarity and 2) the selection of the initial cluster center. For enhancing the quality of the clustering and reducing the clustering uncertainty, a new clustering algorithm called *S_Dbw*-based crow optimization K-means (SDCKM) is proposed.

1) CLUSTER QUALITY OPTIMIZATION

Let $X = \{x_q\}, q = 1, 2, \ldots, Q$ be a dataset with Q instances, and $C_1, C_2, \ldots, C_k$ be the k disjoint clusters of X. The K-means method divides X to minimize the sum of squared errors for all the classes. The clustering result is evaluated using the Euclidean distance after the clustering is completed [21], [24]. The clustering should ensure the smallest intra-cluster distance and the largest distance between clusters. However, the Euclidean distance only considers the intra-cluster similarity and ignores the scattering among clusters. To resolve this deficiency, *S_Dbw*—the sum of the average scattering for the clusters and the inter-cluster density (ID)—is introduced as a clustering evaluation index.

$$S\_Dbw(C_k) = Scat(C_k) + Dens\_bw(C_k) \tag{3}$$

here, $Scat(C_k)$ represents the average scattering for the clusters, and $Dens\_bw(C_k)$ represents the ID. A smaller *S_Dbw* indicates better clustering qualification [32], [33].

2) OPTIMIZATION OF INITIAL CLUSTER CENTER

The K-means methods initializes cluster centers v_k ($k = 1, 2, \ldots,$ K). Different initializations can yield different final clusters because the K-means only converges to the local minimum. The CSA (Algorithm 2) is widely used in optimization owing to its excellent global search capability, which avoids the effect of randomly selecting initial cluster centers on the clustering quality.

Algorithm 2 CSA

function CSA(M, m, MNI, k, fl, P)
  Evaluate the position of the crows
  for m from 1 to MNI do
    Randomly choose one of the crows to follow(j)
    for i from 1 to M
      if $\lambda_j > P^{i,m}$
        $l^{i,m+1} = l^{i,m} + \lambda_i \times fl^{i,m} \times (me^{j,m} - l^{i,m})$
      else
        $l^{i,m+1} = a$ random position
  Check the feasibility of new positions
  Update the memory of crows

3) CSA [34]

It is assumed that there is a k-dimensional environment containing M crows (k is the dimension of the cluster centers). The location of each crow $i(i = 1, 2, \ldots M)$ at time $m(m = 1, 2, \ldots, MNI)$ in the search space is specified by the vector $l^{i,m} = [l_1^{i,m}, l_2^{i,m}, \ldots, l_k^{i,m}]$, and MNI represents the maximum number of iterations. Each crow saves its location in a memory vector $me^{i,m}$, which is similar to $l^{i,m}$. The location matrix **LOC** and memory matrix **MEM** of each crow are defined as follows:

$$\mathbf{LOC} = \begin{bmatrix} l_1^1 & l_2^1 & \cdots & l_k^1 \\ l_1^2 & l_2^2 & \cdots & l_k^2 \\ \cdots & \cdots & \cdots & \cdots \\ l_1^M & l_2^M & \cdots & l_k^M \end{bmatrix}$$

$$\mathbf{MEM} = \begin{bmatrix} me_1^1 & me_2^1 & \cdots & me_k^1 \\ me_1^2 & me_2^2 & \cdots & me_k^2 \\ \cdots & \cdots & \cdots & \cdots \\ me_1^M & me_2^M & \cdots & me_k^M \end{bmatrix} \tag{4}$$

It is assumed that crow i follows crow j and finds the location of crow j, when crow j returns to its location in the mth iteration. At this time, the probability of crow j discovering and replacing its location is P. The new location of crow i is defined as:

$$l^{i,m+1} = \begin{cases} l^{i,m} + \lambda_i \times fl^{i,m} \times (me^{j,t} - l^{i,t}), & \lambda_j \geq P^{j,m} \\ random, & else \end{cases} \tag{5}$$

where λ_i and λ_j are uniformly distributed random numbers between 0 and 1, and fl is the flight distance. If the fitness-function value of the new location is better than that of the original location, the location is updated. Otherwise, the location is not updated.

$$me^{i,m+1} = \begin{cases} l^{i,m+1}, & \text{Fitness}(l^{i,m+1}) \; better \; than \; f(me^{i,m}) \\ me^{i,m}, & else \end{cases} \tag{6}$$

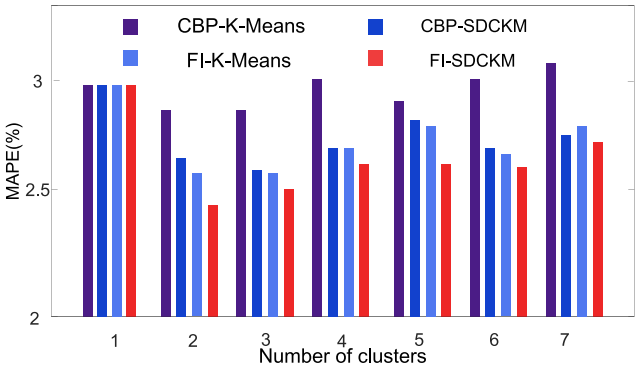


FIGURE 4. MAPE of the predictor for four clustering methods: FI-K-Means, FI-SDCKM, CBP-K-Means, and CBP-SDCKM.

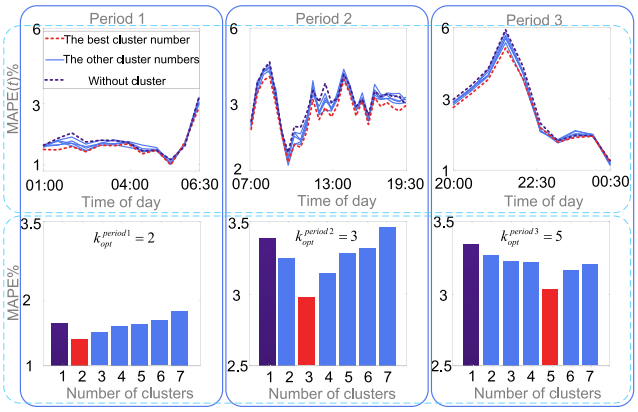


FIGURE 5. MAPE and MAPE(t) of the FI-SDCKM-predictor for three periods at different k values.

Here, "Fitness" represents the fitness function. When *MCN* is 10, the CSA tends to converge, according to statistical experimental analysis.

4) PROCEDURE OF SDCKM

SDCKM combines the excellent global search capability of the CSA and the local search ability of the K-means method. *S_Dbw* is set as a fitness function for evaluating the clustering quality at each time, and the optimal initial cluster center is obtained. The SDCKM process is as follows:

a) Initialization: M, **LOC**, **MEM**, k, MNI, fl, P.
b) Use the K-means to perform clustering based on the memory vector $me^{i,m}$ for each crow, which represents the initial cluster center for obtaining the clustering results.
c) Calculate the fitness according to the clustering result of step a). The fitness function is defined as

$$\text{Fitness} = \sum_{k=1}^{K} S\_Dbw(C_k) \tag{7}$$

d) Update the location using formula (5).
e) Calculate the fitness according to the updated location of each crow. Use formula (6) to determine whether to update the memory.
f) Repeat steps b), c), d) and e), until the MNI is reached. The memory vector with the smallest fitness value is selected as the optimal initial cluster center.
g) The initial cluster center obtained in step f) is treated as the initial K-means clustering center, after the foregoing steps are completed. A final clustering scheme is then generated.

C. FI-SDCKM PRACTICABILITY

The clustering quality for different k values without time segmentation is compared according to the prediction error in Fig. 4.

Fig. 4 shows the MAPE of day-ahead load forecasting for a household-based distribution network with k of 1-7, for the FI-K-Means, FI-SDCKM, CBP-K-Means, and CBP-SDCKM methods. Regarding the K-means methods, for each k, the MAPE of the FI-Predictor is smaller than that of the CBP-Predictor. Additionally, the SDCKM based on the FI is more accurate than the SDCKM based on the CBP. The MAPE of the SDCKM-Predictor is smaller than that of the K-means-Predictor. CBP-K-Means is used as the baseline for the comparison, and FI-SDCKM is the clustering method proposed in this paper. When the number of clusters is 2, the proposed method exhibits the greatest improvement compared with the baseline. The MAPE decrease from 2.814% to 2.431%, and the prediction error decrease by 13.61%. Thus, FI-SDCKM allows a significant reduction of the MAPE in load forecasting and is effective for improving the prediction accuracy. Detailed experimental results are presented in TABLE 4.

Regarding time segmentation, the clustering results of FI-SDCKM in Fig. 5 differ according to the fluctuation periods mentioned in Section II. A detailed explanation for this is presented in Section IV. Clearly, the proposed FI-SDCKM clustering method is effective in both the time-segmented and non-time-segmented cases.

IV. MIXTURE LOAD FORECASTING MODELING

To improve the accuracy of load forecasting in distribution networks, a new forecasting model based on the load fluctuation and FI profile clustering is proposed. Fig. 6 shows flowcharts of the proposed forecasting model. First, RReliefF is used to analyze the importance of each feature. Second, the 24h of the day are divided into periods according to the load fluctuation. Third, the improved SDCKM clustering method based on FI is used to cluster users for the different periods. The optimal number of clusters (k_{opt}) in each period is determined via statistical experiments. Then, the k_{opt} for each period is determined according to the rolling prediction accuracy of the RF. Finally, the final forecast results are obtained by aggregating the results for every cluster.

A. RF-BASED LOAD FORECASTING

Because the RF is composed of multiple classification and regression trees (CARTs), it avoids unstable prediction results

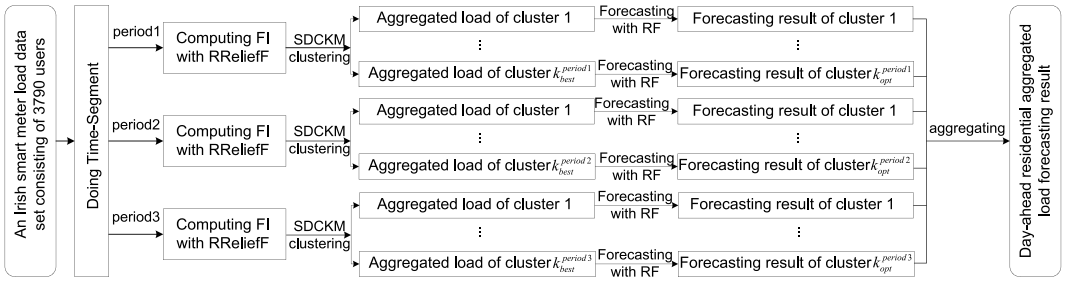


FIGURE 6. Flowcharts of the proposed method.

and overfitting. Additionally, the RF is a data-driven method based on ensemble learning theory; thus, it is effective for analyzing high-dimensional SM data [29].

1) RF

The RF is described as:

$$\{h(x, \Theta_d), d = 1, 2, \ldots, D\} \tag{8}$$

where $h(x, \Theta_d)$ represents the d^{th} DT that constitutes the RF, and x is the input vector of the DT. Each Θ is independently distributed, representing the random process of DT growth and extraction of the sample data of the d^{th} tree in the RF. Thus, the final prediction result can be obtained according to the output of all the CARTs in the model. On the other hand, the bootstrap random sampling of the RF comprises two aspects: training sample extraction and the selection of candidate feature sets from nonleaf nodes in the CART.

1) First, in the process of constructing the RF, a bootstrap is used to randomly generate individual training sets for each CART. The training sets for each CART probably include 2/3 of the entire sample space and the remaining out-of-bag samples for this CART.
2) Second, at each node, rather than choosing the best split among all the features, random samples m_f ($m_f \ll \mathrm{M_f}$, where $\mathrm{M_f}$ is the number of features) of the feature sets are chosen and then the best split among these variables is chosen.

Therefore, the RF is not affected by noise, outliers, or dimensional hazards. The RF is suitable for constructing predictors with complex multiclass high-dimensional input features.

RF requires only two parameters to be set: D and m_f. In this study, these parameters are set as [27]: $D = 500$ and $m_f = \mathrm{M_f}/3$.

2) PREDICTION FEATURES

The load to be predicted is correlated with the historical load, temperature, month, and other features according to previous studies [21], [35]. The cycle variables improve the accuracy of the distribution-network load forecasting [36]. To facilitate the capture of cycles and reflect the cyclical changes in the total load of a household based on the distribution network,

TABLE 1. 345 features for the day-ahead load forecasting.

Category of variable	Label	Variables
Load	F_1~F_{336}	L_{t-1}~L_{t-336} (a week)
Day-type	F_{337}	working days=0, nonworking days =1
Calendar	F_{338}	Monthly variables
Cycle	F_{339}~F_{341} F_{342}~F_{344}	$\cos(2\pi t/48)$, $\cos(2\pi t/336)$, $\cos(2\pi t/17520)$ $\sin(2\pi t/48)$, $\sin(2\pi t/336)$, $\sin(2\pi t/17520)$
Temperature	F_{345}	T_t

two cycle variables $c_1(t)$ and $c_2(t)$ are designed:

$$c_1(t) = \cos(2\pi t/T) \tag{9}$$

$$c_2(t) = \sin(2\pi t/T) \tag{10}$$

where $t (t = 1, 2, \ldots, T)$ is the time index, and T is the cycle period. $T = 48$, $T = 336$, and $T = 17520$ correspond to the day cycle, week cycle, and year cycle, respectively. Working days and non-working days are relevant and are represented by 0 and 1, respectively. [36].

TABLE 1 presents the feature set of day-ahead load forecasting for households based on the distribution network. The dimension of the historical load is large in order to reduce the accumulative error of the rolling forecasting model. This is analyzed in detail in Section IV B.

The root-mean-square error (RMSE) is introduced for evaluating the forecast results. The RMSE is defined as

$$\mathrm{RMSE} = \left(\frac{1}{N_t} \sum_{n_t=1}^{N_t} (L_r - L_p)^2 \right)^{\frac{1}{2}} \tag{11}$$

The RMSE and MAPE are used to evaluate the prediction accuracy of the predictor during the day or three periods of the day. RMSE(t) and MAPE(t) are used to evaluate the prediction accuracy of the predictor at every time point t.

B. PREDICTION ACCURACY OF RF IN ROLLING PREDICTION MODEL

For verifying the accuracy of the RF predictor and analyzing the influence of the accumulative error on different predictors, two types of rolling forecast models are designed using the RF and the ANN [16].

TABLE 2. The composition of the four feature sets.

FS	FN	Composition of features
1	22	L_{t-1}~L_{t-6},L_{t-48}~L_{t-54},L_{t-366}~L_{t-342},T_t,M_t,H_t,D_t
2	63	L_{t-1}~L_{t-48},L_{t-96},L_{t-144},L_{t-192},L_{t-240},L_{t-288},L_{t-336},T_t,c_1,c_2,M_t,D_t
3	157	L_{t-1}~L_{t-144},L_{t-192},L_{t-240},L_{t-288},L_{t-336},T_t,c_1,c_2,M_t,D_t
4	345	L_{t-1}~L_{t-336},T_t,c_1,c_2,M_t,D_t

L = load; T = temperature; H = holiday variables; M = monthly variables; D = day-type variables; FS = feature set; FN = feature number.

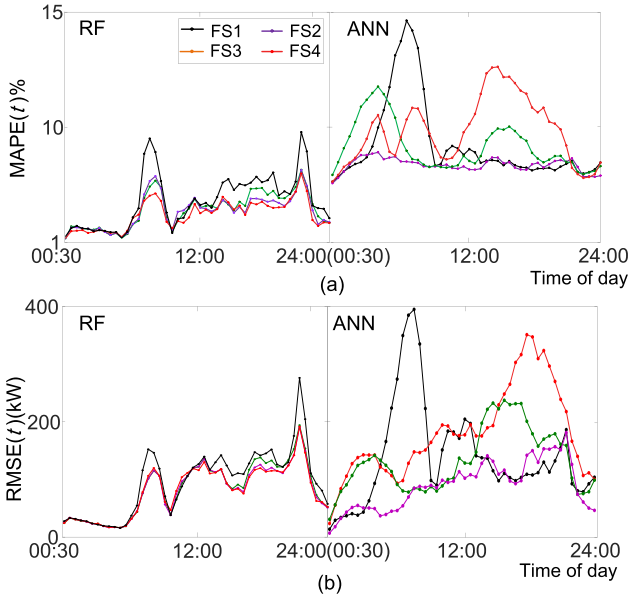


FIGURE 7. MAPE(t) and RMSE(t) for the RF and ANN with four different feature sets.

The prediction error for the rolling forecast is mainly affected by two factors: (1) the predictor performance and (2) the use of the predicted values as the eigenvalues for the subsequent forecast. Next, we conduct experiments on different sources of error. To analyze the effect of accumulative errors, we design four different feature sets, as shown in TABLE 2.

FS1 is constructed according to [21]. FS4 is the feature set constructed in this study. FS2 and FS3 are feature sets formed by reducing part of the historical load features of FS4. The Levenberg–Marquardt approach is employed to select the parameters of the ANN. When the feature dimension is 22, 63, 157, and 345, the number of hidden neurons is 20, 110, 360, and 890, respectively [37].

Fig. 7 presents the MAPE(t) and RMSE(t) of the RF and ANN without clustering and with the splitting of the period for the four feature sets. The results are as follows:

1) The accuracy of the RF does not decrease as the number of features increases and is unaffected by the feature dimension, in contrast to that of the ANN.

2) The RF is less affected by accumulated errors than the ANN, especially with a high feature dimension.

A higher proportion of predicted features in the input features yields a greater impact of accumulated errors [21]. An RF model based on ensemble learning is suitable for the proposed method, according to the conclusions drawn on the basis of Fig. 4. Therefore, we can increase the historical load feature dimension of the input features to reduce the accumulative error.

TABLE 3. Statistial value of total load of the test set with and without time segmentation.

	Mean value (kW)	Peak value (kW)
P1	822.6	1524.1
P2	1812.3	2716.5
P3	2109.6	3027.6
TL	1626.8	3027.6

P1 = Period 1; P2 = Period 2; P3 = Period 3; TL = Total load

V. FORECASTING RESULTS

A. PREDICTION MODEL COMPARISON

We study the application of time segmentation in clustering according to FI based on SM data at the household level for enhancing the load-forecasting performance at the distribution-network level. We verify the effectiveness of the clustering method for different periods after selecting the RF predictor. As explained in the preceding sections, we construct 84 independent RF-based forecasting models, with k varying from 1 to 7. All the models are implemented in MATLAB running on an Intel Core i7 at 3.7 GHz with 16 GB RAM. (Analysis of the load fluctuation in the time domain and the selection method for k_{opt} are presented in Sections II and III, respectively.) According to a comparison of the error evaluation indicators, the statistical values of the total load in the test set with time segmentation are obtained (TABLE 3).

To highlight the advantages of the proposed method, three different methods are employed to conduct residential distribution network load-forecasting experiments. The proposed method is denoted as RRF-SDCKM-RF. CBP-K-Means-RF is a baseline method, and RF without clustering is the traditional method. And, in order to show the promotion of prediction results, two indicators Promoting percentages of mean absolute percentage error (P_{MAPE}) and Promoting percentages of root mean square error (P_{RMSE}) are introduced. The forecast results of the three prediction models for three periods and the aggregated load forecast results with time-segment clustering and non-time-segment clustering are presented (TABLE 4). And the promotion of forecasting for each period is shown in TABLE 5. As shown in TABLE 4 and 5, the proposed load-fluctuation analysis and clustering method improve the accuracy of load forecasting. We compared the predicted increases in different periods. Among them, the promotion effect of P1 is the most significant. Although the fluctuation of P1 is small, there are many outliers, which shows that the prediction error of outliers can be effectively reduced when using time segment analysis of the new approach. The fluctuation of P2 is medium, but the deviation from the outliers is large. There are extreme electric consumption scenarios on individual dates in P2. But time-segment is also helpful for prediction results. P3 has the highest fluctuation with high load in the night, but there is

TABLE 4. Error of different models at different time segments.

	Index	RF(without cluster)		CBP-K-Means-RF		RRF-SDCKM-RF	
		NTS	TS	NTS	TS	NTS	TS
P1	MAPE(%)	2.337	2.041	2.185	2.025	1.874	1.633
	RMSE(kW)	31.046	26.129	25.486	23.247	23.468	20.076
P2	MAPE(%)	3.137	3.145	3.276	3.012	2.935	2.801
	RMSE(kW)	92.493	89.553	90.843	86.348	84.168	80.535
P3	MAPE(%)	3.738	3.494	3.721	3.408	3.347	3.043
	RMSE(kW)	108.761	97.365	100.933	90.356	91.831	86.123
Total	MAPE(%)	2.907	2.864	2.814	2.725	2.431	2.303
	RMSE(kW)	79.362	78.348	77.368	74.452	71.836	69.021

RRF = RReliefF; TS = time-segment clustering; NTS = non-time-segment clustering

TABLE 5. Error improvement between RF-NTS and RRF-SDCKM-RF-TS.

	P1	P2	P3	Total
P_{MAPE}	30.12%	10.71%	18.59%	20.78%
P_{RMSE}	35.33%	12.93%	20.82%	13.03%

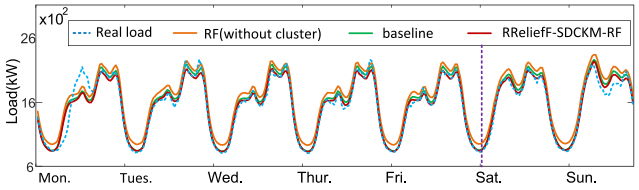


FIGURE 8. Aggregated forecasting load curves of different models for working days and nonworking days.

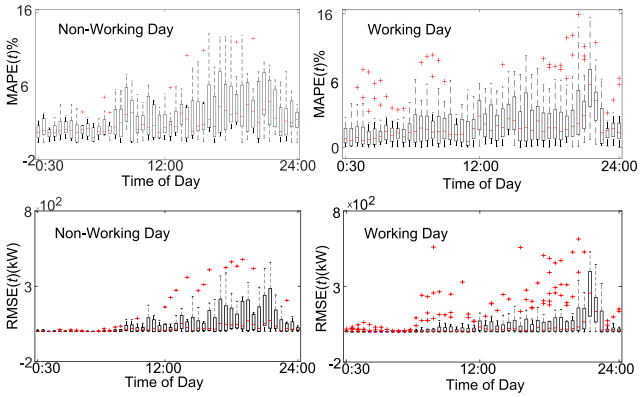


FIGURE 9. Forecast error distribution for working days and non-working days in August 2010.

no outlier, which results in P_{MAPE} and P_{RMSE} being smaller than P1 when time segment is used. Generally, the P_{MAPE} of the proposed method was 20.78% lower than that of the traditional method (MAPE decreased from 2.907% to 2.303%); and the P_{RMSE} was 13.03%. Taking full account of load fluctuation, the forecasting effect of each time period is improved, which is of significance to the dispatching of distribution network.

The results indicate that the proposed clustering method and clustering in different time periods according to the FI positively impact the accuracy of aggregate load forecasting for residents.

B. FORECAST RESULTS FOR WORKING AND NON-WORKING DAYS

Differences in the electricity consumption between working days and non-working days may affect the load forecast. To verify the applicability of the proposed method, the load forecasting results of one week are selected randomly.

Fig. 8 shows the aggregated load-prediction curves of different models for Monday to Sunday. For both date types, the proposed model provides the best fit to the true value. The MAPEs for the working and non-working days are 2.686% and 2.284%, respectively, and the RMSEs are 73.916 kW and 60.3175 kW, respectively.

Fig. 9 shows the forecast error distribution for working and non-working days in August 2010. The prediction error for working days at each interval of the day is larger, but there are fewer outliers. The prediction error for non-working days at each interval of the day is smaller, but there are more outliers. The results indicate that the diversity of the residential electricity consumption on working days is greater than that on non-working days and that load forecasting for working days is more challenging.

C. EFFECTS OF DIFFERENT FI AND PREDICTORS ON FORECAST ERRORS

To verify the improvements of the new method, the Pearson coefficient [38], mutual information (MI) index [39], and Gini index [29] are applied to the least-squares support vector machine (LS-SVM) [40], ANN [21], and extreme learning machine (ELM) [41] for experiments using the aforementioned modeling process. The clustering method is SDCKM. The parameters ($\gamma = 50000, \sigma^2 = 5$) of the LS-SVM (Gaussian radial basis function kernel) are optimized according to [40]. According to [41], the ELM has a sigmoid activation function, and the number of hidden-layer nodes is 1000. The ANN parameters are consistent with 345 dimensions, as previously described.

TABLE 6 presents the k_{opt} and errors of different mixture load-forecasting models. The results are as follows:

a) The MAPE and RMSE are significantly large in Period 2 and Period 3 for the time-segment clustering model. In Periods 2 and 3, the load fluctuation is greater than that in Period 1, as shown in Fig. 2. This confirms that

TABLE 6. Optimal number of clusters and errors for different mixture load forecasting models.

FI	Type	RF			LS-SVM			ANN			ELM		
		MAPE(%)	RMSE(kW)	k_{opt}	MAPE(%)	RMSE(kW)	k_{opt}	MAPE(%)	RMSE(kW)	k_{opt}	MAPE(%)	RMSE(kW)	k_{opt}
RRF	P1	**1.633**	**20.076**	2	2.013	33.439	3	3.836	46.245	2	2.228	36.172	3
	P2	**2.801**	**80.535**	3	2.931	83.352	4	4.825	110.356	2	2.987	85.288	4
	P3	**3.043**	**86.123**	5	3.345	91.132	5	5.635	119.352	4	3.362	91.418	5
	NTS	**2.431**	**71.836**	2	3.029	80.242	3	5.013	110.524	2	3.299	85.257	4
	TS	**2.303**	**69.021**	--	2.937	79.934	--	4.923	107.522	--	3.111	80.163	--
MI	P1	1.693	23.680	3	2.194	35.367	3	4.724	49.572	3	3.231	39.735	3
	P2	2.966	84.227	3	3.132	86.527	4	4.982	115.854	4	3.587	95.362	3
	P3	3.321	90.652	4	3.653	97.367	5	5.745	121.456	4	3.959	100.463	6
	NTS	2.540	73.824	3	3.499	90.989	3	5.391	117.596	4	3.781	98.185	4
	TS	2.492	71.902	--	3.320	89.322	--	5.214	112.355	--	3.652	95.194	--
Gini	P1	1.702	25.141	1	2.063	33.625	2	4.021	48.169	3	2.234	36.537	3
	P2	3.248	89.562	3	3.403	94.256	3	4.682	109.356	3	2.988	86.813	3
	P3	3.368	91.512	3	3.551	95.578	5	5.724	121.355	5	3.399	93.461	6
	NTS	2.592	75.256	2	3.791	98.082	4	5.001	109.348	4	3.213	84.926	3
	TS	2.483	70.913	--	3.954	101.372	--	4.935	108.345		3.127	81.245	--
Pearson	P1	1.679	22.958	2	2.014	34.386	3	4.924	50.367	3	2.195	34.925	4
	P2	3.226	88.053	2	3.441	94.869	3	5.724	130.245	5	3.263	89.245	5
	P3	3.366	91.441	4	3.681	97.356	5	5.924	129.341	5	3.936	104.353	5
	NTS	2.599	75.961	2	3.593	91.253	3	5.781	121.583	3	3.581	92.104	4
	TS	2.456	70.896	--	3.555	90.145	--	5.632	119.035	--	3.459	89.753	--

the load fluctuation is related to the prediction accuracy. Time-segment clustering models improve the accuracy of load forecasting.

b) The RF has the smallest MAPE and RMSE for each Period and each type of FI, indicating the advantage of the RF for prediction using high-dimensional data.

c) When RReliefF is used to calculate the FI, the prediction error for each Period is minimized. This confirms the effectiveness of analyzing the FI with RReliefF.

RReliefF-SDCKM-RF is identified as the most accurate mixture load forecasting model. Using the proposed RReliefF-SDCKM-RF model with time-segment clustering, the smallest prediction error is achieved. The predicted execution time of the proposed method is 3.332394 s. Compared with traditional prediction methods and CBP-K-Means-RF, the MAPE is reduced by 0.604% and 0.511%, respectively, and RMSE is reduced by 10.341 kW and 8.347 kW, respectively.

VI. CONCLUSION

Rational utilization of smart meters could help realize the smart grid. This paper shows how to use smart meter load data to improve the load forecasting accuracy of the entire distribution network according to the difference of load fluctuation. The proposed day-ahead residential distribution network load-forecasting model has the following advantages.

The period fluctuation of different residential load was found, and the overall methodology was improved on this basis. The experimental results show that electricity fluctuation of residential users directly affects load forecasting results. To increase the prediction effect, the 24h of the day are divided into different periods according to the fluctuation of the electricity consumption of the users (the lowest/highest point of standard deviation).

The aspects of the clustering process have been optimized. Breaking the convention of clustering using load shape, this paper uses FI to cluster massive users. It overcomes the inability of traditional clustering methods to analyze the effects of non-load features on the clustering. What's more, the SDCKM can optimize the cluster center selection, and consider the intra-cluster distance and inter-cluster dispersion in clustering.

The input dimension is augmented and the stable forecasting models (RF) are used for rolling prediction. To reduce the error of rolling prediction model, the proportion of predicted values used as subsequent predictive features in the total input is reduced by expanding the historical load features. Additionally, the proposed prediction model solves the problem of high input feature dimension in aggregated residential forecasting, and feature selection is not needed before aggregated forecasting.

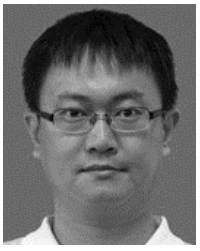
There are some limitations in our study which could be the future research. Firstly, the aggregate forecasting result we obtained is only for the electricity demand of residents in the distribution network, and the impact of renewable energy access to the distribution network on the electricity demand of residents has not been considered. Secondly, as we carry out aggregated load forecasting analysis for residential users in the distribution network, we do not consider the impact of electric vehicles on aggregate residential load forecasting results. Future research could focus on the impact of renewable energy and electric vehicles on aggregated residential load forecasting.

REFERENCES

[1] M. Imani and H. Ghassemian, "Electrical load forecasting using customers clustering and smart meters in Internet of Things," in *Proc. 9th Int. Symp. Telecommun. (IST)*, Dec. 2018, pp. 113–117.

[2] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecasting*, vol. 30, no. 2, pp. 357–363, Apr. 2014.

[3] C. Chen, J. Wang, and S. Kishore, "A distributed direct load control approach for large-scale residential demand response," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2219–2228, Sep. 2014.
[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.
[5] H. Chen, S. Wang, S. Wang, and Y. Li, "Day-ahead aggregated load forecasting based on two-terminal sparse coding and deep neural network fusion," *Electr. Power Syst. Res.*, vol. 177, Dec. 2019, Art. no. 105987.
[6] K. Park, S. Yoon, and E. Hwang, "Hybrid load forecasting for mixed-use complex based on the characteristic load decomposition by pilot signals," *IEEE Access*, vol. 7, pp. 12297–12306, 2019.
[7] N. Huang, D. Wang, L. Lin, G. Cai, G. Huang, J. Du, and J. Zheng, "Power quality disturbances classification using rotation forest and multi-resolution fast S-transform with data compression in time domain," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 22, pp. 5091–5101, Nov. 2019.
[8] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.
[9] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1591–1598, Jul. 2017.
[10] L.-G. Chen, H.-D. Chiang, R.-P. Liu, and N. Dong, "Group-based chaos genetic algorithm and non-linear ensemble of neural networks for short-term load forecasting," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 6, pp. 1440–1447, Apr. 2016.
[11] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 549–556, Mar. 2013.
[12] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3906–3908, Jul. 2018.
[13] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
[14] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Appl. Energy*, vol. 235, pp. 10–20, Feb. 2019.
[15] M. H. Alobaidi, F. Chebana, and M. A. Meguid, "Robust ensemble learning framework for day-ahead forecasting of household based energy consumption," *Appl. Energy*, vol. 212, pp. 997–1012 Jan. 2018.
[16] P. Laurinec, M. Lóderer, M. Lucká, and V. Rozinajová, "Density-based unsupervised ensemble learning methods for time series forecasting of aggregated or clustered electricity consumption," *J. Intell. Inf. Syst.*, vol. 53, no. 2, pp. 219–239, Oct. 2019.
[17] G. Sideratos, A. Ikonomopoulos, and N. D. Hatziargyriou, "A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks," *Electr. Power Syst. Res.*, vol. 178, Jan. 2020, Art. no. 106025.
[18] I. A. Sajjad, G. Chicco, and R. Napoli, "Definitions of demand flexibility for aggregate residential loads," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2633–2643, Nov. 2016.
[19] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.
[20] H. Chen, S. Wang, and Y. Tian, "A new approach for power-saving analysis in consumer side based on big data mining," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Portland, OR, USA, Aug. 2018, pp. 1–5.
[21] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015.
[22] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5196–5206, Sep. 2018, doi: 10.1109/tsg.2017.2683461.
[23] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.
[24] R. Al-Otaibi, N. Jin, T. Wilcox, and P. Flach, "Feature construction and calibration for clustering daily load curves from smart-meter data," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 645–654, Apr. 2016.
[25] S. Wang, H. Chen, L. Wu, and J. Wang, "A novel smart meter data compression method via stacked convolutional sparse auto-encoder," *Int. J. Elect. Power Energy Syst.*, vol. 118, Jun. 2020, Art. no. 105761.
[26] N. Huang, Y. Wu, G. Lu, W. Wang, and X. Cao, "Combined probability prediction of wind power considering the conflict of evaluation indicators," *IEEE Access*, vol. 7, pp. 174709–174724, 2019.
[27] G. Dudek, "Short-term load forecasting using random forests," in *Proc. IEEE Int. Conf. Intell. Syst. (IS)*, vol. 323, Jan. 2015, pp. 821–828.
[28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[29] N. Huang, Y. Wu, G. Cai, H. Zhu, C. Yu, L. Jiang, Y. Zhang, J. Zhang, and E. Xing, "Short-term wind speed forecast with low loss of information based on feature generation of OSVD," *IEEE Access*, vol. 7, pp. 81027–81046, 2019.
[30] (Sep. 20, 2013). *Data from the Commission for Energy Regulation*. [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/
[31] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
[32] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: Finding the optimal partitioning of a data set," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2002, pp. 187–194.
[33] Y. Liu, Z. Li, H. Xiong, X. Gao, J. Wu, and S. Wu, "Understanding and enhancement of internal clustering validation measures," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 982–994, Jun. 2013.
[34] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Comput. Struct.*, vol. 169, pp. 1–12, Jun. 2016.
[35] T. Hong and S. Fan, "Guest editorial: Special section on analytics for energy forecasting with applications to smart grid," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 399–401, Jan. 2014.
[36] N. Ding, C. Benoit, G. Foggia, Y. Besanger, and F. Wurtz, "Neural network-based model design for short-term load forecast in distribution systems," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 72–81, Jan. 2016.
[37] S. Fan, L. Chen, and W.-J. Lee, "Short-term load forecasting using comprehensive combination based on multimeteorological information," *IEEE Trans. Ind. Appl.*, vol. 45, no. 4, pp. 1460–1466, Jul. 2009.
[38] T. Ahmed, D. Vu, K. Muttaqi, and A. Agalgaonkar, "Load forecasting under changing climatic conditions for the city of Sydney, Australia," *Energy*, vol. 142, pp. 911–919, Jan. 2018.
[39] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh, and Z. A. Khan, "An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2587–2596, Oct. 2017.
[40] X. Wang, W.-J. Lee, and H. Huang, "Factors that impact the accuracy of clustering-based load forecasting," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3625–3630, Sep. 2016.
[41] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.



NANTIAN HUANG was born in 1980. He received the B.S. degree from the Harbin Institute of Technology, Harbin, China, in 2003, the M.S. degree from the Dalian University of Technology, Dalian, China, in 2006, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, in 2012. He is currently an Associate Professor with the School of Electrical Engineering, Northeast Electric Power University. His research interests focused on artificial intelligence and machine learning application for data analysis and optimization in smart grid.

WENTING WANG was born in 1994. She received the B.S. degree in electronic information engineering from Northeast Electric Power University, China, in 2017. She is currently pursuing the M.S. degree with the Department of Electrical Engineering, Northeast Electric Power University. Her research interests focuses on applications of machine learning and data science in electrical engineering.

SINING WANG was born in September 1978. She received the master's degree in the major of power system and automation from North China Electric Power University (NCEPU), in 2007. She is currently a Senior Engineer. She also works as a General Manager of the Grid Operation and Monitoring Division, Beijing Zhongdian Puhua Information Technology Company, Ltd. She has been engaged in the research and project management of power system information communication, cloud computing, big data, and other related technologies for a long time.

JUN WANG was born in April 1988. He received the B.E. degree in electrical engineering and automation from Beijing Jiaotong University, in 2015. Now, he lives in Beijing with the postcode of 102200. He is currently a researcher, in grid monitoring department of Beijing Zhongdian Puhua Information Technology Company, Ltd, working on object detection, big data analysis.

GUOWEI CAI was born in 1968. He received the B.S. and M.S. degrees from Northeast Electric Power University, China, in 1990 and 1993, respectively, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 1999. Since 2004, he has been a Professor with the School of Electrical Engineering, Northeast Electric Power University, Jilin, China.

LIANG ZHANG was born in 1985. He received the B.Sc. degree from the Hebei University of Science and Technology, China, in 2007, and the M.S. and Ph.D. degrees in electrical engineering and automation from the Harbin Institute of Technology, Harbin, China, in 2010 and 2015, respectively. He was a Visiting Ph.D. Student with The University of British Columbia, Vancouver, BC, Canada, in 2014. He is currently an Associate Professor with the Department of Electrical Engineering, Northeast Electric Power University, China. His current research interests include ac/dc hybrid microgrids, V2G communication, vehicular communication, the energy Internet, and power line communication.