IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 26, 2021, accepted February 26, 2021, date of publication March 2, 2021, date of current version April 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063123

A Hybrid Short-Term Load Forecasting Model Based on Improved Fuzzy C-Means Clustering, Random Forest and Deep Neural Networks

FU LIU[1], TIAN DONG[1,2], TAO HOU[1], AND YUN LIU[1], (Member, IEEE)
[1]College of Communication Engineering, Jilin University, Changchun 130000, China
[2]State Grid Jilin Electric Power Company, Ltd., Changchun 130021, China

Corresponding author: Yun Liu (liuyun313@jlu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61503151, in part by the Natural Science Foundation of Jilin Province under Grant 20160520100JH, and in part by the Project funded by the China Postdoctoral Science Foundation under Grant 2019M651204.

ABSTRACT Short-term load forecasting (STLF) plays an important role in the secure and reliable operation of the electric power system. Grouping similar load profiles by a clustering algorithm is a common method to reduce the uncertainty of electric consumption data. However, due to the uneven distribution of different date types in a historical data set, the tradition fuzzy c-means clustering (FCM) algorithm cannot identify typical load consumption patterns accurately. To solve this problem, a novel STLF model based on the improved FCM (IFCM) algorithm, random forest (RF) and deep neural networks (DNN) is proposed in this paper. First, IFCM is used to partition the load consumption profiles into several groups, and each group represents a typical load consumption pattern. The optimal number of clusters is determined by a recent clustering validity index. Then, a RF model is trained by the meteorological and calendar features of the historical data set. Finally, a DNN model is established for each group, and is trained using the features of the days that are partition into this group by IFCM. The experimental results on two daily load consumption data sets have showed that the proposed STLF model achieves better prediction performance as compared to other methods. In addition, the load consumption pattern of holidays was extracted from the historical data sets by utilizing IFCM, and the prediction performance of holidays in the testing set therefore has been significantly improved.

INDEX TERMS Clustering, fuzzy c-mean algorithm, load forecasting, random forest, deep neural network.

LIST OF ABBREVIATIONS

FCM,	fuzzy c-means clustering
STLF,	short-term load forecasting
RF,	random forest
DNN,	deep neural network
CNN,	convolutional neural network
RNN,	recurrent neural network
DBN,	deep brief network
LSTM,	long short-term memory
GRNN,	generalized regression neural network
ELM,	extreme learning machine
MAPE,	mean absolute percentage error
MASE,	mean absolute scaled error
RMSE,	root mean square error
SVM,	support vector machine

The associate editor coordinating the review of this manuscript and approving it for publication was Tossapon Boongoen.

I. INTRODUCTION

short term load forecasting (STLF) is a key part of the grid management system, and its accuracy has a direct effect on the results of grid safety inspections. STLF plays a key role in the dynamic estimation of grid states, load dispatch, and the reduction of power generation costs [1]. However, it is a difficult task to perform load forecasting precisely due to the high volatility and uncertainty involved in the power system [2].

Till to now, a number of methods have been proposed to perform the load forecasting, and can be usually divided into two categories: traditional linear prediction methods and artificial intelligence (AI) based methods. Traditional approaches mainly focus on linear predictions by using classical time-series models and their variations, such as the linear regression models [3], [4], the autoregressive moving average method [5] and the Kalman filtering based method [6]. As the load consumption patterns become more dynamic and unpredictable, AI-based methods are applied to improve the prediction precision [7]. Several models have been used to preform the load forecasting, such as the artificial neural network [47], wavelet transform [9], random forest method [1] and so on. Compared with traditional linear prediction approaches, AI-based methods have been showed better prediction performance owing to the advanced models they use.

Recently, deep learning based forecasting models have been observed in literatures [10]. Several types of deep neural networks have been utilized for load forecasting. Qiu *et al.* used the deep brief network (DBN) for load demand time series forecasting [11]. The factored conditional restricted Boltzmann machine was used to forecast the future electric load in [12]. A STLF model was proposed based on the fuzzy time series and convolutional neural networks (CNN) in [13]. The recurrent neural network (RNN) is a powerful model for the feature learning from sequence data, and has been widely used for the STLF. For example, Kong *et al.* proposed a long short-term memory (LSTM) based forecasting model with load consumption sequences [14]. Shi *et al.* proposed a RNN based forecasting model for the household load forecasting [15]. In addition, some studies try to combine two different neural networks for the load forecasting. For example, Kim *et al.* proposed a CNN-LSTM network that can learn spatial and temporary features to predict load consumption [16]; Chen *et al.* constructed a ensemble deep learning model by combining RNN and deep residual network (DRN) for the STLF [17]. Compared with the traditional machine learning methods, the load forecasting performance has been significantly improved by using deep learning models in these studies.

It is a common method to group similar load profiles using a clustering algorithm with the purpose of reducing the variance of the uncertainty within each cluster [18]. For example, Farfar and Khadir identified a set of day classes using the c-means clustering algorithm, and built regression models to forecast each day type independently [19]. Kan *et al.* proposed a forecasting model by combining the c-means clustering algorithm, artificial neural network and k-nearest neighbor method [20]. Quilumba *et al.* applied the c-means clustering algorithm to group customers with similar load consumption patterns prior to perform load forecasting [21]. Oprea and Bara used the c-means clustering algorithm to determine load consumption patterns and then constructed forecasting models [22]. Panapakidis proposed two hybrid forecasting models by combining clustering algorithms and artificial neural networks [23]. FCM clustering method was used to extract the load consumption patterns in [24] and [25]. In these studies, c-means (CM) and FCM clustering algorithms are the most frequently used methods to extract load consumption patterns because of their advantages of simplicity and effectiveness. The prediction precision has been improved owing to the application of clustering algorithms. However, the performance is heavily dependent on the robustness of clustering algorithms.

Previous studies have demonstrated that the load consumption is related to the date category, such as the weekdays, weekends and holidays [26]. The historical load data set is imbalanced because the percents of days in different categories are significantly uneven. Specifically, holidays only occupy a small fraction in a year. However, the clustering results of the CM and FCM algorithms, which are commonly used in existing load forecasting methods, are significantly affected by the imbalanced distributions inter-clusters; this is called ''uniform effect'' in literatures [27], [28]. The reason of the ''uniform effect'' of CM and FCM is that they use a sum-of-squared objective function, and relatively uniform cluster sizes will achieve a smaller value of this objective function. In our previous study, we have proposed an improved FCM (IFCM) by constructing a new objective function [28], in which the ''uniform effect'' of FCM is solved by normalizing the traditional one using cluster volumes. Applied to metagenomic sequences clustering, IFCM exhibited better performance than other methods [29].

Hence, load consumption patterns cannot be effectively extracted from a historical load consumption data set using the traditional CM or FCM algorithm, thus affecting the subsequent load forecasting. To address this issue, a novel STLF model based on the IFCM clustering algorithm, RF method and DNNs is proposed in this paper. First, historical load profiles are grouped using the IFCM algorithm, which is insensitive to the imbalance of days in different categories, to obtain several typical load consumption patterns. Then, the meteorological and calendar features of the training set are used to train a RF model, and the clustering result of the IFCM algorithm is used as the label of a day. Finally, a five-layer DNN is constructed for each group. The DNN model is to fit the relationship between the meteorological and calendar features and the peak and valley load values in each group. The experimental results on two daily load data sets have showed that the proposed STLF model achieves better performance in term of prediction precision than other methods.

The main contributions of this paper are: (1) the load consumption patterns are identified by an advanced clustering algorithm to eliminate the effect caused by the unevenly distribution of different date types; (2) the load consumption pattern of the day to be predicted is determined by a well-trained RF model; (3) the deep neural network is used to fit the relationships between the peak and valley load values and the meteorological and calendar features. To our knowledge, this is the first study that considers the imbalanced property of historical load data sets and solves it by a new clustering algorithm.

The rest of this paper is organized as: the proposed STLF model is described in Section II. The experimental results are

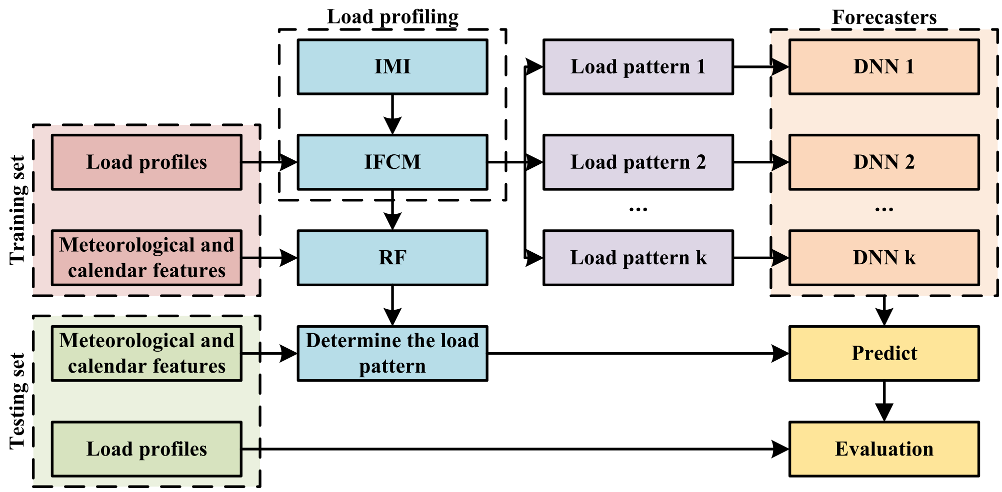


FIGURE 1. **Flowchart of the proposed STLF model.**

showed in Section III. Section IV discusses the performance of the proposed model. A brief conclusion is provided in Section V.

II. MATERIALS AND METHODS

The flowchart of the proposed STLF method is summarized in Fig. 1. First, a historical load data set is randomly partitioned into 2 subsets, namely the training set and testing set. The training set contains 70% of days of this data set, while the rest days are in the testing set. Then the IFCM clustering algorithm [28] is utilized to group the load profiles of the training set with the optimal number of clusters determined by a recent clustering validity index. After that, the meteorological and calendar features, together with the clustering result of IFCM, are used to train a RF model [30]. Meanwhile, a deep neural network [31] based forecaster is constructed for each group. Finally, the testing set is used to evaluate the prediction performance of the proposed STLF model.

A. LOAD PROFILING BASED ON IFCM ALGORITHM

FCM is a centroid-based clustering algorithm, and imbalanced distributions inter-clusters have a negligible effect on its clustering performance [28]. The reason is that FCM uses a sum of squared objective function, which is defined as:

$$J_{FCM} = \sum_{j=1}^{c} \sum_{i=1}^{N} u_{ij}^{q} \parallel \boldsymbol{x}_i - \boldsymbol{\theta}_j \parallel^2 \tag{1}$$

where $\boldsymbol{x}_i$ is the ith data point of $\boldsymbol{X}$, $\boldsymbol{\theta}_j$ is the jth cluster center, u_{ij} is the membership value of $\boldsymbol{x}_i$ to $\boldsymbol{\theta}_j$ and satisfies $\sum_{j=1}^{c} u_{ij} = 1$, q is the fuzziness degree, and c is the number of clusters. Previous studies have demonstrated that minimizing this objective function will equalize the volumes of all clusters [32]–[35], as a result some samples in one class may be classified into its adjacent class incorrectly. Considering the imbalanced number of days of different categories in a historical data set, the IFCM clustering algorithm is used to extract typical load consumption patterns in this paper.

A new objective function is designed in the IFCM clustering algorithm:

$$J_{IFCM} = \sum_{j=1}^{c} \frac{\sum_{i=1}^{N} u_{ij}^{q} DTW\left(\boldsymbol{x}_i, \boldsymbol{\theta}_j\right)^2}{f_j} \tag{2}$$

where f_j represents the volume of jth cluster and is calculated by:

$$f_j = \frac{\sum_{i=1}^{N} u_{ij}}{N} \tag{3}$$

The dynamic time warping (DTW) distance is a famous metric to measure the similarity of time series data, and is used in this paper to calculate the distance between $\boldsymbol{x}_i$ and $\boldsymbol{\theta}_j$. By introducing the volumes of clusters into the objective function, IFCM could normalize the number of samples in different clusters, and is able to improve the clustering performance of the traditional FCM for imbalanced data sets. By using the Lagrange multiplier method, the partial derivative of J_{IFCM} to the membership matrix can be obtained. Defining it to be zero, the membership matrix of IFCM can be calculated by:

$$u_{rs} = \frac{\left(\frac{f_s}{DTW(\boldsymbol{x}_r, \boldsymbol{\theta}_s)^2}\right)^{\frac{1}{q-1}}}{\sum_{j=1}^{c} \left(\frac{f_j}{DTW(\boldsymbol{x}_r, \boldsymbol{\theta}_j)^2}\right)^{\frac{1}{q-1}}} \tag{4}$$

The formula of the cluster center is as the same as the traditional FCM:

$$\boldsymbol{\theta}_j = \frac{\sum_{i=1}^{N} u_{ij}^{q} \boldsymbol{x}_i}{\sum_{i=1}^{N} u_{ij}^{q}} \tag{5}$$

The number of clusters is an important parameter in clustering algorithms. In this paper, it is determined by a recent clustering validity index, IMI [36], which is defined as:

$$IMI(c) = \frac{\sum_{j=1}^{c}\sum_{i=1}^{N} u_{ij}^{q} DTW\left(\boldsymbol{x}_i, \boldsymbol{\theta}_j\right)^2 / \sum_{i=1}^{N} u_{ij}}{\min_{l\neq j}\delta_{lj} DTW\left(\boldsymbol{\theta}_l, \boldsymbol{\theta}_j\right)^2 + \operatorname{median}_{l\neq j}\delta_{lj} DTW\left(\boldsymbol{\theta}_l, \boldsymbol{\theta}_j\right)^2} \tag{6}$$

where $\delta_{lj} = \max(f_l, f_j)/\min(f_l, f_j)$ measures the imbalanced degree of the cluster l and j. IMI is a recently proposed clustering validity index and is robust to the imperfect clustering result of FCM for imbalanced data sets. The result with the minimum IMI value will be determined as the number of clusters:

$$c^* = \arg\min_{2\leq c\leq K} IMI(c) \tag{7}$$

In summary, the pipeline of the load profiling based on the IFCM clustering algorithm is as follows:

1) Set parameters: $q \geq 2$, the number of clusters $c = 2$ to K, iteration error $\varepsilon = 10^{-5}$, max iteration number $r_{max} = 1000$, initial volume of a cluster $f_j = 1/c$, and current iteration number $r = 1$;
2) Select c samples in the training set as the initial cluster centers randomly;
3) Calculate the membership values u_{rs} according to the formula (4);
4) Update the centers of clusters according to the formula (5);
5) Calculate the volumes of clusters f_j according to the formula (3);
6) Set $r = r + 1$. If $\left(J_{IFCM}^{(r)} - J_{IFCM}^{(r+1)}\right)^2 \leq \varepsilon$ or $r \geq r_{max}$, go to the next step; otherwise return to the step 3);
7) If $c = K$, go to the next step; otherwise, set $c = c + 1$ and return to the step 2).
8) Calculate the IMI values of these clustering results and select the optimal one, and get c^* typical load consumption patterns.

B. LOAD CONSUMPTION PATTERN DETERMINATION BASED ON RANDOM FOREST

The RF algorithm is a supervised integrated learning method, and integrates several weak classification decision trees to establish a strong classifier. The fundamental theory of RF is the Bagging ensemble learning theory and the random subspace method. For a classification task, samples to be classified are used as inputs and each decision tree generates an independent classification result. The final classification labels are determined by the majority voting of all the decision trees [30], [37], [38]. Therefore, the RF algorithm can overcome some limitations of single decision tree and has been widely applied in the load forecasting [1], [39].

First, a RF model is developed based on the clustering result of historical load consumption profiles, and the meteorological and calendar features. The meteorological and calendar features are randomized and selectively extracted to establish several individual subsets of samples. Then, an independent decision tree is generated for each subset, and dual classification is applied on each node based on *Gini* gain until one, and only one, load consumption pattern is present in this node. In this case, the node is regarded as a leaf node. Eventually, several decision trees with classification capabilities are established.

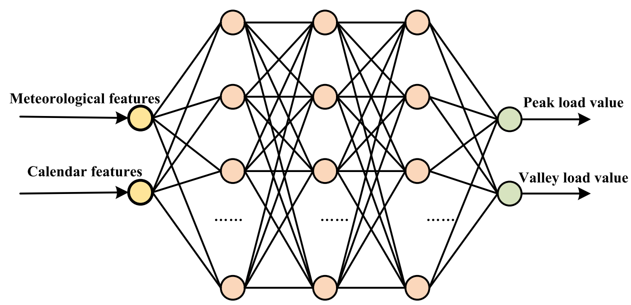


FIGURE 2. Architecture of the DNN-based forecaster.

Then, the meteorological and calendar features of the day to be predicted are employed as the inputs of the RF model and the load consumption pattern of this day is determined by the majority voting of the results of all the decision trees.

C. LOAD FORECASTING BASED ON DEEP NEURAL NETWORKS

Deep neural networks (DNN) are nonlinear representation learning methods, and typically include an input layer, an output layer and several hidden layers. Neuron is the basic component of the layers, and is a multi-input and single-output unit. It receives the information from the previous layer and transmits to the next layer after a nonlinear activation function. Due to its powerful ability of nonlinear complex function fitting, DNN has been applied into many areas.

In this paper, a DNN-based forecaster is constructed for each group. Each neural network contains 5 layers, including 1 input layer, 3 hidden layers and 1 output layer, as showed in Figure 2. The first layer inputs the meteorological and calendar features into the network, and the number of neurons is the sum of the numbers of the two types of features. The hidden layers all contain 10 neurons. The last layer outputs the peak and valley load values of the training samples. Every DNN is trained and optimized by the samples that are clustered into this group by IFCM. The mean squared error between the output values of a network and the true values is used as the loss function, and the Levenberg-Marquardt method was used to train the network.

D. FORECASTING PROCEDURE

The meteorological and calendar features of a day to be predicted are use to forecast its load profile through three steps. First, input the meteorological and calendar features of a day to be predicted into the RF model and get several labels from all the decision trees in the model. The load consumption pattern is determined by the majority voting of these labels, termed as δ_i.

TABLE 1. China federal holidays.

Holiday name	Date
New year's day	January 1
Spring festival	The first to the third day of the first lunar month
Qingming festival	April 4
International Labour Day	May 1
Dragon Boat Festival	the 5th day of the 5th lunar month
Mid-Autumn Festival	the 15th day of the 8th lunar month
National Day	October 1-3

Then, input the meteorological and calendar features of this day into the δ_ith DNN-based forecaster and get the peak and valley load values of this day.

Finally, the load pattern and the peak and valley values are combined to determine the load profile of this day by:

$$\boldsymbol{L}_i = \frac{\boldsymbol{P}_{\delta_i} - min(\boldsymbol{P}_{\delta_i})}{max(\boldsymbol{P}_{\delta_i}) - min(\boldsymbol{P}_{\delta_i})} \times (Peak - Valley) + Valley \quad (8)$$

where δ_i is the label of the day to be predicted and is determined by the RF model, $\boldsymbol{P}_{\delta_i}$ is the δ_ith load consumption pattern, and *Peak* and *Valley* are the output values of δ_ith DNN-based forecaster.

III. RESULTS

A. EXPERIMENTAL DATA SETS

Two data sets were selected to evaluate the prediction performance of the proposed STLF model. The first data set consists of daily load consumption profiles, meteorological and calendar features of a city in Jilin Province, China from 2015 to 2016, totally 710 days. The load profile of a day contains 24 load values with the sampling frequency of one hour. The meteorological features include the maximum and minimum wind speeds, maximum and minimum temperatures, and average surface pressure, while the calendar features are one-hot encoded according to the weekdays, weekends, and holidays. Holidays are determined by the federal holidays of China, as showed in Table 1. Figure 3 shows the load profile of one day in this data set.

The second data set is from a US utility and was used in the Global Energy Forecasting Competition 2014 (GEFCom2014) [40]. Hourly historical load and temperature values from 2005 to 2006, total 730 days, were selected in this paper. The calendar features were determined by the US calendar and US federal holidays, as showed in Table 2. This data set is available at https://www.sciencedirect.com/science/article/pii/S0169207016000133?via%3Dihub#ec000005. Figure 4 shows the load profile of one day in this data set.

For convenience, the two data sets are named as D_China and D_US in the rest of this paper respectively.

B. EVALUATION CRITERIA

The load forecasting performance is evaluated based on three metrics, namely the root mean square error (RMSE), the mean absolute percentage error (MAPE) and the mean absolute scaled error (MASE). RMSE evaluates the general

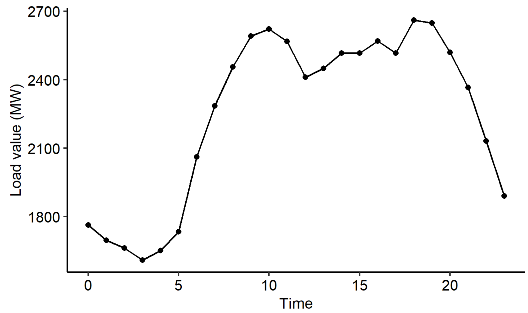


FIGURE 3. The load profile of one day in the D_China data set.

TABLE 2. US federal holidays [40].

Holiday name	Date
New year's day	January 1
Birthday of Martin Luther King, Jr.	Third Monday in January
Washington's birthday (Presidents' day)	Third Monday in February
Memorial day	Last Monday in May
Independence day	July 4
Labor day	First Monday in September
Columbus day	Second Monday in October
Veterans day	November 11
Thanksgiving day	Fourth Thursday in November
Christmas day	December 25

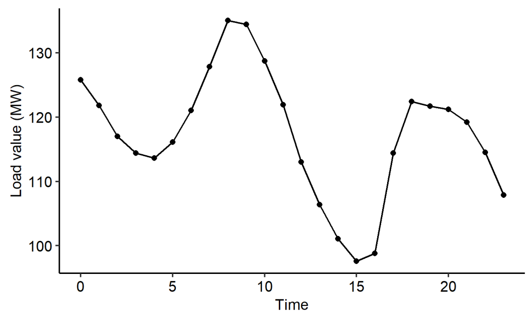


FIGURE 4. The load profile of one day in the D_US data set.

deviation between the real and the predicted load profiles, and is defined as:

$$E_{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} [f(i) - y(i)]^2}{n}}. \quad (9)$$

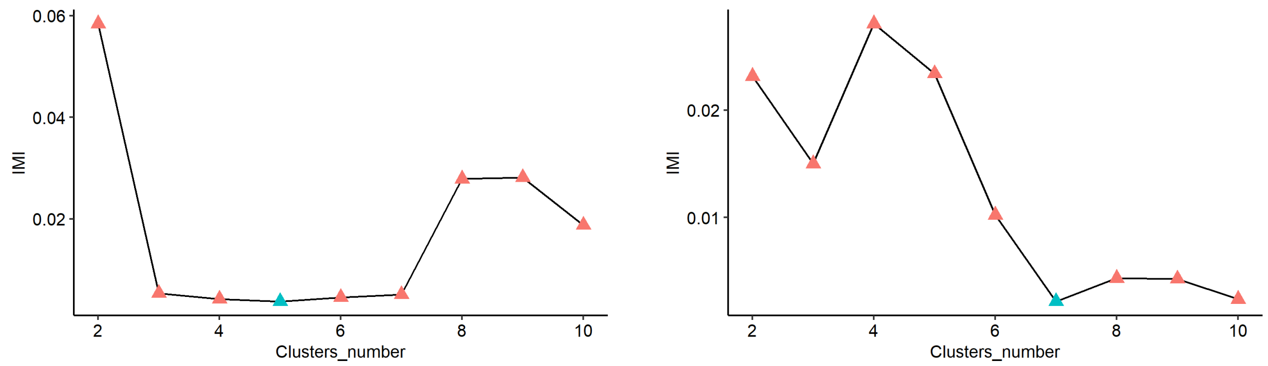
MAPE reflects the accuracy of a prediction method, and can be calculated by:

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f(i) - y(i)}{f(i)} \right|. \quad (10)$$

MASE measures the scaled mean error of the real and predicted load profiles [41] and is defined as:

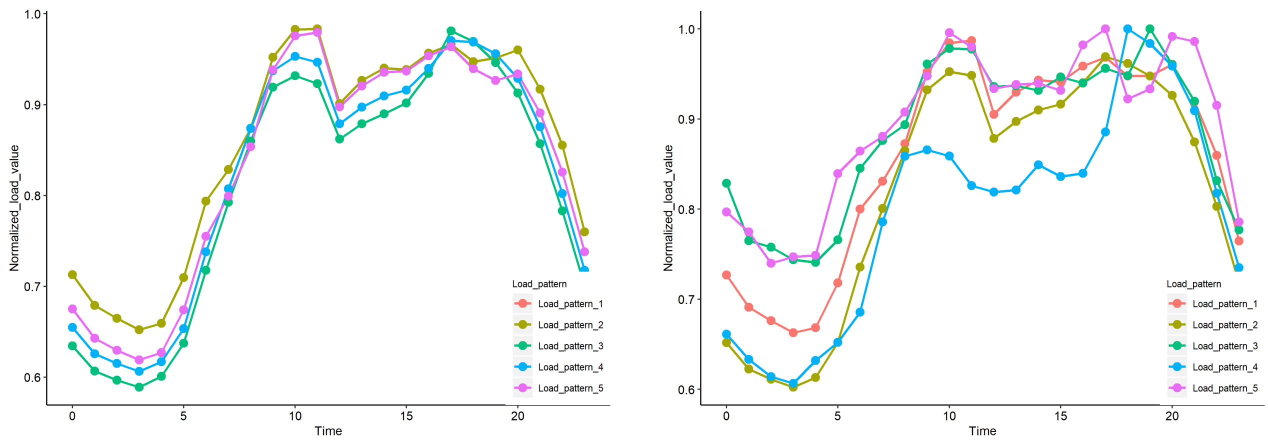
$$E_{MASE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|f(i) - y(i)|}{\frac{1}{n-1} \sum_{j=2}^{n} |f(j) - f(j-1)|}. \quad (11)$$

In the above formulas, $f(i)$ and $y(i)$ are the actual and the predicted load values of ith moment respectively, $\overline{f}$ is



(a) D_China (b) D_US

FIGURE 5. The IMI values of the clustering results of IFCM with different numbers of clusters on the D_China and D_US data sets. The point with the minimum IMI value is determined as the optimal number of clusters, which is labeled in blue.



(a) FCM (b) IFCM

FIGURE 6. Load patterns clustered by the FCM and IFCM clustering algorithms on D_China data set.

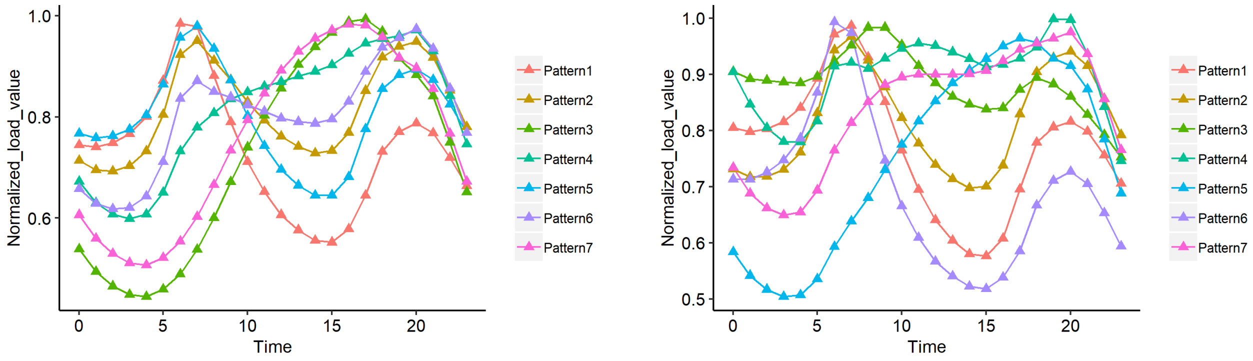
the average value of the real load profile, and n is the number of daily measurements. The smaller these metrics are, the higher the accuracy of the load forecasting model is.

C. RESULTS OF LOAD PROFILING BASED ON IFCM

First, the IFCM clustering algorithm was performed several times with different numbers of clusters (from 2 to 10) and IMI was then used to determine the optimal number of clusters for each data set. Figure 5 shows the IMI values of the clustering results of IFCM on D_China and D_US. The point with the minimum value of IMI is determined as the optimal number of clusters, and there are 5 clusters in the D_China data set and 7 clusters in the D_US data set. The numerical results of the load patterns obtained by FCM and IFCM clustering algorithms are listed in Table S1-S4 of the Supplementary File.

Figure 6 shows the load consumption patterns obtained by the FCM and IFCM clustering algorithms with the optimal number of clusters from the D_China data set. It can be observed from Fig. 6 (a) that the load patterns obtained by FCM have similar shapes. There are only 4 distinct load patterns in Fig. 6 (a), because the load pattern 1 and 4 are exactly the same. The reason lies in that the "uniform effect" of FCM occurred in this experiment, and many samples from a majority cluster might be clustered into its adjacent cluster to make their centers closer. The load pattern 1, 3 and 4 of FCM mainly consist the dates from October to February, representing the load characteristic of winter. The load pattern 2 and 5 of FCM mainly consist the dates from April to September, representing the load characteristic of summer. The difference between them is that the vast majority of weekends in these months are in the load pattern 2, while the load pattern 5 mainly includes weekdays.

IFCM is able to solve the "uniform effect" of FCM and 5 clusters with distinct load consumption patterns were discovered by IFCM, as showed in Fig. 6 (b). The load pattern 3 and 5 of IFCM represent the load characteristic of winter. The load pattern 2 and 4 of IFCM represent the load characteristic of summer; the former load pattern mainly includes weekdays, while majority of weekends in summer are in the latter one. There are 41 days of holidays in the D_China data set, and 24 of them were clustered in the load consumption pattern 1.



(a) FCM (b) IFCM

FIGURE 7. Load patterns clustered by the FCM and IFCM clustering algorithms on D_US data set.

TABLE 3. Comparison of clustering results of IFCM on other algorithms on D_China and D_US data sets. The minimum values are written in bold.

		FCM	csiFCM	siibFCM	IFCM
D_China	XBI	1e15	1e35	5.5805	**0.9309**
	FSI	-0.5917	-0.3237	-1.1917	**-3.1497**
	WLI	9.5397	9.0446	4.6696	**0.985**
	IMI	8.7125	8.2461	3.196	**0.9168**
D_US	XBI	0.5633	**0.4556**	0.4647	0.8993
	FSI	-371.6394	-321.4555	-543.708	**-901.8275**
	WLI	0.4276	0.4061	0.4609	**0.3244**
	IMI	0.0044	0.0044	0.3548	**0.0044**

Figure 7 shows the load consumption patterns obtained by the FCM and IFCM clustering algorithms with the optimal number of clusters from the D_US data set. It can be observed that the load patterns obtained by IFCM are more diverse than that of FCM, such as the load pattern 3 of IFCM.

Because the label information of each historical load data set is not available, the clustering performance of the IFCM clustering algorithm was than evaluated by 4 commonly used internal metrics, FSI [42], XBI [43], WLI [44] and IMI [36], and was compared with the traditional FCM algorithm and two improved FCM algorithms, csiFCM [45] and siibFCM [46]. FSI, XBI, WLI and IMI evaluate a clustering result of a data set according to its structure of clusters with no need for the ground truth of samples in this data set. Smaller values of these indexes indicate better clustering results. Table 3 lists the values of the four indexes on the clustering results of the D_China and D_US data sets by IFCM and other clustering algorithms respectively. We can find that IFCM achieved smaller values of all of these indexes than these of other algorithms on the D_China data set and three indexes on the D_US data set, illustrating that the IFCM algorithm exhibits the ability to improve the performance of the load profiling than the traditional FCM, and outperforms the other two improved FCM algorithms that are also designed to solve the "uniform effect" of the traditional FCM.

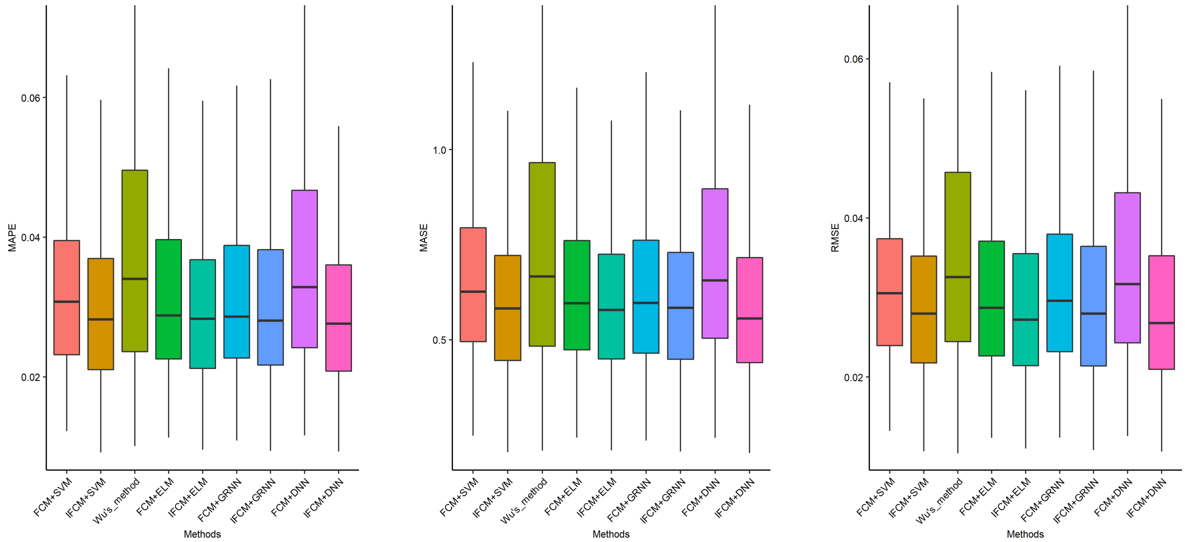
D. RESULTS OF LOAD FORECASTING

First, the IFCM clustering algorithm was combined with four famous regression models, the support vector machine (SVM), general regression neural network (GRNN), extreme learning machine (ELM) and DNN models, and was compared with the traditional FCM clustering algorithm. Figure 8 and 9 show the box diagrams of the MAPE, MASE and RMSE values of the results on the testing sets of the D_China and D_US data sets predicted by these methods respectively. It can be observed that the prediction performance is significantly improved by utilizing IFCM on the D_China data set as the levels of MAPE, MASE and RMSE are much lower than that of FCM. For the D_US data set, a distinct smaller levels of MAPE and MASE were achieved by combining IFCM with SVM, ELM and DNN as compared to the traditional FCM. Table 4 lists the average values of MAPE, MASE and RMSE on the testing sets of the D_China and D_US data sets predicted by these methods respectively. For the D_China data set, the average values of MAPE, MASE and RMSE of the IFCM+SVM model are 0.4%, 6% and 0.3% less than that of the FCM+SVM model respectively, these values of the IFCM+GRNN model are 0.1%, 2% and 0.1% less than that of the FCM+GRNN model respectively, these values of the IFCM+ELM model are 0.3%, 7% and 0.4% less than that of the FCM+ELM model respectively, and these values of the proposed model are 0.9%, 16% and 1.2% less than that of the FCM+DNN model respectively. For the D_US data set, the average values of MAPE, MASE and RMSE of the IFCM+SVM model are 0.6%, 7% and 0.3% less than that of the FCM+SVM model respectively, these values of the IFCM+GRNN model are 0.3%, 5% and 0.1% less than that of the FCM+GRNN model respectively, these values of the IFCM+ELM model are 4%, 83% and 3% less than that of the FCM+ELM model respectively, and these values of the proposed model are 0.6%, 4% and 0.3% less than that of the FCM+DNN model respectively.

Then, the prediction performance for holidays was evaluated. Figure 10 and 11 show the prediction performance of these methods on holidays in the D_China and D_US data sets respectively. It can be observed that the the performance of holidays was significantly improved on both data sets by combining the IFCM clustering algorithm with

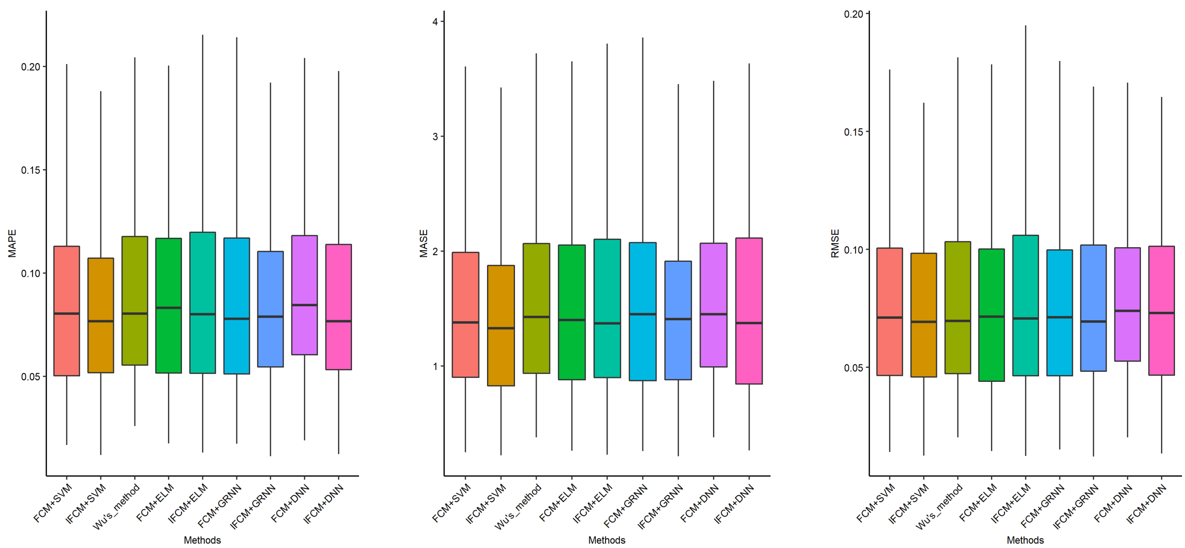
TABLE 4. Average values of MAPE, MASE and RMSE on the testing sets of the D_China and D_US data sets. The best results are written in bold.

		FCM+SVM	IFCM+SVM	FCM+DNN	Wu's method	FCM+GRNN	IFCM+GRNN	FCM+ELM	IFCM+ELM	Proposed
D_China	MAPE	0.035	0.031	0.038	0.04	0.033	0.032	0.035	0.032	**0.031**
	MASE	0.69	0.63	0.74	0.78	0.66	0.64	0.70	0.63	**0.62**
	RMSE	0.033	0.3	0.036	0.037	0.032	0.031	0.034	0.030	**0.029**
D_US	MAPE	0.09	**0.084**	0.094	0.092	0.09	0.087	0.16	0.12	0.088
	MASE	1.56	**1.49**	1.59	1.59	1.59	1.54	2.93	2.1	1.55
	RMSE	0.077	**0.074**	0.08	0.08	0.078	0.077	0.14	0.11	0.077



(a) MAPE (b) MASE (c) RMSE

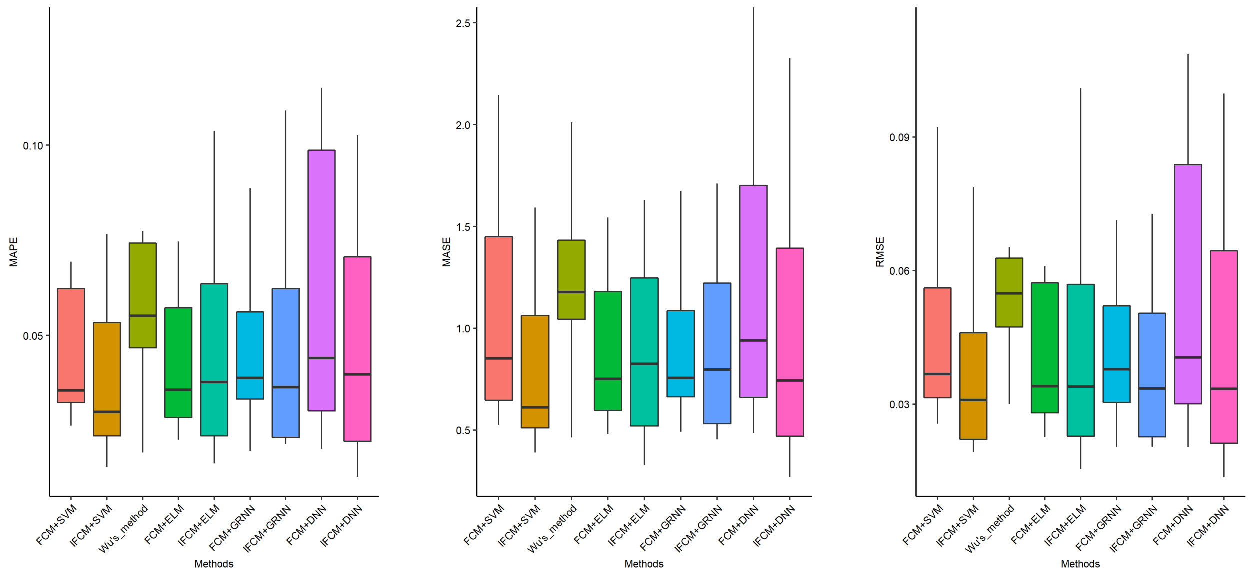
FIGURE 8. Comparison of the prediction performance of different prediction models on the D_China data set. The MAPE, MASE and RMSE values in the testing set are counted and showed in box plots.



(a) MAPE (b) MASE (c) RMSE

FIGURE 9. Comparison of the prediction performance of different prediction models on the D_US data set. The MAPE, MASE and RMSE values in the testing set are counted and showed in box plots.

the four regression methods. Table 5 lists the average values of MAPE, MASE and RMSE on holidays of the testing sets of the D_China and D_US data sets predicted by these methods. For the D_China data set, the average values of MAPE, MASE and RMSE of the IFCM+SVM model are 0.4%, 6% and 0.3% less than that of the FCM+SVM model respectively, these values of the IFCM+GRNN model are all 0.2% less than that of the FCM+GRNN model respectively, the average value of MASE of the IFCM+ELM model is 0.2% than that of the FCM+ELM model, and the average



(a) MAPE (b) MASE (c) RMSE

FIGURE 10. Comparison of the prediction performance of different prediction models on holidays in the D_China data set.

TABLE 5. Average values of MAPE, MASE and RMSE on holidays of the D_China and D_US data sets. The best results are written in bold.

		FCM+SVM	IFCM+SVM	FCM+DNN	Wu's method	FCM+GRNN	IFCM+GRNN	FCM+ELM	IFCM+ELM	Proposed
D_China	MAPE	0.055	0.041	0.061	0.059	0.050	0.048	0.046	0.046	**0.046**
	MASE	1.13	0.87	1.23	1.2	1.02	0.098	0.098	0.096	**0.96**
	RMSE	0.05	0.039	0.055	0.054	0.045	0.043	0.043	0.043	**0.043**
D_US	MAPE	0.097	**0.063**	0.096	0.09	0.093	0.092	0.096	0.073	0.076
	MASE	1.85	**1.25**	1.93	1.79	1.94	1.89	2.00	1.39	1.59
	RMSE	0.084	**0.059**	0.085	0.079	0.086	0.084	0.089	0.066	0.07

values of MAPE, MASE and RMSE of the proposed model are 0.9%, 16% and 1.2% less than that of the FCM+DNN model respectively. For the D_US data set, the average values of MAPE, MASE and RMSE of the IFCM+SVM model are 3.4%, 60% and 2.5% less than that of the FCM+SVM model respectively, these values of the IFCM+GRNN model are 0.1%, 5% and 0.2% less than that of the FCM+GRNN model respectively, these values of MAPE, MASE and RMSE of the IFCM+ELM model are 2.3%, 61% and 2.3% less than that of the FCM+ELM model respectively, and these values of MAPE, MASE and RMSE of the proposed model are 2%, 34% and 1.5% less than that of the FCM+DNN model respectively.

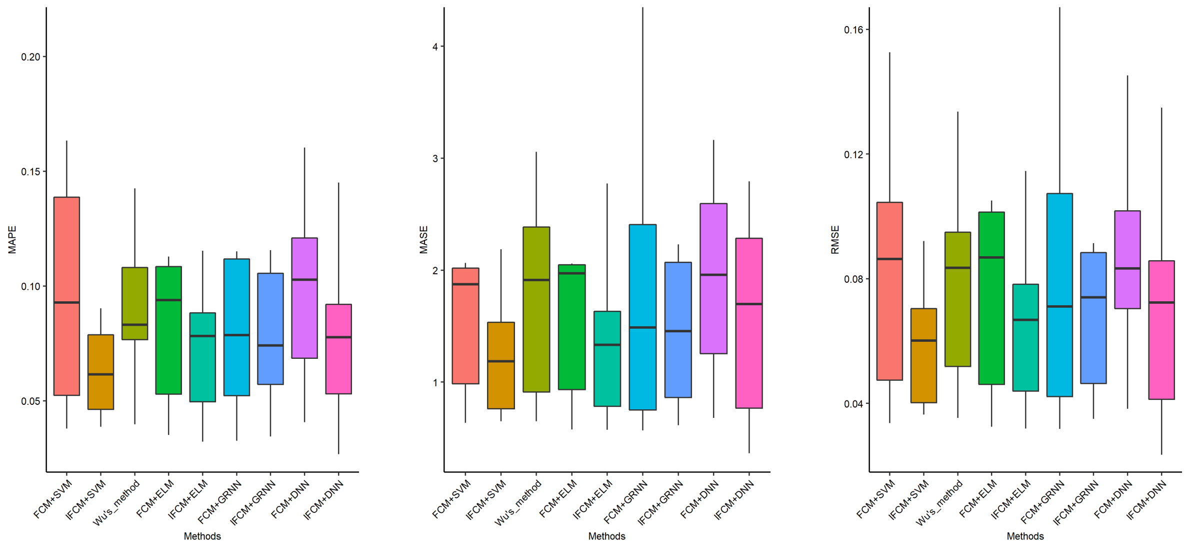
Finally, the proposed STLF model was compared with the method in [1], termed as Wu's method, in which the k-means clustering algorithm and the linear regression model are used. The prediction results on the two data sets have showed the superiority of the proposed STLF model than that of the Wu's method. For the D_China data set, the average values of MAPE, MASE and RMSE of the proposed model are 0.9%, 16% and 1.2% less than that of the Wu's method respectively, and are 0.4%, 4% and 0.3% less than that of the Wu's method for the D_US data set respectively. The prediction performance of holidays was also improved. For the holidays in the D_China data set, the average values of MAPE, MASE and RMSE of the proposed model are 1.3%, 24% and 1.1% less than that of the Wu's method respectively, and are 1.4%, 20% and 0.9% less than that of the Wu's method for the holidays in the D_US data set respectively.

E. STABILITY ANALYSIS

The stability of the proposed STLF model was evaluated by the population stability index (PSI). PSI is the most common criterion to analyze the stability of a model by measuring the distribution difference between the testing samples and the training samples. If the PSI value of a model is less than 0.1, the stability of the model is very high. The proposed STLF model was first to predict the load profiles of the samples in the training set and the testing set of the D_China data set. Then the MAPE values of the results of the two sets were calculated and were grouped into 10 equal intervals respectively. The distributions of the MAPE values of the training set and the testing set were used as the expected and actual distributions respectively. Then the PSI can be calculated by:

$$PSI = \sum_{i=1}^{10} [(Ac_i - Ex_i) \times \ln(Ac_i/Ex_i)] \tag{12}$$

Tabel 6 lists the statistical values and the PSI values of all intervals. The PSI value of the proposed model is 0.0497, illustrating that the model is very stable. Figure 12 shows



(a) MAPE (b) MASE (c) RMSE

FIGURE 11. Comparison of the prediction performance of different prediction models on holidays in the D_US data set.

TABLE 6. The statistical values and the PSI values of all intervals.

Intervals	Actural	Expected	Ac-Ex	ln(Ac/Ex)	Index
0-1%	0.0047	0.0021	0.0027	0.3624	0.0010
1.01%-2%	0.1943	0.2078	-0.01351	-0.02918	0.0004
2.01%-3%	0.3128	0.3992	-0.0864	-0.1059	0.0091
3.01%-4%	0.2512	0.2490	0.0022	0.0038	8.51e-06
4.01%-5%	0.1090	0.0823	0.0267	0.1220	0.0033
5.01%-6%	0.0616	0.03086	0.03075	0.3002	0.0092
6.01%-7%	0.0237	0.0165	0.0072	0.1582	0.0011
7.01%-8%	0.0095	0.0041	0.0054	0.3624	0.0019
8.01%-9%	0.0095	0.0062	0.0033	0.1863	0.0006
>9%	0.02369	0.0021	0.0216	1.0613	0.0229
FSI value					0.0497

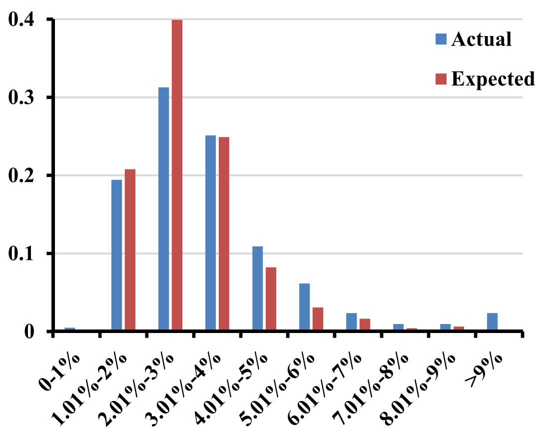


FIGURE 12. The histogram of the percentages of the MAPE values in 10 intervals.

the actual and expected histograms of the percentages of the MAPE values in 10 intervals.

IV. DISCUSSION

In this paper, a novel hybrid STLF model is proposed. The core of this model is that an advanced clustering algorithm is utilized for the discovery of the load consumption patterns.

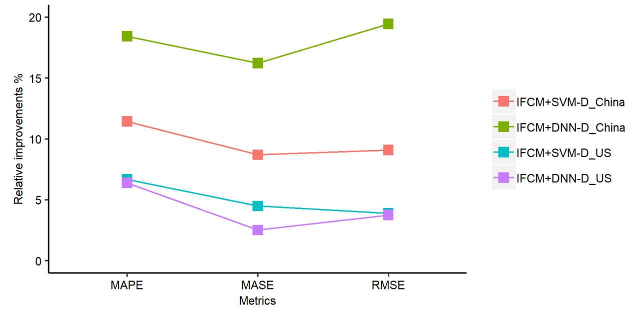


FIGURE 13. The percentages of the relative improvements by combining IFCM with SVM and DNN on the two data sets.

The IFCM clustering algorithm was combined with four regression models and the performance has been improved as compared to the traditional FCM clustering algorithm. The percentages of relative improvements by combining IFCM with SVM and DNN models were calculated and were plotted in Figure 13. There are two interesting findings can be observed. First, the extent improvements by utilizing IFCM on the D_China data set are much larger that on the D_US data set. We infer that this is caused by the different number of holidays in the testing sets of the two data sets. There are 15 holidays in the testing set of the D_China data set, but only 8 holidays of the D_US data set. IFCM is a powerful clustering algorithm for imbalanced data sets. As mentioned above, the load consumption pattern of holidays has been identified by IFCM, which could improve the prediction performance for holidays. As there are much more holidays in the testing set of the D_China data set than that of the D_US data set, the overall performance improvements of the D_China data set is greater than that of the D_US data set.

Secondly, the IFCM+DNN model is better than the IFCM+SVM model for the D_China data set and on the contrary for the D_US data set. This is probable caused by the different number of clusters determined by the IMI index for the two data sets. The samples in the D_China data set were partitioned into 5 clusters, while the samples in the D_US data set were partitioned into 7 clusters. Therefore, in the D_US data set, the number of samples in each cluster used to train the DNN model is much less than that in the D_China data set. The under-fitting may happen when training the DNN models for the D_US data set, and may affect the overall prediction performance.

V. CONCLUSION

In this paper, a novel STLF model is proposed by combining IFCM, RF method and DNNs. The core contributions of the proposed model are the consideration of the imbalanced property of a historical load data set, and the application of an advanced clustering algorithm to perform load profiling. Compared with the traditional FCM, IFCM could provide better results of load profiling. Experimental results on two data sets have showed the superior performance of the proposed model than that of other methods. Additionally, it's worth noting that the prediction performance of holidays has been significantly improved by using IFCM. Therefore, the proposed STLF model is a powerful model and can be applied to the load forecasting in other region of worldwide. IFCM can be used to discover the load consumption patterns from smart meter data and to improve the performance of the meter-level load forecasting.

There are several promising directions in our future study: 1) whether some external factors can be considered in the prediction model to improve the forecasting accuracy, such as the economic data and the population size of a city; 2) how to determine the optimal combinations of meteorological features; 3) how to combine IFCM with other deep neural networks, such as the RNN.

REFERENCES

[1] R. Z. Wu, Z. R. Bao, W. T. Wang, W. Deng, and L. Tang, "Short-term power load forecasting method based on pattern matching in Hadoop framework," *Trans. China Electrotech. Soc.*, vol. 33, no. 7, pp. 1542–1551, 2018.

[2] 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, Sep. 2017.

[3] Q. Wang, B. Zhou, Z. Li, and J. Ren, "Forecasting of short-term load based on fuzzy clustering and improved bp algorithm," in *Proc. Int. Conf. Electr. Control Eng.*, 2011, pp. 4519–4522.

[4] R. Ramanathan, R. Engle, C. W. J. Granger, F. Vahid-Araghi, and C. Brace, "Short-run forecasts of electricity loads and peaks," *Int. J. Forecasting*, vol. 13, no. 2, pp. 161–174, 1997.

[5] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.

[6] S. Sargunaraj, D. P. S. Gupta, and S. Devi, "Short-term load forecasting for demand side management," *IEE Gener., Transmiss. Distrib.*, vol. 144, no. 1, pp. 68–74, Jan. 1997.

[7] X. Wang, W. J. Lee, H. Huang, R. L. Szabados, D. Y. Wang, and P. V. Olinda, "Factors that impact the accuracy of clustering-based load forecasting," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3625–3630, Sep. 2016.

[8] Y. M. Liu, S. L. Lei, C. X. Sun, Q. Zhou, and H. J. Ren, "A multivariate forecasting method for short-term load using chaotic features and RBF neural network," *Eur. Trans. Elect. Power*, vol. 21, no. 3, pp. 1376–1391, 2011.

[9] B. G. Koo, H. S. Lee, and J. Park, "Short-term electric load forecasting based on wavelet transform and gmdh," *J. Electr. Eng. Technol.*, vol. 10, no. 3, pp. 832–837, 2015.

[10] X. Liu, Z. J. Zhang, and Z. Song, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," *Renew. Sustain. Energy Rev.*, vol. 119, 2020.

[11] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017.

[12] G. Hafeez, K. S. Alimgeer, and I. Khan, "Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 114915.

[13] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019.

[14] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018.

[15] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.

[16] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[17] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[18] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.

[19] K. E. Farfar and M. T. Khadir, "A two-stage short-term load forecasting approach using temperature daily profiles estimation," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3909–3919, Aug. 2019.

[20] G. Kan, X. He, J. Li, L. Ding, D. Zhang, T. Lei, Y. Hong, K. Liang, D. Zuo, Z. Bao, and M. Zhang, "A novel hybrid data-driven model for multi-input single-output system simulation," *Neural Comput. Appl.*, vol. 29, no. 7, pp. 577–593, Apr. 2018.

[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] S. V. Oprea and A. Bara, "Machine learning algorithms for short-term load forecast in residential buildings using smart meters, sensors and big data solutions," *IEEE Access*, vol. 7, pp. 177874–177889, 2019.

[23] I. P. Panapakidis, "Application of hybrid computational intelligence models in short-term bus load forecasting," *Expert Syst. Appl.*, vol. 54, pp. 105–120, Jul. 2016.

[24] Y. Y. Song, C. B. Li, and Z. Q. Qi, "Extraction of power load patterns based on cloud model and fuzzy clustering," *Power Syst. Technol.*, vol. 38, no. 12, pp. 3378–3383, 2014.

[25] L. Feng and J. J. Qiu, "Electronical load forecasting based on load patterns," *Power Syst. Technol.*, vol. 29, no. 4, pp. 23–26, 2005.

[26] X. Jin, L. W. Li, J. N. Ji, Z. Q. Li, Y. Hu, and Y.-B. Zhao, "Power short-term load forecasting based on big data and optimization neural network," *J. Commun.*, vol. 37, no. Z1, pp. 36–42, 2016.

[27] J. C. Noordam, W. H. A. M. van den Broek, and L. M. C. Buydens, "Multivariate image segmentation with cluster size insensitive fuzzy C-means," *Chemometric Intell. Lab. Syst.*, vol. 64, no. 1, pp. 65–78, Oct. 2002.

[28] Y. Liu, T. Hou, and F. Liu, "Improving fuzzy c-means method for unbalanced dataset," *Electron. Lett.*, vol. 51, no. 23, pp. 1880–1881, 2015.

[29] Y. Liu, T. Hou, B. Kang, and F. Liu, "Unsupervised binning of metagenomic assembled contigs using improved fuzzy C-Means method," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 6, pp. 1459–1467, Nov. 2017.

[30] G. Gazzola and M. K. Jeong, "Dependence-biased clustering for variable selection with random forests," *Pattern Recognit.*, vol. 96, Dec. 2019, Art. no. 106980.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] J. Liang, L. Bai, C. Dang, and F. Cao, "The K-means-type algorithms versus imbalanced data distributions," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 728–745, Aug. 2012.
[33] H. Xiong, J. Wu, and J. Chen, "K-means clustering versus validation measures: A data-distribution perspective," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 39, no. 2, pp. 318–331, Apr. 2009.
[34] K. L. Zhou and S. L. Yang, "Exploring the uniform effect of FCM clustering: A data distribution perspective," *Knowl.-Based Syst.*, vol. 96, pp. 76–83, Mar. 2016.
[35] K. Zhou and S. Yang, "Effect of cluster size distribution on clustering: A comparative study of k-means and fuzzy c-means clustering," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 455–466, 2019.
[36] Y. Liu, Y. Jiang, T. Hou, and F. Liu, "A new robust fuzzy clustering validity index for imbalanced data sets," *Inf. Sci.*, vol. 547, pp. 579–591, Feb. 2021.
[37] B. Ait Hammou, A. Ait Lahcen, and S. Mouline, "An effective distributed predictive model with matrix factorization and random forest for big data recommendation systems," *Expert Syst. Appl.*, vol. 137, pp. 253–265, Dec. 2019.
[38] A. Nadi and H. Moradi, "Increasing the views and reducing the depth in random forest," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112801.
[39] D. W. Wang and Z. W. Sun, "Big data analysis and parallel load forecasting of electric power user side," *Proc. CSEE*, vol. 35, no. 3, pp. 527–537, 2015.
[40] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, Jul. 2016.
[41] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006.
[42] Y. Fukuyama, M. Sugeno, Y. Fukuyama, and M. Sugeno, "A new method of choosing the number of clusters for the fuzzy c-means method," in *Proc. 5th Fuzzy Syst. Symp.*, 1989, pp. 247–250.
[43] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
[44] C.-H. Wu, C.-S. Ouyang, L.-W. Chen, and L.-W. Lu, "A new fuzzy clustering validity index with a median factor for centroid-based clustering," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 3, pp. 701–718, Jun. 2015.
[45] J. C. Noordam, W. H. A. M. van den Broek, and L. M. C. Buydens, "Multivariate image segmentation with cluster size insensitive fuzzy C-means," *Chemometric Intell. Lab. Syst.*, vol. 64, no. 1, pp. 65–78, Oct. 2002.
[46] P.-L. Lin, P.-W. Huang, C. H. Kuo, and Y. H. Lai, "A size-insensitive integrity-based fuzzy c-means method for data clustering," *Pattern Recognit.*, vol. 47, no. 5, pp. 2042–2056, May 2014.
[47] Y. Liu, S. Lei, C. Sun, Q. Zhou, and H. Ren, "A multivariate forecasting method for short-term load using chaotic features and RBF neural network," *Eur. Trans. Electr. Power*, vol. 21, no. 3, pp. 1376–1391, Apr. 2011.

FU LIU received the B.S. and M.S. degrees from the Jilin University of Technology in 1991 and 1994, respectively, and the Ph.D. degree from the College of Communication Engineering, Jilin University, in 2002. He is currently a Professor with Jilin University. His research interests include machine vision, pattern recognition, bioinformatics, and biometrics.

TIAN DONG received the B.S. and M.S. degrees from the College of Communication Engineering, Jilin University, where he is currently pursuing the Ph.D. degree. His areas of research include machine learning and load forecasting.

TAO HOU received the B.S. degree from the College of Mathematics, and the M.S. and Ph.D. degrees from the College of Communication Engineering, Jilin University. She is currently a Lecturer with Jilin University. Her areas of research include machine learning and bioinformatics.

YUN LIU (Member, IEEE) received the B.S. and Ph.D. degrees from the College of Communication and Engineering, Jilin University, in 2011 and 2016, respectively. He is currently a Lecturer with Jilin University. His areas of research include machine learning and data mining.