Energy 166 (2019) 653–663

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Short term load forecasting based on feature extraction and improved general regression neural network model

Yi Liang [a, *], Dongxiao Niu [a], Wei-Chiang Hong [b]

[a] School of Economics and Management, North China Electric Power University, Beijing 102206, China
[b] Department of Information Management, Oriental Institute of Technology, Panchiao, New Taipei 226, Taiwan

ARTICLE INFO

Article history:
Received 27 May 2018
Received in revised form
16 October 2018
Accepted 20 October 2018
Available online 23 October 2018

Keywords:
Short term load forecasting (STLF)
Empirical mode decomposition (EMD)
Minimal redundancy maximal relevance (mRMR)
General regression neural network (GRNN)
Fruit fly optimization algorithm (FOA)

ABSTRACT

Along with the deregulation of electric power market as well as aggregation of renewable resources, short term load forecasting (STLF) has become more and more momentous. However, it is a hard task due to various influential factors that leads to volatility and instability of the series. Therefore, this paper proposes a hybrid model which combines empirical mode decomposition (EMD), minimal redundancy maximal relevance (mRMR), general regression neural network (GRNN) with fruit fly optimization algorithm (FOA), namely EMD-mRMR-FOA-GRNN. The original load series is firstly decomposed into a quantity of intrinsic mode functions (IMFs) and a residue with different frequency so as to weaken the volatility of the series influenced by complicated factors. Then, mRMR is employed to obtain the best feature set through the correlation analysis between each IMF and the features including day types, temperature, meteorology conditions and so on. Finally, FOA is utilized to optimize the smoothing factor in GRNN. The ultimate forecasted load can be derived from the summation of the predicted results for all IMFs. To validate the proposed technique, load data in Langfang, China are provided. The results demonstrate that EMD-mRMR-FOA-GRNN is a promising approach in terms of STLF.

1. Introduction

Nowadays, increasing attention has been paid to power load forecasting due to the burgeoning development of management modernization in electric power systems. Accurate load prediction not only contributes to contradiction alleviation between power supply and demand, but also administers to stable and reliable operation of the power grid [1]. However, as a random and non-stationary series, electric load is affected by diverse influential factors, such as day types, weather conditions, economic indicators and so on, which brings about lots of challenges to load forecasting [2].

Over the past decades, the development of mathematical theory and modern computational technology boost continuous improvement of load forecasting methods, mainly including traditional prediction approaches and intelligent algorithms. For conventional forecasting techniques such as time series [3–6], regression analysis [7–10] and gray analysis [11–13]. For example, in Ref. [3], authors propose a combination of time series with seasonal auto regressive fractionally integrated moving average. The forecasting results indicates that the proposed method presents higher accuracy than its counterparts. In Ref. [10], authors use two separate forecast processes: seasonal and trend items, for the short term load data from Victoria grid in Australia. After eliminating the seasonal item in the original load demand, the regression model is employed to forecast the trend item. The forecasting results indicate that the proposed model can significantly improve the forecasting accuracy. In Ref. [11], authors also suggests a hybrid method, called Gray-Fuzzy-Markov Chain Method, comprising three stages. In the first stage, daily load is forecasted by Gray model, with its training deviations classified, in a second stage, by fuzzy-set theory, and finally, fed into Markov chain model to predict future relative errors that might be supplied by the Gray model. The obtained forecasts by proposed model proved to have better prediction properties compared to the other forecasting techniques, such as Gray models, specifically GM(1,1) and GM(1,2), time series. There are certain defects that lead to the unideal prediction accuracy despite the simple calculation and mature theory. Wherein, the basic time series methods merely takes time into account and ignores the influence of other influential factors, thus there exist large errors faced with great environmental changes [14].

* Corresponding author.
E-mail address: lianglouis@126.com (Y. Liang).

https://doi.org/10.1016/j.energy.2018.10.119
0360-5442/

Additionally, the selection or expression of input in regression analysis has influence on the diversity and testability, which limits its application in STLF [15]. The gray analysis has a good fitting effect on the smooth discrete series, thus this approach is not suitable for STLF due to the discrete data [16].

At this stage, with the propositions and prosperities of artificial intelligent algorithms, scholars gradually apply these models to load forecasting, such as artificial neural networks (ANNs) [17,18,20] and support vector machine (SVM) [17,19,20]. For example, in Ref. [17], authors reveal the effect of data integrity attacks on the accuracy of four representative load forecasting models (multiple linear regression, support vector regression, artificial neural networks, and fuzzy interaction regression). Then, the four aforementioned load forecasting models are used to generate one-year-ahead ex post point forecasts in order to provide a comparison of their forecast errors. The results show that the support vector regression model is most robust, followed closely by the multiple linear regression model, while the fuzzy interaction regression model is the least robust of the four. In Ref. [20], authors build a power consumption forecasting model using various machine learning algorithms. They propose two electric load forecasting models using artificial neural network and support vector regression. The experimental results show that the two forecasting models can achieve average error rate of 3.46–10% for all clusters.

This kind of approaches improve forecasting accuracy through simulation of human brain mechanism with self-learning and self-optimizing capability [21]. Back propagation neural network (BPNN) is a typical representative of ANNs. Reference [22] introduced BPNN into STLF with the consideration of weather factors. Reference [23] integrated improvement differential evolution with wavelet neural network to forecast the load. However, the drawbacks of slow convergence and easily falling into local optimum limit their application [24]. Therefore, SVM is utilized to avoid network structure selection and local optimization in load forecasting. For Reference [25] and [26], it can be seen that the prediction precision of SVM increases, but there is difficulty to cope with large-scale training samples and multi-classification problems.

Generalized Regression Neural Network (GRNN) is a highly parallel radial basis function network based on one-pass algorithm, which can approximate the implied mapping relationship based on samples [27]. Even if the samples are scare, the output results of GRNN are able to converge to the optimal regression. This algorithm has been applied to many forecasting fields owing to risk reduction of tapping into local optimum and improvement of the learning rate and generalization ability [28–33]. For example, In Ref. [28], authors apply GRNN in wind speed forecasting. The mean absolute percentage error of the forecasting results in two cases are respectively 8.95% and 9.87%, suggesting that the proposed approach outperforms the compared models. In Ref. [29], authors propose that GRNN can also be used in electrical power system forecasting. In view of the effect of the smoothing factor on GRNN performance, fruit fly optimization algorithm (FOA) is employed in this paper to determine its value. FOA is a novel global optimization technique on the basis of foraging behaviors [34]. Fruit fly itself is superior to other species especially in terms of olfaction and vision. The olfactory organs can collect all kinds of odor floating in the air, even the food sources 40 km away. Then the sharp vision can be effectively utilized to find the location of food as well as companions and fly to that direction. It is proved that this approach presents excellent performance and high efficiency in solving complex optimization problems [35–37], especially in optimizing the smoothing factor of GRNN [28,35,38–40].

Multiple factors make difference to power load, such as temperature, meteorology and day types. For instance, the accumulation effect of temperature in summer on load is obviously significant [41]. Meanwhile, large-scale connection of electric vehicles and renewable distributed power as well as the implementation of demand side management based on electricity price and incentive intensify the load forecasting complexity. Thus the existing single intelligent algorithms are hard to obtain satisfactory forecasting results [42]. In order to deal with the challenges, the combination of data processing methods and prediction approaches has been widely applied to STLF. The hybrid prediction techniques are primarily divided into two categories, one is weighted integration of the forecasting results derived from multiple single models. Nie et al. [43] studied STLF based on autoregressive integrated moving average (ARIMA) and SVM. Niu et al. [44] applied BPNN-SVM-KELM model to prediction on the foundation of variance-covariance weight dynamic allocation. This type of approaches give full play to the advantages of each single model, but the accuracy is restricted in virtue of the fact that the complexity of the original data itself is not dealt with. The other is the combination of data pre-processing technology and intelligent algorithm. Wherein, data pre-processing methods contain two aspects, namely the decomposition of original load series and extraction of correlated factors.

The decomposition methods of original load series commonly used at home and abroad consist of wavelet transform (WT) and empirical mode decomposition (EMD). It is notable that WT has the advantage of localization [45] but it's a difficult task to rationalize the wavelet basis and decomposition scale [46]. EMD can effectively decompose nonstationary load series with adaptive ability to further improve prediction precision due to the fact that there is no need to select wavelet basis [47–49]. In terms of extraction of modal associated factors after decomposition, Reference [50] executed conditional mutual information to extract features. The first 50 characteristics were selected as the best feature set by sorting the candidate features. In Reference [51], the input was studied through feature extraction from temperature and historical load based on mutual information. The results demonstrated that the processing could improve the prediction precision. Reference [52] adopted phase space reconstruction to establish the original feature set and analyzed the correlation between load and these characteristics to complete optimal feature selection. However, these techniques merely take the relevance into account and ignore the redundancy among factors. In Reference [53], minimal redundancy maximal relevance (mRMR) applied to pattern recognition feature selection was proposed which not only considered the correlation between characteristics and target variables, but also obtained the redundant information among features. Due to its advantages, mRMR has been applied in various fields, such as wind speed forecasting [54], epigenetic biomarker identification [55] and feature selection of transient stability assessment [56].

This paper proposes an integrated model that combines EMD, mRMR, FOA with GRNN. EMD is firstly exploited to decompose the original load series into regular IMFs and a residue. mRMR is utilized to derive the best feature set through the correlation analysis between each IMF and the features. Then, GRNN optimized by FOA is treated as the forecasting tool. EMD-mRMR-FOA-GRNN can improve the prediction accuracy in load forecasting as a result of input reduction and full consideration of external sensitive factors. The rest of the paper is organized as follows: Section 2 provides a brief description of EMD, mRMR, GRNN, FOA and establishes a complete prediction framework; Section 3 verifies the precision and stability of the developed model through a practical case; Section 4 makes a further validation and the paper is concluded in Section 5.

2. Methodology

2.1. EMD

The load can be regarded as a time series composed of a set of intrinsic mode functions (IMFs) [47]. According to the definition of IMF, there only exists one vibration mode in each cycle and no other complicated singularities. However, the raw data may contain multiple vibration modes at any time Thus, the application of EMD to decompose load series is grounded on two assumptions: (1) The complex signal to be decomposed is made up of IMFs. (2) Each IMF is independent of each other. The specific steps of EMD are summarized as follows:

(1) Identify the local maxima and minima in the original time series $x(t)$ and employ cubic spline function to fit the upper envelope $e_{up}(t)$ and lower envelope $e_{low}(t)$.
(2) Calculate the mean value $m_1(t)$ of the two envelopes $e_{up}(t)$ and $e_{low}(t)$.

$$m_1(t) = \frac{e_{up}(t) + e_{low}(t)}{2} \tag{1}$$

(3) Calculate the difference $h_1(t)$ between the original time series $x(t)$ and $m_1(t)$:

$$h_1(t) = x(t) - m_1(t) \tag{2}$$

(4) If $h_1(t)$ conforms to the conditions of IMFs, it can be treated as the first IMF which consists of the shortest periodic component in the original signal. Otherwise, $h_1(t)$ is regarded as original load series and repeat Step (1) to Step (3) until the difference $h_1^k(t)$ at k-th iteration meets the conditions of IMF.

$$imf_1(t) = h_1^k(t) \tag{3}$$

The standard deviation (SD) is introduced to determine whether the sifting process can be stopped, so as to judge if $h_1^k(t)$ is IMF.

$$SD = \frac{\sum_{t=0}^{T} \left| h_1^{k-1}(t) - h_1^k(t) \right|^2}{\sum_{t=0}^{T} \left| h_1^k(t) \right|^2} \tag{4}$$

where $h_1^{k-1}(t) - h_1^k(t)$ is the mean value of the upper and lower envelopes of $h_1^{k-1}(t)$; SD is between 0.2 and 0.3 [47].

(5) The residue $r_1(t)$ can be obtained when the first IMF $imf_1(t)$ is separated from the original load series $x(t)$:

$$r_1(t) = x(t) - imf_1(t) \tag{5}$$

(6) The residue $r_1(t)$ is taken as the new original load series here. Repeat Step (1) to Step (5) until the amplitude of the residue is smaller than the preset one or the residue is a monotone function or constant. Thus all IMFs and the residue of $x(t)$ can be derived, as shown in Eq. (6):

$$\begin{cases} r_1(t) - imf_2(t) = r_2(t) \\ r_2(t) - imf_3(t) = r_3(t) \\ \vdots \\ r_{N-1}(t) - imf_n(t) = r_n(t) \end{cases} \tag{6}$$

where $r_n(t)$ is a monotone function, the number of modes n rests with the original load series.

(7) The original signal $x(t)$ is reconstructed as described in Eq. (7):

$$x(t) = \sum_{n=1}^{N} imf_n(t) + r_N(t) \tag{7}$$

According to Step (1) to Step (7), the original load series is decomposed into sub-series in diverse frequency, namely IMFs and a residue r. Then, feature correlation analysis is implemented on sub-series to find the influential factors that affect the frequency.

2.2. mRMR

mRMR applies mutual information to measure the dependence between two variables, which not only takes the correlation between characteristics and target variables into account, but also obtains the redundant information among features [53].

2.2.1. Maximum relevance

In view of mRMR, maximum relevance can be expressed as the mean value of mutual information between feature x_i and the target variable y.

$$\max D = \frac{1}{|\mathbf{J}|} \sum_{x_i \in \mathbf{J}} I(x_i, y) \tag{8}$$

where x_i represents the influential factors of the components; y is the component; $\mathbf{J}$ is the feature set of x_i including day types (workdays and weekends, hours, holidays), weather (temperature); $|\mathbf{J}|$ equals the number of features in $\mathbf{J}$; D means the average value of mutual information $I(x_i, y)$ between feature x_i and the target variable y in $\mathbf{J}$.

Mutual information is the intersection of two or multiple random variables. In load forecasting, mutual information method is exploited to capture the linear and nonlinear dependence between the input and the target variables. If they are independent, the value of mutual information equals 0; If not, the value of mutual information corresponds to the positive one of the relationship strength. $I(x_i, y)$ can be calculated as:

$$I(x_i y) = \iint p(x_i, y) \log \frac{p(x_i, y)}{p(x_i) p(y)} dx_i dy \tag{9}$$

where $p(x_i)$ and $p(y)$ represents the marginal probability densities of the random variables x_i and y, partly; $p(x_i, y)$ equals their joint probability density. The value of mutual information $I(x_i, y)$ changes to the positive direction with the correlation between x_i and y. If the two variables are independent with each other, $I(x_i, y)$ equals 0, implying that there exists no interdependence.

2.2.2. Minimum redundancy

The selected features in accordance with maximum relevance may have a certain degree of redundancy, namely overlapping information, which leads to the reduction of accuracy and the increase of calculation. Thus, the redundancy between features needs to be computed.

Minimum redundancy means the minimal dependence between x_i and x_j, as expressed in Eq. (10):

$$\min R = \frac{1}{|\mathbf{J}|^2} \sum_{x_i, x_j \in \mathbf{J}} I(x_i, x_j) \tag{10}$$

Where R means the dependency between each feature.

Eqs. (8) and (10) are integrated with mRMR function, as described in Eq. (11):

$$\max \psi(D, R), \quad \psi = D - R \tag{11}$$

The aim of the mRMR technique is to select the feature that maximizes the relevance and minimize the redundancy simultaneously, thus incremental search can be used here [53]. Suppose the feature set $\mathbf{J}_{n-1}$ consists of $n-1$ features extracted from $\mathbf{F}_m$, the n-th feature selected from $\{\mathbf{F}_m - \mathbf{J}_{n-1}\}$ on the basis of incremental search are shown as:

$$mRMR: \quad \max_{x_j \in \mathbf{F}_m - \mathbf{J}_{n-1}} \left[I(x_j, y) - \frac{1}{n-1} \sum_{x_i \in \mathbf{J}_{n-1}} I(x_j, x_i) \right] \tag{12}$$

According to Eq. (12), the features with the maximum value of mRMR are searched from the remaining characteristics in $\mathbf{F}_m$ successively, which constitute the candidate feature set $\mathbf{J}$.

Due to the different values of mRMR for x_i in $\mathbf{J}$, the input with insufficient information, corresponding to a smaller mRMR, will cause accuracy descent. Hence, it is imperative to select suitable characteristics from $\mathbf{J}$ as the optimal feature set. Additionally, the mean ratio of the absolute error of each component to the actual load is taken as the basis to evaluate the effect of the quantity of input characteristics on the prediction performance. For instance, load forecasting in n previous days can be implemented as:

$$E = \frac{1}{n} \sum_{t=1}^{n} \frac{|imf'(t) - imf(t)|}{y(t)} \times 100\% \tag{13}$$

where $y(t)$ is the actual load at time period t; $imf'(t)$ and $imf(t)$ represents the forecasted and actual value of each component at time period t, respectively. The value of $imf'(t)$ is connected with the prediction approach.

2.3. GRNN

GRNN is a highly parallel radial basis function network based on one-pass algorithm [27]. The diagram of GRNN is illustrated in Fig. 1. The number of neurons in the input layer is equal to the dimension m of the input vector in learning samples. Each neuron is a simple distribution unit which directly transfers the input to the hidden layer [28].

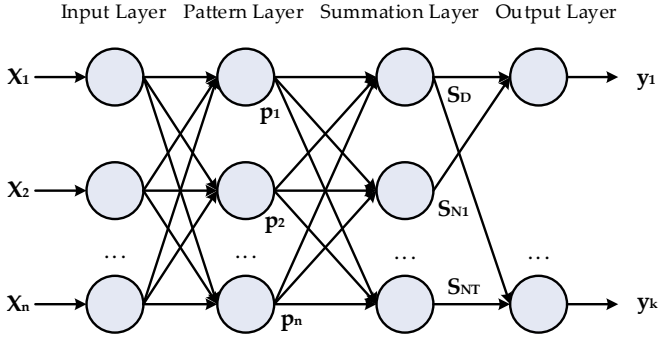

Fig. 1. Topology of GRNN.

The number of neurons in the pattern layer equals the quantity of learning samples n. Each neuron corresponds to a different learning sample. The transfer function of i-th neuron in the pattern layer can be described as:

$$p_i = \exp\left[-\frac{(\mathbf{X} - \mathbf{X}_i)^T (\mathbf{X} - \mathbf{X}_i)}{2\sigma^2} \right], \quad i = 1, 2, \cdots, n \tag{14}$$

where $\mathbf{X}$ and $\mathbf{X}_i$ represent the input and corresponding learning sample of i-th neuron, respectively; σ is the smoothing factor, which pertains to the width coefficient in the Gaussian function.

There are two types of neurons in the summation layer, one of which calculates the arithmetical sum of output in pattern layer. The weights between this neuron and the ones in pattern layer all equal 1. The transfer function is shown in Eq. (15):

$$S_A = \sum_{i=1}^{n} p_i \tag{15}$$

Other neurons calculate the weighted sum of the output in the pattern layer. y_{ij} represents the j-th element in the i-th output Y_i, namely the weight between i-th neuron in the pattern layer and j-th neuron in the summation layer. The transfer function in the summation layer can be expressed as:

$$S_{Nj} = \sum_{i=1}^{n} y_{ij} p_i \quad j = 1, 2, \cdots, k \tag{16}$$

The number of neurons in the output layer is equivalent to the dimension k of the output vector in the learning samples. $\hat{\mathbf{Y}}(\mathbf{X})$ is derived from the division results by the output of two different types in the summation layer, that is:

$$y_j = S_{Nj} / S_A \quad j = 1, 2, \cdots, k \tag{17}$$

Here, $\hat{\mathbf{Y}}(\mathbf{X})$ approximately equals the mean of the whole dependent variables due to large smoothing factor σ. However, if σ tends to 0, $\hat{\mathbf{Y}}(\mathbf{X})$ approaches to the training sample. When the forecasted point belongs to the training sample, the predicted values of the dependent variables will be very close to the corresponding ones in the sample. Otherwise, the poor generalization ability of GRNN will limit its prediction performance. If σ is moderate, all the dependent variables should be taken into account. Thus, the dependent variables corresponding to the predicted point distance are endowed with larger weights. As we can see, the performance of GRNN greatly depends on σ. In this paper, FOA is introduced to determine the optimal value of this parameter.

2.4. FOA

FOA is a novel global optimization technique on the foundation of foraging behaviors [34]. For fruit flies, the excellent nature in smell and vision is conducive to smell collection in the air and correct flight to the food or gathering. The iterative food searching process of fruit fly swarm is displayed in Fig. 2.

The main procedures of FOA are shown as follows:

(1) Initial the position of the fruit fly swarm, namely *InitX_axis* and *InitY_axis*.
(2) Initialization the population. Give the random flight direction and the distance for food finding of an individual fruit fly by using osphresis [34].

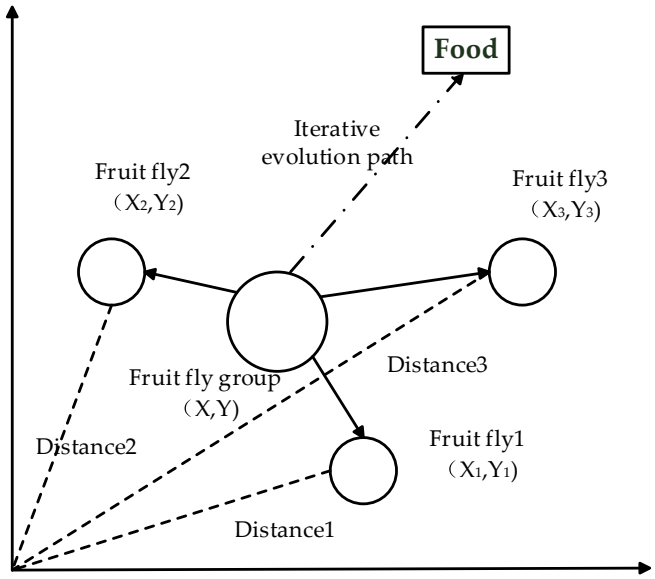

Fig. 2. Iterative food searching process of fruit fly swarm.

$$X_i = InitX\_axis + Random\ Value \tag{18}$$

$$Y_i = InitY\_axis + Random\ Value \tag{19}$$

(3) In virtue of the unknown food location, distance $Dist_i$ between the individual fruit fly and the origin needs to be calculated. Additionally, the reciprocal of $Dist(i)$ is regarded as smell concentration judgment value $S(i)$.

$$Dist_i = \sqrt{X_i^2 + Y_i^2} \tag{20}$$

$$S_i = 1/Dist_i \tag{21}$$

(4) Calculate the smell concentration at the current position as Eq. (22)

$$Smell_i = Function(S_i) \tag{22}$$

(5) The best smell concentration of the fruit fly swarm can be obtained according to Eqs. (23) and (24)

$$bestSmell = \max(Smell_i) \tag{23}$$

$$bestIndex = \text{argmax}(Smell_i) \tag{24}$$

(6) Keep a record of the optimal values including the best smell concentration and the corresponding x, y coordinate. At this time, the fruit flies apply vision to fly towards the food source.

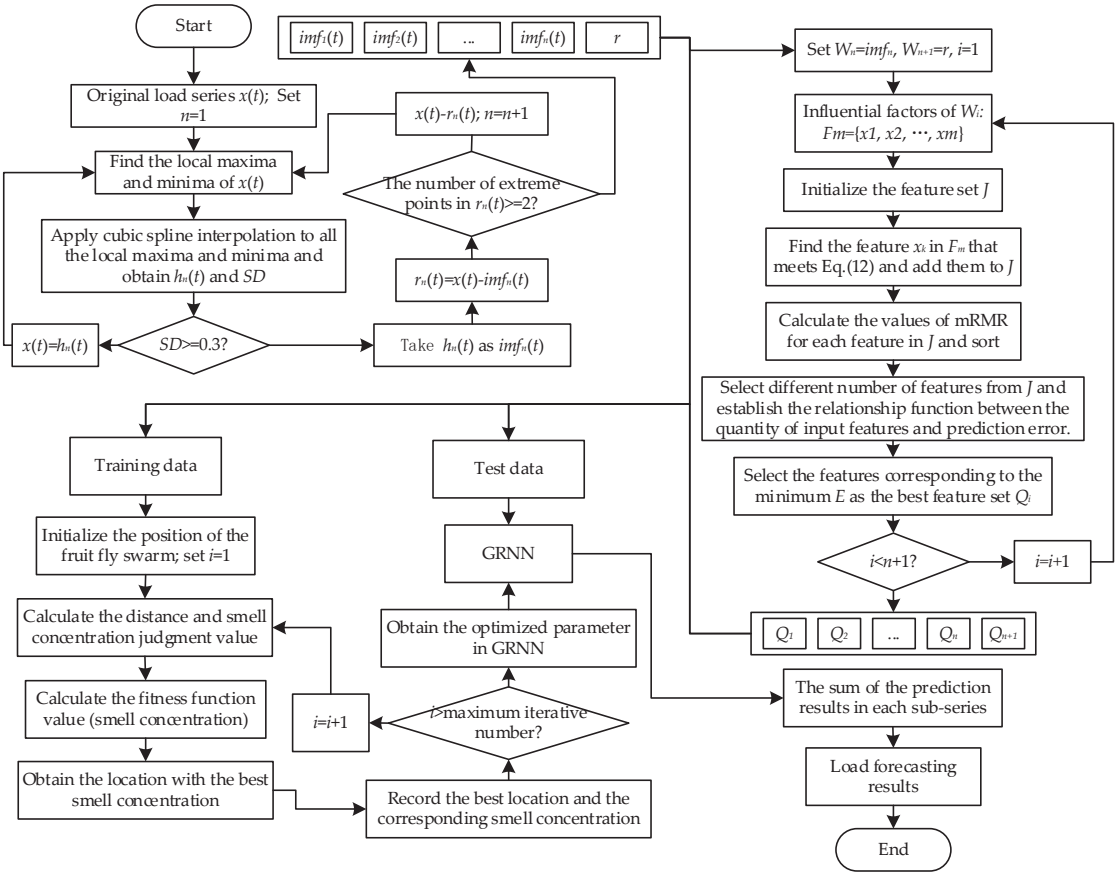

Fig. 3. The flow chart of STLF based on EMD-mRMR-FOA-GRNN.

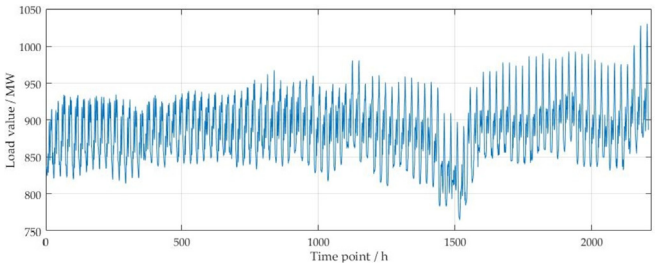

Fig. 4. Original load series.

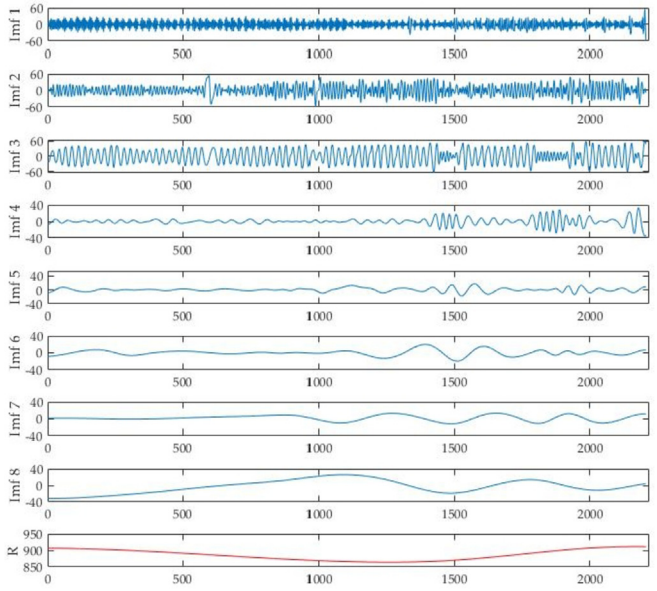

Fig. 5. The EMD results of original load series.

$$Smellbest = bestSmell \tag{25}$$

$$X\_axis = X(bestIndex) \tag{26}$$

$$Y\_axis = Y(bestIndex) \tag{27}$$

(7) Implement iteration optimization. Repeat Step (2) to Step (5) and judge whether the smell concentration is better than the previous one. If it is, go to Step (6).

2.5. STLF model based on EMD-mRMR-FOA-GRNN

The 24 h STLF model combining EMD, mRMR, FOA, and GRNN is constructed as shown in Fig. 3. The original load series is decomposed into a collection of stationary IMFs with different features by EMD. Then mRMR is applied to input selection. The optimal feature set is extracted from the load and its influential factors. GRNN is utilized to predict each sub-series and the forecasted load is a combination. Here, FOA is introduced to optimize the smoothing factor in GRNN so as to improve the forecasting accuracy.

3. Case study

3.1. Load series decomposition with EMD

The load data per hour from August 1, 2017 to October 31, 2017 are collected from the power system in Langfang, China, totally 2208 records. Fig. 4 illustrates the original load series ranges from around 764.69 MW−1030.9 MW with none apparent regularity.

The original load series are decomposed into eight IMFs and one residue, as presented in Fig. 5.

It can be seen that the frequency of imf_1~imf_3 is obviously higher than imf_4~imf_8 and the residue. The prediction is implemented on each component separately and then sum the results up.

3.2. Best feature set extraction by mRMR

The original feature set F_m of each $imf_i(t)$ and $r$can be obtained as listed in Table 1. Incremental search method is employed in this paper to extract characteristics that satisfy Eq. (12) as the candidate feature set. Then calculate the mRMR values of each feature and arrange them in a descending order.

On the foundation of ranking results, the features selected in **J** from the left to the right are input to FOA-GRNN model. The relationship function between the number of input features and errors are established in line with Eq. (13).

As shown in Fig. 6, the best feature numbers of IMFs and the residue are 2, 7, 5, 14, 10, 11, 7, 4 and 6, respectively. The corresponding best input feature sets Q are listed in Table 2.

From Table 1, it can be analyzed the load of imf_1~imf_3 is mainly influenced by historical load, while the load of imf_4~imf_8 and r is closely related to day types, temperature, rainy and snowy days as well as the historical one.

3.3. Load forecasting based on FOA-GRNN

The data from August 8, 2017 to October 30, 2017 are selected as training set and the remaining data on October 31, 2017 are utilized as test set. Based on the determination of the best input features for IMFs and the residue, FOA is employed to optimize the smoothing factor σin GRNN. In FOA, suppose the population size is 20, the maximum iteration number equals 100 and random flight distance ranges in [-10, 10]. The values of smoothing factor in GRNN optimized by FOA are recorded in Table 3.

The optimized smoothing factor is brought into the GRNN for training, and the root mean square error is selected as the training error index. The calculation equation is shown in Eq. (29). The

Table 1
Original feature set.

Feature	Description	Value
PL_{t-n}	Load at the time period t-n	Actual load
TP_t	Temperature at the predicted period t	Actual temperature
HD_t	Whether it is a holiday at the predicted period t	Holidays = 1, otherwise 0
DT_t	Day type at the predicted period t	Monday = 1, Tuesday = 2, and so on
HR_t	Hour at the predicted period t	1 o'clock = 1, 2 o'clock = 2, and so on (according to 24 h a day)
WH_t	Whether there is rain and snow at the predicted period t	Rain and snow days = 1, otherwise 0

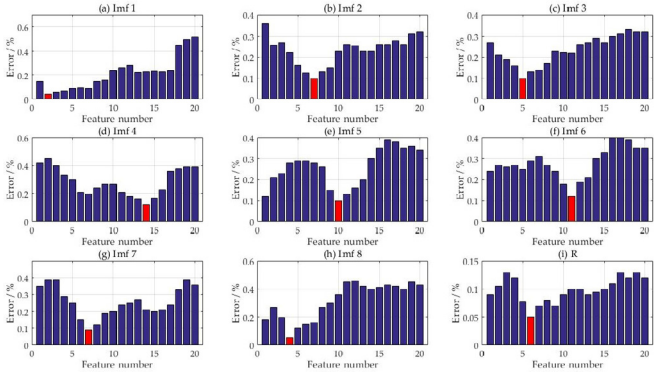

Fig. 6. The relationship between the number of input features and errors.

and residue is low. The IMFs and the residue are superimposed, and the overall training error value is calculated to be 10.2279. The respective components are predicted separately, and the results are shown in Fig. 7. Then, the results of the respective component predictions are superimposed to obtain a final prediction result, and the prediction result is as shown in Fig. 8.

From Fig. 8, it is proved that the STLF model proposed in this paper presents a high fitting degree for the original load series.

3.4. Comparative analysis

In order to verify the performance of the established approach, four other models including mRMR-FOA-GRNN, FOA-GRNN, GRNN and SVM are used for comparison. The parameter settings and input selection of the contrastive models are listed in Table 5.

Table 2
The best input features of IMFs and the residue.

IMFs	The best input feature sets	Number
imf_1	$PL_{t\text{-}96}$, $PL_{t\text{-}128}$	2
imf_2	$PL_{t\text{-}102}$, $PL_{t\text{-}108}$, $PL_{t\text{-}26}$, $PL_{t\text{-}135}$, $PL_{t\text{-}74}$, $PL_{t\text{-}25}$, $PL_{t\text{-}163}$	7
imf_3	$PL_{t\text{-}29}$, $PL_{t\text{-}115}$, $PL_{t\text{-}96}$, $PL_{t\text{-}48}$, $PL_{t\text{-}72}$,	5
imf_4	$PL_{t\text{-}24}$, $PL_{t\text{-}47}$, DT_t, HR_t, TP_t, $PL_{t\text{-}96}$, $PL_{t\text{-}108}$, HD_t, $PL_{t\text{-}144}$, $PL_{t\text{-}25}$, WH_t, $PL_{t\text{-}134}$, $PL_{t\text{-}78}$, $PL_{t\text{-}120}$	14
imf_5	$PL_{t\text{-}48}$, TP_t, $PL_{t\text{-}96}$, HR_t, DT_t, $PL_{t\text{-}114}$, HD_t, $PL_{t\text{-}108}$, $PL_{t\text{-}26}$, $PL_{t\text{-}158}$	10
imf_6	$PL_{t\text{-}24}$, DT_t, $PL_{t\text{-}48}$, TP_t, HR_t, $PL_{t\text{-}96}$, $PL_{t\text{-}168}$, $PL_{t\text{-}86}$, $PL_{t\text{-}74}$, $PL_{t\text{-}58}$, $PL_{t\text{-}166}$	11
imf_7	$PL_{t\text{-}24}$, TP_t, HR_t, $PL_{t\text{-}47}$, $PL_{t\text{-}26}$, $PL_{t\text{-}25}$, WH_t	7
imf_8	$PL_{t\text{-}24}$, TP_t, $PL_{t\text{-}25}$, $PL_{t\text{-}26}$	4
r	$PL_{t\text{-}24}$, DT_t, TP_t, $PL_{t\text{-}25}$, HD_t, WH_t	6

Table 3
The values of smoothing factor for each FOA-GRNN trained by imf_i and r.

IMFs	imf_1	imf_2	imf_3	imf_4	imf_5	imf_6	imf_7	imf_8	r
Smoothing factor	0.0753	0.0435	0.0066	0.0138	0.0128	0.0251	0.0042	0.0087	0.0051

Table 4
Training error statistics for each IMFs and residue.

IMFs	imf_1	imf_2	imf_3	imf_4	imf_5
Error	4.9021	2.9605	7.5617	3.1880	1.0185
IMFs	imf_6	imf_7	imf_8	r	
Error	0.1575	0.0690	0.4044	9.7676	

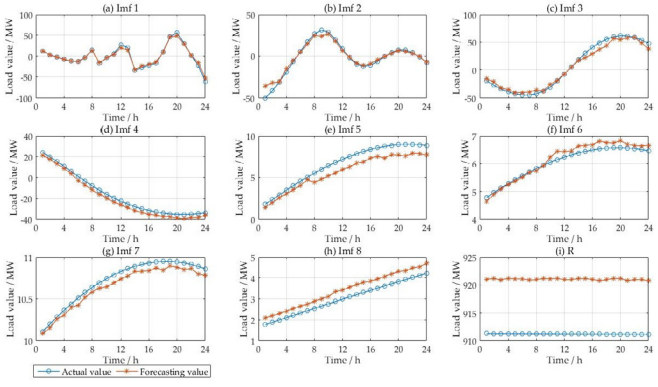

Fig. 7. The prediction results of each IMF and the residue.

training errors of the respective IMFs and residue are shown in Table 4.

It can be seen from Table 4 that the training error of each IMFs

To evaluate which forecasting technique outperforms the other models, five statistical criteria are introduced in this paper, that is relative error (RE), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and Theil's inequality coefficient (TIC). The first four indicators make the errors quantified, and the smaller the values are, the better the forecasting performance is. TIC reflects the predictive capability of the forecasting approaches and it's necessary to keep it as small as possible.

These five error indicators are defined as follows:

$$RE(t) = \frac{\widehat{X}(t) - X(t)}{X(t)} \times 100\% \tag{28}$$

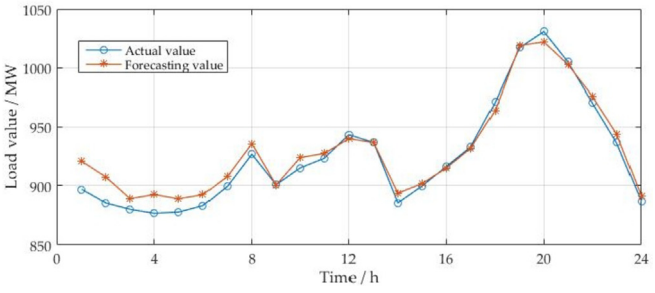

Fig. 8. The final prediction results.

Table 5
Parameter settings and input selection of contrastive models.

Contrastive model	Parameter settings	Input selection
mRMR-FOA-GRNN	The initial population: 20; Maximum number of iterations: 100; Random flight range: [-10, 10]	After mRMR feature extraction: PL_{t-24}, PL_{t-23}, PL_{t-25}, HR_t, PL_{t-1}, PL_{t-2}, PL_{t-48}, DT_t, TP_t, PL_{t-72}, HD_t, WH_t
FOA-GRNN	The initial population: 20; Maximum number of iterations: 100; Random flight range: [-10, 10]	PL_{t-1}, PL_{t-2}, PL_{t-3}, PL_{t-24}, PL_{t-48}, PL_{t-72}, TP_t, HD_t, DT_t, HR_t, WH_t
GRNN	Smoothing factor: 0.05	PL_{t-1}, PL_{t-2}, PL_{t-3}, PL_{t-24}, PL_{t-48}, PL_{t-72}, TP_t, HD_t, DT_t, HR_t, WH_t
SVM	Punish coefficient: 2; kernel parameter: 0.1	PL_{t-1}, PL_{t-2}, PL_{t-3}, PL_{t-24}, PL_{t-48}, PL_{t-72}, TP_t, HD_t, DT_t, HR_t, WH_t

Table 6
Prediction results of load on October 31, 2017 (Unit: MW).

Time/h	Actual data	EMD-mRMR-FOA-GRNN	mRMR-FOA-GRNN	FOA-GRNN	GRNN	SVM
0:00	896.73	920.7653	910.6921	922.3114	927.5110	938.3437
1:00	885.06	907.1785	911.0044	916.8101	923.2213	919.4461
2:00	879.66	888.7714	886.6696	905.7672	913.7961	915.3862
3:00	876.27	892.3184	901.7359	896.2014	904.0564	920.7000
4:00	877.30	888.6179	897.2151	904.6686	918.2835	916.9641
5:00	882.76	892.2087	899.1913	903.6823	916.2875	916.9476
6:00	899.53	907.9637	916.5306	921.6032	932.4783	943.7141
7:00	926.75	935.7983	951.2019	944.6879	962.8595	971.6127
8:00	900.69	900.3874	873.7592	877.5890	860.0119	905.1220
9:00	915.10	923.9240	896.4772	941.5183	955.0073	949.7820
10:00	923.42	927.4278	947.6994	947.8507	960.4056	968.0101
11:00	943.39	940.0241	927.6819	920.6834	908.6150	896.4587
12:00	936.73	936.7086	914.2900	915.3381	901.5665	901.1159
13:00	885.16	893.4164	904.1546	916.6044	912.3943	916.4987
14:00	899.58	901.7147	922.9338	931.2050	937.6752	940.2326
15:00	916.21	914.9140	898.8775	892.0467	886.7614	875.8546
16:00	932.85	931.6636	917.5306	905.7161	899.6218	892.6198
17:00	971.11	964.1098	949.9743	941.1760	933.7591	936.0343
18:00	1017.14	1018.9721	1042.1628	1044.0996	1062.8312	1063.9212
19:00	1030.90	1022.0464	1013.3947	1003.9400	999.6498	978.1489
20:00	1004.88	1002.7658	982.9813	968.8829	961.8059	957.1086
21:00	970.58	976.2053	994.6339	991.8532	1004.6786	1008.5731
22:00	936.68	944.1911	962.3497	969.7707	970.1430	919.2543
23:00	886.61	891.2366	904.0063	906.5342	919.4754	922.6551

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\widehat{X}(t) - X(t)\right| \tag{29}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{X}(t) - X(t)\right)^2} \tag{30}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\widehat{X}(t) - X(t)}{X(t)}\right| \times 100\% \tag{31}$$

$$TIC = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\widehat{X}(t) - X(t)\right)^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}X(t)^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N}\widehat{X}(t)^2}} \tag{32}$$

where $X(t)$ and $\widehat{X}(t)$ are the actual and predicted load at time period t, partly. N represents the number of data points.

The load forecasted values for different models are shown in Table 6.

For more intuitive analysis, the predicted results in Table 5 are presented in Fig. 9. It demonstrates the forecasting results obtained from the proposed technique are most close to the actual load curve with small fluctuation range. From Fig. 10, the values of RE for the

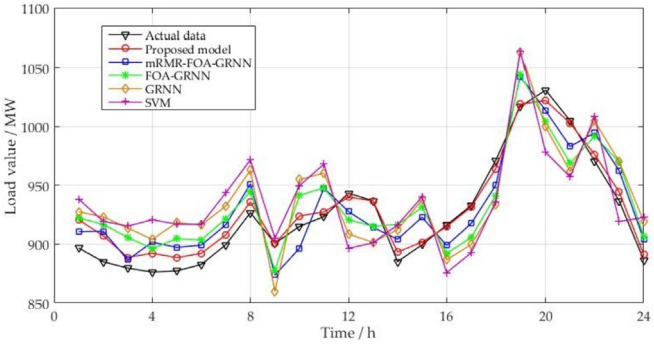

Fig. 9. Load forecasting results (I).

forecasting approaches can be derived. The prediction error range of EMD-mRMR-FOA-GRNN and mRMR-FOA-GRNN are both within [-3%, 3%] while there are only 17 error points of FOA-GRNN in this scope. Among them, 18 error points of EMD-mRMR-FOA-GRNN are controlled in [-1%, 1%] while the corresponding number of mRMR-FOA-GRNN and FOA-GRNN is 1 (at 2:00, RE = −0.8%) and 0, respectively. As we can see, the error points of mRMR-FOA-GRNN, GRNN and SVM mostly ranges in [-2%, 2,%], [-3.5%, 3.5%], [-4%, 4%]. In addition, the minimum RE of EMD-mRMR-FOA-GRNN, mRMR-FOA-GRNN, FOA-GRNN, GRNN and SVM is 0.0023%, −0.8%, −1.94%, 3.03%, −0.49%, respectively, and the maximum RE corresponding

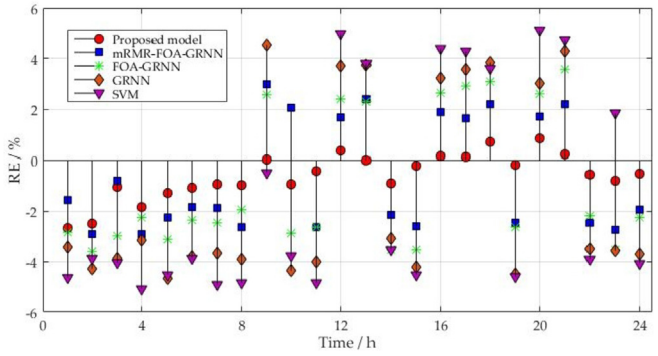

Fig. 10. RE of prediction methods.

equals −2.86%, −2.93%, −3.59%, −4.67% and 5.12%, partly. Hence, the forecasting accuracy from the superior to the inferior can be ranked as: EMD-mRMR-FOA-GRNN, mRMR-FOA-GRNN, FOA-GRNN, GRNN and SVM. Evidently, the decomposition of the original load series with EMD can weaken the volatility of the series influenced by complicated factors. Moreover, the application of mRMR administers to the reduction of input features and efficiency improvement. In contrast with GRNN, FOA improves its forecasting accuracy through parameter optimization. It is without doubt that the prediction precision of some points in the proposed model is inferior to other four techniques, for example, the error of EMD-mRMR-FOA-GRNN is −2.68% at 0:00 that is greater than mRMR-FOA-GRNN.

Fig. 11 displays the comparison results of training set and test set measured by MAE, RMSE, MAPE and TIC. Wherein, PM, m-F-G, F-G, G, S represent EMD-mRMR-FOA-GRNN, mRMR-FOA-GRNN, FOA-GRNN, GRNN and SVM, respectively. It should be noted that the MAE (also for RMSE, MAPE and TIC) of the training set of PM is a comprehensive value of the total sequence after the respective IMFs are trained. It can be seen that the error difference between the training set and the test set of each model is not large, so there is no over-fitting. It confirms that the test set of established model is optimal on the consideration of MAE, RMSE, MAPE and TIC, which equals 7.3550, 9.5823, 0.8093 and 0.0052, distinguishly. Compared with mRMR-FOA-GRNN, more regular input for the prediction model is selected grounded on the decomposition of the load series by EMD. Furthermore, mRMR technology contributes to feature extraction with maximum relevance and minimum redundancy so as to improve the forecasting efficiency. Through the parameter optimization by FOA, the generalization ability and prediction accuracy of GRNN is greatly ameliorated. In contrast with SVM, GRNN can effectively cope with nonlinear and non-stationary series with merely one parameter to be optimized. As a whole, the proposed method EMD-mRMR-FOA-GRNN outperforms other four models.

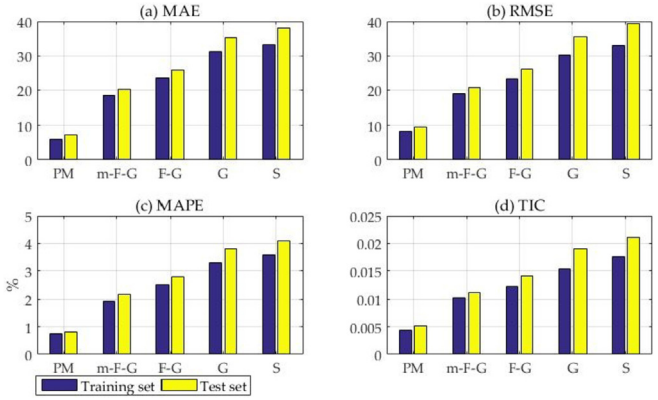

Fig. 11. MAE, RMSE, MAPE and TIC of forecasting models (I).

4. Further study

To further validate whether the proposed model has good adaptability, another case which selects additional load data (from November 10, 2017 to November 16, 2017) in this area is provided in this paper. The above four comparative models are still used here with the same parameter settings. The forecasting results are displayed in Fig. 12. The error criteria are shown in Fig. 13.

As demonstrated in Figs. 12 and 13, the establish model in this paper mostly approximates the actual load curve with the lowest MAE, RMSE, MAPE and TIC. EMD-mRMR-FOA-GRNN is superior to other four methods considering the benefits of EMD, mRMR, FOA and GRNN. Thus, the novel hybrid model EMD-mRMR-FOA-GRNN is capable of the application in STLF with high prediction accuracy.

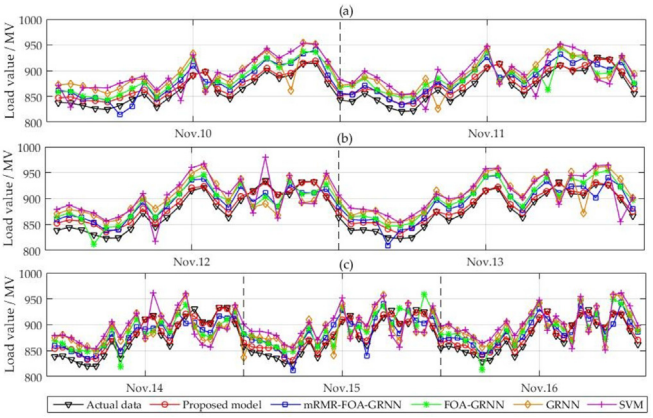

Fig. 12. Load forecasting results (II).

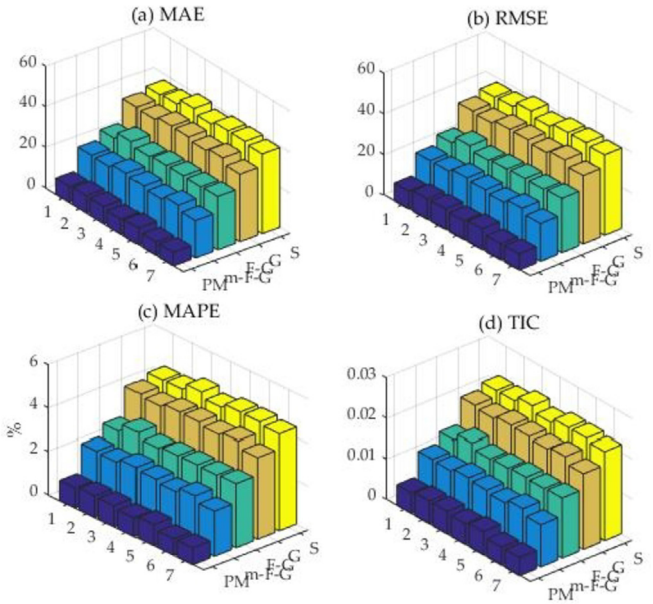

Fig. 13. MAE, RMSE, MAPE and TIC of forecasting models (II).

5. Conclusion

Aiming at the nonlinearity and randomness of power load series, a hybrid model integrating EMD, mRMR, FOA and GRNN is established in this paper. The original load series is firstly decomposed into several IMFs and a residue in order to weaken the volatility of the series affected by complex indicators. Then, mRMR is applied to obtain the best feature set through the correlation analysis between each IMF and the features that contain day types, temperature, meteorology conditions and so on. In the end, FOA is utilized to optimize the smoothing factor in GRNN. The final predicted load can be obtained through the summation of the forecasted results for all IMFs. It has been proved that the proposed model outperforms other four contrast methods (mRMR-FOA-GRNN, FOA-GRNN, GRNN and SVM) with the lowest MAE, RMSE, MAPE and TIC. Thus, the established technique is effective, efficient and practicable in STLF on the power system.

Acknowledgment

This work is supported by the Natural Science Foundation of China (Project No. 71471059 and 7180445). Wei-Chiang Hong thanks the support from Ministry of Science and Technology, Taiwan (MOST 106-2221-E-161-005-MY2).

References

[1] Yang Y, Li S, Li W, et al. Power load probability density forecasting using Gaussian process quantile regression. Appl Energy 2018;213:499–509. https://doi.org/10.1016/j.apenergy.2017.11.035.
[2] Sun W, Liang Y. Research of least squares support vector regression based on differential evolution algorithm in short-term load forecasting model. J Renew Sustain Energy 2014;6(5):1–10. https://doi.org/10.1063/1.4900552.
[3] Sadaei HJ, Guimarães FG, Silva CJD, et al. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. Int J Approx Reason 2017;83(C):196–217. https://doi.org/10.1016/j.ijar.2017.01.006.
[4] Tarsitano A, Amerise IL. Short-term load forecasting using a two-stage sarimax model. Energy 2017;133:108–14. https://doi.org/10.1016/j.energy.2017.05.126.
[5] Taylor JW, Mcsharry PE. Short-term load forecasting methods: an evaluation based on European data. IEEE Trans Power Syst 2007;22(4):2213–9. https://doi.org/10.1109/TPWRS.2007.907583.
[6] Boroojeni KG, Amini MH, Bahrami S, et al. A novel multi-time-scale modeling for electric power demand forecasting: from short-term to medium-term horizon. Elec Power Syst Res 2017;142:58–73. https://doi.org/10.1016/j.epsr.2016.08.031.
[7] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 2017;73:1104–22. https://doi.org/10.1016/j.rser.2017.02.023.
[8] Dudek G. Pattern-based local linear regression models for short-term load forecasting. Elec Power Syst Res 2016;130:139–47. https://doi.org/10.1016/j.epsr.2015.09.001.
[9] Zamo M, Mestre O, Arbogast P, et al. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: probabilistic forecast of daily production. Sol Energy 2014;105:804–16. https://doi.org/10.1016/j.solener.2014.03.026.
[10] Wu J, Wang J, Lu H, et al. Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. Energy Convers Manag 2013;70(70):1–9. https://doi.org/10.1016/j.enconman.2013.02.010.
[11] Asrari A, Javan DS, Monfared M. Application of gray-fuzzy-Markov chain method for day-ahead electric load forecasting. Przeglad Elektrotechniczny 2012;3:228–37.
[12] Li GD, Wang CH, Masuda S, et al. A research on short term load forecasting problem applying improved grey dynamic model. Int J Electr Power Energy Syst 2011;33(4):809–16. https://doi.org/10.1016/j.ijepes.2010.11.005.
[13] Li DC, Chang CJ, Chen CC, et al. Forecasting short-term electricity consumption using the adaptive grey-based approach—an Asian case. Omega 2012;40(6):767–73. https://doi.org/10.1016/j.omega.2011.07.007.
[14] Deb C, Zhang F, Yang J, et al. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24. https://doi.org/10.1016/j.rser.2017.02.085.
[15] Samuel IA, Emmanuel A, Odigwe IA, et al. A comparative study of regression analysis and artificial neural network methods for medium-term load forecasting. Indian J Sci Technol 2017;10(10):7. https://doi.org/10.17485/ijst/2017/v10i10/86243.
[16] Zhao H, Guo S. An optimized grey model for annual power load forecasting. Energy 2016;107:272–86. https://doi.org/10.1016/j.energy.2016.04.009.
[17] Luo J, Hong T, Fang SC. Benchmarking robustness of load forecasting models under data integrity attacks. Int J Forecast 2018;34(1):89–104. https://doi.org/10.1016/j.ijforecast.2017.08.004.
[18] Barrow DK, Crone SF. A comparison of AdaBoost algorithms for time series forecast combination. Int J Forecast 2016;32(4):1103–19. https://doi.org/10.1016/j.ijforecast.2016.01.006.
[19] Ma Y, Sclavounos PD, Cross-Whiter J, et al. Wave forecast and its application to the optimal control of offshore floating wind turbine for load mitigation. Renew Energy 2018;128:163–76. https://doi.org/10.1016/j.renene.2018.05.059.
[20] Moon J, Park J, Hwang E, et al. Forecasting power consumption for higher educational institutions based on machine learning. J Supercomput 2017;(3):1–23. https://doi.org/10.1007/s11227-017-2022-x.
[21] Ren Y, Suganthan PN, Srikanth N, et al. Random vector functional link network for short-term electricity load demand forecasting. Inf Sci 2016;367:1078–93. https://doi.org/10.1016/j.ins.2015.11.039.
[22] Zeng YR, Zeng Y, Choi B, et al. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy 2017;127:381–96. https://doi.org/10.1016/j.energy.2017.03.094.
[23] Gwo-Ching L. Hybrid improved differential evolution and wavelet neural network with load forecasting problem of air conditioning. Int J Electr Power Energy Syst 2014;61(1):673–82. https://doi.org/10.1016/j.ijepes.2014.04.014.
[24] Liang Y, Niu D, Ye M, et al. Short-term load forecasting based on wavelet transform and least squares support vector machine optimized by improved cuckoo search. Energies 2016;9(10):827. https://doi.org/10.3390/en9100827.
[25] Abdoos A, Hemmati M, Abdoos AA. Short term load forecasting using a hybrid intelligent method. Knowl Base Syst 2015;76:139–47. https://doi.org/10.1016/j.knosys.2014.12.008.
[26] Barman M, Choudhury NBD, Sutradhar S. A regional hybrid Goa-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy 2018;145:710–20. https://doi.org/10.1016/j.energy.2017.12.156.
[27] Specht DF. A general regression neural network. IEEE Trans Neural Network 1991;2(6):568–76. https://doi.org/10.1109/72.97934.
[28] Niu D, Liang Y, Hong WC. Wind speed forecasting based on EMD and GRNN optimized by FOA. Energies 2017;10(12):2001. https://doi.org/10.3390/en10122001.
[29] Zhu S, Lian X, Liu H, et al. Daily air quality index forecasting with hybrid models: a case in China. Environ Pollut 2017;231(2):1232–44. https://doi.org/10.1016/j.envpol.2017.08.069.
[30] Debnath KB, Mourshed M. Forecasting methods in energy planning models. Renew Sustain Energy Rev 2018;88:297–325. https://doi.org/10.1016/j.rser.2018.02.002.
[31] Du P, Wang J, Yang W, et al. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renew Energy 2018;122:533–50. https://doi.org/10.1016/j.renene.2018.01.113.
[32] Haidar AMA, Mustafa MW, Ibrahim FAF, et al. Transient stability evaluation of electrical power system using generalized regression neural networks. Appl Soft Comput J 2011;11(4):3558–70. https://doi.org/10.1016/j.asoc.2011.01.028.
[33] Cigizoglu HK. Application of generalized regression neural networks to intermittent flow forecasting and estimation. J Hydrol Eng 2005;10(4):336–41. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:4(336).
[34] Pan WT. A new Fruit Fly Optimization Algorithm: taking the financial distress model as an example. Knowl Base Syst 2012;26(2):69–74. https://doi.org/10.1016/j.knosys.2011.07.001.
[35] Cheng J, Xiong Y. The quality evaluation of classroom teaching based on FOA-GRNN. Procedia Comput Sci 2017;107:355–60. https://doi.org/10.1016/j.procs.2017.03.117.
[36] Iscan H, Gunduz M. An application of fruit fly optimization algorithm for traveling salesman problem. Procedia Comput Sci 2017;111:58–63. https://doi.org/10.1016/j.procs.2017.06.010.
[37] Du TS, Ke XT, Liao JG, et al. DSLC-FOA: an improved fruit fly optimization algorithm application to structural engineering design optimization problems. Appl Math Model 2018;55:314–39. https://doi.org/10.1016/j.apm.2017.08.013.
[38] Li HZ, Guo S, Li CJ, et al. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. Knowl Base Syst 2013;37(2):378–87. https://doi.org/10.1016/j.knosys.2012.08.015.
[39] Niu D, Wang H, Chen H, et al. The general regression neural network based on the fruit fly optimization algorithm and the data inconsistency rate for transmission line icing prediction. Energies 2017;10(12):2066. https://doi.org/10.3390/en10122066.
[40] Zhang Y, Na S, Niu J, et al. The influencing factors, regional difference and temporal variation of industrial technology innovation: evidence with the FOA-GRNN model. Sustainability 2018;10(1):187. https://doi.org/10.3390/su10010187.
[41] Li Y, Bao YQ, Yang B, et al. Modification method to deal with the accumulation effects for summer daily electric load forecasting. Int J Electr Power Energy Syst 2015;73:913–8. https://doi.org/10.1016/j.ijepes.2015.06.027.
[42] Niu D, Wei Y. Short-term power load combinatorial forecast adaptively weighted by FHNN similar-day clustering. Autom Electr Power Syst

2013;37(3):54–7. https://doi.org/10.7500/AEPS201202139.
[43] Nie H, Liu G, Liu X, et al. Hybrid of ARIMA and SVMs for short-term load forecasting. Energy Procedia 2012;16(5):1455–60. https://doi.org/10.1016/j.egypro.2012.01.229.
[44] Niu D, Liang Y, Wang H, et al. Icing forecasting of transmission lines with a modified back propagation neural network-support vector machine-extreme learning machine with kernel (BPNN-SVM-KELM) based on the variance-covariance weight determination method. Energies 2017;10:1196. https://doi.org/10.3390/en10081196.
[45] Li S, Wang P, Goel L. Short-term load forecasting by wavelet transform and evolutionary extreme learning machine. Elec Power Syst Res 2015;122:96–103. https://doi.org/10.1016/j.epsr.2015.01.002.
[46] Fletcher P, Sangwine SJ. The development of the quaternion wavelet transform. Signal Process 2017;136:2–15. https://doi.org/10.1016/j.sigpro.2016.12.025.
[47] Huang NE, Wu MLC, Long SR, et al. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proc Math Phys Eng Sci 2003;459(2037):2317–45. https://doi.org/10.1098/rspa.2003.1123.
[48] Amjady N, Abedinia O. Short term wind power prediction based on improved kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine. Sustainability 2017;9(11):2104. https://doi.org/10.3390/su9112104.
[49] Lahmiri S. Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices. IEEE Syst J 2017;99:1–4. https://doi.org/10.1109/JSYST.2015.2487339.
[50] Li S, Wang P, Goel L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Trans Power Syst 2016;31(3):1788–98. https://doi.org/10.1109/TPWRS.2015.2438322.
[51] Wi YM, Joo SK, Song KB. Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment. IEEE Trans Power Syst 2012;27(2):596–603. https://doi.org/10.1109/TPWRS.2011.2174659.
[52] Kouhi S, Keynia F, Ravadanegh SN. A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection. Int J Electr Power Energy Syst 2014;62(11):862–7. https://doi.org/10.1016/j.ijepes.2014.05.036.
[53] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27(8):1226–38. https://doi.org/10.1109/TPAMI.2005.159.
[54] Wang Q, Guan T, Qin B. Short-term wind speed forecasting of ORELM based on MRMR. Renew Energy Resour 2018;36(01):85–90. https://doi.org/10.13941/j.cnki.21-1469/tk.2018.01.013.
[55] Mallik S, Bhadra T, Maulik U. Identifying epigenetic biomarkers using maximal relevance and minimal redundancy based feature selection for multi-omics data. IEEE Trans NanoBioscience 2017;(99). https://doi.org/10.1109/TNB.2017.2650217. 1-1.
[56] Yang LI, Xueping GU. Feature selection for transient stability assessment based on improved maximal relevance and minimal redundancy criterion. Proc Chin Soc Electr Eng 2013;33(34):179–86. https://doi.org/10.13334/j.0258-8013.pcsee.2013.34.024.