Short Term Load Forecasting Based on SBiGRU and CEEMDAN-SBiGRU Combined Model

TIAN CHEN, WEI HUANG, RUJUN WU, HUABING OUYANG
School of Mechanical Engineering, Shanghai Dianji University, Shanghai 201306, China
Corresponding author: Tian Chen. (e-mail: chent@sdju.edu.cn).
This work was supported in part by the Plateau Disciplines in Shanghai under Grant A157011800703, in part by the Shanghai Natural Science Foundation under Grant 15ZR1417200.

ABSTRACT With the continuous development of global science and technology industry, the demand for power is increasing, so short-term power load forecasting is particularly important. At present, a large number of load forecasting models have been applied to short-term load forecasting, but most of them ignore the error accumulation in the iterative training process. To solve this problem, this paper proposes a combined measurement model which combines stacked bidirectional gated recurrent unit (SBiGRU), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and error correction. In the first stage, SBiGRU model is established to study the time series characteristics of load series under the influence of temperature and holiday types. The error series generated in the prediction process of SBiGRU model reflects the error characteristics of load series; In the second stage, the error sequence is decomposed into several intrinsic mode functions (IMF) components and trend components by CEEMDAN algorithm. The SBiGRU model is established again for each component to learn and predict, and the predicted values of all components are reconstructed to get the error prediction results; Finally, sum the two-stage prediction results to correct the error. The accuracy of SBiGRU-CEEMDAN-SBiGRU combination model is evaluated by two public power load data. The experimental results show that the SBiGRU-CEEMDAN-SBiGRU combination model has better accuracy and stability than the traditional model.

INDEX TERMS short term load forecasting, stacked bidirectional gated recurrent unit, error correction, complete ensemble empirical mode decomposition with adaptive noise.

I. INTRODUCTION
Short term load forecasting is to estimate and forecast the electricity demand in the next 24 hours, days or weeks on the premise of considering the influence of meteorological factors and working day types. Accurate forecasting is of great significance for the dispatching of power system and stable operation of power grid [1].

Common load forecasting algorithms include: extreme learning machine [2], support vector machine [3], Kalman filter [4] and random forest [5]. These methods, through simple internal structure, can rapidly converge the algorithm model, but can not fit the time series relationship between input data and output data, so the prediction accuracy is often not very ideal [6]. In recent years, with the development of deep learning, deep learning algorithms such as convolutional neural network [7]-[9], recurrent neural network (RNN) [10] and deep confidence network [11] have been gradually applied to power load forecasting. Among them, RNN is widely applied in speech recognition and natural language processing. It does not process all the data at one time, but processes the data in a recurrent way. The advantage of the recurrent connection is that it can learn and record the historical information in the sequence, so that it can be very good in the prediction of data Characteristics of fitting data on time series [12]. Based on the time series characteristics of load series, RNN has been introduced into load forecasting by many scholars [13]-[17]. In literature [18], firstly, K-means was used to classify the data sets into the same cluster, and then the EEMD algorithm was used to decompose the load series into relatively stable components, and the BiRNN model was established for each component. At the same time, the network parameters are initialized by deep confidence network. The
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
Where: $\sigma$ represents the offset of hidden layer at time $t$.

**B. STACKED BI-DIRECTIONAL GATED RECURRENT UNIT**

The bi-directional gated recurrent unit (BiGRU) is composed of a forward GRU network and a backward GRU network. Its structure is shown in Fig.2. The BiGRU neural network forecasts based on the whole time series. It divides the hidden layer into two opposite parts, forward and backward, and reads the information of the past and future times respectively. The first level forward GRU calculates the sequence information of the current time series, the second level backward GRU calculates the reverse sequence information of the same time series, and finally the hidden layer state of BiGRU at time $t$ is obtained by weighting the forward hidden layer state and backward hidden layer state. The calculation process is as follows:

$$\tilde{h}_t = \text{GRU}\left(x_t, \tilde{h}_{t-1}\right)$$

$$\tilde{h}_t = \text{GRU}\left(x_t, h_{t-1}\right)$$

$$h_t = W_{\tilde{h}} \tilde{h}_t + V_{\tilde{h}} \tilde{h}_t + b_{\tilde{h}}$$  \hspace{1cm} (7)

Where: $\tilde{h}_t$ represents the forward hidden layer state at time $t$, $\tilde{h}_t$ represents the backward hidden layer state at time $t$, $W_{\tilde{h}}$ and $V_{\tilde{h}}$ represent the weight corresponding to $\tilde{h}_t$ and $\tilde{h}_t$, $b_{\tilde{h}}$ represents the offset of hidden layer at time $t$.

**C. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION WITH ADAPTIVE NOISE**

In the process of EEMD decomposition, CEEMDAN adds adaptive white noise and calculates the unique residual signal to obtain IMF component. It can reconstruct the signal sequence in a few experiments, and eliminate the error caused by adding adaptive white noise in the residual signal produced by EEMD decomposition.

The definition calculation symbol $E_i(t)$ indicates the $i$-th IMF component generated by EMD decomposition. The calculation steps of CEEMDAN algorithm are as follows:

**Step 1** In the $k$ times of experiments, the original signal $x_t + \delta_0 \omega_0$ decomposition, where $\delta_0$ is the standard deviation of white Gaussian noise, $\omega_0$ said Gaussian white noise. The first IMF component is obtained through EMD decomposition and the unique residual signal $r_j(t)$ is obtained:

$$\text{IMF}_j(t) = \frac{1}{k} \sum_{i=0}^{k} \text{IMF}_j^i$$

$$r_j(t) = x(t) - \text{IMF}_j(t)$$  \hspace{1cm} (8)

**Step 2** Continue to acquire the second IMF component:
The proposed algorithm model is verified on two public datasets. The first data set is the power load forecasting....

**Step 1** first stage load forecasting. The training set is used for supervised training of the SBiGRU model in the first stage. After the model converges, the training set and test set are predicted to get the prediction value $P_{b_{train}}$ of training set and the prediction value $P_{b_{test}}$ of test set, and the error sequence $P_{e}$ produced in the training process of SBiGRU is obtained according to the following formula:

$$P_{e} = P_{b_{train}} - T_{train}$$

Where: $T_{train}$ is the real value of the training set.

**Step 2** error prediction in the second stage. The CEEMDAN algorithm is used to decompose $P_{e}$ into several IMF components and one residual component. The SBiGRU model is established for learning and prediction of each component. Finally, the prediction results are accumulated to get the error prediction value $P'_{e}$.

**Step 3** calculates the predicted value of the combined model. The prediction result $P$ of SBiGRU-CEEMDAN-SBiGRU model is the sum of the predicted value of the test set and the predicted value of the error, namely:

$$P = P_{b} + P'_{e}$$

The prediction process of the SBiGRU-CEEMDAN-SBiGRU composite model is shown in Fig.4.
competition data organized by EUNITE. The data set contains historical load, daily average temperature and holiday data, and the load data sampling period is 30 minutes, and the sample number is 17,520. The second data set is the power load data collected in Alberta, Canada in 2016. The sampling period of the data set is 1 h, and the number of samples is 8760.

A. DATA SET PREPROCESSING

(1) Normalization
The MAXMIN method is used to normalize the data, normalize the value range of all data to the range of [0,1], then transfer the normalized data into the model, and finally reverse normalize the output of the model. The calculation formula is as follows:

\[ x'_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

Where: \( x'_i \) and \( x_i \) represent normalized value and data value respectively, \( x_{\text{max}} \) and \( x_{\text{min}} \) represent maximum value and minimum value.

(2) Data quantification
The date type is represented by virtual encoding mode, with \{0, 1, 2\} representing \{working day, rest day, holiday\}.

IV. EXPERIMENT ANALYSIS

A. MODEL TRAINING METHOD
The training process of the model adopts Adam optimization algorithm. Adam combines the advantages of Rmsprop which is good at dealing with sparse gradients, and Adagrad which is good at dealing with non-stationary targets, and AdaGrad which is good at dealing with sparse gradients. The calculation formula is as follows:

\[ m_t = \mu m_{t-1} + (1-\mu) * g_t \]
\[ n_t = \nu n_{t-1} + (1-\nu) * g_t^2 \]
\[ \dot{m}_t = \frac{m_t}{1-\mu^t} \]
\[ \dot{n}_t = \frac{n_t}{1-\nu^t} \]
\[ \Delta \theta_t = -\frac{\dot{m}_t}{\sqrt{\dot{n}_t+\epsilon}} * \eta \]

Where: \( g_t \) is the gradient, \( \epsilon \) is the smoothing index, which is mainly used to prevent the denominator from being zero; \( \mu \) and \( \nu \) are the momentum factors; \( m_t \) and \( n_t \) are the first-order matrix estimates of the gradient, which can be regarded as the estimates of the expected \( E[g_t^2] \) and \( E[g_t^2] \); \( \dot{m}_t \) and \( \dot{n}_t \) are the modifications of \( m_t \) and \( n_t \), so that the expectation can be estimated approximately unbiased.

B. MODEL EVALUATION INDEX
The average absolute percentage error \( E_{\text{MAPE}} \), root mean square error \( E_{\text{RMSE}} \) and maximum absolute percentage error \( E_{\text{Max}} \) are used as evaluation indexes in the model, and the calculation formula is as follows:

\[ E_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x'_i - \hat{x}_i}{x'_i} \right| \times 100\% \]  
\[ E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x'_i - \hat{x}_i)^2} \]  
\[ E_{\text{Max}} = \max(E_{\text{MAPE}}) \]

Where: \( \hat{x}_i \) is the predicted value of neural network at \( i \) time, and the actual value of \( x_i \) neural network at \( i \) time.

C. EXPERIMENT 1
Firstly, the data of EUNITE electric load forecasting competition is used as the test sample. In the experiment, 92d data from August to October were used as training set, 11d data from November 1 to November 11 were used as test set, and 48 historical time points were used to predict the load data of the next 48 time points.

Load data is the result of periodicity, trend and randomness. Seasonal-Trend decomposition procedure based on Loess (STL) decomposition algorithm can be used to decompose the load data from these three characteristics, so as to observe the time series characteristics of load data more intuitively. The experimental results are shown in Fig.5.

FIGURE 5. Time series decomposition of load series
In Fig.5, A, B, C and D are load value, trend component, seasonal component and residual component respectively. As can be seen from Fig.5-B, the overall load value shows an upward trend in the 92-day cycle. Figure 5-C shows the periodicity of the load series with about 48 time points. Fig.5-D shows the distribution of the influence of external random factors on the load time series, and it can be seen that the residual component changes periodically.

In Fig.6, the shaded area represents the 95% confidence interval, the correlation coefficient outside the shadow area indicates significant correlation, the horizontal axis represents the lag order, and the vertical axis represents the correlation coefficient. It can be seen from the figure that the lag orders beyond the confidence area are as follows: 1, 2 and 3, which shows that the load values at historical time t-1, t-2 and t-3 have the highest correlation with the current time t.
FIGURE 6. Autocorrelation graph of load data

The training set of 92d is transformed into a three-dimensional matrix. The dimensional information of the three-dimensional matrix is as follows:

\[
\{D_1 \ldots D_{92}\} \times \{L_1 \ldots L_{48}\} \times \{X_1 \ldots X_3\}
\] (25)

Where: \(\{D_1 \ldots D_{92}\}\) indicates that the data set is 92 days in length; \(\{L_1 \ldots L_{48}\}\) indicates that the data set contains 48 times in a day; \(\{X_1 \ldots X_3\}\) represents the 3-dimensional characteristics of the data set, which represents the load data, average temperature and date type at the current time.

Similarly, the dimension information of the test set is:

\[
\{D_1 \ldots D_{11}\} \times \{L_1 \ldots L_{48}\} \times \{X_1 \ldots X_3\}
\] (26)

1) FIRST STAGE SBIGRU TRAINING AND PREDICTION

The experiment was completed with Pytorch1.10 in Ubuntu 18.04 system. The processor of the experimental PC was Inter Core i5-8300h, the memory was 16GB, and the graphics card was NVIDIA GeForce GTX 1050TI.

In order to design the structure of SBiGRU model scientifically, the test set is used to analyze the SBiGRU model with different layers. Taking single layer SBiGRU, 2+1 layers SBiGRU, 3+1 layers SBiGRU and 4+1 layers SBiGRU as experimental samples, combined with reference [6], the iteration number is set to a larger empirical value of 2000, and the evaluation index is the average absolute percentage error and training time. The experimental results are shown in Tab.1.

| Network structure | \(E_{MAPE}\) | Training time (s) |
|-------------------|-------------|-------------------|
| Single layer      | 0.0405      | 15                |
| 2+1 layer         | 0.0394      | 28                |
| 3+1 layer         | 0.0398      | 37                |
| 4+1 layer         | 0.0412      | 52                |

It can be found from Tab.1 that when the network layer is a single layer structure, the model has weak generalization ability and low prediction accuracy. When the number of layers is 2+1, because its multi-layer network structure improves the generalization ability, the network structure model at this time has the best prediction accuracy. However, with the increase of the network structure, when the model structure is 3+1 layers and 4+1 layers, the accuracy of the model decreases with the increase of the number of network layers. At this time, it is considered that the too complex network structure makes the model over fit.

According to Tab.1, the structure of the first stage SBiGRU model is determined as 2 + 1 layers, which is composed of two layers of BiGRU and one layer of full connection layer. The model structure is shown in Fig.7. Taking the input of the first BiGRU layer as an example, \(\{91,24,3\}\) = \{the length of the input data set is 91 days, 24 time points contained in each day, feature dimension\}.

![Network structure diagram of SBiGRU in the first stage](image)

FIGURE 7. Network structure diagram of SBiGRU in the first stage

The training mode of the model is to forecast the load data of 48 time points in the future through the load data, date type and temperature information of 48 time points in the history. Therefore, the data from August 1 to October 30 is taken as the input of the network, the load data from August 2 to October 31 is taken as the output of the network, and the Adam algorithm is used to optimize the network parameters. In order to obtain the error samples generated in the training process, after the SBiGRU model converges, first input the training set into the trained model, and obtain the prediction result \(P_{\text{train}}\) of the training set, as shown in Fig.8. Then, according to equation (14), the error sequence \(P_e\) generated by the SBiGRU model in the prediction process in the first stage can be obtained, as shown in Fig.9.
2) TRAINING PROCESS OF CEEMDAN-SBiGRU MODEL IN THE SECOND STAGE

The STL decomposition algorithm is used to analyze the error series. The experimental results are shown in Fig.10. It can be seen from the figure that the periodicity of residual components in time series is weak, which increases the difficulty of error series fitting.

FIGURE 10. Time series decomposition of error series

Fig.11 is the autocorrelation diagram of error series. It can be seen from the figure that although the periodicity of the error sequence is weak, there are still historical errors significantly related to the current time point. Therefore, historical errors can be used to predict future errors. In order to improve the prediction accuracy of the error series, the CEEMDAN algorithm is used to decompose the error series into nine IMF components and one residual component. The test results are shown in Fig.12.

FIGURE 12. CEEMDAN decomposition results

It can be seen from Fig.12 that the frequency of IMF1–IMF5 components is high and the periodicity is not obvious, which can be regarded as the high frequency component of error sequence. The periodicity of IMF6–IMF8 component is obvious, which can be regarded as the periodic

FIGURE 9. Error sequence diagram($P_e$)
component of error sequence. The frequency of IMF9 component is low and its periodicity is not obvious, so it can be regarded as the low frequency component of error sequence. IMF10 is the residual component of the error sequence.

The second stage SBiGRU model is established by decomposing eight IMF components and one residual component from CEEMDAN algorithm. The structure design process is similar to the SBiGRU structure in the first stage, and the network structure is determined as 2+1 layers by experiment. However, since the input of the network is error component and its characteristic dimension is 1, the dimension information of SBiGRU input layer in the second stage is \{91, 24, 1\}. At this time, the input of each SBiGRU model is set as the error component from August 1 to October 29, and the output is the error component from August 2 to October 30, and the network parameters are optimized iteratively by Adam algorithm.

3) PREDICTION COMPONENT RECONSTRUCTION
The predicted values of each IMF component \(c_i(t)\) and residual component \(r_n(t)\) are accumulated to obtain the predicted values of errors, as shown in formula (27), where \(p(t)\) is the reconstructed error data.

\[
p(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)
\]  

4) EXPERIMENTAL PROCESS AND RESULT ANALYSIS
In order to display the prediction results intuitively and conveniently, the load data of 96 time points of November 3-4 are predicted through the data of 96 time points of November 2-3.

After the SBiGRU model in the first stage converged with the CEEMDAN-SBiGRU model in the second stage, the first stage SBiGRU model is used to predict the main characteristics \(P_b\) of the load sequence from November 3-4, then the CEEMDAN-SBiGRU model is used to predict the error characteristics \(P_e'\) of November 3-4, and finally the final prediction results are obtained according to equation (15). At the same time, LSTM, SBiGRU, SBiGRU-SBiGRU and SVR, the traditional load forecasting method, are selected as comparative tests, as shown in Fig.13. SBiGRU-SBiGRU is a two-stage forecasting model with SBiGRU as error correction model.

It can be seen from Fig.13 that compared with SVR, LSTM, SBiGRU and SBiGRU-SBiGRU, the combined model proposed in this paper has better performance and can capture the complex nonlinear changes of load series more effectively, so it can make short-term forecast of load series more accurately and stably.

The data of the remaining dates in the test set are predicted by the same method. Fig.14 shows the correlation distribution between the actual and predicted values of the four models in the forecast set. It is seen from Fig.14 that compared with SVR, LSTM, SBiGRU and SBiGRU-SBiGRU, the proposed model in this paper has better performance and can capture the complex nonlinear changes of load series more effectively, so it can make short-term forecast of load series more accurately and stably.

The data of the remaining dates in the test set are predicted by the same method. Fig.14 shows the correlation distribution between the actual and predicted values of the four models in the forecast set. R is Pearson correlation coefficient, which is used to evaluate the correlation between the real value and the predicted value.

**FIGURE 13.** Forecast curve of November 3-4

**FIGURE 14.** Correlation distribution of different models

Fig.14 shows the correlation between the predicted results and the real results. The abscissa is the predicted value and the ordinate is the real value. It can be seen from the figure that, compared with other models, the correlation coefficient between the predicted value and the real value of SBiGRU-CEEMDAN-SBiGRU is closer to 1, so it performs best in the prediction set.

Fig.15 shows the error histograms of the five models. If the error histograms are more concentrated near 0, the performance of the models will be better. It can be seen from the figure that the error histograms of the SBiGRU-CEEMDAN-SBiGRU model proposed in this paper is the most concentrated near 0, which shows that the performance of this combined forecasting model is the best among the five models.
Compare the three indexes in formula (24) - (26), as shown in Tab.2.

| Model                        | $E_{MAPE}$ | $E_{Max}$ | $E_{MAD}$ |
|------------------------------|------------|-----------|-----------|
| SBiGRU-CEEMDAN-SBiGRU       | 0.0275     | 0.0319    | 26.19     |
| SBiGRU- SBiGRU              | 0.0314     | 0.0366    | 30.27     |
| SBiGRU                      | 0.0394     | 0.0506    | 34.40     |
| LSTM                        | 0.0415     | 0.0492    | 36.28     |
| SVR                         | 0.0506     | 0.0705    | 41.91     |

It can be seen from Tab.2. that SVR model has the worst performance in the prediction set, and an average daily accuracy rate in the test set is 94.94%; LSTM records the historical information of load series in the process of forecasting, and the prediction accuracy is 95.85%; The SBiGRU model takes into account the data information of past and future times in the prediction process, and uses the network structure of 2+1 layer to improve the data generalization ability of the model. Therefore, the prediction accuracy of the model is higher than that of the SVR model, and the average accuracy rate of the test set is 96.06%; SBiGRU-SBiGRU model uses SBiGRU as the main feature of load forecasting in the first stage, and SBiGRU is used again as the error correction model in the second stage, so the forecasting accuracy is improved, and the daily average accuracy is 96.86%; in the second stage, SBiGRU-CEEMDAN-SBiGRU model uses CEEMDAN algorithm to decompose the error sequence, which can predict the error sequence on different time scales Therefore, the average daily accuracy is the highest among the four models, which is 97.25%.

Tab.3 shows the training time consumption and prediction time consumption of the five models. The combined forecasting model divides the forecasting process into two stages: load forecasting and error forecasting. Therefore, the consumption of training time and forecasting time is greater than that of single stage model. Especially for the SBiGRU-CEEMDAN-SBiGRU model, because the model needs to decompose the error sequence, and needs to train each error component, the training time consumption and prediction time consumption are far greater than other models. The combined forecasting model sacrifices the time cost of calculation, thus improving the prediction accuracy.

### TABLE III

| Model                        | Training time(s) | Forecast time(s) |
|------------------------------|------------------|------------------|
| SBiGRU-CEEMDAN-SBiGRU       | 146.58           | 8.14             |
| SBiGRU- SBiGRU              | 62.67            | 3.28             |
| SBiGRU                      | 28.24            | 1.22             |
| LSTM                        | 12.08            | 0.97             |
| SVR                         | 0.94             | 0.32             |

C. EXPERIMENT 2

In order to further verify the validity of the model, this paper continues to use the power load data of Alberta in 2016 to evaluate the accuracy of the model.
Firstly, the load data set is analyzed by time series, and the load series is decomposed by STL. The experimental results are shown in Fig.17.

![Autocorrelation graph of load data](image)

**FIGURE 17.** Autocorrelation graph of load data

In Fig.17, A, B, C and D are the load value, trend component, seasonal component and residual component respectively. As can be seen from Fig.17-B, the overall load value increases first and then decreases in the 92-day cycle, that is, the load data presents the development trend of first rising and then decreasing. Fig.17-C shows the periodicity of load series, which takes about 24 time points as the period. Fig.17-D shows the distribution of the influence of external random factors on load and time series.

Then, the autocorrelation of data sets is analyzed experimentally. The experimental results are shown in Fig.18.

![Autocorrelation graph of load data](image)

**FIGURE 18.** Autocorrelation graph of load data

It can be seen from Fig.18 that the load value at the current time point is mainly related to the load value at three historical time points. For the load forecasting model, the load forecasting model uses at least the load data beyond these three historical time points to forecast the future load data.

92d data from July to September were selected as training set, and 11d data from October 1 to October 11 were selected as test set.

Dimension information of training set is:

\[
\{D_1\cdots D_{92}\} \times \{L_1\cdots L_{24}\} \times \{X_1, X_2\}
\]

(28)

Where: \{D_1\cdots D_{92}\} indicates that the length of the data set is 92 days; \{L_1\cdots L_{24}\} indicates that the data set contains 24 times in a day; \{X_1, X_2\} represents the two-dimensional characteristics of the data set. In this case, it represents the load data and date type at the current time.

Similarly, the dimension information of the test set is:

\[
\{D_1\cdots D_{11}\} \times \{L_1\cdots L_{24}\} \times \{X_1, X_2\}
\]

(29)

1) EXPERIMENTAL PROCESS AND RESULT ANALYSIS

In order to determine the network structure, single-layer BiGRU, 2+1 layer SBiGRU, 3+1 layer SBiGRU and 4+1 layer SBiGRU are taken as experimental samples, and the average absolute percentage error and training time are used as evaluation indexes. The experimental results are shown in Tab.4.

| Network structure | \(E_{MAPE}\) | Training time (s) |
|-------------------|-------------|-------------------|
| Single layer      | 0.0204      | 13                |
| 2+1 layer         | 0.0174      | 21                |
| 3+1 layer         | 0.0186      | 29                |
| 4+1 layer         | 0.0193      | 38                |

According to Tab.3, the structure of the first stage BiGRU model is determined as 2+1 layers, which is composed of two layers of BiGRU and one layer of full connection layer. The model structure is shown in Fig.19.

![Network structure diagram of SBiGRU in the first stage](image)

**FIGURE 19.** Network structure diagram of SBiGRU in the first stage

The training mode of the model is as follows: the load data, date type and temperature information of 24 historical time points are used to predict the load data of the future 24 time points. Therefore, the data from July 1 to September 29 are used as the input of the network, and the load data from July 2 to September 30 are taken as the output of the network, and the network parameters are optimized by Adam algorithm.

In order to obtain the error samples generated in the training process, after the SBiGRU model converges, first input the training set into the trained model to obtain the prediction result \(P_{\text{train}}\), of the training set, as shown in Fig.20. Then, according to equation (14), the error sequence \(P_e\) generated in the prediction process of the first stage SBiGRU model can be obtained, as shown in Fig.21.
FIGURE 20. Prediction results of SBiGRU model ($P_{\text{train}}$)

FIGURE 21. Error sequence diagram ($P_e$)

2) TRAINING PROCESS OF THE SECOND STAGE CEEMDAN-SBIGRU MODEL

The STL decomposition algorithm is used to decompose the time series of the error sequence. The experimental results are shown in Fig. 22. It can be seen from the figure that the residual components still show irregular changes.

FIGURE 22. Autocorrelation graph of error data

Fig. 23 is the autocorrelation diagram of the error series. It can be seen from the figure that the error series still has correlation in the time series, so the error series is predictable, in order to improve the accuracy of error prediction, The CEEMDAN algorithm is used to decompose the error sequence. The experimental results are shown in Fig. 24.

FIGURE 23. Autocorrelation graph of error sequence

FIGURE 24. CEEMDAN decomposition results
In Fig.24, IMF1 ~ IMF4 are high frequency components. IMF5 ~ IMF7 are periodic components. IMF8 is a low frequency component. IMF9 is the residual component.

The SBiGRU model was established for each component and supervised training was conducted. The input of each SBiGRU model is set as the error component from July 1 to September 29, and the output from July 2 to September 30. The network parameters are optimized iteratively by Adam algorithm.

3) EXPERIMENTAL PROCESS AND RESULT ANALYSIS

Through the load data of 216 time points from October 2 to 10, the load data of 216 time points from October 3 to 11 are predicted. The experimental results are shown in Fig.25, and the correlation distribution between the actual value and the predicted value is shown in Fig.26.

It can be seen from Fig.25 and Fig.26 that the SBiGRU-CEEMDAN-SBiGRU model can fit the load data more accurately than other models in Experiment 2, and the correlation coefficient between the predicted value and the real value is closer to 1, so the model performs best in Experiment 2.

Fig.27 is the error histogram. It can be seen from the figure that the SBiGRU-CEEMDAN-SBiGRU model is more concentrated near 0 than other models, so it has the best prediction performance.

Fig.28 shows the absolute error box-plot of the five models. Model-1, model-2, model-3, model-4 and model-5 represent SVR, LSTM, SBiGRU, SBiGRU-SBiGRU and SBiGRU-CEEMDAN-SBiGRU respectively. It can be seen from the figure that the SBiGRU-CEEMDAN-SBiGRU model has the smallest box range and the least abnormal points, which further shows that the model has the optimal prediction performance.
The three indexes in formula (24) ~ (26) are compared, as shown in Tab.5.

| Model                        | $E_{MAPE}$ | $E_{Max}$ | $E_{RMS}$ |
|-----------------------------|------------|-----------|-----------|
| SBiGRU-CEEMDAN-SBiGRU       | 0.0105     | 0.0144    | 121.26    |
| SBiGRU-SBiGRU               | 0.0142     | 0.0298    | 172.16    |
| SBiGRU                      | 0.0174     | 0.0250    | 192.15    |
| LSTM                        | 0.0218     | 0.0321    | 267.13    |
| SVR                         | 0.0266     | 0.0508    | 315.00    |

It can be seen from Tab.5 that the SVR model performs the worst with an average daily accuracy of 97.34%; LSTM records the historical information of load series in the process of forecasting, and the prediction accuracy is 97.82%; the SBiGRU model takes into account the data information of past and future times in the prediction process, and uses the 2 + 1 layer network structure to improve the generalization ability of the model, with an accuracy rate of 98.26%; SBiGRU-SBiGRU uses SBiGRU as the error correction model, so the prediction accuracy is improved, and the daily average accuracy rate is 98.58%; The SBiGRU-CEEMDAN-SBiGRU model uses CEEMDAN algorithm to decompose the error series and can predict the error series on different time scales. Therefore, the daily average accuracy rate of the four models is the highest, which is 98.95%.

It can be seen from Tab.5 that the SVR model performs the worst with an average daily accuracy of 97.34%; LSTM records the historical information of load series in the process of forecasting, and the prediction accuracy is 97.82%; the SBiGRU model takes into account the data information of past and future times in the prediction process, and uses the 2 + 1 layer network structure to improve the generalization ability of the model, with an accuracy rate of 98.26%; SBiGRU-SBiGRU uses SBiGRU as the error correction model, so the prediction accuracy is improved, and the daily average accuracy rate is 98.58%; The SBiGRU-CEEMDAN-SBiGRU model uses CEEMDAN algorithm to decompose the error series and can predict the error series on different time scales. Therefore, the daily average accuracy rate of the four models is the highest, which is 98.95%.

TABLE VI

| Model                        | Training time(s) | Forecast time(s) |
|-----------------------------|------------------|-----------------|
| SBiGRU-CEEMDAN-SBiGRU       | 134.58           | 7.95            |
| SBiGRU-SBiGRU               | 58.29            | 3.07            |
| SBiGRU                      | 26.24            | 1.16            |
| LSTM                        | 11.53            | 0.84            |
| SVR                         | 0.88             | 0.28            |

Tab.6 shows the time calculation cost of different models. It can be seen from the table that the training and prediction time of SBiGRU-CEEMDAN-SBiGRU model is much longer than that of other models, indicating that the model improves the prediction accuracy at the expense of calculation cost.

V. CONCLUSION

This paper introduces the combined model of SBiGRU and CEEMDAN-SBiGRU. In the first stage of the combined model, a 2+1 layers SBiGRU model is designed. In the second stage, CEEMDAN-SBiGRU is proposed as the error correction model. Finally, the prediction results of the two stages are added together to correct the error. According to the results of Experiment 1 and Experiment 2, the following conclusions can be drawn:

1. The three evaluation indexes of SBiGRU-CEEMDAN-SBiGRU and SBiGRU-SBiGRU are all better than the other two single prediction models, indicating that the two-stage prediction model has better prediction accuracy than the single-stage prediction model.

2. The combined prediction model of SBiGRU and CEEMDAN-SBiGRU decomposes the error series into several more stable components, and establishes the SBiGRU prediction model for each component, which ultimately improves the accuracy of error prediction in the more stable components.

REFERENCES

[1] Chen Jianqiang, Yang Junjie, Lou Zhibin. Study on a new short-term load forecasting model based on xgboost algorithm [J]. Electrical Measurement & Instrumentation, 2019,56 (21): 23-29.

[2] Gong Gangjun, An Xiaonan, Chen Zhihmin, et al. Load forecasting model of electric vehicle charging station based on sae-elm [J]. Modern Electric Power, 2019,36 (6): 9-15.

[3] Xie min, Deng Jialiang, Jixiang, et al. Cooling load forecasting method of support vector machine based on information entropy and variable precision rough set optimization [J]. Power System Technology, 2017,41 (1): 217-221.

[4] Zhao Hongshan, Tian Tian. Power system dynamic state estimation based on adaptive traceless Kalman filter [J]. Power System Technology, 2014, (1): 3790-3794.

[5] Juan Huo, Tingting Shi and Jing Chang. "Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 1077-1080.

[6] P. Lv, S. Liu, W. Yu, S. Zheng and J. Lv. EGA-STLF: A Hybrid Short-Term Load Forecasting Model[J]. IEEE Access, vol. 8, pp. 31742-31752, 2020.

[7] Lu Jixiang, Zhang Qipei, Yang Zhihong, et al. Short-term load forecasting method based on CNN-LSTM Hybrid Neural Network Model [J]. Automation of Electric Power Systems, 2019, 43(8):131-137.

[8] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu and Z. Zhu, "Multi-Scale Convolutional Neural Network With Time-Cognition for Multi-Step Short-Term Load Forecasting," in IEEE Access, vol. 7, pp. 88058-88071, 2019.

[9] B. Khan, R. Khalid, M. U. Javed, S. Javaid, S. Ahmed and N. Javaid, "Short-Term Load and Price Forecasting based on Improved Convolutional Neural Network," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (CoMET), Sukkur, Pakistan, 2020, pp. 1-6.

[10] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu and J. He. Short-Term Load Forecasting With Deep Residual Networks[J]. IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 3943-3952, July 2019.

[11] M. R. Haq and Z. Ni. A New Hybrid Model for Short-Term Electricity Load Forecasting [J]. IEEE Access, vol. 7, pp. 125413-125423, 2019.

[12] H. Shi, M. Xu and R. Li. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN [J]. IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5271-5280, Sept. 2018.

[13] L. Sehovac and K. Grolinger. Deep Learning for Load Forecasting: A Novel Pooling Deep RNN [J]. IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5271-5280, Sept. 2018.

[14] J. K. Mandal, A. K. Sinha and G. Parthasarathy, "Application of
recurrent neural network for short term load forecasting in electric power system," Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, 1995, pp. 2694-2698 vol.5.

[15] F. Huang, S. Zhuang and Z. Yu, "Power Load Prediction Based On An Improved Clock-Work RNN," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, A. Tokgöz and G. Ünal, "A RNN based time series approach for forecasting turkish electricity load," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[16] A. Tokgöz and G. Ünal, "A RNN based time series approach for forecasting turkish electricity load," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[17] H. Wilms, M. Cupelli, A. Monti and T. Gross, "Exploiting Spatio-Temporal Dependencies for RNN-based Wind Power Forecasts," 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 2019, pp. 921-926.

[18] X. Tang, Y. Dai, Q. Liu, X. Dang and J. Xu. Application of Bidirectional Recurrent Neural Network Combined With Deep Belief Network in Short-Term Load Forecasting [J]. IEEE Access, vol. 7, pp. 160660-160670, 2019.

[19] R. Jiao, T. Zhang, Y. Jiang and H. He. Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network [J]. IEEE Access, vol. 6, pp. 59438-59448, 2018.

[20] 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 [J]. in IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, Jan. 2019.

[21] X. Tang, Y. Dai, T. Wang and Y. Chen. Short-term power load forecasting based on multi-layer bidirectional recurrent neural network [J]. IET Generation, Transmission & Distribution, vol. 13, no. 17, pp. 3847-3854, 3 9 2019.

[22] G. Xiuyun, W. Ying, G. Yang, S. Chengzhi, X. Wen and Y. Yimiao. Short-term Load Forecasting Model of GRU Network Based on Deep Learning Framework[J]. 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (E2), Beijing, 2018, pp. 1-4.

[23] M. Dong and L. Grumbach. A Hybrid Distribution Feeder Long-Term Load Forecasting Method Based on Sequence Prediction. in IEEE Transactions on Smart Grid, vol. 11, no. 1, pp. 470-482, Jan. 2020.

[24] L. Schovac, C. Nesen and K. Grolinger. Forecasting Building Energy Consumption with Deep Learning: A Sequence to Sequence Approach. 2019 IEEE International Congress on Internet of Things (ICIOT), Milan, Italy, 2019, pp. 108-116.

[25] M. Afrasiabi, M. Mohammadi, M. Rastegar and A. Kargarani. Probabilistic deep neural network price forecasting based on residential load and wind speed predictions. IET Renewable Power Generation, vol. 13, no. 11, pp. 1840-1848, 19 8 2019.

[26] X. Tang, Y. Dai, T. Wang and Y. Chen. Short-term power load forecasting based on multi-layer bidirectional recurrent neural network. IET Generation, Transmission & Distribution, vol. 13, no. 17, pp. 3847-3854, 3 9 2019.

[27] P. Dong et al. Short-term Load Forecasting Based on EEMD-approximate Entropy and ELM. 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 2019, pp. 1772-1775.

[28] X. Zheng, H. Xiong and D. Wei. "Short-term Load Forecasting of BP Network Based on EMD," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 1093-1096.

[29] Z. Zhu, Y. Sun and H. Li. "Hybrid of EMD and SVMs for Short-Term Load Forecasting," 2007 IEEE International Conference on Control and Automation, Guangzhou, 2007, pp. 1044-1047.

[30] J. Ye and L. Yang. Short-Term Load Forecasting Using Ensemble Empirical Mode Decomposition and Harmony Search Optimized Support Vector Regression. 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 2019, pp. 851-855.

[31] X. Zhu, H. Qi, X. Huang and S. Sun. A hybrid method for short-term load forecasting in power system. Proceedings of the 10th World Congress on Intelligent Control and Automation, Beijing, 2012, pp. 696-699.

[32] P. Li, H. Liu, M. Li, Q. He, H. Chao and G. Zhang. Power load Forecasting based on VMD Multifrequency Combinations. 2019 IEEE 3rd Advanced Information Management, Communications, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2019, pp. 1214-1218.

[33] L. Tang, S. Wang and L. Yu. EEMD-LSVR-Based Decomposition-and-Ensemble Methodology with Application to Nuclear Energy Consumption Forecasting. 2011 Fourth International Joint Conference on Computational Sciences and Optimization, Yunnan, 2011, pp. 589-593.
TIAN CHEN received the Ph.D. degree from the School of Manufacturing Science and Engineering, Sichuan University (SCU), China, in 2001. She was an Advanced Visiting Scholar with the Department of Mechanical, Materials, & Aerospace Engineering, Illinois Institute of Technology (IIT), Chicago. She is currently a Professor with the School of Mechanical Engineering, Shanghai Dianji University, China. Her current research interests include digital design, intelligent manufacturing, and digital inspection.

WEI HUANG received the B.E. degree in electrical engineering and automation from the Engineering & Technical College of Chengdu University of Technology, Leshan, China, in 2018. He is currently pursuing the master’s degree with the School of Mechanical Engineering, Shanghai Dianji University. His research interests mainly include the application of deep learning in power system detection and control.

RUJUN WU received the Ph.D. degree from School of Aerospace Engineering and Applied Mechanics of Tongji University, China, in 2015. He is currently an associate professor with the School of Mechanical Engineer, Shanghai Dianji University, China. His current research interests include optical fiber sensing technology, numerical simulation, and artificial intelligence.

HUABING OUYANG received the Ph.D. degree from the School of Mechanical Engineering, Tongji University, China, in 2012. He was a Visiting Scholar with the Department of Mechanical Engineering, Boston University, Boston. He is currently an Associate Professor with the School of Mechanical Engineering, Shanghai Dianji University, China. His current research interests include digital design and intelligent manufacturing, topology optimization, and robot technology.