ISSN: 2688-8645

Electricity Price Forecasting Based on BiGRU-Attention Model

Zongyuan An

College of Environmental Science and Engineering, North China Electric Power University, Beijing 102206, China

Abstract

With the continuous advancement of electricity market reform, the electricity spot market has been gradually improved, and electricity price forecasting has become increasingly important in power system operation and market decision-making. Due to the strong nonlinearity, volatility, and temporal dependence of electricity price series, it remains difficult to achieve high forecasting accuracy using conventional methods. To address this issue, this paper proposes an electricity price forecasting model based on a Bidirectional Gated Recurrent Unit (BiGRU) and an Attention mechanism. The electricity price data from a certain province in China are selected as the research object. First, time-dimensional features and temporal features are extracted from the original electricity price series, and logarithmic transformation and normalization are adopted to improve data stability and model training effectiveness. Then, the Attention mechanism is introduced to enhance the model's ability to focus on key temporal information. Finally, the proposed model is compared with XGBoost and GRU models through forecasting experiments. The results show that the BiGRU-Attention model achieves the best performance among the compared models, with a coefficient of determination (R^2) of 0.94, a root mean square error (RMSE) of 9.54, and a mean absolute error (MAE) of 7.61. The experimental results demonstrate that the proposed model can effectively improve electricity price forecasting accuracy in the spot market and has good practical application value.

Keywords

Electricity Price Forecasting; BiGRU; Attention Mechanism; Deep Learning.

1. Introduction

With the acceleration of the global energy transition and the continuous deepening of electricity market reform, the electricity spot market has gradually become an important platform for optimizing the allocation of power resources, promoting renewable energy consumption, and guiding rational competition among market participants. As a core component of market-oriented electricity reform, the electricity spot market plays an essential role in reflecting the real-time balance between supply and demand and discovering market-oriented electricity prices. Compared with traditional planned dispatching mechanisms, the electricity spot market places higher requirements on the responsiveness, accuracy, and flexibility of market decisions.

Electricity prices in the spot market usually exhibit a series of complex characteristics, including strong volatility, randomness, periodicity, and sharp peak fluctuations. Such characteristics make electricity price forecasting much more complicated than forecasting under relatively stable or linear conditions. Accurate electricity price forecasting is of great significance for market participants to optimize bidding strategies, reduce transaction risks, and improve economic returns. At the same time, it is also an important basis for power grid enterprises to achieve safe, stable, and economic dispatching. Therefore, constructing a high-accuracy electricity price forecasting model is of great practical significance for improving market efficiency, reducing market risks, and promoting the optimal allocation of resources.

ISSN: 2688-8645

At present, a large number of studies have been carried out in the field of electricity price forecasting both at home and abroad. Traditional forecasting methods mainly include time series models and regression analysis methods. Although these methods are relatively simple and interpretable, they usually have limited ability in capturing the nonlinear characteristics of electricity price series. With the rapid development of artificial intelligence technology, machine learning and deep learning methods have been increasingly applied to electricity price forecasting tasks. Reference [1] proposed a short-term electricity price forecasting algorithm based on the Long Short-Term Memory (LSTM) neural network. In this method, an electricity price correlation factor matrix was first constructed, and then the LSTM model was employed for forecasting. Reference [2] proposed a forecasting model named BRIM based on bidirectional long short-term memory. Since interconnected electricity exchange markets often exhibit common trends and can provide signal information to each other, the study showed that considering the influence of neighboring markets is beneficial to improving electricity price forecasting accuracy.

In addition, some researchers attempted to combine meteorological data, load forecasting results, and market behavior characteristics to establish multi-dimensional forecasting frameworks, thereby further improving prediction performance. These studies have provided important support for market participants in the electricity spot market. However, in practical applications, many challenges still remain, such as data quality, model adaptability, and computational efficiency. Reference [3] investigated a deep learning-based forecasting model for electricity spot market prices in Spain. By using large-scale market data including demand forecasts, wind power generation, solar generation expectations, and natural gas prices, the study found that the convolutional neural network model achieved the best performance, with a relative mean absolute error of 13.29%, which verified the effectiveness of this model in short-term price forecasting. Reference [4] employed ARIMA and several versions of GARCH models to analyze electricity prices in European markets, and compared the estimated parameters, outlier occurrence rates, and rolling out-of-sample performance across different countries. Reference [5] applied a dynamic time scanning forecasting method to the Brazilian electricity spot market. By searching historical patterns similar to the latest observations and aggregating them for forecasting, this method achieved higher accuracy than traditional statistical models. Reference [6] proposed a hybrid approach based on the combination of parametric and non-parametric models, namely the k-factor GARMA-LLWNN model, and the results indicated that the model had relatively high forecasting accuracy.

In summary, although considerable progress has been achieved in electricity price forecasting, some limitations still exist. On the one hand, some models still have room for improvement in learning electricity price sequence characteristics and temporal dependence relationships. On the other hand, some models fail to sufficiently emphasize key temporal information during feature representation, which affects forecasting accuracy to a certain extent.

To address the above issues, this paper proposes an electricity price forecasting model that combines BiGRU with an Attention mechanism, aiming to improve the representation ability of key features and enhance forecasting performance.

2. Model Construction

(1) GRU Model

The Gated Recurrent Unit (GRU) is a typical recurrent neural network structure. By introducing gating mechanisms, it effectively alleviates the problems of gradient vanishing and gradient explosion that often occur in traditional recurrent neural networks during long-sequence modeling. Compared with the LSTM network, the GRU structure is simpler, contains fewer

ISSN: 2688-8645

parameters, and has higher computational efficiency. Therefore, it has good application value in time series forecasting tasks.

GRU dynamically regulates historical information through an update gate and a reset gate. Its calculation process is as follows.

The update gate is defined as

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

The update gate controls the degree to which the hidden state at the previous moment influences the current moment. When the value of the update gate is large, the model tends to preserve more historical information; otherwise, it relies more on the current input information.

The reset gate is defined as

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

The reset gate is used to regulate the degree to which historical information participates in the current computation, thereby achieving selective forgetting of historical information.

The candidate hidden state is calculated as

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

This equation indicates that after the historical information is filtered by the reset gate, it is combined with the current input to generate the candidate hidden state, which enhances the model's ability to characterize nonlinear relationships.

The final hidden state is expressed as

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

This equation realizes the weighted fusion of historical information and current information, thereby obtaining the hidden state at the current moment.

In the above formulas, x_t denotes the input vector at time t; h_{t-1} and h_t denote the hidden states at the previous and current time steps, respectively; z_t and r_t denote the update gate and reset gate, respectively; $\tilde{h}_t$ denotes the candidate hidden state; W_z, W_r, and W_h are input weight matrices; U_z, U_r, and U_h are hidden-state weight matrices; b_z, b_r, and b_h are bias vectors; $\sigma(\cdot)$ denotes the Sigmoid activation function; $\tanh(\cdot)$ denotes the hyperbolic tangent activation function; and $\odot$ denotes element-wise multiplication.

(2) BiGRU Model

The traditional GRU model only uses one-directional information in a time series and therefore cannot fully reflect the dependency relationships before and after a given moment. However, electricity price series are influenced not only by historical information but also by the changing trends of neighboring moments. Therefore, BiGRU structure is introduced to enhance the modeling ability of temporal dependencies.

BiGRU models the sequence in both forward and backward directions. The forward calculation is

and the backward calculation is

$$\overrightarrow{h_t} = GRU(x_t, \overrightarrow{h_{t-1}})$$

and the backward calculation is

$$\overleftarrow{h_t} = GRU(x_t, \overleftarrow{h_{t+1}})$$

The final output is obtained by concatenating the hidden states from the two directions:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$

This structure can simultaneously use past and future contextual information in the sequence, thereby obtaining more comprehensive feature representations and improving the model's ability to capture complex temporal dependencies.

In the above equations, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ denote the forward and backward hidden states, respectively; h_t denotes the concatenated output vector; $[\cdot;\cdot]$ denotes vector concatenation; and $GRU(\cdot)$ denotes the GRU unit operation.

(3) Attention Mechanism

In time series forecasting, different time steps often contribute differently to the prediction result. Traditional neural networks usually assign equal importance to all time steps and thus may fail to emphasize key temporal information. To address this problem, an Attention mechanism is introduced to weight features from different time steps.

First, the importance score of each time step is calculated as

$$e_t = v^T \tanh(W_a h_t + b_a)$$

This score reflects the importance of the current time step to the final prediction result.

Then, the Softmax function is used to normalize the scores:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}$$

The normalized weight α_t represents the relative importance of each time step.

Finally, the context vector is obtained through weighted summation:

$$c = \sum_{t=1}^{T} \alpha_t h_t$$

This context vector integrates the important information from different time steps and thereby enhances the model's ability to represent key features.

In the above formulas, h_t denotes the output of the BiGRU at time t; e_t denotes the attention score; α_t denotes the attention weight; c denotes the context vector; W_a is the weight matrix; b_a is the bias vector; v is the weight vector; T denotes the sequence length; and $\exp(\cdot)$ denotes the exponential function.

(4) BiGRU-Attention Model Structure

Based on the above components, a BiGRU-Attention electricity price forecasting model is constructed in this paper. The model first uses the BiGRU to perform bidirectional feature extraction on the input sequence so as to obtain complete temporal dependency information. Then, the Attention mechanism assigns different weights to the features at different time steps in order to highlight the key temporal information. Finally, the weighted feature vector is fed into a fully connected layer to obtain the final prediction result.

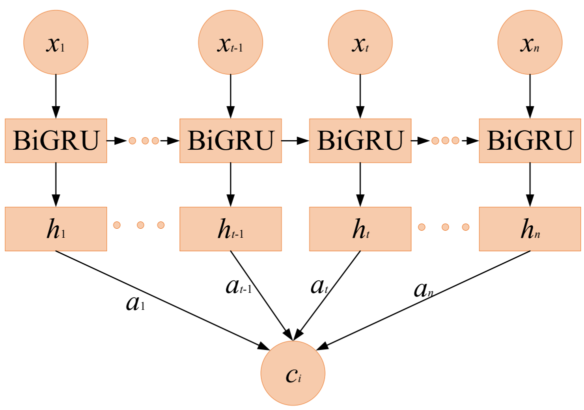

Fig 1. Operating principle of the BiGRU-Attention model

The overall calculation process of the model can be expressed as

$$H = BiGRU(X)$$
$$c = Attention(H)$$
$$y = W_o c + b_o$$

where X denotes the input sequence; H denotes the hidden state sequence output by the BiGRU; c denotes the feature vector after attention weighting; y denotes the final prediction result; W_o denotes the output-layer weight matrix; and b_o denotes the output-layer bias vector.

On the basis of fully utilizing temporal sequence information, this model further strengthens the representation of key features through the Attention mechanism, thereby effectively improving electricity price forecasting accuracy. The overall structure of the model is shown in Fig. 1.

3. Data and Feature Engineering

The electricity price data from a certain province in China are selected as the research object in this paper. The time resolution of the data is 15 minutes, which means that there are 96 sampling points in one day. Since the electricity price series has obvious volatility and nonlinear characteristics, it is necessary to carry out feature construction and data preprocessing before model training in order to improve the forecasting accuracy of the model.

(1) Time Feature Construction

Electricity price series usually exhibit obvious periodic characteristics, such as intra-day periodicity and intra-week periodicity. In order to enable the model to learn variation patterns at different time scales, time-dimensional features are extracted, including hour information and day-of-week information, and these features are represented numerically.

The time feature vector can be expressed as

$$t = [hour_t, day_t]$$

where $hour_t$ denotes the hour information corresponding to time t, and day_t denotes the day-of-week information corresponding to time t.

By introducing time features, the model can better capture the cyclical variation patterns of electricity prices, thereby improving the forecasting accuracy. This step is particularly important for electricity price forecasting because market prices are usually affected by regular changes in power demand, generation schedules, and trading behavior during different hours and days.

(2) Temporal Feature Construction

Electricity price series have strong temporal dependence, and the electricity price at the current moment is often closely related to historical prices. In order to characterize this dependence, a sliding window method is introduced to construct temporal features.

Assume that the electricity price sequence is

$$P = \{p_1, p_2, \dots, p_t\}$$

A sliding window of length $n$is constructed as the input:

$$X_t = [p_{t-n}, p_{t-n+1}, \dots, p_{t-1}]$$

and the corresponding prediction target is

$$y_t = p_t$$

where p_t represents the electricity price at time t, X_t represents the historical electricity price sequence used for prediction, and n denotes the window length.

Through the construction of sliding-window temporal features, the model can effectively make use of historical information and better characterize the changing trend of electricity prices.

This method is widely used in time series forecasting because it transforms the original sequence into supervised learning samples and allows the model to learn the mapping relationship between historical observations and future values.

(3) Data Preprocessing

Since the original electricity price data may have large fluctuations and uneven distribution, directly using them for model training may affect forecasting performance and training stability. Therefore, data preprocessing is required.

First, logarithmic transformation is performed on the electricity price data:

$$p_t' = \log(p_t + 1)$$

where p_t denotes the original electricity price and p_t' denotes the transformed electricity price.

The purpose of logarithmic transformation is to reduce the fluctuation range of the original data, weaken the influence of extreme values, and improve the smoothness of the sequence to a certain extent. This is beneficial for enhancing the stability of subsequent model training.

Then, normalization is carried out:

$$x_t = \frac{p_t' - p_{\min}}{p_{\max} - p_{\min}}$$

where x_t denotes the normalized value, and $p_{\min}$ and $p_{\max}$ represent the minimum and maximum values in the sample, respectively.

Normalization maps the transformed data into a unified interval, which can improve convergence speed, reduce the influence of scale differences among features, and help the deep learning model achieve better training results.

(4) Dataset Splitting

The processed data are divided into a training set and a testing set according to chronological order for model training and performance evaluation. Assume that the dataset is

$$D = \{X, y\}$$

where X denotes the input feature set and y denotes the corresponding prediction target.

The chronological splitting strategy is adopted to avoid future information leakage and to ensure the rationality and reliability of model evaluation. In time series forecasting tasks, this splitting method is more consistent with real forecasting scenarios than random splitting, because the model should always use historical data to predict future observations.

4. Results and Analysis

To verify the effectiveness of the proposed BiGRU-Attention model in electricity price forecasting, the XGBoost model and the GRU model are selected as comparison models. Forecasting experiments are carried out on the electricity price data from a certain province in China, and the results are analyzed from two aspects, namely the prediction curves and evaluation metrics.

(1) Prediction Comparison

The electricity price forecasting results of different models are shown in Fig. 2. From the overall trend, all models can reflect the changing tendency of electricity prices to a certain extent, but there are obvious differences in periods with severe fluctuations.

It can be observed from the figure that the XGBoost model has a relatively good fitting ability in intervals where the electricity price fluctuates smoothly, but it tends to show a lagging phenomenon when the electricity price changes rapidly, making it difficult to accurately capture abrupt changes. Compared with the XGBoost model, the GRU model has certain advantages in temporal sequence modeling, and its forecasting results are closer to the real values in terms of the overall trend. However, there are still certain deviations at local peaks and valleys.

By contrast, the BiGRU-Attention model performs better in fitting both the overall trend and local fluctuations. Especially in the intervals where the electricity price changes sharply, the proposed model can track the real-value changes more accurately. This indicates that the introduction of bidirectional temporal modeling and the Attention mechanism effectively improves the model’s ability to capture key temporal information, thereby enhancing the forecasting effect.

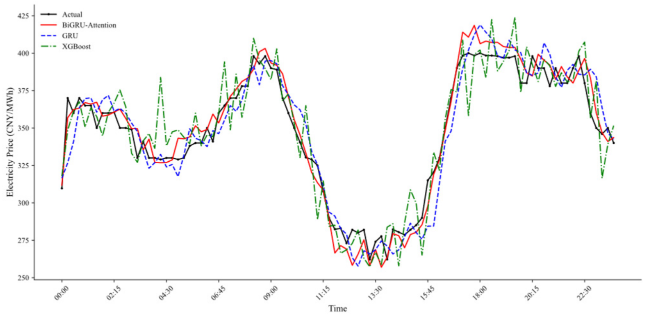

Fig 2. Comparison of forecasting results of different models

(2) Evaluation Metric Analysis

To quantitatively evaluate the forecasting performance of different models, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) are adopted as evaluation metrics. Their calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | y_t - \hat{y}_t |$$

here y_t denotes the true value, $\hat{y}_t$ denotes the predicted value, $\bar{y}$ denotes the mean of the true values, and n denotes the number of samples.

The evaluation results of different models are shown in Table 1.

Table 1. Comparison of evaluation metrics for different electricity price forecasting models

Forecasting Model	R^2	RMSE	MAE
XGBoost	0.84	15.98	12.21
GRU	0.89	13.25	10.33
BiGRU-Attention	0.94	9.54	7.61

It can be seen from the table that the BiGRU-Attention model outperforms the comparison models in all evaluation metrics. Specifically, its coefficient of determination R^2reaches 0.94, indicating that the model has a strong explanatory ability for electricity price variations. At the same time, its RMSE and MAE are 9.54 and 7.61, respectively, both lower than those of the other models, which means that the prediction error is smaller.

Compared with the GRU model, the BiGRU-Attention model further improves the representation ability of temporal sequence features through the introduction of the bidirectional structure and the Attention mechanism, thereby achieving a noticeable improvement in forecasting accuracy. Compared with the XGBoost model, the deep learning-based model exhibits stronger advantages in handling complex nonlinear relationships, and therefore performs better in electricity price forecasting tasks.

Overall, the BiGRU-Attention model demonstrates higher accuracy and stability in electricity price forecasting, which verifies the effectiveness of the proposed method.

5. Conclusion

This paper takes the electricity price data from a certain province in China as the research object and proposes an electricity price forecasting model based on the combination of BiGRU and the Attention mechanism for the problem of strong nonlinearity and high volatility in electricity price series. By introducing the bidirectional GRU structure, the model effectively enhances its ability to characterize forward and backward temporal dependencies. Meanwhile, the Attention mechanism assigns different weights to different time steps, thereby improving the model's ability to represent important features.

The experimental results show that, compared with the XGBoost model and the GRU model, the BiGRU-Attention model performs better in terms of the coefficient of determination, root mean square error, and mean absolute error. This indicates that the proposed model has relatively high accuracy and stability in electricity price forecasting tasks.

Although the proposed model achieves an improvement in forecasting performance, some limitations still exist. For example, the current study mainly constructs the forecasting model based on historical electricity price sequences and does not incorporate external influencing factors such as load and meteorological information. In future research, more comprehensive forecasting models can be developed by integrating multi-source data in order to improve the applicability and prediction capability of the model. In addition, more advanced network structures or optimization algorithms may also be explored to further enhance forecasting accuracy.

References

[1] Gao B, Yan P, Liu X, et al. Short-term Electricity Price Forecast and Analysis Based on LSTM in Spot Electricity Market [J]. Journal of Electrotechnology, Electrical Engineering and Management, 2022, 5(2).

[2] Chen Y, Wang Y, Ma J, et al. BRIM: An Accurate Electricity Spot Price Prediction Scheme Based on Bidirectional Recurrent Neural Network and Integrated Market [J]. Energies, 2019, 12(12): 2241.

[3] Failing J, Tamarit S J, Miota C J, et al. Deep learning-based prediction models for spot electricity market prices in the Spanish market [J]. Mathematics and Computers in Simulation, 2026, 240: 96-104.

[4] Shadi T, Jesús J, Eduardo C. Electricity Spot Price Modeling and Forecasting in European Markets [J]. Energies, 2022, 15(16): 5980.

[5] Gontijo S T, Costa A M, Santis D B R. Electricity price forecasting on electricity spot market: a case study based on the Brazilian Difference Settlement Price [J]. E3S Web of Conferences, 2021, 239: 00002.

[6] Amor B S, Boubaker H, Belkacem L. Forecasting electricity spot price for Nord Pool market with a hybrid k-factor GARMA-LLWNN model [J]. Journal of Forecasting, 2018, 37(8): 832-851.