Research Article

Short-Term Electric Load Prediction and Early Warning in Industrial Parks Based on Neural Network

Guannan Wang, Pei Yang, and Jiayi Chen

State Grid Xiongan Financial Technology Group Co., Ltd., Beijing, China

Correspondence should be addressed to Jiayi Chen; chenjiayi@sgec.sgcc.com.cn

Received 29 June 2021; Accepted 4 August 2021; Published 10 September 2021

Academic Editor: Daqing Gong

Copyright © 2021 Guannan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a load forecasting method based on LSTM model, fully explores the regularity of historical load data of industrial park enterprises, inputs the data features into LSTM units for feature extraction, and applies the attention-based model for load forecasting. The experiments show that the accuracy of our prediction model and early warning model is better than that of the baseline and can reach the standard of application in practice; this model can also be used for early warning of local sudden large loads and identification of enterprise power demand. Therefore, the validity of the method proposed in this paper is verified using the historical dataset of industrial parks, and relevant technical products and business models are formed to provide value-added services to users by combining existing practical cases for the specific scenario of industrial parks.

1. Introduction

Power load forecasting is an important problem in power field. Accurate load forecasting of power system is the basis of efficient management, which provides support for the operation and scheduling of power enterprise [1]. With the development of power market, accurate short-term forecasting of power load can effectively guarantee the safe operation of power grid, reduce the cost of power generation, meet the needs of users, and improve social and economic benefits [2]. However, daily power consumption is affected by many factors, which makes it difficult to accurately predict power consumption [3].

Short-term power load forecasting is to forecast the power load in a short period of time in the future according to the power load in the past and load related data such as temperature, humidity, and date type. Accurate prediction is not only conducive to timely macro control of users' electricity consumption behavior but also to provide scientific guidance for power production [4]. Due to the increasing complexity of power data, the amount of data increases exponentially. Load forecasting based on intelligent forecasting algorithm is widely used in the field of power load

forecasting because of its high stability and forecasting accuracy, strong complex mapping, fault tolerance, and generalization ability [5].

The power system load has a periodic characteristic, and the influencing factors are complex (weather, economy, holidays, observation error, etc.). Therefore, the power system load presents strong randomness and aperiodic components, which brings about great difficulty for short-term forecasting. Currently, short-term power load forecasting methods are mainly divided into three categories: traditional statistics-based models, traditional machine learning-based models, and deep learning-based models [6].

The traditional statistical prediction model can capture time series features, but its nonlinear mapping ability is limited and the generalization ability of unknown data is weak [7]. The prediction model based on machine learning has strong nonlinear mapping ability and generalization ability. However, the model loses part of the time series information, and the prediction accuracy still has room to improve [8]. With the continuous development of deep learning and the strengthening of computer computing power, short-term power load forecasting based on deep learning has gradually become the state of the art. Among

them, recurrent neural network (RNN) has a good performance in dealing with sequence problems, which can learn the change law of sequence according to historical input information [9]. However, when the input sequence becomes long, the gradient will disappear, which results in insufficient forecasting performance. To address this issue, the variants of RNN, such as long short-term memory (LSTM) and gated recurrent unit (GRU), are proposed and utilized in power load forecasting field. Therefore, RNN-based models are the most effective methods currently [10].

Motivated by the aforementioned backgrounds, this paper proposes an LSTM model based on attention mechanism to predict short-term load of power system. On the basis of fully mining the regularity of historical load data, the features such as weather, weekend, and holiday are introduced, and the data feature vectors are input into attention-based LSTM units for feature extraction. The attention mechanism can extract the most relevant information from the features. The proposed method can make full use of the regularity of historical load data and consider the influence of different factors to improve its forecasting efficiency and accuracy. Our findings can provide decision support for the power sector and stakeholders. The model can also be utilized to conduct early warning of local sudden heavy load and identify the power demand of enterprises.

The structure of the article is as follows. Section 2 reviews the related works of power load forecasting; Section 3 shows the methodologies, including the LSTM and attention mechanism; Section 4 provides the experimental setups and experimental results; Section 5 makes a conclusion of this work.

2. Related Works

2.1. Statistical Model. Among the prediction methods based on the traditional mathematical statistical model, the multiple linear regression prediction method has the advantages of simple model, simple calculation, and fast prediction speed. However, the robustness of the model is poor, so it is difficult to obtain high prediction accuracy for complex nonlinear system, and the ability of load forecasting for strong randomness is poor [11]. Time series prediction methods can make accurate prediction for time series with high stationarity and strong periodicity, but it does not consider the weather, holidays, and other factors that affect the load. When the load is affected by complex factors, showing strong randomness or nonstationary characteristics, time series method is difficult to obtain effective prediction results. In the traditional mathematical statistical model, in order to reduce the negative impact of high-frequency components on the prediction results, the load is often filtered [12].

In addition, Hodge et al. [13] made a statistical analysis on the power load forecasting errors of CAISO and NYISO systems in the United States. The results show that the distribution of the forecasting errors has the characteristics of partial normal and nonzero deviation and point out that the magnitude of the errors is related to the load time period.

Song et al. [14] proposed a discrimination method based on sequential trajectory tracking, which discriminates the prediction results of load forecasting model and gives the specific error compensation formula according to the trajectory deviation. Copula theory was used to fit the conditional distribution of prediction error of multiple wind farms, so as to compensate the prediction error of current wind power load according to the prediction error of adjacent wind farms [15]. The simulation method is also the main method for making predictions [16, 17].

2.2. Machine Learning Model. Machine learning models, such as support vector machine (SVM), extreme learning machine (ELM), and random forest (RF), are widely used and gradually improved.

In recent years, due to the fact that the SVR has strong generalization ability and user-defined kernel function, researchers regard it as a research hotspot of power load forecasting. SVR was used to predict the short-term power load demand of office building, which had high accuracy and stability. For instance, Fan et al. [18] presented a hybrid SVM model based on differential-based empirical mode decomposition and autoregression. However, the prediction accuracy of SVR model is greatly affected by the selection of input features, kernel function, and optimization algorithm which demanded complex preexperiments. Wu et al. [19] developed a new model that combined ensemble empirical mode decomposition, an extreme learning machine (ELM), and grasshopper optimization algorithm. Jnr et al. [20] combined a discrete wavelet transform, particle swarm optimization, and radial basis function neural network for load forecasting. Li et al. [21] utilized ensemble empirical mode decomposition, multivariable linear regression, and long short-term memory neural network algorithms for the same purpose. Claude [22] used a hidden Markov model to predict the tourism demands by Google trends.

2.3. Deep Learning Model. Neural network is widely used in nonlinear system prediction with its strong multimapping ability and has achieved better prediction accuracy. The data mining of time series is realized by inputting the state value of the neurons in the last time into the current neurons. Based on RNN, GRU neural network solves the defects of RNN gradient explosion and disappearance by adding the influence degree of the time before the gate structure control, which makes the GRU neural network better deal with and mine the time series than RNN [23].

With its powerful multivariate mapping ability, neural network is widely used in the prediction of nonlinear systems and has achieved good prediction accuracy. Recurrent neural network (RNN) realizes the mining of time series data by inputting the state value of the neuron at the last moment into the current neuron [24]. On the basis of RNN, GRU controls the degree of influence at the previous moment by adding a gate structure, which solves the defects of RNN's gradient explosion and disappearance, so that GRU neural network can process and mine time series better than RNN.

In recent years, some ensemble learning algorithms which combine LSTM, GRU, and decomposition algorithms have been proposed and achieved good prediction results. The sequence decomposition method can decompose the historical load sequence into multiple subsequences and convert the highly nonlinear nonstationary time series forecasting problem into multiple relatively stationary time series forecasting problems. Li et al. [21] combined variational mode decomposition (VMD) with LSTM. The VMD could decompose the wind power load sequence into three different frequency components: high, medium, and low. Then LSTM was used to predict them separately, which effectively improved the accuracy of load forecasting. Ren et al. [25] used ensemble empirical mode decomposition (EEMD) to decompose the original load sequence and divided the original sequence into high-frequency components and low-frequency components according to the zero-crossing rate of each subsequence. Finally, GRU and multiple linear regression were used to make predictions.

Although deep learning algorithms have many advantages over traditional algorithms, their prediction results depend to a large extent on the quality and quantity of data. High-quality and massive data can often make it easier to obtain accurate prediction results. However, in actual load forecasting, due to various reasons such as communication delays, the load data often contains irregular vacancies and noise interference, and it is difficult to fill the irregular vacancies with traditional interpolation methods [26]. In addition, deep learning algorithms generally have the disadvantages of difficulty in determining hyperparameters, high resource consumption, and slow calculation speed.

3. Method

3.1. RNN Model. Recurrent neural network (RNN) is a special kind of neural network with self-connection in the field of deep learning, which can learn complex vector to vector mapping and process time series data and learn the characteristics of time series effectively. The first research on RNN is Hopfield network model proposed by Hopfield, which has strong computing power and associative memory function. However, it is difficult to implement and is replaced by other artificial neural networks and traditional machine learning algorithms. Jordan and Elman proposed the framework of recurrent neural network in 1986 and 1990, respectively, which is called simple recurrent network (SRN). It is considered to be the basic version of RNN, which is widely popular at present. The more complex structures that appear after that can be regarded as its variants or extensions. RNN has been widely used in various time series related tasks.

RNN is connected through a loop on the hidden layer, so that the network state at the previous moment can be transferred to the current moment, and the state at the current moment can also be transferred to the next moment.

The RNN can be regarded as a deep FNN in which all layers share weights, which can be extended by connecting the adjacent time steps. The concept of parameter sharing has already appeared in hidden Markov model (HMM).

HMM is also often used in sequence data modeling and once achieved good results in the field of time series forecasting. Both HMM and RNN use internal states to represent the dependencies in the sequence. When the time series data has long-distance dependence and the scope of the dependence varies with time or is unknown, RNN may be a relatively better solution.

Figure 1 shows the structure of RNN and the forward propagation of RNN can be expressed as follows:

$$\begin{aligned} h_t &= \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h), \\ o_{t+1} &= W_{hy}h_t + b_y, \\ y_t &= \text{softmax}(o_t), \end{aligned} \quad (1)$$

in which W_{xh} is the weight matrix from the input unit to the hidden unit, W_{hh} is the connection weight matrix between the hidden units, W_{hy} is the connection weight matrix from the hidden unit to the output unit, and b_y and b_h are the bias vectors. The parameters needed in the calculation process are shared, so theoretically RNN can handle sequence data of any length. The calculation of h_t requires h_{t-1} , the calculation of h_{t-1} requires h_{t-2} , and so on, so the state at a certain moment in the RNN depends on all the past states. RNN can map sequence data to sequence data output, but the length of the output sequence is not necessarily consistent with the length of the input sequence. According to different task requirements, there can be multiple correspondences.

Although RNN model can deal with nonlinear time series data effectively, for long time series, RNN has the phenomenon of gradient disappearance or gradient explosion, which leads to the loss of historical information. In addition, the length of delay window needs to be set in advance in the process of training RNN model, but the optimal value of this parameter is difficult to obtain in practical application.

3.2. LSTM Model. In current, the most widely used recurrent structure network architecture in practical applications comes from the LSTM model (nonforget gate) proposed by Hochreiter and Schmidhuber [27], which can effectively overcome the problem of gradient disappearance in RNN. Especially in long-distance dependent tasks, the performance of LSTM is far better than RNN. The gradient backpropagation process will no longer be troubled by the problem of gradient disappearance, and accurate modeling of data with short-term or long-term dependence can be carried out. The working mode of LSTM is basically the same as that of RNN. The difference is that LSTM implements a more refined internal processing unit to achieve effective storage and update of time series information.

As shown in Figure 2, there are three types of gates in the LSTM unit: input gates, forget gates, and output gates. Gating can be regarded as a fully connected layer, and the storage and updating of information by LSTM is realized by these gates. More specifically, the gating is implemented by the sigmoid function and the dot multiplication operation, and the gating does not provide additional information. The general form of gating can be expressed as

FIGURE 1: RNN structure.

$$g(x) = \sigma(Wx + b), \quad (2)$$

in which $\sigma(x) = 1/(1 + \exp(-x))$, called the sigmoid function, is a nonlinear activation function commonly used in machine learning. It can map a real value to the interval $0 \sim 1$, which is used to describe how much information passed. When the output value of the gate is 0, it means that no information passes, and when the value is 1, it means that all information can pass. i , f , and o represent input, forget, and output gates, respectively. $\odot$ represents the multiplication of corresponding elements, and W and b represent the weights and biases of the networks, respectively.

The forward calculation process of LSTM can be expressed as equations (3)–(7). At the time step t , the input and output vectors of the hidden layer of LSTM are x_t and h_t , respectively, and the memory unit is c_t . The input gate is used to control the amount of current input data x_t flowing into the memory unit.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i). \quad (3)$$

Forget gate is a key component of LSTM unit, which can control which information should be retained and which information should be forgotten and avoid the problem of gradient disappearance and explosion when the gradient propagates backward. The forget gate controls the self-connecting unit and can determine which parts of the history information will be discarded.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \quad (4)$$

$$c_t = f \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \quad (6)$$

$$h_t = o_t \odot \tanh(c_t). \quad (7)$$

The output gate controls the influence of the memory unit c_t on the current output value h_t , namely, which part of the memory unit will be output at time step t . The value of

the output gate is shown in equation (6), and the output h_t of the LSTM unit at time t can be obtained by equation (7).

3.3. Attention Mechanism-Based Forecasting. The input data of short-term load forecasting involves many types, such as weather data (temperature, humidity, precipitation, wind speed, etc.), daily type data, electricity data, and electricity price information. For example, for the load of a certain day, the impact of three consecutive days of high temperature and sudden high temperature on the load of that day will be significantly different. At the same time, the weather is a combination of many factors to produce a certain effect before acting on the power load, so the coupling effect of weather index should be considered when analyzing the impact. Therefore, the features considered in our paper include holiday, weekend, temperature, humidity, rainfall, wind speed, and the historical power load data. Table 1 shows the description of the features.

The attention mechanism has become an integral part of sequence modeling, and it allows the modeling of dependencies without regard to their distance in sequences [28]. When the sequence proceeds to the output, it generates an attention range to highlight the part of the sequence that should receive much attention from LSTM. Figure 3 shows the deep network architecture for power load forecasting. The feature sequences are regarded as input and the forecasting result is regarded as the output. The attention layer is consisted of a dense layer followed by a SoftMax function, which is utilized to calculate the attention weights. The attention weights denote the importance degree of the corresponding features which can emphasize the most effective information of the data. Then, there is multiplication operation of the attention weights and the input feature.

The overall attention-based LSTM model can be regarded as a solution process of an optimization problem. The decision variables are the parameters, and the objective function is the mean squared error (MSE) of the forecasting power load:

FIGURE 2: Structure of LSTM units.

TABLE 1: Feature description.

Feature	Description
Holiday	Dummy variable: 0, nonholiday; 1, holiday
Weekend	Dummy variable: 0, weekday; 1, weekend
Temperature	Continuous variable: current time temperature
Humidity	Continuous variable: current time temperature
Rainfall	Continuous variable: current period rainfall
Wind speed	Continuous variable: current time wind speed
Power load ($n-1$)	Continuous variable: power load in lag $n-1$
Power load ($n-2$)	Continuous variable: power load in lag $n-2$
...	...

$$\theta = \arg \min_{\theta} \|l_i - \hat{l}_i\|^2, \quad (8)$$

where l_i represents the ground truth of the power load of the target time step and $\hat{l}_i$ represents the forecasting value of the power load of the target time step. θ represents the parameters of the framework, which can be learned through the Adam optimizer via backpropagation.

3.4. Isolation Forest-Based Early Warning. Early warning technologies are widely adopted in various fields [29, 30]. The Isolation Forest (iForest) algorithm is an unsupervised anomaly detection method. The training of the algorithm is mainly by randomly selecting n samples from the training dataset to divide the true binary tree. Namely, a feature is randomly selected, and a split point is randomly selected between the maximum and minimum values of the feature. Those smaller than the split point enter the left branch, and those greater than or equal to the split point enter the right branch. Keep repeating the above process until there is only one sample or the same sample cannot continue to split or the depth limit of the tree is reached. The path length $h(x)$ refers to the number of edges of the binary tree that the sample point x passes from the root node to the external node. Due to their particularity, abnormal samples can usually be separated earlier to reach external nodes, and the path length is small. Normal samples can be separated only after multiple binary tree classifications, and the path length is relatively large. In the same way, an isolated forest

containing multiple isolated trees is constructed, and abnormal events can be detected based on the path length of the sample in each isolated tree.

The degree of data anomaly can be judged by the anomaly score $s(x, n)$. It is defined as follows:

$$c(n) = 2H(n-1) - \left(\frac{2(n-1)}{n}\right), \quad (9)$$

$$s(x, n) = 0.5 - 2^{-(E(h(x))/c(n))}, \quad (10)$$

where n is the number of samples in the sample set, $h(i)$ is the harmonic order, which can be estimated by $\ln(i) + 0.5772156649$ (Eulerian constant), $c(n)$ is the average path length of the binary search tree, used for the standardization of $h(x)$, and $E(h(x))$ is the average of the path lengths of all isolated trees in the isolated forest at the sample point x . When the abnormal score $s(x, n)$ is smaller, the degree of abnormality is higher, and the possibility of being an abnormal point is greater. The training phase returns the isolated tree structure and segmentation conditions, and the out-of-sample data can use the split conditions of the training phase to calculate anomaly scores to determine whether there is anomaly. The isolated forest algorithm is not based on distance and density to judge anomalies, which is suitable for processing high-dimensional data and large-scale data. In this paper, we input the historical power load data and the forecasting value of the target time step to the iForest model and output an early warning of the power system.

FIGURE 3: A deep network architecture for power load forecasting.

4. Experimental Results

4.1. Dataset. The data is selected from the load data of 15 Chinese enterprises from January 1 to March 31, 2021. The sampling interval is 1 hour, and there are 24 data items per enterprise per day. Figure 4 shows hourly power load sample of an enterprise and Figure 5 shows a daily power load sample of an enterprise. The load data does not have complete periodicity and seasonality, and the daily peak and valley values are also different. This paper constructs time series data for training and prediction and uses one day as the window to slide the data; namely, the data at the previous 24 time points are “features” (including features such as weather and date). The data at the next time point are “label,” getting several sub-time-series data. The data corresponding to January 1 to February 28 are utilized as the training set and these are input into the model for training. After the model is trained, the iterative prediction method is used to input the 24 data items from March 1 into the model as “features,” and the load forecasting value at one moment can be obtained and then the model brings which back into the time series and then forecasts the next moment and so on, as well as iteratively forecasting 24 power load data items throughout the day on March 31.

4.2. Experimental Setups

4.2.1. Evaluation Metrics. In our work, we conduct evaluation of the power load forecasting accuracy utilizing MSE and MAPE:

$$\text{MSE} = \frac{1}{T} \sum_{i=1}^T (\hat{l}_i - l_i)^2, \quad (11)$$

$$\text{MAPE} = \frac{1}{T} \sum_{i=1}^T \frac{|l_i - \hat{l}_i|}{l_i},$$

where l_i represents the i th ground truth of the power load, $\hat{l}_i$ represents the i th forecasting value of the power load, and T is the total number of power load forecasting steps.

In terms of early warning, we evaluate the early warning by the precision, recall, and the $F1$ -score, which are calculated in the three following equations:

$$P = \frac{TP}{TP + FP}, \quad (12)$$

where P denotes the precision and TP and FP denote the true positive and false positive rates, respectively.

$$R = \frac{TP}{TP + FN}, \quad (13)$$

where R denotes the recall and TP and FN denote the true positive and false negative rates, respectively.

$$F1 = \frac{2TP}{2TP + TP + FN}, \quad (14)$$

where $F1$ denotes the harmonic mean of precision and recall.

4.2.2. Benchmarks. The benchmark methods of the forecasting model include statistical model, traditional machine learning model, and deep learning model. Statistical model time series models include time series models, autoregressive integrated moving average (ARIMA) model, and ARIMAX. Traditional machine learning models include support vector regression (SVR), regression tree (RT) model, and multilayer perceptron (MLP). Deep learning models include LSTM and GRU.

The benchmark methods of the early warning model include linear regression (LR), decision tree (DT), support vector machine (SVM), Naive Bayes (NB), and MLP.

In the computing environment, we conduct experiments mainly on a vision station with 4 CPU cores (Intel Core i7-7700 CPU@3.60 GHz) and a GPU (NVIDIA GeForce RTX 2080). We use Python 3.5.2 with scikit-learn, TensorFlow, and Keras to build models.

4.3. Experimental Results. Table 2 shows MSE and MAPE under different attention-based LSTM structures. The results show that the model performs best when LSTM cells are 2 and LSTM units are [256, 256].

In order to verify the effect of the model, we tested the prediction accuracy of data in 15 regions. Tables 3 and 4 show the MSE and MAPE of different competition models. The prediction effect of the model with features other than

FIGURE 4: The hourly power load of an enterprise.

FIGURE 5: The daily power load of an enterprise.

TABLE 2: Evaluation of attention-based LSTM structure.

LSTM cells	LSTM units	MSE ($\times 10000^2$)	MAPE (%)
1	64	92.16	8.9
		95.34	8.85
		88.26	8.82
		95.65	8.76
	128	97.93	8.83
		91.42	8.76
		95.75	8.74
		91.61	8.81
	256	92.48	8.77
		87.40	8.68
2	[64, 64]	96.09	8.66
		96.17	8.72
		88.80	8.61
		94.80	8.60
	[128, 128]	97.28	8.9
		94.29	8.85
		91.82	8.82
		90.87	8.60
	[256, 256]		

historical data is better than those of the other models. The prediction performance of the deep learning model is better than those of the traditional machine learning model and time series model, while the results of A-LSTM are the most prominent in the deep learning model.

10-fold cross-validation is used to test the accuracy of the algorithm. It is a common test method. The dataset is divided into ten parts, nine of which are used as training data and one is

used as test data in turn. Each test will get the corresponding correct rate (or error rate). The average value of the correct rate (or error rate) of the results of 10 times is used as the estimation of the accuracy of the algorithm. Generally, it needs to carry out 10-fold cross-validation for many times (e.g., 10-fold cross-validation) and then calculate the average value as the estimation of the accuracy of the algorithm. In this paper, we use 10-fold cross-validation method to conduct the early warning

TABLE 3: MSE ($\times 10000^2$) comparison.

Enterprise no.	A-LSTM	GRU	LSTM	RT	ARIMAX	ARIMA	SVR	MLP
1	92.15	102.68	105.55	107.88	117.91	115.18	122.39	131.94
2	90.43	104.10	105.35	112.25	119.21	118.01	121.20	138.11
3	96.59	100.05	108.64	114.33	109.78	120.75	126.60	121.00
4	92.32	102.64	107.31	114.82	110.77	116.25	127.35	134.28
5	95.47	97.00	102.23	113.35	114.62	121.28	118.80	132.71
6	95.85	95.86	102.73	111.17	116.62	121.98	119.66	132.26
7	91.52	103.34	101.38	112.59	119.77	116.78	121.07	127.24
8	91.65	105.43	103.84	114.55	119.61	117.70	119.47	127.86
9	94.17	100.45	103.52	114.29	111.83	112.33	124.83	124.01
10	97.49	100.29	103.46	110.83	119.61	115.60	119.10	127.88
11	92.13	104.41	104.36	111.83	113.96	117.92	125.66	127.74
12	93.18	102.01	101.61	115.36	114.45	119.93	122.18	132.67
13	87.81	102.89	100.18	105.88	116.56	114.78	121.79	129.82
14	92.15	103.05	104.18	105.31	115.00	118.97	118.86	128.89
15	95.52	99.69	101.83	115.12	114.68	119.82	124.94	128.39

TABLE 4: MAPE comparison.

Enterprise no.	A-LSTM	GRU	LSTM	RT	ARIMAX	ARIMA	SVR	MLP
1	8.90	9.49	9.94	10.70	10.94	11.14	11.96	12.39
2	8.85	9.45	9.89	10.76	10.95	11.10	11.88	12.83
3	8.82	9.50	9.89	10.71	10.93	11.12	11.85	11.91
4	8.85	9.52	9.90	10.64	10.91	11.12	11.77	12.69
5	8.86	9.55	9.88	10.65	10.96	11.16	11.79	12.40
6	8.83	9.56	9.91	10.67	10.94	11.21	11.78	12.54
7	8.80	9.57	9.89	10.67	10.99	11.22	11.75	12.38
8	8.82	9.60	9.91	10.69	10.97	11.27	11.74	11.99
9	8.79	9.62	9.86	10.64	11.03	11.20	11.72	12.27
10	8.75	9.54	9.87	10.64	11.08	11.19	11.72	12.05
11	8.74	9.53	9.87	10.63	11.02	11.21	11.78	12.43
12	8.73	9.46	9.80	10.56	10.98	11.21	11.81	12.64
13	8.64	9.52	9.78	10.56	11.02	11.25	11.76	12.42
14	8.65	9.52	9.77	10.48	11.10	11.20	11.83	12.14
15	8.90	9.49	9.94	10.70	10.94	11.14	11.96	12.39

TABLE 5: Early warning accuracy.

Method	Precision (%)	Recall (%)	F1-score (%)
LR	92.50	93.30	92.90
DT	93.17	95.65	92.63
SVM	92.53	93.86	92.45
NB	94.97	94.57	95.08
MLP	95.68	95.23	94.02
Isolation Forest	98.02	96.34	96.20

experiments. Table 5 shows the accuracy of early warning; and as regards precision, recall, and *F1*-score, the performance of isolation forest is better than those of the other models. The accuracy of our prediction model and early warning model is better than that of the baseline, and it can reach the standard of application in practice.

5. Conclusions

With the development of power market, accurate short-term forecasting of power load can effectively guarantee the safe operation of power grid, reduce the cost of power

generation, meet the needs of users, and improve social and economic benefits. Short-term power load forecasting is to forecast the power load in a short period of time in the future according to the power load in the past and load related data such as temperature, humidity, and date type. Accurate prediction is not only conducive to timely macro control of users' electricity consumption behavior but also to provide scientific guidance for power production.

This paper proposes a short-term load forecasting model of power system based on attention mechanism. On the basis of fully mining the regularity of historical load data, we introduce features such as weather, weekends, and holidays

and input the data feature vector into the attention-based LSTM unit for feature extraction. The attention mechanism can extract the most relevant information from the features. This method can make full use of the regularity of historical load data, consider the influence of various factors, and improve the forecasting efficiency and accuracy. The accuracy of our prediction model and early warning model is better than that of the baseline, and it can reach the standard of application in practice.

Our research results can provide decision support for the power sector and stakeholders. This model can also be used for early warning of local sudden large loads and identification of enterprise power demand.

Data Availability

The data used to support the findings of this study were supplied by State Grid Xiongan Financial Technology Group Co., Ltd., Beijing, China, under license and so cannot be made freely available. Requests for access to these data should be made to Lifeng Li.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Science and Technology Foundation of SGCC (5400-202018223A-0-0-00).

References

- [1] M. Ali, M. Adnan, and M. Tariq, "Optimum control strategies for short term load forecasting in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 113, pp. 792–806, 2019.
- [2] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943–3952, 2019.
- [3] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 127–136, 2019.
- [4] A. Bracale, P. Caramia, P. De Falco, and T. Hong, "Multivariate quantile regression for short-term probabilistic load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 628–638, 2020.
- [5] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.
- [6] T. Ouyang, H. Huang, Y. He, and Z. Tang, "Chaotic wind power time series prediction via switching data-driven modes," *Renewable Energy*, vol. 145, pp. 270–281, 2020.
- [7] A. Heydari, M. M. Nezhad, E. Pirshayan, D. A. Garcia, F. Keynia, and L. D. Santoli, "Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm," *Applied Energy*, vol. 277, 2020.
- [8] C. Fan, C. Ding, J. Zheng, L. Xiao, and Z. Ai, "Empirical mode decomposition based multi-objective deep belief network for short-term power load forecasting," *Neurocomputing*, vol. 388, 2020.
- [9] R. Ye, X. Huang, Z. Chen, Z. Ji, and L. Tan, "Novel fast forecasting method for nodal voltage violations," *IET Generation Transmission & Distribution*, vol. 14, no. 24, pp. 6027–6039, 2020.
- [10] Z. Chen, D. Zhang, H. Jiang, L. Wang, and Y. Chen, "Load forecasting based on LSTM neural network and applicable to loads of "replacement of coal with electricity"," *Journal of Electrical Engineering and Technology*, vol. 6, 2021.
- [11] N. Amral, C. S. Ozveren, and D. King, "Short term load forecasting using multiple linear regression," in *Proceedings of the International Universities Power Engineering Conference*, Brighton, UK, September 2007.
- [12] J. P. Hermias, K. Teknomo, and J. Monje, "Short-term stochastic load forecasting using autoregressive integrated moving average models and Hidden Markov model," in *Proceedings of the 2017 International Conference on Information and Communication Technologies (ICICT)*, Karachi, Pakistan, December 2017.
- [13] B. M. Hodge, D. Lew, and M. Milligan, "Short-term load forecast error distributions and implications for renewable integration studies," in *Proceedings of the 2013 IEEE Green Technologies Conference (GreenTech)*, Denver, CO, USA, April 2013.
- [14] Y. Song, Z. Shen, D. Dai, Y. Qian, and Y. Wang, "Short-term load forecasting in electrical power systems via trajectory tracking and error correcting approach," *Journal of Renewable and Sustainable Energy*, vol. 6, no. 1, Article ID 013112, 2014.
- [15] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1316–1324, 2014.
- [16] Z. P. Tang, Z. X. Chen, J. P. Sun et al., "Noise prediction of traction gear in high-speed electric multiple unit," *International Journal of Simulation Modelling*, vol. 18, no. 4, pp. 720–731, 2019.
- [17] Y. Wang, Y. J. Lu, C. D. Si, and T. C. Sun, "Finite element analysis for rutting prediction of asphalt concrete pavement under moving wheel load," *International Journal of Simulation Modelling*, vol. 16, no. 2, pp. 229–240, 2017.
- [18] G.-F. Fan, L.-L. Peng, W.-C. Hong, and F. Sun, "Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression," *Neurocomputing*, vol. 173, pp. 958–970, 2016.
- [19] J. Wu, Z. Cui, Y. Chen, D. Kong, and Y.-G. Wang, "A new hybrid model to predict the electrical load in five states of Australia," *Energy*, vol. 166, pp. 598–609, 2019.
- [20] E. O.-N. Jnr, Y. Y. Ziggah, and S. Relvas, "Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting," *Sustainable Cities and Society*, vol. 66, no. 4, Article ID 102679, 2021.
- [21] J. Li, D. Deng, J. Zhao et al., "A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2443–2452, 2020.
- [22] U. Claude, "Predicting tourism demands by Google trends: a hidden Markov models based study," *Journal of System and Management Sciences*, vol. 10, no. 1, pp. 106–120, 2020.

- [23] D. P. Kalala and P. Bhargavi, "Hybrid deep learning approach for load forecasting in power systems," *Materials Today: Proceedings*, vol. 3, 2021.
- [24] Z. Li, Y. Li, Y. Liu, P. Wang, and H. B. Gooi, "Deep learning based densely connected network for load forecasting," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 2829–2840, 2020.
- [25] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 236–244, 2017.
- [26] Z. Guo, K. Zhou, X. Zhang, and S. Yang, "A deep learning model for short-term power load and probability density forecasting," *Energy*, vol. 160, pp. 1186–1200, 2018.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio et al., Eds., vol. 30, pp. 5998–6008, Curran Associates, Inc., Red Hook, NY, USA, 2017.
- [29] G. Xue, S. Liu, and D. Gong, "Identifying abnormal riding behavior in urban rail transit: a survey on "in-OUT" in the same subway station," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [30] G. Xue, S. Liu, D. Gong, and Y. Ma, "ATP-DenseNet: a hybrid deep learning-based gender identification of handwriting," *Neural Computing & Applications*, vol. 33, no. 10, pp. 4611–4622, 2021.