RESEARCH ON SHORT-TERM FORECAST OF POWER LOAD BASED ON ADAM-GRU

Pengtao Jia, Yunfeng Miao*
Department of Computer Science and Technology, Xi’an University of Science and Technology, Xi’an 710600, China
614039013@qq.com
*18208208049@stu.xust.edu.cn

Abstract. To forecast the short-term power load accurately and reliably, a short-term power load model based on gate recurrent unit was proposed. In order to improve the accuracy of the model, the Adam algorithm is used to calculate the GRU hyperparameter Optimization, so as to construct an Adam-GRU short-term power load forecasting model. The performance of the forecasting model was evaluated by the Root Mean Square Error and running time, then the forecast results were compared with LSTM method. The experimental results show that Adam-GRU has a higher accuracy than the LSTM model, an increase of 15.54%, and a time complexity reduction of 15.75%. Compared with the RNN model, the accuracy has improved 10.84%. The results show that the short-term power load forecasting model and parameter optimization method based on Adam-GRU can effectively forecast the power load data.

1. Introduction
Short-term forecast of power load is to find the change law from the historical data of power load and to know the impact on the future power load better, so as to realize the scientific forecast of the future power load [1]. The planning and operation of power system requires high precision of power load to ensure that the planned installed capacity and charging margin of the system can meet the needs of the society, ensure the efficient operation efficiency of the system and save investment [2]. Therefore, improving the accuracy of the forecast is the core requirement of the power load forecasting. Besides, the high precision of power load forecasting provides economical and safe guarantee for the operation of the power system. There are many unknown factors on short-term forecast of power load, and the data of short-term power load has the characteristics of timing and nonlinearity.

At present, these methods of power load forecasting can be divided into statistical methods and artificial intelligence methods at home and abroad [3][4].

Methods based on the statistical: Lei ShaoLan et al applied multivariable time series to the linear regression forecast of power short-term load, which has strong adaptive ability and better prediction effect [5]. Qu Diqing et proposed a fuzzy linear regression method to reduce the error of the holidays short term load forecasting, and has a high prediction accuracy [6]. Zhang Shu achieved good results in the research field of short-term power load forecasting by derivative distance into the similarity measure, taking full account of the shape characteristics of the load curve. [7].

Methods based on artificial intelligence are as follows: Bo-Sung using LSTM Layer in deep neural networks to forecast short-time load [8]. By optimizing the grey Elman neural network with genetic algorithm, Wang Zheng obtained the optimal initial value of the network, and verified the superiority
of the algorithm with historical data\cite{9}. Liu Xin used the improved harmonic search algorithm to select the optimal initial weight value of wavelet neural network, and carried on the training study to the fixed initial weight wavelet neural network, then verified its validity based on the actual load data\cite{10}. Jian Xianzhong et al. used the whole-process optimized support vector machine (SVM) model for short-term power load forecasting, and improved the accuracy of local optimization model\cite{11}. Huang Wensi combined the kernel principal component analysis method with the improved echo state network algorithm to improve the prediction accuracy\cite{12}.

Models based on statistical methods and traditional artificial intelligence methods cannot simultaneously take into account the time series and nonlinear characteristics of power load data. The long and short memory network and the deep belief network have more complex network structure, and the time complexity of model training is relatively high\cite{13}.

Therefore, this paper proposes a power load forecasting method based on Adam optimization of gated cycle units. The experimental result shows that the GRU model which is based on Adam optimization has lower error, less training time of models and better prediction effects.

2. Data Source and Data Preprocessing

2.1. Data Source

The data set used in this experiment is a monitoring data of 1075 days from a certain power supply unit to a certain area power supply monitoring point from 2017/01/01 to 2019/12/11, and the data visualization is shown in figure 1.

\[ \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \]  

(1)
where $x$ is the mean of the dataset; $\delta$ is the deviation of the standard dataset.

$$x^* = \frac{(x-x_{\text{min}})}{(x_{\text{max}}-x_{\text{min}})}$$

where $x^*$ represents the standardized value, $x$ represents the pre-standardized value, $x_{max}$ represents the maximum value in the original dataset, and $x_{min}$ represents the minimum value in the original dataset.

3. Model construction

3.1. GRU Model

Due Cho proposed LSTM variant threshold cycle units (GRU) in 2014\cite{14}. The gate cycle unit combines the input gate with the forget gate to make the update gate. The new gate decides to retain or discard the information. It reduces the redundancy of using the input gate and the forget gate simultaneously and maintains features of long-and-short-term memory networks while making the internal structures simple. At the same time, memory units and hidden neurons are combined, so the GRU parameters are less, the training time is shorter, and it is not easy to produce the problem of overfitting. GRU structure is shown in Figure 2.

\begin{align*}
z_t &= \text{sigmoid}(w_z \times [h_{t-1}, x_t]) \quad (4) \\
r_t &= \text{sigmoid}(w_r \times [h_{t-1}, x_t]) \quad (5) \\
'h &= \text{Tanh}(w \times [r_t \times h_{t-1}, x_t]) \quad (6) \\
h_t &= (1 - z_t) \times h_{t-1} + z_t \times 'h \quad (7)
\end{align*}

where $z_t$ is the update gate, $r_t$ is the reset gate, $\sigma$, $\text{Tanh}$ represents the activation function $\text{sigmoid}$ and the hyperbolic tangent function respectively. The update gate represents the received new state, and $(1-z_t)$ represents the retained old state.

3.2. Optimization algorithm

The process of making the model achieve the best prediction effect by using a large number of samples to modify the parameters of the model continuously through the optimization function is the process of model training. The traditional gradient descent algorithm often falls into the local optimal value and does not solve the problem of global optimization. While good learning rate parameters can improve the prediction accuracy of the model after training, we need better optimization algorithm to optimize the learning rate. Adam learned from Adagrad’s ideas, and carried out the adaptive learning
rate change of each parameter. According to the accumulation of the gradient, the mean value of the value of the historical gradient in the window is calculated approximately \[15\].

Adam gradient update rules (8), (9):

\[
\Delta \theta_t = -\frac{\text{RMS}[\Delta \theta_t]}{\text{RMS}[g_t]} g_t 
\]

\[
\theta_{t+1} = \theta_t + \Delta \theta_t 
\]

where \(g\) is the gradient \(\theta\) at \(t\) time. \(\text{RMS}[\Delta \theta]\) means learning rate. Do not need to be set in advance. It is an adaptive update.

3.3. Power Load Forecast Model Based on Adam-GRU

Figure 3 is the three-layer network model framework of the GRU power load forecasting which is based on Adam optimization, including input layer, hidden layer, output layer and Adam optimization. The input layer preprocesses the time series of original power load to meet the input requirements of the GRU model; the hidden layer constructs neural networks with GRU neurons; the Adam optimization continuously optimizes the model; and the output layer outputs the corresponding time series to realize the forecast of power load.

Fig.3 Chart of Model Forecast

4. Experimental Verification

4.1. Experimental Process

We divide 90% of the data set into training set and 10% into test set. The training set is used to train the prediction model, and the final result of the test set is used for evaluation. Here is the experimental process.

(1) Data processing
(2) The data is divided into training set and test set, the proportion of training set is 90%, and the test set is the remaining 10%.
(3) Rebuild the dataset into the data format required by the model.
(4) Determine the parameters of the model; take MSE as the loss function; use the Adam optimization method to optimize.
(5) Determined iteration on the basis of loss function.
(6) Forecast Test Set.
(7) Anti-normalized data set.
(8) Display results.

To verifying the advantages of threshold cycle unit in predicting power load, we use the Keras library to build the model and set one hidden layer. The loss is the lowest when the number of iterations is 15 times and the number of hidden layer nodes is 6. Then the model structure is determined.

4.2. Results and analysis

In order to carry on the contrast experiment, we use RNN and LSTM neural network to build the model. The forecast results of GRU model which is based on the Adam optimization are shown in figure 4 respectively.

![Fig.4 Results of Train Model](image)

(a) Results of RNN Model Forecast  
(b) Results of LSTM Model Forecast  
(c) Results of Model Forecast in this paper
From table 2 it can be seen that the accuracy of this model is 10.84% higher than that of the LSTM model, compared with the LSTM model the accuracy is improved 15.54% and the CPU running time is reduced by 15.75% by using the root mean square error root as the evaluation function.

| Model     | TestRMSE  | CPUruntime |
|-----------|-----------|------------|
| RNN       | 166.074726| 11.408s    |
| LSTM      | 175.438281| 20.966s    |
| Adam-GRU  | 148.165615| 17.663s    |

5. Conclusion
Based on the results and discussions presented above, the conclusions are obtained as below:
(1) In this paper, considering the timing and nonlinear characteristics of the short-term power load data, we propose the short-term forecast of power load based on Adam-GRU model.
(2) By comparing with the prediction results of LSTM model, it is verified that this method is superior to the prediction method based on LSTM neural network model. The experimental results show that the prediction accuracy of this model is improved by 15.54%, and the training time of the model is reduced by 15.75%; compared with the RNN model, the RMSE is reduced 10.84%; so the Adam-GRU model is more suitable for forecasting short-term power load data.

Acknowledgments
This work was supported by Scientific and Technology Program Funded by Xi’an City(Program No. 2020KJRC0069).

References
[1] YUE Yuanbo, HAN Aoyang, ZHANG Zhisheng1. Short-Term Load Forecasting Model of Power System Based on Deep Ridgelet Neural Network [J/OL]. Proceedings of the CSU-EPSA:1-6[2020-06-21].
[2] Dai Yinan. Research on Short-term Power Load Forecasting Method Based on Neural Network[D].Xi’an university of architecture and technology,2019.
[3] Lai Wei. Research on Power Load Modeling Method Based on Component-Based Modeling Approach[D].Zhengzhou University,2019.
[4] Hao Mukai, Su Zheng, Zhao Qi, Wang Yingchun, Zhang Dongxia. Application and Prospect of Artificial Intelligence in the Electric Power Industry [J]. Information Technology & Standardization,2020(Z1):76-82.
[5] Lei Shaolan, Sun Caixin, Zhou Quan, Zhang Xiaoxing. The Research of Local Linear Model of Short-Term Electrical Load on Multivariate Time Series [J]. Proceedings of the CSEE. 2006(02).
[6] Qu Diqing, Chen Hongmin, Lv Lixiang. Holidays short term load forecasting using fuzzy linear regression method [J]. Power Demand Side Management,2013(03).
[7] Zhang Shu. Research on Short-Term Load Forecasting Based on Functional Date Analysis [D].Xiamen University,2018.
[8] Bo-Sung Kwon;;Rae-Jun Park;;Kyung-Bin Song. Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer [J]. Journal of Electrical Engineering & Technology,2020.
[9] Zheng Wang. Short Term Power Load Forecasting based on Cloud Computing and Gray Neural Network[D]. North China Electric Power University,2016.
[10] Liu Xin. Research on Short-term Load Forecasting Methods Based on Machine Learning[D].Beijing University of Posts and Telecommunications,2019.
[11] JIAN Xianzhong, GU Yiting. Short-term electric load forecasting method based on global optimization support vector machine [J]. Electric Power Science and
[12] Huang Wensi. SHORT-TERM POWER LOAD FORECASTING CONSIDERING METEOROLOGICAL FACTORS [J]. Computer Applications and Software, 2020, 37(1):98-105.

[13] Hinton GE, Osindero S, Yee-Whye Teh. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation. 2006;18(7):1527-1554. doi:10.1162/neco.2006.18.7.1527.

[14] Cho, Kyunghyun, et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." Computer Science (2014).

[15] Matthew D. Zeiler. "ADAM: An Adaptive Learning Rate Method." arXiv:1212.5701v1 [cs.LG] 22 Dec 2012.)