GRU-based Encoder-Decoder for Short-term CHP Heat Load Forecast

Kuan LU\(^1\), Xiang Rong MENG\(^1\), Wen Xue SUN\(^2\), Rong Gui ZHANG\(^3\), Ying Kun HAN\(^1\), Song GAO\(^3\), Dongliang Su\(^4\)

\(^1\) State Grid Shandong Electric Power Research Institute, Jinan, Shandong China.
\(^2\) State Grid Zhangqiu Power Supply Company, Jinan, Shandong, China.
\(^3\) Shandong Luneng Software Technology Co., LTD, Jinan, Shandong, China.
\(^4\) Shandong Zhongshi Yitong Group Co., LTD., Jinan, Shandong, China.
\(^a\) Email: lk83@163.com

Abstract. A gated recurrent units (GRU) network based encoder-decoder (E-D) model is proposed for combined heat and power (CHP) heat load forecasting. First, we use the GRU based E-D model as an auto-encoder to map the historical CHP heat load time series into a fixed length representation. Then, concatenate the representation with weather forecasting information as a new input to another multiple GRUs network for heat load prediction. Data collected from Rizhao, Shandong province is used to verify the conclusions. Results illustrate that GRU-based E-D model can give a more accurate forecast to the short-term CHP heat load compared with non-auto-encoded RNN model.

1. Introduction

It is required by the government to determine power generation by heat for CHPs in China. This indicates that CHP heat load forecast should not be neglected in the process of regional power dispatch management to improve the coordination of grids. What is more, subject to the transmission constraint, accurate CHP heat load forecast is also helpful for a more flexible peak load management [1].

Traditional heat load forecasting methods include regression analysis [2-3], exponential smooth method [4], time-series method [5-6], support vector machine (SVM) [7] and BP neural network [8-9] etc. The regression and exponential smooth methods are mainly applied to load data that changes smoothly and has clear trend. Time-series method neglects the analysis for the factors that influence heat loads. Meanwhile, volatility and randomness in short-term heat load data has negative effects on the load forecasting accuracy using traditional methods. As for SVM, there exists arbitrariness in the selection of kernel function. In addition, the increase of the sample data volume and dimension will lead to higher computational complexity [10].

With the success of dealing with temporal and sequential forecasting problem [11-15], more and more recurrent neural network (RNN) models are put into use in the field of energy forecast. Daniel M., Kasun A. and Milos M. [16] compare the short-term electric load forecasting accuracy of standard LSTM using LSTM-based sequence to sequence (Seq2Seq) architecture. Huiting Zheng, Jiabin Yuan, and Long Chen [17] show a generic framework that combines extreme gradient boosting, k-means on SD selection, empirical mode decomposition and LSTM based Seq2Seq model to forecast short-term
electric load. Kuan Lu, Zhenfu Bi and Xin Wang, et al [18] use a concatenated LSTM architecture to make CHP load forecast.

In this work, a GRU based encoder-decoder (E-D) model is proposed for CHP heat load forecasting. First, we use the GRU based E-D model as an auto-encoder to map the historical heat load time series into a fixed length representation. Then, concatenate the representation with weather forecasting information as a new input to another multiple GRUs network for prediction. Data collected from Rizhao, Shandong province is used to verify the improved performance using the proposed model compared with non-auto-encoded GRU model.

2. Data Preprocessing

Min-max scaling method is used to scale the features of input data down to a given range, as follows:

\[ y = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  

(1)

This is because the proposed model uses hyperbolic tangent (tanh) function as the activation function for GRU and the fully connected layer:

\[ \tanh(x) = 2 \cdot \text{sig}(x) - 1 \]  

(2)

where tanh function has a range of \([-1,1]\).

![Figure 1. Hyperbolic Tangent Function](image)

As is shown in Fig1 [19], the neurons of a layer will saturate if the value is too close to 1 or -1. So, we choose the scaled range of \([-0.5, 0.5]\).

3. Models and Forecasting Process

3.1. Gated Recurrent Units

Among all the gated RNNs, gated recurrent units (GRU) architecture introduced in 2014 [20] shows a better computational performance with less parameter [21]. The GRU combines input gate and forget gate of LSTM into an update gate \( z_t \), while, the output gate in LSTM is called a reset gate \( r_t \) in GRU shown in Fig2 [22].

\[ \bar{s}_t = \phi_{\text{tanh}}(W_s(r_t \odot s_{t-1}) + U_s x_t + b_s) \]  

(3)

\[ s_t = (1 - z_t) \odot s_{t-1} + z_t \odot \bar{s}_t \]  

(4)

with two gates:

\[ r_t = \sigma_{\text{sig}}(W_r s_{t-1} + U_r x_t + b_r) \]  

(5)

\[ z_t = \sigma_{\text{sig}}(W_z s_{t-1} + U_z x_t + b_z) \]  

(6)

Where: \( \odot \) is element-wise multiplier; \( W_s, W_r, \) and \( W_z \) are weight matrix for update gate, reset gate and candidate state; \( x_t \) is input data; \( \bar{s}_t \) is candidate state; \( s_t \) is output; \( b_s, b_r, \) and \( b_z \) are constants. \( \sigma_{\text{sig}} \)
and $\phi_{\text{tanh}}$ are sigmoid and tanh activation function. In the GRU architecture, $z_t$ determines how to combine new input with the previous memory while $r_t$ defines how much of the previous memory to let through.

![GRU Structure](image)

**Figure 2. GRU Structure**

### 3.2. GRU based E-D Architecture

In the GRU based E-D structure, shown in Fig3 [23], the encoder reads the historical CHP heat load sequence and maps it into a fixed-length vector representation i.e. the candidate state $\tilde{s}$. After that, the decoder will use $\tilde{s}$ and the value predicted by the previous time step to forecast the next time step. Given input $X = \{x^1, x^2, ..., x^n\}$, $\tilde{s}^E_t$ is the intermediate state of the encoder at step $t$, where $\tilde{s}^E_t \in \mathbb{R}^m$ and $m$ is the number of neurons in the encoder. The decoder decodes $\tilde{s}^E_t$ into the target sequence $Y = \{y^1, y^2, ..., y^n\}$.

![GRU based E-D Architecture](image)

**Figure 3. GRU based E-D Architecture**

### 3.3. GRU-based E-D CHP Heat Load Forecast Model

As is shown in Fig4, the GRU-based E-D CHP heat load forecast model uses the encoder-decoder architecture to perform auto-encoding (AE) to reduce model's misspecification error [24]. During AE process, the fixed-length intermediate state is extracted as an abstract representation of the heat load.
timing relationship. In order to improve the training efficiency, target value equals input value with reverse order. That is to say, given the input value \( \{x^1, x^2, \ldots, x^n\} \), the target value is \( \{x^n, \ldots, x^2, x^1\} \). Here, the auto-encoder use GRU architecture with 2 layers of which the number of neurons for each layer are 128 and 64.

After the AE process, an embedding layer representing the features of heat load time series is obtained. Then combine this embedding layer with future weather data to form a new input and feed it into another GRU network. The intuition is that change in weather patterns is the main reason for the volatility of heat load. If weather time series pattern remains unchanged, heat load historical information has been already learned during the AE process. If there are new changes in weather patterns, the forecast model should consider both of this and features learned from AE process.

Here, we use another 3-layer GRU network as a prediction model, where: input data has a dimension of 68 (including 64 dimensions of the embedded layer as well as wind speed, air temperature, air pressure and humidity) and neurons are set to 128, 64, 32. To train the model, back-propagation through time (BPTT) is used and the loss function has the form as follows:

\[
\text{Loss} = \sum_{t=1}^{M} (y_t - \hat{y}_t)^2
\]

4. Dataset and experimental results

The presented architecture was implemented on a dataset of weather and heat load with 10-minutes interval from a CHP in Rizhao, Shandong province.

The model aims to forecast the heat load of 36, 72, 144 time steps ahead. The whole dataset includes winter time that ranges from 20161218 to 20180309, totally 19164 observations. Divide the dataset into train set, validation set and test set with percentage of 50%, 30% and 20%. Use the deep-learning toolkit Keras to construct the GRU based E-D model. Root Mean Square Error (\( R_{MSRE} \)) for the percent of forecast vs. actual value is selected as evaluating indicator:

\[
R_{MSRE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{\hat{y}_i - y_i}{y_i} \right)^2}
\]
where $y_t$ is the actual value, while $\hat{y}_t$ is the forecasting value.

Validation errors of 144 time-steps-ahead forecasting for 30 training epochs in Fig.5 show the comparison of the proposed model with non-auto-encoded GRU model. Both of the models use the same data min-max scaling method. The result shows that GRU based E-D model has a lower learning error and converges to the optimal solution earlier than non-auto-encoded GRU model.

![Figure 5. Validation Errors for 144 Time-step-ahead Training](image)

Table 1 illustrates that the GRU-based E-D model obviously has a better performance. $R_{MSRE}$ of the test data using the proposed model is 3.4%, 5.2% and 8.3% less than GRU with no AE preprocessing.

| Time     | Test data       |               |
|----------|-----------------|---------------|
|          | GRU-based E-D   | Non-auto-encoded GRU |
| 6-hour   | 3.8             | 7.2           |
| 12-hour  | 5.3             | 10.5          |
| 24-hour  | 9.5             | 17.8          |

This is because the GRU network based on E-D extracts the time series feature of historical heat load during AE process, which takes into account new changes in the weather during the forecast period, such as sudden changes in wind speed and temperature. Therefore, the predictions made by GRU-based E-D model are more sensitive as shown in Fig. 6, whereas the non-auto-encoded GRU model has a delay. The error distribution of two methods can be seen in Fig7.
5. Conclusion
This work presents a GRU-based E-D neural network for CHP heat load forecast. The key idea in this framework is firstly proposing GRU-based E-D architecture as an auto-encoder to extract the feature of the heat load time series to reduce the model’s misspecification risk. Then, the extracted fixed-length intermediate state from AE combined with future weather data are fed into another multi-layered GRU network to make the forecast. Comparison results demonstrate that the proposed model can more effectively forecast the heat loads over a long horizon. Further, we plan on using other deep learning algorithms, activation functions and optimization methods to improve the performance of the model.

References
[1] Zhang Yu, E Zhijun, Liu Weiping, et al. Research and Application of Peak-shaving Capacity of Cogeneration Units[J]. Electric Power Automation Equipment, 2013, 25(3): 115-118.
[2] Kawashima, Minoru, C. E. Dorgan, and J. W. Mitchell. Hourly thermal load prediction for the next 24 hours by ARIMA, EWMA, LR and an artificial neural network[J]. Ashrae Transactions 101.1(1995):186-200.
[3] Dai Jia-yuan, Wang Lin and Zhang You-wei. Application and Research of Prediction Technology for Extremely Short Term Heat Supply[J]. Jiangsu Electrical Engineering, 2014, 5: 74-76.
[4] Christiaanse W R. Short-Term Load Forecasting Using General Exponential Smoothing[J]. Power Apparatus & Systems IEEE Transactions on, 1971, PAS-90(2): 900-911.
[5] Weigui Q I, Zhu X L. A Study of Time Series Prediction Method in Heat Supply Control[J].

Figure 6. 24 Hour-ahead Heat Load Forecast

Figure 7. 24 Hour-ahead Heat Load Forecast Error Distribution
Acta Electronica Sinica, 2003, 31(2): 268-270.

[6] Wei Yanbao, Lin Hongquan, Ma Zengliang, et al. Central heating system load prediction based on ARMAX model[J]. Automation & Instrumentation, 2014, 7, 1-4.

[7] Zhang Jiao, Tian Qi, Wang Meiping. Heating Load Prediction for Heating Systems Based on Support Vector Regression with Cross Validation[J]. Journal of North University of China. Natural science edition, 2014, 35(5): 565-570.

[8] Zhang Jingbo, Guo Lin, Wang chaonia, et al. Thermal Load Forecasting Model of Heating System Based on Genetic Algorithm Optimization BP Neural Network[J]. Journal of Ordnance Equipment Engineering, 2014, 35(4): 152-156.

[9] Li Qi, Zhou Peng. Short-term Heat Load Forecast of Heat-supply Network Based on Improved Neural Network[J]. Control Engineering of China, 2014, 2(4): 506-509.

[10] Burges, Christopher JC, and Bernhard Schölkopf. Improving the accuracy and speed of support vector machines[J]. In Advances in neural information processing systems, 1997: 375-381.

[11] Boulangerlewandowski N, Bengio Y, Vincent P. Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription[J]. Chemistry A European Journal, 2012, 18(13):3981-3991.

[12] Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures[C]. International Conference on International Conference on Machine Learning. JMLR.org, 2015:2342-2350.

[13] Zaremba, W., Sutskever, I., and Vinyals, O. Recurrent Neural Network Regularization. arXiv preprint arXiv:1409.2329, 2014.

[14] Graves A, Jaitly N, Mohamed A R. Hybrid speech recognition with Deep Bidirectional LSTM[C]. Automatic Speech Recognition and Understanding. IEEE, 2014:273-278.

[15] Yongqing Zhu, Xiangyang Li, Xue Li, Jian Sun, Xinhang Song, and Shuqiang Jiang. Joint Learning of CNN and LSTM for Image Captioning[J]. In CLEF (Working Notes), 2016: 421-427.

[16] Marino D L, Amarasinghe K, Manic M. Building Energy Load Forecasting using Deep Neural Networks[J]. 2016: 7046-7051.

[17] Zheng H T, Yuan J, Chen L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation[J]. Energies, 2017, 10(8):1168.

[18] Kuan Lu, Zhenfu Bi and Xin Wang, et al. Short-term CHP heat load forecast method based on concatenated LSTMs[J]. Chinese Automation Congress (CAC), 2017.

[19] Karlik B, Olgac A V. Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks[M]. Quantum theory as an emergent phenomenon: Cambridge University Press, 2010: S42.

[20] Chung J, Gulcehre C, Cho K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. Eprint Arxiv, 2014.

[21] Trofimovich, J. Comparison of neural network architectures for sentiment analysis of russian tweets[J]. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2016, RGGU. 2016.

[22] Kuan Lu, Yan Zhao and Xin Wang. Short-term electricity load forecasting method based on multilayered self-normalizing GRU network[J]. Energy Internet and Energy System Integration (EI2), 2017.

[23] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning[M]. Cambridge, Massachusetts: MIT Press, 2016.

[24] Zhu L, Laptev N. Deep and Confident Prediction for Time Series at Uber[C]. IEEE International Conference on Data Mining Workshops. IEEE Computer Society, 2017:103-110.