Short-Term Load Forecasting Using Modified Sequence-to-Sequence Deep Learning Framework

Jun Luo*, Keyu Zhu and Xiaojia Wang
School of Management, Hefei University of Technology, Hefei, Anhui, China, 230009

*Corresponding author

Abstract. The future of smart grids promises unprecedented flexibility in energy management. Therefore, it is crucial to accurately predict the energy load demand of individual power grid stations and overall levels. In this paper, a Sequence to Sequence learning method based on long short-term memory neural network (Seq2Seq) forecast model is proposed to predict the load of the community microgrid. In particular, this paper studies the effectiveness of Seq2Seq in community microgrid load forecasting, avoiding overfitting through weight attenuation regularization, and reducing the impact of interference and noise on training data. The proposed method is implemented on the base load data set of multi-user. The results of Seq2Seq were compared with the results of two well-known methods, one is traditional LSTM and the other is Support Vector Machine (SVM) on the same datasets. Experimental results show that compared with SVM and traditional LSTM, Seq2Seq has better learning effect. Finally, we calculate the performance index. When using load data to train and test the model, it was noted that compared with traditional LSTM and SVM, the root mean square error (RMSE) of Seq2Seq was decreased by 44.07% and 64.06%, respectively. Our research offers a decent reference for the application of deep learning in the field of energy management.

Keywords: Load forecasting; Seq2Seq; long-short-term memory; deep learning.

1. Introduction

Energy and environment are two major challenges for global supportable development. In most nation state, power supply is still mainly dependent on the burning of traditional fossil fuels, which causes serious environmental and sustainable development problems. In the meantime, with the rapid and sustained growth of population and the speedy progress of economy, the electricity load has also increased significantly. In particular, the electricity load of residents has shown an increasing trend. Guaranteeing sustainability requires more well-organized energy management while minimizing energy waste. Therefore, the future microgrid should provide extraordinary flexibility in energy management and efficient delivery of power generation[1]. Therefore, intelligent control decision system is necessary for the future power grid. It is imperative to accurately predict future demand[2].

This paper explores a deep learning method for energy load forecasting at the micro-community level. In this method, the improved Sequence to Sequence Learning method based on LSTM is used to predict. In this work, multiple RNN layers are used for past load data before presenting the final regression task. The proposed load forecasting method based on Seq2Seq is tested on a benchmark dataset, which contains 101 residential users' load data with a time resolution of 15 minutes. To compare the results with previous deep learning efforts, we used the same dataset. In order to test performance, the same prediction process was performed using traditional LSTM and SVM. Therefore, the performance of Seq2Seq is compared with that of SVM and LSTM.
2. A Sequence to Sequence Learning Method Model

2.1. Model Input

For time-scale variables, Select 24 hours as the input reference sequence length. This is because the conventional pattern of loads corresponds to 24 hours[3]. The resident electricity load data comes from Dataport Web, which provides a huge database of raw and collated data, including efficacy market operation data and consumer behavior research data[4]. In this study, the power load of residential buildings is selected randomly as the research object. These houses are located in Mueller, Austin, Texas, including room details in single-family houses. The power load distribution map with 15- minutes resolution from March 31, 2019 to July 15, 2019 is used in the experiment. Training data corresponding to the morning and afternoon of March 31, 2019 and July 7, 2019. The detection data corresponding to the data above 00:00 a.m. on July 8, 2019.

2.2. Model Structure.

The LSTM is deliberately designed to avoid long-term dependency problems. Keep in mind that long-term information is the default behavior of the LSTM in practice, not a costly capability[5].

Primary, the first step of the LSTM needs to decide what information we need to discard from the cell. This determination is made from the forgot gate in sigmoid. Its inputs are $h_{t-1}$ and $x_t$, and its output $f_t$ is

$$ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). $$

where $W_f$ is weight parameter, $\sigma$ is activation function, $b_f$ is bias parameter, and $h_{t-1}$ is the output from prior LSTM unit.

Next, we need to decide what information should be stored. There are two main steps to this process. First, the sigmoid layer (input gate) determines which values we need to update. Subsequently, tanh layer generates a new candidate vector $C_t$, it can add to the state. Finally, we combine these two values and update the status of the cell.

$$ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). $$

$$ \tilde{C}_t = \tanh(W_c \cdot [h_t, x_t] + b_c). $$

$$ C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t. $$

In (2)(3)(4), $W_i$ and $W_c$ are weight parameters, $b_i$ and $b_c$ are bias parameters, $i_t$ is the vector that establish what value will be modernized, $C_{t-1}$ is the cell state of previous LSTM Unit.

Finally, we need to determine the output $h_t$. This output depends on the state of our cell, expressed as:

$$ o_t = \sigma(W_o [h_{t-1}, x_t] + b_o). $$

$$ h_t = o_t \otimes \tanh(C_t). $$

where $W_o$ is weight parameter, $b_o$ is bias parameter, $o_t$ is a section of cell state that will be output.

Residential power load reveals the process state of domiciliary electrical utilisations, with high real-time and instability, making the power load curve is fleeting and nonlinear. [6] Hence, the predicting model should know how to simulate the nonlinearity and regularity of power load distribution, and consider the time-varying of data.
LSTM has limitations in processing secure length sequence data. To overcome these challenges, we propose Seq2Seq model. Seq2Seq model is propositioned to solve the difficulty of flexible length sequence data.

Sequence to Sequence (Seq2Seq) model is an architecture represented by Encode and Decode. Seq2Seq model generates output sequence \( Y \) according to input sequence \( X \). The classical RNN structure of the input and output sequence must be equal length, its application scenarios are also limited, Seq2Seq to achieve a sequence to another sequence of the conversion, the typical idea is to use two LSTM, one LSTM as encoder, another LSTM as a decoder.

Assume a task, then there are tasks that can be split into \( x_1, \ldots, x_T \) sequences. The hidden state of the next moment can be expressed as

\[
\begin{align*}
h_t &= f(x_t, h_{t-1}) \\
f &= \text{the transformation function of the hidden layer of circular network.}
\end{align*}
\]  

Then we define a function \( q \) that turns the hidden state of each time step into a background vector:

\[
\begin{align*}
c &= q(h_1, \ldots, h_T) \\
q &= \text{function that turns the hidden state of each time step into a background vector.}
\end{align*}
\]  

The previous encoder encodes the information of the entire input sequence into the background vector \( c \) and the decoder outputs the sequence as \( y_1, y_2, \ldots, y_T \). The output of each step of the decoder is based on the output and background vector of the previous step, so it is expressed as:

\[
\begin{align*}
p(y_t | y_{1:t-1}, c) &= g(y_{t-1}, c, s_{t-1}) \\
g &= \text{function of this circular network.}
\end{align*}
\]  

In general, the decoder will also be a circular network. We use \( g \) to represent the function of this circular network, so the hidden state of the current step is

\[
\begin{align*}
s_t &= g(y_t, c, s_{t-1}) \\
g &= \text{function of this circular network.}
\end{align*}
\]  

The prediction model is a hybrid model composed of input layer, RNN-LSTM layer, overall insight layer and output layer. Practice has demonstrated that the hybrid model proposed by us is often better than the non-hybrid model[7]. The model uses six layers network structure of input layer, two RNN-LSTM layers, two simple hidden layers and one output layer to achieve the best forecasting effect for residential power load. The structure of Seq2Seq model is defined as follows:

1. Layer 1: Introduce a 15-minutes resolution input to Layer 1. Each circle characterizes an input, embracing timetable variables and current load. The training data covers \( S \) samples. \( S \) is the numeral of 24-h arrangements.
2. Layers 2 and 3: They are two RNN-LSTM layers, each circle representing an output vector.
3. Layers 4, 5 and 6: Layers 4 and 5 are two simple hidden layers, which correspond to a multi-layer feedforward network.

At the same time, in order to eradicate the inspiration of measurement and speed up the training, we also use the min-max standardization method to alter the training data into scalar values, use the weight attenuation regularization to avoid over-fitting. Adaptively regulate the learning capability of the neural network model through Dropout to enhance the robustness of the model and use the Adam algorithm to deal with the big data set and multi-parameter problems, so that the model has a faster convergence speed[8].

2.3. Model evaluation

In this article, we use Mean absolute percentage error (MAPE), Root Mean Square Error (RMSE) and Average Absolute Error (MAE), which are often used as criteria to measure the predicted results of deep learning models, to evaluate the experimental results. The related
calculations are shown in formulas (10) to (12).

\[
\text{MAPE} = \frac{100\%}{T} \sum_{t=1}^{T} \left| \frac{y_a - y_p}{y_a} \right|, \quad (10)
\]

\[
\text{MSE} = \frac{\sum_{t=1}^{T} (y_a - y_p)^2}{T}, \quad (11)
\]

\[
\text{MAE} = \frac{\sum_{t=1}^{T} |y_a - y_p|}{T}. \quad (12)
\]

Where \( y_a \) is the actual value of the test stage, \( y_p \) is the corresponding predicted value, and \( T \) is the total time step.

3. Results and Discussions

3.1. Results Analysis

Figure 1 show the overall training process of the Seq2Seq model from March 31, 2019 to July 15, 2019 and the comparison of some predicted results with real data respectively, it can be perceived that the fluctuation of power load is large. Yet, the power load outline does not show a significant diurnal trend. This is because residents’ electricity load situation has a high randomness and is affected by environmental interference.

![Figure 1](image1.png)

Figure 1. The training and testing procedure of Seq2Seq model for total residential electricity load data from March 31, 2019 to July 15, 2019 and July 8, 2019 to July 15, 2019 are used.

From Figure 2, in order to visually and clearly show the experimental results, we continue to refine the particle size and select the most representative results of the last three days as the demonstration and comparison. It can be seen from the figure that the prediction curve is basically consistent with the actual load curve.

![Figure 2](image2.png)

Figure 2. Forecasting results of aggregated residential load on July 13, 2019-July 15, 2019.

Figure 3 show the detailed forecast curve of July 13, 14 and 15, 2019 in more detail. Overall, the figures show the close relationship between the predicted value and the actual load demand, as well as the high accuracy of the model for short-term load forecasting. From the figures, we can see that the fitting effect of LSTM model is better than that of SVM model, but both of them are obviously inferior to that of Seq2Seq model. In the area of greater flexibility, the prediction effect of LSTM model and SVM model are not satisfactory.
Figure 3. Forecasting results of comparison method on July 13, 2019-July 15, 2019 (left: SVM; right: LSTM).

By using the traditional LSTM and Support Vector Machine (SVM) to verify the prediction performance of the comprehensive load of residential buildings. We also compare the performance of the model by evaluating indicators. The table shows the relative performance comparison of Seq2Seq, LSTM and SVM in predicting the comprehensive residential power load.

Table 1. The evaluation index values of Seq2Seq, LSTM and SVM in predicting housing integrated electricity load.

| Models    | Seq2Seq | LSTM  | SVM   |
|-----------|---------|-------|-------|
| MAPE      | 7.97%   | 11.76%| 30.47%|
| RMSE      | 3.145   | 5.623 | 8.75  |
| MAE       | 0.487   | 4.961 | 5.79  |

3.2. Discussions and Conclusion

As can be seen from Figure 2 to 3, LSTM is superior to SVM in predicting the total residential load, but still not as good as the proposed Seq2Seq model. Table shows that the RMSE of Seq2Seq model are 3.145 in the test phase, which are memorably smaller than the errors of LSTM and SVM. The MAPE of Seq2Seq model reached 7.97%, which was smaller than LSTM (11.76%) and SVM (30.47%). This is owing to the high haphazardness and real-time of electricity load, which leads to the non-linear and non-periodic characteristics of power load profile.

Based on the above, we propose an improved Seq2Seq model and improve the accuracy of prediction from two aspects.

(i) Reduce the computational burden of processing high-dimensional input data, through structured selection of input subset, reduce the data dimension.

(ii) To remove the false and retain the true, the task processing system is more absorbed on finding the significant useful information connected to the present output in the input data, so refining the quality of the output. Through context vectors, the decoder can query the most relevant source information at each step of decoding, thus avoiding information bottlenecks in Seq2Seq model.

Therefore, the model can obtain useful information between input data and output data more effectively, so as to generate more reasonable output. The accuracy of the results obtained is much higher than that
of SVM and LSTM.

References

[1] Kang, C. and Q.J.A.o.E.P.S. Xia, REVIEW OF POWER SYSTEM LOAD FORECASTING AND ITS DEVELOPMENT. 2004. 28(17): p. 1-11.

[2] Tao, H. and F.J.I.J.o.F. Shu, Probabilistic electric load forecasting: A tutorial review. 2016. 32(3): p. 914-938.

[3] Heghedus, C., A. Chakravorty, and C. Rong. Energy Load Forecasting Using Deep Learning. in 2018 IEEE International Conference on Energy Internet (ICEI). 2018.

[4] Dataport, https://dataport.cloud. Pecan Street Inc, 2019. [accessed 25 July 2019]

[5] Chen, L., et al., Application of LSTM Networks in Short-Term Power Load Forecasting Under the Deep Learning Framework. 2017.

[6] Shi, Z., L. Hao, and V.J.I.T.o.S.E. Dinavahi, Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. 2017. PP (99): p. 1-1.

[7] Jarábek, T., P. Laurinec, and M. Lucká. Energy load forecast using S2S deep neural networks with k-Shape clustering. in 2017 IEEE 14th International Scientific Conference on Informatics. 2018.

[8] Yu, C.N., P. Mirowski, and T.K.J.I.T.o.S.G. Ho, A Sparse Coding Approach to Household Electricity Demand Forecasting in Smart Grids. 2017. 8(2): p. 738-748.