A Regional Integrated Energy System Load Prediction Method Based on Bayesian Optimized Long-Short Term Memory Neural Network

Ang Xuan and Shihao Tian

A Regional Integrated Energy System Load Prediction Method Based on Bayesian Optimized Long-Short Term Memory Neural Network

1st Ang Xuan
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
Shenzhen, China
xa19@mails.tsinghua.edu.cn

2nd Shihao Tian*
Department of Electrical and Computing Engineering
Cornell University
Ithaca, New York, U.S.A.
st689@cornell.edu

Abstract—**In the face of the rapid growth and development of regional integrated energy system (RIES) globally, accurate load prediction technique is increasingly playing a critical role in RIES planning. This paper presents a Bayesian Optimized Long Short-Term Memory (BO-LSTM) neural network to predict the electric, heating and cooling power load for the short and mid-term operation. The Bayesian optimization algorithm is performed to automate hyperparameter tuning to improve results, so avoiding different hyperparameters may lead to considerable differences in the performance of other deep learning network architecture in some sense. The developed model is validated on one actual RIES in China for data collected in a year. The simulation results of the proposed BO-LSTM indicate the effectiveness and excellent prediction accuracy in comparison with other traditional models, such as autoregressive integrated moving average model (ARIMA), long short-term memory (LSTM) and convolutional neural network (CNN).**

Keywords—Bayesian Optimization, long short-term memory, load prediction, deep learning, regional integrated energy system

I. INTRODUCTION

With the rapid socio-economic advancement, the synergistic development of energy use and environmental protection faces serious challenges, and Regional Integrated Energy Systems (RIESs) are considered one of the most promising solutions for future energy systems [1]. Nowadays, many countries, regions and communities have turned their attention to RIESs because they can realize the complementary use of multiple energy sources such as electricity, gas, cooling and heat [2]. Therefore, multi-energy load prediction provides the necessary data to support the planning and operation of RIESs and is of great relevance.

Many studies on load prediction had constructed accurate and effective model architectures, classified as statistical, machine-learning based and deep-learning based. The literature [3] provided a method for predicting next-day electricity prices based on the ARIMA method. The literature [4] proposed an adaptive forecasting method for short-term heating loads based on a linear regression model. The literature [5] proposed a load forecasting model supporting vector regression for office building electrical loads. Meanwhile, so far, deep learning has gradually become the most promising technique for predicting loads in machine learning. The literature [6] used automatic coding techniques to compress historical data and used multilayer GRU to construct a model for predicting electrical loads. The literature [7] proposed a heating load prediction model based on temporal CNN, which implemented the parallel feature processing of CNN capability and the time-domain modeling capability of Recurrent Neural Network (RNN). The literature [8] combined the extraction potential of CNNs with the effective gating structure of multilayer GRUs to form a hybrid model. Similar applications could be found in other artificial neural networks [9]-[10].

The LSTM neural network is an improved RNN and is the most efficient and scalable model among all neural networks in time-series data due to its unique structure avoiding the gradient vanishing and explosion drawbacks of RNNs [11]. The largest experimental runs on LSTM networks [12] showed that none of the corresponding variants could significantly improve the performance of the LSTM and derived a criterion for hyperparameter tuning.

However, LSTM still lacks long-term training effects because it utilizes different modular structures to preserve past data sequences in temporal datasets. LSTM includes various hyperparameters that affect the model architecture, and the prediction performance is highly dependent on these values, which are tedious to achieve optimality. Therefore, in this paper, a Bayesian optimization algorithm is chosen to find more accurate hyperparameters, maintain and update the Gaussian process model with the lowest loss objective function, and find the optimal hyperparameters for the network.

II. PROPOSED FRAMEWORK

In this section, the proposed model framework is described. First, the main components of LSTM and Bayesian optimization theorem are briefly described. In the following subsections, the LSTM module and the Bayesian optimization algorithm are further elaborated.

A. LSTM model

Among the current neural network models, Supported Vector Machine (SVM) is a supervised learning model for linear regression analysis that dismisses the retention of previous data; Artificial Neural Network (ANN) is a feed-forward model that propagates temporal information that loses quickly; CNN is a class of neural networks that are usually applied to analyze visual images, and the method incorporated a shared weight architecture of convolutional kernels and filters that slide along the input features. Time series prediction requires effective

evaluation of past data; thus, LSTM is the most suitable candidate for the above task due to its model structure.

$$H^t = \Theta(X^t) \tag{1}$$

The compact form of LSTM is listed in (1), where H^t denotes the output vector of the initial input vector X^t. Its neural network comprises several modules to create a stable memory sequence to perform better in the longer term. The following

Fig.2 and formula (2)-(9) specifically analyze the internal structure of an LSTM module.

$$f^t = \sigma(W_f \cdot [H^{t-1}, X^t] + b_f) \tag{2}$$

$$i^t = \sigma(W_i \cdot [H^{t-1}, X^t] + b_i) \tag{3}$$

$$C^t = \tanh(W_c \cdot [H^{t-1}, X^t] + b_c) \tag{4}$$

$$o^t = \sigma(W_o \cdot [H^{t-1}, X^t] + o_i) \tag{5}$$

$$C^t = f^t * C^{t-1} + i^t * C^t \tag{6}$$

$$H^t = o^t * \tanh(C^t) \tag{7}$$

$$sigmond(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{9}$$

For a specific time step, the LSTM module takes an input vector X^t, state vector H^{t-1} of time (t-1) and H^t of time t. b_f, b_i, b_c, o_i denote bias in forget gate, input gate, cell state and output gate, W_f, W_i, W_c, W_o depict the corresponding weights, C signifies the cell state vector, which is instinct parameter matrix of LSTM cell. Among the four function areas represented by the dotted frame in Fig.1, forget gate $\sigma(W_f, b_f)$ removes information from the last network, input gate $\sigma(W_i, b_i)$ stores information into the internal state, the input node $\tanh(W_c, b_c)$ (a tanh layer) creates a new state vector C^t, the output gate $\sigma(W_o, o_i)$ determines output information to convey in the LSTM.

All the gates are sigmoidal units with outputs 0 or 1, where 0 means to remove the information and 1 means to retain. Cell state can be calculated with a new state vector C^t as (6) to get the internal state vector C^t. Expression (7) illustrated that the internal state vector C^t is transformed through the tanh layer and then multiplied o^t to obtain the output or next hidden state.
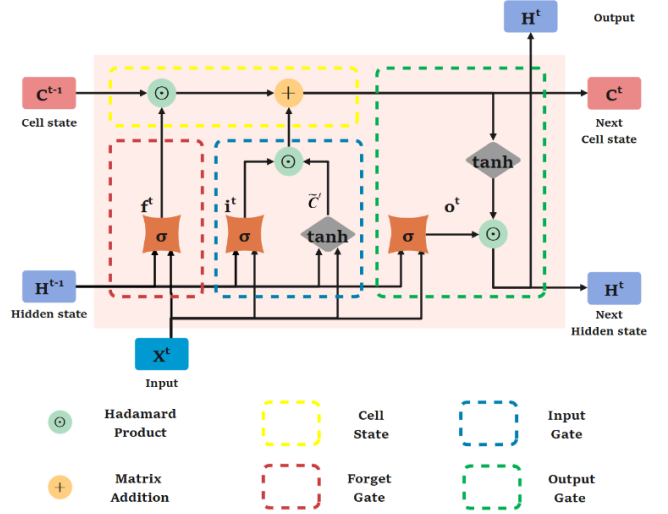

Fig.2 One naïve LSTM module architecture

B. Bayesian optimization theorem

Deep learning models containing LSTMs have several hyperparameters tuned, i.e., empirically or by random attempts to guarantee a perfect fit, making the model more of an art than an exact science. Hyperparameter tuning of deep learning models is generally considered to be a black-box optimization problem. Researchers can only see the inputs and outputs of the model during the tuning process and do not have access to gradient information, nor can they determine whether the hyperparameters of the model meet the convexity conditions. Therefore, there must be accurate and effective methods to tune the hyperparameters. Currently, grid search, stochastic search, and Bayesian optimization [13] are common hyperparameter tuning techniques used in academia and industry.

Bayesian optimization is a powerful sequential strategy for global optimization of black-box functions, which does not assume any functional form and the objective function is estimated as a Gaussian process and interpreted as an agent form [14]. Sequential model-based optimization (SMBO) is the simplest form of Bayesian optimization, and its algorithmic procedure is as follows.

Assuming $X = \{x_1, x_2, \cdots, x_n\}$ is the hyperparameter search space (x_n represents the value of a certain set hyperparameter), f is a black-box objective function taking in a set of hyperparameters $\{x_1, x_2, \cdots, x_n\}$ and getting an output value $\{x_1^*, x_2^*, \cdots, x_n^*\}$.

$$x^* = \underset{x \in X}{\arg\min} f(x) \tag{10}$$

The pseudo-code for finding the minimum f(x) is shown in Listing 1. D represents a dataset composed of several arrays, each pair of arrays is represented as (X, y), X is a set of hyperparameters, and y represents the result corresponding to X. S is the Acquisition Function to select x in formula (10), M is the model type to fit the dataset D, usually takes Gaussian model in Bayesian optimization.

LISTING 1 SEQUENTIAL MODEL-BASED OPTIMIZATION ALGORITHM

Algorithm: Sequential model-based optimization (SMBO)

Stage 1: Input

Input X, f, S, M .

Stage 2: Initsamples

$D_{1:n} \leftarrow \text{INITSAMPLES}(f, X_{1:n})$

Stage 3: Judgment

for $i \leftarrow 1$ to n in $D_{1:n}$

$p(y|x, D_i) \leftarrow \text{FITMODEL}(M, D_i)$

$x_i \leftarrow \underset{x \in X}{\arg\min} S(x, p(y|x, D))$

$y_i \leftarrow f(x_i)$

$D_{i+1} \leftarrow D_i \cup (x_i, y_i)$

end for

The Bayesian strategy treats the objective function as a random function and places a prior on it. After the function is evaluated, the prior is updated to form a posterior distribution, which in turn is used to construct an acquisition function that determines the following hyperparameter combination. The hyperparameters of the model generate an objective function that minimizes the target output.

TABLE I shows the search space for the hyperparameters of the Bayesian optimization algorithm.

TABLE I SEARCH SPACE OF HYPERPARAMETERS

Hyperparameter	Search space
Number of LSTM cells	4-512
Activation function	ReLU, Linear, Sigmoid, Tanh, ELU
Optimization method	SGD, Adam, Nadam, Adamax, Adadelta, Adagrad, RMPSprop
Neurons in hidden layer	4-512
Dropout rate	0-0.8
Batch size	4-256
Epochs	5-100

III. DATA DESCRIPTION

The dataset used in this study is available in [15], which is the annual load data of a RIES in China. The whole dataset is the hourly load data throughout one year (8760h in 2017) including electricity load, heating load and cooling load.

A. Data preprocessing

Normalization, standardization and regularization are commonly used data preprocessing methods. Since the original data in this research belong to the same feature (load), standardization is a more appropriate data preprocessing method to accelerate convergence. The input standardization is conducted using the equation:

$$x_{new} = \frac{x - \mu}{\sigma} \qquad (11)$$

Where μ denotes input data mean, σ denotes input data standard deviation, x and x_{new} denote data actual value and standardized value.

Therefore, the data needs to be sequenced to the size of a 3-dimensional array to feed into the hidden layer, and the dataset is divided into training and test sets using a 10 k cross-validation method to avoid over-fitting.

B. Performance evaluation indices

In this study, mean absolute percentage error (MAPE), root mean squared error (RMSE) and Correlation coefficient (R) are adopted to evaluate the performance of predictions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \qquad (12)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} *100\% \qquad (13)$$

$$R = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (14)$$

Where n is the total number of observations, $\hat{y}$ denotes predicted value and y denotes the actual true value of tested data.

IV. CASE STUDY

In the case study, three cases were used to test the model. Case 1 (short-term prediction): the first 90% of 100 hours of load data was used as the training set and the last 10% (10 hours) was selected as the test set; Case 2 (medium-term prediction): the first 90% of 500 hours of load data was used as the training set and the last 10% (50 hours) was selected as the test set; Case 3 (long-term prediction): the first 90% of 2000 hours of load data was used as the training set and the last 10% (200 hours) as the test set.

In this section, the effectiveness of BO-LSTM is verified and then the performance of the proposed model is evaluated and compared with the performance of other existing prediction models such as ARIMA, LSTM and CNN. All experiments are performed in MATLAB 2020b.

A. The effectiveness of BO-LSTM

For simplicity, we have chosen to show the results for electrical loads. Fig.3 depicts the observed and estimated objective functions for the Bayesian optimization process during 100 iterations in Case 3. The fitted objective function rapidly approaches the observed objective function through the Bayesian optimization process. At the 31st iteration, the optimized objective function reached its minimum value, and in the following 70 iterations, the two objective functions were very close to each other, maintaining a difference of less than 0.001. Overall, the Bayesian optimization process of the objective function ensures the optimality of the hyperparameters.
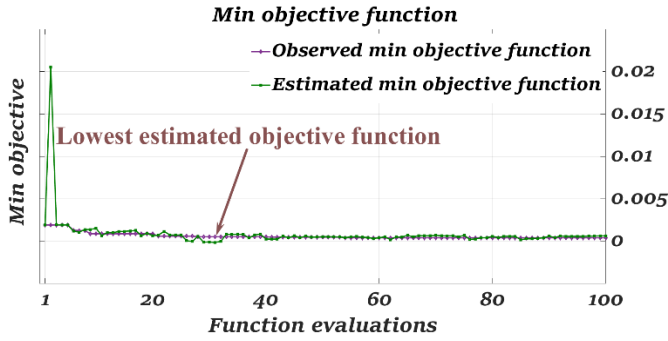
Fig.4 Observed and estimated objective function
in Bayesian optimization process in Case 3

Fig.5 shows the relative errors for Case 3, which further illustrates the effectiveness of the forecast method. In the 200-hour forecast results, the relative error is consistently below ±4% and evenly distributed, which is in line with the current international standard for medium-term forecasting (<5%).
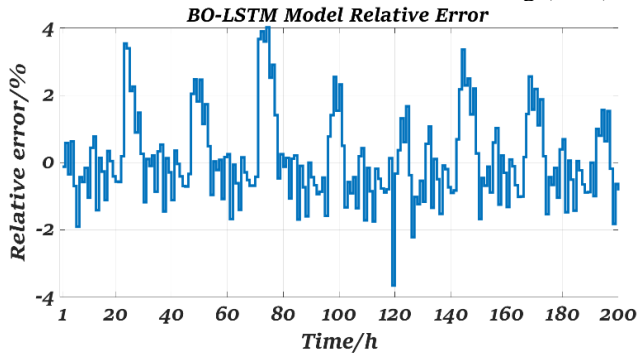

Fig.5 BO-LSTM model relative error in Case 3
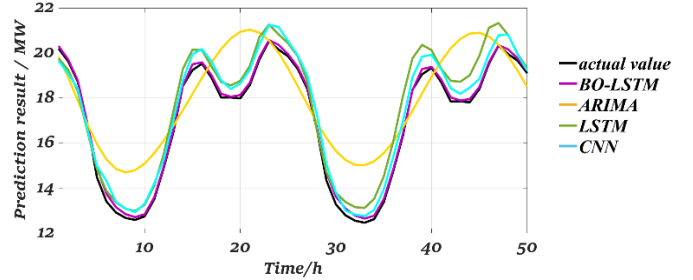
B. Performance comparison

Fig.6 represents the prediction results of the four models from Case 1 to Case 3. BO-LSTM always performs well for short-term and medium-term predictions. The ARIMA model requires time-series data to be stationary, and essentially it can only capture linear relationships, not nonlinear ones. Therefore, whenever the data fluctuates slightly, its performance gets worse. Moreover, the ARIMA model is only effective for short-term prediction, which is determined by its principle. When used for medium-term and long-term prediction, its results are distant from the actual values that making it worthless.

The LSTM model is currently a commonly used model for short-term and medium-term prediction, and it has outstanding results. However, when it is used for long-term prediction, its drawbacks gradually emerge: the cumulative effect of errors caused by the unique structure of the LSTM model makes it incapable of making the long-term prediction.

CNN is widely used for image recognition and processing. The feasibility of prediction is also shown in this paper due to its unique convolutional kernel structure that can mine the potential information of the data. However, because the hyperparameters are constant, the error gradually expands when CNNs are used for long-term prediction (see in Fig. 6 (c)).
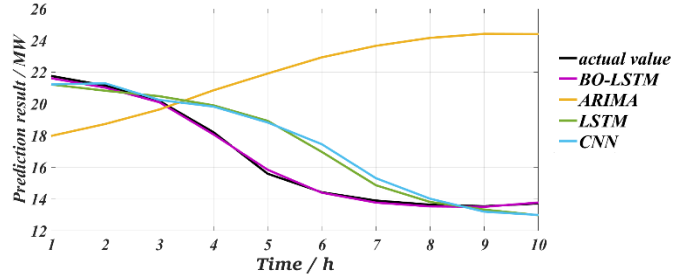
(a) Case 1 (short-term prediction)

(b) Case 2 (medium-term prediction)
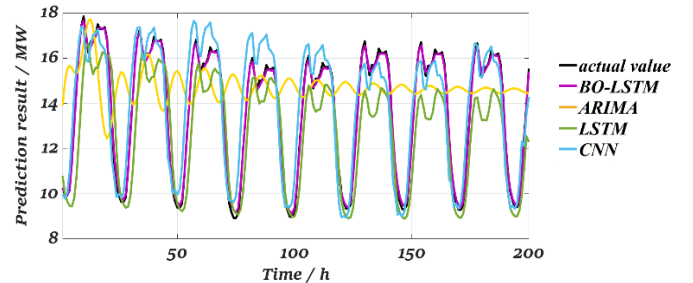
(c) Case 3 (long-term prediction)
Fig. 6 Prediction Results Comparison

In TABLE II, R, MAPE and RMSE are displayed to compare the performance among our proposed model and other established classic models, such as ARIMA, LSTM and CNN. It can be seen that the ARIMA model has gradually become unsuitable for data prediction problems due to its narrow application and simple principle. The accuracy (MAPE, RMAE) and correlation (R) of the LSTM model gradually deteriorate as the prediction time expands. Similarly, CNN models also face these problems. Accordingly, the BO model always shows excellent accuracy and correlation. The case results demonstrate that the accuracy of the data prediction method is significantly improved after introducing the automatic hyperparameter adjustment mechanism.

TABLE II Model Performance Comparison
(a) Case 1 (short-term prediction)

	ARIMA	LSTM	CNN	BO-LSTM
R	-0.947	0.99953	0.99064	0.9986
MAPE	39.278%	-0.12395%	2.0889%	0.124%
RMSE	21.989	0.12986	0.44749	0.0438

(b) Case 2 (medium-term prediction)

	ARIMA	LSTM	CNN	BO-LSTM
R	0.7473	0.99009	0.98794	0.9990
MAPE	10.12%	1.1673%	0.78641%	0.3214%
RMSE	18.119	0.43871	0.17957	0.1203

(c) Case 3 (long-term prediction)

	ARIMA	LSTM	CNN	BO-LSTM
R	$-6.9*10^{-4}$	0.91044	0.99473	0.9992178
MAPE	21.452	-7.6292%	4.7558%	2.5199%
RMSE	14.75	1.6683	0.29195	0.1388

V. CONCLUSION

The BO-LSTM neural network model proposed in this paper demonstrates its effectiveness in long-term prediction. The case study results show stable performance and the ability to capture anomalous trends in data deviations. Moreover, BO-LSTM is compared with other existing models and it outperforms them in terms of MAPE, RMSE and R at different time intervals, demonstrating its superiority. Bayesian optimization helps to tune the hyperparameters to an optimal level, rather than through the randomness and uncertainty of manual tuning. In this sense, the inclusion of adaptive optimization mechanisms in the tuning of machine learning hyperparameters is a future development direction.

Future work will focus on the following areas. 1). Building hybrid models by combining with other neural networks to improve the prediction performance further; 2). Incorporating feature selection techniques into the models; 3). Develop multi-task learning to handle coupled multi-dimensional data.

REFERENCES

[1] Mancarella, Pierluigi. "MES (multi-energy systems): An overview of concepts and evaluation models." Energy, vol.65, 2014, pp. 1-17. doi:10.1016/j.energy.2013.10.041

[2] Wang J, Zhong H, Ma Z, et al. Review and prospect of integrated demand response in the multi-energy system[J]. Applied Energy, 2017, 202: 772-782.

[3] J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, "ARIMA models to predict next-day electricity prices," in IEEE Transactions on Power Systems, vol. 18, no. 3, pp. 1014-1020, Aug. 2003, doi: 10.1109/TPWRS.2002.804943.

[4] Nigitz, Thomas, and Markus Gölles. "A generally applicable, simple and adaptive forecasting method for the short-term heat load of consumers." Applied Energy 241 (2019): 73-81.

[5] Chen, Yongbao, et al. "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings." Applied Energy 195 (2017): 659-670.

[6] Ke, Kang, et al. "Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network." Evolutionary Intelligence 12.3 (2019): 385-394.

[7] Song, Jiancai, et al. "Hourly heat load prediction model based on temporal convolutional neural network." IEEE Access 8 (2020): 16726-16741.

[8] Sajjad, Muhammad, et al. "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting." IEEE Access 8 (2020): 143759-143768.

[9] Chis, Violeta Eugenia, et al. "ANN based Short-Term Load Curve Forecasting." International Journal of Computers Communications & Control 13.6 (2018): 938-955.

[10] Amarasinghe, Kasun, Daniel L. Marino, and Milos Manic. "Deep neural networks for energy load forecasting." 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE). IEEE, 2017.

[11] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[12] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222-2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.

[13] Bergstra, James, et al. "Algorithms for hyper-parameter optimization." Advances in neural information processing systems 24 (2011).

[14] Mockus, Jonas. Bayesian approach to global optimization: theory and applications. Vol. 37. Springer Science & Business Media, 2012.

[15] Xuan, Ang (2021): A Regional Integrated Energy System Load Prediction Method Based on Bayesian Optimized Long-Short Term Memory Neural Network. figshare. Dataset.
https://doi.org/10.6084/m9.figshare.14984247.v1.