SOC Prediction of Lead-acid Battery based on EEMD-LSTM Network

Xinwei Zhang[a], Zhanying Li[b]

Dalian Polytechnic University, Dalian 116000, China

[a]Zxw5118@126.com, [b]l_zy1979@163.com

Abstract

Lead-acid battery is widely used because of its reliable safety. The state of charge (SOC) of the battery is the most direct reflection of the battery's use status. Its accurate estimation helps users to formulate a reasonable battery management plan and maximize the safe use of the battery. In this paper, a data-driven method is proposed to predict the SOC of the battery. The real-time online data of the lead-acid battery cloud data management system of a technology limited company is used, and the isolated forest algorithm is used to deal with the abnormal data. It is proposed to build a deep learning model combining the set empirical mode decomposition (EEMD) and the short-term memory neural network (LSTM), and compare with various models. At the same time, this paper uses the EEMD algorithm to decompose the time series into several subsequences, which enlarges the details of the time series data, making the fluctuation of the subsequences more stable than the original sequence, and solving the prediction lag problem of LSTM network. Through experimental analysis, the EEMD-LSTM model used in this paper has the best prediction effect and the highest prediction accuracy.

Keywords

Lead Acid Battery; EEMD; Long-term and Short-term Memory Network (LSTM); Life Prediction; Neural Network.

1. Introduction

Lead-acid battery is a very important power battery. It is currently widely used in various fields such as communication, power, power, energy storage and emergency.] Especially in the field of automobile starting, lead-acid battery is almost the existence of the strong, while the new energy battery led by lithium battery has a long way to go. Lead-acid battery has mature production technology, high reliability and safety, and plays an important role in all aspects of economic and social development due to its low production cost and strong environmental applicability. [1-3] As a secondary battery, lead-acid battery, such as charging current, voltage, time and other factors will affect the state of charge (SOC) in the process of use. Therefore, it is of great significance for people to accurately predict the SOC of lead-acid battery for rational use of lead-acid battery.

In recent years, with the development of artificial intelligence, a large number of researches have been carried out on the SOC prediction of batteries at home and abroad [4-6], and the research objects are mostly lithium batteries, while the research on lead-acid batteries is less. The prediction methods of battery SOC mainly include physical model and neural network algorithm model [7-9]. Among them, the prediction process of the physical modeling method is complex, and it is difficult to consider the influencing factors fully, resulting in the prediction results are often not accurate enough, and the error is large, while the artificial intelligence network is like a "black box", which can be predicted only by inputting data into the network model, and the prediction accuracy is high.

LSTM neural network is a kind of neural network with memory ability [11]. In recent years, LSTM neural network has received extensive attention in data prediction. At present, there are few cases of using LSTM model to predict lead-acid batteries, and there is a common problem, that is, there is a certain delay in the prediction of SOC, which is intuitively represented by the translation dislocation on the image, and the predicted value is always slightly slower than the true value. The fundamental reason is that LSTM network can not accurately detect the fluctuation of the time series, so that the prediction results of the previous time may be reflected at this time. In this paper, we introduce the Ensemble Empirical Mode Decomposition (EEMD) algorithm, which fundamentally solves this problem. By decomposing a time series into several subsequences, we can better make the fluctuation of subsequences more stable than the original series, obtain the fluctuation of time series, and enlarge the details of time series, This makes the prediction of each LSTM network sub-model more accurate and solves the delay of LSTM network itself. LSTM neural network is a kind of neural network with memory ability [10]. In recent years, LSTM neural network has received extensive attention in data prediction. At present, there are few cases of using LSTM model to predict lead-acid batteries, and there is a common problem, that is, there is a certain delay in the prediction of SOC, which is intuitively represented by the translation dislocation on the image, and the predicted value is always slightly slower than the true value. The fundamental reason is that LSTM network can not accurately detect the fluctuation of the time series, so that the prediction results of the previous time may be reflected at this time. In this paper, we introduce the Ensemble Empirical Mode Decomposition (EEMD) algorithm, which fundamentally solves this problem. By decomposing a time series into several subsequences, we can better make the fluctuation of subsequences more stable than the original series, obtain the fluctuation of time series, and enlarge the details of time series, This makes the prediction of each LSTM network sub-model more accurate and solves the delay of LSTM network itself.

2. Theoretical Basis

2.1 EEMD Method

In 1998, the EMD method, N.E. Huang, proposed a new analytical and processing method for nonlinear signals, empirical mode decomposition (EMD). [11,12] This method is often suitable for analyzing some nonlinear and non-stationary signals. In order to solve the modal mixing problem, Huang E et al. proposed the Ensemble Empirical Mode Decomposition (EEMD), which adds white Gaussian noise satisfying Gaussian distribution to the original signal [13] Then decompose the EMD and repeat the operation until the final result is obtained. Because EEMD not only preserves the decomposition effect of EMD for linear and non-stationary time series data, but also solves the problem of modal aliasing, it has been gradually used in time series in recent years

2.2 Isolation Forest

Isolated forest is a model for detecting outliers in the category of unsupervised learning. It was jointly completed by two professors Fei Tony Liu, Kai Ming Ting and Professor Zhou Zhihua of Nanjing University [14,15]. Just as random forest is composed of a large number of decision trees, iForest forest is also composed of a large number of trees. The tree in iForest is called the isolation tree, which tests every iTree in the iForest. Randomly select points from the training data as sub-samples and put them into the root node of an isolated tree; Randomly specify a dimension, and within the current node data range, randomly generate a cutting point P The cutting point is generated between the maximum and minimum values of the specified dimension in the current node data; The selection of this cutting point generates a hyperplane, which divides the data space of the current node into two subspaces: the points less than the currently selected dimension are placed in the left branch of the current node, and the points greater than or equal to P are placed in the right branch of the current node; Recursive steps continue to construct new leaf nodes until there is only one data on the leaf node (cutting cannot continue) or the tree has grown to the set height.

$$S\ (x,\varphi) = 2^{\frac{E(\mathrm{h}(x))}{c(\varphi)}} \tag{1}$$

When the score is to 1, the more likely it is to be an outlier; And if the score is lower than 0.5, it can be basically determined as normal data;

3. LSTM Neural Network

LSTM neural network is a variant of RNN neural network. The unit state of LSTM will determine which states should be left and which states should be forgotten. At the same time, LSTM can realize the encoding and decoding of time series data. Its design solves the problem of long-time sequence. This method can remember the historical information with very long-time interval.
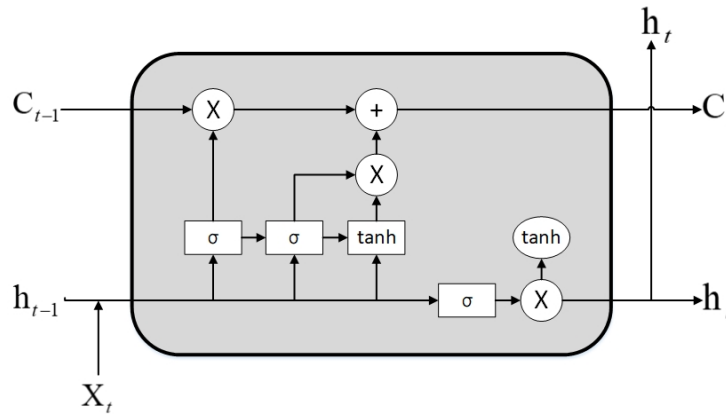
Fig 1. Schematic diagram of LSTM

There are three kinds of protection and control of cell state: forgetting gate, input gate and output gate. [16] This "gate" acts on cells to form the hidden layer of LSTM model. The input gate controls the strength of the new input to the memory cell, the forgotten gate controls the strength of the memory cell to maintain the previous time value, and the output gate controls the strength of the output memory cell. The schematic diagram of LSTM is shown in Fig 1. The specific calculation formula of LSTM is as follows:

$$f_t = \left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

$$i_t = \sigma\left(w_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3}$$

$$\widetilde{C}_t = \tanh\left(w_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

$$O_t = \sigma\left(W_o [h_{t-1}, x_t] + b_o\right) \tag{6}$$

$$h_t = O_t * \tanh(C_t) \tag{7}$$

4. EEMD-LSTM Model

In this paper, an EEMD-LSTM model is proposed to predict the SOC of lead-acid batteries. The specific process is shown in Figure 2. Firstly, the real-time online data of the lead-acid battery pack cloud data management system of a technology limited company is used, and then the data processing is carried out, including data outlier detection. Then, the set empirical mode decomposition model (EEMD) is used for processing, and the complex time series is decomposed into a series of signal IMF with local characteristics and corresponding residual value (Res). For the selected IMF

subsequences, an appropriate LSTM sub-prediction model is established for each subsequence. Finally, the prediction value obtained from LSTM model of subsequence is superimposed. Thus, the final prediction result is the sum of the prediction results of each IMF component. The essence of the prediction model in this paper is to decompose the original time series data by the EEMD method to obtain IMF components that are easy to analyze, and then model, process and predict each component separately. The purpose is to obtain more data features, fully mine the potential information in the data, and improve the prediction of lead acid battery SOC.
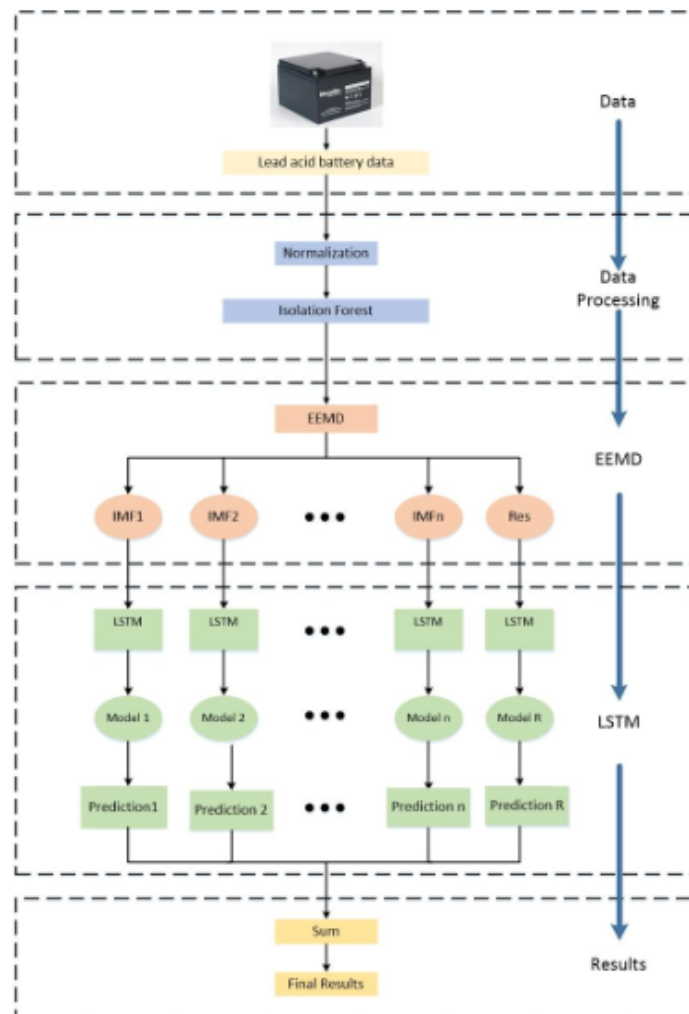
Fig 2. EEMD-LSTM Model Flow Chart

5. Enperiment and Analysis

The source of the data in this paper is the real-time online data of the cloud data management system of the lead-acid battery pack of a technology limited company. A total of 2106 pieces of data were collected in 2021. The data are monitored every hour. There are three characteristic variables, namely, the voltage, temperature and SOC of the lead-acid battery. The change of each characteristic variable can be seen intuitively.

The isolated forest algorithm is not based on distance and density to judge anomalies. It is characterized by small amount of computation, natural distributed training and computation, and is very suitable for massive data scenarios, so it is very suitable for anomaly detection of lead-acid battery time series.A total of about thirty data outliers are found. For the case of abnormal detection,

because the data before and after the change is relatively large, the method of outlier elimination is adopted.

5.1 Evaluating Indicator

There are many evaluation methods for evaluating the good or bad of a model. In order to ensure the accuracy and efficiency of the model, the statistical indicators selected in this paper include RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and the coefficient of determination R^2. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \hat{y}_t\right|}{y_t} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{t=0}^{n-1}\left(y_t - \hat{y}_t\right)^2}{\sum_{t=0}^{n-1}\left(y_t - \bar{y}_t\right)^2} \tag{11}$$

Among, y_t is the actual value of the state of charge of the lead-acid battery, $\hat{y}_t$ Is the predicted value, $\bar{y}_t$ is the sample average.

5.2 Experimental Result

In this laboratory, the training set and verification set are divided by 85%: 15%. The most critical part is parameter setting. On this basis, in order to further improve the prediction accuracy and accuracy of the model, the model test was carried out to find the optimal n, that is, the data of the N+1 day was predicted with the SOC data of the lead-acid battery in the past n days:Therefore, this paper sets the predicted N as and converges to the minimum value. When considering RMSE and MAPE. When N=6, it has the best prediction effect and the smallest error.
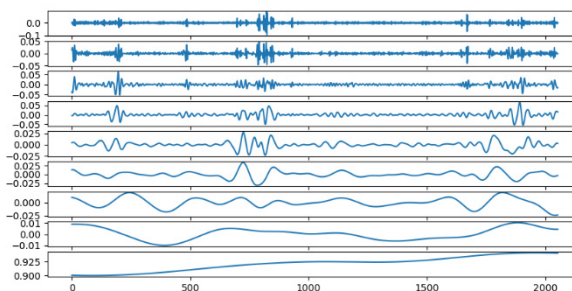

Fig 3. Breakdown of SOC through EEMD

The state of charge data of lead-acid batteries are processed by EEMD, as shown in Figure 3. After decomposition of 8 eigenmode functions and a Res, Res is the representation of the overall trend of

the curve. The comparison model of the experiment is the original LSTM model and the EEMD-LSTM model. The lead acid battery data is processed by the LSTM model. The predicted SOC results of the EEMD-LSTM model are shown in Figure 3, 4. 5.
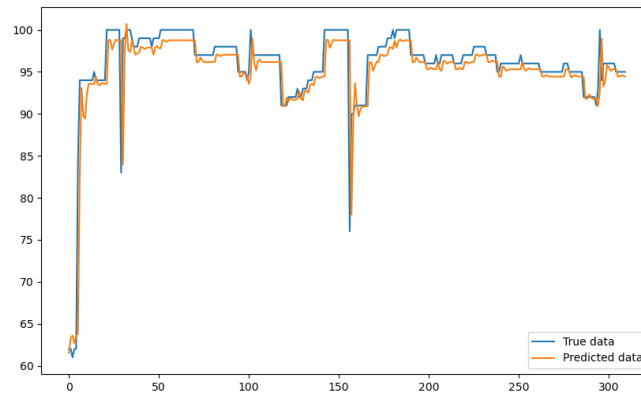
Fig 4. Prediction effect of LSTM model
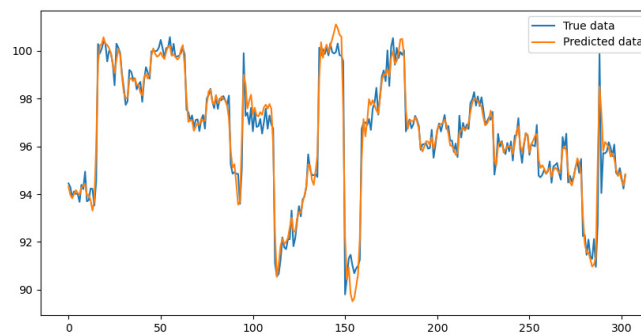
Fig 5. Prediction effect of EEMD-LSTM model

As shown in the figure, lead-acid battery data is best predicted by EEMD-LSTM model, and the evaluation indicators are shown in Table 1 below.

Table 1. Evaluation indicators of each model

Numble	RMSE	MAE	MAPE	R^2
LSTM	0.7475	0.5684	1.3452	0.4536
EEMD-LSTM	0.5390	0.3564	0.3710	0.9539

Acknowledgments

References

[1] Dhundhara, Sandeep, Yajvender Pal Verma, and Arthur Williams. "Techno-economic analysis of the lithium-ion and lead-acid battery in microgrid systems. "Energy Conversion and Management 177 (2018): 122-142.

[2] Xia, B.; Zhang, G.; Chen, H.; Li, Y.; Yu, Z.; Chen, Y. Verification Platform of SOC Estimation Algorithm for Lithium-Ion Batteries of Electric Vehicles. Energies (2022):15, 3221.

[3] Zhang, R., Xia, B., Li, B., Cao, L., Zhang R, Xia B, Li B, Cao L, Lai Y, Zheng W, Wang H, Wang W. State of the Art of Lithium-Ion Battery SOC Estimation for Electrical Vehicles. Energies. 2018; 11(7):1820.

[4] Chen, Xu-Dong, et al. "Life estimation of lithium-ion battery based on a hybrid model."Energy Exploration & Exploitation38.5 (2020): 1854-1878.

[5] Zhao, R.; Kollmeyer, P.J.; Lorenz, R.D.; Jahns, T.M. A compact methodology via a recurrent neural network for accurate equivalent circuit type modeling of lithiumion batteries. IEEE Trans. Ind. Appl. 2019, 55, 1922–1931.

[6] Zhao, J.; Mao, X.; Chen, L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed. Signal Process. Control (2019), 47, 312–323.

[7] SHEIKHAN M, PARDIS R, GHARAVIAN D. State of charge neural compumfional models for high energy density batteries inelectric vehicles [J]. Neural Computing and Applications, (2013), 2 (6): 1171-1180.

[8] Yang Zhanlu, Zhang Guoqing, Wang Zongliang Modeling method for large-capacity power lead-acid battery [J]. Marine Power Technology, 2012, 32 (6): 35-37.

[9] Shu Chengcai. Study on collaborative estimation of SOC and SOH of lead-acid battery and charging and discharging strategy [D]. Hefei: Hefei University of Technology, 2018.

[10] Hu Chen. Health state estimation of lead-acid batteries based on in-depth learning [J] 1001-1579 two thousand and twenty-one (in chinese).

[11] Chen Liang, Wang Zhen, Wang Gang. Application of LSTM network in short-term power load forecasting under the framework of deep learning [J]. Power Information and Communication Technology, 2017,15 (5): 8-11.

[12] A. Rosato, M. Panella, R. Araneo and A. Andreotti. A Neural Network Based Prediction System of Distributed Generation for the Management of Microgrids[J]. IEEE Transactions on Industry Applications, 2019, 55(6): 7092-7102.

[13] HUANG NE, SHNE Z, LONG S R, et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis[J].Proceeding of the Royal so-ciety of London.Series A:mathematical,physical and engineering sciences,(1998),454(1971):903-995.

[14] WU Z, HUANG NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method [J]. Advances in adaptive data analysis,(2009),1(1):1-41.

[15] Zhao Chenxiao, Xue Huifeng, Wang Lei, et al. Detection method of abnormal water intake data based on isolated forest algorithm [J] Journal of China Academy of Water Resources and Hydropower Sciences 2020, 18 (1): 31-39.

[16] Zhou Suqing. Multi-data stream fusion anomaly detection method based on cuckoo search algorithm [J] Journal of Inner Mongolia University for Nationalities: Natural Science Edition, (2020), 35 (3): 227-232.