

Lithium Battery SOC Estimation Based on Multi-Head Attention Mechanism and GRU Algorithm

Xueguang Li^{1, a}, Menchita F. Dumlao^{1, 2, 3, *}

¹ Graduate School, Angeles University Foundation, Angeles, Philippines

² Graduate School, Philippine Women's University, Manila, Philippines

³ Visiting Professor, Graduate School, Angeles University Foundation, Angeles, Philippines

^a li.xueguang@auf.edu.ph, * Corresponding author: mfdumlao@pwu.edu.ph

Abstract: Pure electric vehicles have been widely used due to their non-pollution, low noise, high energy conversion efficiency and other advantages. SOC (State of Charge) is a crucial indicator for lithium batteries and pure electric vehicles. SOC cannot be directly measured. This article designs a new network structure. It is the GRU-Attention network structure. The stacked GRU algorithm in GRU-Attention network extracts the temporal characteristics of lithium battery test data, and the stacked multi-head self-attention network extracts the global information. The GRU-Attention network can avoid long-term dependency and gradient disappearance problems. The proposed network utilizes Stacked FFN as the dense layer. This article will test the network designed in the public data set at the University of Maryland. Simultaneously, this article compares the effects of different BatchSize on the performance of the algorithm. The network training process converges more effectively with a smaller BatchSize. Both too large and too small BatchSize have a negative impact on the generalization performance of the network. The extraction of the time-order character, however, may be hampered if the timestamp is too small. At the same time, the paper also compares the GRU-Attention network horizontally with the GRU and Attention networks. Eventually, the GRU-Attention network proposed in this article could better meet the estimate of the lithium battery SOC.

Keywords: Machine Learning, Lithium battery, Attention Mechanisms, State of Charge, GRU.

1. Introduction

Global energy shortage[1] and environmental problems have become urgent research hotspots. Conventional fuel-powered vehicles are fueled by fossil fuels. It is not only a non-renewable resource, and the exhaust gas is polluting the environment. The carbon dioxide emissions will also bring a greenhouse effect. Pure electric vehicles[2] have been widely used because of their pollution-free, low-noise, high energy conversion efficiency[3], simple structure, and easy maintenance. Lithium battery[4] is the main component of pure electric vehicles. It accounts for a third of the total price of an electric car. Therefore, the safety and performance of lithium batteries directly determine the performance and safety of pure electric vehicles. SOC is a key indicator of lithium batteries. It determines the driving range of the lithium battery. At the same time, it is the basic parameter of SOH[5, 6], SOE, RUL and other indicators. Therefore, its accurate estimation plays a very important role in the performance and safety of lithium batteries. However, the SOC of lithium batteries cannot be measured directly. In the process of charging, lithium batteries convert electrical energy into chemical energy. During discharge[7], lithium batteries convert chemical energy into electrical energy. And in the static process, the lithium battery itself has a self-discharge effect. All these characteristics lead to the difficulty in estimating the SOC of lithium batteries. The only parameters that can be measured directly are current, voltage, temperature, internal resistance, and open circuit voltage. This brings great difficulty to the accurate estimation of SOC. However, the accurate estimation of SOC directly determines the performance of lithium battery and the safety of lithium battery. The performance and safety of lithium batteries directly determine the driving range of pure electric vehicles

and the safety of pure electric vehicles. If the SOC of lithium battery cannot be accurately estimated, it will increase the risk of safety accidents in pure electric vehicles. The inaccuracy of SOC estimation will have a great impact on thermal management and over-discharge protection[8, 9]. These two are the two biggest causes of safety accidents of pure electric vehicles. The occurrence of pure electric vehicle accidents will not only bring property losses, but also bring people's lives.

In view of the importance of SOC and the difficulty of estimation, researchers all over the world are interested in this direction.

At present, the research on lithium battery SOC estimation is a very popular direction. Researchers all over the world have carried out more in-depth research on this aspect. There are three main directions to study the estimation of SOC in lithium batteries. The first direction is based on the simple measurement method. The main basis of this method is that there is a very large positive correlation between SOC and open circuit voltage of lithium batteries. And with the discharge process of the car driving process, the current will continue to discharge, the integral of the current discharge over time is the amount of electricity released. In this way, SOC can be measured by OCV[7, 10] and ampere-hour integration method. However, the disadvantages of this method are also very obvious. First, the open-circuit voltage measurement itself is very difficult. Only very specialized research institutions can measure it. Number two. The process of open circuit voltage measurement itself is time-consuming. Each short discharge requires 2 hours of standing before the open circuit voltage value can be measured. Third, ampere-hour integration has cumulative error. As time goes by, the cumulative error of ampere-hour integration will become larger and larger, which will lead to the accuracy of SOC

estimation will become lower and lower. Fourth, the battery will age with the increase of the service life. Simple measures do not take into account the process of battery aging. The second direction is model-based estimation methods. The approach is done by modeling the battery. Through the accurate establishment of the model, the SOC can be better estimated. The quality of the model directly determines the quality of the estimation. The value estimated by the model is used as an estimate of the SOC of the lithium battery. In general, the following models are commonly used: Electrochemical Model (EM)[11]. EM calculates the terminal voltage and SOC of the battery according to the electrochemical reaction process, and is a battery model based on the theory of porous electrodes and solution concentration. The P2D model[12] describes the solid-phase Li-ion concentration and potential of Li-ion batteries using a set of coupled partial differential equations, and is one of the most widely used EM in Li-ion battery SoC estimation. The electrochemical model mainly reflects the chemical reaction mechanism inside the battery, and the model has high accuracy, but it is difficult to determine all the parameters, which has huge computational complexity and time consumption. Electrochemical Impedance Model (EIM). Electrochemical Impedance Spectroscopy (EIS) [13] is a battery measurement technique proposed by Heaviside in 1894, which is commonly used to model electrochemical impedance circuits. EIM obtains the AC equivalent impedance model in the frequency domain by EIS method, determines the composition of the equivalent circuit and the size of each component by EIS, and then uses a complex equivalent network to match the impedance spectrum. EIM can accurately describe the battery characteristics, but the matching process is difficult, complex and not intuitive in practical application. Moreover, the impedance model is only useful for specific SoC and temperature, and cannot predict DC reaction and battery operation time. Equivalent Circuit Model (ECM). ECM[14] is used to describe and simulate the dynamic characteristics of the battery. It regards the battery as a two-port network, and uses voltage sources[15], resistors, capacitors and other components to simulate the internal characteristics of the battery. Various ECM models have been proposed for SOC estimation of lithium-ion batteries. Common models include Rint model, Thevenin model, second-order resistor and capacitor parallel equivalent model, Partnership for a New Generation of Vehicle (PNGV) model[16]. The third direction is based on data-driven approaches. Data-driven based estimation methods refer to the direct estimation of SOC using battery data by measuring battery parameters such as current, voltage, temperature, and internal resistance. In recent years, with the rapid development of machine learning, data-driven SOC estimation methods often use machine learning platforms to automatically learn network parameters through intelligent algorithms and obtain the relationship between battery parameters and SOC. Machine learning methods commonly used for SOC estimation include neural networks and deep learning, Support Vector Machine (SVM)[17] and recurrent neural networks. Estimation methods based on data-driven SOC also have their own shortcomings. It has a large amount of data and high data requirements. At the same time, it needs the support of more powerful hardware platform in the training process. All these bring great difficulties to the application of data-driven methods in engineering practice. Fortunately, recently, GRU computing power and distributed

cloud platforms (Hadoop, Spark) are widely used. These improvements in hardware performance bring great opportunities for data-driven SOC applications.

2. Related Works

There are three approaches for estimating a Li-ion battery's SOC. Direct measurement is the first technique. Although the OCV measurement approach is more exact, it still necessitates the use of specialist measurement equipment and resting between measurements. As a result, it is rarely utilized in practice and is often measured by manufacturers using specialized equipment. Using lithium batteries can also result in erroneous OCV measurements. This is due to the fact that the earlier observed OCV values are no longer correct due to the rise in polarization internal resistance. According to Wei Zhongbao's paper[18], a low-current technique would be superior for determining OCV values. The coulomb counting approach is used by Movassagh [19] to estimate SOC. Wang Zuolu[20] developed a technique for calculating the value of SOC using an ECM model. The lithium-ion battery is used as an analogous circuit or electrochemical model in the approach for estimating SOC that is based on models. Choi Woosung models[21] the lithium-ion battery with the EIS approach. Yang Fangfang[22] offers a fresh concept by combining UKF and LSTM. The model then makes an estimation of the SOC. A data-driven strategy is the foundation of the third approach. The current trend in lithium-ion battery SOC estimation is toward data-driven methods. A convolutional informer network-based method to calculate the SOC of lithium-ion batteries is suggested by Zou Runmin[23]. CNN+LSTM is the proposed algorithm by Li Jiarui[24]. The CNN+LSTM algorithm performs SOC estimation by combining CNN and LSTM to extract features from lithium-ion battery data. El Fallah, Saad employs a very intricate GRU network strategy. Alkaita and Eyad proposed utilizing the LSTM algorithm to estimate the SOC of a collection of lithium-ion batteries. A technique based on Temporal Convolutional Network and Transfer Learning was proposed by Liu Yuefeng[25].

3. Methodology

3.1. Data preprocessing

The CACLE dataset from the University of Maryland is used in this paper. It is a public dataset. The data set selected in this paper contains columns such as step-time and step-index, which do not affect the change of SOC[26] during processing. Second, there is a correlation between step time and test time, so only one column of features is needed to express it. Typically, null values are filled in at this stage. Typically, mean filling is used.

3.2. Selection of hyperparameters

This paper is a regression problem, so the selection of the cost function follows the basic loss function of regression problems. The mean squared error is used in this paper. Sigmoid and tanh functions are used as activation functions. Other hyperparameters are the width of the network and the length of BatchSize[27]. There is also the size of the EPOCH[28]. There is also the depth of the network.

3.3. GRU networks

GRU algorithm is a typical algorithm in recurrent neural networks. Figure 1 represents the basic structure of the GRU network. It solves the problem of long-term dependence on

recurrent neural networks. It is also much faster than LSTM. And the GRU algorithm is simpler than the LSTM[29] algorithm in structure and idea. LSTM is a special kind of RNN model. The memory unit of RNN[30] is improved to overcome the problem of vanishing gradient. It mainly includes three gates, an input gate, an output gate, and a forget gate. The input is the current time x_t , and the hidden state from the previous time. GRU is an improved model based on LSTM, and the two principles are similar. Compared with LSTM, GRU[1, 31] has a simpler network structure, fewer parameters, and faster convergence. Specifically, GRU lacks forgetting gates and units, which are updated to control the information transfer at the previous moment and the hidden layer computed at the current moment.

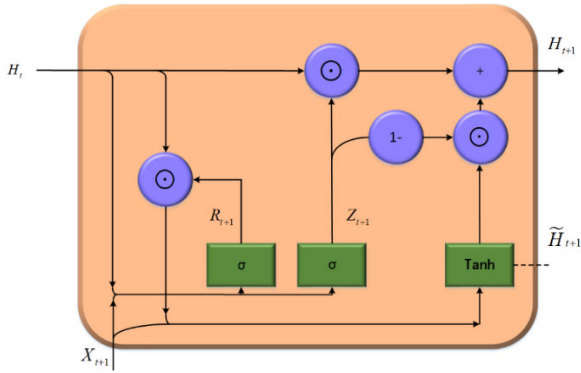


Figure 1. GRU network.

$$R_{t+1} = \sigma(X_{t+1}W_{xr} + H_tW_{hr} + B_r) \quad (1)$$

$$Z_{t+1} = \sigma(X_{t+1}W_{xz} + H_tW_{hz} + B_z) \quad (2)$$

$$\tilde{H}_{t+1} = \tanh(X_{t+1}W_{xh} + (R_{t+1} \odot H_t)W_{hh} + B_h) \quad (3)$$

$$H_{t+1} = Z_{t+1} \odot H_t + (1 - Z_{t+1}) \odot \tilde{H}_{t+1} \quad (4)$$

GRU network is able to alleviate the problem of long-term dependence and gradient disappearance. However, GRU networks have the following drawbacks. First, it is based on recurrent neural network algorithm, so it does not fully solve the problem of long-term dependence. Second, due to its own properties, the value of H at the current time depends on the value of H at the previous time. This leads to the GRU network can not carry out large-scale parallel computing. Third, the GRU network itself is not enough to pay attention to the influence of ultra-long time variables simultaneously. That is, it does not have a global attention mechanism by itself. The value of H at the current time is the result obtained by combining the factors of the values of H at all previous times. When the time series are too long apart, it will perform insufficient comprehensive transmission ability of influence.

3.4. Self-attention mechanisms

The attention mechanism is actually an attempt to apply human perception, and attentional behavior, to the machine so that the machine learns to perceive the important and unimportant parts of the data. The weight assigned to each input item depends on the interaction between the input items, i.e., a "vote" within the input item determines which input items should be attended to. Compared to the first two, this has the advantage of parallel computing when dealing with very long inputs. Figure 2 represents the computational

process of the self-attention mechanism.

Suppose the input sequence is $X = [x_1, x_2, \dots, x_N]$, and the output sequence is $H = [h_1, h_2, \dots, h_N]$. The calculation procedure of the self-attentive mechanism is shown below.

1) The three weight matrices are multiplied by the input matrix X. The transformation of X into three distinct areas. It acquires the three matrices Q, K, and V. The equivalents of the X matrix's retrieval matrix, keyword matrix, and value matrix are the K matrix for the X matrix, the Q matrix for the X matrix, and the V matrix for the X matrix. The Q matrix and the K matrix are used to calculate similarity. There are several methods for calculating similarity. By somewhat normalizing the calculated similarity, the similarity is calculated. The normalized values are multiplied by the V matrix's weights. The attentional qualities result from this.

$$Q = W_Q X \quad (5)$$

$$K = W_K X \quad (6)$$

$$V = W_V X \quad (7)$$

2) Calculate the similarity between K and V. The Q matrix formula has query vectors composed of. That is, $Q = [q_1, q_2, \dots, q_N]$. k matrix is also composed of keyword vectors. That is, $K = [k_1, k_2, \dots, k_N]$. v-matrix is also composed of value vectors, that is, $V = [v_1, v_2, \dots, v_N]$. The similarity is compared by using q-vectors with k-vectors. By comparing the similarity of q-vectors with different k-vectors. These similarities are normalized with softmax. Finally, it is possible to derive the weight coefficients α .

$$h_N = att((K, V), q_n) \quad (8)$$

$$h_N = \sum_{s=1}^N \alpha_{ns} v_s \quad (9)$$

$$h_N = \sum_{s=1}^N softmax(score(k_s, q_n)) * v_s \quad (10)$$

3) Scale the dot product to be the attention score function.

$$H = V * softmax\left(\frac{K^T Q}{\sqrt{D_k}}\right) \quad (11)$$

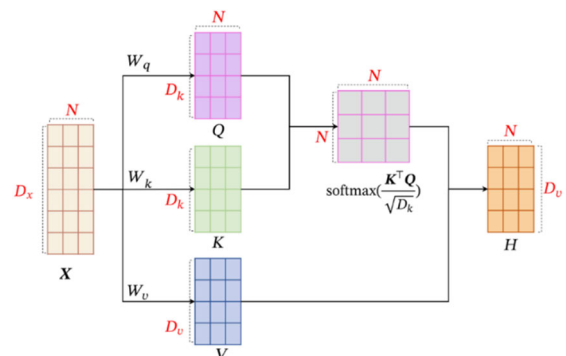


Figure 2. Self-attention mechanism calculation process

3.5. GRU-Attention network

The GRU network is an improvement of the LSTM network. The GRU network can not only extract time series information but also ensure the update of the new state. However, the GRU network still has the problem of long-term dependence. And when the GRU network is too deep, there will be gradient explosion and gradient disappearance. The

multi-head attention mechanism is derived from the Transformer network, which is the basis for the outstanding performance of the Transformer algorithm. The multi-head attention mechanism used in this paper can search for the characteristics of global attention. The significant advantages of the attention mechanism are to focus on relevant information and ignore irrelevant information, to directly establish the dependence between input and output without looping, to enhance the degree of parallelism, to greatly improve the running speed, and to directly act the value of ultra-long path on the value of output.

The algorithm designed in this paper first extracts the temporal features and underlying information of Li-ion battery information using a multi-layer GRU network. After passing the extracted features into the multi-layer attention mechanism layer algorithm, the multi-layer attention mechanism is used to further focus on the features that greatly impact the lithium battery SOC and ignore the irrelevant information. Also, increase the parallel computing capability where $Volt(t), Volt(t-1), Volt(t-i)$ represent the voltages at different moments. $Cur(t), Cur(t-1), Cur(t-i)$ are the current values at different times. the magnitude of TimeStep determines the value of I . the value of TimeStep is a hyperparameter. $tem(t), Tem(t-1), Tem(t-i)$ are the temperature values at different times. values. Temperature is crucial in the SOC estimation of Li-ion batteries. $SOC(t), SOC(t-1), SOC(t-$

$i)$ represent the SOC values at different times. It does not make sudden changes. Therefore, the value of the previous moment of SOC will have a greater impact on the current moment of SOC. These parameters are also the values of charging capacity, discharging capacity, charging energy, discharging energy, and so on.

The network designed in this paper takes the parameter data of the lithium battery detected by the lithium battery test platform as a vector and turns these vectors into a matrix according to the TimeStep size. Then we add BatchSize to make it a 3D matrix. However, in the network design process, the influence brought by BatchSize is not considered. The current, voltage, temperature, and SOC vectors are combined into a matrix, which is then used as the input feature X . The input feature X will be the input of the Stacked GRU network. Stacked GRU networks represent multiple layers of GRU. The output of Stacked GRU is used as input to the Stacked Attention layer. Attention layer uses multi-head self-attention layer. Different multiple heads are used to extract different attention features. Multi-layer Attention layers are used to extract higher-order semantic information. The output of the Attention layer is fed to the input features of the neural network. Feedforward neural networks compress the width of the network step by step. The final output is an estimate of the SOC. Figure 3 represents the GRU+Attention network structure designed in this paper.

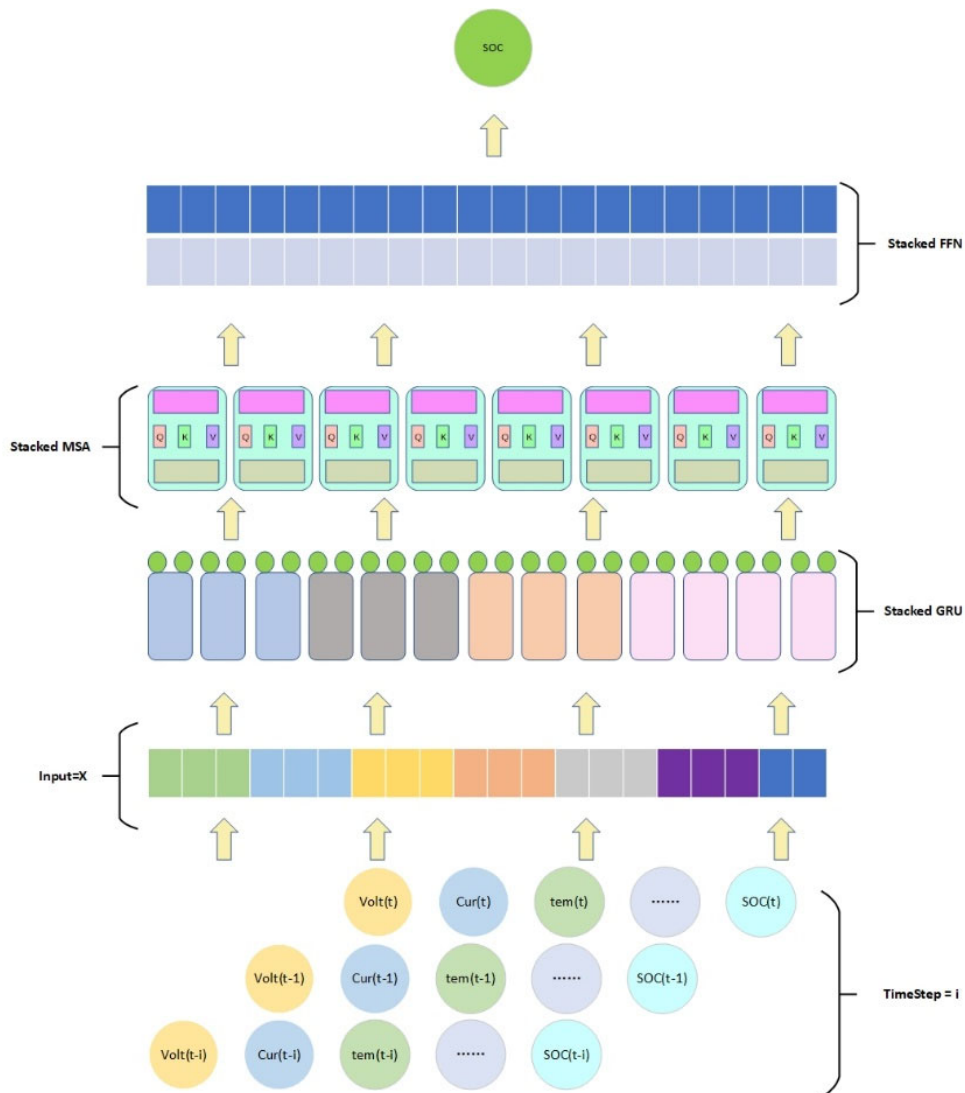


Figure 3. GRU + Attention Network

4. Experiment

The dataset used in this paper is a public dataset. It is the CALCE data provided by the University of Maryland. The battery model used in this paper is 18650-20R. The calculation of SOC of Li-ion battery is highly dependent on temperature and initial capacity. Therefore, in-depth testing is needed specifically for the initial capacity. In order to simulate the real situation of Li-ion battery in normal operation, each country has different standards to test Li-ion battery. This paper mainly uses BJDST. In the BJDST data set, the initial capacity of the lithium battery is first tested. Secondly, the lithium battery is tested under different temperatures. Table 1 shows the basic parameters of lithium batteries.

Table 1. Parameters of lithium batteries

Battery Parameters	Specifications Value
Battery Type	INR 18650-20R
Capacity Rating	2000mAh
Cell Chemistry	LNMC/Graphite
Weight	45.0g
Dimensions(mm)	18.33±0.07mm
Length	64.85±0.15mm

In the CALCE test set, the data were tested separately according to different temperatures, and the BJDST condition simulation was first used to test the lithium battery at 0 degrees Celsius. At the same time, the battery capacity was tested at 80% SOC and 50% SOC. Then BJDST tests were conducted at 25 degrees C and 45 degrees C. The 80% SOC and 50% SOC tests were also conducted separately.

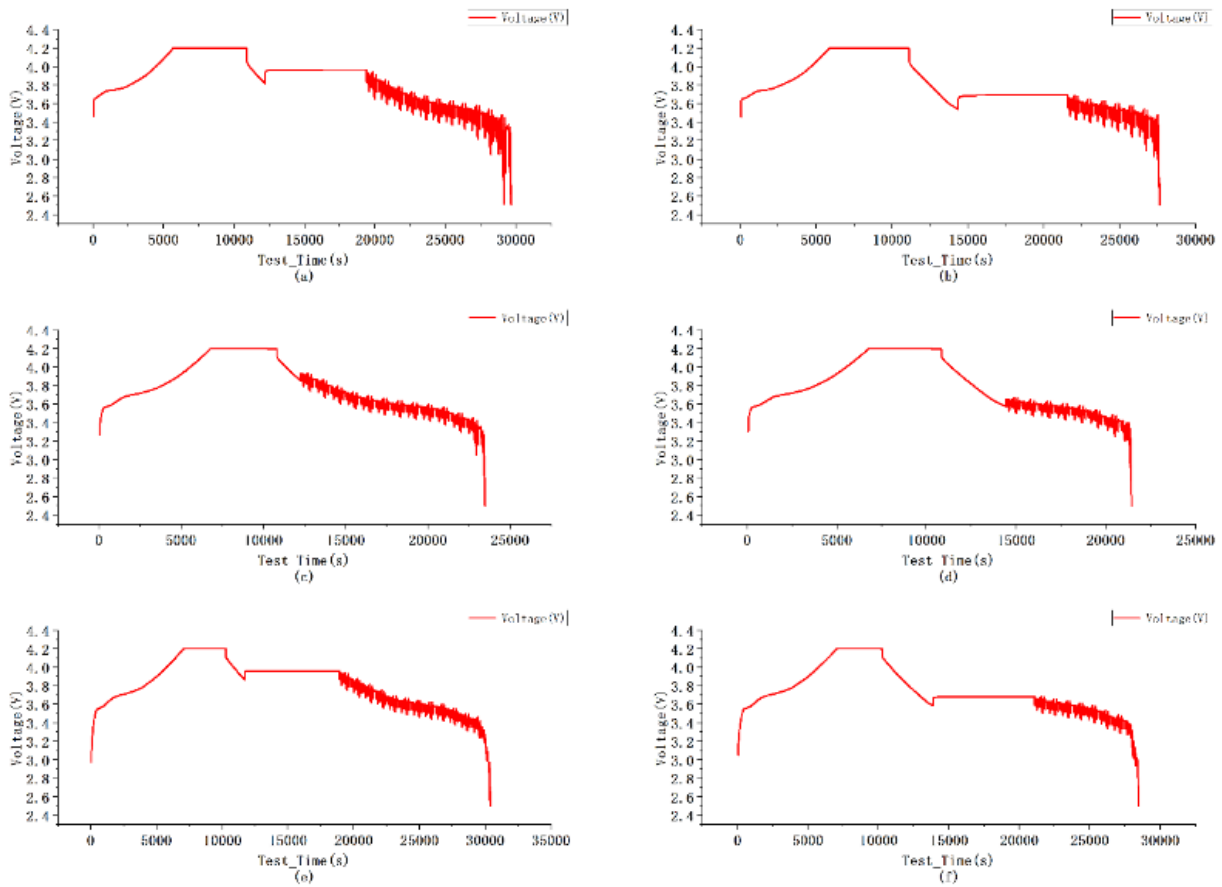


Figure 4. Voltage testing under BJDST operating conditions.

Subplot a in Figure 4 represents the BJDST test voltage at 80% SOC at 0 degrees Celsius. It indicates that the battery is charged first. When the charging current reaches a certain cutoff current, the battery is then discharged and discharged to the OCV voltage corresponding to 80% SOC without starting frequent charging and discharging until it is discharged to the minimum cutoff voltage. Subfigure b in Figure 4 shows the BJDST test voltage at 50% SOC at 0°C. It shows the BJDST test voltage at 50% SOC at 0°C. It indicates that the battery is charged first. When the charging

current reaches a certain cutoff current, the battery is then discharged and discharged to the OCV voltage corresponding to 50% SOC without starting frequent charging and discharging until it is discharged to the minimum cutoff voltage. Subfigure c in Figure 4 shows the BJDST test at 80% SOC at 25 degrees Celsius. Subplot d in Figure 4 shows the BJDST test at 50% SOC at 25 degrees C. Subplot e in Figure 4 shows the BJDST test at 80% SOC at 45 degrees C. Subplot f in Figure 4 shows the BJDST test at 50% SOC at 45 degrees.

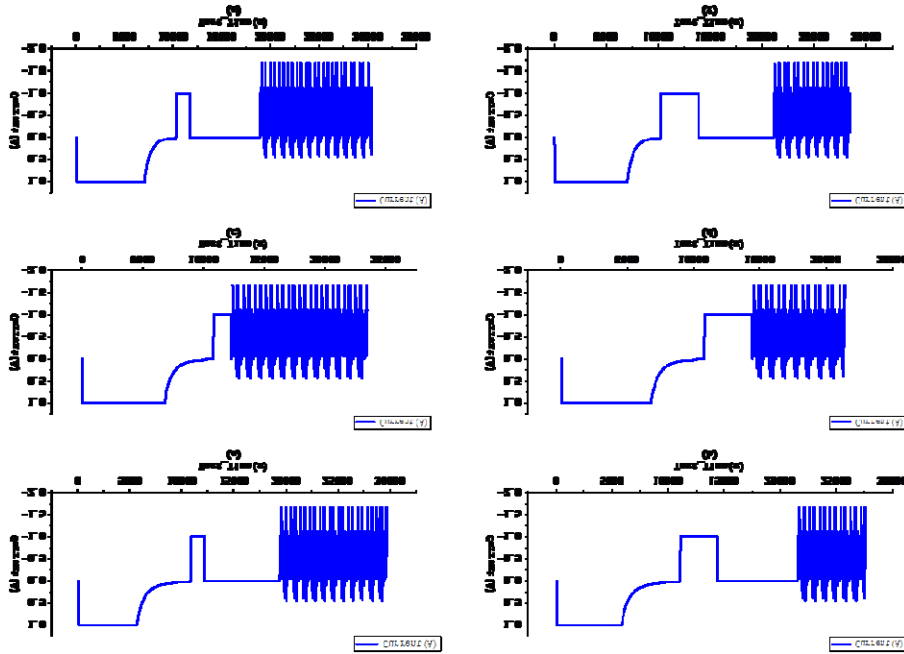


Figure 5. SOC test data at 50 percent under BJDST conditions

Subplot a in Figure 5 represents the BJDST operating condition current data of 80%SOC test at 0 degree Celsius. Subplot b in Figure 5 represents the BJDST operating condition current data of 50%SOC test at 0 degree Celsius state. Subplot c in Figure 5 represents the BJDST operating condition current data of 80%SOC test at 25 degrees Celsius. The d subplot in Figure 5 represents the BJDST operating condition current data of the 50%SOC test at 25 degrees Celsius. The e subplot in Figure 5 represents the BJDST operating condition current data of 80%SOC test at 45 degrees Celsius. The f subplot in Figure 5 represents the BJDST operating condition current data of 50%SOC test at 45 degrees Celsius state.

According to the definition of SOC, it can be seen that when estimating SOC, the accurate initial capacity also directly determines the accuracy of SOC value estimation. So, in the CALCE data set of the University of Maryland, the

initial capacity was re-tested further in the dynamic testing phase.

In this paper, the TimeStep is selected as 3,5,7,10. The epoch is selected as 50, and the number of GRU layers is selected as 3. The multi-head attention layer is selected as three layers. The feedforward neural network is chosen as 2 layers. The network width of GRU is selected as 64, and the head of the multi-head attention mechanism is selected as 3.

5. Results and discussion

In this paper, the mean square error is used for both the training error and the testing error. In this paper, the model is first trained according to different TimeStep. Since the TimeStep is different, it contains different lengths of temporal information.

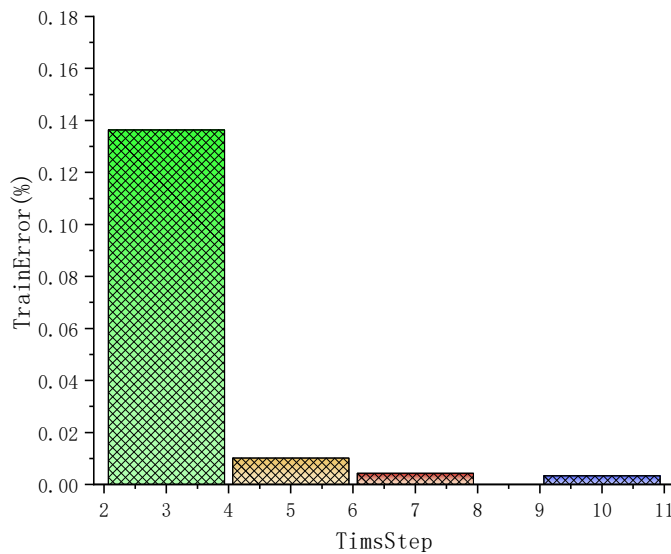


Figure 6. Training errors in different TimeStep states

The selection of hyperparameters often determines the performance of this designed network. Figure 6 represents the

training error for different TimeStep states. In this paper, a more suitable TimeStep is selected by comparing the training

error when the TimeStep is 3, 5, 7 and 10, respectively. If the TimeStep is selected too large, firstly, it will affect the training performance of the designed algorithm. It causes the network training too slow. Moreover, too much input of useless information will lead to the problem of long-term dependency more prominent. If the TimeStep is chosen too small, it will cause the designed algorithm network to perform incorrectly and will also lead to an underfitting phenomenon. Moreover, when the TimeStep is selected small, the timing information in it is not adequately reflected. The extraction of timing features becomes more difficult. The estimation of Li-ion battery SOC is mainly using Li-ion battery in the charge and discharge itself is a time series. Therefore, looking at Figure 6, there is not much difference between the training error when TimeStep is 5 and when TimeStep is 10. According to Occam's razor principle. In this paper, we choose TimeStep of 5 as the hyperparameter. And it is the TimeStep of 5 that is selected in the later experiments and results.

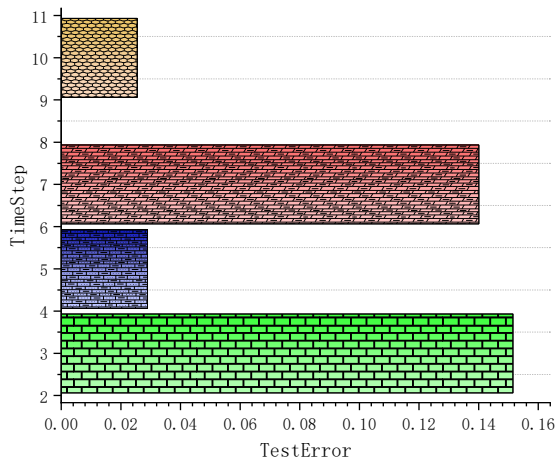


Figure 7. Test errors in different TimeStep states

The test error best reflects the generalization of the network. Figure 7 shows the test error for different TimeStep states. Selecting TimeStep only by training error may not be able to meet the requirements of prosperity. In this paper, experiments are carried out to test the error for different timesteps. TimeStep 3,5,7,10 were selected for comparison. Figure 7 shows the result of the test error. If we look at the training error, the larger the TimeStep selection, the better. A TimeStep of 10 resulted in the smallest training error, and a TimeStep of 7 resulted in the second smallest training error. However, it is evident in the test error. The test error with a TimeStep of 5 is smaller than that with a TimeStep of 7. There is little difference between the test error when TimeStep is 5 and TimeStep is 10. According to Occam's razor principle, TimeStep 5 is selected in this paper. More importantly, it is found in the training process that when TimeStep 7 and TimeStep 10, the training time increases exponentially. This increases the training time and training cost and does not lead to more efficient results. Therefore, 5 is used in the selection of TimeStep in this paper.

BatchSize is also one of the hyperparameter metrics that affect the convergence of the data. Figure 8 represents the training error for different BatchSize states. This paper compares five groups of data from BatchSize 4, BatchSize 8, BatchSize 16, BatchSize 32 and BatchSize 64. Pass the test when the TimeStep of the previous test is 5. It can be seen from the Figure that as the number of epochs increases, they

all eventually tend to converge, and there is little difference between the training errors of the five groups of data when EPOCH is 50. Therefore, BatchSize has little effect on the training error from the convergence point of view. However, in terms of the fast and slow rate of convergence, it is obvious that the convergence rate of BatchSize 4 is larger than that of BatchSize 8. The convergence rate with BatchSize 8 is greater than the convergence rate with BatchSize 16, and so on. In terms of the fast and slow convergence rate, the smaller BatchSize is, the faster the convergence rate is. The larger the BatchSize, the slower the convergence. The reason for this is because the number of iterations is different. When the total amount of data is constant, the smaller BatchSize is, the more iterations are needed in an EPOCH loop. Conversely, the larger BatchSize, the smaller the number of iterations within an EPOCH loop. The more iterations, the more times the gradient is updated, which can better approximate the convex function in the direction of the extreme value. However, with fewer iterations, the gradient update is slower, and the correction towards the extreme value of the convex function will be slower. However, the gradient descent algorithm ensures that the approximation tends to be in the direction of the minimum of the convex function. Therefore, when the number of epochs is large, the difference between the two is not obvious. When the number of epochs is small, the value of BatchSize is appropriate. In order to save resources and time, under the same condition, selecting a smaller BatchSize can accelerate the convergence.

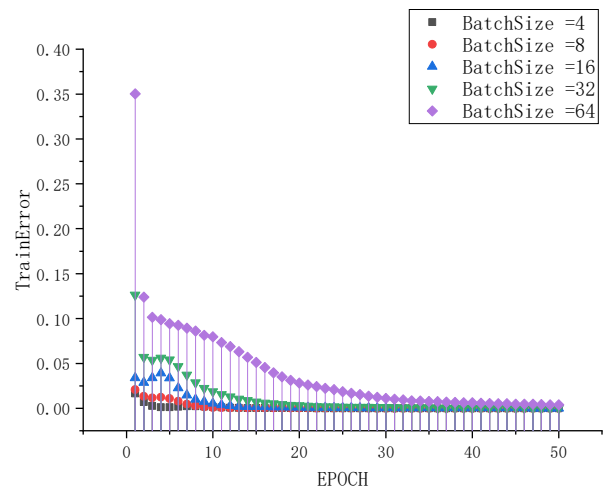


Figure 8. Training error in different BatchSize states

Training error can only be used to measure convergence and convergence speed. Figure 9 represents the training error for different BatchSize states. What really allows the model algorithm to be applied to engineering practice is testing error. The test error can better express the generalization performance of the algorithm. When facing unknown data, the test error shows the accuracy and robustness of the algorithm. The comparison of the test errors in different BatchSize states is shown in Figure 8. According to the test error when the last EPOCH is 50, For BatchSize, no clear-cut rule says bigger is better or smaller is better. When BatchSize is too small, the test error at the beginning of the EPOCH is the largest, which indicates that the BatchSize is too small, and the model algorithm is easy to enter the local minimum value, which will lead to falling into the local minimum value and the error jumps out. The true minimum point is not reached. When BatchSize is too large, it may cross the region of the minimum point and enter another local minimum point.

And because BatchSize is too large, it leads to too large interpolation in the gradient update process, which may bring a concussion effect and never enter the minimum area. Therefore, it can be concluded that in the selection process of BathSize, more experiments are needed to select a value suitable for the algorithm. From the point of view of this paper, the value of BatchSize of 16 is selected for the algorithm model in this paper.

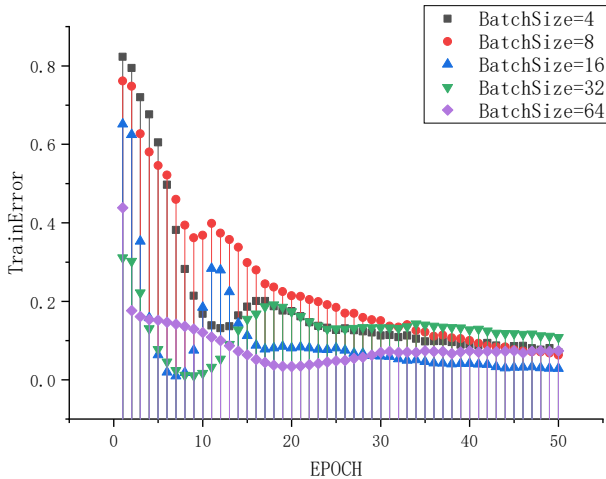


Figure 9. Training error in different BatchSize states

In this paper, the designed GRU-Attention network is compared with the GRU network and the Attention network in a cross-sectional manner. The table 2 shows the results of the algorithm comparison. The comparison reveals that the GRU-Attention network architecture proposed in this paper can indeed better meet the needs of lithium battery SOC estimation. It has better improvement in accuracy. However, in terms of training speed, the GRU-Attention network is slower than the pure Attention network. However, the GRU-Attention network is faster than the pure GRU network.

Table 2. Performance comparison of different algorithms.

Id	Algorithm	TrainError(%)	TestError(%)
1	GRU	2.32%	4.81%
2	Attention	1.96%	3.72%
3	GRU-Attention	0.014%	1,3%

6. Conclusion

SOC cannot be measured directly. The estimation of SOC will become less accurate as the running time of the lithium battery increases. The GRU-Attention algorithm network proposed in this paper can solve these two problems effectively. The GRU-attention algorithm network has a Stacked GRU subnetwork. It can effectively extract the timing information of lithium batteries. The GRU-Attention network also contains the Attention sub-network. It can pay more attention to global information characteristics. The publicly available dataset from the University of Maryland is used in this paper. In optimizing network hyperparameters, BatchSize determines the speed of network convergence. However, in the test network performance, whether BatchSize is too large or too small will negatively impact the network performance. The value of another hyperparameter, TimeStep, also affects the network's performance. The smaller the value of TimeStep, the faster the network designed in this paper converges. However, too small of TimeStep will degrade the

generalization performance of the network. This paper compares the proposed algorithm with the GRU network and Attention network. Finally, it is concluded that the accuracy of the GRU-Attention network proposed in this paper is higher. Future research must combine multiple public data sets to extract lithium batteries' intrinsic characteristics better.

References

- [1] Pan, C., et al., Adaptive Neural Network-Based Prescribed-Time Observer for Battery State-of-Charge Estimation. *IEEE Transactions on Power Electronics*, 2023. 38(1): p. 165-176.
- [2] Yang, X., et al., Battery state of charge estimation using temporal convolutional network based on electric vehicles operating data. *Journal of Energy Storage*, 2022. 55.
- [3] Li, X. and M.F. Dumlaio, Deep Learning Based Electric Vehicle BMS Intelligent Cloud Monitoring System, in *2022 3rd Asia-Pacific Conference on Image Processing, Electronics and Computers*. 2022. p. 452-457.
- [4] Brieske D M, Warnecke A, Sauer D U. Modeling the volumetric expansion of the lithium-sulfur battery considering charge and discharge profiles. *Energy Storage Materials*, 2023, 55: 289-300.
- [5] Park, Minjun, et al. "Data-Driven Capacity Estimation of Li-Ion Batteries Using Constant Current Charging at Various Ambient Temperatures." *IEEE Access* 11 (2023): 2711-2720.
- [6] Hendricks C, Sood B, Pecht M. Lithium-ion battery strain gauge monitoring and depth of discharge estimation. *Journal of Electrochemical Energy Conversion and Storage*, 2023, 20(1): 011008.
- [7] Huang C S. An Online Condition-Based Parameter Identification Switching Algorithm for Lithium-Ion Batteries in Electric Vehicles. *IEEE Transactions on Vehicular Technology*, 2022.
- [8] de Souza A K, Hileman W, Trimboli M S, et al. A Control-Oriented Reduced-Order Model for Lithium-Metal Batteries. *IEEE Control Systems Letters*, 2022.
- [9] Li, X., et al., SOC Estimation of Lithium-Ion Battery for Electric Vehicle Based on Deep Multi-layer Perceptron. *Comput Intell Neurosci*, 2022. 2022: p. 3920317.
- [10] Çelikten, B., O. Eren, and Y.S. Karataş, An execution time optimized state of charge estimation method for lithium-ion battery. *Journal of Energy Storage*, 2022. 51.
- [11] Han B, Harding J R, Goodman J K S, et al. End-of-Charge Temperature Rise and State-of-Health Evaluation of Aged Lithium-Ion Battery. *Energies*, 2022, 16(1): 405.
- [12] Wassiliadis N, Kriegler J, Gamra K A, et al. Model-based health-aware fast charging to mitigate the risk of lithium plating and prolong the cycle life of lithium-ion batteries in electric vehicles. *Journal of Power Sources*, 2023, 561: 232586.
- [13] He J, Meng S, Li X, et al. Partial charging-based health feature extraction and state of health estimation of lithium-ion batteries. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2022.
- [14] Konz Z M, Wirtz B M, Verma A, et al. High-throughput Li plating quantification for fast-charging battery design. *Nature Energy*, 2023: 1-12.
- [15] Guo, S. and X. Li, Computer Aided Art Design and Production Based on Video Stream. *Computer-Aided Design and Applications*, 2020. 18(S3): p. 70-81.
- [16] Zhou L, Lai X, Li B, et al. State Estimation Models of Lithium-Ion Batteries for Battery Management System: Status, Challenges, and Future Trends. *Batteries*, 2023, 9(2): 131.

- [17] Wadi A, Abdel-Hafez M, Hashim H A, et al. An Invariant Method for Electric Vehicle Battery State-of-Charge Estimation Under Dynamic Drive Cycles. *IEEE Access*, 2023, 11: 8663-8673.
- [18] Martuyshv N V, Malozyomov B V, Sorokova S N, et al. Mathematical Modeling of the State of the Battery of Cargo Electric Vehicles. *Mathematics*, 2023, 11(3): 536.
- [19] Wong K L, Chou K S, Tse R, et al. A Novel Fusion Approach Consisting of GAN and State-of-Charge Estimator for Synthetic Battery Operation Data Generation. *Electronics*, 2023, 12(3): 657.
- [20] Chen S, Pan T, Jin B. State of Charge Estimation of Lithium-Ion Battery Using Energy Consumption Analysis. *International Journal of Automotive Technology*, 2023, 24(2): 445-457.
- [21] Lee J, Won J. Enhanced Coulomb Counting Method for SoC and SoH Estimation Based on Coulombic Efficiency. *IEEE Access*, 2023, 11: 15449-15459.
- [22] Wu L, Zhang Y. Attention-based encoder-decoder networks for state of charge estimation of lithium-ion battery. *Energy*, 2023: 126665.
- [23] Zou, R., et al., A novel convolutional informer network for deterministic and probabilistic state-of-charge estimation of lithium-ion batteries. *Journal of Energy Storage*, 2023. 57.
- [24] Li, J., et al., The state-of-charge predication of lithium-ion battery energy storage system using data-driven machine learning. *Sustainable Energy, Grids and Networks*, 2023. 34.
- [25] Liu, Y., et al., A review of lithium-ion battery state of charge estimation based on deep learning: Directions for improvement and future trends. *Journal of Energy Storage*, 2022. 52.
- [26] Mi Y Q, Deng W, He C, et al. In Situ Polymerized 1, 3-Dioxolane Electrolyte for Integrated Solid-State Lithium Batteries. *Angewandte Chemie*, 2023, 135(12): e202218621.
- [27] Zhao, L. and P. Qin, Accurate SOC Prediction and Monitoring of Each Cell in a Battery Pack Considering Various Influencing Factors. *IEEE Transactions on Industrial Electronics*, 2023. 70(1): p. 1025-1035.
- [28] Xiao, L., et al., Online state-of-charge estimation refining method for battery energy storage system using historical operating data. *Journal of Energy Storage*, 2023. 57.
- [29] Wang, Z., et al., Adaptive self-attention LSTM for RUL prediction of lithium-ion batteries. *Information Sciences*, 2023.
- [30] Park, M., et al., Data-Driven Capacity Estimation of Li-Ion Batteries Using Constant Current Charging at Various Ambient Temperatures. *IEEE Access*, 2023. 11: p. 2711-2720.
- [31] Khaleghi Rahimian, S. and Y. Tang, A Practical Data-Driven Battery State-of-Health Estimation for Electric Vehicles. *IEEE Transactions on Industrial Electronics*, 2023. 70(2): p. 1973-1982.