State of charge estimation of lithium-ion battery based on double deep Q network and extended Kalman filter

Guodong You 1*, Xue Wang 1, Chengxin Fang 1, Shang Zhang 1 and Xiaoxin Hou 1

1 College of Electronic Information and Automation, Tianjin University of Science and Technology, Hexi District, Tianjin, 300222, China

*Corresponding author’s e-mail: yougdong@tust.edu.cn

Abstract. The state of charge (SOC), as an important parameter of the lithium-ion battery management system (BMS), is an important factor affecting the current battery detection and safety issues. Due to the non-linear characteristics of the lithium-ion battery charging and discharging process, it is difficult to model the BMS and estimate the SOC accurately. This paper takes lithium-ion battery as the research object, designs the Markov decision process (MDP) model of lithium-ion battery energy storage prediction, establishes the second-order RC equivalent circuit model (ECM) and extended Kalman filter (EKF), and proposes a method based on double deep Q network (double DQN) to optimize the EKF parameters SOC estimation method to solve the problem of dimensionality disaster and value function overestimation generated in deep reinforcement learning (DRL). Through the effective verification of the lithium-ion battery SOC estimation algorithm, the simulation results show that compared with the traditional deep Q network (DQN) algorithm, double DQN has better convergence, adaptive ability and better estimation accuracy. It can provide new ideas for the accurate estimation of SOC.

1. Introduction

The state of charge (SOC) of a lithium-ion battery can directly reflect the remaining capacity of the battery. It is an important basis for energy management and control strategies and a key parameter to ensure battery reliability[1]. The factors affecting the estimated SOC are capacity decay, self-discharge, consistency, temperature and discharge current. These factors make it difficult to model the energy storage system and accurately estimate the SOC. The traditional open-circuit voltage and ampere-hour integration method cannot estimate the SOC online due to its low accuracy. An improved SOC estimation method is needed to improve its prediction effect. Therefore, accurate SOC estimation is of great significance to the efficient utilization and energy management of lithium-ion batteries[2].

In the past few years, in order to improve the SOC estimation accuracy as much as possible, various intelligent algorithms have been applied to the SOC estimation of lithium-ion batteries. Among them, Literature [3] proposes a hybrid method based on DQN to optimize EKF parameters to estimate SOC, which solves the problem of "dimensionality disaster" in Q-learning, but there is the problem of Q value being overestimated. Literature [4] an improved BP neural network algorithm is proposed to simulate the human brain and neurons to process nonlinear systems, which can significantly reduce the lithium-ion battery SOC prediction error, but the amount of calculation is huge. Literature [5] proposed a method based on experimental data. Compared with the EKF algorithm, the cubic Kalman filter-based method for estimating the SOC of lithium-ion batteries can obtain higher estimation accuracy when computational resources are consumed.
Aiming at the problems of the above-mentioned lithium-ion battery SOC estimation method, this paper proposes a lithium-ion battery SOC estimation method based on double DQN and EKF method. And based on the lithium-ion battery energy storage prediction, a Markov Decision Process (MDP) model was designed to achieve the optimization of EKF parameters. The simulation results show that, compared with the DQN algorithm, the proposed double DQN can effectively solve the problem of dimensionality catastrophe and value function overestimation, and has good robustness, adaptive optimization capabilities, and significantly improves the estimation accuracy of SOC.

2. ECM

2.1. Lithium-ion battery modeling
The SOC estimation accuracy of lithium-ion batteries and the accuracy of simulating the actual charge and discharge characteristics of lithium-ion batteries depend on the identification accuracy of model parameters[6]. The ECM uses electrical components such as resistors, capacitors, and voltage sources to form a circuit to simulate the dynamic characteristics of the battery's dynamic response. It has the advantages of simple structure and good dynamic adaptability[7]. In order to accurately reflect the charge and discharge characteristics of the battery, this article uses a second-order RC model, as shown in figure 1.

![Figure 1. Equivalent circuit model of lithium-ion battery.](image)

According to the second-order RC model in figure 1, the basic electrical equation can be obtained as

\[
\begin{align*}
U_1 & = U_i + U_2 + I_c R_c + U_{oc} \\
I_c & = U_1 / R_1 + C_1 \frac{dU_1}{dt} = U_2 / R_2 + C_2 \frac{dU_2}{dt}
\end{align*}
\]

(1)

The SOC is the ratio of the battery's remaining capacity to the nominal capacity at a certain discharge rate. The formula is as follows:

\[
SOC(t) = SOC(t_0) - \frac{1}{C_n} \int_{t_0}^{t} \eta I_c(t) dt
\]

(2)

According to the open circuit voltage method, there is a linear relationship between the voltage source voltage in the ECM and the SOC of the battery. It’s expression is

\[U_{oc} = f(SOC)\]

(3)

2.2. Battery model parameter identification
The selected research object is the 18650EC battery of Tianjin Li’shen Company, and the SOC-OCV characteristic curve is obtained through charge and discharge experiments.
3. Lithium-ion battery SOC estimation algorithm based on double DQN and EKF

3.1. Build MDP model
Reinforcement learning (RL) continuously interacts with the environment through the agent, and performs trial and error learning according to the feedback information of the environment, and then optimizes the decision-making action, with the goal of finding the optimal strategy or the maximum reward[8]. From the point of view of mathematics, RL is a MDP with Markov properties. By searching for an optimal strategy, the overall reward function in the decision-making process is expected to be optimal. Usually expressed as \((S, A, P, R)\), where \(S\) is the state space, \(A\) is the control action, \(P\) is the transition probability between states, and \(R\) is the reward function. In this paper, the ECM and EKF of lithium-ion battery are regarded as the environment, and the double DQN algorithm based on lithium-ion battery energy storage prediction is used as an agent. Therefore, the MDP model of lithium-ion battery energy storage prediction is shown in figure 3. Its goal task is to change the parameters of the EKF, thereby improving the accuracy of the estimated value of the state of charge of the lithium-ion battery.

![Figure 3. The MDP model of lithium-ion battery energy storage prediction.](image)

3.2. EKF
The charging and discharging process of lithium-ion batteries is non-linear, so to solve this non-linear problem can only use EKF. The EKF is based on the Kalman filter (KF), linearizes the nonlinear link, expands the coefficient matrix of the state equation and the observation equation by Taylor series, and ignores or approximates the high-order terms of the second and above. Since the system handled by EKF is discretized, the basic electrical equations of the second-order RC model are discretized to form the state equation and the observation equation, which are brought into the EKF system, namely...
Among them, $\Delta t$ is the sampling time, $\eta$ is the charging and discharging efficiency, $Q_r$ is the rated capacity of the battery, $w_k$ is the system noise, and $v_k$ is the observation noise.

### 3.3. Algorithm design based on double DQN

In actual lithium-ion battery work, the energy storage prediction system of lithium-ion battery has a large-scale state space and action space. Deep reinforcement learning (DRL) uses RL to define problems and optimization goals, uses deep learning (DL) to solve strategy functions or value functions, and uses backpropagation algorithms to optimize objective functions. It can solve the problem of the "curse of dimensionality" in traditional Q-learning, that is, the state space of the environment increases exponentially with the increase in the number of system features, which leads to the problem of solving the optimal strategy and optimal value function[9]. The DQN algorithm proposed in 2015 increases the target network, and reduces the correlation between the current Q value and the target Q value by using a dual network structure, greatly improving the stability of the DQN algorithm. Since there is only one neural network in the DQN algorithm, the greedy strategy is adopted when updating the Q value and action selection, so the Q value is often overestimated[10].

In response to the above problems, this article proposes an improved version of the DQN algorithm—double DQN. That is, by using two Q networks to solve the overestimation problem generated in the algorithm learning process. Double DQN makes use of the current Q network to select actions and the target Q network to calculate the target Q value, so as to reduce the problem of overestimates of value function caused by the calculation deviation caused by max Q value calculation. The relationship is as shown in the following formula:

$$U_i(k) = U_{c,k}(k) - R_i I_c(k) - U_{1,k} - U_{2,k} + v_k$$ (5)

3.3. Algorithm design based on double DQN

In actual lithium-ion battery work, the energy storage prediction system of lithium-ion battery has a large-scale state space and action space. Deep reinforcement learning (DRL) uses RL to define problems and optimization goals, uses deep learning (DL) to solve strategy functions or value functions, and uses backpropagation algorithms to optimize objective functions. It can solve the problem of the "curse of dimensionality" in traditional Q-learning, that is, the state space of the environment increases exponentially with the increase in the number of system features, which leads to the problem of solving the optimal strategy and optimal value function[9]. The DQN algorithm proposed in 2015 increases the target network, and reduces the correlation between the current Q value and the target Q value by using a dual network structure, greatly improving the stability of the DQN algorithm. Since there is only one neural network in the DQN algorithm, the greedy strategy is adopted when updating the Q value and action selection, so the Q value is often overestimated[10].

In response to the above problems, this article proposes an improved version of the DQN algorithm—double DQN. That is, by using two Q networks to solve the overestimation problem generated in the algorithm learning process. Double DQN makes use of the current Q network to select actions and the target Q network to calculate the target Q value, so as to reduce the problem of overestimates of value function caused by the calculation deviation caused by max Q value calculation. The relationship is as shown in the following formula:

$$Q(s_i, a_i, \theta) \leftrightarrow r_i + \gamma \max_{a'} Q(s_{i+1}, a', \theta')$$ (6)

Where $\theta$ is the parameter of the current Q network and $\theta'$ is the parameter of the target Q network. If the Q value overestimates action, select it. Then the Q value will provide the appropriate value. If the Q value is overrated, the Q value will not select an action.

The target Q value based on double DQN is calculated as follows:

$$Y_{i, \text{Double DQN}} = R_{i+1} + \gamma Q(s_{i+1}, \max_{a} Q(s_{i+1}, a, \theta'), \theta')$$ (7)

The algorithm has two major characteristics: one is to use the BP neural network parameterized approximation value function to achieve generalization, and the other is to train two BP neural networks to complete the action selection and action evaluation to solve the problem of overestimation of the value function. The neural network parameters are updated by the gradient descent method, and the relationship is as follows:

$$\theta' = \theta + [\gamma \max_{a} Q(s_i, a, \theta') - Q(s_i, a_i, \theta')] \nabla Q(s, a, \theta)$$ (8)
3.4. Operation process of lithium-ion battery SOC estimation based on double DQN and EKF

3.4.1 Setting of state variable parameters
The choice of state variable type has an important influence on performance control. In order to prevent the depth neural network from falling into local optimization after training, it is generally necessary to observe and train fewer state variables. However, in the absence of key state variables, the double DQN algorithm may not fully understand the interaction between the state transition process and the environment, resulting in iteration and convergence difficulties. Therefore, the state function of the equivalent circuit model and the function that caused the SOC estimation error in the EKF were selected as the state variables, that is,

\[ s_k = \left\{ Q_k, R_k, \hat{x}_k, P_k, U_k, (k), U_{1,k}, U_{2,k} \right\} \]  

3.4.2 Setting of action parameter
When setting the action parameters, discretization should be performed and appropriate discretization steps should be set. If the discretization step size is too small, the training time of the algorithm will be increased, and if the discretization step size is too large, it will result in the local optimization of the algorithm. According to the assumption, EKF provides the observed variance in EKF for the double DQN module, and the action parameters are as follows:

\[ \text{Action} = \left\{ \text{increasing } R \text{ 10 times, increasing } R \text{ 5 times, maintaining } R, \text{ decreasing } R \text{ 5 times, decreasing } R \text{ 10 times} \right\} \]  

3.4.3 Parameter setting of reward function
The reward function is used to evaluate the quality of the action value in a given state, which mainly depends on the estimation error of SOC, which is defined as:

\[ \text{reward} = \frac{10}{|\text{SOC error}| + \text{learning rate}} \]  

4. Simulation results and analysis
The simulation data comes from figure 2, and the EKF absolute estimation error estimation curve is shown in figure 4. The true initial value of SOC is 1, and the experimental initial value is 0.9. For the double DQN algorithm, the hyperparameter was set according to the simulation requirements and artificial experience, the learning rate was 0.001 and the initial probability of the greedy strategy was 0.4. The average value of the cumulative reward in the training iteration based on the DQN and double DQN algorithm is shown in figure 5. Compared with DQN, the strategy based on double DQN shows better performance and higher average return value in terms of convergence speed, indicating that double DQN can produce more accurate value estimation and better strategy at the same time.
5. Conclusion
In this paper, DRL is used to optimize the parameters of the EKF method, and the original SOC estimation method is improved to improve the SOC estimation performance. Through experimental testing and data fitting, a lithium-ion battery OCV-SOC relationship curve model is proposed. A block diagram of double DQN for optimizing the parameters in the EKF for energy storage prediction is established. The state function of the equivalent circuit model and the function that causes the error of the SOC estimation in the EKF are selected as the state variables. The state variables are used to change the observed variance R in EKF to estimate the state of charge. The results prove the effectiveness and applicability of the proposed method to improve the SOC convergence performance. Although the accuracy has improved, the computational complexity has increased. Therefore, it needs to be further simplified and improved in subsequent research.

Acknowledgments
This paper was funded by the National Natural Science Foundation (No.60771014); Tianjin Science and Technology Support Foundation of China (No.17YFZCNC00230); Tianjin Natural Science Foundation of China (No.13JCZDJC29100); Tianjin Sci-tech Commissioner Foundation of China (No.15JCTPJC64100).

References
[1] Wu, T., Ji, F., Liao, L. & Chang, C. (2019) Voltage-SOC balancing control scheme for series-connected lithium-ion battery packs. J. Energy Storage, 25: 100-895.
[2] Zhang, L., Hu, X., Wang, Z., Sun, F. & Dorrell, D. G. (2018) A review of supercapacitor modeling, estimation, and applications: A control/management perspective. Renewable and Sustainable Energy Reviews, 81:1868-1878.
[3] Kim, M., Kim, K., Kim, J., Yu, J. & Han, S. (2018) State of Charge Estimation for Lithium Ion Battery Based on Reinforcement Learning. IFAC-PapersOnLine, 51: 404–408.
[4] Guo, Y., Zhao, Z. & Huang, L. (2017) SOC Estimation of Lithium Battery Based on Improved BP Neural Network. Energy Procedia, 105: 4153-4158.
[5] Luo, J., Peng, J. & He, H. (2019) Lithium-ion battery SOC estimation study based on Cubature Kalman filter. Energy Procedia, 158: 3421–3426.
[6] Lai, X., Zheng, Y. & Sun, T. (2018) A comparative study of different equivalent circuit models for estimating state-of-charge of lithium-ion batteries. Electrochimica Acta,259: 566-577.
[7] Rodriguez, A. & Plett, G. L. (2017) Controls-oriented models of lithium-ion cells having blend electrodes. Part 1: Equivalent circuits. J. Energy Storage,11: 162–177.
[8] Barrett, E., & Linder, S. (2015) Autonomous HVAC Control, A Reinforcement Learning
Approach. Lecture Notes in Computer Science, 3–19.

[9] Yan, Z., & Xu, Y. (2018) Data-driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method with Continuous Action Search. IEEE Transactions on Power Systems, 1–1.

[10] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015) Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.