International Journal of Research in Interdisciplinary Studies
Volume 2, Issue 12, December 2024
https://www.ijris.com | ISSN (Online): 2584-1017

Research on Battery Remaining Capacity Estimation Based on Radial Basis Kernel Function-Support Vector Regression Model

Bin Xiao*
School of Information Engineering, Guangzhou Vocational College of Technology & Business, Guangzhou, China

Abstract— **Battery remaining capacity estimation is a critical indicator in battery management systems. Accurate estimation of battery remaining capacity, i.e., State of Health (SoH), can guide the timely recycling and cascade utilization of LiFePO4 batteries, contributing to economic savings and environmental protection. This paper proposes a Support Vector Regression (SVR) model using the Radial Basis Kernel Function (RBF) to estimate battery SoH. Sample data is rapidly obtained from aged batteries through the Hybrid Pulse Power Characterization (HPPC) test, and features are constructed using the minimum, maximum, and average values of the hysteresis curve. Hyperparameters of the RBF-SVR model are determined through literature review and empirical analysis. To validate the proposed method, the RBF-SVR model is trained and tested using LiFePO4 battery samples with varying degrees of aging, demonstrating the method's accuracy and effectiveness.**

Index Terms— **Lithium-ion battery, Support Vector Machine, State of Health estimation.**

1. Introduction

With the rapid development of renewable energy and electric vehicle technologies, batteries, as the core components for energy storage and conversion, significantly influence the overall efficiency and reliability of the system. The State of Health (SOH) of a battery is a critical indicator measuring the difference between the current performance and the performance of a new battery, directly affecting its lifespan, safety, and economic viability. Therefore, accurate estimation of battery SOH is of paramount importance for optimizing Battery Management Systems (BMS), extending battery life, enhancing energy utilization efficiency, and reducing maintenance costs.

A. Research Background and Significance

Estimating battery SOH is a complex and challenging problem influenced by various factors, including the number of charge-discharge cycles, depth of discharge, temperature, and self-discharge rate. Traditional SOH estimation methods primarily rely on capacity and internal resistance tests, which are straightforward but have limitations in practical applications, such as long testing times, complex operations, and difficulty in real-time monitoring. In recent years, with the rapid development of data-driven methods and machine learning technologies, model-based and data-driven SOH estimation methods have become research hotspots. These methods can more accurately estimate battery SOH by establishing battery aging models or training prediction models using historical data, providing real-time health monitoring. Accurate SOH estimation has multiple research implications:

1) *Optimizing Battery Management Systems:* Real-time monitoring of battery SOH enables BMS to more effectively control charge-discharge, thermal management, fault diagnosis, and maintenance decisions, thereby improving battery efficiency and safety.
2) *Extending Battery Life:* Accurate SOH estimation helps identify battery aging trends, allowing timely adjustments to usage strategies to avoid overcharging, over-discharging, and high temperatures, thereby extending battery life.
3) *Improving Energy Utilization Efficiency:* Optimizing battery usage and management can enhance energy storage and conversion efficiency, reduce energy waste, and promote the widespread application of renewable energy.
4) *Reducing Maintenance Costs:* Accurate SOH estimation can help predict the remaining lifespan of batteries, enabling rational maintenance and replacement planning, reducing maintenance costs and downtime. In summary, battery SOH estimation is an important research topic in the field of battery management, with broad application prospects and profound research significance. This study aims to explore and develop data-driven SOH estimation methods, achieving accurate, real-time, and reliable estimation of battery health status by combining machine learning techniques and battery aging samples, providing theoretical support and practical guidance for the optimization and application of battery management systems.

B. Literature Review

With the rapid development of electric vehicles, accurate estimation of the health status of lithium-ion batteries, as the core power source, is crucial for ensuring the safe and efficient operation of battery systems [1]. Currently, SOH estimation

*Corresponding author: binxiao84@126.com

methods are mainly divided into direct measurement methods and indirect analysis methods [2]. Direct measurement methods include capacity measurement, internal resistance measurement, and impedance measurement. Although these methods are highly accurate, they are primarily suitable for laboratory environments and difficult to implement in practical applications [3]. Indirect analysis methods mainly include model-based methods and data-driven methods. Model-based methods require high-fidelity battery models to describe the nonlinear characteristics of batteries, such as Equivalent Circuit Models (ECM), Electrochemical Models (EM), and empirical models [4].

In recent years, with the development of big data and artificial intelligence technologies, data-driven methods based on machine learning have gained widespread attention in the field of SOH estimation [5]. These methods treat the battery as a black box, requiring no detailed knowledge of battery characteristics, and mainly include shallow neural networks, deep learning, Support Vector Machines (SVM), and Gaussian Process Regression (GPR) algorithms [6]. Shallow neural networks, including BP neural networks, RBF neural networks, and extreme learning machines, have the characteristics of simple structure and high computational efficiency [7]. Deep learning algorithms such as DNN, RNN, and CNN can automatically extract features and perform well in processing large-scale data [8]. SVM transforms nonlinear problems into linear problems through kernel functions, with good generalization ability [9]. GPR, based on probabilistic statistical theory, can provide uncertainty estimation for predictions [10].

Currently, machine learning in battery SOH estimation still faces the following challenges: (1) data quality issues, including data collection, cleaning, and labeling; (2) model structure selection and hyperparameter tuning; (3) practical deployment and computational efficiency of algorithms. Future research directions can focus on developing hybrid algorithms to improve estimation accuracy, enhancing the generalization ability of algorithms, and reducing computational complexity.

2. Problem Description

The capacity degradation of LiB mainly comes from three aspects: loss of active material in the negative electrode, loss of active material in the positive electrode, and loss of lithium inventory. This includes aging mechanisms such as the formation of the Solid Electrolyte Interface (SEI) and lithium plating, where recyclable lithium is consumed by side reactions.

As an indicator of battery degradation, there is no unified definition of battery SoH to date. Several concepts represent SoH, such as battery capacity, internal resistance, and cycle count. The capacity ratio is commonly used to define SoH, as shown in the following equation.

$$SoH_i = \frac{C_i}{C_0} \tag{1}$$

where SoH_i is the SoH value after the i cycle, C_i is the battery capacity after the i-th cycle, and C_0 is the initial battery capacity.

3. Modeling Method

A. Feature Construction

Due to battery aging, the voltage response varies with the degree of aging, which is similar to the battery hysteresis phenomenon. Therefore, features are constructed using the voltage response under the HPPC test, which includes both charging and discharging. The results of the experimental test are shown in Figure 1. Figure 1(a) shows the current pulse, Figure 1(b) shows the voltage response, and Figure 1(c) shows the features constructed from the hysteresis curve. In the HPPC test of the battery, the hybrid pulse current discharges at a current rate of 0.06C(A) and charges at a current rate of 0.06C(A), followed by discharging and charging at current rates of 0.12C(A) and 0.18C(A), respectively. In Figure 1(a), the duration of the current pulse is 10s, the relaxation time is 30s, and the amplitudes are 0.06C(A), 0.12C(A), and 0.18C(A), respectively. Additionally, there is a relaxation time after each pulse charge and discharge. Through the HPPC battery pulse test, the battery response voltage shown in Figure 1(b) can be obtained. Then, the hysteresis curve shown in Figure 1(c) is obtained from the difference between the charging and discharging response voltages. Therefore, the battery hysteresis curve can be obtained by subtracting the charging and discharging curves of the same rate. Finally, features (F1i, F2i, F3i, i = 1, 2, 3) are constructed using the minimum, maximum, and average values of the hysteresis curve.

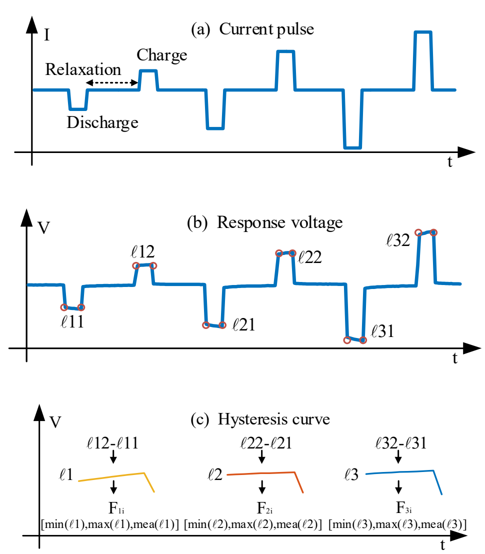

Fig. 1. The battery voltage response of current pulse test

B. RBF-SVR Model for Battery SoH Estimation

The classifier Support Vector Machine (SVM) was developed by Vapnik and his colleagues [11] and is currently widely used for classification and regression tasks. Compared to other machine learning algorithms, SVM has demonstrated superior performance in handling small sample sizes, nonlinear, and high-dimensional datasets. The theoretical foundation of SVM is based on the VC dimension theory and the principle of minimizing structural risk, which consists of empirical risk and confidence intervals. The training samples are represented as:

$$\{(x_1, y_1), \cdots, (x_i, y_i), \cdots (x_n, y_n)\} \tag{2}$$

where $x_i \in X$ is the feature variable, $y_i \in R$ is the true label value, and $i = 1, 2, \cdots, n$ is the number of samples. There exists an optimal hyperplane to separate the training dataset:

$$w \cdot \varphi(x) + b = 0 \tag{3}$$

where $\varphi(x)$ is the mapping function that can improve model performance, w is the weight vector, and b is the bias term.

For the RBF-SVR regression model, the hyperplane can be transformed into the following optimization problem:

$$\begin{cases} \min R(w, C, \xi) = \frac{1}{2}\|w\|^2 + \frac{1}{2} C \sum_{i=1}^{N} \xi_i^2 \\ s.t.\ y_i = w^T \varphi(x_i) + b + \xi_i, i = 1, 2, ..., N \end{cases} \tag{4}$$

where ξ is slack variables introduced to measure the estimation error of the RBF-SVR model, and C is the penalty factor used to balance the model flatness and empirical risk. By solving the above constrained problem, the regression model can be written in the following form:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{5}$$

where $K(x_i, x)$ is defined as:

$$K(x_i, x) = \varphi(x_i) \cdot \varphi(x) \tag{6}$$

Thus, the structure of the RBF-SVR model for battery SoH prediction is illustrated in Figure 2.

4. Experiments and Analysis

A. Experiment Introduction

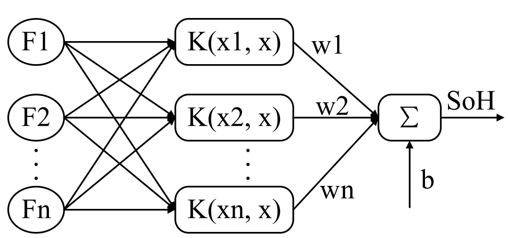

Fig. 2. Schematic diagram of RBF-SVM model for SoH estimation

The battery test platform is shown in Figure 3. The test platform consists of LiFePO4 batteries, a self-developed battery tester, and a computer. The cycle life of the LiFePO4 batteries used in the test ranges from 1000 to 2500 cycles.

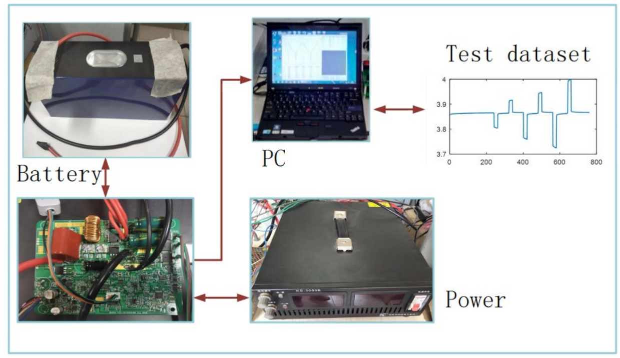

Fig. 3. The experiment setup of test platform

First, the test batteries are rapidly aged using the CC-CV charging/discharging method, with a charging and discharging current of 0.18C (42.12A), and the number of battery cycles is set to 10, 30, 50, 100, and 150, respectively. Therefore, by using a large charging and discharging rate, the battery aging can be accelerated, and batteries with different degrees of aging can be obtained after multiple cycles. Second, after the battery aging test, battery sample data is collected using the HPPC test method. The test battery is first discharged at a current of 0.06C (14.04A), then charged at a current of 0.06C (A), followed by discharging and charging tests at currents of 0.12C (28.08A) and 0.18C (A), respectively. The duration of the pulse charge and discharge is 10s, and a 30s relaxation time is set after each pulse charge and discharge. To accurately verify the model, the number of battery cycles and Depth of Discharge (DoD) are selected as influencing factors of the battery test. After the battery aging test, the HPPC test is used to collect battery sample data at DoDs of 5%, 10%, 15%, 20%, and 25%, respectively. The sampling time is set to 1 second, and a total of 12 batteries are tested, collecting 60 data samples for model training and 60 data samples for model validation.

B. Model Training

The experiment is conducted on a Windows 10 + PyCharm tool platform. The sequence of data samples fed into the RBF-SVR model is modified based on a random function. To prevent the different dimensions of the original data from affecting the model training process, data preprocessing must be performed. Data normalization is applied to eliminate this influence and improve the convergence speed during model training. The minimum-maximum normalization scales the feature data to a range between 0 and 1 according to the equation:

$$x' = (x - \min_x) / (\max_x - \min_x) \tag{7}$$

where x is the original data, x' is the normalized data, and $\max_x$ and $\min_x$ are the maximum and minimum values of the data, respectively.

Adaptability evaluation plays a crucial role, and the generality of the SVR is selected as the evolution criterion. The performance of the SVR is evaluated using cross-validation

Mean Squared Error (MSE). Specifically, the cross-validation MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(SoH_i - SoH_i^*)^2 \tag{8}$$

where SoH_i is the measured value, SoH_i^* is the estimated value, and n is the number of samples.

To select an appropriate kernel, the SVR model is configured with linear, polynomial, and RBF kernels and evaluated using 5-fold cross-validation. Table 1 shows the cross-validation results, which display the best RBF-SVR parameters and cross-validation errors for the three models. The model configured with a linear kernel produces the largest cross-validation error, indicating the worst evaluation performance on the validation set. The model configured with a polynomial kernel produces a cross-validation error of 6.33%, slightly higher than the cross-validation error of the model with an RBF kernel. This indicates that among the three models, the SVR model configured with an RBF kernel has the best generalization performance. However, due to the introduction of the Gaussian function, the structure of the RBF model is more complex compared to other models. Therefore, a trade-off between model complexity and generalization performance is made, and the polynomial kernel may be the best choice for the implementation of the RBF-SVR model on embedded devices.

Table 1
Cross-validation results

Kernel Function	**Parameter Settings**	**Validation Error**
Linear Kernel	gamma=200	11.91%
Polynomial Kernel	gamma=200, d=6, p=2	6.33%
Radial Basis Function	gamma=200, sig2=1.2	0.49.5%

C. Model Performance Metrics

Five statistical parameters, namely R2, RMSE, MAE, MAPE, and SoH Estimation Error, are used as performance indicators to verify the comprehensive performance of the SVR estimation model. R2 represents the explanatory power of the input variables on the output variables. RMSE represents the sample standard deviation of the difference between the estimated and actual values. MAE, MAPE, and Error are used to determine the error range of the SoH estimation results. An R2 of 1 is an ideal model, and the closer the values of RMSE, MAE, and Error are to 0, the better the model performance. Each performance indicator is defined as follows:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(SoH_k - SoH_k^*)^2}{\sum_{k=1}^{n}(SoH_k - M_{soh})^2} \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{n}(SoH_k - SoH_k^*)^2} \tag{10}$$

$$MAE = \frac{1}{N}\sum_{k=1}^{n}| SoH_k - SoH_k^* | \tag{11}$$

$$MAPE = \frac{1}{N}\sum_{k=1}^{n}| \frac{SoH_k - SoH_k^*}{SoH_k} | \tag{12}$$

$$Error = \frac{SoH_k - SoH_k^*}{SoH_k} \tag{13}$$

Where SoH_k is the actual measured value, SoH_k^* is the model estimated value, M_{soh} is the mean value of SoH, and n is the number of test samples.

D. Model Validation

To demonstrate the comprehensive performance of the estimation model and the rationality of the kernel function selection, the RBF-SVR model was trained using the training set and then validated using three test sets. The measured battery SoH was taken as the actual SoH and normalized to a range of 0% to 100%. As shown in Table 2, the statistical test results indicate that the estimation results on Test Set 3 were the worst, with RMSE, MAE, MAPE, and R2 values of 0.5279%, 0.3499%, 0.4199%, and 99.94%, respectively. The estimation results obtained on Test Set 2 were better, with RMSE, MAE, MAPE, and R2 values of 0.3834%, 0.2822%, 0.3766%, and 99.95%, respectively, exceeding the estimation accuracy from Test Set 3. The estimation results on Test Set 1 were slightly better than those from Test Sets 2 and 3, with RMSE, MAE, MAPE, and R2 values of 0.3875%, 0.2546%, 0.3233%, and 99.95%, respectively. Table 2 Statistical results of SoH estimation input variables and battery SoH through the Gaussian function.

Table 2
Statistical results of SoH estimation

Data	**RMSE**	**MAE**	**MAPE**	**R^2 (%)**	**Error (%)**
Test 1	0.3875	0.2546	0.3233	99.95	[-0.82 1.34]
Test 2	0.3834	0.2822	0.3766	99.95	[-0.82 1.09]
Test 3	0.5279	0.3499	0.4199	99.94	[-1.61 0.99]
aver	0.4329	0.2956	0.3733	99.95	[-1.61 1.34]

As shown in Table 2, the best test results obtained from the RBF-SVR model were on Test Set 1. In this case, the R2 metric was the best among all tests, indicating that the battery SoH estimation curve showed the best fit. In contrast, the worst test results were obtained from Test Set 3. Although the results of the three tests were different, the overall performance indicators were still ideal, especially the total error range of battery SoH estimation was between [-1.61% 1.34%]. Therefore, the above test results are sufficient to show that the RBF-SVR model can overcome the nonlinear relationship between

E. Impact of Kernel Function on the Model

To illustrate the rationality of the kernel function selection, two additional SVR models with linear and polynomial kernel functions were trained using the training set and then tested using the three test sets, taking their average values. The SoH estimation results for each model are shown below.

Table 3
Statistical results of SoH estimation

Kernel	RMSE	MAE	MAPE	R^2	Error
Linear Kernel	11.77	9.98	14.46	59.27	[-37.91 43.85]
Polynomial	6.38	4.94	6.95	88.76	[-15.36 24.51]

As shown in Table 3, the statistical test results indicate that the SVR model using a linear kernel produced the worst average estimation results on the test sets, with RMSE, MAE, MAPE, and R2 values of 11.77%, 9.98%, 14.46%, and 59.27%, respectively. The SVR model using a polynomial kernel produced better estimation results, with RMSE, MAE, MAPE, and R2 values of 6.38%, 4.94%, 6.95%, and 88.76%, respectively, which clearly exceeded the estimation accuracy of the model with a linear kernel. According to Table 2, the estimation results of the RBF-SVR model were superior to those of the models with polynomial and linear kernels. Although the SVR model with a linear kernel did not obtain satisfactory estimation results, the other two models achieved high-precision battery SoH estimation. In particular, the RBF-SVR model achieved accurate estimation across the entire test set, indicating that the trained model has good robustness and generality.

5. Conclusion

In this study, we focused on model design and feature construction to establish a battery SoH estimation method based on machine learning models. On one hand, the RBF-SVR was chosen to build the battery SoH estimation model because the SVR model is widely used to solve classification and regression problems and has superior performance on small, nonlinear, and high-dimensional datasets. On the other hand, the response voltage measured using the HPPC method was used for feature construction, which is a simple, effective, and convenient non-destructive, non-invasive method for short-term feature acquisition in engineering applications. Based on this study, the following conclusions can be drawn: This paper proposes a new method for estimating the SoH of power batteries, which requires only a few short-term feature samples and can effectively perform on-site rapid measurements without relying on ECM, complex mathematical calculations, or time-consuming parameter adjustments. Finally, through model validation, the superior comprehensive performance of the proposed method is demonstrated.

Funding

The research is funded by the Scientific Research Project (2022YB14) of Guangzhou Vocational College of Technology & Business.

References

[1] Zuolu Wang, Xiaoyu Zhao, Lei Fu, Dong Zhen, Fengshou Gu, Andrew D. Ball. A review on rapid state of health estimation of lithium-ion batteries in electric vehicles. Sustainable Energy Technologies and Assessments, Volume 60, December 2023, 103457.

[2] Javier Urquizo, Pritpal Singh. A review of health estimation methods for Lithium-ion batteries in Electric Vehicles and their relevance for Battery Energy Storage Systems. Journal of Energy Storage, Volume 73, Part D, 20 December 2023, 109194.

[3] Wang Y, et al. A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. Renewable and Sustainable Energy Reviews, 2020.

[4] Maheshwari Adaikkappan, Nageswari Sathiyamoorthy. Modeling, state of charge estimation, and charging of lithium-ion battery in electric vehicle: A review. Int J Energy Res. 2022;46:2141–2165.

[5] Kaushik Das, Roushan Kumar. Electric vehicle battery capacity degradation and health estimation using machine-learning techniques: a review. Clean Energy, Volume 7, Issue 6, December 2023, Pages 1268–1281.

[6] Deepak Kumar, M. Rizwan, Amrish K. Panwar, Robust state of health estimation of commercial lithium-ion batteries based on enhanced hybrid machine learning model for electrified transportation, Electrical Engineering, Oct 2024.

[7] Hongqian Zhao, Zheng Chen, Xing Shu, Jiangwei Shen, Zhenzhen Lei, Yuanjian Zhang. State of health estimation for lithium-ion batteries based on hybrid attention and deep learning. Reliability Engineering & System Safety, Volume 232, April 2023, 109066.

[8] Xin Sui, Shan He, Søren B. Vilsen, Jinhao Meng, Remus Teodorescu, Daniel-Ioan Stroe. A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery. Applied Energy, Volume 300, 15 October 2021, 117346.

[9] Y Zhang, Y Liu, J Wang, T Zhang - Energy. State-of-health estimation for lithium-ion batteries by combining model-based incremental capacity analysis with support vector regression. Energy, vol 239, Part B, Jan 2022, 121986.

[10] R Zhang, CH Ji, X Zhou, T Liu, G Jin, Z Pan, Y Liu. Capacity estimation of lithium-ion batteries with uncertainty quantification based on temporal convolutional network and Gaussian process regression. Energy. Volume 297, 15 June 2024, 131154.

[11] Cortes C., Vapnik V. N., Support-vector network, [J] Machine Learning, 1995, 20: 273-297.