Article

A Two-State-Based Hybrid Model for Degradation and Capacity Prediction of Lithium-Ion Batteries with Capacity Recovery

Yu Chen [1,2,3], Laifa Tao [1,2,3], Shangyu Li [1,2,3], Haifei Liu [1,2,3] and Lizhi Wang [4,*]

1 Institute of Reliability Engineering, Beihang University, Beijing 100191, China; 13141043@buaa.edu.cn (Y.C.); taolaifa@buaa.edu.cn (L.T.); lishangyu@buaa.edu.cn (S.L.); phoebeliu@buaa.edu.cn (H.L.)
2 Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing 100191, China
3 School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
4 Unmanned System Institute, Beihang University, Beijing 100191, China
* Correspondence: wanglizhi@buaa.edu.cn

Abstract: The accurate prediction of Li-ion battery capacity is important because it ensures mission and personnel safety during operations. However, the phenomenon of capacity recovery (CR) may impede the progress of improving battery capacity prediction performance. Therefore, in this study, we focus on the phenomenon of capacity recovery during battery degradation and propose a hybrid lithium-ion battery capacity prediction framework based on two states. First, to improve the density of capacity-related information, the simultaneous Markov blanket discovery algorithm (STMB) is used to screen the causal features of capacity from the initial feature set. Then, the life-long cycle sequence of batteries is partitioned into global degradation regions and recovery regions, as part of the proposed prediction framework. The prediction branch for the global degradation region is implemented through a long short-term memory network (LSTM) and the other prediction branch for the recovery region is implemented through Gaussian process regression (GPR). A support vector machine (SVM) model is applied to identify recovery points to switch the branch of the prediction framework. The prediction results are integrated to obtain the final prediction results. Experimental studies based on NASA's lithium battery aging data highlight the trustworthy capacity prediction ability of the proposed method considering the capacity recovery phenomenon. In contrast to the comparative methods, the mean absolute error and the root mean square error are reduced by up to 0.0013 Ah and 0.0043 Ah, which confirms the validity of the proposed method.

Keywords: lithium-ion battery; causal feature; capacity recovery phenomenon; capacity prediction; long short-term memory; Gaussian process regression

1. Introduction

As an increasingly important energy source, lithium-ion batteries offer many advantages, including high energy density, rechargeability, recyclability, and low environmental impact, and they are widely used in key areas such as electric vehicles, consumer electronics, and aerospace [1]. However, battery capacity gradually decreases with use, and when the capacity drops to 80% of the rated capacity, lithium-ion batteries are considered faulty batteries [2,3]. Accurately characterizing and predicting the aging state of lithium-ion batteries assists in the development of maintenance and replacement strategies for them, thereby reducing the utilization risk [4].

There have been numerous publications focusing on remaining useful life (RUL) prediction methods for lithium-ion batteries. Model-based methods and data-driven methods are the two main strategies [5]. For model-based methods, the goal is to model the internal degradation process of lithium batteries. Model-based methods commonly include mechanistic models, equivalent circuit models (ECMs), and empirical models [6]. In addition, algorithms such as the Kalman filter and particle filter have been applied as prediction models [7,8]. However, the aging process of a battery is the result of the

interaction of various factors during the battery's use phase. The complex state changes within the battery increase the difficulty of building the model and reduce the prediction accuracy [9]. Data-driven methods, on the other hand, achieve battery life prediction based on monitoring parameters only. For example, Patil et al. [10] proposed a multilevel SVM-based capacity prediction method for lithium batteries. Zhou et al. [11] utilized Box–Cox transformation for battery feature extraction and a capacity degradation model with an RVM for battery life prediction. In addition, deep learning models have been widely used for lithium battery life prediction tasks. Zhang et al. [12] and Ren et al. [13] utilized recurrent neural networks to obtain the long-term correlation of the battery degradation process and realized accurate prediction results. In the battery life prediction task, capacity and impedance are usually selected as health indicators. Capacity data are widely used in battery life prediction tasks because they are easy to calculate and do not require additional measurement work. In this case, the capacity recovery phenomenon is a major challenge for the prediction task.

The phenomenon of capacity recovery (CR) (also known as capacity regeneration) refers to that of battery capacity recovery after a suspension of charge/discharge cycles. In early studies [14,15], this phenomenon was regarded as unpredictable perturbation information, and the lithium battery RUL was predicted by separating the effects of CR. However, in practice, this approach leads to an increase in prediction error due to the occurrence of the CR phenomenon [16–18]. Eddahech et al. [19] revealed the predictability of the CR phenomenon by discovering the potential connection between the CR phenomenon and relaxation time as well as discharge depth. Based on this, Qin et al. [20] and Deng et al. [21] clarified the resting time as an indicator to evaluate the CR phenomenon. In addition, Ma et al. [22] combined particle filtering and the Mann–Whitney U test and proposed a CR point detection method to decrease the prediction error caused by the CR phenomenon. Pang et al. [23] and Qiao et al. [24] regarded the capacity recovery phenomenon and the battery global degradation process as different prediction tasks.

Based on the above work, we take the CR effect during the battery degradation process into account. To begin with, we focus on an interesting phenomenon: Taking NASA lithium-ion battery experimental data as an example, although there are significant differences in the global degradation trends of #5, #6, and #7 batteries, the occurrence of the CR phenomenon has astonishing consistency. The research conclusions in reference [19] also prove that the CR phenomenon is more related to relaxation time. This phenomenon indicates that although the degradation process of batteries has strong specificity [25], there is still an opportunity to use relaxation time features to stably predict the CR phenomenon. Intuitively, attempting to predict the CR phenomenon and global degradation trend separately can help obtain more accurate prediction results. Two studies [23,24] followed this approach and achieved good prediction results. However, we further considered a situation where a certain difference exists between the accelerated degradation after the CR point [20] and the global degradation trend, in which case the use of a single degradation model for capacity prediction leads to unanticipated errors. To the best of our knowledge, research on analyzing and utilizing this difference to improve prediction performance has not yet been addressed. Therefore, in this paper, we consider designing separate capacity prediction models for the two different degradation states due to CR, and integrating the prediction results of the two prediction models through the CR recognition method. In summary, this paper designs a hybrid prediction framework that considers the CR effect, attempting to improve the prediction accuracy of battery capacity degradation and, to some extent, improve the prediction stability that is influenced by the CR effect.

Specifically, we first propose a capacity causal feature extraction method to mine capacity-related causal features from battery lifetime series data, which are used to construct a stable relationship between battery capacity and feature sets to build a more accurate prediction model. Then, due to the accelerated degradation characteristics of the capacity recovery region and the stability of the global degradation region, this paper focuses on addressing the impact of this difference on the prediction method. A hybrid prediction

framework based on LSTM and GPR models that considers the two degradation states of capacity is proposed. Considering that the global degradation region accounts for a large proportion of the battery degradation process, there are sufficient data available for training, whereas the recovery region has a shorter duration and occurs infrequently, resulting in a smaller amount of data being available for training. Therefore, we utilize the good learning capability of LSTM to capture the evolutionary patterns of long-term global degradation sequences. At the same time, we take advantage of the GPR model's good adaptability to the regression problem with small sample data to learn the recovery region degradation information [26]. Moreover, by utilizing the prediction method of CR points, the capacity prediction task is divided into two prediction branches, namely, global degradation prediction and restoration region degradation prediction. By integrating the prediction results of both branches, the final capacity degradation prediction result is obtained. Prediction results of NASA lithium batteries show that the proposed method can accurately predict capacity despite the influence of the CR effect.

The contributions of this study are listed below:

(1) We focus on the negative impact of the CR phenomenon on battery capacity degradation prediction. A hybrid prediction framework considering dual degradation states of battery capacity is proposed. By decoupling the global degraded state prediction process from the recovery region degraded state prediction process, high-precision prediction of Li-ion battery capacity that is not affected by the recovery phenomenon is realized.

(2) A causal feature extraction method based on Bayesian network with MB is introduced to extract the causal feature set of capacity to improve the stability of capacity prediction.

(3) The validation of the proposed method is conducted by using four NASA lithium batteries. Several comparative experiments show that the proposed method has higher prediction accuracy than the existing methods in terms of the CR effect.

The rest of this paper is arranged as follows. The feature set in lithium battery capacity prediction; relevant theoretical and methodological models including STMB, LSTM, and GPR; and the steps of the proposed method are introduced in Section 2. Section 3 presents the dataset for the case application and the prediction results. Section 4 confirms the efficiency of our suggested approach in addressing the capacity prediction issue while considering the CR phenomenon. We demonstrate its superiority over existing methods by using a range of comparative cases. Section 5 provides the conclusions.

2. Materials and Methods

2.1. Feature Settings

When batteries are in constant use, the aging phenomenon gradually becomes apparent. Specific indicators of aging include capacity decrease and impedance increase. The manufacturing characteristics of a battery and the environment in which it is used affect the aging process. Charging and discharging voltage and current levels, discharge depth, and ambient temperature are the main external factors of battery aging [1]. This paper refers to the setup described in the literature [27] and takes current, temperature, and time as the initial characteristics of the battery life cycle. Specific features are as follows: charging current, charging voltage, discharging current, discharging voltage, charging temperature, discharging temperature, the constant-current charging time, the constant-voltage charging time, and the constant-current discharging time. We selected these features as the starting point for studying the battery capacity degradation phenomenon for two reasons. First, these features are easy to monitor and acquire during battery use, avoiding complicated measurement processes (e.g., impedance). Second, sufficient research has already demonstrated the high sensitivity of the above features to the battery capacity degradation trend, which means that these features contain a wealth of battery degradation information [19,20].

In addition, the CR effect is relatively stable compared with the strong specificity of the degradation process of batteries, and the results of related studies have illustrated the important influence of the relaxation time of batteries on the CR phenomenon. Therefore, four features are added to characterize the relaxation time of the battery: same-cycle

charging–discharging interval time, adjacent-cycle charging interval time, adjacent-cycle discharging interval time, and adjacent-cycle discharging–charging interval time. We predict the CR point with the help of these four features and use the results as the basis for switching prediction branches of the hybrid model to realize the prediction of capacity degradation of lithium batteries based on two states.

2.2. Causal Feature Selection

Causal feature selection methods are closely related to Bayesian theory. A Bayesian network is usually represented as a directed acyclic graph (DAG), where the nodes of the graph represent the variables and the edges between the nodes represent the dependencies between the variables [28]. As shown in Figure 1, when the edge $T{\rightarrow}Y$ exists, the variable T is called the direct cause (parents) of Y, and the corresponding Y is the direct effect (children) of T. In addition, the other parents of the variable Y are called the spouses of T.
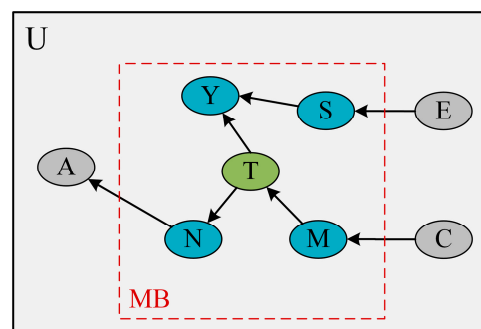

Figure 1. DAG of a Bayesian network.

For a target variable T, its parent, child, and spouse nodes form a Markov blanket of the variable T. Definition 1 describes the Markov blanket.

Definition 1 ([28]). *A Markov blanket of any element $T \in U$ is any subset $MB(T)$ of elements for which*

$$T \perp\!\!\!\perp U \backslash (MB(T), T) | MB(T) \tag{1}$$

Based on the Markov blanket $MB(T)$, other variables $U \backslash (MB(T), T)$ are independent of the target variable T [29]. Therefore, we can obtain the MB of the target variable and use it to construct a causality-based feature subset [30]. There are many research results related to MB discovery, among which the simultaneous Markov blanket method [31] is distinguished by its high recognition accuracy and computational efficiency. As a result, the STMB method is selected to filter the $MB(C)$ of capacity from the initial features of batteries to construct the causal feature set of capacity.

First, the set of parent nodes of the target node is searched, and the standard forward–backward selection strategy is used to filter the parents and children (PC) variables from the full set of variables. Specifically, the set Z is first selected from the set $U \backslash \{T\}$, and a conditional independence test is applied to determine whether $X \perp\!\!\!\perp T | \mathbf{Z}$ exists or not, and if the conditional independence test is passed, the variable X is removed from the set. During the search process, $|Z|$ increases from 1 to the maximum value. After determining the PC set, the STMB enters the second search phase by a single traversal to simultaneously find the mates of the target node and remove the false-positive samples from the PC set, i.e., the non-MB variables that share V-structured paths with the target variable. When the second search step is over, the conditional independent test is supplemented to remove the false-positive variant in the mate set. The MB on the target variable is finally obtained.

2.3. The Long Short-Term Memory Network

Considering the gradient disappearance and explosion after multistage propagation in RNNs when processing time-series data, a long short-term memory network (LSTM) [32] is proposed to solve the problem of long-term dependence. Memory cell state $C(t)$ is used to store information on time series. Gate mechanisms (including input gate, output gate, and forget gate) are introduced to selectively determine which information should be removed from or added to the cell state.

At moment t, the LSTM cell receives three inputs: the cell state $C(t-1)$ and the hidden state $h(t-1)$ at moment $t-1$, and the input vector $x(t)$ at moment t. The cell produces two outputs: the cell state $C(t)$ and the hidden state $h(t)$ at moment t. The memory cell acquires the input $x(t)$ and, via gates, determines whether to receive information from $C(t-1)$ or $h(t-1)$. This allows for the transportation of features between layers within the network structure, facilitating the integration of information at different time points to obtain high-quality features.

Specifically, vectors $f(t)$ and $i(t)$ act first as the outputs of the forget gate and input gate, respectively, through Equations (2) and (3). The cell state $C(t)$ can be obtained via Equation (4).

$$f(t) = \sigma\left(W_f \odot concat(x(t), h(t-1)) + b_f\right) \tag{2}$$

$$i(t) = \sigma(W_i \odot concat(x(t), h(t-1)) + b_i) \tag{3}$$

$$C(t) = f(t) \odot C(t-1) + i(t) \odot \tanh(W_c \odot concat(x(t), h(t-1)) + b_c) \tag{4}$$

Here, W and b are weights and biases, respectively, and $\sigma(\cdot)$ and $tanh(\cdot)$ represent activation functions.

The output gate is then used to obtain the second output $h(t)$ through Equations (5) and (6).

$$o(t) = \sigma(W_o \odot concat(x(t), h(t-1)) + b_o) \tag{5}$$

$$h(t) = o(t) \odot \tanh(C(t)) \tag{6}$$

2.4. Gaussian Process Regression

Gaussian process regression is a non-parametric model used in regression analysis [33]. It utilizes the Gaussian process which ensures that any random variable sample satisfies Gaussian distribution, i.e., $y_i \sim N\left(\mu_i, \sigma_i^2\right)$.

Suppose that $\mathbf{X} = [x_1, x_2, \ldots, x_N]$ is the lithium battery feature vector used for training the model and $\mathbf{y}$ is the corresponding battery capacity. Then, the joint distribution of the sequence satisfies the multivariate Gaussian distribution. Suppose that x^* is the test sample with random noise $\varepsilon \sim N(0, \sigma^2)$ and f^* is the prediction result. The joint probability distribution of the corresponding outputs $\mathbf{y}$ of the training samples and the corresponding outputs f^* of the test samples is as follows [20]:

$$\begin{bmatrix} \mathbf{y} \\ f^* \end{bmatrix} \sim N\left[\begin{bmatrix} \overline{\mathbf{X}} \\ \overline{x}^* \end{bmatrix}, \begin{bmatrix} k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & k(\mathbf{X}, x^*) \\ k(x^*, \mathbf{X}) & k(x^*, x^*) \end{bmatrix}\right] \tag{7}$$

where $\overline{\mathbf{X}}$ is the mean value of $\mathbf{X}$ and $k(\mathbf{X}, \mathbf{X})$ is the covariance function. Under the assumption of zero-mean distribution, $\overline{\mathbf{X}}$ and $\overline{x}^*$ are usually set to 0. And $k(x, x)$ is achieved by selecting the appropriate kernel function.

The posterior distribution of f^* can be derived from the following equation.

$$p(f^* | \mathbf{X}, \mathbf{y}, x^*) = N(\mu^*, \Sigma^*) \tag{8}$$

$$\mu^* = k(\boldsymbol{x}^*, \mathbf{X})[k(\mathbf{X}, \mathbf{X})]^{-1}\boldsymbol{y} \tag{9}$$

$$\Sigma^* = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - k(\boldsymbol{x}^*, \mathbf{X})[k(\mathbf{X}, \mathbf{X})]^{-1}k(\mathbf{X}, \boldsymbol{x}^*) \tag{10}$$

In addition, we use the radial basis function (RBF) as the kernel function.

$$k(x_i, x_j) = \sigma^2 \exp\left(-\frac{\|x_i - x_j\|_2^2}{2l^2}\right) \tag{11}$$

where σ and l are hyperparameters of the kernel function.

To optimize the hyperparameters, i.e., σ and l, the objective of the training phase is to maximize the marginal log-likelihood function as follows.

$$\log p(\boldsymbol{y}|\sigma, l) = -\frac{1}{2}\left[\boldsymbol{y}^T k(\mathbf{X}, \mathbf{X})^{-1}\boldsymbol{y} + \log|k(\mathbf{X}, \mathbf{X})| + N\log(2\pi)\right] \tag{12}$$

2.5. Steps of the Proposed Method

This paper proposes a hybrid capacity prediction model based on LSTM and Gaussian process regression under two states. We note that the battery capacity tends to show an accelerated degradation trend when the CR effect occurs, and this trend disappears after approaching the capacity value before the recovery effect. Based on a previous study [19], we categorize the process of Li-ion battery capacity degradation into two states, i.e., the global degradation state and the recovery region degradation state, and predict the CR point based on the relaxation time in the battery cycle. Further, to exploit its ability to mine temporal correlation properties, the LSTM model is applied to predict the global degradation state. Gaussian process regression is used to model the degradation trajectory of the battery in the capacity recovery region. The set of capacity-related causal features filtered based on other causal features helps to improve the prediction performance. The proposed method is shown in Figure 2.

- Step 1: Selecting Initial Features

First, the monitoring data of the full life cycle of the battery are divided into n charge/discharge cycles. A feature set consisting of nine initial features of the battery is sequentially extracted according to the number of cycles, which is defined as $\{F_C(k)\}_{k=1}^n$, a feature set consisting of four relaxation time features $\{F_T(k)\}_{k=2}^n$, and the capacity value $\{C(k)\}_{k=1}^n$. The relaxation time feature of the kth cycle is the difference between the start time of charging and discharging of the kth cycle, on the one hand, and the k-1th cycle, on the other. Therefore, the relaxation time features are obtained from the second cycle.

- Step 2: Extracting Causal Features of Capacity

The battery capacity values and nine initial features are modeled as a Bayesian network, and the causal features are filtered using the STMB method. Considering the specificity of battery degradation, we analyze each battery cell separately to obtain the causal features corresponding to each cell, defined as $\left\{MB_C^{(i)}(k)\right\}_{k=1}^n$. To exclude the randomness caused by a single battery cell, the concatenated set of causal features of each battery cell is taken as the final feature set $\{MB_C(k)\}_{k=1}^n$ for capacity prediction.

- Step 3: Life Cycle Data Segmentation Based on CR

When the CR phenomenon occurs, the capacity of lithium batteries will significantly increase. In addition, the rate of capacity decline accelerates in the following cycles until the capacity approaches the level before the CR phenomenon. The two states, with significant differences, restrict the accuracy of the unified model for battery capacity prediction. Therefore, we attempt to divide the battery life cycle data into two types of state and model them separately to avoid the decrease in prediction accuracy caused by state transitions.

Specifically, a recovery region is established between the CR point and the point at which the capacity returns to its pre-recovery level, defined as $\{C_{re}(j)\}_{j=1}^{m}$, where m is the number of samples in the regeneration region. The remaining loops form the global degradation region, i.e., $\{C_g(l)\}_{l=1}^{n-m}$ [20]. At the same time, we retain the number of cycles in the original data for the regeneration region and the global degradation region, namely, $\{cycle_{re}(j)\}_{j=1}^{m}$, and $\{cycle_g(l)\}_{l=1}^{n-m}$, to achieve the recombination of the prediction results of the two regions.
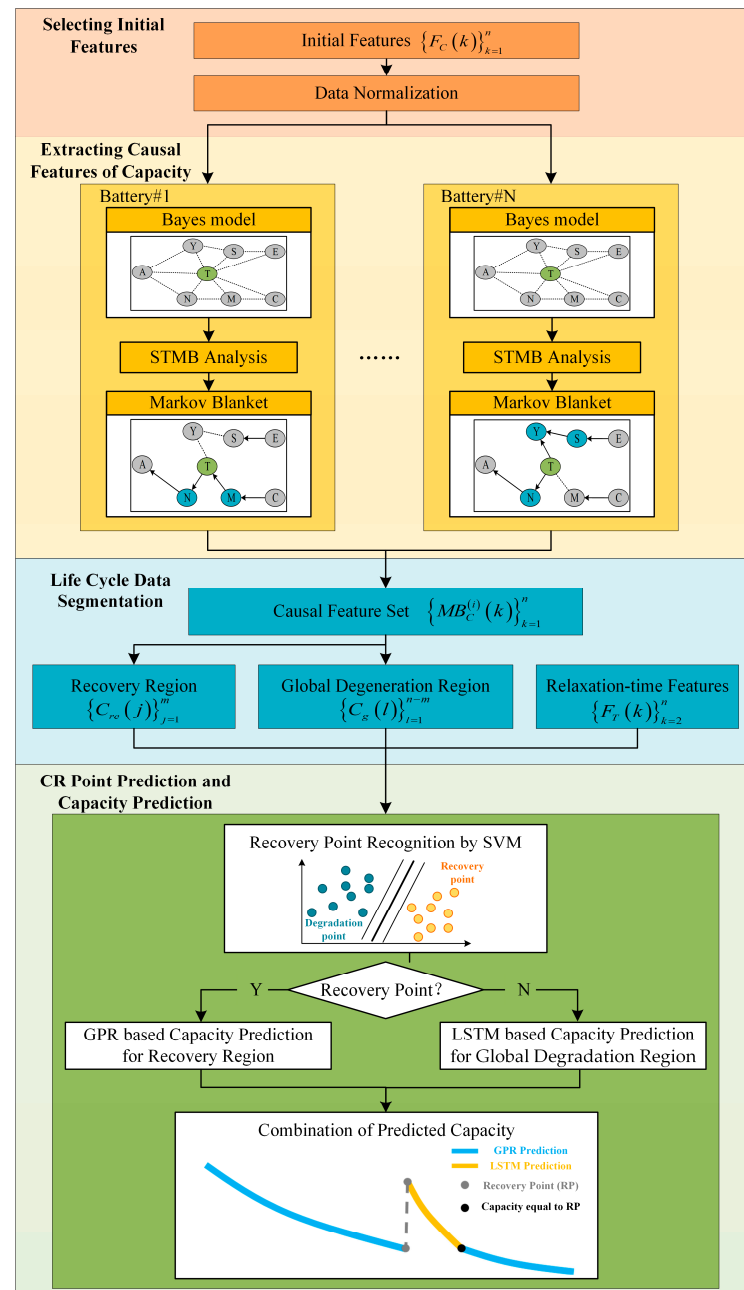

Figure 2. Steps of the proposed method.

- Step 4: CR Point Prediction Based on Relaxation Time

SVM is used to perform CR point prediction. We use the relaxation time feature between the cycle where the CR point is located and the previous cycle as the discriminant feature, i.e., $\{F_T(k)\}_{k=2}^{n}$, and apply a binary classification strategy to set the label of each cycle as 0 or 1, where 0 means that the CR phenomenon has not occurred and 1 means that the CR phenomenon has occurred. Since the number of CR point samples in a single

battery cell is small, and the relationship between the CR phenomenon and relaxation time is relatively stable in different batteries, we constructed uniform training data with CR point samples from all batteries to avoid the adverse effect of an insufficient number of samples on the training process. In addition, according to the number of CR point samples in the training set, an equal amount of normal cycle data is selected as negative samples to participate in the training.

- Step 5: Capacity Prediction Based on Hybrid Model Considering Two Degradation States

In this paper, two types of state (global degradation region and regeneration region) in the life cycle of lithium batteries are modeled using an LSTM network and a GPR model, respectively. The LSTM model is structured as follows: an input layer, two LSTM layers, a dense layer, and finally, an output layer. The feature set contains the causal features of the global degradation region and is constructed using a sliding window with a many-to-one structure. The feature set $\left\{ MB_C^g(l) \right\}_{l=1}^{n-m}$ contains the selected causal features. The input samples are constructed by using a sliding step with a "many-to-one" strategy, i.e., $\left[MB_C^G(t-L), \ldots, MB_C^G(t-1) \right]$, to receive the prediction result $\hat{C}_g(t)$ of the corresponding loop, where L is the window length. During the prediction process, the SVM model continuously predicts whether the CR phenomenon occurs in the next loop. Once the prediction result is 1, the capacity prediction model switches to GPR. The GPR model uses the causal features $MB_C^{re}(t-1)$ of the previous loop in the regeneration region as input and the capacity value $\hat{C}_{re}(t)$ of the current loop as output. The RBF is selected to be the kernel function of the GPR model. Hyperparameters of the kernel function are optimized with the training samples. Until the current capacity value is close to the capacity value before the CR phenomenon, the capacity prediction model is switched back to LSTM. Finally, the prediction results of the two models are combined to obtain a complete battery capacity prediction curve. The capacity prediction process considering the two states is shown in Figure 3.
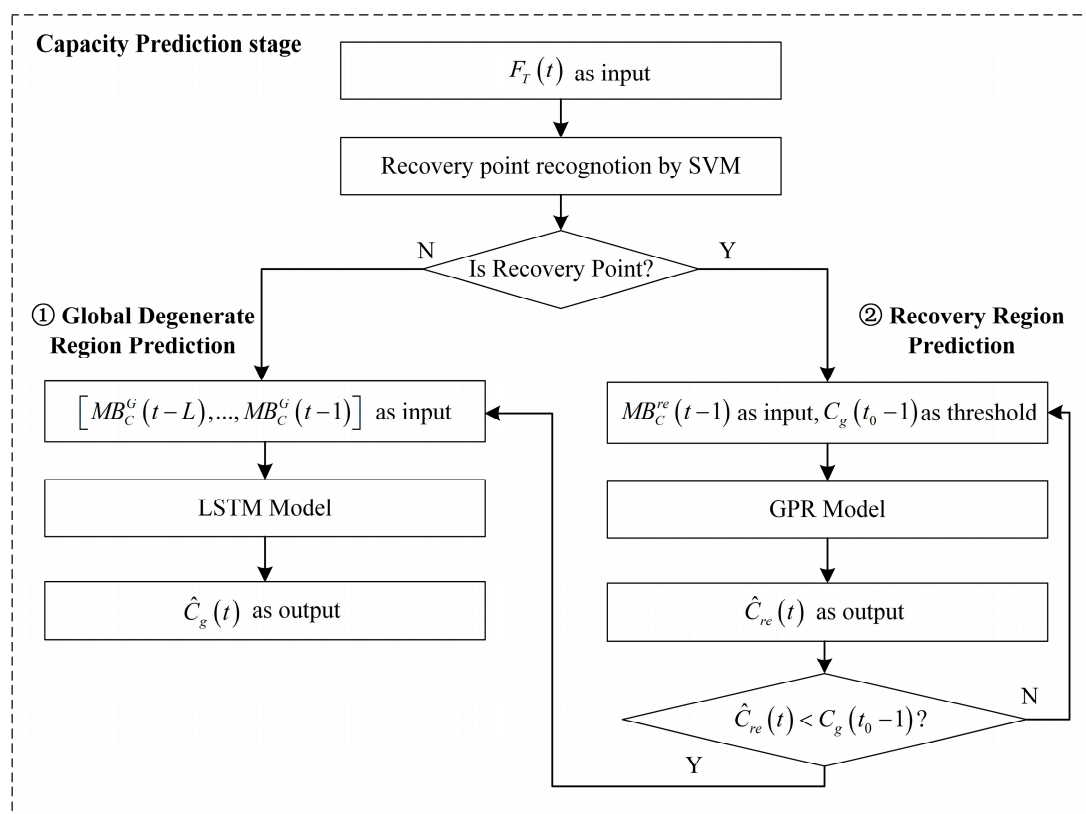

Figure 3. Flowchart of capacity prediction process.

3. Experimental Results

3.1. Dataset Description

We would like to reiterate that this paper focuses on capacity prediction methods for Li-ion batteries in relation to the CR phenomenon. Thus, observable and substantial CR is the key criterion for selecting validation datasets. The battery cycle aging dataset from NASA Ames Research Center [34] provides test data for several battery cells with CR. Thus, test data from four individual cells, namely, cells #5, #6, #7, and #18, are used in this paper. These cells are all 18650-type general-purpose size and have a rated capacity of 2 Ah. The aging experiments were conducted at 24 °C and consisted of three operational profiles, namely, charging, discharging, and impedance measurements. During the charging process, a constant current (1.5 A) was first applied, until the charging voltage increased to 4.2 V, and the voltage was maintained at 4.2 V until the charging current decreased to 20 mA. During the discharging process, the discharge current was maintained at 2 A until the discharge voltage decreased to a specific level, which is 2.7 V, 2.5 V, 2.2 V, and 2.5 V for cells #5, #6, #7, and #18, respectively. Capacity trends and capacity change rates of the four cells in the aging experiment are presented in Figure 4.
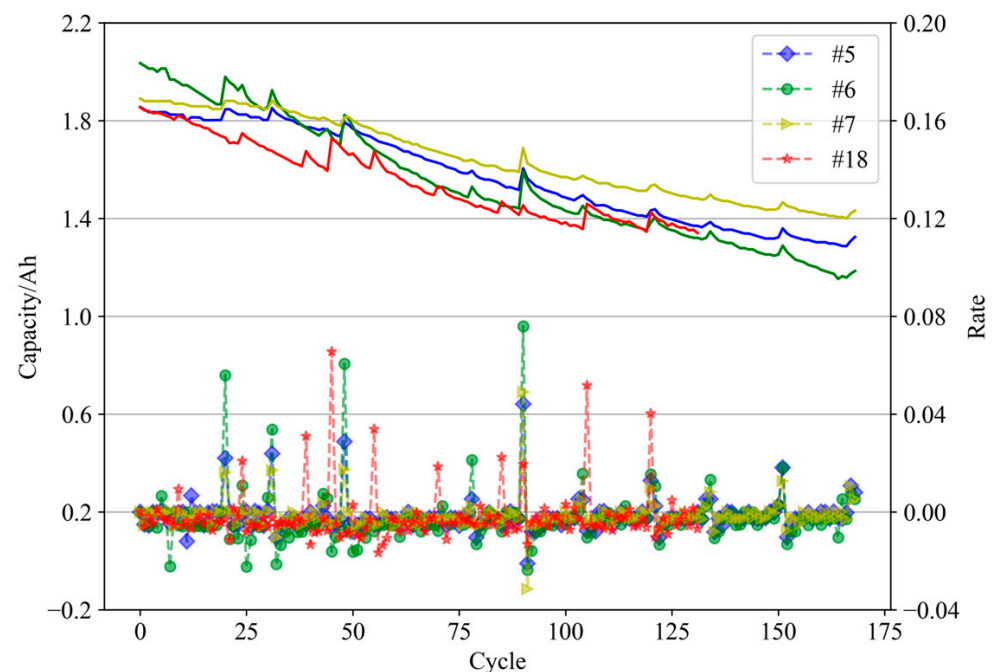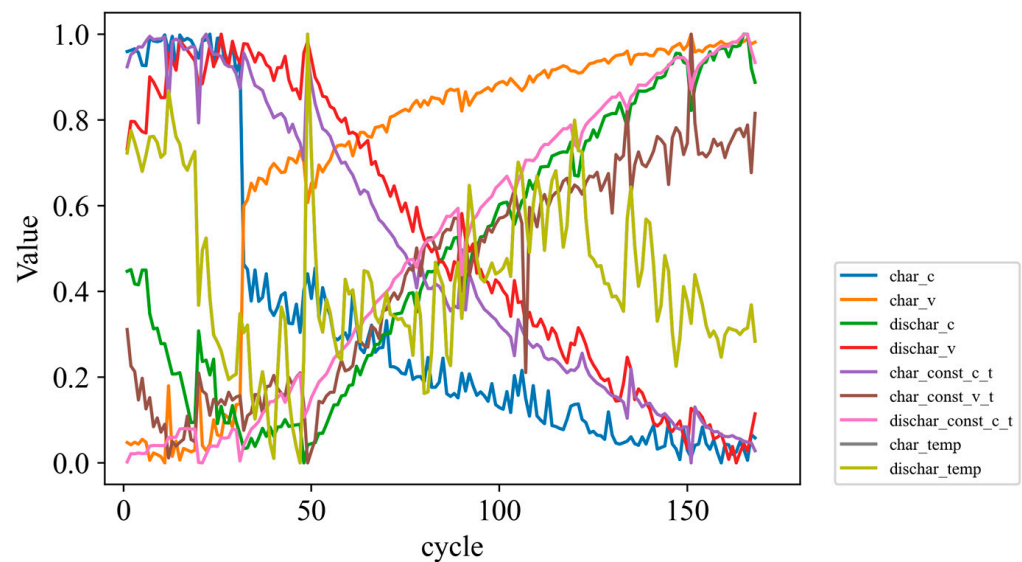

Figure 4. Battery degradation trends. The top solid line is the capacity trend curve and the bottom dashed line is the capacity change rate curve.

From the battery capacity change rate curve, the decay process of the battery presents a non-smooth characteristic. In general, the capacity declines at a low level, but along with the occurrence of the CR phenomenon, the capacity shows accelerated degradation after a sharp increase. This characteristic also restricts the further improvement of the accuracy of the capacity prediction method. On the other hand, it also proves the necessity of our proposed idea of considering dual degradation states of batteries and modeling them separately. To describe the degradation pattern of battery capacity, some charging and discharging feature parameters are extracted and used simultaneously for capacity degradation modeling. Features used in this paper and their corresponding codes are given in Table 1, while the normalized features are shown in Figure 5.

Table 1. Explanation of the features of lithium battery capacity.

Category	Code	Explanation
	'char_c'	Average charging current
	'char_v'	Average charging voltage
	'dischar_c'	Average discharging current
	'dischar_v'	Average voltage during discharging
Degradation features	'char_const_c_t'	Duration of constant current during charging
	'char_const_v_t'	Duration of constant voltage during charging
	'dischar_const_c_t'	Duration of constant current during discharging
	'char_temp'	Average temperature during charging
	'dischar_temp'	Average temperature during discharging
	're_char_dischar'	Interval time between charging and discharging (within the same cycle)
Relaxation features	're_char'	Interval time between charging (adjacent cycles)
	're_dischar'	Interval time between discharging (adjacent cycles)
	're_dischar_char'	Interval between discharging and next charging (adjacent cycles)

Figure 5. Features after standardization.

3.2. Capacity Curve Segmentation

Before capacity curve segmentation, the CR points are first derived from the curve. We refer to the experimental results in the literature [19] and combine them with the degradation characteristics of the NASA dataset, and use a change rate greater than +0.5% as the discriminating criterion. In addition, after the occurrence of the CR phenomenon, we select the point where the capacity degrades again to a level close to the pre-CR point as the end point of the recovery region. To summarize, we divide the capacity curve, taking battery #5 as an example, and the segmentation result is shown in Figure 6. From this figure, it can clearly be observed that the global degradation curve shows a smooth degradation trend after stripping the recovery regions, which is important for the prediction model in summarizing the degradation process of lithium-ion batteries. Meanwhile, it illustrates the necessity of the proposed separate prediction strategy for Li-ion batteries with dual degradation states.
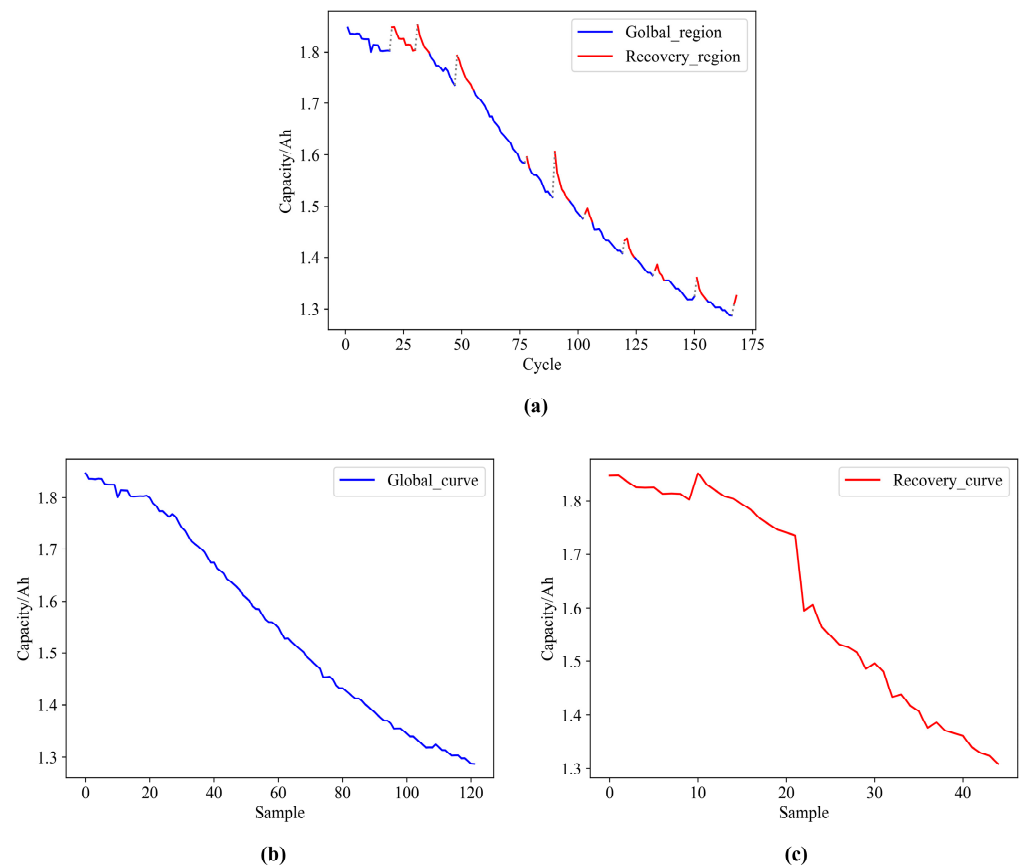
(a)

(b)

(c)

Figure 6. Segmentation results of #5 capacity curves. (**a**) Original curve and distribution of global degradation regions and recovery regions in it; (**b**) global degradation curve after stripping the recovery regions; (**c**) curve after splicing of each recovery region.

3.3. Causal Feature Selection

Use the nine feature sequences from $\{F_C(k)\}_{k=1}^{n'}$ and capacity $\{C(k)\}_{k=1}^{n'}$ sequences as a dataset to construct a DAG containing 10 nodes, which represents the number of battery cycles used for training, which is determined by the prediction start point. Then, the MB with the target variable (i.e., capacity value) is screened from the DAG by the STMB method, and the significance level of the independence test is set to 0.05. Table 2 provides the MB screening results for all four batteries.

Table 2. Screening results for features.

Battery	1	2	3	4	5	6	7	8	9
#5			√		√		√		
#6			√		√		√		
#7	√		√				√		
#18							√		

For all of these batteries, the constant current duration during the discharge phase is the common MB variable. Moreover, average current during charging and average current during discharging are MB variables for battery #7, and average current during discharging and constant current duration during charging are common MBs for batteries #5 and #6. To improve the broad adaptability of the causal features to different batteries, we select the concatenated set of MB variables of each battery as the set of dependent variables for battery capacity prediction.

3.4. Battery Capacity Prediction Results Based on the Proposed Method

In this section, the proposed hybrid prediction model is applied to predict the capacity of the four batteries from the NASA dataset. First, two evaluation metrics, i.e., mean absolute error (MAE) and root mean square error (RMSE), are introduced as evaluation metrics for the proposed method. The two metrics are calculated using the following equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{C}_i - C_i\right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{C}_i - C_i\right)^2} \tag{14}$$

where C_i is the true capacity and $\hat{C}_i$ is the predicted result given by the prediction model, and m is the number of predicted samples.

For batteries #5, #6, and #7, the prediction starting point is set to 100, and for battery #18, it is set to 80. The results of the capacity prediction are shown in Figure 7. It should be mentioned that, since the SVM-based CR point prediction method only determines whether the recovery phenomenon occurs, the default CR ratio is 1% of the initial capacity if the CR phenomenon is judged to have occurred when the prediction is performed. This ratio setting is derived from the experimental results in reference [19]. From Figure 7, it can be observed that the prediction results of both the global degradation region and the recovery region maintain a high degree of consistency. Although setting the CR ratio to 1% affects the capacity prediction accuracy at some points, it has little effect on the global prediction results. The prediction results of the four batteries are listed in Table 3, which shows that the proposed method can predict the capacity degradation trend of lithium batteries with high accuracy.
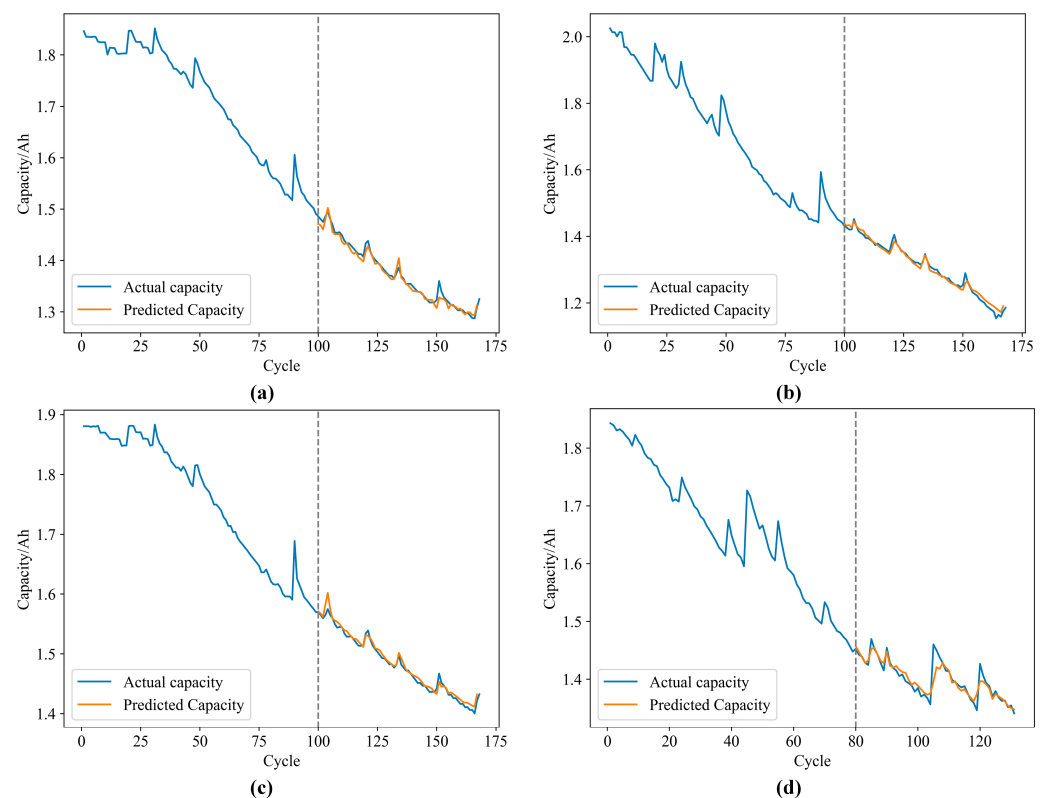

Figure 7. Capacity prediction results: (**a**) prediction results of #5; (**b**) prediction results of #6; (**c**) prediction results of #7; (**d**) prediction results of #18.

Table 3. Prediction results of the proposed method.

Battery	#5		#6		#7		#18	
Criteria	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)
Value	0.0079	0.0103	0.0101	0.0121	0.0053	0.0069	0.0082	0.0135

4. Discussion

In this section, we compare the prediction results of the proposed method with those of other cases, including a comparison with the prediction results based on the whole features (not screened), a comparison with the prediction results of a one-state-based model, and a comparison with the prediction results of other published methods. The development of the proposed method is also discussed.

4.1. Comparison with Initial Feature Sets

One of the innovations of the capacity prediction framework proposed in this paper is the introduction of causal features based on Bayesian networks with MB for lithium battery capacity prediction. In Section 3, the causal feature set for capacity is selected and excellent prediction results are obtained. In this section, we compare the capacity prediction results based on causal features with those based on the original feature set to demonstrate the positive significance of selecting the causal feature set to improve the capacity prediction results. A comparison of the prediction results is shown in Figure 8.
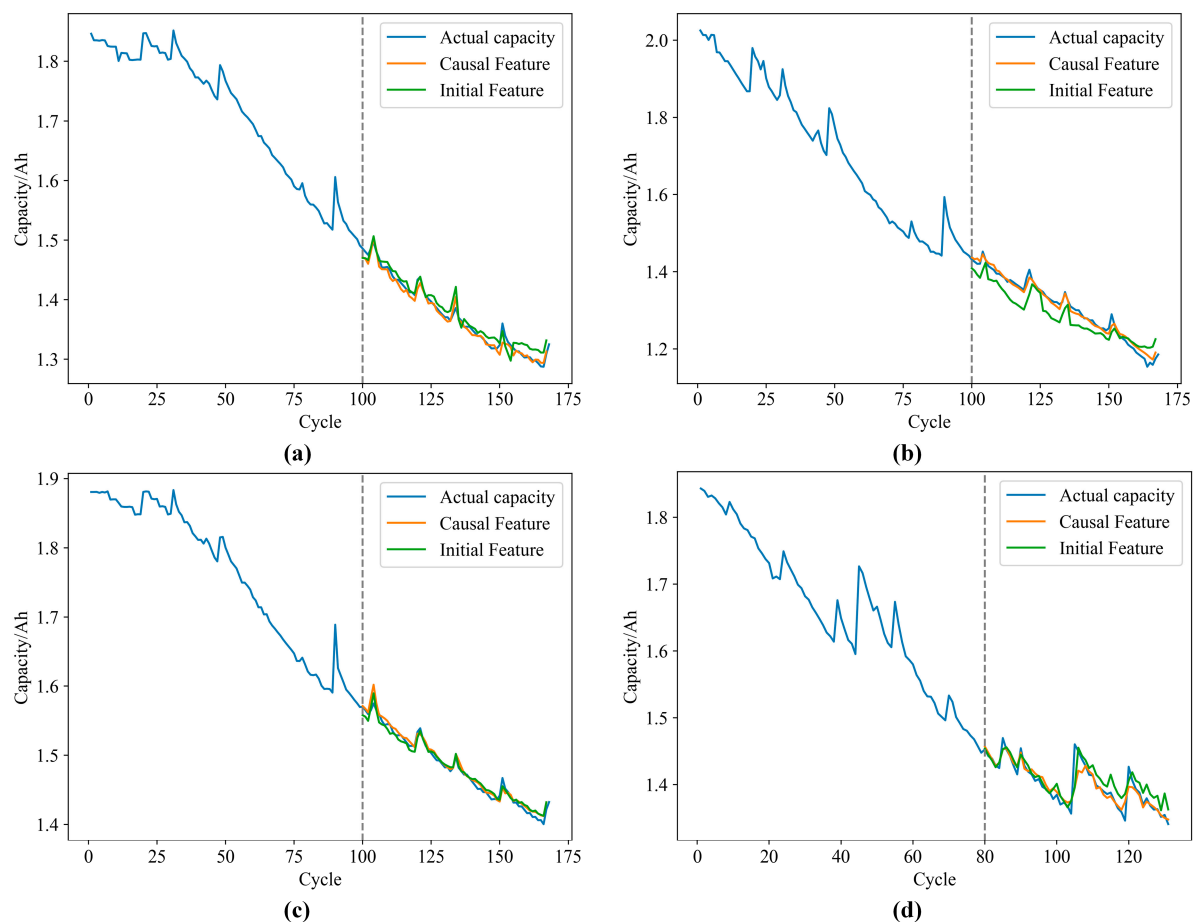

Figure 8. Comparison of the prediction results between causal feature sets and initial feature sets: (**a**) prediction results of #5; (**b**) prediction results of #6; (**c**) prediction results of #7; (**d**) prediction results of #18.

The results in Figure 9 show that for different batteries, the causal feature set for capacity prediction can obtain more accurate prediction results compared with the original feature set. To visualize the relative differences between different prediction results, we show the evaluation indexes of the prediction results, i.e., MAE and RMSE, in Figure 9. In the prediction results of batteries #6 and #18 with high-capacity volatility, the errors of the capacity prediction results based on causal features do not significantly increase. This also shows that capacity prediction based on causal features helps to circumvent the negative impact of battery specificity.
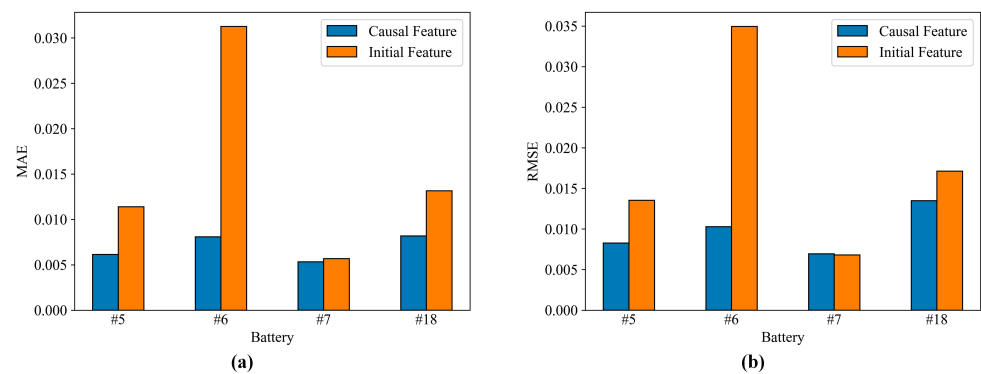

Figure 9. Comparison of prediction results between initial features and causal features: (**a**) prediction results assessed by MAE; (**b**) prediction results assessed by RMSE.

4.2. Comparison with the One-State-Based Model

Another innovation of the capacity prediction framework proposed in this paper is the provision of dual prediction branches to separately predict the two typical states of the capacity degradation process of Li-ion batteries, i.e., global degradation versus degradation of the CR region. To illustrate the necessity of this operation to improve capacity prediction. We select the LSTM model as a single-state prediction model to compare with the proposed method. The prediction results of both methods are shown in Figure 10.
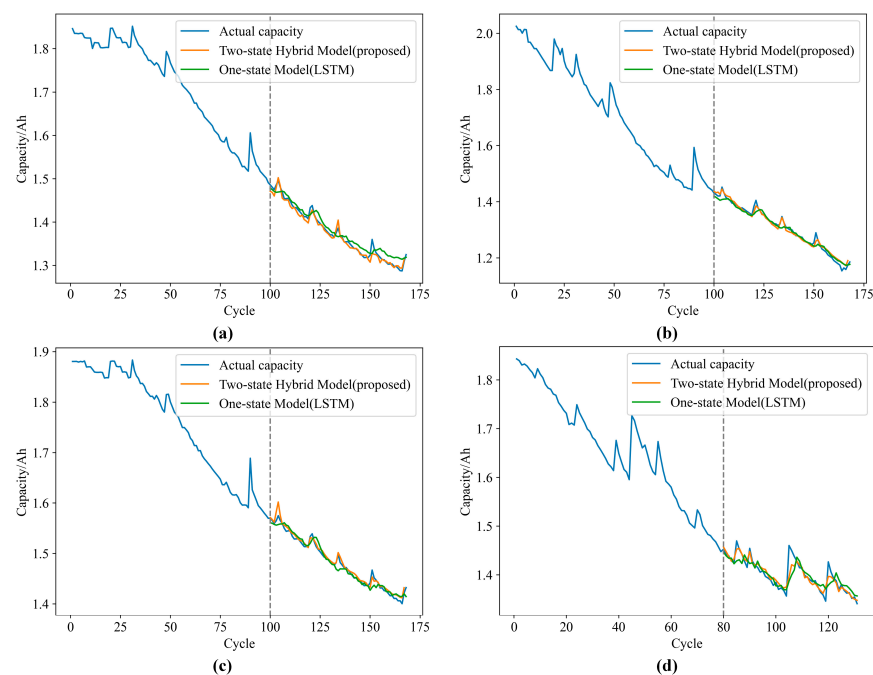

Figure 10. Prediction results comparison between the proposed two-state-based hybrid model and the one-state-based model: (**a**) prediction results of #5; (**b**) prediction results of #6; (**c**) prediction results of #7; (**d**) prediction results of #18.

It can be seen in Figure 11 that the one-state-based model struggles to make accurate predictions in the fluctuation region caused by the CR phenomenon, and there is a lag in the prediction of the CR point, which has a negative impact on the timely and accurate prediction of battery capacity. In contrast, the hybrid two-state-based prediction model proposed in this paper always maintains a good performance during the fluctuation caused by the CR phenomenon through its excellent identification mechanism and two relatively independent prediction branches. This also indicates the positive significance of the proposed method for battery capacity prediction in relation to the CR phenomenon.
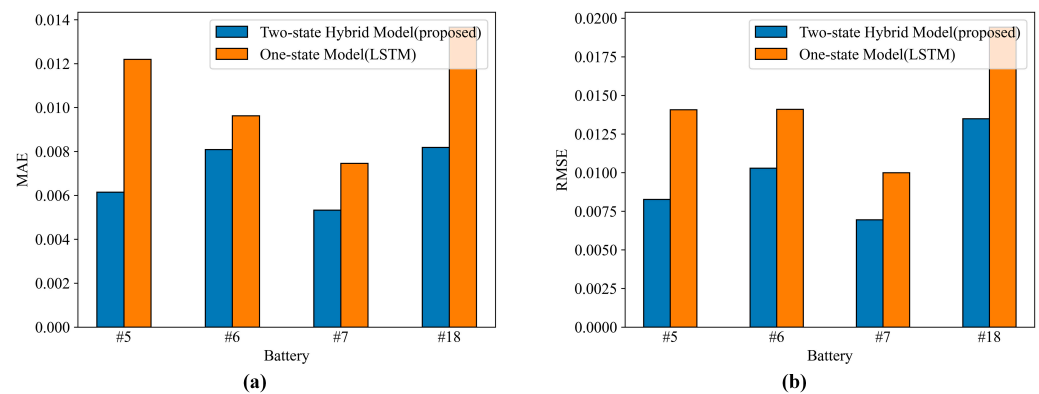

Figure 11. The prediction results comparison between the two-state-based hybrid model and the one-state-based model: (**a**) prediction results assessed by MAE; (**b**) prediction results assessed by RMSE.

4.3. Comparison with Different Methods

Table 4 presents the results of different methods for predicting the capacity of #5, #6, #7, and #18 cells. The capacity prediction results of the established methods are taken from references [27,35]. The proposed method has better prediction performance for all cell data. However, it is worth noting that the RMSE of the proposed method is slightly larger than that of GC-LSTM in the capacity prediction results for cell #18. Considering that the RMSE is more affected by outliers compared to the MAE, one possible reason is the fixed-value prediction method of the capacity at the recovery point adopted in this paper, i.e., after predicting the CR phenomenon. The predicted capacity of the next cycle is set to an increase of 1% of initial capacity (0.2 Ah in this case) from that of the previous cycle to simulate the CR phenomenon, which does not follow the several CR points with a high percentage of CR value in #18 well. This problem can be avoided with the improvement of the capacity prediction method at the recovery point.

Table 4. Comparison of prediction results between different methods.

Battery	#5		#6		#7		#18	
Criteria	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)	MAE/ (Ah)	RMSE/ (Ah)
IOWA [1]	0.0178	0.0233	0.0286	0.0362	0.0157	0.0198	0.0235	0.0300
GA-BP [1]	0.0116	0.0131	0.0436	0.0484	0.0117	0.0151	0.0436	0.0484
GC-LSTM [2]	0.0065	0.0093	0.0095	0.0110	**0.0045**	0.0112	0.0095	**0.0087**
Proposed	**0.0061**	**0.0083**	**0.0081**	**0.0103**	0.0053	**0.0069**	**0.0082**	0.0135

[1] Results are obtained from [35]. [2] Results are obtained from [27].

5. Conclusions

To address the issue of insufficient accuracy in current lithium battery capacity prediction methods when considering the CR phenomenon, this paper proposed an improved lithium-ion battery capacity prediction method that combines causal feature extraction and considers a two-state-based hybrid prediction model. The MB discovery algorithm based on Bayesian networks was first used to filter the MB variable set of battery capacity in the

original dataset as a causal variable set for predicting capacity, to improve the representation ability of the feature set in relation to capacity and reduce interference caused by other features. Then, based on our understanding of the CR phenomenon, we divided the life-long cyclic data of batteries into global degradation regions and recovery regions. On this basis, a two-state-based hybrid prediction model was proposed. This model framework provides a decoupling between the prediction process of the global degradation region and the recovery region, which not only reduces the interference of the CR phenomenon in the prediction of the global degradation process but also facilitates the improvement of the prediction accuracy of the recovery region. In addition, the SVM model based on relaxation time features is used for binary recognition of the CR phenomenon, providing information for model switching in hybrid prediction model frameworks. A case study based on the NASA lithium battery dataset demonstrates that (1) in the feature extraction stage, the filtered causal features are more helpful in improving the prediction accuracy of lithium battery capacity compared to the initial features; (2) the two-state-based hybrid prediction model successfully models two different degradation characteristics: the global degradation region and the recovery region. In the prediction phase, compared with the one-state-based model and published battery capacity prediction methods, the framework also provides higher prediction accuracy.

Author Contributions: Conceptualization, Y.C. and L.T.; methodology, Y.C.; software, Y.C.; validation, Y.C., S.L. and H.L.; formal analysis, L.T.; investigation, Y.C. and S.L.; resources, L.T. and L.W.; data curation, H.L.; writing—original draft preparation, Y.C.; writing—review and editing, L.T.; visualization, Y.C.; supervision, L.W.; project administration, L.T.; funding acquisition, L.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#battery (accessed on 9 October 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Che, Y.; Hu, X.; Lin, X.; Guo, J.; Teodorescu, R. Health prognostics for lithium-ion batteries, mechanisms, methods, and prospects. *Energy Environ. Sci.* **2023**, *16*, 338–371. [CrossRef]
2. Yu, J. State of health prediction of lithium-ion batteries, Multiscale logic regression and Gaussian process regression ensemble. *Reliab. Eng. Syst. Saf.* **2018**, *174*, 82–95. [CrossRef]
3. Zhao, J.; Ling, H.; Liu, J.; Wang, J.; Burke, A.F.; Lian, Y. Machine learning for predicting battery capacity for electric vehicles. *ETransportation* **2023**, *15*, 100214. [CrossRef]
4. Zhang, L.; Ji, T.; Yu, S.; Liu, G. Accurate Prediction Approach of SOH for Lithium-Ion Batteries Based on LSTM Method. *Batteries* **2023**, *9*, 177. [CrossRef]
5. Richardson, R.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for forecasting battery state of health. *J. Power Sources* **2017**, *357*, 209–219. [CrossRef]
6. Hasib, S.A.; Islam, S.; Chakrabortty, R.K.; Ryan, M.J.; Saha, D.K.; Ahamed, M.H.; Moyeen, S.I.; Badal, F.R. A comprehensive review of available battery datasets, RUL prediction approaches, and advanced battery management. *IEEE Access* **2021**, *9*, 86166–86193. [CrossRef]
7. Xu, X.; Chen, N. A state-space-based prognostics model for lithium-ion battery degradation. *Reliab. Eng. Syst. Saf.* **2017**, *159*, 47–57. [CrossRef]
8. Wei, J.; Dong, G.; Chen, Z. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5634–5643. [CrossRef]
9. Safari, M.; Morcrette, M.; Teyssot, A.; Delacourt, C. Multimodal physics-based aging model for life prediction of Li-ion batteries. *J. Electrochem. Soc.* **2008**, *156*, A145. [CrossRef]
10. Patil, M.A.; Tagade, P.; Hariharan, K.S.; Kolake, S.M.; Song, T.; Yeo, T.; Doo, S. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. *Appl. Energy* **2015**, *159*, 285–297. [CrossRef]
11. Zhou, Y.; Huang, M.; Chen, Y.; Tao, Y. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. *J. Power Sources* **2016**, *321*, 1–10. [CrossRef]

12. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [CrossRef]
13. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3478–3487. [CrossRef]
14. Olivares, B.E.; Munoz, M.A.C.; Orchard, M.E.; Silva, J.F. Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena. *IEEE Trans. Instrum. Meas.* **2012**, *62*, 364–376. [CrossRef]
15. Orchard, M.E.; Lacalle, M.S.; Olivares, B.E.; Silva, J.F.; Palma-Behnke, R.; Estévez, P.A.; Severino, B.; Calderon-Muñoz, W.; Cortés-Carmona, M. Information-theoretic measures and sequential monte carlo methods for detection of regeneration phenomena in the degradation of lithium-ion battery cells. *IEEE Trans. Reliab.* **2015**, *64*, 701–709. [CrossRef]
16. Tong, Z.; Miao, J.; Mao, J.; Wang, Z.; Lu, Y. Prediction of Li-ion battery capacity degradation considering polarization recovery with a hybrid ensemble learning model. *Energy Storage Mater.* **2022**, *50*, 533–542. [CrossRef]
17. Li, A.; Tian, H.; Li, K. Remaining useful life prediction of lithium-ion batteries using a spatial temporal network model based on capacity self-recovery effect. *J. Energy Storage* **2023**, *67*, 107557. [CrossRef]
18. Lyu, G.; Zhang, H.; Zhang, Y.J.; Miao, Q. An interpretable remaining useful life prediction scheme of lithium-ion battery considering capacity regeneration. *Microelectron. Reliab.* **2022**, *138*, 114625. [CrossRef]
19. Eddahech, A.; Briat, O.; Vinassa, J.M. Lithium-ion battery performance improvement based on capacity recovery exploitation. *Electrochim. Acta* **2013**, *114*, 750–757. [CrossRef]
20. Qin, T.; Zeng, S.; Guo, J.; Skaf, Z. A rest time-based prognostic framework for state of health estimation of lithium-ion batteries with regeneration phenomena. *Energies* **2016**, *9*, 896. [CrossRef]
21. Deng, L.; Shen, W.; Wang, H.; Wang, S. A rest-time-based prognostic model for remaining useful life prediction of lithium-ion battery. *Neural Comput. Appl.* **2021**, *33*, 2035–2046. [CrossRef]
22. Ma, Q.; Zheng, Y.; Yang, W.; Zhang, Y.; Zhang, H. Remaining useful life prediction of lithium battery based on capacity regeneration point detection. *Energy* **2021**, *234*, 121233. [CrossRef]
23. Pang, X.; Huang, R.; Wen, J.; Shi, Y.; Jia, J.; Zeng, J. A lithium-ion battery RUL prediction method considering the capacity regeneration phenomenon. *Energies* **2019**, *12*, 2247. [CrossRef]
24. Qiao, J.; Liu, X.; Chen, Z. Prediction of the remaining useful life of lithium-ion batteries based on empirical mode decomposition and deep neural networks. *IEEE Access* **2020**, *8*, 42760–42767. [CrossRef]
25. Couture, J.; Lin, X. Novel image-based rapid RUL prediction for li-ion batteries using a capsule network and transfer learning. *IEEE Trans. Transp. Electrif.* **2022**, *9*, 958–967. [CrossRef]
26. Lally, N.; Hartman, B. Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping. *Insur. Math. Econ.* **2018**, *82*, 124–140. [CrossRef]
27. Tian, Y.; He, J.; Peng, Z.; Guan, Y.; Wu, L. Lithium-Ion Battery Degradation and Capacity Prediction Model Considering Causal Feature. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 3630–3647. [CrossRef]
28. Pearl, J. *Probabilistic Reasoning in Intelligent Systems, Networks of Plausible Inference*; Morgan Kaufmann: Burlington, MA, USA, 1988.
29. Tsamardinos, I.; Aliferis, C.F. Towards Principled Feature Selection, Relevancy, Filters and Wrappers. In Proceedings of the International Workshop on Artificial Intelligence and Statistics, Key West, FL, USA, 3–6 January 2003; pp. 300–307.
30. Yu, K.; Guo, X.; Liu, L.; Li, J.; Wang, H.; Ling, Z.; Wu, X. Causality-based feature selection, Methods and evaluations. *ACM Comput. Surv.* **2020**, *53*, 1–36. [CrossRef]
31. Gao, T.; Ji, Q. Efficient Markov blanket discovery and its application. *IEEE Trans. Cybern.* **2016**, *47*, 1169–1179. [CrossRef]
32. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
33. Seeger, M. Gaussian processes for machine learning. *Int. J. Neural Syst.* **2004**, *14*, 69–106. [CrossRef]
34. Saha, B.; Goebel, K. *Battery Data Set*; National Aeronautics and Space Administration (NASA), Ames Prognostics Data Repository: Moffett Field, CA, USA, 2007.
35. Sun, T.; Wang, S.; Jiang, S.; Xu, B.; Han, X.; Lai, X.; Zheng, Y. A cloud-edge collaborative strategy for capacity prognostic of lithium-ion batteries based on dynamic weight allocation and machine learning. *Energy* **2022**, *239*, 122185. [CrossRef]