Reliability Engineering and System Safety 230 (2023) 108947

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Prediction of state of health and remaining useful life of lithium-ion battery using graph convolutional network with dual attention mechanisms

Yupeng Wei [a,*], Dazhong Wu [b]

[a] Department of Industrial and Systems Engineering, San Jose State University, San Jose, 95192, CA, USA
[b] Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, 32816, FL, USA

ARTICLE INFO

Keywords:
Lithium-ion battery
State-of-health
Remaining useful life
Graph convolutional network
Dual attention mechanism

ABSTRACT

Prediction of state-of-health and remaining useful life is crucial to the safety of lithium-ion batteries. Existing state-of-health and remaining useful life prediction methods are not effective in revealing the correlation among features. Establishing the correlation can help identify features with high similarities and aggregate them to improve the accuracy of predictive models. Moreover, existing methods, such as recurrent neural networks and long short-term memory, have limitations in state-of-health and remaining useful life predictions as they are not capable of using the most relevant part of time-series data to make predictions. To address these issues, a two-stage optimization model is introduced to construct an undirected graph with optimal graph entropy. Based on the graph, the graph convolutional networks with different attention mechanisms are developed to predict the state-of-health and remaining useful life of a battery, where the attention mechanisms enable the neural network to use the most relevant part of time series data to make predictions. Experimental results have shown that the proposed method can accurately predict the state-of-health and remaining useful life with a minimum root-mean-squared-error of 0.0104 and 5.80, respectively. The proposed method also outperforms existing data-driven methods, such as gradient-boosting decision trees, long short-term memory, and Gaussian process.

1. Introduction

Lithium-ion batteries have been widely used in many industries such as unmanned aerial vehicles, electric vehicles, and portable electronics [1,2]. The performance of a lithium-ion battery will deteriorate with repeated charge and discharge cycles, which is also known as battery aging [3]. Battery aging results in severe economic losses and even catastrophic disasters such as fire hazards or explosions because of flammable electrolytes. For example, Sony recalled 9.6 million laptops due to abnormal battery aging issues in 2007, which led to around $430 million in economic losses [4]. In 2018, a battery fire happened in a cement plant in South Korea, which resulted in over $3 million in damage [5]. Therefore, it is important to monitor lithium-ion battery health conditions.

State-of-health (SOH) and remaining useful life (RUL) are commonly used to measure the health condition of lithium-ion batteries. SOH refers to the ratio of the maximum battery capacity to its rated capacity [6]. RUL refers to the number of remaining charge and discharge cycles that a battery reaches its minimum acceptable capacity. Numerous methods have been introduced to predict the SOH and RUL of lithium-ion batteries, and these methods can be classified into two groups: model-based and data-driven methods. The model-based methods are built upon equivalent circuit models [7] and electrochemical models [8] to make predictions. In comparison with model-based methods, data-driven methods can be classified into three categories: filter-based [9], machine learning [10], and deep learning methods [11]. While existing data-driven methods are effective in predicting SOH and RUL, these methods are not capable of identifying the correlation among features [12,13]. This correlation can be used to identify features with high similarities as well as aggregate similar features to improve the robustness of a predictive model. Moreover, most of the existing deep learning methods with recurrent characteristics, such as long short term memory (LSTM) [14] and recurrent neural network (RNN) [15] based methods, have limitations in SOH and RUL predictions as these methods are not capable of using the most relevant part of time-series data [16]. To address these issues, we first construct an undirected graph and use the topological structure of the graph to better reveal the feature correlation. To construct such an undirected graph, a two-stage optimization model is introduced to obtain a graph with optimal entropy by maximizing graph entropy and minimizing

* Corresponding author.
E-mail address: yupeng.wei@sjsu.edu (Y. Wei).

https://doi.org/10.1016/j.ress.2022.108947
Received 16 June 2022; Received in revised form 25 October 2022; Accepted 28 October 2022
Available online 3 November 2022

graph density simultaneously, where maximizing the graph entropy can maximize the information communicated rate among different features and minimizing the graph density can reduce the redundant edges. Next, the spectral graph convolution operation is introduced to develop a graph convolutional network (GCN) to aggregate features with high similarities based on the constructed undirected graph. In addition, in order to use the most important part of a time series to make predictions, an LSTM network with dual attention mechanisms is introduced to predict the SOH and RUL of a battery, where the local attention mechanism and global attention mechanism are incorporated. These attention mechanisms can generate attention vectors with weights representing the importance of the time series data at different time points, and using these attention vectors can allow an LSTM network to use the most important part of a time series to make predictions. The contributions of this work can be summarized as follows:

- A two-stage optimization method is introduced to construct an undirected graph with optimal entropy so that the correlation among different features can be identified by using the topological structure of the constructed graph.
- Two types of graph convolutional networks with different attention mechanisms are developed to respectively predict the SOH and RUL of a lithium-ion battery, where these attention mechanisms incorporate the local attention mechanism and global attention mechanism.

The remainder of this paper is organized as follows. Section 2 reviews the data-driven methods for SOH and RUL predictions. Section 3 introduces the optimal entropy-enabled graph convolution LSTM network with dual attention mechanisms for SOH and RUL predictions. Section 4 demonstrates the effectiveness of the proposed method using a case study and compares the proposed method with other predictive models of SOH and RUL reported in the literature. Section 5 discusses the impact of the proposed optimal entropy graph and the availability of data from charge cycles. Section 6 concludes this research work and directs future work.

2. Data-driven methods for SOH and RUL predictions

The data-driven methods for SOH and RUL predictions fall into three categories: (1) filter-based methods, (2) machine learning methods, and (3) deep learning methods. The filter-based methods mainly include extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filter (PF). For example, Zhang et al. [17] proposed an adaptive unscented Kalman filter method to estimate the SOH and RUL of a lithium-ion battery, where a three-dimensional open circuit voltage model was used to reduce prediction errors. The effectiveness of the proposed Kalman filter was evaluated by a lithium-ion battery cell operated under different operating conditions. The results have shown that the proposed methodology can achieve a less than 2% prediction error. Yan et al. [18] introduced an extended Kalman filter based on Lebesgue sampling to predict the SOH and RUL of a battery. The experimental results have shown that the proposed Kalman filter can not only reduce computational complexities, but also enable a good performance in SOH prediction. Zhang et al. [19] presented a particle filter method based on optimal sampling methodology to enable a better prediction performance. A Kalman filter was used to generate a distribution for the proposed particle filter, and the optimal sampling method was used to deal with the inadequacy of particle variety. Zhang et al. [20] presented a novel method for RUL prediction of lithium-ion batteries based on the Kalman filter. The proposed Kalman filter approach considered the parametric uncertainty by using the Wiener process. Experiment results have shown that the proposed Kalman filter approach outperforms the traditional data-driven methods.

In comparison with the filter-based methods, machine learning methods have the potential to achieve better prediction performance as they are more effective in dealing with complex data. Therefore, machine learning methods such as support vector machine, ensemble learning, and Gaussian process have been introduced for battery SOH and RUL predictions [21,22]. For example, Li et al. [23] proposed a battery SOH predictive model based on support vector regression. The battery health attributes were extracted and fed into the support vector regression method for SOH predictions. The experimental results have shown that the proposed method can improve the prediction accuracy by 30% in terms of mean absolute errors and root mean squared errors. Gou et al. [24] presented an ensemble learning method to predict the RUL of the lithium-ion battery, where multiple random vector functional link networks were combined to improve the prediction performance. The numerical results have demonstrated that the proposed ensemble learning methodology can predict the RUL with high precision. Richardson et al. [25] proposed a Gaussian process regression method to predict the SOH of a battery. An explicit mean function was used to exploit the prior health information of lithium-ion batteries. The numerical results have demonstrated that the proposed Gaussian process can accurately predict the SOH in both short and long runs.

Deep learning methods with recurrent characteristics such as RNN and LSTM have been increasingly used for SOH and RUL predictions [26,27]. For example, Tang and Yuan [28] used a bidirectional gated recurrent neural network to predict the SOH and RUL of the battery. The empirical model decomposition was used to decompose the sensor measurements into low and high-frequency signals, the low and high signals were respectively fed into a deep neural network and a bidirectional gated recurrent neural network for predictions. The numerical results have shown that the proposed method can make accurate predictions under both constant and random charging and discharging modes. Park et al. [29] introduced a LSTM-based battery RUL prediction method to reduce the risk of battery failures. The proposed method employed a multi-channel architecture for more flexibility and generalization while processing different types of variables. The experimental results have shown that the proposed method can significantly improve the prediction performance by 63.7% in terms of mean absolute percentage error. Qu et al. [15] combined the LSTM network with the particle swarm optimization method for SOH and RUL prediction. The empirical mode decomposition method was implemented to eliminate the noise effects of sensor measurements. The experimental results have shown that the proposed method can achieve a higher prediction accuracy in comparison with methods reported in the literature. Wang et al. [30] introduced an improved feedforward-LSTM network to predict the state-of-charge and reliability of a battery cell by considering different types of variations. An optimal sliding window and a steady-state screening method were proposed to reduce the state-of-charge prediction redundancy. The experimental results have shown that the proposed method can predict the state of charge with a prediction error of 0.0353. However, these deep learning methods with recurrent features have difficulties in predicting the SOH and RUL with long time-series data as they are highly dependent on the final vector output and do not allow one to use the entire time series while predicting. Moreover, the existing data-driven methods are not effective in identifying the correlation of features.

In summary, numerous studies have been reported on SOH and RUL predictions. However, most of these methods are not able to reveal the correlation of extracted features, and the existing deep learning methods are not able to use the most relevant part of the series data to make predictions. To address these issues, we introduce an optimal entropy-enabled graph convolution LSTM network with dual attention mechanisms to consider the correlation of the temporal features extracted from sensor measurements and use the most important parts of the time series to make predictions.

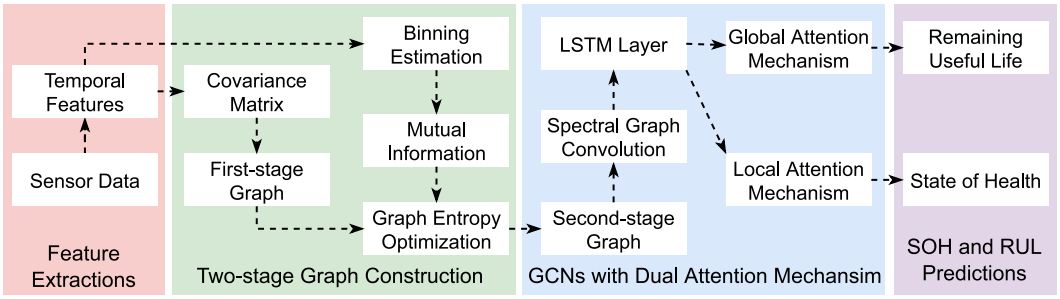

Fig. 1. Flow diagram of the proposed optimal entropy-enabled graph convolution network with dual attention mechanisms.

3. Optimal entropy graph convolutional LSTM network with dual attention mechanisms

Fig. 1 shows a flow diagram of the proposed optimal entropy graph convolutional LSTM network with dual attention mechanisms. First, an undirected graph with optimal entropy is constructed with a two-stage method. Second, the constructed graph is fed into the graph convolutional networks (GCNs) for mining the correlations among different features by using the topological structure of the constructed graph. Third, the outputs of the GCNs are fed into two LSTM networks with different attention mechanisms to predict the SOH and RUL of a battery respectively. More specifically, temporal features are extracted from sensor measurements, and then the extracted features are used to derive a covariance matrix and mutual information of features; a masking function is applied to the covariance matrix to construct a first-stage graph; the mutual information of temporal features and the first-stage graph are fed into the proposed graph entropy optimization model to construct the second-stage graph by maximizing graph entropy and minimizing graph density simultaneously. The topological structure of the constructed second-stage graph is used to perform spectral graph convolution operations to better reveal the feature correlations; and the outputs of the GCNs are fed into an LSTM network with a local attention mechanism and an LSTM network with a global attention mechanism to predict the SOH and RUL of a battery, respectively. More details of these steps are provided in Sections 3.1 and 3.2.

3.1. Two-stage graph construction

The most commonly used sensor measurements for battery health management are voltage, current, and temperature. In this work, we extract multiple temporal features from these three sensor measurements in both charge and discharge cycles as these features have been demonstrated as effective in SOH and RUL predictions [31]. The extracted temporal features include time to the maximum voltage in charge cycles, time charged under a constant voltage mode, time charged under a constant current mode, time to the minimum current in charge cycles, time to the maximum temperature in charge cycles, time to the minimum voltage in discharge cycles, time discharged under a constant current mode, and time to the maximum temperature in discharge cycles.

The optimization model of establishing a graph with optimal entropy is a nonlinear integer optimization problem. Directly solving such an optimization problem may result in establishing a local optimization graph with very less edges. To address this issue, we propose a two-stage method in which a first-stage graph is pre-constructed and the first-stage graph is used as the initial graph to construct a second-stage graph with optimal graph entropy. To construct a first-stage graph, a covariance matrix $C \in \mathbb{R}^{m\times m}$ is constructed for these extracted features, the element $c_{k,k'}$ of the matrix can be written as Eq. (1),

$$c_{k,k'} = \frac{1}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} \sum_{j=1}^{t_i} (x_{i,j,k} - \bar{x}_{..k})(x_{i,j,k'} - \bar{x}_{..k'}) \quad i = 1,\dots,n;\ j = 1,\dots,t_i;\ k = 1,\dots,m. \tag{1}$$

where $x_{i,j,k}$ refers to the kth extracted feature for battery unit i in the charge and discharge cycle j; n is the total number of battery units; m refers to the total number extracted features; t_i is the total number of charge and discharge cycles for battery unit i. $\bar{x}_{..k}$ represents the expectation of the feature k, which can be written as $\bar{x}_{..k} = \sum_{i=1}^{n}\sum_{j=1}^{t_i} x_{i,j,k} / \sum_{i=1}^{n} t_i$. A masking function $M(.)$ is applied to the covariance matrix to derive an adjacency matrix A_1 of the first-stage graph G_1, and this function can be written as $a^1_{k,k'} = M(c_{k,k'}) = 1$ if $c_{k,k'} \geq \varepsilon$ and $a^1_{k,k'} = M(c_{k,k'}) = 0$ if $c_{k,k'} < \varepsilon$, where $a^1_{k,k'}$ is the element of the adjacency matrix A_1, $a^1_{k,k'} = 1$ refers to an edge between the features k and k', $a^1_{k,k'} = 0$ represents that there is not an edge between features k and k', and ε is the threshold to determine the edges in the first-stage graph. In this work, we set ε equals to zero for simplification, meaning that there is an edge if two features are positively correlated.

Next, the mutual information of the temporal features is derived, which aims at measuring the mutual dependency among these features [32]. The mutual information between the features k and k' can be written as Eq. (2), which can be explained as the KL-divergence between the joint distribution and the multiplication of two marginal distributions.

$$w_{k,k'} = D_{KL}(P_{(k,k')} \parallel P_k \otimes P_{k'}) \tag{2}$$

The binning method is used to estimate the joint and marginal distributions, where the extracted features are grouped into multiple bins for probability estimations [33]. By using the binning method, Eq. (2) can be written as Eq. (3) which is in a format of discrete distributions,

$$w_{k,k'} = \sum_{b_k=1}^{B_k} \sum_{b_{k'}=1}^{B'_k} p_{b_k,b_{k'}}(k,k') \log[p_{b_k,b_{k'}} / p_{b_k}(k) p_{b_{k'}}(k')] \tag{3}$$

where B_k is the number of bins used to estimate the probability distribution of the feature k; and $B_{k'}$ is the number of bins used to estimate the probability distribution of the feature k'. The derived mutual information of the temporal features is used as the edge weight of the graph. The mutual information and the first-stage graph are used to derive the second-stage graph, where an optimization model is introduced to obtain a graph enables a balance between the graph entropy and graph density. Graph entropy is a measurement of the information rate of a graph [34]. Maximizing the graph entropy can maximize the information communicated rate among different features, meaning that every single feature can utilize as well as aggregate the data of all other features while predicting so that the prediction performance can be largely improved. However, solely maximizing the graph entropy may lead to obtaining a complete graph with the highest graph density [34]. Such a complete graph with the highest graph density may incorporate some redundant edges resulting in excessive and redundant feature aggregations while predicting. Therefore, the constructed graph should simultaneously maximize the graph entropy to increase the information communicated rate and minimize the graph density to reduce the redundant edges.

The entropy of a weighted graph $G = (V, E, \mathbf{w})$ has been introduced in some studies [35], which can be written as Eq. (4),

$$I(G,\mathbf{w}) = -\sum_{uv\in E} d_{uv} \log(d_{uv}),\ d_{uv} = w_{uv} / \sum_{uv\in E} w_{uv} \tag{4}$$

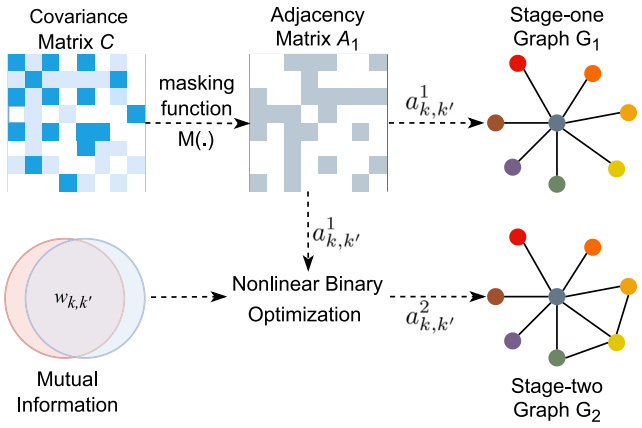

Fig. 2. The framework of the proposed two-stage graph construction method.

where V, E, w refer to the vertex set, edge set, and edge weights; u and v are vertex points in the vertex set; w_{uv} is the weight between two vertex points u and v. Similarly, we use the definition of graph entropy for a weighted graph to formulate an optimization model in order to construct the second-stage graph with maximum graph entropy and minimum graph density. The formulated optimization problem is written as Eq. (5),

$$\max \; I(G_2, \mathbf{w}) - \alpha \|\mathbf{w}^T \mathbf{a}\|_1 \tag{5}$$

where $I(G_2, \mathbf{w})$ is the graph entropy for the second-stage graph G_2; $\mathbf{w}$ is the weight vector stores the mutual information of different features; α is the penalty parameter determining the level of the sparse connections of the second-stage graph; and $\mathbf{a}$ is a vector that stores binary variables $a^2_{k,k'}, \forall k$. More specifically, Eq. (5) can also be rewritten as Eq. (6),

$$\begin{aligned}
&\max \; \sum_{k=1}^{m}\sum_{k'=1}^{m} d_{k,k'} \log(d_{k,k'}) - \alpha \sum_{k=1}^{m}\sum_{k'=1}^{m} |w_{k,k'} \cdot a^2_{k,k'}| \\
&d_{k,k'} = w_{k,k'} \cdot a^2_{k,k'} / \sum_{k=1}^{m}\sum_{k'=1}^{m} w_{k,k'} \cdot a^2_{k,k'}; \\
&a^2_{k,k'} = a^2_{k',k};\; a^2_{k,k'} = 0 \text{ or } 1; \\
&a^2_{k,k'} = 0, \text{ if } k = k';\; a^2_{k,k'} = a^1_{k,k'}, \text{if } a^1_{k,k'} = 1; \\
&k = 1, \ldots, m;\; k' = 1, \ldots, m.
\end{aligned} \tag{6}$$

where $w_{k,k'}$ is the mutual information between the extracted features k and k'; $a^1_{k,k'}$ is element of the adjacency matrix A_1 and denotes the edge status of the first-stage graph; $a^1_{k,k'} = 1$ represents that there is an edge between the features k and k' in the first-stage graph; $a^1_{k,k'} = 0$ represents that there is not an edge between the features k and k' in the first-stage graph; $a^2_{k,k'}$ is a binary variable that denotes the status of edges of the second-stage graph; $a^2_{k,k'} = 1$ represents that there is an edge between the features k and k' in the second-stage graph, $a^2_{k,k'} = 0$ represents that there is not an edge between the features k and k' in the second-stage graph.

The formulated optimization problem is a nonlinear integer optimization problem with nonlinear constraints, many algorithms can be used to solve this problem such as the genetic algorithm, particle swarm optimization, and spatial branch and bound [36]. In this work, we use the most often used algorithm, the genetic algorithm, to derive a local optimal solution for simplification. After solving the proposed optimization problem, the value of the variable $a^2_{k,k'}$ for all k and k' can be used to derive the second-stage graph. Fig. 2 shows the framework of the proposed two-stage graph construction method. In summary, a covariance matrix C and mutual information $w_{k,k'}$ are derived from the extracted temporal features; a masking function is applied to the covariance matrix C to derive the adjacency matrix A_1; and the element $a^1_{k,k'}$ of the adjacency matrix is used to construct the stage-one graph G_1. Moreover, both the mutual information $w_{k,k'}$ for all features and the element of the adjacency matrix are fed into the proposed nonlinear binary optimization problem to derive the optimal value of variables $a^2_{k,k'}, \forall k$; and these variables are used to construct the stage-two graph G_2.

3.2. Graph convolutional LSTM network with dual attention mechanisms

Fig. 3 illustrates the framework of the proposed graph convolutional LSTM network with dual attention mechanisms. The extracted feature matrices X_i for all battery unit i are added in the stage-two graph G_2, where each node of the graph G_2 represents each feature. The spectral graph convolution operations are performed to utilize the topological structure of the stage-two graph, and the outputs of the GCNs are fed into an LSTM network to further extract temporal correlations of features. The output of the LSTM network is connected with a local attention mechanism and a global attention mechanism to predict the SOH and RUL, respectively. More detailed descriptions of the proposed graph convolutional LSTM network with dual attention mechanisms can be founded in the remaining paragraphs of Section 3.2.

To use the topological structure of the stage-two graph, the convolutional operation on the graph G_2 can be defined as Eq. (7), where $\mathscr{F}$ represents the graph Fourier transform and $\mathscr{F}^{-1}$ is the inverse graph Fourier transform; f refers to the graph filter provided by the stage-two graph G_2. Moreover, $X_i \in \mathbb{R}^{m \times t_i}$ is the feature matrix for battery unit i, where m is the number of features and t_i is the number of charge and discharge cycles for battery unit i.

$$f(\cdot_{G_2}) X_i = \mathscr{F}^{-1}(\mathscr{F}(f) \odot \mathscr{F}(X_i)) \tag{7}$$

The graph Fourier transform for graph filter f and feature matrix X_i can be written as $Q^T f$ and $Q^T X_i$ respectively, where Q is the eigenvector of the Laplacian matrix. Then, Eq. (7) can be written as Eq. (8), where Λ is the diagonal matrix that stores the eigenvalue and $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$; ϕ refers to a collection of trainable parameters in the graph filter.

$$f(\cdot_{G_2}) X_i = Q(Q^T f \odot Q^T X_i) = Q f_\phi(\Lambda) Q^T X_i \tag{8}$$

For the constructed stage-two graph G_2, the normalized Laplacian matrix $\mathcal{L}$ is defined as Eq. (9), where $I \in \mathbb{R}^{m \times m}$ refers to an identity matrix; A_2 is the adjacency matrix for the second-stage graph G_2; and D represents the degree matrix and the entry of the degree matrix can be written as $D_{k,k'} = \sum_{k'=1}^{m} a^2_{k,k'}$.

$$\mathcal{L} = I - D^{-1/2} A_2 D^{1/2} \tag{9}$$

The eigendecomposition of the Laplacian matrix can be written as $\mathcal{L} = Q \Lambda Q^{-1}$. The eigendecomposition of the Laplacian matrix can also be written as $\mathcal{L} = Q \Lambda Q^T$ as the Laplacian matrix is a real symmetric matrix. Therefore, Eq. (7) can also be written as Eq. (10).

$$f(\cdot_{G_2}) X_i = f_\phi(\mathcal{L}) X_i \tag{10}$$

Solving the Eq. (10) can be computationally expensive due to the eigendecomposition process. It has been suggested that a truncated Chebyshev polynomials can be used to approximate the convolutional operation on a graph [37], and the Eq. (10) can be rewritten as Eq. (11). In this equation, $C_k(\tilde{\mathcal{L}})$ refers to the Chebyshev polynomial in the order of p, and $\tilde{\mathcal{L}}$ is the scaled Laplacian matrix.

$$f_\phi(\mathcal{L}) X_i = \sum_{p=0}^{P-1} \phi_k C_k(\tilde{\mathcal{L}}) X_i \tag{11}$$

To further reduce the computational complexity, the first order Chebyshev polynomials are only taken into account [38]. Then, Eq. (11) can be written as Eq. (12).

$$f_\phi(\mathcal{L}) X_i = \phi_0 C_0(\tilde{\mathcal{L}}) X_i + \phi_1 C_1(\tilde{\mathcal{L}}) X_i = (\phi(D^{-1/2} A_2 D^{1/2} + I)) X_i \tag{12}$$

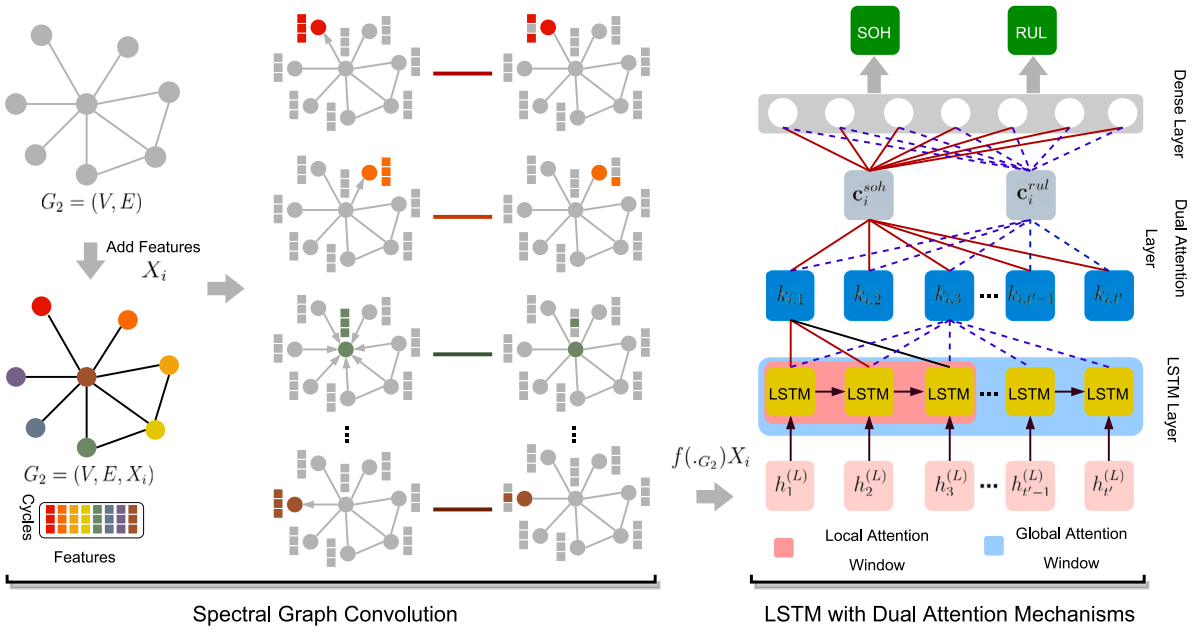

Fig. 3. The framework of the proposed graph convolutional networks with dual attention mechanisms.

To avoid the gradient vanishing or explosion problem, the adjacency matrix $\tilde{A}_2 \in \mathbb{R}^{m\times m}$ with a self connection is introduced, which can be written as Eq. (13), where the parameter β denotes the level of self connection for extracted features and $I \in \mathbb{R}^{m\times m}$ is an entity matrix.

$$\tilde{A}_2 = A_2 + \beta \cdot I \tag{13}$$

By aggregating all equations, we can obtain $f(\cdot_{G_2})X_i = \phi\tilde{D}^{-1/2}\tilde{A}_2\tilde{D}^{1/2}X_i$, where $\tilde{D}_{k,k'} = \sum_{k'=1}^{m}\tilde{a}^2_{k,k'}$. Next, we write the ϕ in the matrix format that is $\boldsymbol{\Phi} \in \mathbb{R}^{t_i\times t'_i}$, and the spectral convolutional operational can be written as Eq. (14), where $\hat{A}_2$ equals $\tilde{D}^{-1/2}\tilde{A}_2\tilde{D}^{1/2}$.

$$f(\cdot_{G_2})X_i = \hat{A}_2 X_i \boldsymbol{\Phi} \tag{14}$$

The graph convolutional layer is obtained based upon the graph convolution operation in Eq. (14). In this work, we add bias weight into the graph convolutional layer to increase the robustness and performance, and such a layer can be written as Eq. (15), where σ_a refers to the activation function, $\boldsymbol{\Phi}^{(l-1)}$ is the weight matrix at layer $l-1$, $b^{(l-1)}$ is the bias weight matrix at layer $l-1$, and $h^{(l-1)}$ refers to the output of the GCNs at the layer $l-1$.

$$h^{(l)} = \sigma_a(\hat{A}_2 h^{(l-1)}\boldsymbol{\Phi}^{(l-1)} + b^{(l-1)}) \tag{15}$$

Next, the output of the GCNs $h^{(L)}$ at the last (Lth) GCN layer is fed into an LSTM network with dual attention mechanisms for further mining temporal correlations of features. The general idea of attention-based models is to simulate the data retrieval process in a data management system [39]. In this work, we introduce an LSTM network with a local attention mechanism and an LSTM network with a global attention mechanism to predict the SOH and RUL of a battery, respectively. The SOH predictions are highly dependent on the most recent sensor measurements, thus, a local attention mechanism is used to calculate the local attention value $k^{(soh)}_{i,j}$ for SOH predictions. The formula used to calculate local attention values can be written as Eq. (16),

$$k^{(soh)}_{i,j} = exp(sc^{(soh)}_{i,j}) / \sum_{q=j-s}^{j+s} sc^{(soh)}_{i,q} \tag{16}$$

where s refers to the window size of the local attention mechanism; and $sc^{(soh)}_{i,j}$ is the alignment score between the final output $\mathbf{v}^{(soh)}_{i,t'_i}$ of the LSTM network and the hidden state $\mathbf{v}^{(soh)}_{i,j}$ at time j for SOH predictions.

We use the general alignment function [40] to calculate the alignment score, which can be written as Eq. (17),

$$sc^{(soh)}_{i,j} = \mathbf{v}^{(soh)^T}_{i,t'_i} W^{(soh)}_k \mathbf{v}^{(soh)}_{i,j} \tag{17}$$

where $W^{(soh)}_k \in \mathbb{R}^{z_{soh}\times z_{soh}}$ is the trainable weight matrix for SOH predictions; and z_{soh} refers to the number of hidden nodes in the LSTM network for SOH predictions. Next, a context vector $\mathbf{c}^{(soh)}_i$ is created and fed into a single layer neural network with the linear activation function to predict the SOH $y^{(soh)}_i$ of the battery unit i. Eq. (18) shows the creation of the context vector.

$$\mathbf{c}^{(soh)}_i = \sum_{j=1}^{t_i} k^{(soh)}_{i,j}\mathbf{v}^{(soh)}_{i,j} \tag{18}$$

The RUL predictions are dependent on all historical sensor measurements as historical measurements include usage pattern information. To consider all historical measurements, a global attention mechanism is used to calculate the global attention value $k^{(rul)}_{i,j}$ for RUL predictions. The formula is used to calculate global attention values can be written as Eq. (19).

$$k^{(rul)}_{i,j} = exp(sc^{(rul)}_{i,j}) / \sum_{q=1}^{t'_i} sc^{(rul)}_{i,q} \tag{19}$$

Similarly, we use general alignment function to calculate the alignment score, which can be written as Eq. (20),

$$sc^{(rul)}_{i,j} = \mathbf{v}^{(rul)^T}_{i,t'_i} W^{(rul)}_k \mathbf{v}^{(rul)}_{i,j} \tag{20}$$

where $W^{(rul)}_k \in \mathbb{R}^{z_{rul}\times z_{rul}}$ is the trainable weight matrix, and z_{rul} refers to the number of hidden nodes in another LSTM network for RUL prediction. Next, a context vector $\mathbf{c}^{(rul)}_i$ is created and fed into a single layer neural network to predict the RUL $y^{(rul)}_i$ of the battery unit i. Eq. (21) shows the creation of the context vector.

$$\mathbf{c}^{(rul)}_i = \sum_{j=1}^{t'_i} k^{(rul)}_{i,j}\mathbf{v}^{(rul)}_{i,j} \tag{21}$$

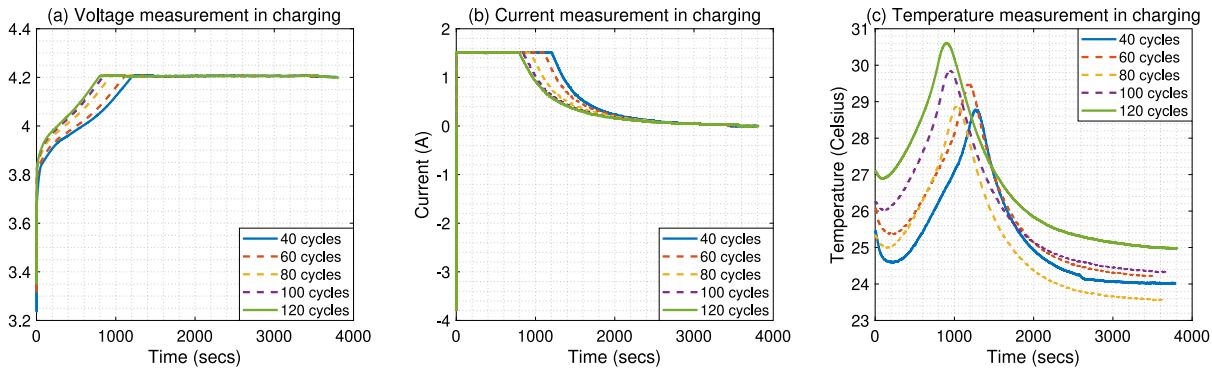

Fig. 4. The voltage, current, and temperature measurements evolve with the increment of charge cycles for battery No. 5.

Table 1
Extracted temporal features from current, voltage, and temperature measurements.

Symbol	Description
x_1	Time to the max voltage in charging
x_2	Time charged under a constant voltage mode
x_3	Time charged under a constant current mode
x_4	Time to the minimum current in charging
x_5	Time to the max temperature in charging
x_6	Time to the minimum voltage in discharging
x_7	Time discharged under a constant current mode
x_8	Time to the max temperature in discharging

4. Case study

4.1. Data description

A battery dataset generated by the NASA Ames Prognostics Center of Excellence [41] was used to demonstrate the effectiveness of the proposed method. Battery No. 5, No. 6, No. 7, and No. 18 in this dataset were used in this case study, and these batteries were operated under different operating conditions: charging, discharging, and impedance. The temperature, current, and voltage were measured during charge and discharge cycles. In charge cycles, a constant current mode at 1.5 A was carried out until the measured voltage reached 4.2 V, and resumed in a constant voltage mode until the measured current was dropped to 20 mA. In discharge cycles, a constant load mode at 2 A was applied until the voltage measurement dropped to 2.7 V, 2.5 V, 2.2 V, and 2.5 V for batteries No. 5, No. 6, No. 7, and, No. 18 respectively. The experiment was terminated when the battery capacity had been reduced by 30%. The maximum of these batteries was 2Ahr, and the end-of-life capacity was 1.4Ahr.

Fig. 4 shows the voltage, current, and temperature measurements evolve with the increment of charge cycles for battery No. 5. We could observe from this figure that the trajectories of the voltage, current, and temperature evolve with the increment of charge cycles. To capture these trajectory changes, temporal features were extracted from these measurements in both charge and discharge cycles. Table 1 shows the symbol and the descriptions of these extracted temporal features.

4.2. Graph constructions

The extracted temporal features were used to derive the covariance matrix and mutual information, and the covariance matrix was used to construct the first-stage graph. Both the constructed first-stage graph and the extracted mutual information were fed into the proposed non-linear binary optimization model to construct the second-stage graph. In order to derive the second-stage graph, the only hyperparameter that should be determined is the penalty parameter α, we set up $\alpha = 0.1$ in this case study to optimize the performance of the proposed method. Fig. 5 shows the constructed first-stage graph and second-stage graph for this case study. Next, the second-stage graph was fed into the proposed graph convolutional networks with dual attention mechanisms to predict SOH and RUL.

Table 2
The detailed structure used in the case study for SOH predictions.

No. of layers	Descriptions	Output dimensions
1	Input	30 × 165 × 8
2	Graph convolutional layer	30 × 165 × 8
3	LSTM layer	30 × 165 × 100
4	Local attention layer	30 × 5 × 100
5	Flatten layer	30 × 500
6	Dense layer	30 × 1

Table 3
The detailed structure used in the case study for RUL predictions.

No. of layers	Descriptions	Output dimensions
1	Input	30 × 165 × 8
2–11	Graph convolutional layers	30 × 165 × 8
12	LSTM layer	30 × 165 × 100
13	Global attention layer	30 × 165 × 100
14	Flatten layer	30 × 16500
15	Dense layer	30 × 1

4.3. Structure and hyperparameters

In this section, the structure and hyperparameters of the proposed method are presented to optimize the SOH and RUL prediction performance. Table 2 shows the detailed structure for SOH predictions, and Table 3 shows the detailed structure for RUL predictions. In Tables 2 and 3, 30 refers to the batch size, 165 refers to the second dimension of the parameter Φ, 8 represents the number of features, and 100 refers to the number of hidden nodes in the LSTM layer. Moreover, the number 5 in Table 2 refers to the window size of the local attention layer. The learning rate of this case study is 0.00005 and we add the l_2 norm regularization to the parameter matrix Φ with a penalty coefficient of 0.01.

4.4. SOH prediction

The SOH of a battery refers to the ratio between the current capacity and the maximum capacity, which can be written as Eq. (22), where $SOH_{i,j}$ is the SOH for battery unit i in the charge and discharge cycle j, C_i^{max} refers to the maximum capacity of the battery unit i, and $y_{i,j}$ refers to the current capacity for battery unit i in cycle j.

$$SOH_{i,j} = \frac{y_{i,j}}{C_i^{max}} \tag{22}$$

Fig. 6 shows the SOH predictions for Battery No. 5, No. 6, No. 7, and No. 18 when at least 20 charge and discharge cycles have been

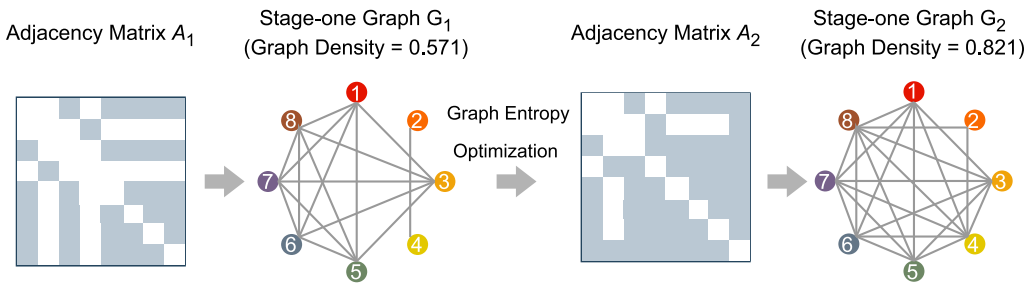

Fig. 5. The constructed first-stage graph and second-stage graph.

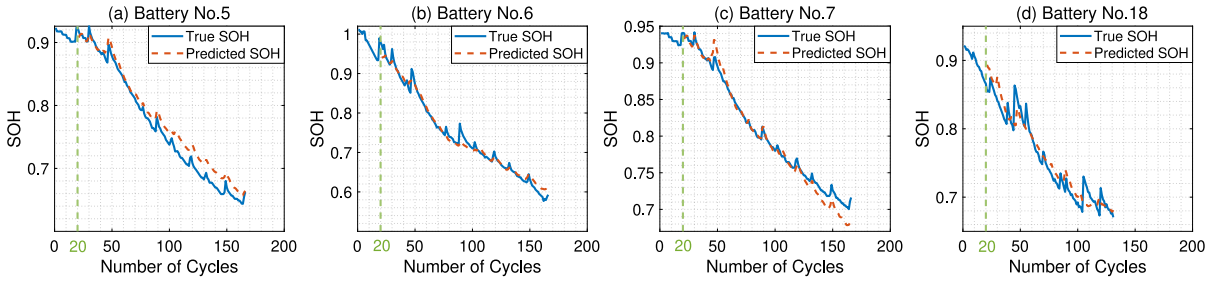

Fig. 6. The SOH predictions for Battery No. 5, No. 6, No. 7, and No. 18.

Table 4
Symbols and descriptions of the proposed method (OEGC-LSTM-DA) and comparable methods (LSTM-DA, OEGC-LSTM, and LSTM) in SOH predictions.

Method symbol	Description
OEGC-LSTM-DA	**optimal entropy graph convolutional LSTM network with dual attention mechanisms**
LSTM-DA	LSTM network with dual attention mechanisms
OEGC-LSTM	optimal entropy graph convolutional LSTM network without attention mechanism
LSTM	long short term memory network

observed. It can be observed from this figure that the proposed method can predict the SOH of batteries with high accuracy as the trajectory of the predicted SOH is closed to the trajectory of the true SOH.

To demonstrate the effectiveness of the proposed method, the proposed method is compared with other methods listed in Table 4. OEGC-LSTM-DA is the proposed method which is the optimal entropy graph convolutional LSTM network with dual attention mechanisms. LSTM-DA is a comparable method which is the LSTM network with dual attention mechanisms. OEGC-LSTM is also a comparable method which is the optimal entropy graph convolutional LSTM network without attention mechanism. LSTM serves as a benchmark method which is the traditional long short term memory network. The comparison between OEGC-LSTM-DA (proposed method) and LSTM-DA is to demonstrate the effectiveness of the proposed optimal entropy graph convolution operation. The comparison between OEGC-LSTM-DA (proposed method) and OEGC-LSTM is to demonstrate the effectiveness of the proposed dual attention mechanisms.

Fig. 7 shows the SOH prediction performance of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, and LSTM in terms of root mean squared error (RMSE). Table 5 shows the average RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, and LSTM in SOH predictions for all battery units. Based on this figure and table, we can conclude that the proposed method (OEGC-LSTM-DA) can improve the SOH prediction performance. For example, when at least 20 cycles have been observed for battery No. 7, the RMSE of the proposed method is 0.0109. However, the RMSE of LSTM-DA, OEGC-LSTM, and LSTM ranges from 0.0114 to 0.0530. When at least 40 cycles have been observed, the average RMSE of the proposed method is 0.0135. However, the average RMSE of LSTM-DA, OEGC-LSTM, and LSTM ranges from 0.0158 to 0.0307. Although LSTM-D outperforms the proposed method (OEGC-LSTM-DA) when the prediction starting point is 60, the proposed method still achieves a lower

Table 5
The average RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, and LSTM in SOH predictions for all battery units.

	Number of observed cycles				Average
	20	40	60	80	
OEGC-LSTM-DA	**0.0146**	**0.0135**	**0.0165**	**0.0168**	**0.0154**
LSTM-DA	0.0205	0.0181	0.0158	0.0212	0.0189
OEGC-LSTM	0.0154	0.0158	0.0167	0.0146	0.0156
LSTM	0.0511	0.0307	0.0301	0.0237	0.0339

average RMSE of 0.0154 in comparison with that LSTM-DA has an average RMSE of 0.0189. Moreover, another interesting finding is that the SOH prediction performance of methods listed in Table 4 does not increase with the increment of the observed charge and discharge cycles. The primary reason for such a finding is that the battery cells used in this case study have more frequent rest time when approaching their end-of-life (EOL), the rest time is completely random and leads to the capacity regeneration phenomenon. The capacity regeneration phenomenon brings more uncertainties while predicting the SOH so the methods listed in Table 4 achieve a worsened SOH performance when more charge and discharge cycles have been observed.

Moreover, the proposed method is compared with the existing methods reported in the literature to further demonstrate the effectiveness of the proposed method. Table 6 shows the RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, LSTM, multiple Gaussian regression model(MGPR), gradient boosting decision tree (GBDT), logic regression (LR) method, Gaussian process regression (GPR) method, and logic regression with Gaussian process regression (LR-GPR) method [10,42,43] when over 40 cycles have been observed. Based on this table, we can conclude that the proposed method outperforms the existing methods reported in the literature. For example,

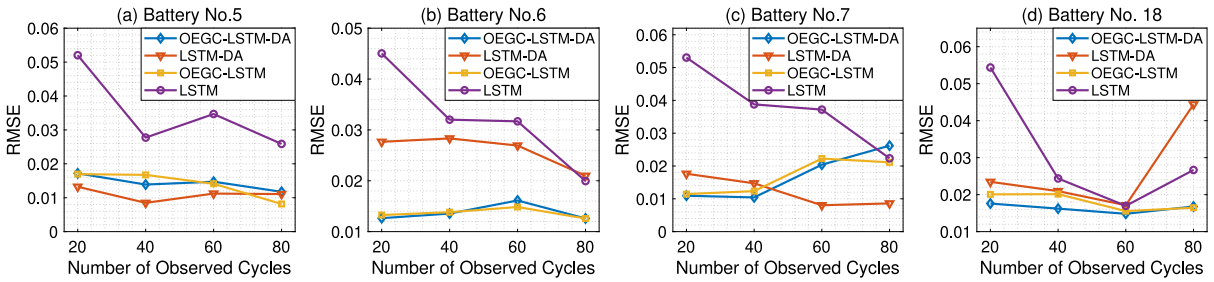

Fig. 7. The SOH prediction performance of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, and LSTM in terms of RMSE.

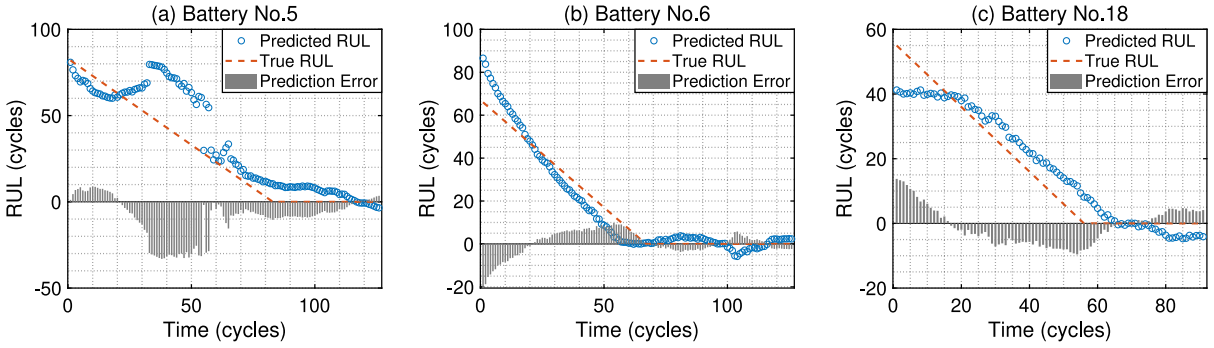

Fig. 8. RUL predictions for Battery No. 5, No. 6, and No. 18 when at least 40 cycles have been observed.

Table 6
The RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, LSTM, and other methods reported in the literature in SOH predictions.

	Battery number			
	No. 5	No. 6	No. 7	No. 18
OEGC-LSTM-DA	**0.0139**	**0.0136**	**0.0104**	**0.0162**
LSTM-DA	0.0085	0.0283	0.0147	0.0209
OEGC-LSTM	0.0167	0.0138	0.0123	0.0201
LSTM	0.0277	0.0320	0.0388	0.0244
MGPR [42]	0.0096	0.0167	0.0129	0.0228
LR-GPR [10]	0.0168	0.0292	–	0.0169
GBDT [43]	0.0192	0.0281	0.0157	–
GPR [10]	0.0751	0.0406	–	0.0323
LR [10]	0.0289	0.0252	–	0.0221

Table 7
The RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, LSTM, and other methods reported in the literature in RUL predictions.

	Battery number			
	No. 5	No. 6	No. 7	No. 18
OEGC-LSTM-DA	**14.93**	**5.80**	**31.71**	**5.84**
LSTM-DA	32.56	42.00	33.01	8.97
OEGC-LSTM	19.26	6.20	31.64	6.35
LSTM	27.52	40.20	33.62	6.46
LR-GPR [10]	16.60	21.10	–	13.50
GPR [10]	40.70	29.50	–	20.6
LR [10]	24.00	22.40	–	14.30

for battery No. 5, the RMSE of LSTM is 0.0277, but the RMSE of the proposed method is only 0.0139. For battery No. 18, the RMSE of the proposed method is 0.0162. However, the RMSE of the other methods ranges from 0.0169 to 0.0323.

4.5. RUL prediction

In this case study, we assume that a battery reaches its end of life when it has lost 30% capacity. Fig. 8 shows examples of the RUL predictions for Battery No. 5, No. 6, and No. 18 when at least 40 cycles have been observed. It can be observed from these figures that the proposed method can predict the RUL of batteries with relatively high precision. For example, for battery No. 5, the predicted RUL is 80.8 cycles when the true RUL is 82 cycles. For battery No. 6, the predicted RUL is 49.0 cycles when the true RUL is 51 cycles. For battery No. 18, the predicted RUL is 39.2 cycles when the true RUL is 37 cycles.

To further demonstrate the effectiveness of the proposed method, the proposed method is compared with methods listed in Table 4 and other methods reported in the literature. Table 7 shows the RMSE of the proposed method (OEGC-LSTM-DA), LSTM-DA, OEGC-LSTM, LSTM, logic regression (LR) method, Gaussian process regression (GPR) method, and logic regression with Gaussian process regression (LR-GPR) when at least 40 cycles have been observed. Based on this table, we can observe that the performance of the proposed method (OEGC-LSTM-DA) outperforms other methods. For example, for battery No. 18, the RMSE of the proposed method is 5.84. However, the RMSE of LSTM-DA, OEGC-LSTM, and LSTM ranges from 6.35 to 8.97, and the RMSE of other methods reported in the literature ranges from 13.50 to 20.6. For battery No. 6, the RMSE of the proposed method is 5.80. However, the RMSE of LSTM-DA, OEGC-LSTM, and LSTM range from 6.20 to 42.00, and the RMSE of other methods reported in the literature ranges from 21.10 to 29.50.

5. Discussion

5.1. Impact of optimal entropy graph

In this section, the impact of the proposed optimal entropy graph is demonstrated. We compare the performance of the optimal entropy graph with the complete graph and a randomly generated graph in both SOH and RUL predictions. The complete graph refers to an undirected graph in which every combination of feature nodes is connected with a single undirected edge, and a random graph refers to an undirected graph in which the edge of feature nodes are determined randomly. Table 8 shows the RMSE of using the proposed optimal entropy graph, the complete graph, and the randomly generated graph in SOH and RUL predictions when at least 40 cycles have been observed. Based on this

Table 8
The RMSE of using the proposed optimal entropy graph, the complete graph, and the randomly generated graph in SOH and RUL predictions.

	SOH prediction				RUL prediction			
	No. 5	No. 6	No. 7	No. 18	No. 5	No. 6	No. 7	No. 18
Optimal entropy graph	**0.0139**	**0.0136**	**0.0104**	**0.0162**	**14.93**	**5.80**	**31.71**	**5.84**
Complete graph	0.0141	0.0162	0.0134	0.0173	15.27	7.63	31.78	7.15
Random graph	0.0139	0.0164	0.0122	0.0170	20.41	7.23	31.77	6.67

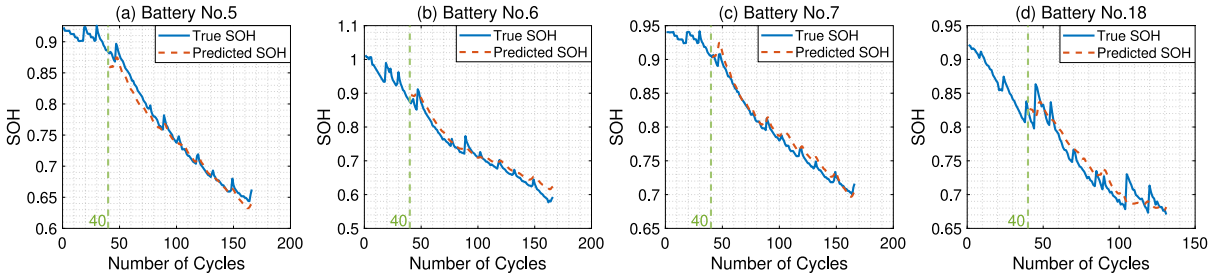

Fig. 9. The SOH predictions for Battery No. 5, No. 6, No. 7, and No. 18 with only using condition monitoring data collected from discharge cycles.

table, we can conclude that using the proposed optimal entropy graph enables better SOH and RUL predictions. For example, the RMSE of the optimal entropy graph is 0.0136 in SOH prediction for battery No. 6. However, the RMSE of the complete graph and the randomly generated graph are 0.0162 and 0.164, respectively. The RMSE of the optimal entropy graph is 5.84 in RUL prediction for battery No. 18. However, the RMSE of the complete graph and the randomly generated graph are 7.15 and 6.67, respectively.

5.2. SOH and RUL predictions without data from charge cycles

Condition monitoring data in charge cycles are not always available in many applications in practice [44]. For example, the condition monitoring data in charge cycles is not available for unmanned aerial vehicles during flight. The data in charge cycles is also not available for electric vehicles as many of the existing electric vehicle charging stations do not provide such data. To demonstrate the wider applicability of the proposed method in real conditions, we only use condition monitoring data collected from discharge cycles to make both SOH and RUL predictions in this section. Fig. 9 shows the SOH predictions with only using condition monitoring data collected from discharge cycles for Battery No. 5, No. 6, No. 7, and No. 18 when over 40 charge and discharge cycles have been observed. Based on this figure, we can observe that the proposed method can accurately predict the SOH for all batteries even only using the data collected from the discharge cycles as the trajectory of the predicted SOH is close to the trajectory of the true SOH.

Table 9 shows the RMSE of the SOH prediction with using data collected from charge cycles and without using data collected from charge cycles, where over 40 cycles have been observed. Based on this table, we can conclude that the proposed method is not sensitive to the availability of data collected from charge cycles in SOH predictions. For example, the RMSE for battery No. 18 is 0.0162 with using data from charge cycles for SOH predictions, and the RMSE for battery No. 18 is 0.0170 without using data from charge cycles for SOH predictions. The average RMSE difference between using data collected from charge cycles and without using data collected from charge cycles for SOH predictions is only 0.0003. Moreover, even without using the data collected from charge cycles, the proposed method still outperforms the other methods listed in Table 6, where the other methods use data collected from both charge and discharge cycles. For example, without using data from charge cycles, the RMSE of the proposed method for battery No. 7 is 0.0088. However, with using data from charge cycles, the RMSE of other methods listed in Table 6 for battery No. 7 ranges from 0.0147 to 0.0388. Without using data from charge cycles, the RMSE of the proposed method for battery No. 5 is 0.0110. However, with using data from charge cycles, the RMSE of LR-GPR for battery No. 5 is 0.0168.

Table 10 shows the RMSE of the RUL prediction with using data collected from charge cycles and without using data collected from charge cycles, where over 40 cycles have been observed. Based on this table, we can conclude that the proposed method is not sensitive to the availability of data collected from charge cycles in RUL predictions. For example, the RMSE for battery No. 18 is 5.84 with using data from charge cycles for SOH predictions, and the RMSE for battery No. 18 is 7.00 without using data from charge cycles for SOH predictions. The average RMSE difference between using data collected from charge cycles and without using data collected from charge cycles for SOH predictions is only 1.94. Moreover, even without using the data collected from charge cycles, the proposed method still outperforms the other methods listed in Table 7, where the other methods use data collected from both charge and discharge cycles. For example, without using data from charge cycles, the RMSE of the proposed method for battery No. 5 is 7.16. However, with using data from charge cycles, the RMSE of other methods listed in Table 6 for battery No. 5 ranges from 16.60 to 40.70. Without using data from charge cycles, the RMSE of the proposed method for battery No. 18 is 7.00. However, with using data from charge cycles, the RMSE of LSTM-DA for battery No. 5 is 8.97.

6. Conclusion and future work

In this paper, a two-stage optimization model was introduced to construct an undirected graph with optimal graph entropy to maximize graph entropy and minimize graph density simultaneously. The first-stage graph was established based on the feature similarity, then the constructed first-stage graph was fed into the proposed nonlinear integer optimization model to establish the second-stage graph. Next, the spectral graph convolution operation was introduced to develop a graph convolutional network (GCN) to utilize the topological structure of the graph to better reveal the correlation among extracted features. Last, the output of GCN was connected to an LSTM network with dual attention mechanisms to respectively predict the SOH and RUL of a battery, where dual attention mechanisms enabled an LSTM network to use the most relevant part of time series data to make predictions so that time series can be better analyzed. A battery dataset collected by the NASA Ames Prognostics Center of Excellence was used to demonstrate the effectiveness of the proposed methodology. The experimental results have shown that the proposed method can accurately predict the SOH and RUL with a minimum RMSE of 0.0104 and 5.80, respectively. The proposed method also outperforms existing

Table 9
The RMSE of the SOH prediction with using data collected from charge cycles and without using data collected from charge cycles.

	Battery Number				Average
	No. 5	No. 6	No. 7	No. 18	
With using data from charge cycles	0.0139	0.0136	0.0104	0.0162	0.0135
Without using data from charge cycles	0.0110	0.0185	0.0088	0.0170	0.0138

Table 10
The RMSE of the RUL prediction with using data collected from charge cycles and without using data collected from charge cycles.

	Battery number				Average
	No. 5	No. 6	No. 7	No. 18	
With using data from charge cycles	14.93	5.80	31.71	5.84	14.57
Without using data from charge cycles	7.16	20.48	31.42	7.00	16.51

data-driven methods, such as gradient boosting decision trees, LSTM, and the Gaussian process in both SOH and RUL predictions. We also demonstrated that the proposed optimal graph entropy can better reveal the feature correlations of condition monitoring data as using the proposed graph enables a more accurate SOH and RUL prediction in comparison with using the complete graph and a randomly generated graph. Moreover, as condition monitoring data from charge cycles are not always available in practice, we also demonstrated that the proposed method is not sensitive to the availability of data in charge cycles. In the case study, we also observe that the SOH prediction performance does not increase with the increment of the observed charge and discharge cycles. The primary reason for such a finding is that the battery cells used in this case study have more frequent rest time when approaching their end-of-life (EOL), the rest time is completely random and leads to the capacity regeneration phenomenon. The capacity regeneration phenomenon brings more uncertainties and difficulties when predicting.

In the future, datasets for different lithium-ion battery types will be used to evaluate the effectiveness of the proposed method so that the wider applicability of this work can be demonstrated. In addition, there are some challenges in dealing with the battery data in practical driving conditions, such as the availability of data from charge cycles, the uncertainty brought by the capacity regeneration phenomenon, and the randomness resulting from complicated operating conditions. In this work, we have demonstrated the proposed method can accurately predict the SOH and RUL even without using data from charge cycles. However, the uncertainty brought by the capacity regeneration phenomenon and the randomness resulting from complicated operating conditions has not been addressed yet. Therefore, future work will also include the development of efficient and effective deep learning methods to deal with the uncertainty brought by the capacity regeneration phenomenon. Moreover, we will also devote ourselves to addressing the randomness issue resulting from the complicated operating conditions to increase the accuracy and robustness of the SOH and RUL prediction model.

CRediT authorship contribution statement

Yupeng Wei: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Dazhong Wu:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research was in part supported by the National Science Foundation under Grant No. 2131619.

References

[1] Xu Fan, Yang Fangfang, Fei Zicheng, Huang Zhelin, Tsui Kwok-Leung. Life prediction of lithium-ion batteries based on stacked denoising autoencoders. Reliab Eng Syst Saf 2021;208:107396.
[2] Xu Xiaodong, Tang Shengjin, Yu Chuanqiang, Xie Jian, Han Xuebing, Ouyang Minggao. Remaining useful life prediction of lithium-ion batteries based on wiener process under time-varying temperature condition. Reliab Eng Syst Saf 2021;214:107675.
[3] He Wei, Williard Nicholas, Chen Chaochao, Pecht Michael. State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation. Int J Electr Power Energy Syst 2014;62:783–91.
[4] Frost Stephen, Welford Richard, Cheung Dennis. CSR Asia News Review: October–December 2006. Corp Soc Responsib Environ Manag 2007;14(1):52–9.
[5] Baird Austin R, Archibald Erik J, Marr Kevin C, Ezekoye Ofodike A. Explosion hazards from lithium-ion battery vent gas. J Power Sources 2020;446:227257.
[6] Rezvanizaniani Seyed Mohammad, Liu Zongchang, Chen Yan, Lee Jay. Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. J Power Sources 2014;256:110–24.
[7] Tian Huixin, Qin Pengliang, Li Kun, Zhao Zhen. A review of the state of health for lithium-ion batteries: Research status and suggestions. J Clean Prod 2020;261:120813.
[8] Prasad Githin K, Rahn Christopher D. Model based identification of aging parameters in lithium ion batteries. J Power Sources 2013;232:79–85.
[9] Park Jinhyeong, Lee Munsu, Kim Gunwoo, Park Seongyun, Kim Jonghoon. Integrated approach based on dual extended Kalman filter and multivariate autoregressive model for predicting battery capacity using health indicator and SOC/SOH. Energies 2020;13(9):2138.
[10] Yu Jianbo. State of health prediction of lithium-ion batteries: Multiscale logic regression and Gaussian process regression ensemble. Reliab Eng Syst Saf 2018;174:82–95.
[11] Khaleghi Sahar, Hosen Md Sazzad, Karimi Danial, Behi Hamidreza, Beheshti S Hamidreza, Van Mierlo Joeri, et al. Developing an online data-driven approach for prognostics and health management of lithium-ion batteries. Appl Energy 2022;308:118348.
[12] Li Tianfu, Zhao Zhibin, Sun Chuang, Yan Ruqiang, Chen Xuefeng. Multireceptive field graph convolutional networks for machine fault diagnosis. IEEE Trans Ind Electron 2020;68(12):12739–49.
[13] Wei Yupeng, Wu Dazhong, Terpenny Janis. Constructing robust and reliable health indices and improving the accuracy of remaining useful life prediction. J Nondestruct Eva Diagn Progn Eng Sy 2022;5(2):021009.
[14] Eddahech Akram, Briat Olivier, Bertrand Nicolas, Delétage Jean-Yves, Vinassa Jean-Michel. Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks. Int J Electr Power Energy Syst 2012;42(1):487–94.
[15] Qu Jiantao, Liu Feng, Ma Yuxiang, Fan Jiaming. A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery. IEEE Access 2019;7:87178–91.

[16] Wei Yupeng, Wu Dazhong, Terpenny Janis. Learning the health index of complex systems using dynamic conditional variational autoencoders. Reliab Eng Syst Saf 2021;216:108004.
[17] Zhang Weige, Shi Wei, Ma Zeyu. Adaptive unscented Kalman filter based state of energy and power capability estimation approach for lithium-ion battery. J Power Sources 2015;289:50–62.
[18] Yan Wuzhao, Zhang Bin, Zhao Guangquan, Tang Shijie, Niu Guangxing, Wang Xiaofeng. A battery management system with a lebesgue-sampling-based extended Kalman filter. IEEE Trans Ind Electron 2018;66(4):3227–36.
[19] Zhang Heng, Miao Qiang, Zhang Xin, Liu Zhiwen. An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction. Microelectron Reliab 2018;81:288–98.
[20] Zhang Jiusi, Jiang Yuchen, Li Xiang, Huo Mingyi, Luo Hao, Yin Shen. An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty. Reliab Eng Syst Saf 2022;222:108357.
[21] Li Sai, Fang Huajing, Shi Bing. Remaining useful life estimation of lithium-ion battery based on interacting multiple model particle filter and support vector regression. Reliab Eng Syst Saf 2021;210:107542.
[22] He Jiabei, Tian Yi, Wu Lifeng. A hybrid data-driven method for rapid prediction of lithium-ion battery capacity. Reliab Eng Syst Saf 2022;226:108674.
[23] Li Xiaoyu, Yuan Changgui, Wang Zhenpo. State of health estimation for Li-ion battery via partial incremental capacity analysis based on support vector regression. Energy 2020;203:117852.
[24] Gou Bin, Xu Yan, Feng Xue. State-of-health estimation and remaining-useful-life prediction for lithium-ion battery using a hybrid data-driven method. IEEE Trans Veh Technol 2020;69(10):10854–67.
[25] Richardson Robert R, Osborne Michael A, Howey David A. Gaussian process regression for forecasting battery state of health. J Power Sources 2017;357:209–19.
[26] Nagulapati Vijay Mohan, Lee Hyunjun, Jung DaWoon, Brigljevic Boris, Choi Yunseok, Lim Hankwon. Capacity estimation of batteries: Influence of training dataset size and diversity on data driven prognostic models. Reliab Eng Syst Saf 2021;216:108048.
[27] Ardeshiri Reza Rouhi, Liu Ming, Ma Chengbin. Multivariate stacked bidirectional long short term memory for lithium-ion battery health management. Reliab Eng Syst Saf 2022;224:108481.
[28] Tang Ting, Yuan Huimei. A hybrid approach based on decomposition algorithm and neural network for remaining useful life prediction of lithium-ion battery. Reliab Eng Syst Saf 2022;217:108082.
[29] Park Kyungnam, Choi Yohwan, Choi Won Jae, Ryu Hee-Yeon, Kim Hongseok. LSTM-based battery remaining useful life prediction with multi-channel charging profiles. IEEE Access 2020;8:20786–98.
[30] Wang Shunli, Takyi-Aninakwa Paul, Jin Siyu, Yu Chunmei, Fernandez Carlos, Stroe Daniel-Ioan. An improved feedforward-long short-term memory modeling method for the whole-life-cycle state of charge prediction of lithium-ion batteries considering current-voltage-temperature variation. Energy 2022;124224.
[31] Ren Lei, Zhao Li, Hong Sheng, Zhao Shiqiang, Wang Hao, Zhang Lin. Remaining useful life prediction for lithium-ion battery: A deep learning approach. IEEE Access 2018;6:50587–98.
[32] Thomas MTCAJ, Joy A Thomas. Elements of information theory. Wiley-Interscience; 2006.
[33] Ross Brian C. Mutual information between discrete and continuous data sets. PLoS One 2014;9(2):e87357.
[34] Dehmer Matthias, Mowshowitz Abbe, Emmert-Streib Frank. Advances in network complexity. John Wiley & Sons; 2013.
[35] Kwun Young Chel, Yousaf Muhammad, Nazeer Waqas, Kang Shin Min, et al. The entropy of weighted graphs with atomic bond connectivity edge weights. Discrete Dyn Nat Soc 2018;2018.
[36] Jünger Michael, Liebling Thomas M, Naddef Denis, Nemhauser George L, Pulleyblank William R, Reinelt Gerhard, et al. 50 Years of integer programming 1958–2008: From the early years to the state-of-the-art. Springer Science & Business Media; 2009.
[37] Hammond David K, Vandergheynst Pierre, Gribonval Rémi. Wavelets on graphs via spectral graph theory. Appl Comput Harmon Anal 2011;30(2):129–50.
[38] Kipf Thomas N, Welling Max. Semi-supervised classification with graph convolutional networks. 2016, arXiv preprint arXiv:1609.02907.
[39] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[40] Xu Kelvin, Ba Jimmy, Kiros Ryan, Cho Kyunghyun, Courville Aaron, Salakhudinov Ruslan, et al. Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning. PMLR; 2015, p. 2048–57.
[41] Saha Bhaskar, Goebel Kai. Battery data set. NASA AMES Prognostics Data Repository 2007.
[42] Zheng Xueying, Deng Xiaogang. State-of-health prediction for lithium-ion batteries with multiple gaussian process regression model. IEEE Access 2019;7:150383–94.
[43] Qin Pengliang, Zhao Linhui, Liu Zhiyuan. State of health prediction for lithium-ion battery using a gradient boosting-based data-driven method. J Energy Storage 2022;47:103644.
[44] Shi Junchuan, Rivera Alexis, Wu Dazhong. Battery health management using physics-informed machine learning: Online degradation modeling and remaining useful life prediction. Mech Syst Signal Process 2022;179:109347.