Heliyon 10 (2024) e31119

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Optimizing energy storage plant discrete system dynamics analysis with graph convolutional networks

Yangbing Lou [a,*], Fengcheng Sun [b], Jun Ni [a]

[a] S.M. Wu Manufacturing Research Center, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, 48109, United States
[b] Shanghai Anmai Future Energy Ltd, Shanghai, China

ARTICLE INFO

Keywords:
GCN
Packet switching
Temporal depth-separated convolutional modules
Energy storage plants

ABSTRACT

Addressing the challenges of suboptimal model performance and excessive parameters and operations in the optimization of energy storage power plants utilizing Graph Convolutional Network (GCN), this paper introduces a novel approach – the packet-switched graph convolutional network. Initially, a GCN extreme learning machine is established. Drawing inspiration from this solid foundation, we have innovatively crafted a group exchange graph convolution module. This module leverages group graph convolution techniques to amalgamate unique node feature information, tailored to diverse topology graph matrices based on various groupings. This innovative approach ensures that information flows freely and effectively among distinct groupings. Furthermore, we have designed a cutting-edge timing depth separation convolution module, comprising two innovative components. The first component introduces timing depth separation convolution, revolutionizing the original timing convolution module. The second component, the packet-switching graph convolutional network, revolutionizes the time sequence depth separation convolution process. It achieves this by employing 1×1 convolutional layers between different feature fusion packets, enabling seamless information exchange between distinct packets. Experimental results demonstrate the efficacy of the proposed model, with root mean square error (RMSE) metrics and root mean square error (MAE) metrics for single-step prediction reaching 46.08 and 26.22 at 60 min, respectively. In multi-step testing, the proposed model exhibits a 14.71 % reduction in RMSE error at the 15-min scale and a 9.29 % reduction at the 60-min scale compared to the benchmark model. This performance improvement enhances the operational efficiency and reliability of the energy storage plant, particularly under dynamic changes in the time series.

1. Introduction

In tandem with the global transformation of the energy landscape, the share of renewable energy in the overall energy supply is steadily increasing. Energy storage technology has emerged as a pivotal solution to address the intermittency challenges posed by renewable energy sources. Notably, electric energy, as a burgeoning energy source, has garnered widespread attention and application in the development of energy storage power plants. As the world promotes the adoption of distributed energy systems to meet escalating demands for electric energy, these decentralized discrete energy storage plant systems encounter challenges characterized

* Corresponding author.
E-mail address: yangbinglou@163.com (Y. Lou).

https://doi.org/10.1016/j.heliyon.2024.e31119
Received 7 March 2024; Received in revised form 17 April 2024; Accepted 10 May 2024
Available online 10 May 2024

by significant fluctuations in power loads and stringent reliability requirements.

Optimizing the discrete system of energy storage power plants assumes paramount importance in advancing energy transition objectives, enhancing power system stability and flexibility, propelling reform and development in the power market, fostering the growth of distributed energy resources, and contributing to environmental protection and sustainable development [1]. Firstly, the optimization of discrete systems in energy storage power stations can effectively address the severe environmental pollution and hazards faced by power plants during production and operation. By improving the unreasonable utilization of resources and reducing the adverse impact of production operations on the environment and ecosystem, power plants can significantly enhance their ecological and environmental protection benefits. This not only helps alleviate the current severe environmental protection pressure but also further improves the ecological environment [2]. Secondly, system optimization contributes to achieving the strategic goal of sustainable development in power plant production and operation. By enhancing the social and economic benefits of power plant production and operation, as well as prolonging the service life of thermal power systems, power plants can ensure excellent system operational safety and stability during production [3,4]. This helps to promote the continuous development and progress of power plants towards a healthy, positive, and sustainable direction, improving their competitiveness in the industry. Furthermore, system optimization not only improves the performance of mechanical equipment related to the thermal power system in power plants but also effectively enhances the requirements for energy conservation and consumption control in production operations. This helps to reduce production costs, improve energy utilization efficiency, and achieve a win-win situation in terms of economic and environmental benefits.

Given the above demands, the optimization of energy storage power stations based on graph convolutional networks (GCN) has become an emerging research field, aiming to manage and optimize energy storage power stations more efficiently through the utilization of GCN [5]. GCN, specialized neural networks designed for processing data with graph structures, prove adept at learning feature representations and patterns from graph-structured data [6]. Given the abundance of non-Euclidean data inherent in energy storage plants, encompassing battery status, power flow, device connections, and more, which can be conceptualized as graph structures, GCN offer the capability to process these irregular data structures and extract valuable features and patterns.

Currently, GCN methods are primarily categorized into two types: static and dynamic. In the static approach, the topological structure derived from the graph structure data within the energy storage power station remains unchanged throughout the entire network training process [7]. While this method is straightforward, it fails to effectively represent the connection relationships among graph structure data under specific circumstances, leading to poor generalization ability. On the other hand, dynamic methods have emerged to address these limitations. In these methods, most works utilize globalized operations to capture the associative relationships between any two types of graph structure data, effectively compensating for the shortcomings of the static approach [8]. However, these methods suffer from a shared topology across channels, limiting the ability to capture richer topological structures. Furthermore, the direct application of GCN for optimizing energy storage power stations often faces issues of excessive parameter and computational complexity, rendering it challenging to deploy on practical devices. In addition, there are also relevant studies to optimize energy storage power stations based on Bitcoin mining [9]. Therefore, there is a need to develop more sophisticated and efficient GCN methods that can capture the dynamic and complex interactions within the energy storage power station while maintaining reasonable computational requirements for practical deployment.

To address these issues, based on GCN, this paper first constructs a graph convolution overlimit learning machine, and then constructs a packet switched graph convolution module. Its advantage is that it can aggregate different node feature information according to different topological matrix under different groups, and fuse the results of packet graph convolution on the channel, so that the information of different groups can flow. In the final proposed sequential deep separation convolution module, sequential deep separation convolution is used to replace the original sequential convolution module, and packet switching is used to carry out feature fusion between different groups of the 1×1 convolution layer in the sequential deep separation convolution, so that information between different groups can flow and improve the operation efficiency and reliability of the energy storage power station under dynamic changes in the sequential sequence. The key contributions of this paper are outlined as follows.

(1) Construction of a GCN extreme learning machine. This approach operates on graph-structured data by leveraging neighbor relationships and node features, providing random graph convolutional embeddings.
(2) Development of a packet-switched graph convolution network tailored to the graph structure data of energy storage power plants. This includes the construction of corresponding topological graph matrices to efficiently represent association relationships between each graph-structured data. Features are then aggregated in different subgroups based on distinct topological graph matrices, and information from these subgroups is fused using packet switching.
(3) Introduction of an improved Timing Depth Separation Convolution Module, replacing the conventional timing convolution module. This module comprises two parts: timing depth-separated convolution and the packet-switched graph convolutional network. The expansion of the time series dependence range significantly reduces the total number of network parameters.

2. Related works

In recent years, graph-based network models have garnered substantial attention due to their efficacy in processing graph-structured data. Such models, designed for graphs comprising numerous nodes, exhibit noteworthy effectiveness. The primary configurations of graph networks include the fusion of graph networks with long short-term memory (LSTM) networks. Literature [10] integrated structure mirrors the architecture of an LSTM, featuring input gates, oblivion gates, and output gates. Notably, these gates deviate from conventional ones, incorporating graph convolutional modules that process data with a graph structure. Moreover, the

incorporation of the attention mechanism [11] into graph networks endows them with the capability to autonomously discern relationships among nodes within the graph. Another manifestation of graph networks arises through their integration with convolutional neural networks (CNN), denoted as GCN [12]. [13] pioneers the application of GCN in node-based recognition, introducing the ST-GCNI network for modeling relationships among human skeletal nodes. This method classifies nodes into categories based on their proximity to the center node, distinguishing between those near the center, the current node, and those distant from the center. Additionally, a parameterized edge weight matrix is introduced to represent the node strength of the links.

Addressing the occlusion challenge in the task [14], proposes a multi-stream graph convolutional neural network [15]. designs the MS-G3D model, employing a complex architecture with a multilevel neighbor graph convolutional network and null temporal convolution to tackle long-distance dependency issues. This work introduces a multilevel graph convolution module to enhance the modeling capability without introducing supplementary operations. To capture global information from all nodes [16], incorporates the Non-local method [17] into the two-stream graph convolutional neural model AGCN. AGCN, a two-stream network featuring a node feature network and a graph feature network, merges the results from both streams to yield the final outcome [18]. Analogous to AGCN, the GCNL network proposed in Ref. [19] adeptly captures the correlation relationships among input nodes based on their intrinsic connections.

All aforementioned methods initially execute graph convolution operations in spatial dimensions and subsequently perform temporal convolution operations within the time sequence. The spatial graph convolution and temporal convolution are commonly associated either in series or in parallel. GCN-based methods for dynamic behavior recognition fall into two categories: static methods [20] and dynamic methods [21], categorized by the structural characteristics of the graph. In static methods, the topological graph matrix remains unchanged throughout network training. The ST-GCN network, as proposed in Ref. [22], defines a pre-established topological map matrix describing the static structure, maintaining constancy during both training and testing phases [23]. introduces a multi-scale topology capable of modeling nodes at various distances.

In contrast, dynamic methods involve a dynamically changing topological graph matrix during training [24]. presents an A-links module to explore correlations between nodes in specific cases. Additionally [25], integrates a self-attention mechanism to augment the learning of graph topology, adept at modeling associative relationships between particular nodes. The Dynamic GCN network, outlined in Ref. [26], incorporates the feature information of all nodes into computations to establish modeled relationships between any pair of nodes. Compared to the static approach, the dynamic approach automatically derives association relationships between nodes based on input node data, showcasing superior generalization ability.

The system architecture diagram of the energy storage plant resembles a graph structure, vividly depicting interactions and relationships among its components. Information sharing, aligned with requirements, is crucial, and current practices heavily involve topology sharing methods. These methods facilitate the dissemination of the static graph structure across all channels, enabling the GCN network to aggregate node features based on a common topology. However, this approach imposes a performance ceiling. Contrarily, the topology non-sharing approach employs distinct topologies on various channels or groups, circumventing the limitations associated with sharing. Some researchers have introduced different topology map matrices for subgroups within DCGCN networks, but this introduces optimization challenges due to a proliferation of parameters at the subgroup level. This paper designs the topology at the channel level based on the existing GCN network to achieve dynamic behavior analysis of the discrete system of the energy storage power station.

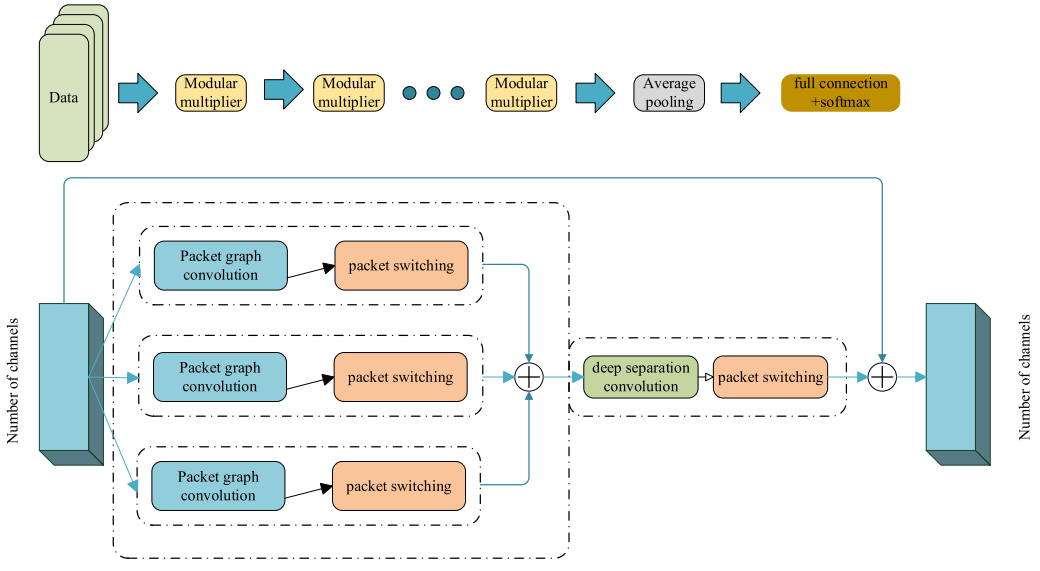

Fig. 1. Model framework.

3. Methods

The discrete system optimization model for energy storage plant based on packet switched graph convolutional network proposed in this paper consists of several modular units, each of which is shown in Fig. 1.

The modular unit comprises a packet-switched graph convolution module and a timing depth-separated convolution module. Within the packet-switched graph convolution module, the graph convolution operation undergoes a tripartite division, with each part constituting a complete packet-switched graph convolution. The outcomes of these three packet-switched graph convolution computations are ultimately aggregated. Each packet-switched graph convolution module integrates packet graph convolution and packet swapping. Packet graph convolution dissects the dynamic topological graph matrix into multiple packets, aggregating diverse graph structure data features based on distinct topological graph matrices. Packet swapping evenly distributes channel information from different packets to other packets, facilitating the circulation of information among them.

Post spatial graph convolution computation, the output results feed into the timing depth-separated convolution module. This module encompasses depth separation convolution and group exchange. Depth-separated convolution primarily involves depth convolution and 1×1 convolution, strategically devised to curtail the overall number of parameters and operations in the network. Packet switching channels the outcomes of packet computations in the 1×1 convolution, enabling seamless information flow across different packets.

3.1. Extreme learning machine

In this section we construct a graph convolutional superconvolutional learning machine in conjunction with a traditional superconvolutional learning machine, where the goal of the random graph convolutional layer is to generate a random embedding matrix H for the input undirected graph G. Given a filter parameter matrix $W \in \mathbb{R}^{m \times L}$ containing L filters, each element of which is generated according to a random probability distribution. Thus, the mathematical expression for the random graph convolutional embedding is as follows:

$$H = \sigma(\overline{A}XW) \tag{1}$$

Equation (1) utilizes the random map to multiply the normalized adjacency matrix $\overline{A}$ pair by pair. Therefore, the random map convolutional embedding theoretically does not degrade the performance of the random mapping. In addition σ denotes the nonlinear activation function, and the nonlinear activation function is Sigmoid:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{2}$$

where $\overline{A}$ represents the normalized adjacency matrix, which is given by:

$$\overline{A} = D^{-1/2}AD^{-1/2} \tag{3}$$

Equation (4) can be calculated by combining Equations (1) and (3):

$$H = \sigma\left(D^{-1/2}AD^{-1/2}XW\right) \tag{4}$$

where D is denoted as the ith element on the diagonal of the degree matrix D. The random graph convolutional embedding of the i-th node is:

$$h_i = \sigma\left[\sum_{j \in N_k(x_i) \cup i} \left(\frac{x_i W}{\sqrt{D_{ii}}\sqrt{D_{jj}}}\right)\right] \tag{5}$$

From Equqtion (5), it is evident that the graph embedding (h) of the i-th node is contingent on the feature vector of the node and its k first-order neighbor nodes. Considering the association of node attributes with the graph, there is a heightened likelihood that these neighboring nodes belong to the same class, mitigating the risk of corruption by noisy data in the obtained random graph convolutional embedding. The construction of the random graph convolutional embedding, as outlined above, leverages both neighbor relationships and node features, allowing it to effectively operate on graph-structured data.

Furthermore, owing to the inherent characteristics of graph structures, which exhibit resilience against noisy data points, random graph convolutional embeddings have a propensity to yield more robust transformations.

3.2. Group convolution module

Building upon the GCN extreme learning machine, this section introduces the grouped switched graph convolution module to enhance the generation of more diverse dynamic topologies. Comprising two components, the first part is the grouped graph convolution, which primarily consolidates distinct node feature information in different groups based on various topology graph matrices. The second part is packet switching, employed to amalgamate results from packet graph convolution on the channel, facilitating the circulation of information across different packets.

Traditional graph convolution methods rely on a topological graph matrix A, often parameterized or computed through global operations. Furthermore, feature X utilizes the same topological graph matrix A across all channels, resulting in a uniform topology for aggregating node information on all channels. However, this approach is limited to representing only generalized features on each channel. To address this limitation, the paper introduces the grouped graph convolution module, as depicted in Fig. 2.

The topological graph matrix A is partitioned into multiple subgroups, mirroring the division of input features X into an equivalent number of subgroups. Each subgroup of features corresponds to a distinct topological graph matrix, enabling aggregation with different node information in various subgroups. This design enhances the extraction of more informative features.

The dynamic topology map matrix $A_d \in \mathbb{R}^{C\times V\times V}$ is first grouped on the channel, i.e:

$$A_d = A_1 \| A_2 \| ... \| A_g \tag{6}$$

All dynamic topological map matrices of the same group are identical and C is the total number of channels. This topological map matrix is a parameterized adjacency matrix. It also groups the input features $X \in \mathbb{R}^{C\times T\times V}$ on the channels:

$$X = X_1 \| X_2 \| ... \| X_g \tag{7}$$

Both the dynamic topological map matrix and the input features are segmented into an equal number of subgroups on the channel. Each subgroup of input features aligns with a dynamic topological map matrix, and the graph convolution operation is independently executed within each subgroup, as illustrated below:

$$Y_i = (A_i + A_s) X_i W_i, i = 1, 2, ..., g \tag{8}$$

$$Y = Y_1 \| Y_2 \| ... \| Y_g \tag{9}$$

Where, $Y_i \in \mathbb{R}^{C\times T\times V}$ is the computational result of the ith grouping, $W_i \in \mathbb{R}^{C\times C}$ is the parameter matrix of the ith grouping different groupings have different parameter matrices, $Y \in \mathbb{R}^{C\times T\times V}$ combines all channels to get the graph convolution output. Group graph convolution retains the static topological map matrix, the input features are grouped on the channel each grouping corresponds to a dynamic topological map matrix, so that each grouping will be aggregated to different node feature information according to different topologies.

The group graph convolutional computation is independently conducted on different subgroups, with the feature information of each subgroup exclusive to that subgroup, precluding information exchange between subgroups. To facilitate the integration of feature information across subgroups, this paper introduces a subgroup exchange mechanism, depicted in Fig. 3. This mechanism evenly distributes feature information from different subgroups to others, ensuring that each subgroup incorporates information from other subgroups, fostering comprehensive information integration. Importantly, this approach eliminates redundant convolution calculations, consequently reducing the overall number of network parameters.

Assuming a total of g subgroups, each comprising n channels, resulting in a total of g*n channels. Following this, the channels of each subgroup are further divided into g subgroups, and these subgroups are evenly distributed to other subgroups. As a result, each subgroup assimilates feature information from other subgroups. Post-subgroup exchange, node information aggregated by different subgroups, based on distinct topological graph matrices, undergoes fusion.

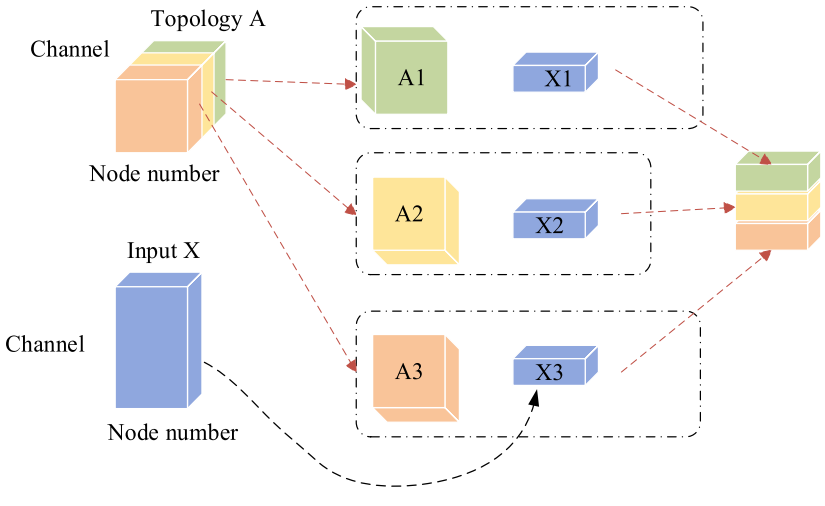

Fig. 2. Group graph convolution process.

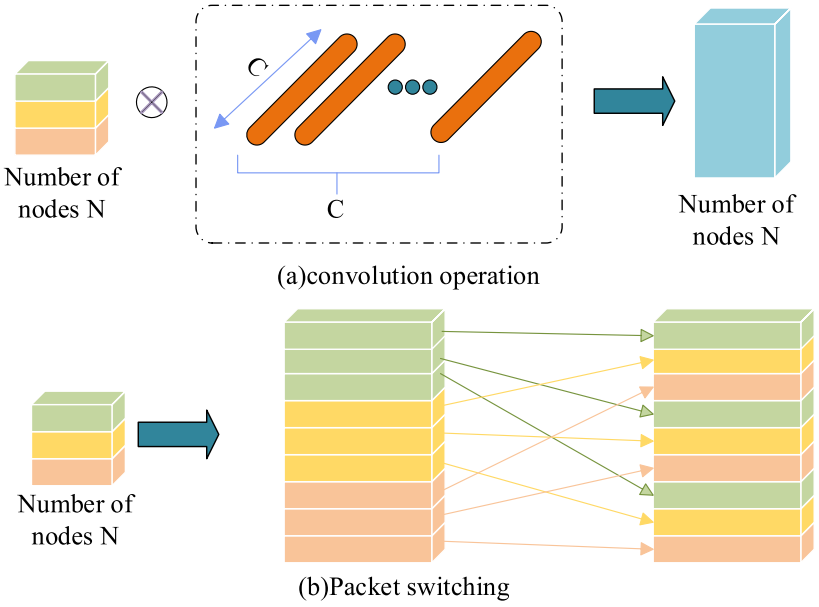

Fig. 3. Packet switching diagram.

3.3. Depthwise separable convolution module

The group convolution module primarily conducts spatial aggregation operations on nodes, yet the sequence of nodes within an energy storage plant encapsulates crucial information about temporal changes. This temporal sequence information is pivotal for energy storage optimization. Existing studies often resort to using a large timing convolution kernel to capture rich timing feature information, thereby expanding the range of temporal dependencies. However, the adoption of a larger timing convolution kernel introduces more parameters, elevating the overall model complexity. To address these challenges, this subsection introduces the timing depth separation convolution module, comprising two components. The first part is the timing depth separation convolution, serving as a replacement for the original timing convolution module. The second part involves packet switching, where this mechanism facilitates feature fusion between different packets of the 1×1 convolutional layer within the temporal depth separation convolution, enabling the flow of information between different packets.

Assuming that both the total number of input and output channels are denoted as C, the number of nodes is N, the total number of tensors is T, and the size of the temporal convolution is represented by Ω, this paper adopts deep convolution to mitigate the total number of network parameters. Deep convolution independently performs convolution operations on each channel, isolating different channels from each other and subsequently merging them after completing the convolution operation. The use of deep convolution on separate channels enables the configuration of a relatively large timing convolution kernel to expand the temporal dependency range. Additionally, a packet-switching module is added after the 1×1 packet convolution. Leveraging the graph convolution overrun learner, packet-switched graph convolution network module, timing depth-separated convolution module, and a cross-entropy loss function, an optimized networ for the energy storage plant, is obtained. The specific cross-entropy loss function is as follows:

$$Loss = \frac{1}{N}\sum_{i=1}^{N} -y_i \log y_i^* - (1-y_i)\log(1-y_i^*) \tag{10}$$

Where N is the total number of energy storage plant node sequences and y_i, y_i^* corresponds to the predicted and true values of the i-th energy storage plant node sequence, respectively. In the training phase of the network, a fixed number of energy storage plant node sequences are input to the network each time, followed by the calculation of the loss value:

$$f_i = \sigma\left(W_f \times x_t + U_f \times h_{t-1} + b_f\right) \tag{11}$$

Finally the parameter values in the network are updated by a gradient descent algorithm.

4. Experiments and analysis

In this section, we conduct a comparative analysis of the proposed optimized network for energy storage plants, employing various time scales for prediction, namely, 15 min, 30 min, 45 min, and 60 min. The objective is to assess the model's performance across different time intervals. For comparison, we consider the BP neural network (BPNN) [27], the attentional mechanism-based

spatio-temporal graph convolutional network (ASTGCN) [28], and the diffused convolutional recurrent neural networks (DCRNN) [29]. These models are evaluated on the energy storage power plant dataset from the Datong Energy Sunshan wind farm in China, providing insights into the role of each module.

In the experiments, 60 % of the data served as the training set, 20 % as the validation set, and the remaining 20 % as the test set. All experiments were conducted on an NVIDIA GeForce RTX 3090 server within a server environment featuring PyTorch deep learning library version 1.9 and Scikit-learn machine learning library version 1.0.1. Optimal parameters for all deep learning models were determined through tuning on the validation set. The convolution kernels for time-series graph convolution were uniformly set to 64. The models underwent training using the Adam optimizer with a learning rate of 0.001, a batch size of 16, and 100 training epochs.

4.1. Feature processing

Let the data information (t,y,z) of battery status, power flow, equipment connection in the energy storage power station is a three-dimensional vector, with the change of time t, and its corresponding x, y is a continuous change with time t. Therefore, it is assumed that the data of each graphical structure in the energy storage power station is a sequence of coordinates. Specifically the coordinates of all nodes under a certain time sequence are denoted as $J^t \in R^{3\times N}$, N means there are N discrete energy storage power stations.

As the coordinates of different energy storage plants vary over time, the coordinate information emerges as a crucial feature. To streamline the input format and facilitate network data processing, all input sequence data undergo processing into coordinate data within a 300-s consecutive time window. This processed data can be denoted as:

$$Sequence = \{J^t\}_{t=1}^{300} \tag{12}$$

To capture information regarding the temporal evolution of data characteristics for each energy storage plant, it is essential to leverage the variations in coordinates over time in the time series.

$$M_J^t = J^{t+1} - J^t \tag{13}$$

Where M_J^t is the change of each energy storage power station in adjacent time. This feature is used to represent the change of position of each coordinate of the energy storage power station in the time series, which is a kind of time series information.

4.2. Evaluation indicators

The experiment employs three distinct evaluation metrics to gauge the predictive effectiveness of the model: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These three evaluation indicators are calculated as follows:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_t - y_t'| \tag{14}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - y_t'}{y_t}\right| \tag{15}$$

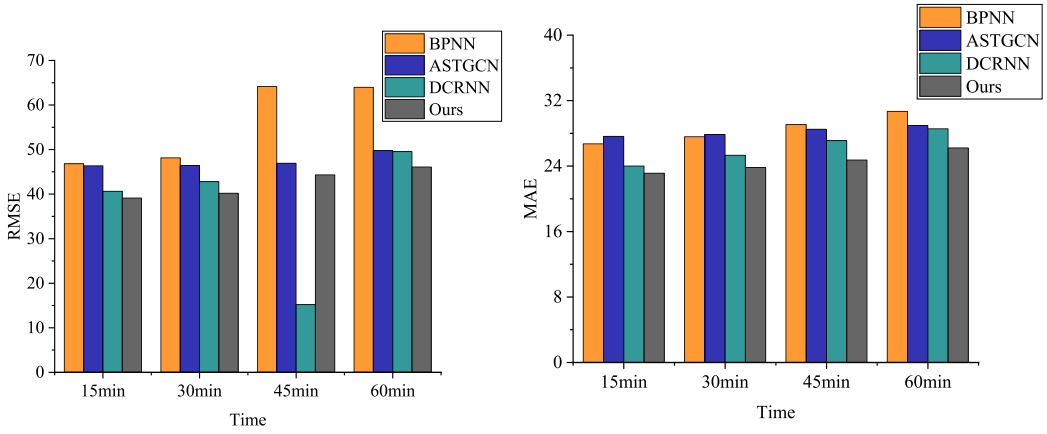

Fig. 4. The optimization performance of different models.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - y_t')^2} \tag{16}$$

Where y is the actual data, y' represents the predicted value, and T is the length of the time dimension of the observation sample. The smaller value of each of the above indicators indicates a better effect.

4.3. Model comparison

The results of the experimental comparison are depicted in Fig. 4, wherein the RMSE and MAE of the four models are compared and analyzed. BPNN, comprising simple multilayer neurons, exhibits similar metric values to ASTGCN for the 15-min prediction length. However, the disparity between BPNN and ASTGCN becomes more pronounced in long sequence prediction tasks. ASTGCN demonstrates superior effectiveness in learning trend patterns in the time dimension. In comparison with BPNN, this paper's model achieves a reduction of approximately 12.53 % and 9.32 % in RMSE and MAE indices, respectively, for single-step prediction, reaching 46.08 and 26.22 at 60 min. The utilization of the attention mechanism in ASTGCN, although slightly inferior to the GRU model trained with multiple parameterization, presents valuable insights for constructing deep learning models encompassing both temporal and spatial dimensions.

In contrast to ASTGCN, which learns features separately from temporal and spatial dimensions before fusion, the model in this paper utilizes hypergraph convolution incorporating the hyperedge relationship embedded in a recurrent neural network. Additionally, group swapping is employed to synchronize the learning of spatio-temporal features. Consequently, the RMSE prediction error of this paper's model for single-step prediction is 14.71 % lower than that of ASTGCN. As a representative model for spatio-temporal prediction, DCRNN employs diffusion convolution to learn spatial features. However, this paper's model outperforms DCRNN by considering both local and global t.

In this study, to further demonstrate the optimization effect of the models in this paper, multi-step predictions were conducted for the four models, and the results were presented through visualization. These visual representations of the prediction results provide an intuitive display of the optimization ability of the four models at different time scales, enabling a comprehensive comparison of their performance advantages and disadvantages. Fig. 5 illustrates the four-step prediction images for each benchmark model within 400

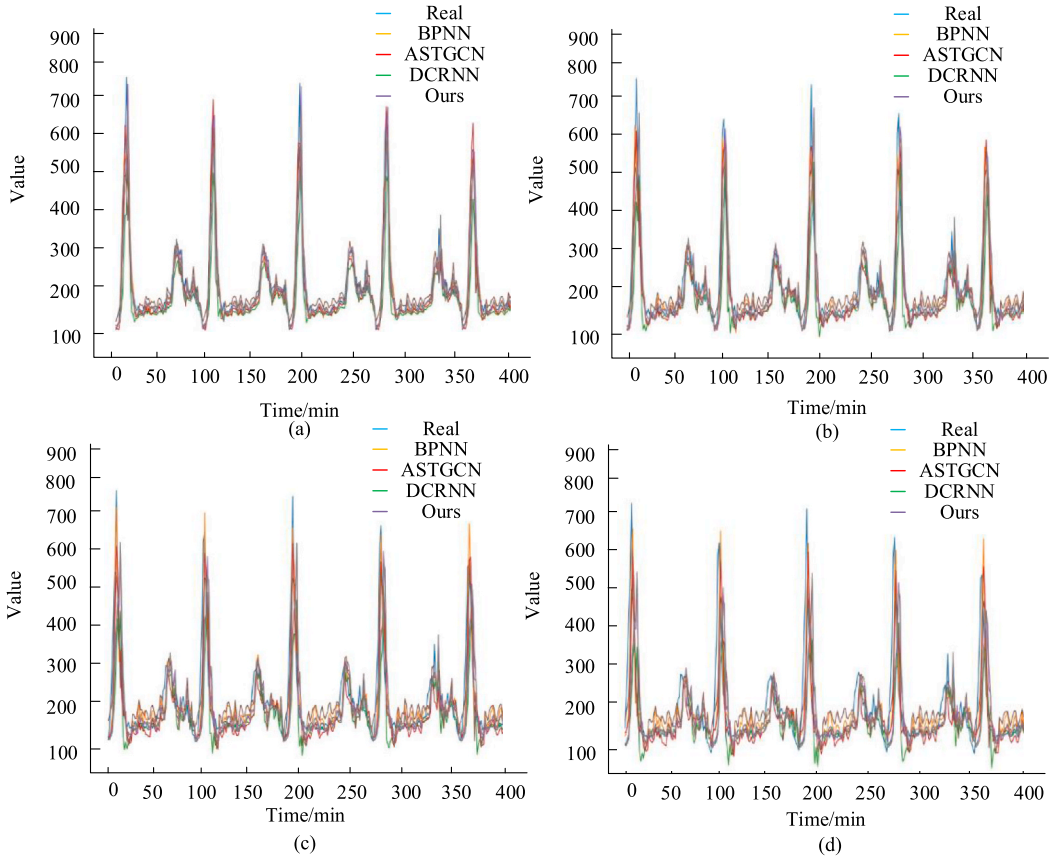

Fig. 5. Comparison of four-step prediction performance.

min. emporal features in the temporal dimension.

From the first step prediction in subfigures (a), it is evident that the overall prediction is better, with less deviation from the actual values. Among the benchmark models, the model proposed in this paper exhibits a smaller error between predicted and actual values in the one-step prediction. This advantage is attributed to the adoption of hypergraph convolution, allowing the model to capture relationships between multiple nodes simultaneously using the graph convolution hyperlocalized learning machine. This approach is well-suited for the spatially-dependent discrete system data of the energy storage power plant. In contrast, models like ASTGCN fail to fully leverage the relationships between nodes.

Furthermore, subfigures (b) and (c) show that the models in this paper have smaller errors in the second step prediction, further highlighting the advantages of employing hypergraph convolution, multi-scale temporal information, group switching, and temporal depth-separated convolution modules.

On the fourth step prediction, all models exhibit worse performance compared to the previous predictions, but the proposed model in this paper still demonstrates a partial advantage. Compared to the benchmark model, the RMSE error of the model in this paper is reduced by 14.71 % at the 15-min scale and 9.29 % at the 60-min scale. This emphasizes the superior performance and stronger generalization ability of the model proposed in this paper for time series prediction problems.

Although ASTGCN is also a spatio-temporal prediction model based on GCN, it primarily achieves prediction by learning temporal and spatial attentional weights. While it considers temporal and spatial influences to some extent, there are still shortcomings in temporal and spatial feature learning and fusion. In contrast, the model constructed in this paper evidently captures features and patterns in time series more effectively.

4.4. Analysis of modular experiments

In the proposed GCN extreme learning machine outlined in this paper, the objective of its random graph convolutional layer is to generate a random embedding matrix H for the input undirected graph G. The matrix dimension of the embedding, treated as a hyperparameter, has an impact on the prediction effect of the model. To explore the effect of the embedding dimension on the flow prediction of energy storage power plants, we compare the prediction results under different embedding dimensions. In the experiment, the size of the embedding matrix is set from 5 to 10, and the specific prediction results are shown in Fig. 6.

According to the experimental results, the prediction errors of the model for the two metrics, RMSE and MAPE, gradually decrease with the increase of the embedding dimension. The decrease is particularly noticeable when the embedding dimension increases from 5 to 8. When the embedding dimension is set to 8, the RMSE and MAPE are 39.45 and 13.26 %, respectively. However, as the embedding dimension continues to increase, the decreasing trend of RMSE and MAPE slows down, indicating that the influence of embedding dimension on the prediction effect gradually diminishes. When the embedding dimension is set to 9 and 10, the decreases in RMSE and MAPE are relatively small. This suggests that increasing the embedding dimension can improve the prediction effect within a certain range, but excessively high embedding dimensions may lead to overfitting and reduce the model's generalization ability.

In summary, the results demonstrate that the prediction model employing the graph convolutional surjective learning machine can enhance the prediction effect to a certain extent, with the optimal embedding dimension found in the experiment being 8.

To assess the impact of a specific component of the proposed model on the overall network, this subsection conducts experiments focusing on the Multi-Channel Graph Convolutional Networ. This module autonomously learns the topology on channels from the global nodes based on the input data using a Non-local-like method. To reduce the computational effort in generating the channel topology graph matrix, the module first downscales the input data. This downsizing involves setting the channel coefficients (denoted as r), indicating that the total number of channels before downsizing is r times the total number of channels after downsizing.

In order to determine the appropriate size of the channel coefficient, ablation experiments are conducted under different channel coefficients. Fig. 7 illustrates the experimental results under various channel coefficients. As the channel coefficient increases, the

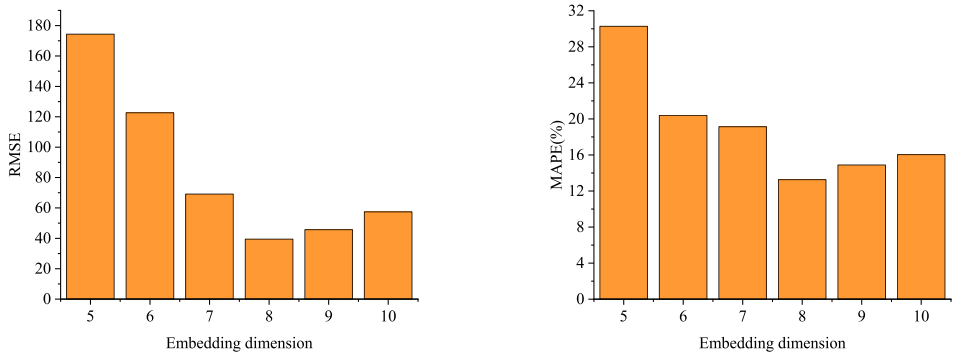

Fig. 6. Prediction results for different embedding dimensions.

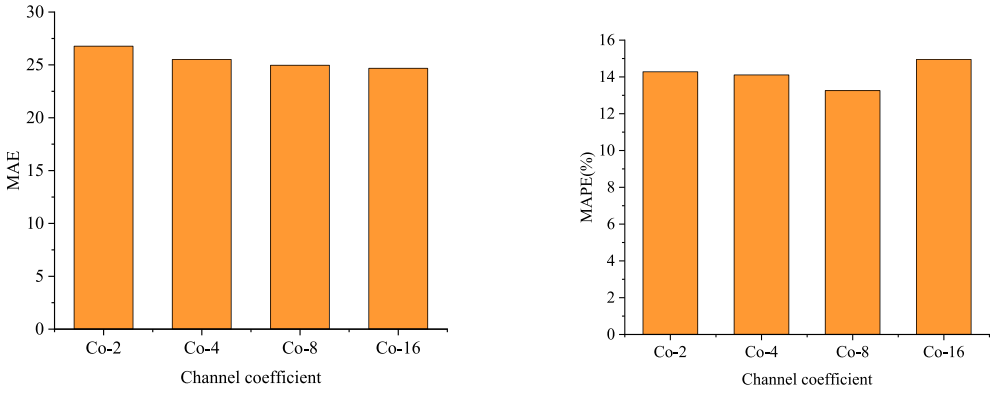

Fig. 7. Results of ablation experiments.

network parameters gradually decrease, suggesting that adjusting the channel coefficients can achieve the goal of reducing network parameters. However, purely reducing network parameters may lead to a decrease in network performance. Thus, it is necessary to strike a balance between network parameters and performance.

When the channel coefficient is set to 8, although the network parameters are not the smallest, the network performance reaches its best. In the final experimental results, the MAE and MAPE are 24.96 and 13.26 %, respectively. The difference in the number of parameters between the channel coefficients 8 and 16 is not substantial. This emphasizes the importance of finding a balance between network parameters and performance, and the optimal channel coefficient is identified as 8.

4.5. Discussion

This paper predominantly adopts the concept of group exchange, dividing the topological map matrix and input features into multiple subgroups. Each subgroup feature corresponds to a topological map matrix, enabling different subgroups to converge distinct node information based on different topological map matrices. This approach enhances feature richness and contributes to improved optimization and prediction performance in the final constructed model. By utilizing GCN to model and analyze battery state data with graph structures in energy storage power plants, the model efficiently predicts key indices such as charging and discharging state, health condition, and lifespan of the battery. This provides decision support for the optimal operation of the power plant. Learning and modeling power flow data through GCN enable intelligent scheduling and optimal distribution of power, improving operational efficiency and reliability. Furthermore, GCN monitors and analyzes equipment operation data, enabling early warning and prediction of faults, reducing equipment failure rates, and enhancing the stability and reliability of the power plant. GCN optimizes energy management in energy storage plants, encompassing energy scheduling, conversion, and consumption, facilitating efficient energy use and savings.

An energy storage power plant, composed of interconnected energy storage units, is influenced by various factors affecting its operational state. The group-switched graph convolution module in the model efficiently utilizes graph-structured data to capture spatial and temporal correlations between energy storage units, enhancing the accurate sensing and prediction of dynamic system behavior. Moreover, in an energy storage plant, faults may occur in any of the energy storage units or connection lines. By analyzing abnormal patterns in the graph structure data, the group exchange graph convolution module swiftly and accurately detects and locates faults, providing more precise system state information for optimization algorithms and thereby improving optimization effectiveness. The packet-switching graph convolution module exhibits good scalability and flexibility, adaptable to different sizes and structures of energy storage power plants. Adjusting module parameters and structure allows easy application to various types of energy storage systems and scenarios.

5. Conclusion

In this paper, we introduce the innovative concept of the group-switching graph convolution network, designed to process and fuse node features effectively under different groupings through graph convolution techniques. Leveraging the principles of graph convolutional surjective learning, we develop a packet-switched graph convolution module that employs a unique packet graph convolution strategy to aggregate features based on distinct topological graph matrices. To promote information circulation among subgroups, we fuse the grouped graph convolution results at the channel level, thereby enhancing the model's representational power and generalization capabilities. Furthermore, we introduce the temporal depth separation convolution module, which comprises two components: temporal depth separation convolution and a group exchange mechanism. The former efficiently replaces traditional time-series convolution, reducing computational complexity, while the latter facilitates feature fusion between groupings in the 1×1 convolutional layer, enriching feature diversity.

Experimental results demonstrate that our model achieves impressive performance, with RMSE and MAE metrics of 46.08 and

26.22, respectively, for single-step prediction at 60 min. During multi-step testing, our model outperforms the benchmark, reducing RMSE errors by 14.71 % at the 15-min scale and 9.29 % at the 60-min scale. Additionally, the GCN module achieves MAE and MAPE of 24.96 and 13.26 %, respectively, with a channel factor of 8. This performance enhancement has the potential to significantly improve the operational efficiency and reliability of energy storage plants, optimize their discrete systems, reduce operational costs and maintenance challenges, and promote the further development and widespread application of energy storage technology. Moreover, the application of GCN in other domains, such as recommender systems, social network analysis, and bioinformatics, also presents promising prospects.

Funding statement

This work was supported by Shanghai Science and Technology Plan Project, (Project No.23PJ1420100).

Data availability statement

Data in this article will be provided upon reasonable request.

Ethics declarations

This article does not contain any studies with human participants or animals performed by any of the authors.

CRediT authorship contribution statement

Yangbing Lou: Writing – original draft, Investigation, Formal analysis. **Fengcheng Sun:** Project administration, Methodology, Conceptualization. **Jun Ni:** Visualization, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests.

Acknowledgements

The authors would like to thank the anonymous reviewers whose comments and suggestions helped improve this manuscript.

References

[1] B. Zhou, J. Fang, X. Ai, et al., Storage right-based hybrid discrete-time and continuous-time flexibility trading between energy storage station and renewable power plants, IEEE Trans. Sustain. Energy 14 (1) (2022) 465–481.
[2] A.K. Ipadeola, K. Eid, A.M. Abdullah, Porous transition metal-based nanostructures as efficient cathodes for aluminium-air batteries, Curr. Opin. Electrochem. 37 (2023) 101198.
[3] Q. Peng, J. Rehman, K. Eid, et al., Vanadium carbide (V4C3) MXene as an efficient anode for Li-ion and Na-ion batteries, Nanomaterials 12 (16) (2022) 2825.
[4] F. Ma, Y. Liu, T. Huang, et al., Facile in situ polymerization synthesis of poly (ionic liquid)-based polymer electrolyte for high-performance solid-state batteries, Energy Convers. Manag. X (2024) 100570.
[5] T. Fu, C. Wang, N. Cheng, Deep-learning-based joint optimization of renewable energy storage and routing in vehicular energy network, IEEE Internet Things J. 7 (7) (2020) 6229–6241.
[6] S. Zhang, H. Tong, J. Xu, et al., Graph convolutional networks: a comprehensive review, Computational Social Networks 6 (1) (2019) 1–23.
[7] L Ji, X Li, W Huang, et al., Study on site selection combination evaluation of pumped-storage power station based on cycle elimination—Based on the empirical analysis of North China[J], J. Energy Storage 52 (2022) 104824.
[8] A. Kodba, T. Puksec, N. Duic, P-Graph approach for the economical optimization of biomass supply network that meets requirements on greenhouse gas emissions savings - a case study of rural areas, J. Clean. Prod. (2023) 416.
[9] F. Liu, L. Wang, D. Kong, et al., Is there more to bitcoin mining than carbon emissions? Heliyon 9 (4) (2023) e15099.
[10] R. Li, H. Wang, Graph convolutional networks and LSTM for first-person multimodal hand action recognition, Mach. Vis. Appl. 33 (6) (2022) 84.
[11] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.
[12] Y. Ding, Z. Zhang, X. Zhao, et al., Multi-feature fusion: graph neural network and CNN combining for hyperspectral image classification, Neurocomputing 501 (2022) 246–257.
[13] E.Y. Yu, Y.P. Wang, Y. Fu, et al., Identifying critical nodes in complex networks via graph convolutional networks, Knowl. Base Syst. 198 (2020) 105893.
[14] G. Yue, Y. Zhai, M. Shen, et al., MF-net: encrypted malicious traffic detection based on multi-flow temporal features[C], in: International Conference on Blockchain and Trustworthy Systems., Springer Nature Singapore, Singapore, 2023, pp. 58–71.
[15] Y. Liu, H. Zhang, D. Xu, et al., Graph transformer network with temporal kernel attention for skeleton-based action recognition, Knowl. Base Syst. 240 (2022) 108146.
[16] E. Şahin, H. Yüce, Prediction of water leakage in pipeline networks using graph convolutional network method, Appl. Sci. 13 (13) (2023) 7427.
[17] Y.L. Liu, J. Wang, X. Chen, et al., A robust and fast non-local means algorithm for image denoising, J. Comput. Sci. Technol. 23 (2) (2008) 270–279.
[18] F. Liu, Z. Fu, Y. Wang, et al., TACFN: transformer-based adaptive cross-modal fusion network for multimodal emotion recognition, CAAI Artificial Intelligence Research 2 (2023).
[19] L. Yuan, Y. Cai, J. Wang, et al., Joint multimodal entity-relation extraction based on edge-enhanced graph alignment network and word-pair relation tagging [C], Proc. AAAI Conf. Artif. Intell. 37 (9) (2023) 11051–11059.
[20] C. Chen, X. Zhao, J. Wang, et al., Dynamic graph convolutional network for assembly behavior recognition based on attention mechanism and multi-scale feature fusion, Sci. Rep. 12 (1) (2022) 7394.
[21] F. Manessi, A. Rozza, M. Manzo, Dynamic graph convolutional networks, Pattern Recogn. 97 (2020) 107000.

[22] W Wu, F Tu, M Niu, et al., STAR: An STGCN ARchitecture for Skeleton-Based Human Action Recognition, IEEE Transactions on Circuits and Systems I: Regular Papers 70 (June 2023) 2370–2383.
[23] J. Wu, O. Sigmund, J.P. Groen, Topology optimization of multi-scale structures: a review, Struct. Multidiscip. Optim. 63 (2021) 1455–1480.
[24] L. Russo, V. Casella, A. Marabotti, et al., Trophic hierarchy in a marine community revealed by network analysis on co-occurrence data, Food Webs 32 (2022) e00246.
[25] J. Lee, I. Lee, J. Kang, Self-attention Graph pooling[C]//International Conference on Machine Learning, pmlr, 2019, pp. 3734–3743.
[26] J. Chen, X. Wang, X. Xu, GC-LSTM: graph convolution embedded LSTM for dynamic network link prediction, Appl. Intell. (2022) 1–16.
[27] J. Chen, Z. Liu, Z. Yin, et al., Predict the effect of meteorological factors on haze using BP neural network, Urban Clim. 51 (2023) 101630.
[28] J. Perevozcikova, D. Pavlyuk, Attention-based spatio-temporal graph convolutional networks-A systematic review[C], in: International Conference on Reliability and Statistics in Transportation and Communication, Springer International Publishing, Cham, 2022, pp. 26–33.
[29] T. Mallick, P. Balaprakash, E. Rask, et al., Graph-partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting, Transport. Res. Rec. 2674 (9) (2020) 473–488.