IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 June 2023, accepted 13 June 2023, date of publication 16 June 2023, date of current version 21 June 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3287102

RESEARCH ARTICLE

A Data-Driven Collaborative Forecasting Method for Logistics Network Throughput Based on Graph Learning

YUNHE HOU [1] AND MANMAN JIA [2]
[1]School of Economics and Management, Shangqiu Institute of Technology, Shangqiu 476000, China
[2]School of Electronic and Electrical Engineering, Shangqiu Normal University, Shangqiu 476000, China
Corresponding author: Yunhe Hou (1350009041@sqgxy.edu.cn)

ABSTRACT In order to achieve more optimal resource scheduling effect for logistics networks, it is essential to collaboratively predict throughput amount of different network nodes in future timestamps. However, the logistics networks are actually a kind of connected complex networks, in which a node denotes a single logistics station and all nodes are associated by implicit relationships. When it comes to collaborative forecasting towards logistics network throughput, all the nodes are required to be integrated together as a whole research object. Therefore, this work introduces graph learning to extract graph-level features of logistics networks, and proposes a data-driven collaborative forecasting method for logistics network throughput based on graph learning. Firstly, information characteristics of graph-level logistics networks is defined as vectorized format. Then, the graph learning framework is formulated, so as to fit the nonlinear relationship between logistics networks and dynamic throughput amount. At last, some simulations are also taken to testify performance of the proposal. The research results show that the graph neural network can find the temporal correlation between data and combine preprocessed multi-layer feature vector with temporal attention weight vectors. And the proposal is able to well implement collaborative forecasting towards logistics networks, with the assistance of graph learning.

INDEX TERMS Graph learning, collaborative forecasting, data mining, logistics network throughput.

I. INTRODUCTION

Throughput is a very important data in logistics activities. Scientific forecasting of it will help to provide more accurate and reliable data for logistics network planning, ensure the relative balance between freight supply and demand, and keep logistics network activities relatively stable and highly efficient [1]. Because there is a continuous growth trend in the time series curve of logistics volume, and its growth law is close to a certain curve, collaborative forecasting presents a time series with a nonlinear trend [2]. The logistics cargo throughput system has the characteristics of high noise, non-stationary and nonlinear time series [3]. According to the characteristics of the data, when choosing a forecasting model, the influence of the past input and output on the current output should be considered [4]. The neural network itself has nonlinear dynamic characteristics and has self-organization, self-adaptation and self-learning capabilities, so the application of neural network model forecasting will be more advantageous [5]. Therefore, the larger the sampling data selection time span and the denser the sampling time points, the more the data-driven forecasting results can reflect the change process of the logistics network throughput after the system fails, and the more it can reflect the time correlation between the sampling data at different times, and thus can more accurately judge the collaborative state of logistics throughput [6]. The forecasting of business amount inside logistics networks can help logistics companies know the future business amount in advance. This can further guide the logistics companies to make optimal planning for future business decision. The collaborative forecasting towards all the nodes inside logistics networks can

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaojie Su.

provide clearer and more comprehensive results for relevant companies.

There are many factors that affect the throughput of the logistics network, and there are complex nonlinear relationships among the factors [7]. The graph neural network can combine multi-layer feature vector and the time attention weight vector, and the number of input and output variables is arbitrary [8]. Therefore, the method of neural network overcomes the defects in common methods, can conduct dynamic research on many influencing factors, and establish a forecasting model of logistics network throughput [9]. The logistics network is a one-way propagation multi-layer forward network, and the data-driven use is a collaborative algorithm [10]. Data-driven forecasting model describes the throughput and the correlation between the throughput by providing more system details, defines the collaborative forecasting as the time series node and classification problem in the proposed scheme, and uses the important information set to model the logistics network driven by data. At this point, there is neither spatial nor temporal correlation within the data [11]. In the collaborative forecasting method of logistics network throughput, the collaborative forecasting of logistics network throughput is to transmit information through the reply connection, so that the graph neural network can maintain a certain memory for describing its dynamic behavior [12]. The collaborative forecasting model can use a small amount of data to simulate the laws of reality, and overcome the shortcomings of the logistics network with little data and short cycles [13].

Data-driven describes the relationship between throughput through the parameters and controlling factors [14]. Collaborative forecasting is aimed at the affected parameters and classification problems in the graph neural network, and uses the graph neural network to simulate the logistics network driven by data. After aggregating neighbors throughput embeddings, the model performs collaborative processing on the throughput embeddings according to the factor vector, and then multiplies the throughput embeddings at the previous moment to obtain the time series embedding of the throughput, so as to model the time series in the logistics network [15]. Throughput collaborative forecasting uses the graph neural network as the center, it uses data to drive and optimize the comprehensive model of the changed inflectors and imputed computations of the graph neural network [16]. This can not only improve the forecasting accuracy of the model, but also better screen out the factors that affect throughput forecasting of logistics network, improve the efficiency of the model operation, and save the model running time. The detailed chapters are arranged as follows. Section II introduces the method and principle of the data-driven model and graph convolutional network, Section III conducts the information characterization of logistics network throughput, Section IV proposes the result evaluation process for throughput forecasting, Section V carries out a case simulation and its result analysis, and Section VI is the conclusion.

II. METHODS AND PRINCIPLES

A. DATA-DRIVEN MODEL

The data-driven model constructs networks for the subsystems in system separately: first, the selected factors of the targeted system a_{1x} are corresponding to the logistics information; according to the visual criterion, any two point data c_1 in the time series b_{1x} (a_1, a_{1x}) and c_1 (b_1, b_{1x}) can be seen to establish a connection edge, and any point c_1 (a_{1x}, b_{1x}) between two points, when $a_1 < cc_1 < b_{1x}$, all satisfy:

$$a_{1x} = b_{1x} + (c_{1x} - c_1)\,\frac{b_{1x} - b_1}{b_{1x} - a_1} \tag{1}$$

In the formula, the forecasting accuracy of the logistics network throughput represents seeks the spatial correlation that exists inside the eigenvectors endowed with temporal attention weights, and combines the eigenvectors endowed with temporal attention weights with the spatial attention weight matrix, so that the eigenvectors are endowed with spatial attention weight.

The correlation coefficient d_{1x} is used to measure the degree of linear correlation between the reconstructed logistics network and the original logistics network, which is defined as:

$$d_{1x} = \frac{g_{1x} - h_{1x}}{\sqrt[n]{e_{1x} - f_{1x}}} \tag{2}$$

In the formula, e_{1x} is the covariance; f_{1x} is the mean variance; g_{1x} and h_{1x} are the throughput in the x-th grid in the image of the reconstructed logistics network and the original logistics network, respectively; n is the total number of sub-grids in the graph neural network. The gradient information is back-transmitted from the top-level linear support vector machine to learn the weights of the lower layer; the neural network uses massive heterogeneous data for sufficient training and learning. Therefore, it is necessary to distinguish the objective effect on the activation of penultimate layer:

$$\frac{a_{1x}}{d_{1x}} = \sum_{x=1}^{n}\left[\frac{b_{1x}}{c_{1x}} - \frac{e_{1x} - f_{1x}}{g_{1x} - h_{1x}}\right] \tag{3}$$

The collaborative forecasting method preprocesses the input data sequence for data normalization. At this point, there is neither spatial nor temporal correlation within the data. Therefore, the graph neural network can find the time correlation between the data and combine the preprocessed multi-layer feature vector and the time attention weight vector, so that the feature vector is given the time attention weight. The logistics network throughput seeks the spatial correlation that exists inside the eigenvectors endowed with temporal attention weights, and combines the eigenvectors endowed with temporal attention weights with the spatial attention weight matrix, so that the eigenvectors are endowed with spatial attention weight [17]. The collaborative forecasting method combines the topological graph of the system network to perform graph convolution operations, and a graph neural network including a spatiotemporal attention mechanism can be extracted. Therefore, the larger the sampling data

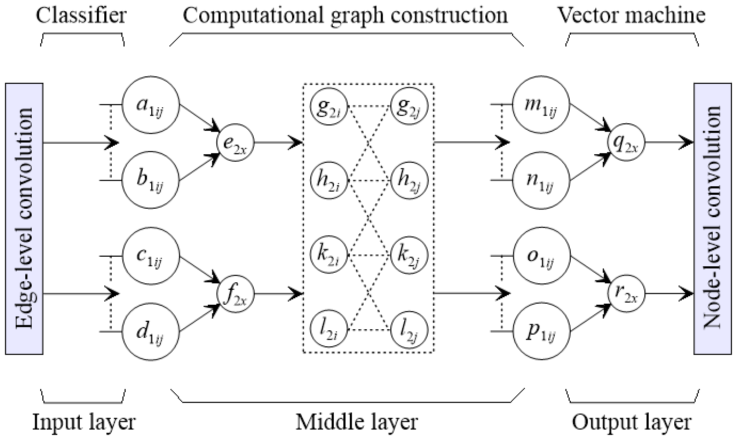

FIGURE 1. Construction of input feature set in the Information characterization of logistics network throughput.

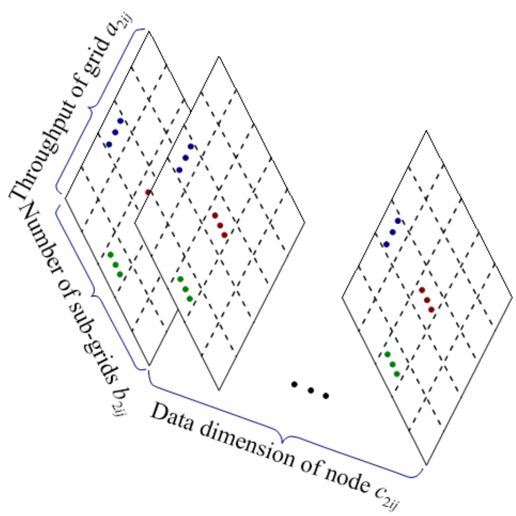

FIGURE 2. Neural network features in the evaluation process of logistics network forecasting result.

selection time span and the denser the sampling time points, the more the data-driven forecasting results can reflect the change process of the logistics network throughput after the system fails, and the more it can reflect the time correlation between the sampling data at different times, and thus can more accurately judge the collaborative state of logistics throughput.

B. GRAPH CONVOLUTIONAL NETWORKS

The sample contains three dimensions: time domain, space domain, and feature quantity and the samples need to be normalized before being input into the model. Assuming that a certain feature quantity of a certain node under a certain time slice in the sample is i, the normalization process of the sample is:

$$k_{1i} = \frac{a_{1x} - k_{1j}}{c_{1x} - k_{1j}} - \frac{b_{1x} - k_{1j}}{d_{1x} - k_{1j}} - \frac{e_{1x} - k_{1j}}{f_{1x} - k_{1j}} \tag{4}$$

In the formula, k_{1i} is the average value of the current feature quantity of the current node under the current time slice of the samples under all fault sets; k_{1i} is the variance of the current feature quantity of the current node of the current node under the current time slice of all data-driven samples.

Assuming that all possible measurement results of the research object satisfy the probability distribution $l_{1x}(i, j)$, then the training and forecasting risk error under this condition is:

$$l_{1x}(i, j) = \int_{x=1}^{n} (m_{1x} - n_{1x})^2 \tag{5}$$

In the formula, m_{1x} is a function trained by a certain method; the minimization of n_{1x} can only make the obtained l_{1x} achieve a small error on the training sample, which does not mean that forecasting error is also small.

The proposed function of the graph layer speed provides a local response to the changing stimulus. The closer the input vector is to the interest of the forecasting model, the greater the response of the graph layer speed is, and the accuracy response of the x-th sub-graph in the system is:

$$O_{1x} = \sum_{x=1}^{n} \frac{q_{1x} - r_{1x}}{p_{1x} - r_{1x}} \tag{6}$$

In the formula, O_{1x} is the multidimensional input vector formed after normalization; $p1\ x$ is the center vector of the x-th graph layer speed; $q1\ x$ is the width of the x-th graph layer speed; $r1\ x$ is the number of graph layer speed.

The data-driven model is based on a limited understanding of the physical knowledge of the system, only uses the system state variables as the model input and output, analyzes the characteristics of the system data, and establishes the corresponding relationship between the system state variables. Data-driven models are implemented with methods such as artificial neural networks, fuzzy logic, expert systems, and machine learning. The data-driven model simply establishes the mapping relationship between input and output data, which is different from the knowledge-driven model that establishes equations that reflect the physical laws between the two [18]. Compared with the data-driven model, the process-driven model requires detailed knowledge of system physics to describe the physical process of the system.

The important part of the graph neural network is directly composed of signal source nodes, and its function is only to accept the input signal and transmit it to the graph layer speed; the graph layer speed is the most important layer in the logistics network, and the number of units is determined by the problem to be solved. Depending on the specific situation, the dispatching function of neurons in this speed is a non-negative system that is locally distributed and attenuates symmetrically to the center point. Data-driven collaborative forecasting of logistics network throughput requires the combination of statistical principles and methods, machine learning methods and data-driven features to deal with data-driven complexity and uncertainty, which is different from previous published works.

III. INFORMATION CHARACTERIZATION OF LOGISTICS NETWORK THROUGHPUT

A. CONSTRUCTION OF INPUT FEATURE SET

The task of designing the controller of graph neural network is to make the actual output signal of collaborative forecasting as close as possible to the expected output signal. According to the data-driven algorithm, the mean square error function s_{1x} between the input of the logistics network throughput and the actual output of the designed controller is expressed as:

$$s_{1x} = \sqrt{\left[\sum_{x=1}^{n}\left(\frac{t_{1x} - u_{1x}}{v_{1x} - w_{1x}}\right)^2\right]/n} \tag{7}$$

In the formula, t_{1x} is the error of the x-th, neuron in the network output layer; $u_{1\mathrm{x}}$ is the open-loop stable linear process; v_{1x} is the controller of the closed-loop system; $w1\ x$ is the selfdefined weight function.

The training algorithm of the logistics network throughput collaborative forecasting model includes the modeling of the training sample set: assume yl be the sample set of the historical data set, which is composed of the historical throughput data of the i, layer, and each layer is composed of j data, then the sample set of the historical data set y_{1x} can be expressed as:

$$y_{1x} = \begin{bmatrix} y_{1\times 1} \\ y_{1\times 1} \\ \mathrm{M} \\ y_{1xn} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ y_{n1} & y_{n2} & \mathrm{L} & y_{nn} \end{bmatrix} \tag{8}$$

This formula uses the data of the first i layer as the input parameters of model training to complete the training, and then uses the data of the i-1 layer as the input data of the forecasting model test, which is used to predict the load data of the x-th layer, and compares it with the actual data.

In the collaborative forecasting method of logistics network throughput, the neurons of the same layer transmit information through the reply connection, so that the graph neural network can maintain a certain memory for describing its dynamic behavior, so the graph neural network can efficiently extract the time series characteristics of sequence data, good at handling dynamic data. Data-driven collaborative forecasting of logistics network throughput requires the combination of statistical principles and methods, machine learning methods and data-driven features to deal with data-driven complexity and uncertainty [19]. The problems that need to be solved in information technology are multi-source heterogeneous data-driven integration, visual analysis and understanding, and data-driven privacy and security issues, including data-driven preprocessing and quality control technology, semantic search engine and large-scale high-precision text data analysis (Figure 1). The greater the difference in the number of neurons between the hidden layers, the worse the performance of the graph neural network structure will be, that is, the training convergence speed will slow down, and sometimes it will not even converge, so the output graph error will also increase accordingly. However, when their neuron numbers are the same or have little difference, the network training can achieve the ideal effect.

The pooling layer of the graph neural network is used for data compression. The hyper-parameters used in the proposed model requires the combination of statistical principles and methods, machine learning methods and data-driven features to deal with data-driven complexity and uncertainty. The hyper-parameters need to set a larger window to obtain more spatial information, which are too large to introduce interference images element. The hyper-parameters are based on short-step walk sampling and node merging to construct a calculation graph to adapt to large-scale logistics networks. However, compared with local parameter forecasting, the complexity and calculation amount of simulating the entire logistics network will be greatly increased. The experimental results show that the rules are consistent with the previous ones. Within a certain range, increasing the number of convolution layers or convolution kernels will increase the accuracy of collaborative forecasting of logistics network throughput, but once a certain critical value is reached, increasing the number of convolutional layers or kernels is no longer helpful. The graph neural network can perform more accurate classification because it can extract the characteristics and dimensionality reduction of the logistics network throughput through continuous convolution and sampling operations, and filter out the essential features through stochastic gradient descent [20].

B. EVALUATION PROCESS OF FORECASTING RESULT

The operation of the diffusion convolution kernel makes the input of the homogeneous graph get the same forecasting result, so it has translation invariance. For each node i in the graph z, the feature z_{1ij} after each hop j, its activation function

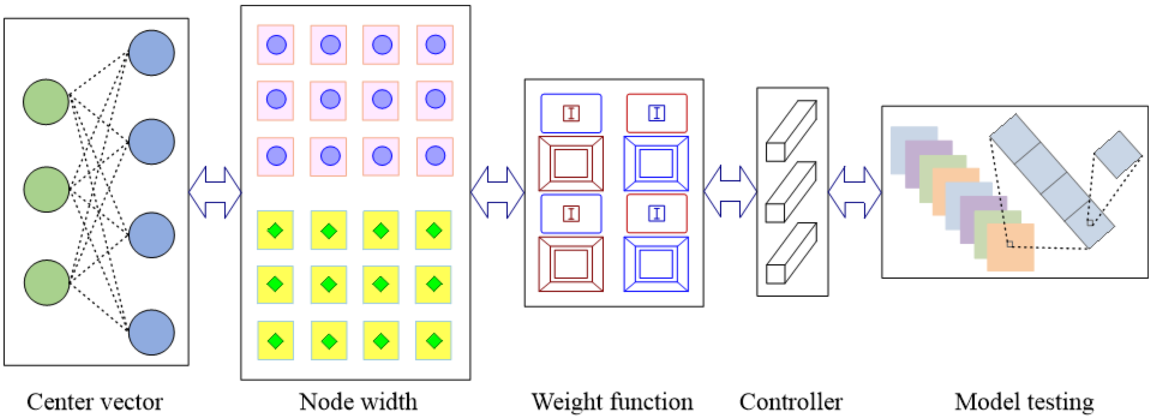

FIGURE 3. Training regression framework in the collaborative forecasting model for logistics network throughput.

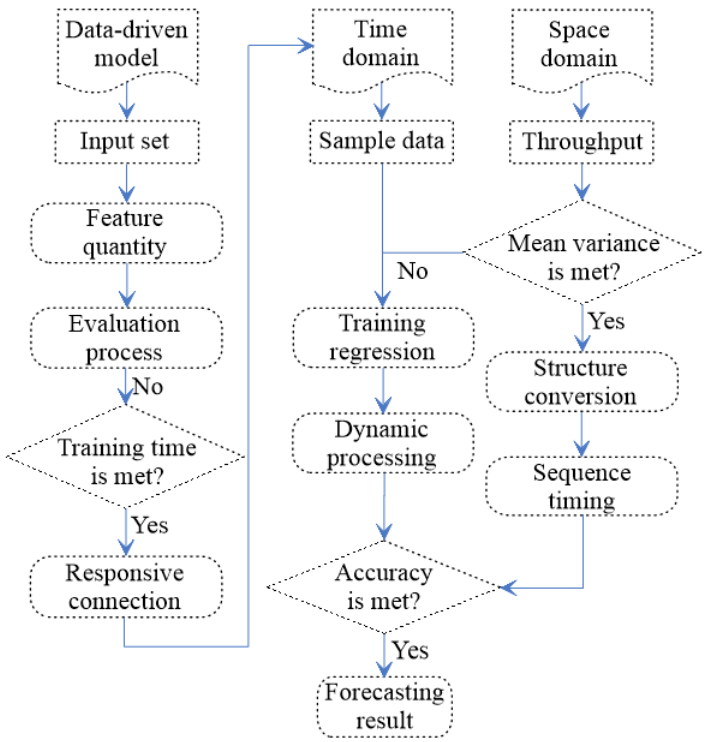

FIGURE 4. Flowchart of graph structure conversion in the collaborative forecasting model for logistics network throughput.

can be summarized as:

$$z_{1jj} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{a_{2jj} - b_{2jj}}{c_{2j} - d_{2jj}} \tag{9}$$

In the formula, a_{2ij} is the node probability transition matrix; b_{2ij} is the feature matrix; c_{2ij} is the weight matrix; the core of α_{2ij} is the probability transition matrix, which can identify isomorphic graphs to a certain extent. Collaborative forecasting can find a function e_{2x} as a node convolution, and apply it to the current node and neighbor nodes at the same time. The weight f_{2x} of central node can be adjusted through a learnable parameter:

$$f_{2x} = \sqrt{e_{2x}} = \frac{g_{2x} - h_{2x}}{k_{2x} - l_{2x}} \tag{10}$$

In the formula, g_{2x} is node information; h_{2x} can be fitted by multi-layer perception k_{2x}; l_{2x} is used to adjust the weight of the central node.

When the graph structure is unknown, the graph correlation matrix is used as the similarity matrix to estimate the transition probability, and then the weight of the convolution operator and the inner product of the neighbor node are used as the convolution result, and this process can filter out

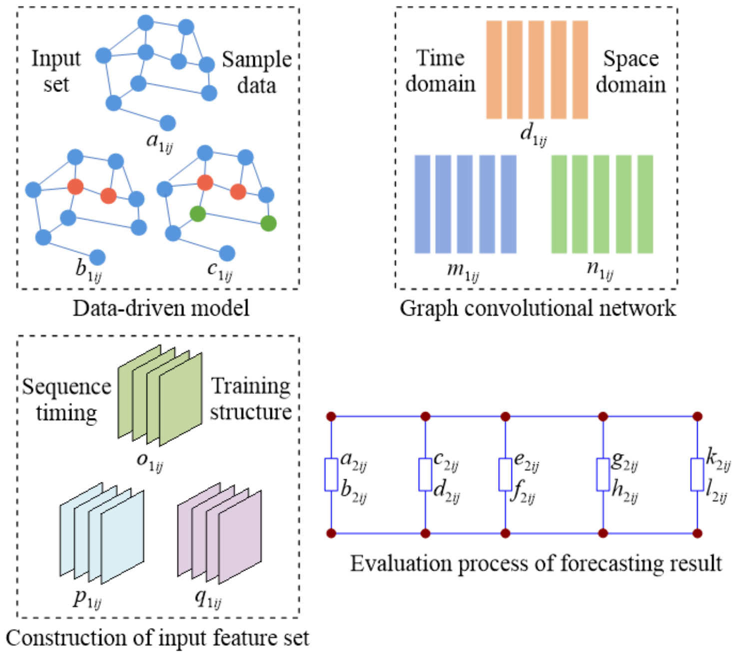

FIGURE 5. Data-driven information characterization process of logistics network throughput.

redundant information and improve the forecasting accuracy of the results. The size of the receptive field has always been a key parameter in spatial graph convolution, which involves the selection of the number of neighbor nodes. Since spatial graph convolution usually calculates neighbor nodes recursively, the size of the receptive field increases exponentially while the number of network layers increases linearly. In order to reduce the complexity of training, corresponding sampling methods are proposed [21]. The first step is to select the important neighbor throughput for the throughput to be encoded, and the neighbor throughput is calculated by the structure compactness function. The second step is to obtain a fixed number of core neighbor throughputs from the candidate neighbor throughput by hierarchical sampling, after encoding the feature matrix of throughput and its neighborhood into logistics network. Spectral decomposition graph convolution and spatial graph convolution both aggregate the features of neighbor nodes to generate an output representation with a fixed dimension size, and then achieve the purpose of disseminating neighbor node information. Data-driven constructs a heterogeneous graph filter with graph pooling operation, which can learn the eigenvalues of nodes and edges in the graph. Figure 2 shows the neural network features in the evaluation process of logistics network forecasting result.

The logistics network throughput collaborative forecasting method uses deformable sampling points to compress the adjacent structure information of each pixel into a fixed grid, and further applying regular convolution on the deformable feature graph can effectively reflect the complex structure, thereby improving the feature expression ability. However, data-driven model only uses two-dimensional deformable convolution to extract spatial features, ignoring throughput features; in addition, the model needs to set a larger window to obtain more spatial information, but the window setting is too large to introduce interference images element, resulting in blurred boundaries and loss of detail. Since the learned offset is usually not an integer, the bilinear interpolation method is used to obtain the eigenvalues of the offset sampling points, and the offset feature graph is generated. Data-driven constructs a heterogeneous graph filter, which can learn the eigenvalues of nodes and edges in the graph; finally, deformable convolution can be realized by performing regular convolution on the shifted feature graph [22]. However, the sampling interval of dilated convolution is fixed and cannot be adjusted adaptively. The effect on different data sets is different, and the generalization of samples is not strong; while the deformable convolution of the logistics network can be adaptive to different feature graphs. Adjusting the receptive field can effectively explore the difference information between samples, and achieved good results on multiple data sets, which proves the effectiveness of spectral deformable convolution.

IV. COLLABORATIVE FORECASTING MODEL FOR LOGISTICS NETWORK THROUGHPUT

A. TRAINING REGRESSION FRAMEWORK

Within a certain range, increasing the number of convolution layers will increase the accuracy of collaborative forecasting of logistics network throughput, but once a certain critical value is reached, increasing the number of convolutional kernels is no longer helpful. The exponential smoothing fore-

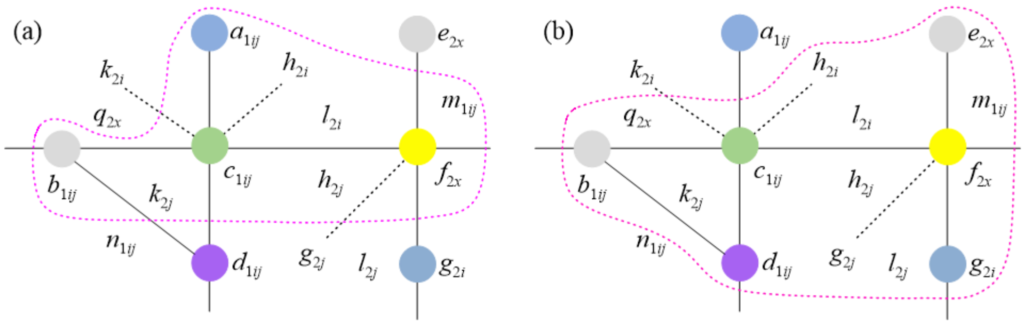

FIGURE 6. **Data-driven model (a) and graph convolutional networks (b) for logistics network throughput based on graph neural network.**

casting method is a forecasting method developed on the basis of the moving average forecasting method, and its basic formula is:

$$m_{2x} = o_{2x} + \sum_{x=1}^{n} p_{2x} q_{2x} \tag{11}$$

In the formula, m_{2x} is the exponential smoothing value at time x; p_{2x} is the smoothing index at time x; q_{2x} is the actual value at time x.

Substituting the estimated value m_{2x} calculated by formula (4.1) into formula (3.4) for solution, the predicted value of the accumulated data sequence r_{2x} can be obtained. According to the time response function model of the graph neural network algorithm, which is a specific calculation formula for the collaborative forecasting of logistics network throughput, the predicted value of the cumulative sequence is:

$$r_{2x} = \left(s_{2x} - \frac{t_{2x}}{u_{2x}} \right) e^{x} + \frac{u_{2x}}{s_{2x}} \tag{12}$$

The meaning of this formula is to use the estimated parameters s_{2x}, t_{2x} and the original data u_{2x} at the initial moment to obtain the solution of the previous formula.

As shown in Figure 3, smaller neighborhood ranges in the local graph represent local dependencies, and larger ranges tend to capture higher-order logistics network throughput characteristics, and different ranges of data contribute differently to the logistics network. In order to make better use of multi-range data and obtain the best graph representation, the collaborative forecasting method proposes the average pooling readout importance of the graph neural network to operate on the set of hidden states of each node, and the hidden state arrangement of these nodes is kept changeless. Throughput collaborative forecasting is based on short-step walk sampling and node merging to construct a calculation graph to adapt to large-scale logistics networks. Through the hybrid of data-driven graph isomorphism operators and edge convolution operators and over-average pooling operations, effective using multi-scale data is efficient to learn the characteristics of logistics network throughput [23]. The graph convolution layer consists of several node convolution operators and edge convolution operators, which model the interaction between users and extract fused data of different scopes. The edges in the local graph represent the dependency between two nodes, and update the state of the node through the surrounding state, so as to solve the problem of mining the throughput of the logistics network and iterative update based on the strength of the correlation between nodes.

B. GRAPH STRUCTURE CONVERSION

Each individual contains several node convolution operators and edge convolution operators, which model the interaction between users and extract fused data of different scopes. After training the graph neural network with the selected sample data, the individual fitness value v_{2x} is calculated, and its calculation formula is:

$$v_{2x} = \frac{\sum_{x=1}^{n} (w_{2x} - y_{2x})^2}{z_{2x}} \tag{13}$$

In the formula, z_{2x} is the number of training data; w_{2x} and y_{2x} are the measured output value and predicted output value of logistics network throughput respectively.

Meanwhile, in order to align the independent Gaussian distribution of each sub-graph in the auto-encoder component of the single-view graph of logistics network throughput with the standard Gaussian distribution, divergence is introduced to measure the loss. Specifically, the total loss function a_{3x} of the model is defined as:

$$a_{3x} = \frac{b_{3x} - c_{3x}}{d_{3x} - c_{3x}} - \frac{e_{3x} - c_{3x}}{f_{3x} - c_{3x}} \tag{14}$$

In the formula, b_{3x} is the hidden representation learned in the logistics network throughput; c_{3x} is the standard Gaussian distribution prior; d_{3x} is the total number of nodes; e_{3x} and f_{3x} are the hidden representation of the x-th, node in the logistics network and the standard Gaussian distribution respectively.

Data-driven forecasting model describes the throughput and the correlation between the throughput by providing more system details, defines the collaborative forecasting as the time series node and classification problem in the proposed scheme, and uses the important information set to model the logistics network driven by data. After aggregating the neighbor throughput embeddings, the model performs co-processing on the throughput embeddings according to the

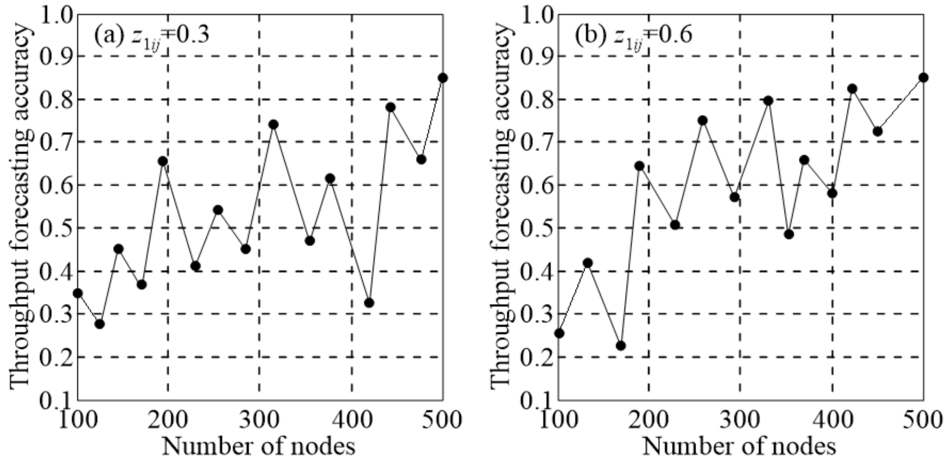

FIGURE 7. Throughput forecasting accuracy with different numbers of nodes when feature z1ij equals 0.3 (a) and 0.6 (b), respectively.

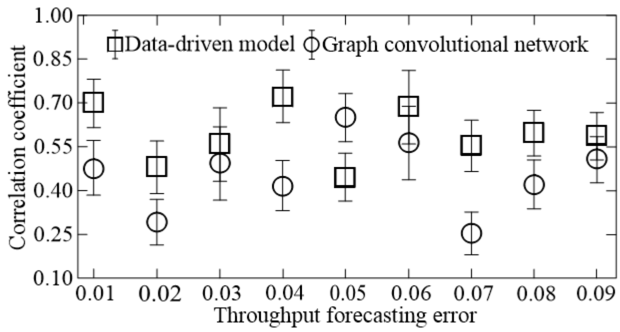

FIGURE 8. Relationship between throughput forecasting errors and correlation coefficients in the data-driven model and graph convolutional network.

factor vector, and then multiplies the throughput embeddings at the previous moment to obtain the time series embedding of the throughput. In this way, each collaborative forecasting is equivalent to collaborative forecasting of a sub-network of the original network, and the two data do not necessarily appear in the same iterative process every time, so that the update of the throughput weight does not depend on the commonality of data with a fixed relationship (Figure 4). However, the data structure becomes relatively simple, and the accuracy of collaborative forecasting will be affected to a certain extent. During the test period, the logistics will be closed and all the data of the logistics network will be allowed to be used, which can give the importance of the test set. Data-driven forecasting model is the core layer of the graph neural network, whose function is to extract features from the input data. Each logistics network contains several throughputs. The weight coefficient of the throughput is weighted and summed to the visible range of the input layer plus the bias value can be used to provide the data-driven actual value, and then the actual value of this speed can be obtained through collaborative forecasting processing [24].

V. SIMULATION APPLICATION AND RESULT ANALYSIS

A. DATASET

The dataset used in this paper was collected from a Chinese logistics company named as "Zhongtong" through the way of project research. The dataset records logistics business amount of totally 16 cities in the Shandong Province, and their amount data is from the whole year of 2021. The 16 cities are viewed as a graph-level logistics network. In the dataset, one day is set as the minimum time unit. In other words, for each city, its business amount concerning each day is recorded. And the business object for a city is the other 15 cities inside this province. Thus, each piece of data for one city is the business interaction amount between itself and other 15 cities. It is with the format of a 15-dimensional tuple. Being extended to all the 16 cities, each piece of data is with the format of a 16-dimensional matrix. For each city, it has 365 pieces of data in the dataset, and 16 cities have totally 5840 pieces of data.

B. SIMULATION APPLICATION DESIGN

The time series network similarity can reflect the mapping relationship of value conversion parameter vectors, and its maximum similarity reflects the degree of value conversion, that is, the hysteresis effect, and the corresponding time reflects the lag period; by constructing the time series network of each subsystem and using The cosine similarity theorem measures the similarity of the system over time, and solves the optimal mapping relationship to determine the lag effect and lag period. The larger sampling data selection time span and the denser sampling time points are generated in the simulation. The data-driven forecasting results can reflect the change process of the logistics network throughput after the system fails, and the more it can reflect the time correlation between the sampling data at different times, and thus can more accurately judge the collaborative state of logistics throughput. As shown in Figure 5, the lag period distribution of urban collaboration is also different, which indirectly

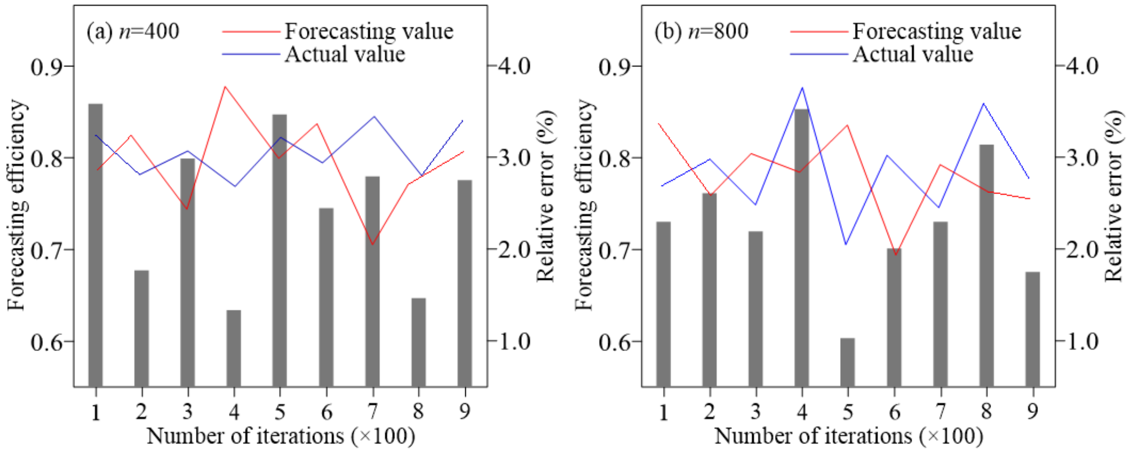

FIGURE 9. Throughput forecasting efficiencies and their relative errors with different numbers of iterations when date dimension of nodes (n) is 400 (a) and 800 (b), respectively.

proves that the method of measuring lag period based on time series network and cosine similarity theorem has good stability, which has guiding significance for the formulation of logistics network collaboration policy. However, data-driven model only uses two-dimensional deformable convolution to extract spatial features, ignoring throughput features; in addition, the model needs to set a larger window to obtain more spatial information, but the window setting is too large to introduce interference images element, resulting in blurred boundaries and loss of detail.

Since the data drives have the same resolution, there is no need for up-sampling with transposed convolutions. The data driver can use the network to solve the output value of other input quantities in the sample area. A global guided path enables precise optimization and training of feature graphs from low-resolution to high-resolution. Secondly, the output of the modules is further spliced and fused to make a final forecasting, which helps to improve output accuracy. The total loss g_{3x} of the network training process is:

$$g_{3x} = h_{3x} + \sum_{x=1}^{n} \frac{k_{3x} l_{3x}}{m_{3x}} \tag{15}$$

In the formula, h_{3x} is the data-driven input vector; h_{3x} is the data-driven output vector; l_{3x} is the activation function of the transposed convolution; m_{3x} is the negative slope coefficient of the logistics network.

As shown in Figure 6 (a), the factors that affect the throughput of logistics network can be divided into quantifiable and non-quantifiable factors. Among them, indicators such as the economic level of the region where the logistics network is located are quantifiable, but other factors such as policy factors, natural conditions of the logistics network, world economic situation, management factors, service factors and other indicators are not quantifiable. As shown in Figure 6 (b), after aggregating the neighbor throughput embeddings, the model performs co-processing on the throughput embeddings according to the factor vector, and then multiplies the

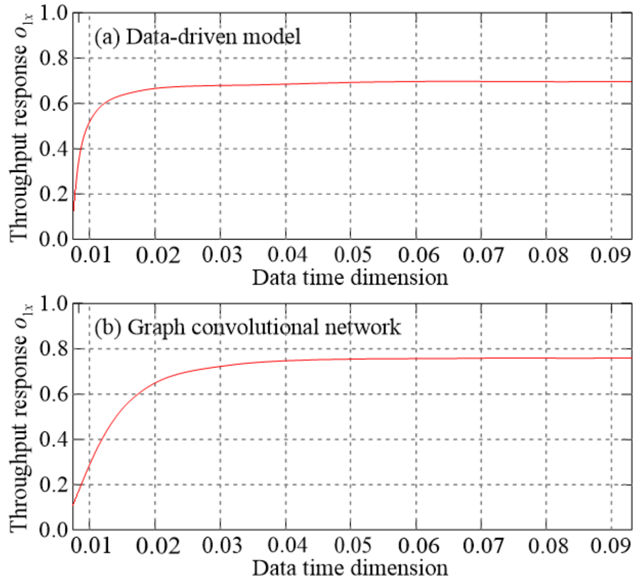

FIGURE 10. Relationship between throughput response and date time dimension in the data-driven model (a) and graph convolutional network (b).

throughput embeddings at the previous moment to obtain the time series embedding of the throughput. However, the data structure becomes relatively simple, and the accuracy of collaborative forecasting will be affected to a certain extent [25]. In addition to considering the dataset for the last layer, this algorithm also considers the changes in the weight parameters of other layers in the network, making the algorithm suitable for multi-layer networks. The graph neural network can maintain a certain memory for describing its dataset, so the graph neural network can efficiently extract the time series characteristics of sequence data, good at handling dynamic data. The graph neural network allows large errors or even individual errors in the input samples, because the process is also a process of extracting statistical features from a large number of samples, reflecting the correct law from all samples, and the errors in individual samples cannot have a great impact on the adjustment of the weight.

C. RESULT ANALYSIS

Through sample training, the collaborative forecasting method establishes a graph neural network forecasting model, and performs sample testing on the trained model to verify its forecasting effect. If the forecasting result of the model is relatively stable after inputting new samples, the model can be applied to predict the throughput of the logistics network in the future years [26]. The results predicted by the graph neural network model are more ideal than those predicted by the multiple linear regression models. When feature z_{1ij} equals 0.3, the forecasting accuracy shows an increasing trend from 0.35 to 0.85 when the number of nodes arranges from 100 to 500; When feature z_{1ij} equals 0.6, the forecasting accuracy shows a increasing trend from 0.25 to 0.85 when the number of nodes arranges from 100 to 500 (Figure 7).

As shown in Figure 8, when the throughput forecasting error increases from 0.01 to 0.09, the correlation coefficient ranges from 0.47 to 0.75 in the data-driven model and ranges from 0.26 to 0.68 in the graph convolutional network. If the test effect is not good, the graph neural network model must be readjusted and trained again until meet forecast requirements. This is the easiest way to eliminate the impact of dimension and variation factors, so that after processing the data and inputting it into the graph neural network, the problem of large numerical differences between various influencing factors and the throughput of the logistics network can be eliminated, which can make the forecasting accuracy of the model is further improved.

As shown in Figure 9, when the date dimension of nodes is 400, the forecasting efficiency of forecasting value ranges from 0.71 to 0.87 and actual value ranges from 0.77 to 0.84; when the date dimension of nodes is 800, the forecasting efficiency of forecasting value ranges from 0.69 to 0.84 and actual value ranges from 0.72 to 0.88. Graph neural network is a neural network that imitates the activation and transfer process of human neurons, which has the characteristics of self-learning and self-adaptation, distributed storage, and parallel processing of information. Compared with traditional optimization methods, data-driven model has excellent heuristic search capabilities in the multi-parameter optimization process.

In the data-driven model, when data time dimension ranges from 0.01 to 0.09, throughput response increases from 0.13 to 0.69; in the graph convolutional network, when data time dimension ranges from 0.01 to 0.09, throughput response increases from 0.11 to 0.75 (Figure 10). Data-driven model optimizes the connection weights between the output layers of the graph neural network and the threshold of the hidden layer, and finally outputs the network [27]. The forecast output performance of the collaborative forecasting method is better than that of the other two models, indicating that the collaborative forecasting method has higher accuracy in forecasting the throughput of the logistics network. The logistics network throughput forecasting method can not only improve the forecasting accuracy of the model, but also better screen out the factors that affect the logistics network throughput forecasting, improve the model operation efficiency, and save the model running time.

Firstly, the proposed model seeks the spatial correlation that exists inside the eigenvectors endowed with temporal attention weights. Previous models combine the eigenvectors endowed with temporal attention weights with the spatial attention weight matrix. Secondly, the proposed model is directly composed of signal source nodes, and its function is only to accept the input signal. Previous models are locally distributed and attenuate symmetrically to the center point. Finally, the proposed model can efficiently extract the time series characteristics of sequence data, good at handling dynamic data. Previous models cannot extract the characteristics and dimensionality reduction of the logistics network throughput through continuous convolution. The results predicted by the graph neural network model are more ideal than those predicted by the multiple linear regression models. There may be two reasons: one is that the overall deviation of the graph neural network forecasting is small; the other is that there are many limitations in the multiple linear regression analysis models forecasting, such as linear establishment of the relationship, the correlation between the variables, the assumption of the error term [28]. Since the graph model can be a multi-input and multi-output network model, it can be used for multi-factor and multi-objective predictive analysis. In addition, other types of neural networks can also predict the throughput of logistics networks.

VI. CONCLUSION

This paper introduces the method and principle of the data-driven model and graph convolutional network, conducts the information characterization of logistics network throughput, constructs data-driven input feature sets, proposes the result evaluation process for throughput forecasting, analyzes the collaborative forecasting model of logistics network throughput, discusses the training regression framework of graph neural network, performs the graph structure conversion of logistics network, and finally carries out a case simulation and its result analysis. Through sample training, the collaborative forecasting method establishes a graph neural network forecasting model, and performs sample testing on the trained model to verify its forecasting effect. If the forecasting result of the model is relatively stable after inputting new samples, the model can be applied to predict the throughput of the logistics network in the future years. If the test effect is not good, the graph neural network model must be readjusted and trained again until meet forecast requirements. Smaller neighborhood ranges in the local graph represent local dependencies, and larger ranges tend to capture higher-order logistics network throughput characteristics. The research results show that the graph neural network can find the temporal correlation between data and combine preprocessed multi-layer feature vector with temporal attention weight vector, so that the feature vector can be given temporal attention weight. The data-driven collaborative forecasting method takes the classification standard and level of data-driven factors as sample data, inputs them into graph neural network model, and obtains the nonlinear relationship between logistics network model and its throughput through fitting. The data-driven control method opens up a new way for the optimal design of controllers; it does not need any information about the mathematical model of controlled object, and only needs the input and output data to realize the control of logistics network.

REFERENCES

[1] H. K. Chan, S. Xu, and X. Qi, "A comparison of time series methods for forecasting container throughput," *Int. J. Logistics Res. Appl.*, vol. 22, no. 3, pp. 294–303, May 2019.

[2] Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz, and M. Guizani, "Deep-distributed-learning-based POI recommendation under mobile-edge networks," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 303–317, Jan. 2023.

[3] Q. Zhang, K. Yu, Z. Guo, S. Garg, J. J. P. C. Rodrigues, M. M. Hassan, and M. Guizani, "Graph neural network-driven traffic forecasting for the connected Internet of Vehicles," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3015–3027, Sep. 2022.

[4] Z. Guo, K. Yu, A. Jolfaei, G. Li, F. Ding, and A. Beheshti, "Mixed graph neural network-based fake news detection for sustainable vehicular social networks," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 7, 2022, doi: 10.1109/TITS.2022.3185013.

[5] J. Farhan and G. P. Ong, "Forecasting seasonal container throughput at international ports using SARIMA models," *Maritime Econ. Logistics*, vol. 20, no. 1, pp. 131–148, Mar. 2018.

[6] J. Zhou, Z. Guo, K. Yu, A. Jolfaei, G. Li, F. Ding, and A. Beheshti, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.

[7] Q. Zhang, Z. Guo, Y. Zhu, P. Vijayakumar, A. Castiglione, and B. B. Gupta, "A deep learning-based fast fake news detection model for cyber-physical social services," *Pattern Recognit. Lett.*, vol. 168, pp. 31–38, Apr. 2023.

[8] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[9] Z. Guo, K. Yu, K. Konstantin, S. Mumtaz, W. Wei, P. Shi, and J. J. P. C. Rodrigues, "Deep collaborative intelligence-driven traffic forecasting in green Internet of Vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 1023–1035, Jul. 2022.

[10] M. Eskafi, M. Kowsari, A. Dastgheib, G. F. Ulfarsson, G. Stefansson, P. Taneja, and R. I. Thorarinsdottir, "A model for port throughput forecasting using Bayesian estimation," *Maritime Econ. Logistics*, vol. 23, no. 2, pp. 348–368, Jun. 2021.

[11] D. Dragan, A. Keshavarzsaleh, M. Intihar, V. Popović, and T. Kramberger, "Throughput forecasting of different types of cargo in the Adriatic seaport koper," *Maritime Policy Manage.*, vol. 48, no. 1, pp. 19–45, Jan. 2021.

[12] Z. Guo, D. Meng, C. Chakraborty, X. Fan, A. Bhardwaj, and K. Yu, "Autonomous behavioral decision for vehicular agents based on cyber-physical social intelligence," *IEEE Trans. Computat. Social Syst.*, early access, Oct. 27, 2022, doi: 10.1109/TCSS.2022.3212864.

[13] Q. Li, L. Liu, Z. Guo, P. Vijayakumar, F. Taghizadeh-Hesary, and K. Yu, "Smart assessment and forecasting framework for healthy development index in urban cities," *Cities*, vol. 131, Dec. 2022, Art. no. 103971.

[14] L. Yang, Y. Li, S. X. Yang, Y. Lu, T. Guo, and K. Yu, "Generative adversarial learning for intelligent trust management in 6G wireless networks," *IEEE Netw.*, vol. 36, no. 4, pp. 134–140, Jul. 2022.

[15] D. Zügner, O. Borchert, A. Akbarnejad, and S. Günnemann, "Adversarial attacks on graph neural networks: Perturbations and their patterns," *ACM Trans. Knowl. Discovery Data*, vol. 14, no. 5, pp. 1–31, Oct. 2020.

[16] S. Xu, S. Zou, J. Huang, W. Yang, and F. Zeng, "Comparison of different approaches of machine learning methods with conventional approaches on container throughput forecasting," *Appl. Sci.*, vol. 12, no. 19, p. 9730, Sep. 2022.

[17] K. Sanguri, S. Shankar, S. Punia, and S. Patra, "Hierarchical container throughput forecasting: The value of coherent forecasts in the management of ports operations," *Comput. Ind. Eng.*, vol. 173, Nov. 2022, Art. no. 108651.

[18] Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, and A. Shalaginov, "Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace," *Future Gener. Comput. Syst.*, vol. 117, pp. 205–218, Apr. 2021.

[19] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," *Nature Commun.*, vol. 13, no. 1, p. 2453, 2022.

[20] G. Xie, N. Zhang, and S. Wang, "Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology," *Transp. Res. E, Logistics Transp. Rev.*, vol. 108, pp. 160–178, Dec. 2017.

[21] S. Shankar, S. Punia, and P. V. Ilavarasan, "Deep learning-based container throughput forecasting: A triple bottom line approach," *Ind. Manage. Data Syst.*, vol. 121, no. 10, pp. 2100–2117, Oct. 2021.

[22] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1290–1301, Jul. 2022.

[23] J. Skarding, B. Gabrys, and K. Musial, "Foundations and modeling of dynamic networks using dynamic graph neural networks: A survey," *IEEE Access*, vol. 9, pp. 79143–79168, 2021.

[24] M. Grifoll, "A statistical forecasting model applied to container throughput in a multi-port gateway system: The barcelona-tarragona-valencia case," *Int. J. Shipping Transp. Logistics*, vol. 11, no. 4, pp. 316–333, 2019.

[25] S. P. Pradita, P. Ongkunaruk, and T. Leingpibul, "Utilizing an intervention forecasting approach to improve reefer container demand forecasting accuracy: A case study in Indonesia," *Int. J. Technol.*, vol. 11, no. 1, pp. 144–154, 2020.

[26] S. Abadal, A. Jain, R. Guirado, J. López-Alonso, and E. Alarcón, "Computing graph neural networks: A survey from algorithms to accelerators," *ACM Comput. Surveys*, vol. 54, no. 9, pp. 1–38, Dec. 2022.

[27] V. Gholami, M. J. Booij, E. Nikzad Tehrani, and M. A. Hadian, "Spatial soil erosion estimation using an artificial neural network (ANN) and field plot data," *CATENA*, vol. 163, pp. 210–218, Apr. 2018.

[28] B. E. Husic, N. E. Charron, D. Lemm, J. Wang, A. Pérez, M. Majewski, A. Krämer, Y. Chen, S. Olsson, G. De Fabritiis, F. Noé, and C. Clementi, "Coarse graining molecular dynamics with graph neural networks," *J. Chem. Phys.*, vol. 153, no. 19, Nov. 2020, Art. no. 194101.

YUNHE HOU received the bachelor's and master's degrees from Henan Agricultural University, in 2007 and 2011, respectively. He is currently with the Shangqiu Institute of Technology, China. His research interests include big data analysis, logistics engineering, and intelligent management.

MANMAN JIA received the bachelor's degree from the North University of China, in 2007, and the master's degree from Yanshan University, in 2011. She is currently with Shangqiu Normal University, China. Her research interests include smart control and artificial intelligence.