A joint network of non-linear graph attention and temporal attraction force for geo-sensory time series prediction

Hongbin Dong[1] · Shuang Han[1] · Jinwei Pang[2] · Xiaodong Yu[3]

Abstract

Geo-sensory time series, such as the air quality and water distribution, are collected from numerous sensors at different geospatial locations in the same time interval. Each sensor monitors multiple parameters and generates multivariate time series. These time series change over time and vary geographically; hence, geo-sensory time series contain multi-scale spatial-temporal correlations, namely inter-sensor spatial-temporal correlations and intra-sensor spatial-temporal correlations. To capture spatial-temporal correlations, although various deep learning models have been developed, few of the models focus on capturing both correlations. To solve this problem, we propose simultaneously capture the inter- and intra-sensor spatial-temporal correlations by designing a joint network of non-linear graph attention and temporal attraction force(J-NGT) consisting two graph attention mechanisms. The non-linear graph attention mechanism can characterize node affinities for adaptively selecting the relevant exogenous series and relevant sensor series. The temporal attraction force mechanism can weigh the effect of past values on current values to represent the temporal correlation. To prove the superiority and effectiveness of our model, we evaluate our model in three real-world datasets from different fields. Experimental results show that our model can achieve better prediction performance than eight state-of-the-art models, including statistical models, machine learning models, and deep learning models. Furthermore, we conducted experiments to capture inter- and intra-sensor spatial-temporal correlations. Experimental results indicate that our model significantly improves performance by capturing both inter- and intra-sensor spatial-temporal correlations. This fully shows that our model has a greater advantage in geo-sensory time series prediction.

Keywords Geo-sensory time series · Spatial-temporal correlation · Non-linear graph attention · Temporal attraction force

1 Introduction

Multivariate time series data are gathered from a single sensor, which monitors multiple parameters, including target series and exogenous series. Target series can be defined as a time series designated for making predictions. Exogenous series are other series that influence target series. The monitoring range of a single sensor is limited and cannot represent the environmental information of an entire city or country. Therefore, it is necessary to develop multiple sensors in

different geospatial locations to simultaneously detect the environment of the entire region. These sensors with geospatial correlations collect time series called geo-sensory time series. Geo-sensory time series covers a broad scope of applications, such as traffic prediction [1], weather prediction [2], air quality prediction [3], and urban water distribution prediction [4]. Hence, forecasting geo-sensory time series has grown in importance since there is a rising demand for geo-sensory time series prediction.

One of the most significant challenges to realize accurate prediction is mining valuable knowledge from geo-sensory time series. Geo-sensory time series data are collected from certain geospatial and temporal scenarios; therefore, there are multi-scale spatial-temporal correlations in both spatial and temporal dimensions:

(1) Inter-sensor spatial-temporal correlations. Massive sensors with geospatial locations are deployed in different regions. The series generated by other sensors directly impact the series generated by the target sensor, as

✉ Shuang Han
hshanshuang@hrbeu.edu.cn

[1] College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang, China

[2] School of Computer and Control Engineering, Yantai University, Yantai City, Shandong Province, China

[3] Computer Science and Technology College, Harbin Normal University, Harbin, Heilongjiang, China

depicted in Fig. 1a. To distinguish from the target series (target sensor series), we define the target series of other sensors as the relevant sensor series. The environmental information of neighbouring sensors is usually affected due to the distance advantage. For instance, wind can blow air pollutants to neighbouring regions. In addition, some regions have similar exogenous features, such as meteorological features and traffic occupancy rates, which can generate similar environmental information. Therefore, spatial correlations are highly dynamic, changing over time. Hence, selecting relevant sensor series is of great significance to geo-sensory time series prediction. Additionally, historical information influences the current value differently, as depicted in Fig. 1b.

(2) Intra-sensor spatial-temporal correlations. For a certain (target) sensor, the intra-sensor time series contains the target series and exogenous series [5]. In the spatial dimension, the target series is usually affected by exogenous series, as depicted in Fig. 1a. In the temporal dimension, the current value is usually affected by historical information, as depicted in Fig. 1c.

In recent years, many deep learning models have been proposed for geo-sensory time series prediction. For instance, the hDS-RNN [4] model for water flow and pressure forecasting, a novel multi-channel attention model (MCAM) [6] for fine-grained air quality inference, the adaptive spatial-temporal graph attention network (ASTGAT) [7] for traffic flow forecasting, and the causal-based graph Neural network (CausalGNN) [8] for COVID-19 pandemic forecasting have focused more on extracting spatial-temporal correlations between sensors and have achieved state-of-the-art results. The drawback is that these models only consider sensor time series and ignore exogenous series. For geo-sensory time series prediction, the target series is affected by various exogenous series, such as meteorological conditions. Existing works have focused on blending the information of sensor series and exogenous series [9–11]. Some researchers have attempted to employ multilayer perceptron (MLP) [12] and parametric-matrix-based methods [13] to extract exogenous features and then fuse them into the sensor features. Although these studies confirm that prediction performance can be improved by considering various exogenous series, they cannot explicitly and dynamically select relevant sensor series and exogenous series to make predictions. Hence, it is becoming challenging to distinguish the contribution of exogenous series and relevant sensor series to the predictive value [14].

Recently, multi-level attention networks (GeoMAN) [15] have employed local spatial attention to obtain the dynamic correlations between target series and each exogenous series and have used global spatial attention to capture the dynamic correlations between different sensors. GeoMAN employs a temporal attention mechanism in the temporal dimension to model the dynamic temporal correlations in a time series. The model is a typical geo-sensory time series prediction model for capturing inter- and intra-sensor spatial correlations. However, existing works first incorporate all time series, which consist of target series, exogenous series, and relevant series, and capture temporal correlations between different time intervals, and GeoMAN is no exception. Sensor series change over time and vary geographically; hence, blindly blending the information of relevant sensor series and exogenous series makes it impossible to explore deep-seated temporal correlation. Although works for geo-sensory time series prediction tasks have been significant, the aforementioned challenges are still not addressed well.

To overcome the aforementioned challenges, our focus is on capturing inter- and intra-sensor spatial-temporal correlations. Inspired by graph attention networks that capture spatial dependencies by assigning different weights to different neighbourhood nodes, we propose a joint network of non-

Fig. 1 **a** Spatial correlation. The graph structure inside the rectangle indicates the intra-sensor spatial correlations. The graph structure outside the rectangle represents the inter-sensor spatial correlations. **b** Temporal correlation. The contributions of the exogenous series and relevant sensor series to the predictive value are different

(a) (b) (c)
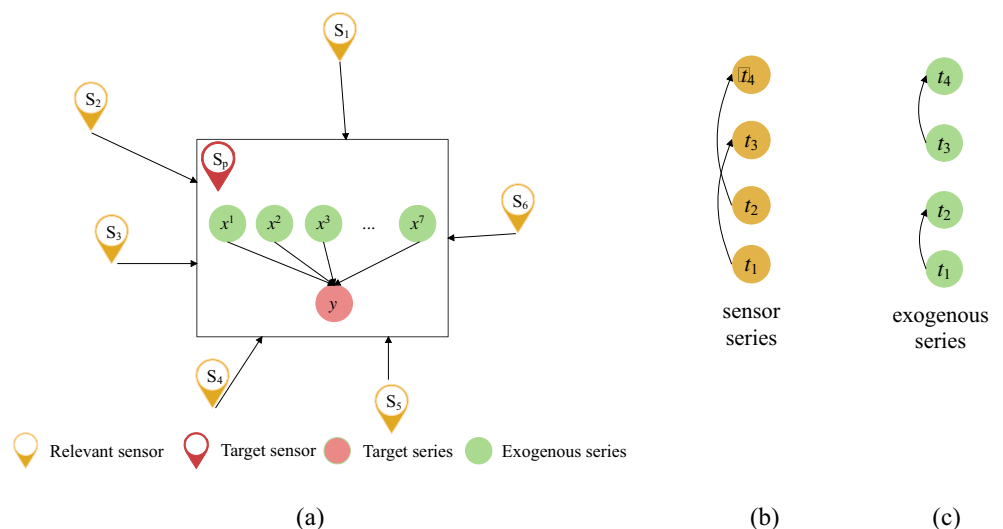
linear graph attention and temporal attraction force (J-NGT) to achieve geo-sensory time series prediction. Our model can simultaneously capture inter-sensor and intra-sensor spatial-temporal correlations by two graph attention mechanisms.

The main contributions of our study are summarized as follows:

(1) We propose a joint network containing two graph attention mechanisms, i.e., a non-linear graph attention mechanism and a temporal attraction force mechanism.

(2) We design a non-linear graph attention mechanism to obtain the inter- and intra-sensor spatial correlations. In the non-linear graph attention mechanism, first, we specifically employ IMV-LSTM, which is a tensor LSTM, to transform the input features into higher-level features to gain sufficient expressive power and then calculate the weights between nodes.

(3) Inspired by the Law of universal gravitation, we propose a temporal attraction force mechanism to sufficiently capture inter- and intra-sensor temporal correlations.

(4) We conduct extensive experiments on three real-world datasets to evaluate our model. The results show that our model outperforms comparison models in previous works. The geo-sensory time series data are split into multivariate time series and sensor series, and we perform our model over the sub-datasets. The results show that considering both inter- and intra-sensor spatial-temporal correlations can enhance prediction accuracy. We replace each component with state-of-the-art methods in our model framework to demonstrate the effectiveness of different components. The result suggests that our proposed components play a role in improving prediction accuracy.

The remainder of this paper is organized as follows. We provide a literature review on time series prediction methods and graph neural networks for geo-sensory time series prediction in Section 2. We define the notations and problem formulation in Section 3. We describe the non-linear graph attention mechanism and temporal attraction force mechanism to capture spatial-temporal correlations in Section 4. We design experiments to test the validity of our model in different fields and analyse the experimental results in Section 5. We summarize our work and future work in Section 6.

2 Related work

2.1 Geo-sensory time series prediction

The existing time series prediction methods can be divided into statistical models, machine learning models, and deep learning models. Statistical models are the classical model,

employed to predict stable and autocorrelated time series data. Although autoregressive (AR) [16], moving average (MA) [17], the autoregressive moving average (ARMA) [18], and the autoregressive integrated moving average (ARIMA) [19] have advantages in dealing with the univariate time series, they cannot model the dynamic spatial-temporal correlations of multivariate time series. The vector autoregressive (VAR) [20] model considers the relationships between stable and autocorrelated time series, and the information is limited.

Machine learning models have advantages in dealing with nonlinear and non-stationary data. Among them, artificial neural networks (ANN) [21], Gaussian process regression [22], support vector regression [23], and ensemble learning [24] have achieved respectable results in small numbers of multivariate time series prediction tasks. For example, Zhang et al. [25] presented the least squares support vector regression to achieve stock index and bond index prediction. Wang et al. [26] introduced sparse Gaussian conditional random fields for multivariate time series prediction. Although statistical models and machine learning models are widely used in time series forecasting, they do not scale well to geo-sensory time series because of multi-scale spatial-temporal correlations. [27]

Deep learning models have proven to be reliable for time series prediction. The recurrent neural network (RNN) with the capacity to gain short-term dependencies can perform time series prediction tasks. Due to the vanishing gradient, it is becoming difficult for standard RNN to gain long-term dependencies. To overcome the vanishing gradient, LSTM [28] and GRU [29], which are successful variants of RNN, were constructed to learn long-term dependencies. Recently, some advanced RNN variants were proposed. For instance, Feng et al. [30] introduced the clockwork RNN, which runs the hidden layer at different clock speeds to solve the long-term dependency problem. Zhang et al. [31] modified the GRU architectures, in which gates explicitly regulate two distinct types of memories to predict medical records and multi-frequency phonetic time series. Ma et al. [32] designed a temporal pyramid RNN to gain long-term and multi-scale temporal dependencies. The above RNN methods focus on capturing long-term dependencies; hence, they cannot fully exploit spatial relationships between variables. Therefore, the above existing works were far from satisfactory in geo-sensory time series prediction.

Modelling spatial-temporal correlation is the key to achieving a better prediction performance for geo-sensory time series. Ge et al. [10] calculated the sensor's similarity matrix and selected k similar sensors, which were combined with exogenous series, to extract spatial features. Liang et al. [15] introduced a multi-level attention mechanism to gain local and global spatial-temporal correlations. Although the above studies consider the inter- and intra-sensor spatial correlations, which were most relevant to ours, they ignored the unique

temporal characteristics of geo-sensory time series, e.g., sensors with geospatial correlation vary over time. In contrast, our model simultaneously models both inter- and intra-sensor spatial-temporal correlations to distinguish the contribution of the relevant sensor series and exogenous series in the spatial and temporal dimensions.

2.2 Graph neural network

Recently, graph neural networks (GNNs) have become popular due to their success in graph structure data. Many studies formulate geo-sensory time series on graphs to utilize spatial information fully. In particular, graph convolutional networks (GCNs) and graph attention mechanisms have become widespread in practice. The idea of existing studies that designed graph convolutional networks is to select the neighboruhood of sensors to capture the spatial correlations. Yu et al. [33] introduced a spatial-temporal GCN to gain bidirectional spatial-temporal dependencies from the neighbourhood of central nodes for traffic forecasting. Wang et al. [34] designed a GCN to learn the topological structure of sensor networks to capture spatial correlations for traffic safety prediction. Song et al. [35] developed a spatial-temporal synchronous mechanism to obtain localized spatial-temporal correlations for traffic flow prediction. These dynamic spatial-temporal correlations are localized due to the restriction of the range of neighbourhoods. To address these issues, Wang et al. [36] introduced a geographical spatial convolution to obtain complex spatial relationships among regions for traffic accident risk forecasting.

The graph attention mechanism is a novel graph neural network architecture for node classification of graph data [37]. The goal of the graph attention mechanism is to judge the relationships between nodes. Kong et al. [7] designed the graph talking-heads attention layer to capture the highly dynamic relationships between nodes in the traffic network. Lu et al. [12] explored multi-layer graph spatial attention networks to capture the dependencies between inbound and outbound flows in metro passenger flow. Shi et al. [38] introduced graph attention evolving networks that preserve similarities between nodes to evolve graph attention network weights across all temporal graphs. Han et al. [39] employed a graph attention mechanism to calculate the weight between nodes for representing the temporal dependence. In our work, we extend the graph attention network with IMV-LSTM to capture inter- and intra-sensor spatial correlations. IMV-LSTM transforms the individual variables into higher-level features, which can obtain sufficient expressive power.

2.3 Law of universal gravitation

Chi et al. [40] introduced the Law of universal gravitation to calculate the attraction force between nodes, representing the similarity between links for link prediction and improved prediction accuracy. Motivated by the above study, considering that the temporal correlations between different time intervals can be expressed as the gravitational force between nodes, we propose a temporal attraction force mechanism to calculate inter- and intra-sensor temporal correlations.

3 Preliminaries

Assume there are M sensors, each of which generates N kinds of time series. We specify one sensor as the target sensor for making predictions, while other sensors are used as relevant sensors. The target sensor generates N-1 kinds of exogenous series and one target series. We first construct two types of directed graphs to describe the spatial relationships of the time series. Both take the target series as the central node and the other series as neighbouring nodes. One graph structure indicates the spatial correlations of the target (sensor) series and relevant sensors, and the numbers of nodes are M, as depicted in Fig. 2a. We employ $\mathbf{Y} = (\mathbf{y}^1, \mathbf{y}^2, ..., \mathbf{y}^{M-1}) = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T) \in \mathbb{R}^{(M-1) \times T}$ to denote relevant sensor series, $\mathbf{y}_t = (y_t^1, y_t^2, ..., y_t^{M-1}) \in \mathbb{R}^{(M-1)}$ to denote the M-1 relevant sensor time series at time t, and $y_t^i \in \mathbb{R}$ to represent the node feature. Let $\mathbf{y}^p = (y_1^p, ..., y_T^p)$ with $y_t^p \in \mathbb{R}$ represent the target series of target sensors during the past T time points.

Another graph indicates the spatial correlations of the target series and exogenous series, as depicted in Fig. 2b. Among N kinds of time series of the target sensor, one time series is the target series, and the other time series, i.e., $\mathbf{x}_t = (x_t^1, x_t^2, ..., x_t^{N-1}) \in \mathbb{R}^{N-1}$, is the ($N$-1) exogenous series at time t, and $x_t^j \in \mathbb{R}$ represents the feature of a node. We employ $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T) = (\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^{N-1}) \in \mathbb{R}^{(N-1) \times T}$ as the exogenous series of window sizes T.

Given the previous readings of all sensors and the exogenous series, the model aims to predict the current value of the target sensor, denoted as

$$\widehat{y}_{T+1}^p = \mathcal{F}(\mathbf{X}, \mathbf{Y}, \mathbf{y}^p) \tag{1}$$

4 Model

Two-stage attention-based encoder-decoder networks are currently the most popular method for time series prediction. The encoder with spatial attention selects the relevant features, while the decoder with temporal attention captures the long-term dependencies. We propose a joint network of non-linear graph attention and temporal attraction force for geo-sensory time series prediction. First, we propose a non-linear graph attention mechanism to capture the inter-sensor spatial
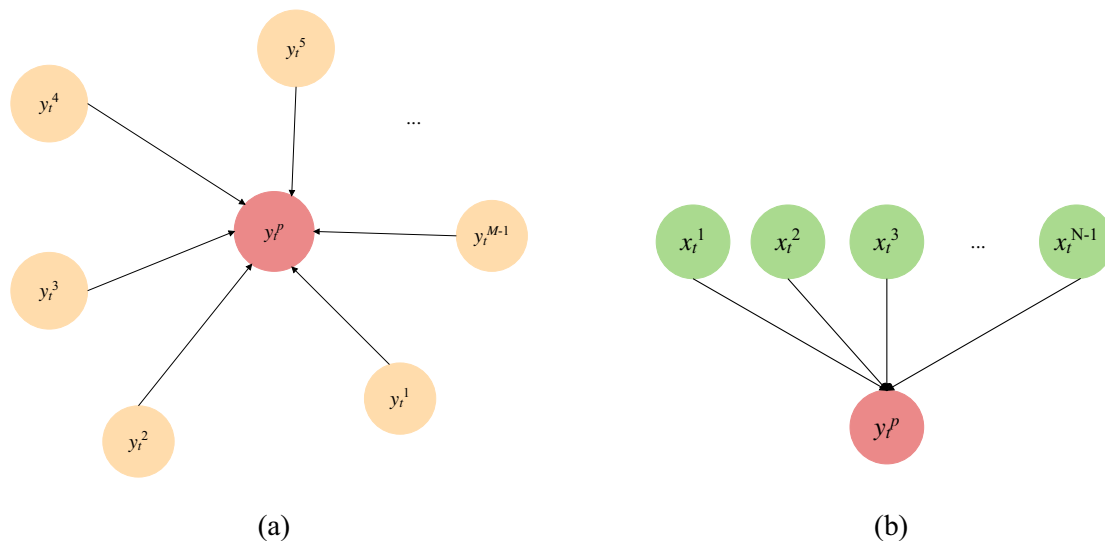
Fig. 2 Graph structure. Nodes represent the target series of sensors(orange circles), exogenous series(green circles), and target series(pink circles). **a** Graph structure between sensors, **b** Graph structure between target series and exogenous series

correlations for multiple sensors and the intra-sensor spatial correlations for a single sensor, as depicted in Fig. 3a. Then, we design a temporal attraction force to capture the inter- and intra-sensor temporal correlations between the current value and the previous values, as depicted in Fig. 3b.

4.1 IMV-LSTM

IMV-LSTM, which is a standard LSTM, uses hidden state tensors to update gate control units and memory cells. Hence IMV-LSTM can directly encode the individual variables into hidden states to reduce the time complexity. For example, given a two-variable input sequence, IMV-LSTM can encode the input feature into the hidden matrix of size 4×2, i.e., 4-dimensional hidden state per variable. IMV-LSTM, which is a standard LSTM neural network, contains an input gate $\widetilde{\mathbf{i}}_t$, forget gate $\widetilde{\mathbf{f}}_t$, output gates $\widetilde{\mathbf{o}}_t$, and memory cells $\mathbf{s}_t$. For multiple variables, the iterative update process of an MV-LSTM unit is as follows:

$$\left.\begin{array}{c}\widetilde{\mathbf{i}}_t \\ \widetilde{\mathbf{f}}_t \\ \widetilde{\mathbf{o}}_t\end{array}\right\} = \sigma\big(\mathbf{W} \circledast \mathbf{D}_{t-1} + \mathbf{U} \circledast \mathbf{y}_t] + \mathbf{b}\big) \tag{2}$$

$$\mathbf{j}_t = \tanh\big(\mathbf{W}_j \circledast \mathbf{D}_{t-1} + \mathbf{U}_j \circledast \mathbf{y}_t + \mathbf{b}_j\big) \tag{3}$$

$$\mathbf{s}_t = \widetilde{\mathbf{f}}_t \odot \widetilde{\mathbf{c}}_{t-1} + \widetilde{\mathbf{i}}_t \odot \mathbf{j}_t \tag{4}$$

$$\mathbf{D}_t = \widetilde{\mathbf{o}}_t \odot \tanh(\mathbf{s}_t) \tag{5}$$

where $\mathbf{D}_{t-1} \in \mathbb{R}^{(M-1) * m}$ and $\mathbf{y}_t \in \mathbb{R}^{M-1}$ are the previous hidden state and the current input, respectively. The cell update matrix $\mathbf{j}_t \in \mathbb{R}^{(M-1) * m}$ is used to update the memory cell $\mathbf{s}_t$. The transition tensor $\mathbf{W} \in \mathbb{R}^{(M-1) * m * m}$, $\mathbf{W}_j \in \mathbb{R}^{(M-1) *}$

$^{m * m}$, $\mathbf{U} \in \mathbb{R}^{(M-1) * m}$ and $\mathbf{U}_j \in \mathbb{R}^{(M-1) * m}$ are parameters to learn. $\circledast$ is a tensor-dot operation, which is the product of two tensors. σ and $\odot$ are a logistic sigmoid function and an element-wise multi-plication.
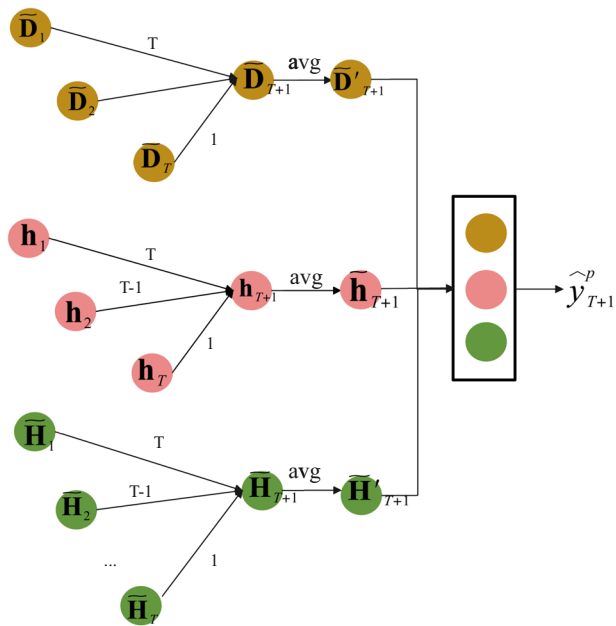
4.2 Non-linear graph attention

Geo-sensory time series contains two spatial correlations, namely, the inter-sensor spatial correlations for multiple sensors and the intra-sensor spatial correlations for a single sensor. A non-linear graph attention mechanism is developed to capture the two kinds of spatial correlations. Different from the graph attention network (GAT), the non-linear graph attention mechanism depends on knowing the graph structure up front. We formulate the Geo-sensory time series as two types of graph structures, as depicted in Fig. 2, where nodes represent the relevant sensor series (orange circles), exogenous series (green circles), and target series (pink circles). We take the target series as the central node and the other series as neighbourhood nodes.

4.2.1 Inter-sensor spatial correlation

Relevant sensors have decisive impacts on a target sensor. Hence, the non-linear graph attention mechanism aims to calculate the impacting weights between the target sensor and relevant sensors, namely, the inter-sensor spatial correlation, as depicted in Fig. 3a. The input of NGAT is a set of node features, $\mathbf{y} = \{y_t^1, y_t^2, \dots, y_t^{M-1}, y_t^P\}$, where $\mathbf{y}_t = \big(y_t^1, y_t^2, \dots, y_t^{M-1}\big)$, where $y_t^i \in \mathbb{R}$ represents the target series of relevant sensors, $y_t^p \in \mathbb{R}$ represents the target series of target sensors y_t^p, and M is the number of nodes.
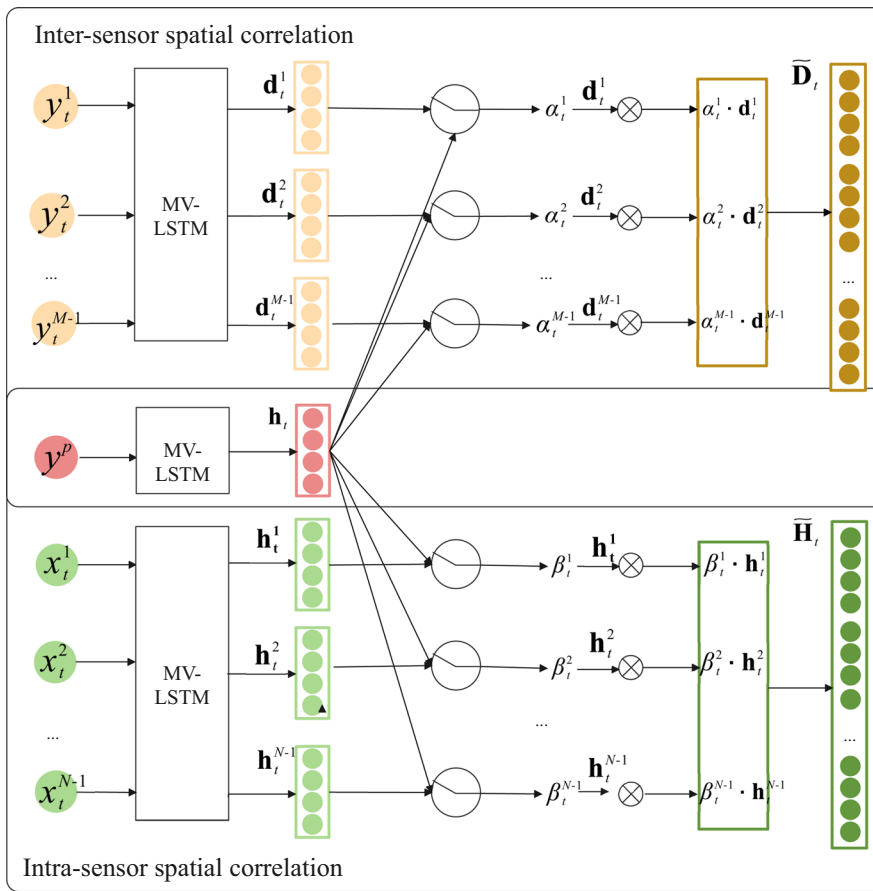
(a)

(b)

◀ **Fig. 3** Graphical illustration of the joint network of non-linear GAT and temporal attraction force. **a** Non-linear graph attention mechanism. The non-linear GAT employs IMV-LSTM to transform the input features into higher-level features. Green, pink and orange circles, such as input data x_t^1, y^p, and y_t^1, which are used as the input, represent a one-dimensional input feature. Blocks containing rectangles with circles inside represent higher-level features. **b** Temporal attraction force mechanism. The output of the non-linear graph attention mechanism, i.e., $\widetilde{\mathbf{D}}_t$, $\mathbf{h}_t$, and $\widetilde{\mathbf{H}}_t$, are used as the input to the temporal attraction force mechanism

The node feature is an individual variable, which does not have sufficient expressive power. In the original GAT, a learnable linear transformation is applied to transform the input features into higher-level features [37]. Each node is parametrized by a weight matrix, $\mathbf{W}y_t^i$, $\mathbf{W} \in \mathbb{R}^m$. However, the expressive power of linear transformation is limited. Therefore, we employ IMV-LSTM as transformation. We obtain specific transformation representations by IMV-LSTM:

$$\mathbf{D}_t = f_{imv}^{sen}(\mathbf{D}_{t-1}, \mathbf{y}_t) \tag{6}$$

$$\mathbf{h}_t = f_{imv}^{pre}(\mathbf{h}_{t-1}, y_t^p) \tag{7}$$

where $\mathbf{D}_t \in \mathbb{R}^{(M-1)*m}$ and $\mathbf{h}_t \in \mathbb{R}^m$ are the hidden state tensors of the relevant sensor series and target series, respectively. f_{imv}^{sen} and f_{imv}^{pre} are IMV-LSTM units that can be computed according to Eq. (2)–(5). IMV-LSTM uses tensor to represent higher-level features of sensor series at time t, such that each row vector of the hidden state tensor represents the higher-level feature of an individual variable:

$$\mathbf{D}_t = \left(\mathbf{d}_t^1, \mathbf{d}_t^2, ..., \mathbf{d}_t^{M-1}\right) \tag{8}$$

where the element $\mathbf{d}_t^i \in \mathbb{R}^m$ of $\mathbf{D}_t$ is the hidden state vector specific to relevant sensor i.

We exploit GAT to calculate the spatial correlations between the target series and relevant sensors. In other words, assigning different weights to different relevant sensors. Here, we calculate the weight coefficients between the target series and relevant sensor i. First, a weight vector is employed to parametrize, and then the LeakyReLU nonlinearity(with negative input slope $\alpha = 0.2$) is applied. The calculated correlation coefficient is expressed as:

$$e_t^i = \text{LeakyReLU}\left(a^\top\left[\mathbf{d}_t^i; \mathbf{h}_t\right]\right) \tag{9}$$

$$\alpha_t^i = \text{softmax}\left(e_t^i\right)$$
$$= \frac{\exp\left(\text{LeakyReLU}\left(a^\top\left[\mathbf{d}_t^i; \mathbf{h}_t\right]\right)\right)}{\sum_{\mathbf{d}_t^i \in \mathbf{D}_t}\exp\left(\text{LeakyReLU}\left(a^\top\left[\mathbf{d}_t^i; \mathbf{h}_t\right]\right)\right)} \tag{10}$$

where $\left[\mathbf{d}_t^i; \mathbf{h}_t\right]$ is a concatenation operation, and $*^\top$ represents transpose. The attention weight α_t^i represents the importance of the relevant sensor i to the target series, namely, the intersensor spatial correlations. A softmax function is applied to normalize them across all relevant sensors. Once obtained, the attention weights are assigned to different relevant sensors:

$$\widetilde{\mathbf{D}}_t = \left(\alpha_t^1\mathbf{d}_t^1, \alpha_t^2\mathbf{d}_t^2, ..., \alpha_t^{M-1}\mathbf{d}_t^{M-1}\right)$$
$$= \left(\widetilde{\mathbf{d}}_t^1, \widetilde{\mathbf{d}}_t^2, ..., \widetilde{\mathbf{d}}_t^{M-1}\right) \tag{11}$$

where $\widetilde{\mathbf{d}}_t^i$ is the new higher-level feature of relevant sensor i.

4.2.2 Intra-sensor spatial correlation

Inside the target sensor, there are complex correlations between the target series and exogenous series. For instance, an air quality sensor reports different time series, such as PM2.5, PM10, and CO. In the real world, PM2.5 concentration is affected by the concentration of PM10 and CO. To address this issue, we also apply the non-linear graph attention mechanism to calculate correlations between the target series and the exogenous series, namely, intra-sensor spatial correlation. Given the target series, $y_t^p \in \mathbb{R}$ represents the feature of the central node, as well as the exogenous series, $\mathbf{x}_t = \left[x_t^1, x_t^2, ..., x_t^{N-1}\right] \in \mathbb{R}^{N-1}$, where $x_t^j \in \mathbb{R}$ represents the feature of neighbourhood node j. We can construct a non-linear graph attention mechanism to calculate intra-sensor spatial correlations, as depicted in Fig. 3a. First, transforming the input features into higher-level features via IMV-LSTM,

$$\mathbf{H}_t = f_{imv}^{exo}(\mathbf{H}_{t-1}, \mathbf{x_t}) \tag{12}$$

$$\mathbf{H}_t = \left[\mathbf{h}_t^1, \mathbf{h}_t^2, ..., \mathbf{h}_t^{N+1}\right] \tag{13}$$

where $\mathbf{H}_t \in \mathbb{R}^{(N-1)*n}$ represents the hidden state tensors of the exogenous series and $\mathbf{h}_t^j \in \mathbb{R}^n$ is the higher-level features of the j-th node. f_{imv}^{exo} is an IMV-LSTM unit that can be computed according to Eq. (2)–(5) with the newly input $\mathbf{H}_{t-1}$ and $\mathbf{x_t}$. The correlation coefficients between the target series and exogenous series are calculated as follows:

$$l_t^j = \text{LeakyReLU}\left(a^\top\left[\mathbf{h}_t^j; \mathbf{h}_t\right]\right) \tag{14}$$

$$\beta_t^j = \text{softmax}\left(l_t^j\right)$$
$$= \frac{\exp\left(\text{LeakyReLU}\left(a^\top\left[\mathbf{h}_t^j; \mathbf{h}_t\right]\right)\right)}{\sum_{\mathbf{h}_t^j \in \mathbf{D}_t}\exp\left(\text{LeakyReLU}\left(a^\top\left[\mathbf{h}_t^j; \mathbf{h}_t\right]\right)\right)} \tag{15}$$

where the attention weight β_t^j is used to measure the importance of the j-th exogenous series at time t. $\mathbf{h}_t$ has been computed according to Eq. (7). With these attention weights, the new features are produced as follows:

$$\widetilde{\mathbf{H}}_t = \left[\beta_t^1\mathbf{h}_t^1, \beta_t^2\mathbf{h}_t^2, ..., \beta_t^{N-1}\mathbf{h}_t^{N-1}\right]$$
$$= \left[\widetilde{\mathbf{h}}_t^1, \widetilde{\mathbf{h}}_t^2, ..., \widetilde{\mathbf{h}}_t^{N-1}\right] \tag{16}$$

4.3 Temporal attraction force mechanism

Most research on the temporal dependence of geo-sensory time series has focused on blindly blending the information of relevant sensors and exogenous series and then calculating the temporal dependence by a temporal attention mechanism. Hence, these studies rarely distinguish the contribution of the historical target series, exogenous series, and relevant sensor series into the predictive value.

In the temporal dimension, there may be temporal correlations between pairs of time points, which can be regarded as the diverse attractive factors between time points. Therefore, we assume that a dynamic attraction force always exists between any pair of time points. To facilitate the understanding of attraction force, we analyse it as follows: In the universe, planets revolve around a star in a galaxy. The universal gravitation between the planet and the star prevents the planet from flying away from the star. Similar to the universe, the current time point gains information from the historical time points, which can be taken as the historical time points revolving around the current time point. Hence, the gravitational force between the current time point and the historical time points is named the temporal correlation. The temporal correlations are used to measure the importance of historical time points to the current time point.

Given the mass of the planet and the star, i.e., m_1 and m_2, universal gravitation is presented as follows:

$$F = G \frac{m_1 m_2}{r^2} \tag{17}$$

where r denotes the distance between the planet and the star, and G is the gravitational constant.

The gravitation formula was demonstrated by [40] to measure the correlations between nodes for link prediction in the social network. The social network is a typical graph-structured data. People have a real link if the attraction force reaches a specified threshold. We can map time series to a graph structure in the temporal dimension, where a time point is regarded as a node, and the edge weight is the temporal correlation (attraction force). Inspired by the Law of universal gravitation, we propose a modified gravitation formula, namely the temporal attraction force mechanism, for measuring the temporal correlation. In our work, the feature of time point is defined as its mass, which is matrix rather than constant, e.g., $\widetilde{\mathbf{D}}_t$ and $\widetilde{\mathbf{H}}_t$. The distance is the difference between two time points, e.g., $r = t - i$, t is the current time point, and i is the historical time point. In addition, the gravitational constant G is replaced by a learnable weight matrix.

From another point of view, we score the temporal correlation between the current and historical time points by performing a learnable linear transformation (i.e., G) on the fused features, which is further scaled by the squared distance of the two-time points. Here, the features of two time points are fused by element-wise multiplication. The purpose of scaling is to improve sensitivity to time intervals. Moreover, the softmax function is employed to gain the weights.

The temporal attraction force mechanism takes the output of the non-linear graph attention mechanism as input, as depicted in Fig. 3b. The inter-sensor temporal correlation for relevant sensors is given by

$$p_{ti} = \mathbf{G}_p \frac{\widetilde{\mathbf{D}}_t * \widetilde{\mathbf{D}}_i}{(t-i)^2} \tag{18}$$

$$\alpha_{ti} = \frac{\exp(p_{ti})}{\sum_{k=1}^{t-1} \exp(p_{tk})} \tag{19}$$

where $\widetilde{\mathbf{D}}_t$ is the feature of relevant sensors at time point t, $\widetilde{\mathbf{D}}_i$ is the feature of relevant sensors at time point i ($t > i$), and $\mathbf{G}_p$ is a parameter to learn. $(t - i)$ is the temporal distance, which is the difference between time point t and time point i. * is element-wise multiplication. Then, a softmax function is applied to p_{ti} to make correlations easily comparable across all time points. The correlation coefficient α_{ti} represents the importance of the relevant sensor features at time point i for the relevant sensor features at time t. Once obtained, the correlation coefficients are utilized to compute a weighted sum of the features corresponding to them as the new temporal features at time t:

$$\widetilde{\mathbf{D}}_t{}' = \sum_{i=1}^{t-1} \alpha_{ti} \widetilde{\mathbf{D}}_i \tag{20}$$

Similarly, the intra-sensor temporal correlation for two time points is given by

$$q_{ti} = \mathbf{G}_q \frac{\widetilde{\mathbf{H}}_t * \widetilde{\mathbf{H}}_i}{(t-i)^2} \tag{21}$$

$$\beta_{ti} = \frac{\exp(q_{ti})}{\sum_{k=1}^{t-1} \exp(q_{tk})} \tag{22}$$

where $\widetilde{\mathbf{H}}_t$ is the feature of exogenous series at time point t, $\widetilde{\mathbf{H}}_i$ is the feature of relevant sensors at time point i ($t > i$), and $\mathbf{G}_q$ is a parameter to learn. $(t - i)$ is the distance, which is the difference between time point t and time point i. * is element-wise multiplication. We normalize these weight coefficients by a softmax function. With these weight coefficients, the new temporal feature of the exogenous series is computed as

$$\widetilde{\mathbf{H}}_t{}' = \frac{1}{t-1} \sum_{i=1}^{t-1} \beta_{ti} \widetilde{\mathbf{H}}_i \tag{23}$$

The univariate time series prediction models analyse the temporal dependence between the previous values and the current values of the prediction series. However, multivariate time series prediction models generally ignore the long-term dependencies of the target series. To ensure that the model no longer suffers from the loss of historical information, we

consider the temporal correlations of the target series. The temporal correlations can be obtained by the temporal attraction force mechanism as follows:

$$o_{ti} = \mathbf{G_o} \frac{\mathbf{h}_t * \mathbf{h}_i}{(t-i)^2} \tag{24}$$

$$\gamma_{ti} = \frac{\exp(o_{ti})}{\sum_{k=1}^{t-1} \exp(o_{tk})} \tag{25}$$

Then, the new temporal feature of the target series is defined as follows:

$$\widetilde{\mathbf{h}}_t^{'} = \frac{1}{t-1} \sum_{i=1}^{t-1} \gamma_{ti} \mathbf{h}_i \tag{26}$$

Considering all the temporal correlations, we briefly integrate the new temporal features of the inter-sensor, intra-sensor, and target series as follows:

$$\mathbf{C}_t = \left[\widetilde{\mathbf{D}}_t^{'} ; \widetilde{\mathbf{H}}_t^{'} ; \widetilde{\mathbf{h}}_t^{'}\right] \tag{27}$$

The temporal attraction force mechanism sufficiently selects the historical information of the target series, exogenous series, and relevant sensor series. It takes advantage of inter- and intra-sensor features to strengthen the temporal correlation. Finally, we use a linear transformation to make the final output as follows:

$$\hat{y}_{T+1}^{p} = \mathcal{F}(\mathbf{X}, \mathbf{Y}, \mathbf{y}^p) = \mathbf{v}_y^{\top}\left(\mathbf{W}_y \mathbf{C}_t + \mathbf{b}_w\right) + b_v \tag{28}$$

4.4 Complexity analyses

In this section, we analyse the complexity of our model. Our model has five parts: two non-linear graph attention mechanisms and three temporal attraction force mechanisms. For IMV-LSTM, the time complexity is $O(D^2/N + N*D)$, where D is the number of IMV-LSTM neurons and N is the input variable. The time complexity of a non-linear graph attention mechanism computing N nodes may be expressed as $O(D^2 + N*D)$. The time complexity for the temporal attraction force mechanism is $O(N*D^2)$. Therefore, the total time complexity is $O((N + M + 3)*D^2 + (N + M)*D)$. In the actual training process, we use a GPU to improve the training speed.

5 Experiment

5.1 Dataset description

We utilize three real-world geo-sensory time series datasets to evaluate our model, including Beijing air quality datasets,

traffic flow (PEMS08) datasets, and Ireland Weather datasets, as shown in Table 1.

1. Air quality dataset[1]

This dataset collected hourly the concentration of many different pollutants (i.e., PM2.5, PM10, SO2 NO2, CO, O3) as well as some meteorological readings (i.e., temperature, pressure, dew point temperature, precipitation, wind direction, wind speed) from 12 nationally controlled air quality monitoring sites. The time period is from March 1st, 2013, to February 28th, 2017. Among them, we employ the Aotizhongxin station as the target sensor and the others as relevant sensors. Generally, PM2.5 is the primary pollutant of air quality; thus, we take it as the target series and the concentration of other pollutants as the exogenous series.

2. PEMS08 dataset[2]

This dataset consists of the data of 170 sensors collected from the Caltrans Performance Measurement System, which is aggregated into 5-minute windows from the raw data [41]. The dataset ranges from July 1, 2016, to August 31, 2016. We choose a station as the target sensor and the others as relevant sensors. We set the traffic flow as the target series and other key attributes of traffic observations (i.e., occupy and speed) as exogenous series. Constrained by our experimental environment, we decided to simplify and select 20 sensors to verify our model.

3. Weather dataset[3]

The dataset records hourly weather data from 23 Met Éireann weather stations in Ireland. The time period is from January 1, 2018, to February 1, 2022. We take the temperature as the target series and choose 8 relevant features as exogenous series. The Cork station is set as the target sensor, and the other 22 stations are set as relevant sensors.

In fact, there are missing values in all datasets due to sensor power outages or communication errors. We employ linear interpolation to fill the missing values. We partition the datasets into the training, validation, and test sets by a ratio of 6:2:2.

5.2 Methods for comparison

We select statistical models, machine learning models, and state-of-the-art deep learning models as comparison models. The modes are introduced as follows:

---

[1] https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data
[2] http://pems.dot.ca.gov
[3] http://www.met.ie

Table 1 Details of the datasets

Dataset	Air Quality	PEMS08	Weather
Target series	PM2.5	flow	temperature
Sensors	12	20	23
Exogenous series	11	2	8
Time intervals	1 hour	5 minutes	1 hour
Time spans	3/1/2013–2/28/2017	7/1/2016–8/31/2016	1/1/2018–2/1/2022
Number of instances	35,065	17,857	35,810

ARIMA [42] ARIMA is a typical statistical model for univariate time series prediction. The ARIMA model converts nonstationary time series to stationary data by utilizing difference processing.

SVR [43] SVR is an application of support vector machines to time series prediction problems. A significant advantage of SVR is that it can deal with small numbers of high-dimensional datasets.

DA-RNN [5] The attention-based encoder-decoder network for time series prediction employs an input attention mechanism to obtain spatial correlations and temporal attention to capture temporal dependencies.

DSTP [44] The model employs a two-phase attention mechanism to strengthen the spatial correlations and temporal attention mechanism to capture temporal dependencies for long-term and multivariate time series prediction.

hDS-RNN [4] The model develops a hybrid spatial-temporal attention mechanism, which can enhance spatial-temporal correlation learning.

DAQFF [45] DAQFF is a hybrid deep learning model that employs one-dimensional CNNs and bidirectional LSTM to extract trend features and possible spatial correlation features of multiple stations.

GeoMAN [15] GeoMAN is a multi-level attention-based recurrent neural network for geo-sensory time series prediction. The model considers local spatial correlations between target series and exogenous series as well as global spatial correlations between sensors.

IMV-LSTM [14] IMV-LSTM is an extension of LSTM. The model utilizes tensorized hidden states and an associated updating scheme to update gate control units and memory cells.

5.3 Experimental settings

We execute a grid search strategy and choose the best values for three types of key hyperparameters in J-NGT. For the number of time windows T, we set $T \in \{5, 10, 15, 20, 25\}$. To determine the dimensions of higher-level features for the relevant sensor series and the exogenous series, we set $m = n \in \{16, 32, 64, 128\}$. For all RNN-based models (i.e., DA-RNN, DSTP, hDS-RNN, DAQFF, GeoMAN, and IMV-LSTM), we similarly adopt a grid search strategy to determine the best performance of these models for a fair comparison. We take T as the window and m as the size of the hidden states for RNN. For DAQFF, three convolution layers have different filter sizes, and we set them to 64, 32, and 16. The models are trained for 10 epochs with a batch size of 128. The initial learning rate is set as 0.001 and decays by 10% every 3 epochs.

To assess the performances of J-NGT and comparison models, we adopt three evaluation metrics: mean absolute error (MAE), root mean square error (RMSE), and R squared (R2). MAE and RMSE are employed to measure the error between the predicted and observed values. R Squared (R2) is chosen as the indicator to measure the fitting effect of the model. The range of R2 is determined as [0,1]. If R2 is close to 1, it means that the prediction accuracy of our model is high. MAE, RMSE, and R2 are defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_t^i - \widehat{y}_t^i| \tag{29}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_t^i - \widehat{y}_t^i\right)^2} \tag{30}$$

$$\text{R2} = 1 - \frac{\sum_{i=1}^{N} |y_t^i - \widehat{y}_t^i|}{\sum_{i=1}^{N} |y_t^i - \overline{y}|} \tag{31}$$

where $\widehat{y}_t$ and y_t are the predicted value and observed value at time t, $\overline{y}$ is the average value of the observed values, and N represents the number of samples.

Table 2 Prediction results on geo-sensory time series datasets with sensor series and exogenous series

Datasets	Air Quality			PEMS08			Weather		
	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE	R2
ARIMA	83.5743	61.0411	0.0006	88.2807	69.7207	−0.0003	5.3406	4.3357	−0.1163
SVR	32.8259	25.8426	0.8378	41.3123	26.2529	0.9202	1.1951	0.9135	0.9458
DA-RNN	24.3446	14.0042	0.9152	30.3231	21.8391	0.9524	0.8455	0.6069	0.972
DSTP	29.4257	16.3356	0.8761	31.1639	22.6251	0.9497	1.0972	0.7921	0.9528
hDS-RNN	24.505	14.1457	0.914	30.0176	21.6926	0.9533	0.9717	0.6117	0.9768
DAQFF	24.3589	14.0216	0.9151	30.2502	22.0139	0.9526	0.8356	0.613	0.9726
Geo-MAN	16.0528	9.4563	0.9621	28.5449	21.1778	0.9572	0.6299	0.4596	0.9844
IMV-LSTM	16.9356	10.1891	0.9589	29.1881	21.2693	0.9559	0.567	0.4273	0.9874
J-NGT	**15.1621**	**8.5449**	**0.9671**	**27.9916**	**20.3754**	**0.9594**	**0.5492**	**0.3909**	**0.9881**

5.4 Comparison and result analysis

In this section, we give the experimental results on three real-world datasets, as shown in Tables 2, 3 and 4. The best results of each dataset are marked. In addition, we show the fitting results with bar charts in Fig. 4 to clearly observe their differences.

Table 2 displays the performances of each model on three datasets with sensor series and exogenous series. As seen in the table, J-NGT significantly outperforms all the comparison models on three evaluation metrics. In this paper, we discuss MAE. J-NGT has a relatively lower MAE from 3.7% to 90.1% than other models on the three datasets. In the table, it can be seen that the performance of ARIMA is the worst. The MAE value of J-NGT is 8.5449 on the air quality dataset, which is approximately 86.0% less than that of ARIMA(61.0411) since the ARIMA only model target series ignores the sensor series and exogenous series of geo-sensory time series. SVR takes into account all information of geo-sensory time series, so the MAE of SVR is significantly lower than that of ARIMA. Although SVR improves prediction accuracy to a certain extent, it requires considerable computational cost for large numbers of multivariate time series

prediction tasks. Because geo-sensory time series have long-term dependencies, J-NGT achieves better performance than SVR by considering a much longer dependency.

The attention-based RNNs, such as DA-RNN, DSTP, and hDS-RNN, employ various attention mechanisms to obtain spatial correlations and use temporal attention mechanisms to capture temporal correlations. Although these models outperform SVR and ARIMA, they mixed sensor and exogenous series to capture spatial correlations resulting in the MAE values of J-NGT being smaller than that of the attention-based RNNs. For example, J-NGT shows 39.0%, 47.7%, and 39.6% improvements in MAE compared to the above attention-based RNN models on the air quality dataset. The above methods focus on blending the information of relevant sensor series and exogenous series and hardly distinguish the contribution of the exogenous series and relevant sensor series into predictive value.

DAQFF employs one-dimensional convolutional neural networks to extract the local trend features and spatial correlation features of multiple stations, but implicit features do not contain geospatial spatial correlations. The results reveal that it and DA-RNN achieve comparable performance. GeoMAN outperforms the above models since it is capable of capturing

Table 3 Prediction results on geo-sensory time series datasets with sensor series

Datasets		SVR	LSTM	DA-RNN	DSTP	hDS-RNN	IMV-LSTM	J-NGT-sen
Air Quality	RMSE	27.0609	27.2278	25.0326	29.2611	24.5605	19.2578	**16.4409**
	MAE	21.4397	16.9189	14.6028	16.3656	13.8782	10.8971	**9.1654**
	R2	0.8898	0.8939	0.9103	0.8770	0.9136	**0.9661**	0.9613
PEMS08	RMSE	41.5968	32.7752	29.9048	31.7543	30.0894	**28.8341**	29.1482
	MAE	26.6663	24.4684	21.4637	23.2507	21.7637	**20.9297**	21.2057
	R2	0.9191	0.9432	0.9537	0.9478	0.9531	**0.9569**	0.9560
Weather	RMSE	1.1208	1.1667	0.8449	1.1424	0.8477	0.7160	**0.5845**
	MAE	0.8458	0.8336	0.6106	0.8115	0.6153	0.5292	**0.4229**
	R2	0.9523	0.9467	0.9720	0.9489	0.9718	0.9724	**0.9866**

Table 4 Prediction results on geo-sensory time series datasets with exogenous series

Datasets		SVR	LSTM	DA-RNN	DSTP	hDS-RNN	IMV-LSTM	J-NGT-exo
Air Quality	RMSE	45.8400	31.6367	28.5680	29.2838	28.2590	21.7608	**17.2126**
	MAE	37.5141	18.3932	15.9200	16.1676	15.8076	11.8384	**9.6426**
	R2	0.6838	0.8567	0.8832	0.8773	0.8857	0.9568	**0.9576**
PEMS08	RMSE	78.9279	32.7621	31.4969	31.7785	31.9767	**28.7016**	29.4861
	MAE	52.6300	24.3410	23.0749	23.3721	23.6615	**21.0869**	21.6857
	R2	0.7087	0.9433	0.9486	0.9477	0.9470	**0.9573**	0.9550
Weather	RMSE	1.7166	1.0439	1.0673	1.1384	1.0452	0.6336	**0.6315**
	MAE	1.3755	0.7531	0.7757	0.8161	0.7613	0.4615	**0.4486**
	R2	0.8882	0.9573	0.9554	0.9492	0.9572	0.9784	**0.9843**

spatial correlations of intra-sensor and spatial correlations of inter-sensor. Since J-NGT not only considers both spatial correlations but also captures the temporal correlations of inter- and intra-sensor, it achieves better performance than GeoMAN. For instance, J-NGT shows a 9.6% improvement in MAE compared to GeoMAN on the air quality dataset. The IMV-LSTM network outperforms the attention-based RNN model (i.e., DA-RNN, DSTP, and hDS-RNN) by 46.1% at most since the IMV-LSTM network models individual variables can capture different dynamics to make accurate predictions.

In summary, the J-NGT can outperform the comparison models. This illustrates that capturing the inter- and intra-sensor spatial-temporal correlations can provide more reliable input features for accurate prediction. The non-linear graph attention mechanism can calculate the correlation weights between the target series and other series(relevant sensor series and exogenous series). The temporal attraction force mechanism can sufficiently select the historical information of the target series, exogenous series, and relevant sensor series. For visual comparison, Fig. 4 provides the R2 of all models. J-NGT achieves the best fitting effect across the three datasets.

5.5 Evaluation of the sub-dataset

To verify the importance of both the inter- and intra-sensor spatial-temporal correlations, we compare J-NGT with comparison models on the sub-datasets. We divide our dataset into two sub-datasets, one containing sensor series and target series and the other containing exogenous series and target series. J-NGT can be separated into two parts: J-NGT-sen, which captures inter-sensor spatial-temporal correlations for the sensor series dataset, and J-NGT-exo, which captures intra-sensor spatial-temporal correlations for the exogenous series dataset. Tables 3 and 4 compare the performances of the submodule and comparison models on the six sub-datasets. We do not compare the experimental results of ARIMA, DAQFF, and GeoMAN on sub-datasets since ARIMA only considers the target series, and DAQFF, as well as GeoMAN, are introduced to predict geo-sensory time series.

In Tables 3 and 4, we discover that J-NGT-sen and J-NGT-exo exceed almost all comparison models on both the sub-datasets of the Air Quality dataset and Weather dataset. For the sub-datasets of the PEMS08 dataset, J-NGT consistently achieves the best performance compared with SVR, LSTM,
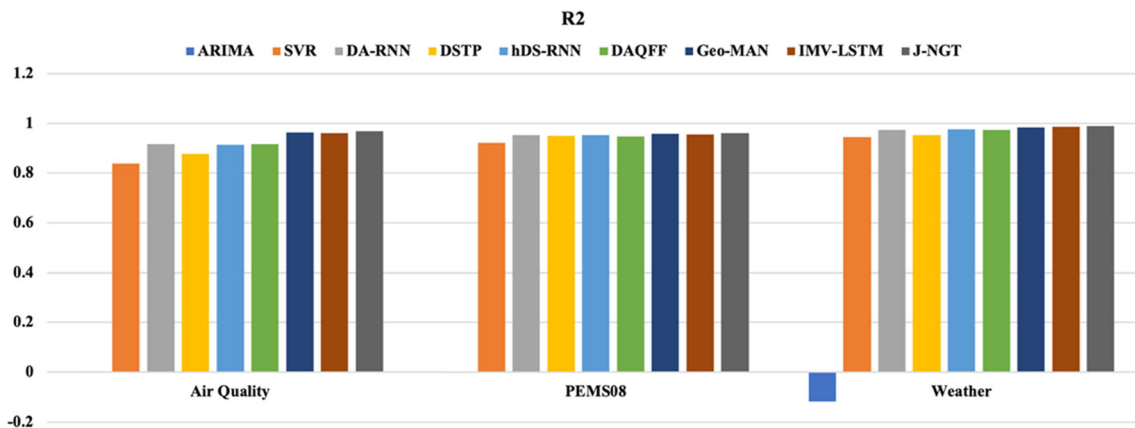
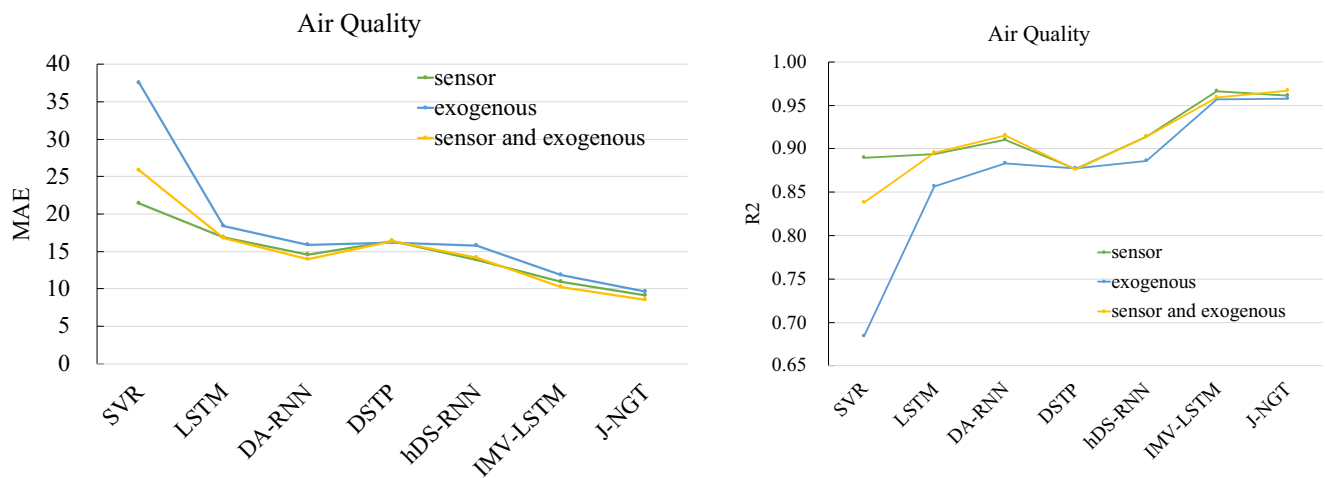**Fig. 4** Fitting results on three datasets

Fig. 5 Prediction results on the Air Quality dataset

DA-RNN, DSTP, and hDS-RNN. The IMV-LSTM model achieves competitive results on the sub-datasets of the PEMS08 dataset, outperforming J-NGT-sen and J-NGT-exo. However, J-NGT will not suffer from this issue since we focus on capturing both inter-sensor and intra-sensor spatial-temporal correlations.

To show the necessity of capturing both inter-sensor and intra-sensor spatial-temporal correlations, we compare J-NGT with comparison models on the sub-datasets and the full datasets. J-NGT has a relatively lower MAE from 3.9% to 12.8% than J-NGT-sen and J-NGT-exo on the three datasets. For visual comparison, we provide the experimental results on
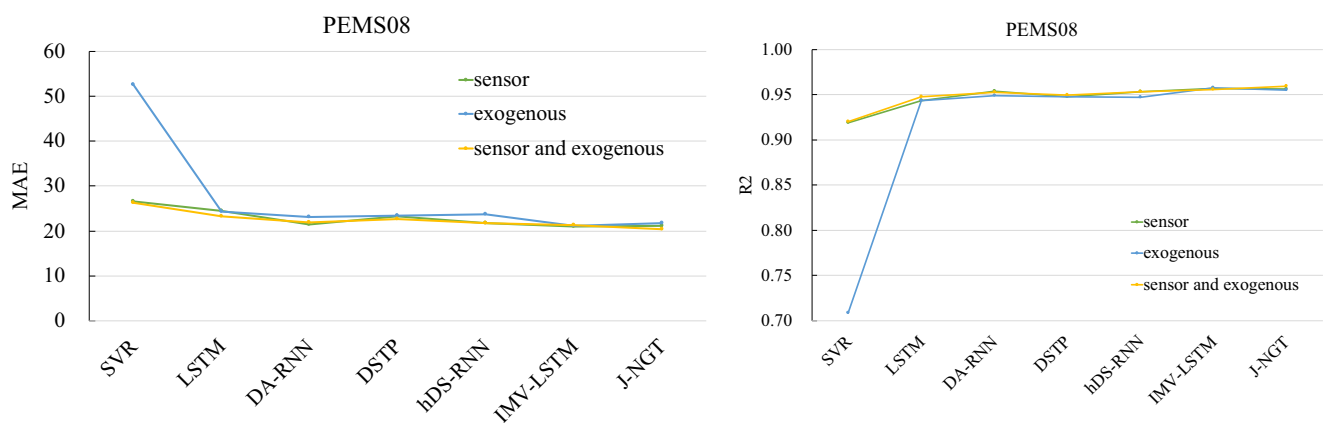
Fig. 6 Prediction results on the PEMS08 dataset
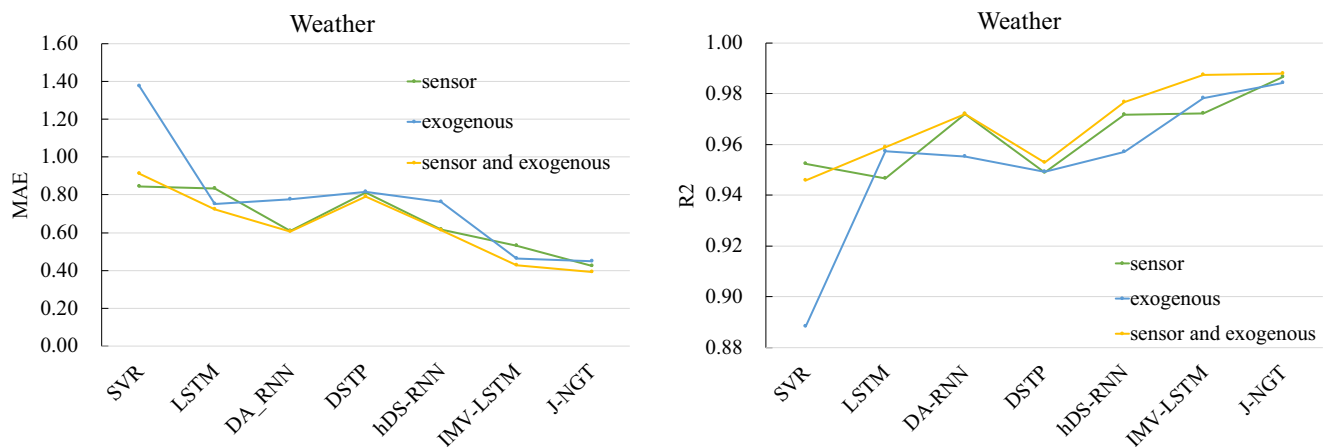
Fig. 7 Prediction results on the Weather dataset

Table 5 Prediction results of J-NGT and its variants

Datasets		J-NGT/linear transformation	J-NGT/ temporal attention	J-NGT
Air Quality	RMSE	16.7553	16.1929	**15.1621**
	MAE	9.2265	9.2141	**8.5449**
	R2	0.9598	0.9624	**0.9671**
PEMS08	RMSE	30.5208	29.4253	**27.9916**
	MAE	22.7668	21.5116	**20.3754**
	R2	0.9518	0.9552	**0.9594**
Weather	RMSE	0.6312	0.6030	**0.5492**
	MAE	0.4544	0.4356	**0.3909**
	R2	0.9844	0.9857	**0.9881**

Table 6 Two-tailed T-test table

degree	0.5	0.2	0.1	0.05	0.02
1	1	3.078	6.314	12.706	31.821
2	0.816	1.886	2.92	4.303	6.965
3	0.765	1.638	2.353	3.182	4.541
4	0.741	1.533	2.132	**2.776**	3.747
5	0.727	1.476	2.015	2.571	3.365
6	0.718	1.44	1.943	2.447	3.143
7	0.711	1.415	1.895	2.365	2.998
8	0.706	1.397	1.860	2.306	2.896

MAE and R2, as depicted in Figs. 5, 6 and 7. From the data in the figures, it is apparent that J-NGT obtains the highest performance on three geo-sensory time series datasets with sensor series and exogenous series.

5.6 Effects of different components

We conduct a detailed experiment to demonstrate the effectiveness of different components. Specifically, we replace each component with state-of-the-art models in our model framework. First, the J-NGT with replaced components is named as follows.

J-NGT/linear transformation We employ the learnable linear transformation, i.e., a weight matrix, $\mathbf{W} \in \mathbb{R}^F$, instead of the IMV-LSTM of J-NGT.

J-NGT/temporal attention We replace the temporal attraction force mechanism with the temporal attention mechanism. The temporal attention mechanism is commonly utilized to capture the long-term dependencies and has demonstrated outstanding performance.

Table 5 presents the prediction result of J-NGT and its variants. We highlight several observations from these results:

(1) The best results on all datasets are obtained with a non-linear graph attention mechanism and temporal attraction force mechanism.
(2) Replacing the non-linear graph attention mechanism from the full model leads to a degradation of model performance on all datasets. The result suggests that non-linear graph attention mechanism components play a role in improving prediction accuracy.
(3) The temporal attraction force mechanism has not been used to obtain temporal correlations. To verify its

validity, we replace it with a temporal attention mechanism. As Table 5 shows, the performance of J-NGT/temporal attention drops on all datasets. The result shows that the temporal attraction force mechanism is effective in capturing temporal correlations.

Taken together, these results suggest that J-NGT fully captures both inter- and intra-sensor spatial-temporal correlations and still achieves the best prediction performance.

5.7 Statistical test

To accurately evaluate the performance of J-NGT, we perform the statistical test in the MAE of three datasets. Based on the previous work that uses a Two-tailed T-test [46], we mainly discuss it. We set the significance degree α as 0.05. When the value of the two-tailed T-test is larger than the critical value of the table lookup, i.e., $p \leq 0.05$, the assumption that $m_i < m_0$ cannot be rejected suggests that there is a significant difference.

On the Air Quality dataset, the average MAE value of J-NGT is computed as $\mu = \sum_{i=1}^{k} m_i / k = 8.5449$, where m_i is the i-th MAE value, and the value of k is 5. The variance is computed as $\delta^2 = \sum_{i=1}^{k}(m_i - \mu)^2/(k-1) = 2.1 \times 10^{-4}$. The critical value is computed as $t = \sqrt{k}|\mu - m_0|/\delta = 3.608$, where $m_0 = 8.5683$ is the assumed maximum MAE value. In the same way, on PEMS08 and Weather, the critical value t is 4.124 ($m_0 = 20.4772$) and 3.350 ($m_0 = 0.4001$), respectively. We can observe that the critical value t of three datasets is all greater than that given 2.776 (that are marked in black in Table 6) from the two-tailed T-test table, which indicates that the MAE of J-NGT is smaller than the assumed maximum MAE value with a confidence degree ($1-\alpha$=0.95).

From the above statistical tests, it can be seen that the performance of J-NGT on the three datasets is significantly better than that of the comparison models.

6 Conclusion

We propose a joint network of non-linear graph attention and temporal attraction force (J-NGT) for geo-sensory time series prediction. Specifically, we propose two graph attention mechanisms to strengthen both inter- and intra-sensor spatial-temporal correlations, which can improve prediction accuracy. To investigate the effectiveness of J-NGT, we construct three-part experiments:

(1) Compared with statistical models, machine learning models, and deep learning models on three real-world datasets, the experimental results show that J-NGT can outperform them. In particular, the comparison with the GeoMAN and DQAFF models suggests that J-NGT enhances the ability to obtain inter- and intra-sensor spatial-temporal correlations.

(2) The geo-sensory time series data are split into multivariate time series and sensor series, and we perform our model over the sub-datasets. The results show that it is necessary to simultaneously model both spatial-temporal correlations.

(3) The experiments replacing our proposed component from the full model validate the necessity of each component in J-NGT.

The advantages of J-NGT are summarized as follows:

(1) Considering the exogenous series and sensor series, it is beneficial to overcome the limitation of lacking spatial-temporal information.

(2) We propose the non-linear graph attention mechanism to learn the spatial correlations of inter- and intra- sensors. To capture the temporal correlations, we design the temporal attraction force mechanism to sufficiently select the historical information of the target series, exogenous series, and relevant sensor series. A joint network of non-linear graph attention and temporal attraction force can distinguish the contribution of exogenous series and relevant sensor series into predictive value.

There are several promising directions for future work. First, we will design a similarity measure algorithm to calculate the temporal distance between temporal points to replace the distance of the temporal attraction force mechanism. The temporal distance is a crucial hyper-parameter. As the temporal distance increases, the easier it is for the gradient explosion. How to calculate the temporal distance will be a significant problem. Second, we will employ a novel graph attention mechanism to recover consecutive missing values of the geo-sensory time series. Currently, we linearly fill the missing values. How to fill the consecutive missing values with contextual information is an important problem for geo-sensory time series prediction.

Declarations

References

1. Huang X, Tang J, Yang X, Xiong L (2022) A time-dependent attention convolutional LSTM method for traffic flow prediction. Appl Intell 52:17371–17386. https://doi.org/10.1007/s10489-022-03324-7

2. Zhu Q, Chen J, Zhu L, Duan X, Liu Y (2018) Wind speed prediction with spatio-temporal correlation: a deep learning approach. Energies 11(4):1–18. https://doi.org/10.3390/en11040705

3. Han J, Liu H, Zhu H, Xiong H, Dou D (2021) Joint Air Quality and Weather Predictions Based on Multi-Adversarial Spatiotemporal Networks. In: 35th AAAI Conf. Artif. Intell. AAAI 2021, vol. 5A, pp. 4081–4089

4. Ma Z, Liu S, Guo G, Yu X (2022) Hybrid attention networks for flow and pressure forecasting in water distribution systems. IEEE Geosci Remote Sens Lett 19:1–5. https://doi.org/10.1109/LGRS.2020.3030839

5. Qin Y, Song D, Cheng H, Cheng W, Jiang G, Cottrell GW (2017) A dual-stage attention-based recurrent neural network for time series prediction. In: IJCAI Int Jt Conf Artif Intell, vol. 0, pp. 2627–2633. https://doi.org/10.24963/ijcai.2017/366

6. Han Q, Lu D, Chen R (2021) Fine-grained air quality inference via multi-channel attention model. In: IJCAI Int Jt Conf Artif Intell, pp. 2512–2518. https://doi.org/10.24963/ijcai.2021/346

7. Kong X, Zhang J, Wei X, Xing W, Lu W (2022) Adaptive spatial-temporal graph attention networks for traffic flow forecasting. Appl Intell 52(4):4300–4316. https://doi.org/10.1007/s10489-021-02648-0

8. Wang L, Adiga A, Chen J, Sadilek A, Venkatramanan S, Marathe M (2022) CausalGNN: causal-based graph neural networks for Spatio-temporal epidemic forecasting. Proc AAAI Conf Artif Intell 36(11):12191–12199. https://doi.org/10.1609/aaai.v36i11.21479

9. Huang Y, Ying JJC, Tseng VS (2021) Spatio-attention embedded recurrent neural network for air quality prediction. Knowledge-Based Syst 233:107416. https://doi.org/10.1016/j.knosys.2021.107416

10. Ge L, Zhou A, Li H, Liu J (2019) Spatially fine-grained air quality prediction based on DBU-LSTM. In: ACM Int Conf Comput Front 2019, CF 2019 - Proc, pp. 202–205. https://doi.org/10.1145/3310273.3322829

11. Zheng H, Lin F, Feng X, Chen Y (2020) A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. IEEE Trans Intell Transp Syst 22(11):1–11

12. Lu Y, Ding H, Ji S, Sze NN, He Z (2021) Dual attentive graph neural network for metro passenger flow prediction. Neural Comput & Applic 0123456789:13417–13431. https://doi.org/10.1007/s00521-021-05966-z

13. Ali A, Zhu Y, Zakarya M (2021) Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks. Inf Sci (NY) 577:852–870. https://doi.org/10.1016/j.ins.2021.08.042

14. T. Guo, T. Lin, and N. Antulov-Fantulin (2019) Exploring interpretable LSTM neural networks over multi-variable data. In: 36th Int Conf Mach Learn ICML 2019, vol. 2019-June, pp. 4424–4440

15. Liang Y, Ke S, Zhang J, Yi X, Zheng Y (2018) Geoman: Multi-level attention networks for geo-sensory time series prediction. In: IJCAI Int Jt Conf Artif Intell, vol. 2018-July, pp. 3428–3434. https://doi.org/10.24963/ijcai.2018/476

16. Xu W, Peng H, Zeng X, Zhou F, Tian X, Peng X (2019) Deep belief network-based AR model for nonlinear time series forecasting. Appl Soft Comput J 77:605–621. https://doi.org/10.1016/j.asoc.2019.02.006

17. Yang J et al (2022) Crack classification of fiber-reinforced backfill based on Gaussian mixed moving average filtering method. Cem Concr Compos 134:104740. https://doi.org/10.1016/j.cemconcomp.2022.104740

18. Li Y, Lang J, Ji L, Zhong J, Wang Z, Guo Y, He S (2021) Weather forecasting using Ensemble of Spatial-Temporal Attention Network and Multi-Layer Perceptron. Asia-Pac J Atmos Sci 57(3):533–546. https://doi.org/10.1007/s13143-020-00212-3

19. Mahmoudi MR, Baroumand S (2022) Modeling the stochastic mechanism of sensor using a hybrid method based on seasonal autoregressive integrated moving average time series and generalized estimating equations. ISA Trans 125:300–305. https://doi.org/10.1016/j.isatra.2021.07.013

20. Guefano S, Tamba JG, Azong TEW, Monkam L (2021) Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models. Energy 214:118791. https://doi.org/10.1016/j.energy.2020.118791

21. Hamzaçebi C (2008) Improving artificial neural networks' performance in seasonal time series forecasting. Inf Sci (NY) 178(23):4550–4559. https://doi.org/10.1016/j.ins.2008.07.024

22. Rai A, Upadhyay SH (2018) An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence. Appl Soft Comput J 71:36–50. https://doi.org/10.1016/j.asoc.2018.06.038

23. Liu J, Vitelli V, Zio E, Seraoui R (2015) A novel dynamic-weighted probabilistic support vector regression-based Ensemble for Prognostics of time series data. IEEE Trans Reliab 64:1203–1213. https://doi.org/10.1109/TR.2015.2427156

24. Li Z, Wu D, Hu C, Terpenny J (2019) An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction. Reliab Eng Syst Saf 184:110–122. https://doi.org/10.1016/j.ress.2017.12.016

25. Zhang H-C, Wu Q, Li F-Y (2022) Application of online multitask learning based on least squares support vector regression in the financial market. Appl Soft Comput 121:108754. https://doi.org/10.1016/j.asoc.2022.108754

26. Wang X, Zhang M, Ren F (2018) Sparse Gaussian conditional random fields on top of recurrent neural networks. In: 32nd AAAI Conf Artif Intell AAAI 2018, pp. 4219–4226

27. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C (2020) Connecting the dots: multivariate time series forecasting with graph neural networks. In: Proc ACM SIGKDD Int Conf Knowl Discov Data Min, pp. 753–763. https://doi.org/10.1145/3394486.3403118

28. Li W, Wang X, Han H, Qiao J (2022) A PLS-based pruning algorithm for simplified long–short term memory neural network in time series prediction. Knowledge-Based Syst 254:109608. https://doi.org/10.1016/j.knosys.2022.109608

29. Ozdemir AC, Buluş K, Zor K et al (2022) Res Policy 78:102906. https://doi.org/10.1016/j.resourpol.2022.102906

30. Feng X, Chen J, Zhang Z, Miao S, Zhu Q (2021) State-of-charge estimation of lithium-ion battery based on clockwork recurrent neural network. Energy 236:121360. https://doi.org/10.1016/j.energy.2021.121360

31. Zhang Y, Peng N, Dai M, Zhang J, Wang H (2021) Memory-gated recurrent networks. Thirty-Fifth AAAI Conf Artif Intell 35(12):10956–10963

32. Ma Q, Lin Z, Chen E, Cottrell GW (2020) Temporal pyramid recurrent neural network. In: AAAI 2020 - 34th AAAI Conf Artif Intell, pp. 5061–5068. https://doi.org/10.1609/aaai.v34i04.5947

33. Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: IJCAI Int Jt Conf Artif Intell, vol. 2018-July, pp. 3634–3640. https://doi.org/10.24963/ijcai.2018/505

34. Wang J, Chen Q, Gong H (2020) STMAG: A spatial-temporal mixed attention graph-based convolution model for multi-data flow safety prediction. Inf Sci (NY) 525:16–36. https://doi.org/10.1016/j.ins.2020.03.040

35. Song C, Lin Y, Guo S, Wan H (2020) Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: AAAI 2020 - 34th AAAI Conf Artif Intell, pp. 914–921. https://doi.org/10.1609/aaai.v34i01.5438

36. Wang B, Lin Y, Guo S, Wan H (2021) GSNet: learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting. In: AAAI 2021

37. Veličković P, Casanova A, Liò P, Cucurull G, Romero A, Bengio Y (2018) Graph attention networks. In: 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf Track Proc, pp. 1–12

38. Shi M, Huang Y, Zhu X, Tang Y, Zhuang Y, Liu J (2021) GAEN: Graph attention evolving networks. In: IJCAI Int Jt Conf Artif Intell, pp. 1541–1547. https://doi.org/10.24963/ijcai.2021/213

39. Han S, Dong H, Teng X, Li X, Wang X (2021) Correlational graph attention-based Long Short-Term Memory network for multivariate time series prediction. Appl Soft Comput 106:107377. https://doi.org/10.1016/j.asoc.2021.107377

40. Chi K, Yin G, Dong Y, Dong H (2019) Link prediction in dynamic networks based on the attraction force between nodes. Knowledge-Based Syst 181:104792. https://doi.org/10.1016/j.knosys.2019.05.035

41. Li M, Zhu Z (2021) Spatial-temporal fusion graph neural networks for traffic flow forecasting. In: 35th AAAI Conf Artif Intell AAAI 2021, vol. 5A, pp. 4189–4196

42. Geurts M, Box GEP, Jenkins GM (1977) Time series analysis: forecasting and control. J Mark Res. https://doi.org/10.2307/3150485

43. Sun Y et al (2021) Hotspot temperature prediction of dry-type transformers based on particle filter optimization with support vector regression. Symmetry (Basel) 13(8):1320. https://doi.org/10.3390/sym13081320

44. Liu Y, Gong C, Yang L, Chen Y (2020) DSTP-RNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Syst Appl 143:113082. https://doi.org/10.1016/j.eswa.2019.113082

45. Du S, Li T, Yang Y, Horng SJ (2021) Deep air quality forecasting using hybrid deep learning framework. IEEE Trans Knowl Data Eng 33(6):2412–2424. https://doi.org/10.1109/TKDE.2019.2954510

46. Hu J, Zheng W (2020) Multistage attention network for multivariate time series prediction. Neurocomputing 383:122–137. https://doi.org/10.1016/j.neucom.2019.11.060

Hongbin Dong received the B.S. and M.S. degrees in computer science and technology from the Harbin Ship Engineering College, China, in 1986 and 1995, respectively, and the Ph.D. degree in computer science and information technology from Beijing Jiaotong University, Beijing, China. Since 2005, he has been a Professor with the Computer Science and Technology Department, Harbin Engineering University, China. His research interests include natural computation, multi-agent systems, machine learning, and data mining. He is a member of the Committee on Artificial Intelligence and Pattern Recognition of the Chinese Computer Society and the Committee on Granular Computing and Knowledge Discovery of the Chinese Academy of Artificial Intelligence, and a Senior Member of the Computer Society.

Jinwei Pang received the B.S. degree in computer science and technology from China University of Petroleum (East China), Shandong, China and the M.S. degree and Ph.D. degree in computer science and technology from Harbin Engineering University, Heilongjiang, China. Since 2020, she is a lecturer within in School of Computer and Control Engineering, Yantai University, Shandong, China. Her research interest includes the development of optimization algorithm and time series prediction techniques using neural network.

Shuang Han received the B.S degree in software engineering from Harbin Engineering University, Heilongjiang, in 2014. She is currently pursuing the Ph.D. degree in software engineering in Harbin Engineering University. Her research interests include machine learning, deep learning, time series forecasting.

Xiaodong Yu received the B.S. and M.S. degrees in computer science and technology from the Harbin Normal University, China, in 2002 and 2007, respectively, and the Ph.D. degree in computer science and information technology from Harbin Engineering University, China. Since 2021, he has been a Professor with the Computer Science and Technology Department, Harbin Normal University, China. His research interests include natural computation, machine learning, and data mining.