



## 融合双注意力机制的GNN多维时间序列预测

范航舟, 梅红岩, 赵勤, 张兴, 程耐

引用本文:

范航舟, 梅红岩, 赵勤, 张兴, 程耐. 融合双注意力机制的GNN多维时间序列预测[J]. *智能系统学报*, 2024, 19(5): 1277–1286.

FAN Hangzhou, MEI Hongyan, ZHAO Qin, et al. Multivariate time series forecasting with a graph neural network and dual attention mechanism[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(5): 1277–1286.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202305020>

## 您可能感兴趣的其他文章

### 用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction

智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

### 结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics

智能系统学报. 2021, 16(4): 729–738 <https://dx.doi.org/10.11992/tis.202008007>

### 面向推荐系统的分期序列自注意力网络

Recommendation system with long-term and short-term sequential self-attention network

智能系统学报. 2021, 16(2): 353–361 <https://dx.doi.org/10.11992/tis.202005028>

### 基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

### 层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

### 基于注意力融合的图片描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

DOI: 10.11992/tis.202305020

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240828.0932.006>

# 融合双注意力机制的 GNN 多维时间序列预测

范航舟, 梅红岩, 赵勤, 张兴, 程耐

(辽宁工业大学 电子与信息工程学院, 辽宁 锦州 121000)

**摘要:** 针对现有多维时间序列数据 (multivariate time series, MTS) 预测中变量间依赖关系捕获能力不足和时间序列数据多通道信息利用不充分的问题, 提出一种融合双注意力机制的多维时间序列预测模型 (feature fusion and dual attention mechanism based GNN, FFDA-GNN)。该模型将图神经网络与空间注意力机制融合, 用于增强多变量之间依赖关系捕获能力; 利用并行的多层膨胀卷积和通道注意力机制, 对时间序列数据进行多通道的特征提取, 实现对时间序列数据多通道信息的充分利用, 从而提升预测性能。在经济、电力、交通 3 个领域数据集上与基准模型进行对比实验, 该模型预测精度优于其他基准方法, 有良好的可行性。

**关键词:** 多维时序预测; 图神经网络; 注意力机制; 特征融合; 时间卷积网络; 深度学习; 卷积神经网络; 时空特征  
**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2024)05-1277-10

中文引用格式: 范航舟, 梅红岩, 赵勤, 等. 融合双注意力机制的 GNN 多维时间序列预测 [J]. 智能系统学报, 2024, 19(5): 1277-1286.

英文引用格式: FAN Hangzhou, MEI Hongyan, ZHAO Qin, et al. Multivariate time series forecasting with a graph neural network and dual attention mechanism[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1277-1286.

## Multivariate time series forecasting with a graph neural network and dual attention mechanism

FAN Hangzhou, MEI Hongyan, ZHAO Qin, ZHANG Xing, CHENG Nai

(School of Electronics &amp; Information Engineering, Liaoning University of Technology, Jinzhou 121000, China)

**Abstract:** To address the issues of insufficient capture of dependency relationships between variables and inadequate utilization of multichannel temporal information in multivariate time series, a forecasting model fused with a dual attention mechanism (FFDA-GNN) is proposed. This model integrates a graph neural network with a spatial attention mechanism to enhance the capture of dependencies among multiple variables. Additionally, parallel multilayer inflation convolution and channel attention mechanisms are used for multi-channel feature extraction from temporal data. This approach fully utilizes multi-channel information and enhances forecasting performance. Comparative experiments conducted on datasets from the fields of economics, electric power, and traffic show that the proposed model achieves better prediction accuracy than other benchmark methods and demonstrates superior feasibility.

**Keywords:** multivariate time series forecasting; graph neural network; attention mechanism; feature fusion; temporal convolutional network; deep learning; convolutional neural network; spatiotemporal features

时间序列数据在电力<sup>[1-2]</sup>、金融<sup>[3]</sup>、交通<sup>[4-5]</sup>等多个行业中有广泛应用。时间序列预测, 能够根据过去一组恒定间隔的数据序列, 通过挖掘其潜在的规律, 对未来的决策提供前沿性的指导意

见。但该任务中存在一些相互关联的变量, 使得模型不能对时间序列进行准确的预测, 因而这类预测任务也被称作多变量的时序预测任务。

解决多维时序预测问题的关键在于如何处理好多变量之间的依赖关系, 只有准确清晰地表达各个变量之间的相互影响, 才能实现好的预测效果。传统的时序预测的一些方法仅仅关注单变量

收稿日期: 2023-05-20. 网络出版日期: 2024-08-28.

基金项目: 辽宁省教育厅科学研究项目 (JZL202015404, LJKZ0625, JYTMS20230869).

通信作者: 梅红岩. E-mail: 715014795@qq.com.

对预测结果的影响。最常见的单变量平稳模型是自回归模型 (autoregression, AR)<sup>[6]</sup>、滑动平均模型 (moving average, MA)<sup>[7]</sup>、自回归滑动平均混合模型 (autoregressive moving average, ARMA)<sup>[8]</sup>。在非平稳时序预测中, 一个典型的模型是差分整合移动平均自回归模型 (autoregressive integrated moving average, ARIMA)<sup>[9]</sup>, 该模型可以通过差分实现非平稳的序列平稳化, 之后通过 ARMA 模型拟合。在多维的时序预测方法中, 高斯过程 (Gaussian process, GP)<sup>[10]</sup> 和支持向量机 (support vector machines, SVM)<sup>[11]</sup> 模型为典型的基线方法。此后又有学者针对 AR 模型对单变量时序数据进行预测的问题, 提出向量自回归模型 (vector auto regression, VAR)<sup>[12]</sup>, 广泛应用于多维时序预测中。传统的时序预测方法虽然取得不错效果, 但仍存在很多问题。在深度学习时序预测中, 循环神经网络 (recurrent neural network, RNN)<sup>[13]</sup> 广泛应用于多维时序预测中, 但是由于 RNN 存在梯度消失或梯度爆炸的问题, 学者研究出 RNN 的 2 种常用变体, 分别是长短期记忆网络 (long short-term memory networks, LSTM)<sup>[14]</sup> 和门控循环单元 (gated recurrent unit, GRU)<sup>[15]</sup>, 这 2 个变体引入门机制, 有效地缓解了上述的问题。随后学者将循环神经网络与注意力机制融合, 提出基于循环神经网络的双阶注意力网络 (dual-stage attention based recurrent neural network, DA-RNN)<sup>[16]</sup>, 该网络能够自适应地选择最相关的输入特征, 捕捉时间序列的长期时间依赖性。又有学者在结合 CNN 和 LSTM 基础上, 提出长期和短期时间序列网络 (long and short-term time-series network, LSTNet)<sup>[17]</sup>, 该模型使用 CNN 来捕捉序列间的短期局部依赖关系, 使用 RNN 来发现序列间的长期依赖关系, 并且与传统的自回归模型相融合, 取得了很好的预测效果。文献 [18] 利用并行卷积和注意力机制, 并融合传统自回归模型, 提出适用于动态周期序列或非周期序列的 DSANet (dual self-attention network)。文献 [19] 提出时间卷积网络 (temporal convolutional networks, TCN), 它通过结合因果卷积、残差卷积和膨胀卷积等, 增大时域感受野, 在时序预测问题上效果更好。此后, 文献 [20] 提出的 SCINet (sample convolution and interaction network) 通过多分辨率样本卷积和交互进行时间建模, 有效增强对复杂时间动态序列的可预测性。另一方面, 由于 Transformer<sup>[21]</sup> 在机器翻译领域的良好效果, Transformer 也很快被应用到时间序列预测任务上。文献 [22] 尝试将 Transformer 结构应用在时

间序列预测任务中并取得较好效果。但同时 Transformer 也存在诸如计算时间长、内存占用高等方面问题, 使其无法直接应用于更长时间序列预测任务当中。文献 [23] 针对计算开销大的问题, 提出 LogSparse Transformer 结构, 将 Transformer 模型的计算复杂度从  $O(L^2)$  降低到  $O(L \log L)$ 。Informer<sup>[24]</sup> 主要针对长序列预测, 提出 ProbSparse self-attention 结构, 降低了注意力层的时间和空间复杂度, 同时通过生成式解码器直接产生长期预测序列, 避免在使用单步前向预测进行长期预测时出现累积误差, 提高长序列预测的效率。此外有学者针对频域自注意力机制时序建模, 提出 Autoformer<sup>[25]</sup>, 设计了一种短期趋势分解架构, 能够更为准确分解时态数据的趋势和季节成分, 使模型能够比标准注意力更好地学习时间依赖性。FEDformer (frequency enhanced decomposed transformer)<sup>[26]</sup> 通过傅里叶变换和小波变换, 提出了一种在频域应用注意力机制的模型结构, 以增加对噪声的鲁棒性。然而在多维时序预测中, 由于其只关注数据的时序特征, 现有方法无法完全利用变量之间的潜在空间依赖性。近年来, 图神经网络 (graph neural networks, GNN)<sup>[27]</sup>, 成为对图结构数据中的复杂模式进行建模的一种成功方法。文献 [28] 提出了一种图神经网络架构 Graph WaveNet, 主要用于解决时空建模问题上图结构具有不确定性的问题, 通过自适应的可学习的邻接矩阵从数据中自动学习图结构, 从而精确地捕获数据中隐藏的空间依赖关系, 同时采用一维膨胀因果卷积来替代 RNN, 提升对时序数据长期关系的提取能力。文献 [29] 提出一种新的联合推断和利用时序关系的图推理方法, 该方法推断出一个潜在的全连接图来对时间序列之间的相互作用进行建模, 将图模型中复杂程度由  $O(N^2)$  降低为  $O(NK)$ , 实现精度和计算复杂度之间的平衡。随后, 为了更好地建模多变量时间序列数据中成对变量之间的潜在空间依赖, 文献 [30] 提出了一种为多元时间序列数据设计的通用图神经网络模型, 取得不错效果。

经过上述分析, 多维时间序列预测研究已经取得一定的成果, 但仍存在一些不足: 1) 不同通道的时间序列呈现不同走势, 其对应的不同通道特征对预测发挥着不同程度的作用, 但上述模型没有很好地利用时序数据多通道的特征信息。2) 仍然存在时序数据空间依赖关系捕获不足的问题。

为了实现对多维时间序列准确和稳健的预

测, 本文将空间注意力机制和通道注意力机制引入到多维时序预测问题中, 通过多个并行的多层膨胀卷积和通道注意力机制挖掘多通道时序特征信息, 空间注意力机制增强图卷积空间依赖特征, 提出一种新的基于双注意力机制的网络结构, 该网络能够在解决多维数据空间依赖不足和多通道信息利用不足问题的基础上提高预测的效果。本文主要贡献为:

1) 通过空间注意力机制对时序数据空间依赖特征进行提取, 空间依赖特征与图卷积特征融合, 增强变量间依赖关系, 主要解决变量间依赖关系捕获能力不足的问题。

2) 利用膨胀卷积和通道注意力机制, 对时序数据进行多通道的特征提取, 解决了时序数据多通道信息利用不充分的问题。

## 1 融合双注意力机制的 GNN 网络模型

### 1.1 问题描述

多维时间序列表示由多个相关的单维时间序列构成, 接下来的研究方法, 主要聚焦于多变量的时序预测的任务, 具体可表示为如下定义:

时间序列是指由一组时间间隔固定且按有序的时间顺序组成的观测值序列  $\mathbf{X}^{(i)} = [x_1^{(i)} x_2^{(i)} \cdots x_T^{(i)}]$ ,

其中  $i \in \{1, 2, \dots, D\}$ ,  $x_t^{(i)}$  表示在  $T$  时刻记录第  $i$  个观测值。多变量时间序列表示为  $\mathbf{X} = [\mathbf{X}^{(1)} \mathbf{X}^{(2)} \cdots \mathbf{X}^{(D)}]$ ,  $\mathbf{X} \in \mathbf{R}^{D \times T}$ , 其中  $D$  表示多维变量的维度,  $T$  为输入的时间步总数。

多维时间序列预测表示有一组时序数据  $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \cdots \mathbf{X}_T]$ , 目标是预测  $\mathbf{X}_{T+h}$ , 其中  $\mathbf{X}_T$  表示  $T$  时刻的记录,  $T$  表示序列长度,  $h$  表示未来预测的步长,  $\mathbf{X}_T$  表示为

$$\mathbf{X}_T = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_T^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_T^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(D)} & x_2^{(D)} & \cdots & x_T^{(D)} \end{bmatrix}$$

本文采用滚动预测的方式进行多变量时间序列预测, 即基于  $[\mathbf{X}_{1+k} \mathbf{X}_{2+k} \cdots \mathbf{X}_{T+k}]$ , 预测未来的、的值  $\mathbf{X}_{T+h+k}$ , 其中  $k$  为滑动值且  $k \in \mathbf{R}^+$ 。

### 1.2 FFDA-GNN 模型结构

本文基于图神经网络和双注意力机制, 提出一种针对多变量的时间序列数据的预测模型, 简记为 FFDA-GNN, 其整体的模型结构如图 1 所示。FFDA-GNN 模型由图学习层、 $m$  个时序卷积模块 (temporal convolution module, TCM)、 $m$  个图卷积模块、空间注意力模块 (spatial attention module, SAM) 和输出部分共 5 个部分组成。

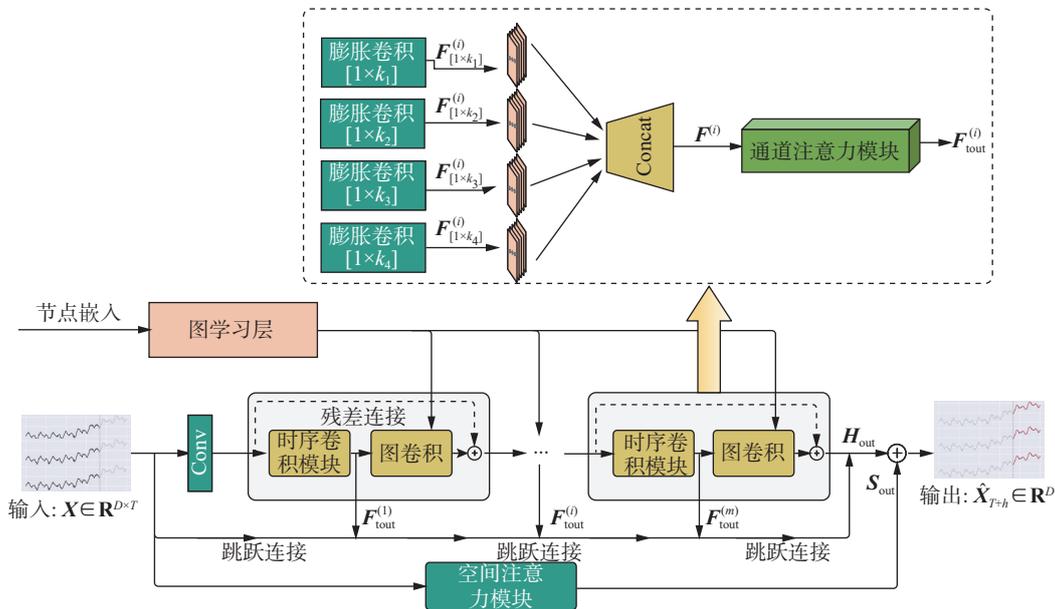


图 1 FFDA-GNN 模型结构  
Fig. 1 Architecture of FFDA-GNN

给定样本  $\mathbf{X} \in \mathbf{R}^{D \times T}$ , 模型 FFDA-GNN 首先对每个变量进行节点嵌入表示, 图学习层根据多维时序数据计算余弦相似度来学习图的邻接矩阵。同时样本  $\mathbf{X}$  经过一维卷积, 映射为对应的三维特征  $\mathbf{F}_0 \in \mathbf{R}^{D \times T' \times C}$ , 其中  $C$  为特征通道数,  $T'$  为卷积之

后的时序特征长度。之后将  $\mathbf{F}_0$  输入到时序卷积模块中, 经过时序卷积模块获取多维时序数据的多个通道特征。将经过图学习层学习的邻接矩阵和时序卷积模块输出特征输入到图卷积模块中, 通过图卷积模块来获取多维时序数据之间的空间

依赖特征  $\mathbf{H}_{\text{out}} \in \mathbf{R}^D$ 。同时将样本  $\mathbf{X}$  输入到空间注意力机制中, 得到特征权重  $\mathbf{S}_{\text{out}} \in \mathbf{R}^D$ , 最后将图卷积结果与空间注意力机制结果相加得到整个模型的预测值  $\hat{\mathbf{X}}_{T+h} \in \mathbf{R}^D$ 。

### 1.3 图学习层

图学习层的主要作用是能够根据每个节点生成的节点嵌入, 学习图结构的时间序列数据之间的隐含关系。每个节点则代表预测任务中的每一个变量。本文采用余弦相似度来评估 2 个向量的相似程度。当余弦相似度的值为 0 时, 代表 2 个向量无相似性, 当值为 1 的时候表示 2 个向量最相似。具体的方法公式为

$$C_{ij} = \frac{\mathbf{E}_i \cdot \mathbf{E}_j^T}{\|\mathbf{E}_i\|_2 \cdot \|\mathbf{E}_j\|_2}$$

$$\mathbf{A} = \text{ReLU}(\tanh(\mathbf{C} - \mathbf{I})) \quad (1)$$

$$A_{ij} = \begin{cases} 1, & j \in \text{argtopk}(\mathbf{A}[i,:]) \\ 0, & j \notin \text{argtopk}(\mathbf{A}[i,:]) \end{cases} \quad (2)$$

式中:  $\mathbf{C}$  表示 2 个矩阵的余弦相似度矩阵,  $C_{ij}$  为第  $i$  节点和第  $j$  个节点的余弦相似值,  $\mathbf{E}_i$  和  $\mathbf{E}_j$  分别表示在第  $i$  个节点和第  $j$  个节点的节点嵌入,  $\mathbf{I}$  表示单位矩阵。  $\text{argtopk}(\mathbf{A}[i,:])$  表示第  $i$  行中向量的前  $k$  个最大值的索引集合。该模块通过式 (1) 利用 ReLU 激活函数实现邻接矩阵的归一化处理。式 (2) 得到稀疏邻接矩阵同时减少图卷积的计算成本,  $A_{ij}$  表示第  $i$  节点和第  $j$  节点对应的特征矩阵的值, 这里选择节点的前  $k$  个最近邻节点作为邻居, 并将其未连接的节点权重置为 0。

### 1.4 时序卷积模块

时序卷积模块主要解决没有充分利用多维时序数据多通道信息的问题。FFDA-GNN 模型的时序卷积模块由不同尺寸大小卷积核的一维膨胀卷积和通道注意力机制组成, 用于对多维时序数据进行多通道的时序特征提取。时序卷积模块参考 MTGNN 模型<sup>[30]</sup>, 共有 5 层, 每一层都会使用尺寸为  $1 \times 2$ 、 $1 \times 3$ 、 $1 \times 6$ 、 $1 \times 7$  的 4 种卷积核进行卷

积操作。首层膨胀卷积的膨胀系数为  $q = 1$ , 随着隐藏层数目的增加, 膨胀系数会以  $q$  的指数倍速率增长, 其对应网络的感受野大小也会以  $q$  的速率呈指数增加。因此, 使用这种策略可以捕获更长的序列。给定第  $i$  层的输入特征为  $\mathbf{F}_{\text{in}}^{(i)} \in \mathbf{R}^{D \times T \times C}$ , 特征提取过程公式为

$$d_i = q^{i-1}$$

式中  $d_i$  表示第  $i$  层的膨胀率。对于计算变量的隐状态, 公式为

$$\mathbf{h}^n = \mathbf{F}_{\text{in}}^{(i)}(n)$$

式中  $\mathbf{h}^n$  表示第  $n$  个变量的隐状态。对于计算表示第  $i$  层对第  $n$  个变量  $t$  时刻的隐状态, 公式为

$$\mathbf{F}_{1 \times k}^{(i)}(n, t, d_i) = \sigma \left( \sum_{s=0}^{k-1} f_{1 \times k}^{(i)}(s) \mathbf{h}_{t-d_i+s}^n \right)$$

式中:  $f_{1 \times k}^{(i)}$  表示第  $i$  层尺寸为  $1 \times k$  的膨胀卷积核,  $\mathbf{F}_{1 \times k}^{(i)}$  表示第  $i$  层对第  $n$  个变量  $t$  时刻的隐状态,  $\sigma(\cdot)$  为激活函数。

通道注意力机制结构如图 2 所示, 首先将各个卷积核卷积运算结果经过拼接操作得到特征  $\mathbf{F}^{(i)} \in \mathbf{R}^{D \times T \times C}$ , 通过平均池化层和最大池化层将特征  $\mathbf{F}^{(i)}$  的各个通道时序信息进行压缩, 生成 2 组特征  $\mathbf{F}_{\text{max}}^{(i)}$ ,  $\mathbf{F}_{\text{avg}}^{(i)} \in \mathbf{R}^C$ , 之后将 2 组特征相加, 经过处理之后由 Sigmoid 函数获取通道注意力的权重。通道注意力机制的公式为

$$\mathbf{C}^{(i)} = \sigma \left( \mathbf{F}_{\text{chan}}^{(i)} (\text{AvgPool}(\mathbf{F}^{(i)}) + \text{MaxPool}(\mathbf{F}^{(i)})) \right)$$

式中:  $\text{AvgPool}(\cdot)$  和  $\text{MaxPool}(\cdot)$  分别表示对时序特征的平均池化和最大池化操作,  $\mathbf{C}^{(i)}$  表示第  $i$  层的通道注意力权重矩阵,  $\mathbf{F}_{\text{chan}}^{(i)}$  表示第  $i$  层进行卷积和 ReLU 激活函数计算的输出, 公式为

$$\mathbf{F}_{\text{chan}}^{(i)} = \mathbf{W}_{\text{chan1}}^{(i)} \left( \sigma \left( \mathbf{W}_{\text{chan2}}^{(i)} (\mathbf{F}^{(i)}) \right) \right)$$

式中  $\mathbf{W}_{\text{chan1}}^{(i)}$  和  $\mathbf{W}_{\text{chan2}}^{(i)}$  分别表示参数矩阵。经过通道注意力权重与输入特征相乘生成时序卷积模块的输出  $\mathbf{F}_{\text{tout}}^{(i)} \in \mathbf{R}^{D \times T \times C}$ 。

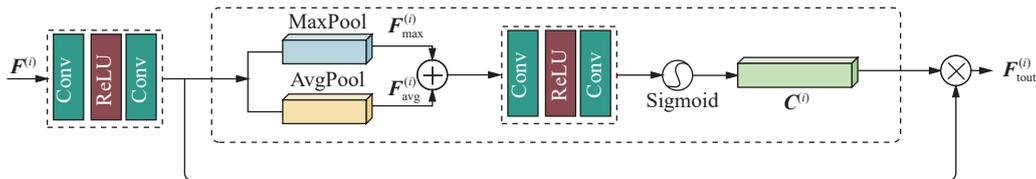


图 2 通道注意力机制结构

Fig. 2 Structure of the channel mechanism

### 1.5 图卷积模块

图卷积模块主要作用是根据图学习层的邻接矩阵和时序卷积模块传递过来的多通道的时序特征  $\mathbf{F}_{\text{tout}}^{(i)}$ , 提取数据中的空间依赖关系。图卷积模

块<sup>[30]</sup>由 2 个混合跳传播层组成, 第 1 个用于处理经过每个节点的流入信息, 第 2 个用于处理节点的流出信息。之后将 2 层的输出相加得到最终流入信息。图 3 分别显示了图卷积模块整体结构和

混合跳传播层的结构。图 3(a) 中的  $A$  为学习到的邻接矩阵,  $A^T$  为邻接矩阵的转置矩阵。图 3(b) 为混合跳传播层的网络结构。混合跳传播层由信息传播和信息选择 2 个步骤组成。

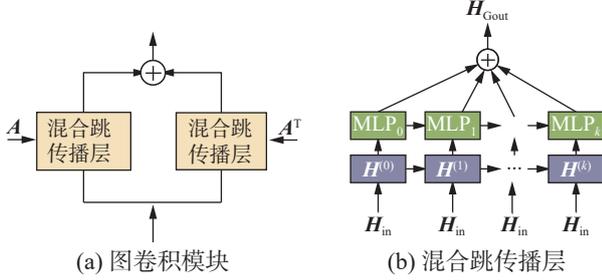


图 3 图卷积和混合跳传播层

Fig. 3 Graph convolution and mix-hop propagaion layer

混合跳传播层的信息传播过程公式为

$$H^{(k)} = \beta H_{in} + (1 - \beta) \tilde{A} H^{(k-1)}$$

式中:  $\beta$  是一个超参数, 控制着根节点保持原始状态的比例。  $k$  为传播的深度,  $H_{in}$  表示由前一层输出的输入隐藏状态,  $H^{(0)} = H_{in}$ ,  $\tilde{A} = \tilde{D}^{-1}(A + I)$ ,  $\tilde{D}_{ii} = 1 + \sum_j A_{ij}$ 。混合跳传播层的信息选择过程公式为

$$H_{Gout} = \sum_{i=0}^K H^{(i)} W^{(i)}$$

式中:  $H_{Gout}$  表示图卷积当前层的输出;  $W^{(i)}$  表示参数矩阵, 当图结构没有空间依赖时, 将  $W^{(i)} = 0$  仅保留初始节点信息。在本文提出的传播层中, 水平方向为信息传播的过程, 垂直方向为信息选择的过程。经过 2 个混合跳传播层处理, 第  $i$  层图卷积模块最终输出  $H_{GCN}^{(i)} \in \mathbf{R}^{D \times T \times C}$ 。

### 1.6 空间注意力模块

空间注意力模块主要是针对空间依赖关系捕获不足的问题, 用于增强空间依赖特征。该模块参考 CBAM<sup>[31]</sup> 的空间注意力模块部分, 由 2 个卷积核大小为  $3 \times 3$  的卷积神经网络层、非线性激活函数 ReLU 和 Sigmoid 层组成。首先, 将样本  $X$  的多维时序数据作为输入数据, 经过空间注意力机制模块的首个  $3 \times 3$  卷积运算, 提取时序数据在空间上的特征, 之后经过 ReLU 激活函数处理, 将处理后的特征再次经过卷积核大小为  $3 \times 3$  的卷积运算, 生成各个变量中每个元素的重要程度, 最后, 由 Sigmoid 激活函数生成在空间上对应的权重矩阵, 经过全连接层得到  $S_{out} \in \mathbf{R}^D$ 。

### 1.7 预测结果的输出

在 FFDA-GNN 模型的预测输出阶段, 主要是将图卷积输出的结果  $H_{GCN}^{(i)}$  经过跳跃连接组合相加, 与空间注意力机制输出的结果  $S_{out}$  进行和运算作为最终的预测结果。图卷积层输出结果  $H_{GCN}^{(i)}$ ,

经过激活函数 ReLU、卷积核大小为  $1 \times 1$  的卷积运算, 将输入的通道维度转为所需的输出维度, 再组合跳跃连接, 得到输出特征  $H_{out} \in \mathbf{R}^D$ , 然后将其与空间注意力机制模块输出  $S_{out}$  进行相加运算, 得到该模型的最终预测结果  $\hat{X}_{T+h} \in \mathbf{R}^D$ , 公式为

$$\hat{X}_{T+h} = S_{out} + H_{out}$$

## 2 实验设置与结果分析

### 2.1 数据集介绍

实验内容参考同类实验<sup>[11-22]</sup>, 选取加州大学欧文分校 (UCI) 机器学习库中真实存在的原始数据集。分别为电量数据集 (electricity)、汇率数据集 (exchange-rate) 和交通数据集 (traffic)。

**电量数据集** 包含 321 个葡萄牙用户自 2012 年到 2014 年的每小时电力消耗数据。

**汇率数据集** 包含 8 个国家从 1990 年到 2016 年的每日汇率数据。

**交通数据集** 来自加州交通部 2015 年到 2016 年共计 48 个月的数据, 数据记录了高速公路上 862 个传感器每小时测量的道路占用率, 数据值介于 0 和 1 之间。

在本文进行的实验中, 所有的数据集按时间顺序分为训练集 (60%)、验证集 (20%) 和测试集 (20%)。为了避免数据尺度对实验结果的影响, 数据集训练之前会采用最大最小值归一化 (min-max normalization) 的方法进行数据预处理。

### 2.2 评估指标

本文模型同样是沿用了 LSTNet 和 MTGNN 实验所使用的评价指标, 即相对平方误差 (relative squared error, RSE) 和经验相关系数 (empirical correlation coefficient, CORR) 这 2 个常用评估指标。RSE 是均方根差 (root mean square error, RMSE) 的一个变体, 其针对模型中预测变量的数量进行调整, 用于评估模型预测结果与真实值的误差, 对于 RSE, 值越低表示预测模型的性能越好, 其公式为

$$E_{RSE} = \frac{\sqrt{\sum_{(i,j) \in \Omega_{test}} (Y_{it} - \hat{Y}_{it})^2}}{\sqrt{\sum_{(i,j) \in \Omega_{test}} (Y_{it} - \text{mean}(Y))^2}}$$

CORR 是通过评估真实值与预测值的相关情况来评估模型效果, CORR 的取值范围是 0~1, 当 CORR 为 0, 表示模型预测不准确, 当 CORR 为 1, 表示预测效果非常好, 即 CORR 值越大说明模型的预测效果越好, 其公式为

$$E_{\text{CORR}} = \frac{1}{D} \sum_{i=1}^D \frac{\sum_t (Y_{it} - \text{mean}(Y_i))(\hat{Y}_{it} - \text{mean}(\hat{Y}_i))}{\sqrt{\sum_t (Y_{it} - \text{mean}(Y_i))^2(\hat{Y}_{it} - \text{mean}(\hat{Y}_i))^2}}$$

式中  $Y$  和  $\hat{Y}$  分别是真实值和模型的预测结果。

### 2.3 实验参数设置

本次实验在 Python3.8 环境下实现, 使用 PyTorch 深度学习框架完成。模型实验运行在 Ubuntu 操作系统, 所用 GPU 为 NVIDIA RTX A5000 24 GB。同时实验采用滑动窗口 (sliding window) 的预测方法, 预测未来步长的数据。针对本次实验所使用的超参数, 参考同类的深度学习模型参数设置相关参数范围。

表 1 为参数范围和调优的具体参数设置, 另外, epoch 为 100。

表 1 模型主要参数设置  
Table 1 Model main parameter settings

参数名	参数取值
激活函数	{Sigmoid, ReLU}, ReLU, Sigmoid
优化器	{SGD, Adagrad, Adam}, Adam
学习率	{0.000 1, 0.000 5, 0.001}, 0.001
损失函数	{ $L_1, L_2$ }, $L_1$
参数丢弃系数	{0.1, 0.2, 0.5, 0.6, 0.8}, 0.8
窗口大小	32
预测步长	{3, 6, 12, 24}
图卷积模块数	5
每个图卷积模块层数	2
梯度裁剪数	5
批处理大小	{16, 32, 64, 128, 256}, 32
空间注意力卷积核数	3
空间注意力卷积步长	1

### 2.4 实验结果及分析

#### 2.4.1 对比实验

为了验证模型的有效性, 实验选择 7 种深度学习中主流的时序预测模型作为对比模型, 分别为:

AR<sup>[6]</sup>, 一种单变量平稳自回归模型;

VARMLP<sup>[32]</sup>, 一种向量自回归模型和多层感知器的组合模型;

GP<sup>[10]</sup>, 一种为时序数据设计的高斯过程;

RNN-GRU<sup>[33]</sup>, 一种循环神经网络使用 GRU 作为隐层的时间序列预测模型;

MTGNN<sup>[30]</sup>, 一种为多维时间序列数据设计的通用图神经网络模型;

LSTNet-skip<sup>[17]</sup>, 一种结合卷积神经网络和长

短期记忆网络的时序预测模型;

AttnAR<sup>[34]</sup>, 一种基于注意力和自回归的多维时间序列预测模型。

表 2~4 为本文模型与对应的几种对比方法在测试集上的 RSE 和 CORR 的评估结果。其中, 输入时序数据的窗口大小为 32, 预测未来步长的取值为 {3,6,12,24}, 分别表示数据集 exchange-rate 的第 3、6、12、24 天, 数据集 electricity 和 traffic 的第 3、6、12、24 小时的预测结果, 这里将产生的指标数据默认保存小数点后 3 位。在实验结果上, 本文提出的模型在指定的 3 种数据集 (exchange、electricity 和 traffic) 上 RSE 指标相对最低, CORR 指标相对最高, 说明该模型的预测性能优于其他基准模型, 在指定数据集进行预测任务时有很好的预测效果。

表 2 在 exchange-rate 数据集预测结果对比  
Table 2 Comparison of prediction results in the exchange-rate dataset

模型	指标	预测步长			
		3	6	12	24
AR	RSE	0.022 8	0.027 9	0.035 3	0.044 5
	CORR	0.973 4	0.965 6	0.952 6	0.935 7
VARMLP	RSE	0.026 5	0.039 4	0.040 7	0.057 8
	CORR	0.860 9	0.872 5	0.828 0	0.767 5
GP	RSE	0.023 9	0.027 2	0.039 4	0.058 0
	CORR	0.871 3	0.819 3	0.848 4	0.827 8
RNN-GRU	RSE	0.019 2	0.026 4	0.040 8	0.062 6
	CORR	0.978 6	0.971 3	0.953 1	0.922 3
LSTNet-skip	RSE	0.022 6	0.028 0	0.035 6	0.044 9
	CORR	0.973 5	0.965 8	0.951 1	0.935 4
MTGNN	RSE	0.019 4	0.025 9	0.034 9	0.045 6
	CORR	0.978 6	0.970 8	0.955 1	0.937 2
AttnAR	RSE	—	0.024 0	0.033 6	0.044 8
	CORR	—	0.967 2	0.953 6	0.924 8
FFDA-GNN	RSE	0.015 7	0.023 7	0.028 5	0.044 2
	CORR	0.981 3	0.973 0	0.959 4	0.937 5

表 3 在 electricity 数据集预测结果对比  
Table 3 Comparison of prediction results in the electricity dataset

模型	指标	预测步长			
		3	6	12	24
AR	RSE	0.099 5	0.103 5	0.105 0	0.105 4
	CORR	0.884 5	0.863 2	0.859 1	0.859 5
VARMLP	RSE	0.139 3	0.162 0	0.155 7	0.127 4
	CORR	0.870 8	0.838 9	0.819 2	0.867 9

续表 3

模型	指标	预测步长			
		3	6	12	24
GP	RSE	0.150 0	0.190 7	0.162 1	0.127 3
	CORR	0.867 0	0.833 4	0.839 4	0.881 8
RNN-GRU	RSE	0.110 2	0.114 4	0.118 3	0.129 5
	CORR	0.859 7	0.862 3	0.847 2	0.865 1
LSTNet-skip	RSE	0.086 4	0.093 1	0.100 7	0.100 7
	CORR	0.928 3	0.913 5	0.907 7	0.911 9
MTGNN	RSE	0.074 5	0.087 8	0.091 6	0.095 3
	CORR	0.947 4	0.931 6	0.927 8	0.923 4
AttnAR	RSE	—	0.087 1	0.094 2	0.098 9
	CORR	—	0.916 0	0.910 8	0.908 9
FFDA-GNN	RSE	0.074 2	0.087 2	0.091 2	0.100 3
	CORR	0.950 1	0.931 7	0.928 1	0.919 5

表 4 在 traffic 数据集预测结果对比

Table 4 Comparison of prediction results in the traffic dataset

模型	指标	预测步长			
		3	6	12	24
AR	RSE	0.599 1	0.621 8	0.625 2	0.630 0
	CORR	0.775 2	0.756 8	0.754 4	0.751 9
VARMLP	RSE	0.558 2	0.657 9	0.602 3	0.614 6
	CORR	0.824 5	0.769 5	0.792 9	0.789 1
GP	RSE	0.608 2	0.677 2	0.640 6	0.599 5
	CORR	0.783 1	0.740 6	0.767 1	0.790 9
RNN-GRU	RSE	0.535 8	0.552 2	0.556 2	0.563 3
	CORR	0.851 1	0.840 5	0.834 5	0.830 0
LSTNet-skip	RSE	0.477 7	0.489 3	0.495 0	0.497 3
	CORR	0.872 1	0.869 0	0.861 4	0.858 8
MTGNN	RSE	0.416 2	0.475 4	0.446 1	0.453 5
	CORR	0.896 3	0.866 7	0.879 4	0.881 0
AttnAR	RSE	—	0.428 7	0.437 0	0.439 6
	CORR	—	0.886 5	0.881 9	0.880 0
FFDA-GNN	RSE	0.415 6	0.472 5	0.445 4	0.454 7
	CORR	0.896 9	0.866 9	0.879 2	0.878 9

从表 2 可知, 对 exchange-rate 数据集进行预测时, 本文模型相对 MTGNN 模型在不同预测步长下的 RSE 误差分别降低了 19.07%、8.49%、18.33%、3.07%。对比早期的 AR 模型效果提升较大, 原因是统计方法 AR 模型无法捕获多维时序数据的非线性关系; VARMLP 和 RNN-GRU 模型只能捕获单一变量的时序特征, 无法对多变量特征依赖关系进行提取, 导致误差相对较大; 而 LST-

Net 只能处理定长的短期和长期特征, 无法很好处理时序数据的多通道特征; 此外虽然 AttnAR 引入注意力机制, 但只关注时序维度特征, 忽略空间维度, 对预测效果提升有限。从表 3 可知, 对 electricity 数据集进行预测时, 当预测步长为 3 时, 该模型比 MTGNN 模型在评价指标 CORR 提升明显, 当预测步长为 6 时, 模型在该评价指标有部分提升。总体来说, 当预测步长为 3、6、12 时, 该模型的预测效果最好, 说明当预测步长为 3、6 和 12 时, 通道注意力模块能够对时序数据的每个通道特征进行进一步捕获, 并且经过空间注意力机制的特征融合, 该模型能够更好地利用多维的空间依赖关系提升预测精度。从表 4 可知, traffic 数据集中该模型预测效果提升并不显著, 原因在于模型在 traffic 数据集中提取时序特征方面并不具有良好的表现。针对 3 种数据集, 当预测步长为 3、6 和 12 时, 该模型的性能是逐渐变差的, 说明该数据集时间分布特征非常明显, 对于短期预测性能较好。当预测步长为 24 时, 模型预测效果最低, 说明本文模型对于长期预测的能力相对较低。综上, 该模型双注意力机制在多变量的时间序列预测问题上有良好的可靠性。

### 2.4.2 消融实验

由 1.2 节内容可知, FFDA-GNN 模型包括时序卷积模块和空间注意力模块。为了验证各部件在该模型中的有效性, 本文设计如下网络: 1) Base; 2) Base+TCM; 3) Base+SAM; 4) FFDA-GNN。其中, Base 模型为在 FFDA-GNN 模型的基础上除去 TCM 和 SAM 模块; Base+TCM 为在 Base 模型的基础上添加时序卷积模块 TCM; Base+SAM 为在 Base 模型的基础上添加空间注意力模块 SAM。选取有代表性的 2 个数据集 exchange-rate 和 electricity 进行消融实验。

针对 exchange-rate 数据集实验结果如图 4 所示。实验结果表明, 在 exchange-rate 数据集中, 添加空间注意力机制 SAM 模块的模型, 在预测步长为 3、6 和 12 时, 其评价指标 RSE 分别降低了 18.69%、13.10% 和 5.86%。主要原因是空间注意力机制 SAM 模块准确地挖掘了多个变量之间的关联关系, 增强了由图卷积模块生成的多变量之间依赖关联程度。对于增加了时序卷积模块 TCM 的模型, 在不同的预测步长下, CORR 评估指标分别提高了 0.000 7、0.000 3、0.000 1, 表明通道注意力机制对多个通道的时序特征进行了聚焦, 提高了模型的预测能力。

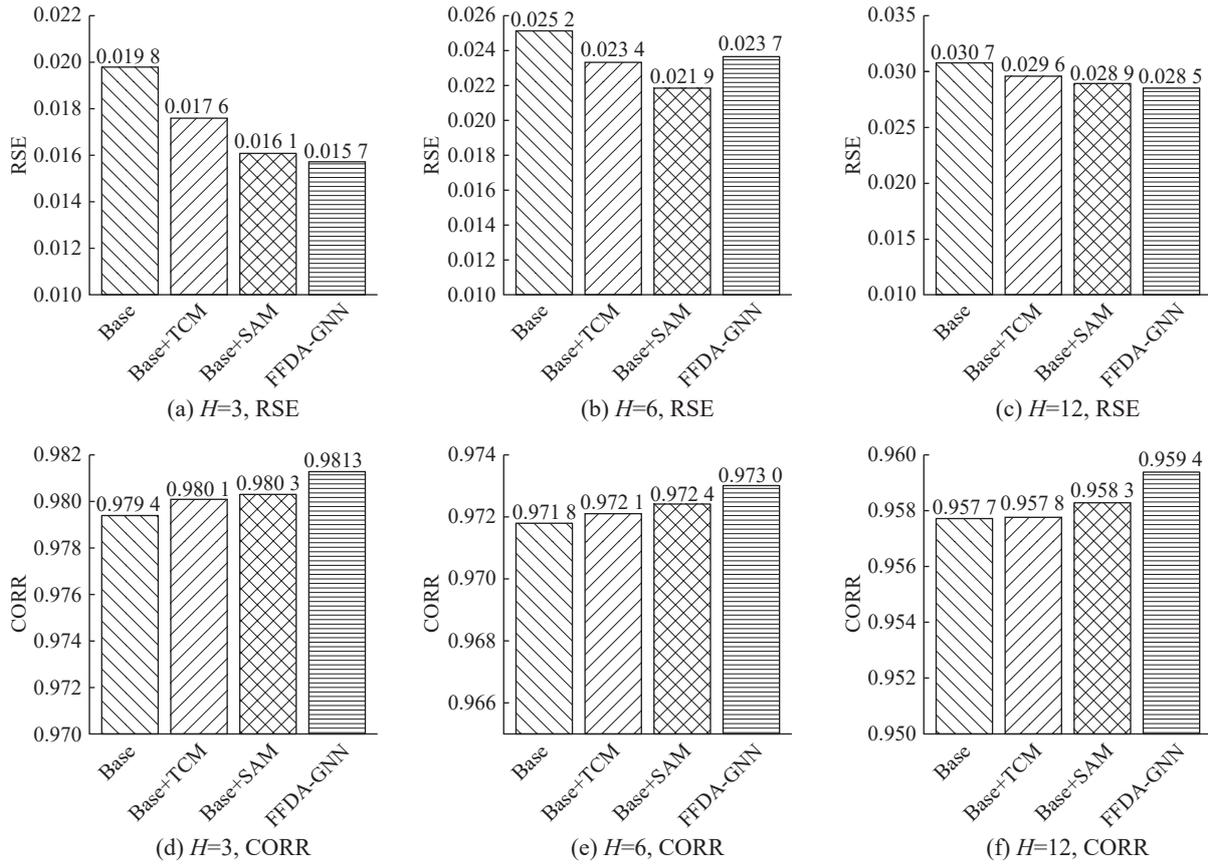


图 4 在 exchange-rate 数据集消融实验对比

Fig. 4 Comparison of ablation study in the exchange-rate dataset

而在 electricity 数据集上, 通过表 5 可得, 添加 TCM 和 SAM 2 个模块后, 本文提出的 FFDA-GNN 模型在 RSE 和 CORR 2 个指标上取得了最好的效果。相比 Base 模型, FFDA-GNN 模型在不同的预测步长上 RSE 评估指标分别降低了 15.39%、13.27%、11.74%。综上, 在 Base 模型的基础增加了基于通道注意力机制的 TCM 和基于空间注意力的 SAM 模块后模型达到了最好的结果, 注意力机制的添加提升了模型整体的预测能力。

表 5 在 electricity 数据集消融实验对比

Table 5 Comparison of ablation study in the electricity dataset

模型	指标	预测步长		
		3	6	12
Base	RSE	0.0877	0.1190	0.1260
	CORR	0.9254	0.8956	0.8854
Base+TCM	RSE	0.0899	0.1227	0.1258
	CORR	0.9258	0.9027	0.8936
Base+SAM	RSE	0.0870	0.1137	0.1255
	CORR	0.9261	0.8996	0.8860
FFDA-GNN	RSE	0.0742	0.1032	0.1112
	CORR	0.9501	0.9317	0.9281

### 3 结束语

本文提出了一种基于图神经网络和双注意力机制的混合网络模型。该模型利用多个卷积核的一维膨胀卷积提取多维时序数据多个范围的时序特征, 通过增加通道注意力机制, 捕获多通道的时序数据特征, 经图卷积神经网络提取多维时间序列变量之间的空间依赖关系特征, 同时利用空间注意力机制进一步提取多变量之间的空间依赖关系, 然后融合多维时间序列的变量间依赖关系特征进行预测。在经济、电力和交通等领域的数据集上的实验结果表明, FFDA-GNN 在相对平方误差 RSE 和经验相关系数 CORR 2 种指标上取得了较好的预测结果, 且优于其他基准模型, 更加适合解决多变量时间序列预测任务。未来的工作重点将集中在如何使用更优的算法来自适应地学习图的邻接矩阵, 进一步挖掘时间序列数据之间的隐含关系, 同时在模型超参数优化方面做进一步探究。

### 参考文献:

[1] GASPARIN A, LUKOVIC S, ALIPPI C. Deep learning for time series forecasting: The electric load case[J]. CAAI

- transactions on intelligence technology, 2022, 7(1): 1–25.
- [2] 王锋华, 成敬周, 文凡. 快速双非凸回归算法及其电力数据预测应用[J]. 智能系统学报, 2018, 13(4): 665–672. WANG Fenghua, CHENG Jingzhou, WEN Fan. Fast double nonconvex regression algorithm for forecast of electric power data[J]. CAAI transactions on intelligent systems, 2018, 13(4): 665–672.
- [3] 程鹏超, 杜军平, 薛哲. 基于多路交叉的用户金融行为预测[J]. 智能系统学报, 2021, 16(2): 378–384. CHENG Pengchao, DU Junping, XUE Zhe. Prediction of user financial behavior based on multi-way crossing[J]. CAAI transactions on intelligent systems, 2021, 16(2): 378–384.
- [4] FANG Weiwei, ZHUO Wenhao, YAN Jingwen, et al. Attention meets long short-term memory: a deep learning network for traffic flow forecasting[J]. Physica A: statistical mechanics and its applications, 2022, 587: 126485.
- [5] 李伯涵, 郭茂祖, 赵玲玲. 基于分割注意力机制残差网络的都市区域客流量预测[J]. 智能系统学报, 2022, 17(4): 839–848. LI Bohan, GUO Maozu, ZHAO Lingling. Passenger flow prediction in urban areas based on residual networks with split attention mechanism[J]. CAAI transactions on intelligent systems, 2022, 17(4): 839–848.
- [6] YULE G U. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sun-spot numbers[J]. Philosophical transactions of the royal society of London series A, containing papers of a mathematical or physical character, 1927, 226(1): 267–298.
- [7] WALKER G T. On periodicity in series of related terms [J]. Proceedings of the royal society of London series A, containing papers of a mathematical and physical character, 1931, 131(818): 518–532.
- [8] JENKINS G M, BOX G E P. Time series analysis: forecasting, and control[J]. Journal of time, 1976, 31(2): 238–242.
- [9] 杨海民, 潘志松, 白玮. 时间序列预测方法综述[J]. 计算机科学, 2019, 46(1): 21–28. YANG Haimin, PAN Zhisong, BAI Wei. Review of time series prediction methods[J]. Computer science, 2019, 46(1): 21–28.
- [10] 陈艳, 王子健, 赵泽, 等. 传感器网络环境监测时间序列数据的高斯过程建模与多步预测[J]. 通信学报, 2015, 36(10): 252–262. CHEN Yan, WANG Zijian, ZHAO Ze, et al. Gaussian process modeling and multi-step prediction for time series data in wireless sensor network environmental monitoring[J]. Journal on communications, 2015, 36(10): 252–262.
- [11] LI Peixian, TAN Zhixiang, YAN Lili, et al. Time series prediction of mining subsidence based on a SVM[J]. Mining science and technology (China), 2011, 21(4): 557–562.
- [12] GOSSÉ J B, GUILLAUMIN C. L'apport de la représentation VAR de Christopher A. Sims à la science économique[J]. L'économie en perspective, 2014, 89(4): 305–319.
- [13] ELMAN J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179–211.
- [14] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735–1780.
- [15] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. (2014–12–11)[2023–05–20]. <https://arxiv.org/abs/1412.3555>.
- [16] QIN Yao, SONG Dongjin, CHEN Haifeng, et al. A dual-stage attention-based recurrent neural network for time series prediction[EB/OL]. (2017–04–07)[2023–05–20]. <https://arxiv.org/abs/1704.02971>.
- [17] LAI Guokun, CHANG Weicheng, YANG Yiming, et al. Modeling long and short-term temporal patterns with deep neural networks[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor: ACM, 2018: 95–104.
- [18] HUANG Siteng, WANG Donglin, WU Xuehan, et al. DSANet: dual self-attention network for multivariate time series forecasting[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019: 2129–2132.
- [19] BAI Shaojie, KOLTER J Z, KOLTUN V, et al. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB/OL]. (2018–03–04)[2023–05–20]. <https://arxiv.org/abs/1803.01271>.
- [20] LIU Minhao, ZENG Ailing, CHEN Muxi, et al. SCINet: Time series modeling and forecasting with sample convolution and interaction[J]. Advances in neural information processing systems, 2022, 35: 5816–5828.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. California: ACM, 2017: 6000–6010.
- [22] WU N, GREEN B, BEN Xue, et al. Deep transformer models for time series forecasting: the influenza prevalence case[EB/OL]. (2020–01–23)[2023–05–20]. <https://arxiv.org/abs/2001.08317>.
- [23] LI Shiyang, JIN Xiaoyong, XUAN Yao, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[EB/OL]. (2019–07–29)

- [2023-05-20]. <https://arxiv.org/abs/1907.00235>.
- [24] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. In-former: beyond efficient transformer for long sequence time-series forecasting[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2021, 35(12): 11106–11115.
- [25] WU Haixu, XU Jiehui, WANG Jianmin, et al. Auto-former: decomposition transformers with auto-correlation for long-term series forecasting[J]. *Advances in neural information processing systems*, 2021, 34: 22419–22430.
- [26] ZHOU Tian, MA Ziqing, WEN Qingsong, et al. FED-former: frequency enhanced decomposed transformer for long-term series forecasting[EB/OL]. (2022-01-30) [2023-05-20]. <https://arxiv.org/abs/2201.12740>.
- [27] 吴博, 梁循, 张树森, 等. 图神经网络前沿进展与应用[J]. *计算机学报*, 2022, 45(1): 35–68.  
WU Bo, LIANG Xun, ZHANG Shusen, et al. Advances and applications in graph neural network[J]. *Chinese journal of computers*, 2022, 45(1): 35–68.
- [28] WU Zonghan, PAN Shirui, LONG Guodong, et al. Graph WaveNet for deep spatial-temporal graph modeling [EB/OL]. (2019-05-31)[2023-05-20]. <https://arxiv.org/abs/1906.00121>.
- [29] SATORRAS V G, RANGAPURAM S S, JANUSCHOWSKI T. Multivariate time series forecasting with latent graph inference[EB/OL]. (2022-03-07)[2023-05-20]. <https://arxiv.org/abs/2203.03423>.
- [30] WU Zonghan, PAN Shirui, LONG Guodong, et al. Connecting the dots: multivariate time series forecasting with graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event: ACM, 2020: 753–763.
- [31] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Lecture Notes in Computer Science. Cham: Springer, 2018: 3–19.
- [32] ZHANG G P. Time series forecasting using a hybrid AR-IMA and neural network model[J]. *Neurocomputing*, 2003, 50: 159–175.
- [33] JIA Pengtao, LIU Hangduo, WANG Sujian, et al. Research on a mine gas concentration forecasting model based on a GRU network[J]. *IEEE access*, 2020, 8: 38023–38031.
- [34] YOO J, KANG U. Attention-based autoregression for accurate and efficient multivariate time series forecasting [C]//Proceedings of the 2021 SIAM International Conference on Data Mining. Philadelphia: Society for Industrial and Applied Mathematics, 2021: 531–539.

### 作者简介:



范航舟, 硕士研究生, 主要研究方向为时序序列预测、大数据分析、深度学习。E-mail: [fanhangzhou@163.com](mailto:fanhangzhou@163.com)。



梅红岩, 教授, 博士, 中国计算机学会会员, 主要研究方向为深度学习、数据挖掘与大数据分析、网络服务。发表学术论文 40 余篇。E-mail: [715014795@qq.com](mailto:715014795@qq.com)。



赵勤, 硕士研究生, 主要研究方向为时序序列预测、深度学习。E-mail: [1139574265@qq.com](mailto:1139574265@qq.com)。