

王明胜. 基于时序双通道图注意力的工业过程故障时序预测模型[J]. 智能计算机与应用, 2024, 14(12): 26-35. DOI: 10.20169/j. issn. 2095-2163. 241204

# 基于时序双通道图注意力的工业过程故障时序预测模型

王明胜

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 工业过程是工业生产的基石, 工业过程故障会导致整个工业流程中断, 造成巨大经济损失, 但由于工业过程多元特征的高耦合和时变等特性, 传统机器学习方法难以实现对工业过程故障时序的有效预测。针对此问题, 本文利用共享的多头图注意力机制提取特征的高耦合关联关系, 并结合双向门控循环单元(Bi-directional Gated Recurrent Unit, BiGRU)和多尺度时间卷积网络(Multi-Scale Temporal Neural Network, MSTCN)构成时序学习层(Temporal Learning Module, TLM)提取特征的多尺度时序信息, 提出一种基于时序双通道图注意力(Temporal Dual-Channel Graph Attention, TDCGAT)的工业故障时序预测模型。最后, 通过堆叠时序学习层和共享的图注意力实现对工业过程故障时序的有效预测。

**关键词:** 多元时序预测; 工业过程故障; 图注意力; BiGRU; TLM

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)12-0026-10

## A time series prediction model for industrial process failures based on temporal dual-channel graph attention network

WANG Mingsheng

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**Abstract:** Industrial process is the cornerstone of industrial manufacturing, whose failure can lead to the interruption of the entire industrial processes, resulting in huge economic losses. However, due to the high-coupling and time-varying characteristic of the multivariate features of industrial processes, traditional machine learning methods have difficulty in achieving precise prediction of industrial process failures time series. To address this problem, a time-series prediction model for industrial faults based on Temporal Dual-Channel Graph Attention is proposed in this paper. In the model, a shared multi-head graph attention is used to extract the highly coupled correlations of industrial features, a BiGRU and MSTCN are combined as a Temporal Learning Module to extract the multi-scale temporal characteristics of features. Finally, the effective prediction of industrial process faults multivariate time series is achieved by stacking TLM and shared graph attention.

**Key words:** multivariate time series prediction; industrial process failure; graph attention; BiGRU; TLM

## 0 引言

随着工业信息化的发展, 工业过程积累了大量与产品质量、设备性能和过程故障相关的时序数据<sup>[1]</sup>。工业过程的多元时序数据可以展现多变量工业过程的动态演化过程, 对其进行有效处理可以提取到丰富的多层次特征信息<sup>[2]</sup>, 为工业过程产品质量优化、设备性能改进和过程故障诊断提供数据支撑。

传统的机器学习中一些方法只能进行单特征时序预测, 如差分移动平均自回归和高斯过程; 还有一些则以特征间的线性关系假设为前提进行时序预

测, 如向量自回归和支持向量自回归; 另外一些则只依赖前一状态信息预测下一状态, 无法有效捕捉长时序特征, 如隐马尔可夫模型。因此传统机器学习方法难以对多元复杂高耦合的工业过程故障时序进行有效预测。

当前工业过程的时序预测多采用神经网络进行, 如循环神经网络(Recurrent Neural Network, RNN)、卷积神经网络(Convolution Neural Network, CNN)和Transformer模型等神经网络。但是目前的很多研究集中于对工业过程影响较大的特征预测上, Li等学者<sup>[2]</sup>利用动态时滞重构方法提取工业过程特征间的动态时滞关系, 并结合LSTM对工业蒸

馏过程和甲醇生产过程多元时序中的单个重要特征进行了预测,模型未给出对工业过程所有变量的预测效果。Xie 等学者<sup>[3]</sup>利用改进隐藏状态生成方式的 GRU 模型对聚酯纤维酯化过程影响较大的 MVI 指标进行了预测。Zhai 等学者<sup>[4]</sup>利用 XGBoost 对融合冶金加热炉多元特征进行处理以生成的新特征,并和原特征一起输入 GRU 模型中以预测加热炉温度。

目前对于工业过程所有变量进行预测的研究相对较少,Shi 等学者<sup>[5]</sup>针对图卷积神经网络的平滑假设无法有效处理高动态和噪声环境复杂工业过程的问题,结合 Taubin 算法提出一种混合高通滤波和低频滤波的图卷积神经网络,并在聚酯纤维聚合过程数据集上进行了测试,但提出的模型预测结果误差较大,使用 30 采样预测未来 10 采样数据时 MAPE 误差达到了 50%。Zheng 等学者<sup>[6]</sup>在时序维对图卷积神经网络进行了拓展,综合考虑了滞后一个时间步和滞后两个时间步的采样数据,并以此建立了时空因果发现图卷积网络,最后在聚酯纤维聚合过程数据集上进行了测试。

但是上述对工业过程多元时序的预测研究只考虑了正常运行状态工业过程,未综合考虑复杂多变的故障状态下的设备运行情况,对故障状态工业过程多元时序的预测依然有待进一步研究。因此,本文提出一种基于图注意力的时序双通道图注意力模型(Temporal Dual-Channel Graph Attention Model,

TDCGAT)对工业过程的多元故障时序进行预测。针对工业过程故障时序特征难以提取的问题,提出一种结合多尺度时间卷积(Multi-Scale Temporal Convolutional Network, MSTCN)和双向 GRU(BiGRU)的时序学习模块(Temporal Learning Module, TLM);针对工业过程间的高耦合关系难以提取的问题,提出使用图注意力机制提取特征间的高耦合关系,并提出利用自适应学习的节点嵌入构建节点间的邻接矩阵以简化图注意力计算。

## 1 基于 TDCGAT 的工业过程故障时序预测

### 1.1 模型整体结构

TDCGAT 模型的整体结构如图 1 所示,输入数据  $X \in R^{P \times N \times D}$ ,其中  $P$  为输入数据长度, $N$  为工业过程传感器节点数量, $D$  为每个节点采集的数据维度;输出数据  $Y \in R^{Q \times N}$ ,其中  $Q$  为输出数据长度。模型主要由 TLM 和图注意力模块(Graph Attention Module, GAM)构成,GAM 模块主要由多头 GAT 以及 GAT 的融合过程组成。模型首先将输入数据卷积到指定维度,接着通过串联 TLM 和 GAM 依次捕获特征的时序信息和空间依赖信息。为避免梯度消失,本文在每个 TLM-GAM 模块组之间添加了一维卷积以传递梯度信息。模型的最终输出通过跳层连接输入数据、每个 TLM 的输出数据以及最后一层 GAM 输出数据实现。

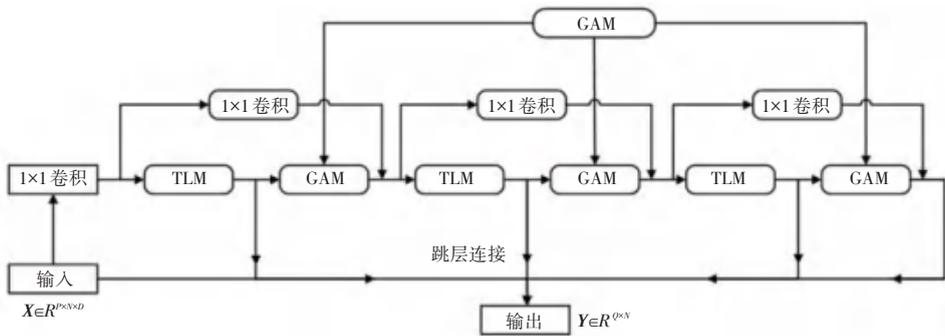


图 1 TDCGAT 模型整体结构

Fig. 1 The overall framework of TDCGAT model

### 1.2 时序学习模块

TLM 通过残差连接的 MSTCN 通道和 BiGRU 通道来捕获时序依赖性,模块整体结构如图 2 所示。

图 2 中,BiGRU 模块为简单的双层 BiGRU,分别提取前向时序信息和反向时序信息,并通过 Hadamard 积进行融合,BiGRU 模块计算过程的数学公式为:

$$y_t = \vec{h}_t \odot \overleftarrow{h}_t \quad (1)$$

其中,“ $\odot$ ”表示 Hadamard 积; $\vec{h}_t$  表示前向层隐藏状态; $\overleftarrow{h}_t$  表示反向层隐藏状态。在 MSTCN 模块中,不同于 Wu 等学者<sup>[7]</sup>使用的  $1 \times 2$ 、 $1 \times 3$ 、 $1 \times 6$  和  $1 \times 7$  卷积核,本文使用  $1 \times 2$ 、 $1 \times 3$ 、 $1 \times 5$  和  $1 \times 9$  卷积核提取时序特征信息,所有卷积核的输出结果经过拼接后输

出。MSTCN 模块的多尺度卷积如图 3 所示。

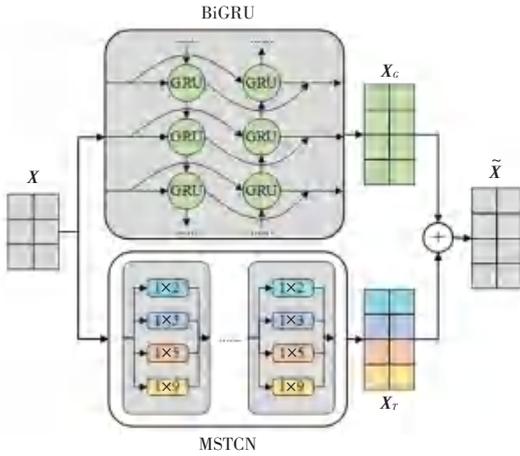


图 2 时序学习模块结构

Fig. 2 The structure of Temporal Learning Module

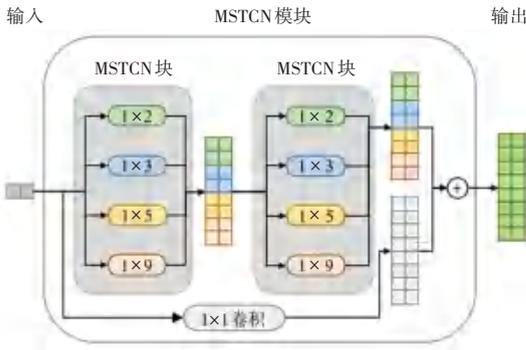


图 3 MSTCN 模块的多尺度卷积

Fig. 3 The multi-scale convolutions of MSTCN module

因此, TLM 模块整体计算过程的数学公式具体如下:

$$X_c = BiGRU(X) \quad (2)$$

$$X_r = MSTCN(X) \quad (3)$$

$$\tilde{X} = \text{Tanh}(X_c \oplus X_r) \quad (4)$$

其中,  $\text{Tanh}(\cdot)$  表示双曲正切函数; “ $\oplus$ ” 表示残差连接;  $X \in R^{P \times h}$ , 这里  $h$  表示经过  $1 \times 1$  卷积后的特征维度,  $P$  表示历史数据长度;  $X_c$ 、 $X_r$  和  $\tilde{X} \in R^{P \times h}$ , 分别表示经过 BiGRU、MSTCN 处理和 TLM 层的输出数据。

### 1.3 图注意力模块

Atwood 等学者<sup>[8]</sup>提出 DCNN (Diffusion - Convolution Neural Network) 将空域的图卷积视为信息扩散过程, 但其计算复杂度过高, 难以拓展到大图上, 因此 Veličković 等学者<sup>[9]</sup>将自注意力机制拓展到图上, 提出了图注意力网络 (Graph Attention Network, GAT), 图注意力可以直接获得全局特征间的关联信息。

由于工业过程特征的时变和高耦合特性, 特征间的邻接矩阵往往难以构建, 本文基于 GAT 提出 GAM 模块, 通过多头 GAT 多层次提取工业流程特征间的复杂关系。原始的 GAT 在隐藏层拼接多头 GAT 以输入到下一层进行计算, 在输出层使用取平均值的方式融合多头 GAT, 而拼接的图注意力会导致数据维度膨胀而增加计算成本, 取平均值的 GAT 对所有注意力头赋予相同权重, 会增加无效信息权重而忽略重要信息, 因此本文利用单层前馈神经网络自适应学习各注意力权重以融合 GAT 的多头注意力, 从而自适应学习重要信息并忽略无效信息。融合 GAT 的计算方法的数学公式可表示为:

$$h_i^k = \text{Concat}(h_i^1, \dots, h_i^K) \quad (5)$$

$$h_i' = \text{LeakyReLU}(h_i^k \cdot \beta_i) \quad (6)$$

其中,  $h_i^m = \sigma(\sum_{j \in N_i} \alpha_{ij}^m W^m h_j) \in R^{F'}$  ( $m \in \{1, 2, \dots, K\}$ ) 表示  $i$  特征的第  $m$  个图注意力;  $h_i^k \in R^{F' \times K}$  由  $K$  个  $i$  特征的图注意力拼接而成;  $\beta_i \in R^K$  表示自适应注意力权重向量, 由前馈连接实现;

$\text{LeakyReLU}$  激活函数对融合后的节点数据进行非线性化处理。

当节点数量过多, 利用 GAT 直接计算节点间的关联关系会导致计算量过大, 模型训练缓慢, 通过在 GAT 中加入额外的邻域节点信息可有效降低计算量, 加快模型训练。但 GAT 存在求解  $i$  节点邻域  $N_i$  的问题, 本文利用可学习的嵌入矩阵构建邻接矩阵  $A^{adj}$ , 具体计算过程见下式:

$$M_1 = \text{Tanh}(E_1(N)) \quad (7)$$

$$M_2 = \text{Tanh}(E_2(N)) \quad (8)$$

$$A = \text{ReLU}(\text{Tanh}(M_1 M_2^T)) \quad (9)$$

$$A_i^{adj}[i, \text{argtopk}(A[i, :])] = 1 \quad (10)$$

$$A^{adj} = [A_1^{adj}, A_2^{adj}, \dots, A_N^{adj}] \quad (11)$$

其中,  $N$  表示节点数量;  $E_1$  和  $E_2 \in R^{N \times E}$  表示随机初始化的节点嵌入矩阵;  $E$  表示节点嵌入维度;  $A \in R^{N \times N}$  表示节点邻域关系矩阵;  $A_i^{adj}$  表示保留了  $k$  个最大相关节点的  $i$  节点列邻接向量, 当存在  $i$  节点对  $j$  节点的单向关系时  $A_{ij}^{adj} = 1$ ; 所有  $N$  个节点的列邻接向量组成邻接矩阵  $A^{adj}$ 。

因此 GAM 模块主要过程如图 4 所示, 首先利用随机初始化的节点嵌入生成可学习的邻接矩阵  $A^{adj}$ , 接着将邻接矩阵和输入数据输入到多头图注意力中计算邻域节点权重并根据邻域节点生成多个候选节点, 然后通过前馈连接融合多头图注意力生成的多个新节点以生成下一节点数据。

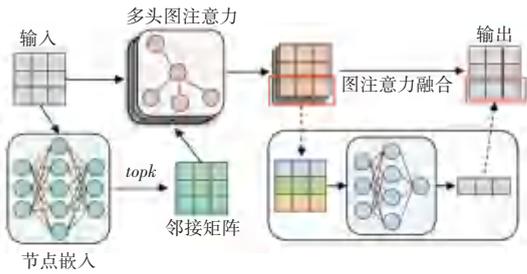


图 4 GAM 主要过程

Fig. 4 The main process of GAM

## 2 实验及结果分析

### 2.1 田纳西伊斯曼过程数据集

本文模型在田纳西伊斯曼过程 (Tennessee Eastman Process, TEP) 数据集上进行了测试, TEP 是伊斯曼化工公司创建的模拟工业过程模型<sup>[10]</sup>, 其多元特征具有强耦合和时变等非线性特征<sup>[11]</sup>, 广泛应用于评价模型对多元复杂工业过程故障的诊断能力。该仿真模型由 Downs 等学者<sup>[12]</sup> 提出, 并由 Ricker<sup>[13]</sup> 进行了优化。本文使用 Rieth 等学者<sup>[14]</sup> 提出的扩展的 TEP 数据集进行训练, 该数据集使用随机种子生成了更多样本。

TEP 有 21 种运行状态, 包括 1 个正常运行状态和 20 种故障状态, 故障状态中包含了不同类型的过程故障, 所有运行状态的详细信息见表 1。

表 1 TEP 的不同运行状态

Table 1 Different operation status of TEP

故障编号	故障描述	故障类型
0	正常运行状态	无
1	A/C 物料进料流量比改变, B 组分含量不变	阶跃
2	B 组分含量改变, A/C 进料流量比不变	阶跃
3	物料 D 温度改变	阶跃
4	反应器冷却水入口温度改变	阶跃
5	冷凝器冷却水入口温度改变	阶跃
6	物料 A 进料损失	阶跃
7	物料 C 压力损失	随机变量
8	物料 A、B 和 C 的组成改变	随机变量
9	物料 D 温度改变	随机变量
10	物料 C 温度改变	随机变量
11	反应器冷却水入口温度改变	随机变量
12	冷凝器冷却水温度改变	随机变量
13	反应动力学特性改变	缓慢漂移
14	反应器冷却水阀门粘滞	粘滞
15	冷凝器冷却水阀门粘滞	粘滞
16~20	未知	未知

TEP 数据集的 52 个特征包括 22 个过程测量值、19 个成分测量值和 11 个操作变量。其中, 过程测量值以 3 min 为间隔采样, 成分测量值和操作变量采样间隔不固定, TEP 的过程测量值、成分测量值和操作变量的详细信息分别见表 2~表 4。

表 2 TEP 的过程测量值

Table 2 Process measurements of TEP

变量	描述	单位
XMEAS(1)	A 进料(流 1)	km <sup>3</sup> /h
XMEAS(2)	D 进料(流 2)	kg/h
XMEAS(3)	E 进料(流 3)	kg/h
XMEAS(4)	A、C 进料(流 4)	10 <sup>3</sup> · m <sup>3</sup> /h
XMEAS(5)	再循环流量(流 8)	10 <sup>3</sup> · m <sup>3</sup> /h
XMEAS(6)	反应器进料速度(流 6)	10 <sup>3</sup> · m <sup>3</sup> /h
XMEAS(7)	反应器压力	kPa(表值)
XMEAS(8)	反应器等级	%
XMEAS(9)	反应器温度	°C
XMEAS(10)	排放速度(流 9)	km <sup>3</sup> /h
XMEAS(11)	产品分离器温度	°C
XMEAS(12)	产品分离器液位	%
XMEAS(13)	产品分离器压力	kPa(表值)
XMEAS(14)	产品分离器塔底低流量(流 10)	m <sup>3</sup> /h
XMEAS(15)	汽提器等级	%
XMEAS(16)	汽提器压力	kPa(表值)
XMEAS(17)	汽提器塔底流量(流 11)	m <sup>3</sup> /h
XMEAS(18)	汽提器温度	°C
XMEAS(19)	汽提器流量	kg/h
XMEAS(20)	压缩机功率	kW
XMEAS(21)	反应器冷却水出口温度	°C
XMEAS(22)	分离器冷却水出口温度	°C

TEP 数据集中, 操作变量设置值为 0~100, 单位为“%”, 是设置的操作变量值占最大值的百分比。

由于 TEP 数据集不同特征取值范围差异较大, 如特征 XMEAS(3), 即过程测量值 E 进料的取值范围为 3 540.7~5 175.8 kg/h, 而特征 XMEAS(37), 即流 11 的成分测量值成分 D, 取值范围为 0.001~0.067 mol%, 如果不对数据进行预处理会导致模型在进行时序预测时偏向优化大数值特征而忽略小数值特征, 因此本文对 TEP 数据集的各个特征分别进行均值-方差归一化以统一各个特征数据尺度, 经过归一化的 XMEAS(3) 取值范围为 -8.871~6.180, XMEAS(37) 的取值范围为 -4.534~4.811, 数据尺度差异显著缩小。

表3 TEP的成分测量值

Table 3 Component measurement of TEP

变量	描述	流	采样间隔/min	变量	描述	流	采样间隔/min
XMEAS(23)	成分 A	6	6	XMEAS(33)	成分 E	9	6
XMEAS(24)	成分 B	6	6	XMEAS(34)	成分 F	9	6
XMEAS(25)	成分 C	6	6	XMEAS(35)	成分 G	9	6
XMEAS(26)	成分 D	6	6	XMEAS(36)	成分 H	9	6
XMEAS(27)	成分 E	6	6	XMEAS(37)	成分 D	11	15
XMEAS(28)	成分 F	6	6	XMEAS(38)	成分 E	11	15
XMEAS(29)	成分 A	9	6	XMEAS(39)	成分 F	11	15
XMEAS(30)	成分 B	9	6	XMEAS(40)	成分 G	11	15
XMEAS(31)	成分 C	9	6	XMEAS(41)	成分 H	11	15
XMEAS(32)	成分 D	9	6				

注:成分测量值的单位均为 mol%。

表4 TEP的操作变量

Table 4 Manipulated variables of TEP

特征	描述	最小值	最大值	单位
XMV(1)	D 进料量(流 2)	0	5 811.000	kg/h
XMV(2)	E 进料量(流 3)	0	8 354.000	kg/h
XMV(3)	A 进料量(流 1)	0	1.017	$10^3 \cdot \text{m}^3/\text{h}$
XMV(4)	A,C 进料(流 4)	0	15.250	$10^3 \cdot \text{m}^3/\text{h}$
XMV(5)	压缩机再循环阀	0	100	%
XMV(6)	排放阀(流 9)	0	100	%
XMV(7)	分离罐液流量(流 10)	0	65.710	$\text{m}^3/\text{h}$
XMV(8)	汽提器液体产品流量(流 11)	0	49.100	$\text{m}^3/\text{h}$
XMV(9)	汽提器水流阀	0	100	%
XMV(10)	反应器冷却水流量	0	227.100	$\text{m}^3/\text{h}$
XMV(11)	冷凝器冷却水速度	0	272.600	$\text{m}^3/\text{h}$

## 2.2 实验环境设置

本文中所有实验在 Windows10 系统及 Pytorch 1.12.0+cu116 环境下,使用 PyCharm 2020.2 完成,CPU 为 AMD 5600G@4.4 GHz, GPU 为 NVIDIA RTX 2060S@8 GB,内存为 128 G@3 200 MHz,硬盘为致钛 PC005@1 TB。

模型使用 TEP 数据集的训练集进行训练,使用 24 个历史采样数据预测未来 24 个采样数据。数据集以 6:2:2 的比例划分训练集、验证集和测试集并随机打乱,模型的训练损失通过对预测数据的反均值-方差归一化计算实现。

模型使用 Adam 算法进行优化。在进行反归一化计算模型训练损失时,如果使用平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE) 计算损失,则不同尺度的特征对训练损失的贡献各不相同,数据尺度大的特征会对模型训练损失做出更大贡献,因此有必要使用统一尺度的损失函数计算模型训练损失。本文使

用去百分比的 MAPE 函数计算训练损失以优化模型,模型使用 MAE、RMSE 和 MAPE 指标评估模型对工业过程故障多元时序的预测性能。MAE、RMSE 和 MAPE 的计算方法的数学公式分别如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (13)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (14)$$

## 2.3 模型参数设置

TDCGAT 模型参数见表 5。模型首先通过  $1 \times 1$  卷积将输入的 TEP 数据在特征维映射到高维以输入 TLM 中。TLM 层中,两层 BiGRU 的输入和输出均为 128 维,MSTCN 将输入 128 维特征划分成 4 份,分别输入到  $1 \times 2$ 、 $1 \times 3$ 、 $1 \times 5$  和  $1 \times 9$  卷积核的双通道 TCN 中,然后将每个 TCN 的 64 维输出拼接为 256 维并通过  $1 \times 1$  卷积为 128 维,最后将 MSTCN 层输出与 GRU

的输出取 Hadamard 积并通过 Tanh 函数激活作为 TLM 模块的输出。GAM 模块中首先根据给定的 128 个节点生成可学习的邻接矩阵  $A^{adj}$ , 再取邻接矩阵每个节点的 16 个邻域节点求取 4 头图注意力, 最后以融合的图注意力结果输出。残差连接层通过  $1 \times 1$  卷积将每个 TLM-GAM 模块组前后的数据进行残差连接以跨层传递历史信息, 加快模型拟合。跳层连接对输入数据使用  $1 \times 1$  卷积将特征维卷积到 128 维以与其他层输入数据相加, 最后的输出通过  $1 \times 1$  卷积将特征维卷积到 52 维实现。

表 5 TDCGAT 参数设置

Table 5 Parameters' setting of TDCGAT

结构	模块/参数	设置值
输入卷积	$1 \times 1$ 卷积	104→128
TLM	BiGRU	128→128
	MSTCN	$4 \times (32 \rightarrow 64)$
GAM	节点嵌入 $E$	$R^{128 \times 64}$
	邻接矩阵 $A^{adj}$	$R^{128 \times 128}$
	邻域节点数量 $N$	16
	注意力头 $K$	4
	特征变换矩阵 $W$	$R^{128 \times 128}$
残差连接	$1 \times 1$ 卷积	128→128
跳层连接	输入 $1 \times 1$ 卷积	104→128
	输出 $1 \times 1$ 卷积	128→52

## 2.4 对比模型

本文将模型与经典的多元时序预测模型进行了对比, 包括 STID、MTGNN、STGNN、BiGRU 和 TCN 等模型, 其中 STID 是由 Shao 等学者<sup>[15]</sup>提出的简单基准多元时序预测模型。该模型对输入数据附加了额外的空间和时间信息嵌入, 并利用残差连接的三层全连接层进行

多元时序预测。MTGNN 由 Wu 等学者<sup>[16]</sup>提出, 通过在堆叠的多尺度时间卷积层中插入共享的图嵌入学习层实现多元时序预测。STGNN 由 Wang 等学者<sup>[17]</sup>提出, 首先通过对每个时间采样应用谱域的图卷积提取特征的空间相关性, 接着利用 GRU 提取特征间的局部时序依赖, 最后通过 Transformer 提取特征的全局时序依赖并通过全连接层实现多元时序预测。STID、MTGNN 和 STGNN 相关参数均与其开源模型默认设置保持一致。BiGRU 为普通的双层 GRU, 隐藏层单元 128 个; TCN 为 TDCGAT 中的 MSTCN 模块, 堆叠层数与 TDCGAT 中保持一致。

## 2.5 模型对比实验

TDCGAT 模型与其他模型对比实验结果见表 6。表 6 中, “@  $N$ ” 为模型在  $N$  长度处的预测情况。加粗字体为最佳预测结果, 下划线字体为次优预测结果。由表 6 可见, TDCGAT 模型在所有长度预测任务上均取得了最佳预测效果。加入时间和空间编码的简单 STID 模型超过了 STGNN、MSTCN 和 BiGRU 模型, 并且在预测 3 和 6 长度的 MAPE 获得了次优结果, 但整体逊于 MTGNN 模型。MTGNN 在大多数预测长度上取得了次优结果, 但逊于 TDCGAT 模型。MTGNN 的图卷积仅考虑单个邻域节点, 无法有效处理高耦合的工业过程多元特征。TDCGAT 在单层图注意力中考虑多邻域节点, 通过多次图注意力可保证对多数节点的有效覆盖, 保证了对远距离节点隐藏关联关系的有效提取。此外, 由表 4~6 可见, BiGRU 模型对于多元时序预测整体略优于 MSTCN 模型, MTGNN 中仅使用 MSTCN 进行时序预测, 未结合使用 BiGRU。

表 6 各个模型在 TEP 数据集上的实验结果

Table 6 Experimental results of different models on TEP dataset

指标		STID	MTGNN	STGNN	MSTCN	BiGRU	TDCGAT
MAPE/%	@ 3	2.562 3	2.741 5	2.844 9	4.597 8	4.567 9	<b>2.391 9</b>
	@ 6	2.938 5	2.954 6	3.261 5	4.859 5	4.694 5	<b>2.609 1</b>
	@ 12	3.455 6	3.226 4	3.991 8	5.377 3	4.847 2	<b>2.734 7</b>
	@ 24	4.827 2	4.014 5	5.923 7	6.521 9	6.545 6	<b>3.253 5</b>
MAE	@ 3	1.752 3	1.609 6	1.844 4	2.562 4	2.582 2	<b>1.309 8</b>
	@ 6	1.960 1	1.747 1	2.033 9	2.803 0	2.630 2	<b>1.569 7</b>
	@ 12	2.334 4	1.934 3	2.380 9	3.083 0	2.736 8	<b>1.728 9</b>
	@ 24	2.998 9	2.377 9	3.012 7	3.540 5	3.128 0	<b>2.069 8</b>
RMSE	@ 3	7.506 6	6.624 9	7.475 5	9.117 2	8.914 7	<b>5.707 8</b>
	@ 6	8.153 1	7.015 9	7.944 9	9.820 2	9.357 2	<b>6.215 8</b>
	@ 12	9.566 4	7.418 7	8.973 6	10.826 0	9.943 3	<b>7.009 0</b>
	@ 24	10.994 5	9.108 0	11.472 8	12.691 9	11.034 9	<b>8.143 4</b>

## 2.6 各故障和各特征预测情况

为详细探究模型预测性能, 本节实验探究了 TDCGAT 模型对于各类型故障和各个特征的预测情

况。模型对各故障类型的预测情况见表 7。表 7 中, 下划线数据为所有故障中的最大值, 加粗数据为所有故障中的最小值。模型对数据集各特征预测情况见表 8。

表7 TDCGAT模型对不同故障的MAPE、MAE、RMSE

Table 7 MAPE、MAE、RMSE of TDCGAT model on different faults

故障	MAPE/ %				MAE				RMSE			
	@ 3	@ 6	@ 12	@ 24	@ 3	@ 6	@ 12	@ 24	@ 3	@ 6	@ 12	@ 24
0	<b>2.177</b>	<b>2.310</b>	<b>2.336</b>	<b>2.391</b>	<u>1.591</u>	<u>1.810</u>	<u>1.956</u>	2.108	<u>7.087</u>	<u>7.914</u>	<u>8.172</u>	<u>8.468</u>
1	2.391	2.610	2.789	3.262	1.344	1.598	1.771	2.125	5.789	6.756	7.075	8.229
2	2.433	2.611	2.812	3.232	1.342	1.605	1.772	2.121	5.795	6.781	7.067	8.283
3	2.386	2.629	2.719	3.264	1.343	<b>1.589</b>	1.756	2.107	5.795	6.730	<b>7.025</b>	8.176
4	2.398	2.625	2.813	3.218	1.342	1.607	1.777	2.125	5.807	6.844	7.160	8.307
5	2.410	2.605	2.752	3.321	1.340	1.600	1.767	2.132	5.785	6.773	7.043	8.267
6	2.441	2.643	2.804	3.389	1.334	1.597	1.761	2.106	5.770	6.777	7.049	8.178
7	2.390	2.646	2.784	3.282	1.346	1.612	1.780	<u>2.146</u>	5.787	6.814	7.072	8.339
8	2.387	2.624	2.789	3.263	<b>1.331</b>	1.591	<b>1.755</b>	<b>2.097</b>	5.768	6.771	7.041	<b>8.168</b>
9	2.436	2.623	2.770	3.355	1.337	1.595	1.766	2.112	5.771	6.756	7.084	8.249
10	2.417	2.592	2.783	3.201	1.346	1.610	1.780	2.132	5.793	6.793	7.097	8.241
11	2.441	2.674	2.822	3.345	1.341	1.602	1.774	2.121	5.804	6.787	7.102	8.220
12	2.443	2.656	2.811	3.202	1.332	1.591	1.765	2.116	5.755	6.732	7.056	8.249
13	<u>2.501</u>	<u>2.684</u>	<u>2.856</u>	<u>3.410</u>	1.340	1.605	1.765	2.115	5.794	6.814	7.059	8.239
14	2.450	2.683	2.799	3.287	1.333	1.593	1.765	2.114	<b>5.744</b>	<b>6.728</b>	7.054	8.230
15	2.406	2.639	2.789	3.245	1.338	1.598	1.766	2.128	5.793	6.779	7.053	8.265
16	2.399	2.615	2.786	3.312	1.335	1.597	1.761	2.113	5.754	6.755	7.028	8.213
17	2.399	2.621	2.783	3.295	1.340	1.599	1.767	2.127	5.790	6.759	7.031	8.257
18	2.428	2.656	2.852	3.244	1.344	1.606	1.768	2.133	5.798	6.790	7.068	8.301
19	2.427	2.649	2.826	3.285	1.342	1.598	1.770	2.124	5.789	6.762	7.053	8.275
20	2.415	2.668	2.791	3.173	1.334	1.595	1.768	2.121	5.761	6.766	7.075	8.227

表8 TDCGAT模型对52个特征预测值的MAPE、MAE、RMSE

Table 8 MAPE、MAE、RMSE of the prediction values of TDCGAT model on 52 features

特征	MAPE / %	MAE	RMSE	特征	MAPE / %	MAE	RMSE
1	<u>12.834 8</u>	0.025 3	0.036 6	27	1.193 4	0.222 7	0.282 8
2	0.687 1	<u>25.167 7</u>	<u>32.578 6</u>	28	1.411 8	0.022 8	0.029 0
3	0.609 5	<u>27.477 1</u>	<u>35.512 6</u>	29	0.801 1	0.260 1	0.378 6
4	0.773 3	0.072 2	0.102 0	30	0.701 7	0.096 8	0.128 5
5	0.649 8	0.174 8	0.230 9	31	1.134 5	0.273 6	0.406 1
6	0.420 2	0.178 0	0.228 4	32	7.000 4	0.085 8	0.108 4
7	<b>0.172 7</b>	4.680 0	9.117 0	33	1.291 4	0.237 0	0.305 5
8	0.592 0	0.443 5	0.586 5	34	1.043 7	0.023 0	0.029 3
9	<b>0.026 7</b>	0.032 1	0.070 6	35	1.201 5	0.057 1	0.080 8
10	3.411 3	<b>0.011 7</b>	<b>0.016 9</b>	36	2.121 6	0.047 6	0.062 2
11	0.282 1	0.225 2	0.429 6	37	<u>29.763 3</u>	<b>0.008 1</b>	<b>0.010 3</b>
12	1.569 1	0.783 7	0.994 8	38	1.684 1	0.014 1	0.022 8
13	0.192 7	<u>5.082 0</u>	<u>9.637 4</u>	39	9.517 8	<b>0.008 9</b>	<b>0.011 3</b>
14	3.193 3	0.800 0	1.021 4	40	0.738 6	0.396 9	0.508 1
15	1.601 7	0.799 3	1.018 1	41	0.911 5	0.398 8	0.512 6
16	<b>0.145 6</b>	4.523 8	8.187 7	42	0.744 2	0.473 6	0.607 6
17	2.135 8	0.489 1	0.627 0	43	0.676 1	0.365 1	0.467 8
18	0.414 7	0.272 9	0.446 7	44	<u>13.373 7</u>	2.041 9	2.977 2
19	2.829 5	3.308 2	5.495 4	45	1.606 1	0.991 2	1.270 9
20	0.295 1	1.000 8	1.795 5	46	2.466 5	0.458 2	1.084 6
21	0.217 5	0.200 9	1.591 7	47	9.663 5	1.260 2	1.715 3
22	0.368 9	0.282 7	0.521 8	48	6.160 1	2.316 4	2.928 2
23	0.747 9	0.238 4	0.324 9	49	4.010 9	1.852 6	2.359 1
24	1.035 3	0.091 9	0.118 3	50	3.858 1	0.698 2	1.104 6
25	0.927 8	0.246 6	0.340 6	51	1.932 0	0.725 5	1.299 8
26	1.276 2	0.087 7	0.110 0	52	6.587 0	1.172 7	1.500 7

由表 7 可知, TDCGAT 模型对于工业过程不同故障时序的预测性能相差不大。从表 7 中可见, 模型对 TEP 过程正常运行状态预测值的 *MAPE* 显著低于其他故障类型, 且对预测长度不敏感, @ 24 的 *MAPE* 与 @ 6 的 *MAPE* 差异较小, 但模型对正常状态预测值的 *MAE* 和 *RMSE* 显著高于其他故障类型, 说明模型对正常状态特征的预测值与真实数据偏差更大, 但总体误差保持相对平稳。

以图 5 中故障 0 和故障 1 对 42 号特征 XMV

(1)、即操作变量 D 进料量的预测情况为例, 故障 0 的 D 进料量总体保持在 62~64 kg/h 之间震荡, 随机震荡幅度较小且整体趋势平稳, 模型可以准确把握进料量的整体趋势, 预测误差仅和时序震荡情况相关, 因此在不同长度处的预测误差都相对平稳; 模型也能捕捉故障 1 的 42 号特征的整体运行趋势, 但相比远处, 模型对近处的预测效果更好, 因此对故障 1 预测值的误差在近处更低, 对远处预测值的误差相对更大。

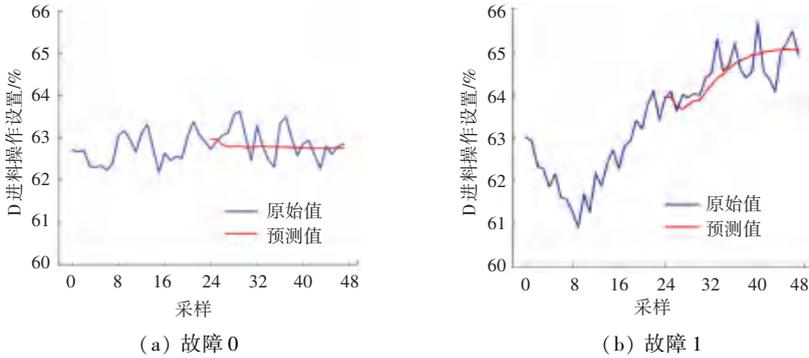


图 5 故障 0 和故障 1 对特征 42 的预测情况

Fig. 5 Prediction results of fault 0 and fault 1 on feature 42

模型对所有 52 个特征的预测情况, 具体见表 8。表 8 中, 分别以加粗和下划线标记所有特征 *MAPE*、*MAE* 和 *RMSE* 指标中最小和最大的 3 个值。

由表 8 可知, 模型能较好预测大部分特征, 但部分特征预测值的 *MAE*、*RMSE* 和 *MAPE* 的误差较大。 *MAE* 和 *RMSE* 较大的特征包括特征 2、3、13 和 16, 分别对应过程测量值 D 进料、E 进料、产品分离

器压力和汽提器压力, 但这几项特征的 *MAPE* 均较小; *MAPE* 较大的特征包括特征 1、37、44、47 等, 分别对应过程测量值 A 进料、成分测量值成分 E、操作变量 A 进料量和排放阀, 且几个特征的 *MAE* 和 *RMSE* 均较大。接下来分别对 2 种预测情况进行具体分析, *MAE* 和 *RMSE* 较大的特征的预测情况如图 6 所示, *MAPE* 较大的特征预测情况如图 7 所示。

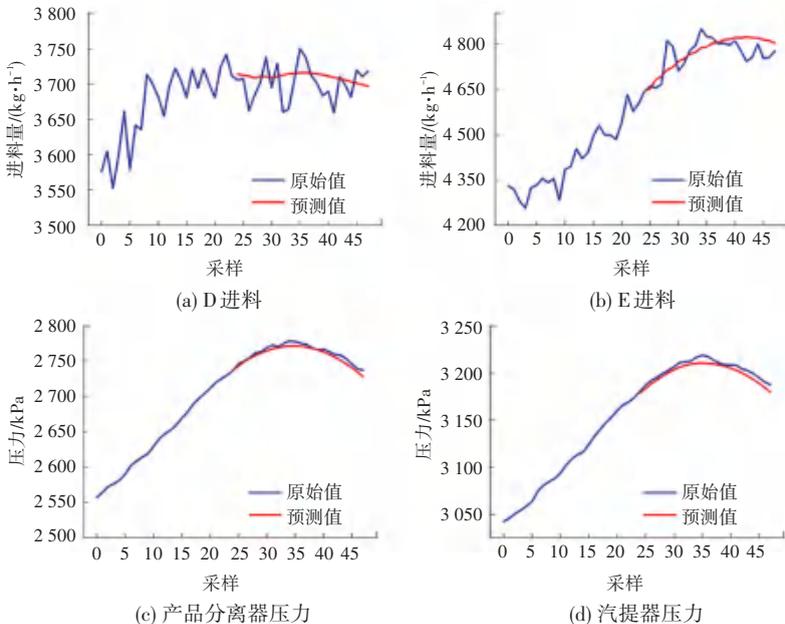


图 6 TDCGAT 模型对故障 2 特征 2、3、13、16 的预测情况

Fig. 6 The prediction results of TDCGAT model of fault 2 on feature 2, 3, 13, 16

由图6可知,特征2、3、13和16依次对应(a)~(d)四个子图,所有特征的数值变化较大,尤其是特征2和特征3,有较大幅度震荡,但TDCGAT模型能捕捉特征的整体时序趋势。由于这几项特征数值较大,所以即使模型预测值与真实值存在一定偏差也不会导致较大MAPE;另一方面,由于特征数值大且存在一定震荡,因此即使预测值的MAPE较小也会导致预测值和真实值的MAE和RMSE误差较大,所以如果使用MAE作为损失函数来优化模型会导致模型偏向优化较大数值特征。

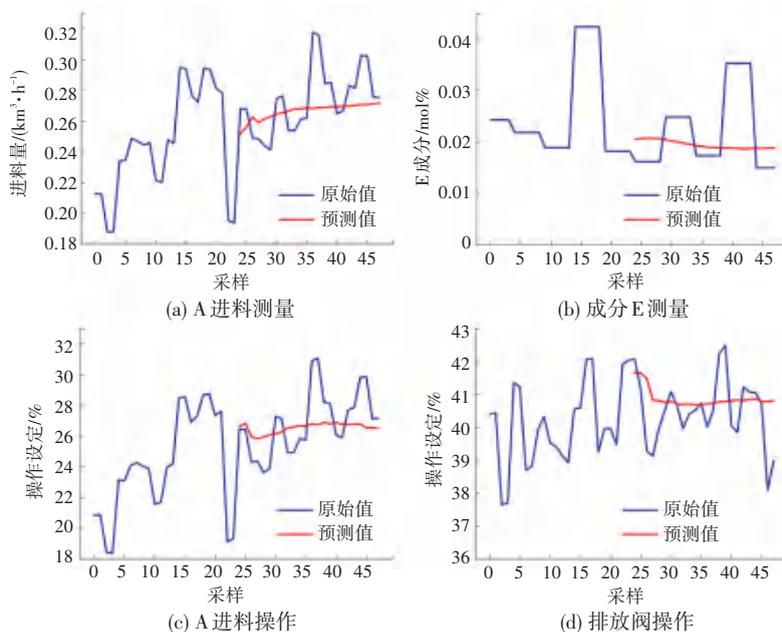


图7 TDCGAT模型对故障2特征1、37、44、47的预测情况

Fig. 7 The prediction results of TDCGAT model of fault 2 on feature 1, 37, 44, 47

综上所述,在TEP数据集中,不同特征数值差异巨大,使用MAE和RMSE作为模型的损失函数会导致模型偏向优化大数值特征,而使用MAPE作为损失函数可以消除特征数值尺度的影响,降低模型对大数值特征的优化偏好。

### 3 结束语

对工业过程故障复杂高耦合的多元时序数据的预测一直是研究难点。本文在图注意力基础上,结合MSTCN和BiGRU,提出针对工业过程故障多元时序数据的预测模型,并且针对原始图注意力中的多头注意力融合可能会忽略重要特征信息的问题,利用自适应前馈连接对学习图注意力权重以融合图注意力。针对工业过程多元特征数据尺度不同的问题,提出对各个特征分别做归一化处理,并利用取百分比的MAPE损失计算模型损失以降低特征数据尺

度影响。由图7可知,特征1、37、44和47分别对应图7中(a)~(d)四个子图,图7中所有特征均剧烈抖动,其中特征1和特征37数值较小,特征44和47数值相对较大,但远小于图6中的特征。由于特征44是特征1的操作变量但数据尺度不同,所以2个特征的MAE和RMSE损失差距较大,但使用MAPE损失函数来优化模型不会导致模型偏向优化大数值特征,所以模型对2个特征的预测情况类似,效果均较差。特征37的预测情况较差,但由于数值较小,所以预测值的MAE和RMSE在所有特征中甚至是最小值。

度影响。在TEP数据集上的测试表明,本文提出的模型对工业过程故障的多元时序数据具有较好预测效果,对工业过程故障的研究具有重要现实意义。

### 参考文献

- [1] HU Jie, WU Min, CHEN Xin, et al. A multilevel prediction model of carbon efficiency based on the differential evolution algorithm for the iron ore sintering process[J]. IEEE Transactions on Industrial Electronics, 2018, 65(11): 8778-8787.
- [2] LI Jince, YANG Bo, LI Hongguang, et al. DTDR—ALSTM: Extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models[J]. Knowledge-Based Systems, 2021, 211: 106508.
- [3] XIE Ruimin, HAO Kuangrong, HUANG Biao, et al. Data-driven modeling based on two-stream  $\lambda$  gated recurrent unit network with soft sensor application [J]. IEEE Transactions on Industrial Electronics, 2020, 67(8): 7034-7043.
- [4] ZHAI Naji, YAO Peifu, ZHOU Xiaofeng. Multivariate time series forecast in industrial process based on XGBoost and GRU [C]//2020 IEEE 9th Joint International Information Technology

- and Artificial Intelligence Conference (ITAIC). Piscataway, NJ: IEEE, 2020; 1397–1400.
- [5] SHI Xun, HAO K R, CHEN Lei, et al. Multivariate time series prediction of complex systems based on graph neural networks with location embedding graph structure learning [J]. *Advanced Engineering Informatics*, 2022, 54:101810.
- [6] ZHENG S, HAO K, SHI X, et al. Spatiotemporal causal discovery graph convolutional networks for multivariate time series forecasting of industrial process[C]//2023 IEEE 6<sup>th</sup> Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Piscataway, NJ: IEEE, 2023; 264–271.
- [7] WU Zonghan, PAN Shirui, LONG Guodong, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks [C]//Proceedings of the 26<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2020; 753–763.
- [8] ATWOOD J, TOWSLEY D. Diffusion – convolutional neural networks[C]//Proceedings of the 30<sup>th</sup> International Conference on Neural Information Processing Systems. Cambridge, UK: MIT Press, 2016; 2001–2009.
- [9] VELICKOVIĆ P, CUCRURLL G, CASANOVA A, et al. Graph attention networks [C] //International Conference on Learning Representations( ICLR 2018). Ithaca, USA: arXiv.org, 2018; 1–12.
- [10] REINARTZ C, KULAHCI M, RAVN O. An extended tennessee eastman simulation dataset for fault–detection and decision support systems[J]. *Computers & Chemical Engineering*, 2021, 149: 107281.
- [11] YIN S, DING S X, HAGHANI A, et al. A comparison study of basic data–driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process [J]. *Journal of Process Control*, 2012, 22(9): 1567–1581.
- [12] DOWNS J J, VOGEL E F. A plant – wide industrial process control problem[J]. *Computers & Chemical Engineering*, 1993, 17(3): 245–255.
- [13] RICKER N L. Optimal steady–state operation of the Tennessee Eastman challenge process [J]. *Computers & Chemical Engineering*, 1995, 19(9): 949–959.
- [14] RIETH C A, AMSEL B D, TRAN R, et al. Issues and advances in anomaly detection evaluation for joint human–automated systems [C]//International Conference on Applied Human Factors and Ergonomics. Berlin, Germany: Springer–Verlag, 2017; 52–63.
- [15] SHAO Zezhi, ZHANG Zhao, WANG Fei, et al. Spatial – temporal identity: A simple yet effective baseline for multivariate time series forecasting [C]//Proceedings of the 31<sup>st</sup> ACM International Conference on Information & Knowledge Management. New York, USA: ACM, 2022; 4454–4458.
- [16] WU Zonghan, PAN Shirui, LONG Guodong, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks [C]//Proceedings of the 26<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2020; 753–763.
- [17] WANG Xiaoyang, MA Yao, WANG Yiqi, et al. Traffic flow prediction via spatial temporal graph neural network [C]//Proceedings of the Web Conference 2020 ( WWW' 20 ). New York, USA: ACM, 2020; 1082–1092.