

基于增强型多尺度图 Transformer 的工业不规则多变量 时间序列分类

顾岩, 金锋[†], 赵珺, 王伟

(大连理工大学 控制科学与工程学院, 辽宁 大连 116024)

摘要: 多变量时间序列分类在工业状态识别与决策支持中发挥关键作用, 但异步采样与随机缺失形成的不规则数据严重削弱了其分类精度. 基于插补的方法易引入噪声和伪影, 而直接建模又易使关键信息被稀释或扭曲. 鉴于此, 提出一种基于增强型多尺度图 Transformer 的“序列-图像-图”统一建模方法. 首先, 增强型多通道图像转换将不规则序列编码为无需插补的 RGB 图像, 直接利用数值动态、缺失模式与采样信息. 其次, 动态扇形图构建将图像块映射为图节点, 在局部扇形邻域内自适应建立邻接边, 以覆盖跨时间和跨变量依赖并减少冗余. 最后, 自适应多尺度相对图卷积, 在邻居聚合中引入多尺度差分特征, 并结合注意力加权突出关键邻居, 从而缓解图卷积过平滑. 实验在钢铁企业高炉煤气数据及四个公开不规则基准上验证所提出方法在多种不规则模式下的有效性.

关键词: 工业不规则数据; 多变量时间序列分类; 增强型图像; 动态扇形图; 相对图卷积; 图 Transformer
中图分类号: TP311.13 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.1199

引用格式: 顾岩, 金锋, 赵珺, 等. 基于增强型多尺度图 Transformer 的工业不规则多变量时间序列分类 [J]. 控制与决策.

Industrial irregular multivariate time series classification based on an enhanced multi-scale graph Transformer

GU Yan, JIN Feng[†], ZHAO Jun, WANG Wei

(School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Multivariate time series classification plays a key role in industrial state identification and decision support, but irregular data formed by asynchronous sampling and random missingness severely degrades its classification accuracy. Imputation-based methods are prone to introducing noise and artifacts, whereas direct modeling tends to dilute or distort critical information. This paper proposes a unified “sequence-image-graph” modeling method based on an enhanced multi-scale graph Transformer. First, enhanced multi-channel image Transformation encodes irregular sequences into imputation-free RGB images, directly leveraging value dynamics, missingness patterns, and sampling information. Second, dynamic sector graph construction maps image patches to graph nodes and adaptively builds adjacency edges within local sector-based neighborhoods to cover cross-time and cross-variable dependencies while reducing redundancy. Finally, adaptive multi-scale relative graph convolution introduces multi-scale difference features into neighbor aggregation and combines attention weighting to emphasize critical neighbors, thereby alleviating over-smoothing in graph convolution. Experiments on a blast furnace gas dataset from a steel enterprise and four public irregular benchmarks validate the effectiveness of the proposed method under diverse irregular patterns.

Keywords: industrial irregular data; multivariate time series classification; enhanced image; dynamic sector graph; relative graph convolution; graph Transformer

0 引言

多变量时间序列 (Multivariate time series, MTS) 分类在状态监测^[1]、故障诊断^[2]、气体识别^[3]等工业

领域具有重要作用. 然而, 实际工业场景中采集到的数据往往是不规则 MTS^[4], 其采样时间不均、变量间不同步, 并伴随缺失. 若直接输入传统模型, 常因插

收稿日期: 2025-11-20; 录用日期: 2026-02-19.

基金项目: 国家科技重大专项项目 (2025ZD1601800); 国家自然科学基金项目 (62125302, 62394344); 大连市科技人才创新支持计划项目 (2022RG03); 中央高校基本科研业务费专项资金项目 (DUTZD25108); 教育部基础学科和交叉学科突破计划项目 (JYB2025XDXM414).

[†]通信作者. E-mail: jin_feng@dlut.edu.cn.

补伪影、特征丢失难以准确表征复杂的动态规律,严重制约了分类性能和决策可靠性.因此,如何在充分考虑工业场景固有属性的前提下,有效处理不规则数据,对提升分类决策的准确性和可靠性至关重要.

为了充分挖掘 MTS 中丰富的时空信息,已经有许多模型被提出.例如,长短期记忆网络^[5]等序列模型被广泛用于建模时间依赖性,图神经网络^[6]则侧重于刻画变量之间的交互关系.近年来,基于注意力机制的 Transformer 结构^[7]在序列建模任务中展现出强大能力,并被用于提取时间和交互信息.然而,上述方法都假定采样率和时间戳保持恒定,采用固定大小的数值输入,无法处理时间间隔不均的不规则 MTS.插补是一种直观的处理不规则性的手段,IP-Net^[8]结合了采样强度信息和门控循环单元(Gated Recurrent Unit, GRU),通过插补将不规则 MTS 恢复为规则序列.然而,插补往往会扭曲潜在的不规则模式,并不可避免地引入噪声或伪影,甚至导致异常值,从而削弱模型性能.为避免依赖预处理,GRU-D^[9]、ODE-RNN^[10]等通过衰减机制和连续常微分方程来直接建模不规则动态,但这种基于 RNN 的序列建模并未显式捕捉多个变量间的相关性,难以从全局层面刻画不规则性的复杂特征.

为直接学习不规则 MTS,一些研究将时间依赖和变量交互的特征学习分为两个阶段.例如,SeFT^[11]和 Raindrop^[12]将一个变量的所有样本点组合在一起,再为每个变量提取统一特征,但这种一次性处理的方式未能充分挖掘单个不规则样本中的细粒度信息.为此,mTAND^[13]设计了时间注意力机制,通过时间嵌入建模时间相似性;STraTS^[14]融合时间、数值与稀疏掩码形成上下文嵌入,并利用多头注意力进行信息交互;Warpformer^[15]则将动态时间规整引入 Transformer 架构,统一非均匀时间序列并生成多尺度嵌入;MTSFormer^[16]从局部、时间、空间及不规则性四个视角进行建模,并通过不规则性门控机制自适应利用采样模式信息.尽管这些方法在一定程度上提升了对不规则序列的刻画能力,但它们依然依赖固定的网格或序列结构.在面对采样频率不一致、缺失分布复杂以及跨变量非对齐的工业不规则 MTS 时,仍易造成关键信息被稀释或扭曲,从而限制了对真实动态模式的准确建模.

针对上述问题,本文提出了一种基于增强型多尺度图 Transformer(Enhanced Multi-scale Graph Transformer, EMGT)的“序列-图像-图”统一建模框架.其中,增强型多通道图像转换方法(Enhanced Multi-channel Image Transformation, EMIT)将不规

则信息结构化编码,以避免插补扭曲.自适应多尺度图 Transformer 通过动态扇形图构建(Dynamic Sector Graph Construction, DSGC)将图像划分为块并映射为节点,在局部扇形邻域内自适应连边,从而覆盖跨时间和跨变量依赖并减少冗余连接.进而,自适应多尺度相对图卷积(Adaptive Multi-scale Relative Graph Convolution, AMRGC)在邻居聚合中引入多尺度差分特征,并结合逐维差值与注意力加权突出关键邻居,以捕捉局部与全局时空关系并缓解过平滑.在实际工业数据与公开基准上进行实验,结果表明 EMGT 在工业不规则 MTS 分类中的有效性.

1 相关理论

1.1 不规则多变量时间序列定义

设数据集 $\mathcal{D} = \{(\mathbf{X}_i, y_i) \mid i = 1, \dots, B\}$, 其中, B 表示样本数, \mathbf{X}_i 为第 i 个 MTS, $y_i \in \{1, \dots, C\}$ 为对应的类别标签, C 为类别总数. 一个 \mathbf{X}_i 最多包含对 N 个变量的观测(部分变量可能完全缺失). 对于变量 n , 其观测数据由带时间戳和值的元组序列表示为 $[(t_1^n, v_1^n), (t_2^n, v_2^n), \dots, (t_{T_n}^n, v_{T_n}^n)]$, 其中, T_n 表示变量 n 的观测点数量. 若不同变量或不同样本间的观测时间序列 $[t_1^n, t_2^n, \dots, t_{T_n}^n]$ 间隔不一致, 则称 \mathbf{X}_i 为不规则采样时间序列; 反之, 则为规则采样时间序列.

1.2 图神经网络

图神经网络(Graph Neural Network, GNN)^[17]是一类专门处理图结构数据的深度学习方法,其核心思想是通过消息传递机制在节点之间传播和聚合信息.设有一张图 $G = (\mathcal{V}, E)$, 其中 \mathcal{V} 为节点集合, E 为边集合. 每个节点 $v_i \in \mathcal{V}$ 对应一个特征向量. 在第 l 层图神经网络中, 节点特征通过以下过程更新

$$\mathbf{h}_i^{(l)} = \sigma(\mathbf{W}^{(l)} \cdot \text{AGG}(\{\mathbf{h}_j^{(l-1)} : j \in \mathcal{N}(i) \cup \{i\}\})). \quad (1)$$

其中, $\mathbf{h}_i^{(l)}$ 表示节点 i 在第 l 层的表示, $\mathcal{N}(i)$ 表示节点 i 的邻居集合, $\text{AGG}(\cdot)$ 为聚合函数(如求和、平均或注意力加权), $\mathbf{W}^{(l)}$ 为可学习的权重矩阵, $\sigma(\cdot)$ 为非线性激活函数. 通过这种迭代式的聚合与更新, 节点特征能够逐层融合邻居的局部信息, 并在更深的层次捕获全局依赖关系.

2 增强型多尺度图 Transformer

EMGT 采用“由稀疏到稠密、由图像到图结构”的整体流程. 如图 1 所示, EMGT 将不规则 MTS 依次经由 EMIT 表示构造、DSGC 图结构生成与 AMRGC 多尺度聚合, 最终得到分类结果.

2.1 增强型多通道图像转换

受传感器采样频率差异、设备故障与采集不均

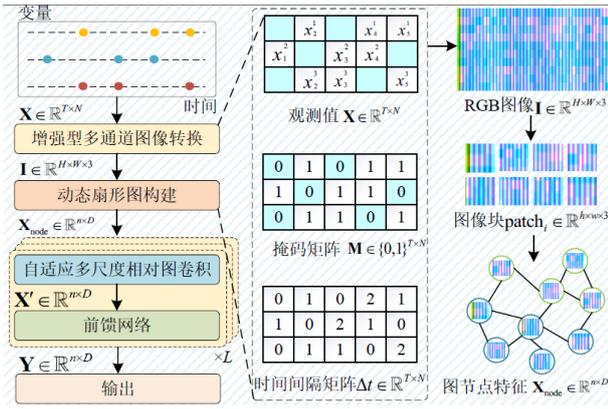


图1 EMGT 的整体结构

等因素影响,MTS 常呈现异质采样间隔与严重缺失。EMIT 将不规则 MTS 的观测值、缺失掩码与采样间隔信息统一编码为 RGB 图像表示,为后续图结构建模提供结构化输入。

具有异质的采样间隔和缺失点的观测值 $\mathbf{X} \in \mathbb{R}^{T \times N}$, 其对应的掩码矩阵 $\mathbf{M} \in \{0, 1\}^{T \times N}$ 用于指示观测值与缺失值,时间戳为 $\{t_i\}_{i=1}^T$. 该方法沿变量维度构造图像的高度 $H = N$, 沿相同时间窗口构造宽度 $W = T$, 最终得到形状为 $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ 的 RGB 图像. 三个通道分别编码不同的信息: 在 R 通道中, 对观测值进行 min-max 归一化, 空缺位置用 0 填充, 从而在列方向上直接反映数值随时间的变化趋势. 在 G 通道中, 利用对应的掩码矩阵 \mathbf{M} , 将观测与缺失点编码为 0/1 序列, 直观反映缺失分布. 在 B 通道中, 利用时间戳计算相邻观测点之间的采样间隔 Δt , 在变量内归一化至 $[0, 1]$ 并通过前向保持填充至全局时间网格, 以反映变量在时间轴上的非均匀采样特性。

相比之下,渲染式图像编码往往需要把不规则采样点映射到固定分辨率的像素网格, 并伴随重采样、插值或抗锯齿等处理. 这样一来, 原始序列中多个不同的采样点或时间间隔细节可能会落在同一纹理区域里, 导致缺失位置与间隔变化只能以模糊的形态特征呈现. 另一方面, 图像化过程中通常会做归一化与强度映射, 细微的间隔差异或局部缺失模式容易被平滑, 使模型更难准确感知不规则机制本身. 此外, EMIT 的二维坐标直接对齐变量和时间索引, 从而使动态图构建能够在几何上定义横向、纵向和斜向邻域, 进一步支撑后续相对差分聚合对关键邻居的突出. 通过这种方式, EMIT 以三通道编码观测值、掩码与时间间隔, 在不进行插补的前提下保留数值动态及其观测过程信息, 为后续跨变量关系建模提供更清晰的输入。

2.2 自适应多尺度图 Transformer

自适应多尺度图 Transformer 包含两个核心环节: 一是构建动态图结构以突破固定网格限制并建模跨变量与跨时间依赖; 二是在该结构上设计图卷积单元, 引入多尺度差分与注意力机制以增强特征聚合效果。

2.2.1 动态扇形图构建

DSGC 以图像块为中心, 在由行列边界限定的 90° 扇形邻域内沿右、下及右下斜向生成候选连接, 以反映跨变量间的时空依赖。

具体而言, 将输入图像 $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ 沿空间维度划分为大小为 $h \times w$ 的 n 个图像块, 由此得到块级网格尺寸 $H_g = H/h$, $W_g = W/w$, 节点总数 $n = H_g \cdot W_g$. 每个图像块 $\text{patch}_i \in \mathbb{R}^{h \times w \times 3}$ 通过线性映射转化为特征向量 $\mathbf{x}_i \in \mathbb{R}^D$, 得到节点特征 $\mathbf{X}_{\text{node}} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, 其中 D 是特征维度, $i = 1, 2, \dots, n$. 在得到块级网格后, 将每个图像块特征向量视为图节点, 表示为 $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$. 为了保证依赖覆盖的同时避免重复比较与冗余连边, 在扇形候选域内进行局部搜索与动态筛选. 如图 2 所示, 设节点 v_i 的块级坐标为 (p_i, q_i) , $p_i \in \{1, \dots, H_g\}$, $q_i \in \{1, \dots, W_g\}$, 以其为圆心定义行列为边界的 90° 扇形候选域. 给定最大范围 K (步长固定为 1), 枚举三个前向方向集合 $\mathcal{D}_{\text{fwd}} = \{(0, 1), (1, 0), (1, 1)\}$, 分别对应向右, 向下与右下斜向, 邻域枚举、坐标定义与范围截断均以块级网格 (H_g, W_g) 为准. 由此得到节点 v_i 的候选邻居集合定义为

$$\begin{aligned} \mathcal{N}_i^{\text{cand}} = \{ & v_j \mid (a, b) \in \mathcal{D}_{\text{fwd}}, r \in \{1, \dots, K\}, \\ & p_j = p_i + ra, q_j = q_i + rb, \\ & 1 \leq p_j \leq H_g, 1 \leq q_j \leq W_g \}. \end{aligned} \quad (2)$$

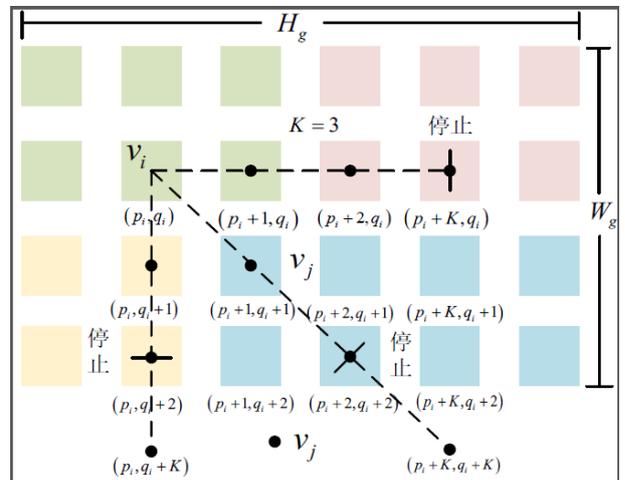


图2 动态扇形图构建

需要说明的是, r 的有效取值由“最大范围 K ”

和“网格边界”共同截断,当网格剩余长度小于 K 时枚举在边界处停止;当网格剩余长度大于等于 K 时枚举在 $r = K$ 处停止. 设 $\mathcal{S} \subseteq \{(i, j) \mid j \in \mathcal{N}_i^{\text{cand}}\}$ 为候选对的子采样集合,对任一候选对 (i, j) ,计算特征距离 $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$,为避免全局两两比较,在候选距离集合中进行子采样估计其均值 μ 与标准差 σ ,并设置动态阈值 $\tau = \mu - \sigma$. 其中

$$\mu = \frac{1}{|\mathcal{S}|} \sum_{(i,j) \in \mathcal{S}} d_{ij}, \sigma = \sqrt{\frac{1}{|\mathcal{S}|} \sum_{(i,j) \in \mathcal{S}} (d_{ij} - \mu)^2}. \quad (3)$$

引入动态门控掩码 m_{ij} ,在算子层面隐式决定是否交互,当满足 $d_{ij} < \tau$,记录候选对 (i, j) 被选中,此时 $m_{ij} = 1$,否则 $m_{ij} = 0$. 仅当 $m_{ij} = 1$ 时, (i, j) 被允许进入后续消息传递,从而在保持结构候选覆盖的同时抑制噪声和冗余邻居. 由于我们仅在前向扇形中生成候选对,为保证图结构在消息传递阶段对称可用,本文进行对称化处理. 令

$$B_{ij} = \begin{cases} m_{ij}, & j \in \mathcal{N}_i^{\text{cand}}, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

则最终无向邻接矩阵 $\mathbf{A}_{ij} = \max(B_{ij}, B_{ji})$,边集合为 $E = \{(i, j) \mid \mathbf{A}_{ij} = 1\}$,最后形成图 $G = (\mathcal{V}, E)$,用于后续消息传递. 为在无向结构下仍保留方向与先后信息,为每条边引入相对位置向量

$$\mathbf{r}_{ij} = \left[\frac{p_j - p_i}{H_g}, \frac{q_j - q_i}{W_g} \right]^\top, \mathbf{r}_{ji} = -\mathbf{r}_{ij}. \quad (5)$$

后续自适应多尺度相对图卷积利用 \mathbf{r}_{ij} 在注意力打分与聚合中区分不同方向和跨度的依赖关系.

DSGC以图像块坐标 (p_i, q_i) 构图,坐标本身携带变量轴和时间轴的位置信息. 前向 90° 扇形仅沿向右、向下与右下斜向枚举邻域,在避免全局重复比较与冗余连边的前提下获得关键的节点关联关系. 为保证消息传递稳定性,我们对候选边进行对称化得到无向邻接,同时引入相对位置向量 \mathbf{r}_{ij} 作为边属性,在注意力打分中区分不同方向与跨度,从而在无向传播下仍保留时间先后与方向信息.

2.2.2 自适应多尺度相对图卷积

虽然DSGC已建立跨变量与跨时间连接,但若直接采用传统图卷积进行特征聚合,容易出现邻居信息平均化导致的过平滑问题. AMRGC在单一邻域内同时引入差值、绝对差值与交互项等多种相对特征,从不同尺度刻画邻居关系,并结合相对位置 \mathbf{r}_{ij} 的注意力加权,自适应选择关键邻居并强化方向性与幅值差异的表达.

对于节点特征 $\mathbf{X}_{\text{node}} \in \mathbb{R}^{n \times D}$,采用图卷积来聚合来自邻居节点的特征并进行交换信息,表示如下:

$$\mathbf{x}'_i = h(\mathbf{x}_i, g(\mathbf{x}_i, \mathcal{N}(\mathbf{x}_i), \mathbf{W}_{\text{agg}}), \mathbf{W}_{\text{update}}). \quad (6)$$

其中, $\mathcal{N}(\mathbf{x}_i)$ 是 \mathbf{x}_i 的邻居节点集合,由2.2.1节得到的无向邻接 \mathbf{A} 确定; \mathbf{W}_{agg} 和 $\mathbf{W}_{\text{update}}$ 分别是聚合操作和更新操作的可学习权重. 聚合操作通过聚合邻居节点的特征来计算节点的表示,而更新操作进一步合并聚合的特征. 具体而言,对于中心节点 \mathbf{x}_i 及其邻居 \mathbf{x}_j ,计算多尺度相对差值特征 $\mathbf{h}_{ij} \in \mathbb{R}^{4D}$,并与中心节点拼接:

$$\mathbf{h}_{ij} = [\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i, |\mathbf{x}_j - \mathbf{x}_i|, \mathbf{x}_j \odot \mathbf{x}_i]. \quad (7)$$

其中, $|\cdot|$ 表示逐维绝对值, \odot 表示逐元素乘法. 这样能够同时建模方向性差异、幅值差异和交互关系. 在此基础上,通过可学习的注意力池化聚合邻居信息

$$g(\mathbf{x}_i, \mathcal{N}(\mathbf{x}_i)) = \sum_{j \in \mathcal{N}(\mathbf{x}_i)} \alpha_{ij} \mathbf{h}_{ij}, \quad (8)$$

$$\alpha_{ij} = \frac{\exp(\phi([\mathbf{h}_{ij}, \mathbf{r}_{ij}]))}{\sum_{k \in \mathcal{N}(\mathbf{x}_i)} \exp(\phi([\mathbf{h}_{ik}, \mathbf{r}_{ik}]))}. \quad (9)$$

其中, $g(\mathbf{x}_i) \in \mathbb{R}^{4D}$ 表示节点 i 的聚合特征, $\phi(\cdot)$ 表示打分函数,本文采用轻量的MLP, α_{ij} 表示注意力权重. 将相对位置 \mathbf{r}_{ij} 与相对差分特征共同用于注意力打分,使得在无向图结构下仍可区分不同方向与跨度的依赖,从而缓解时序先后信息被弱化的问题. 这种方式能自适应地选择“重要邻居”提升灵活性,避免均值聚合造成的过平滑. 最后,将所有节点聚合后的特征 $\mathbf{G} \in \mathbb{R}^{n \times 4D}$ 经过线性映射到目标空间进行更新,为了进一步增强特征表达能力并避免过平滑,引入全连接层与非线性激活,并与原始输入通过残差连接得到最终输出,其过程表示为

$$\mathbf{X}' = \sigma(\mathbf{G}\mathbf{W}_{\text{update}})\mathbf{W}_{\text{out}} + \mathbf{X}_{\text{node}}. \quad (10)$$

其中, $\mathbf{W}_{\text{update}} \in \mathbb{R}^{4D \times D}$, \mathbf{W}_{out} 是可学习参数, $\sigma(\cdot)$ 表示非线性激活函数GELU. 为了进一步增强特征转换能力并缓解过平滑,在每个节点上利用前馈网络,采用一个简单的多层感知器,具有两个全连接层:

$$\mathbf{Y} = \sigma(\mathbf{X}'\mathbf{W}_1)\mathbf{W}_2 + \mathbf{X}'. \quad (11)$$

其中, $\mathbf{Y} \in \mathbb{R}^{n \times D}$, \mathbf{W}_1 和 \mathbf{W}_2 是全连接的权重层,并且省略偏置项,前馈网络的隐藏维度通常大于 D . EMGT算法的整体流程如Algorithm 1所示.

3 数据集与实验设置

3.1 工业数据集

采用钢铁企业高炉煤气(Blast Furnace Gas, BFG)系统^[18]数据,将其运行状态识别建模为工业不规则多变量时间序列分类问题. 合理的能源调度和分配对企业节能减排、提高能源利用率至关重要^[19],

同时 BFG 系统也是典型的多输入多输出复杂系统, 由高炉、输送网络和各类用户单元组成. 为应对产耗不平衡, 调度人员可调节可控单元的消耗量以恢复稳定, 煤气柜则作为关键缓冲装置, 在需求增加时补充气体或在产气过剩时削峰. 如图 3 所示, BFG 系统包括生产过程中的产气、储气和用气三个主要环节, 具有强耦合性与明显的动态模式切换特征. 运行状态包括“煤气即将过剩”、“煤气即将短缺”和“煤气平衡”三类, 准确识别 BFG 能源系统运行状态, 能够辅助操作人员给出增加煤气消耗量、减少煤气消耗量或不调整的调度指令, 从而对实际生产决策具有直接指导意义. 其中, 干熄焦和余压透平单元为能量回收设备, 不直接消耗煤气, 因此未纳入变量范围. 表 1 展示了系统各设备的类型和生产过程, 这些用户变量相互联系, 是运行状态的关键影响因素.

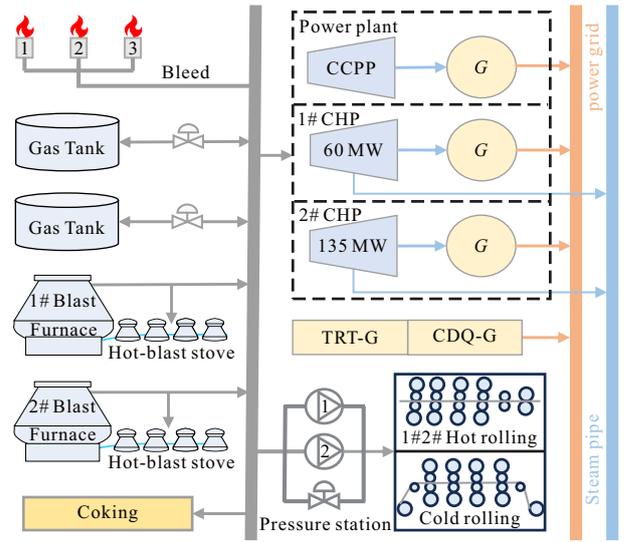


图3 钢铁高炉煤气系统

表1 高炉煤气能源系统各设备和过程

编号	类别	用户名称	单位	气量范围	缩写
1	产气	1#高炉	km ³ /h	220~500	1#BF
		2#高炉	km ³ /h	220~500	2#BF
3		1#热风炉	km ³ /h	100~280	1#HBT
4		2#热风炉	km ³ /h	100~280	2#HBT
5	用气	焦化	km ³ /h	100~230	CK
6		1#热轧	km ³ /h	0~120	1#HR
7		2#热轧	km ³ /h	0~80	2#HR
8		冷轧	km ³ /h	0~25	CR
9	自备电厂	联合循环发电机	km ³ /h	120~260	CCPP
10		1#热电联产	km ³ /h	0~100	1#CHP
11		2#热电联产	km ³ /h	70~200	2#CHP
12	存储	1#煤气柜	km ³	40~180	1#GT
		2#煤气柜	km ³	40~180	2#GT

1#和 2#热风炉、1#和 2#热轧机组这几个同类型设备按单位时间气量求和合并为 10 个关键变量. 为表征一次典型调度过程的动态演化, 每个样本采用 90 min 的时间窗口, 窗口内采样间隔为 1 min. 此外, 为避免高重叠滑动窗口导致样本高度相似进而造成评估乐观, 本文在同一调度事件附近仅保留少量代表性窗口, 未进行密集滑窗扩充. 最终共获得 1400 个带标签的工况样本, 并按 7:3 比例随机划分为训练集和测试集. 所有变量原始取值均为非负值, 故在建模前统一归一化至 $[0, 1]$ 区间. BFG 数值变化不仅反映了工业能源系统的运行状态, 也体现了实际场景下典型的多变量动态特性. 图 4 展示了高炉煤气系统在一个样本时间段内各工序的煤气量实际运行曲线. 这些曲线表现出一定不同类型的周期性, 并且在某些时刻存在噪声和离群值.

Algorithm 1 EMGT算法的整体流程

Input: 训练集 $\{(\mathbf{X}^{(n)}, \mathbf{M}^{(n)}, \Delta \mathbf{t}^{(n)}, y^{(n)})\}_{n=1}^N$, 训练轮数 $Epoch$, 批量大小 \mathcal{B} , 学习率 η

Output: 训练后的模型参数 Θ

1: 初始化模型参数 Θ , 并以学习率 η 初始化优化器

Adam;

2: for $e = 1$ to $Epoch$ do

3: for 每个批次 \mathcal{B} do

4: 对批中每个样本由 $(\mathbf{X}^{(n)}, \mathbf{M}^{(n)}, \Delta \mathbf{t}^{(n)})$ 构建 RGB 图像 I

5: 将 I 划分为图像块并映射为节点特征 $\{\mathbf{x}_i\}$, 构建动态图 $G = (\mathcal{V}, \mathcal{E})$;

6: 在 G 上执行 L 层多尺度相对图卷积, 得到节点表征 \mathbf{Y} ;

7: 经全连接层得到并进行预测;

8: 计算交叉熵损失, 并通过反向传播更新参数 Θ

9: end for

10: end for

11: return Θ

本研究选取某钢铁企业 BFG 系统 2022 年 1 月至 2022 年 6 月的历史运行数据, 由监测控制与数据采集 (SCADA) 系统获取, 以覆盖不同负荷水平下的调度工况. 需要说明的是, 上述时间跨度用于保证工况覆盖的多样性, 并非将连续数据按固定时间间隔进行稀疏抽样. 本文采用调度事件驱动的样本构建方式, 首先依据调度日志与操作记录定位具备明确状态标签的调度事件, 然后在对应连续序列中截取代表性时间窗口构成样本, 并结合现场资深调度工程师经验对运行片段进行人工标注与核验. 为便于说明系统结构突出主要物理单元, 将 1#和 2#高炉、

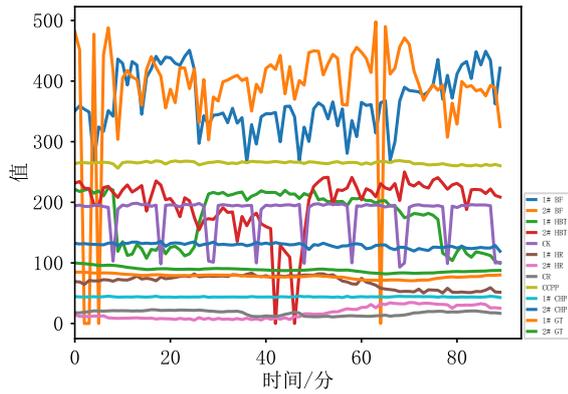


图4 高炉煤气能源系统各工序煤气量

3.2 工业不规则数据创建

异步采样. 针对工业过程中的传感器采样频率差异与记录不同步问题, 预设每个变量不同的采样间隔, 并在其余时间点将对应观测值置空, 以模拟异步采样特性. 这样可形成不同变量在不同时间戳下的观测密度差异, 更加符合实际场景中传感器异频采样和非对齐记录的情况. 对于 BFG 能源系统, 考虑的热风炉、焦化、热轧、冷轧消耗煤气的稳定型, 实际中可按照增加采样间隔来减少密集采样带来数据传输不稳定性. 其余变量的采样间隔均为 1 步, 具体来说:

- 不同变量设置相同采样间隔. 将热风炉、焦化、热轧、冷轧等消耗煤气稳定型单元的采样间隔统一设为 2、4、6、8、10 步中的一个固定值, 并分别开展实验.

- 不同变量设置不同采样间隔. 共 5 类组合: 组合 1 为 {2, 4}, 组合 2 为 {2, 4, 6}, 组合 3 为 {2, 4, 6, 8}, 组合 4 为 {4, 6, 8}, 组合 5 为 {6, 8}. 为不同单元随机分配上述预定义的间隔组合, 以体现工业系统中采样频率的多样性.

随机缺失. 在总体数据中随机遮蔽一定比例的数据点, 以模拟工业数据采集中由于电磁干扰、通信冲突或传感器瞬时故障等因素引起的孤立性缺失. 在实验中, 缺失率从 10% 逐步增加至 70%, 用于考察各模型在不同缺失强度下的鲁棒性.

3.3 基准数据集

为进一步评估所提方法在不同领域的天然不规则采样场景下的分类能力, 本文选用中国国家电网公司 (State Grid Corporation of China, SGCC) 窃电检测数据集和 PhysioNet 2012(P12), PhysioNet 2019 (P19) 和 PAMAP2(PAM) 三个被广泛使用的公开基准. SGCC^[20] 电力窃电检测数据集是智能电表场景下的真实用电数据, 被广泛用于窃电检测的监督分

类研究. 该数据集以用户为样本单位, 每个样本对应一个用户的日用电量序列, 并附带是否窃电的标签. 该数据集包含 2014–2016 年 42372 个用户在连续 1035 天内的日用电量记录, 时间步长为 1 天, 数据的缺失率为 25.64%. 训练和测试按 5:1 的方式划分. P12^[21] 每个样本包含 36 个传感器的 48 小时内的不规则采样记录, 共 11988 个样本, 每个样本对应一个二分类标签, 用于预测重症监护病房 (ICU) 患者的死亡率. P19^[22] 每个样本包含 34 个传感器的 60 小时内的不规则采样记录数据, 共 38803 个样本, 二元标签用于预测败血症. PhysioNet 数据的不规则性主要来源于按需检查和记录的临床流程, 监测频率随病情变化而改变, 导致同一患者序列内部以及不同患者之间均存在明显的时间间隔不均与异步观测. PAM^[23] 基于 PAMAP2 可穿戴活动识别数据. 筛选记录较短的受试者与样本不足的活动后, 形成 8 类分类任务, 共 5333 个样本, 每个样本为长度 600 的连续窗口. 由于可穿戴传感器数据在真实采集中容易受到佩戴、遮挡、数据传输与运动状态变化等因素影响, 样本中常呈现不同程度的缺失与有效观测不均, 这些数据具有非均匀采样、异步观测与高缺失率等特点, 能够检验模型对不同数据场景的性能.

3.4 实验实现与设置

所有实验均在 PyTorch 1.9.1 框架下实现, 并在 Nvidia GeForce RTX5070Ti GPU 上进行训练, 优化器为 Adam. 为减小随机性影响, BFG 实验结果均按照文献 [12, 16] 的统计方式, 取 5 次独立实验的分类准确率 (Accuracy) 及其标准差作为最终评价指标. 在 P19 和 P12 上遵循文献 [12] 的评估指标, 即 ROC 曲线下面积 (Area Under ROC Curve, AUROC) 和 PR 曲线下面积 (Area Under Precision-Recall Curve, AUPRC) 作为度量标准; 对 PAM 则采用准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 分数 (F1). 同时, 本文还统计了模型显存峰值 (MiB)、参数量 (M)、FLOPs(M)、训练时间 (s/epoch) 与测试时间 (s/epoch), 以相同硬件条件下的运行效率与资源开销. 在超参数设置方面, 训练次数为 500. BFG 数据集批量大小为 20, 学习率为 0.001, 在四个基准数据集中使用批量大小为 128、学习率为 0.0001, 采用 $L = 1$ 作为默认配置. 对比方法设置与原论文保持一致, 对于依赖插补的模型 (如 IP-Net), 按照其原始设计在不规则时间戳上进行插补处理后再输入模型.

4 实验验证与结果分析

4.1 消融研究

为评估 EMGT 内部各模块的影响, 在异步采样

的统一设定下,进行了系统的消融实验。

不同的堆叠层数 L 。如图 5 所示,当 $L = 1$ 时,模型取得最高准确率且方差最小,说明在 BFG 这一规模相对有限的工业数据集上,浅层结构已能充分利用由 EMIT 与 DSGC 提供的判别性表示,随着层数加深分类准确率逐渐降低。这是由于小规模数据条件下多轮邻居聚合会引发过平滑,逐步抹平节点间差异。而随层数增加参数量逐渐增加,但并未换来更高的准确率。

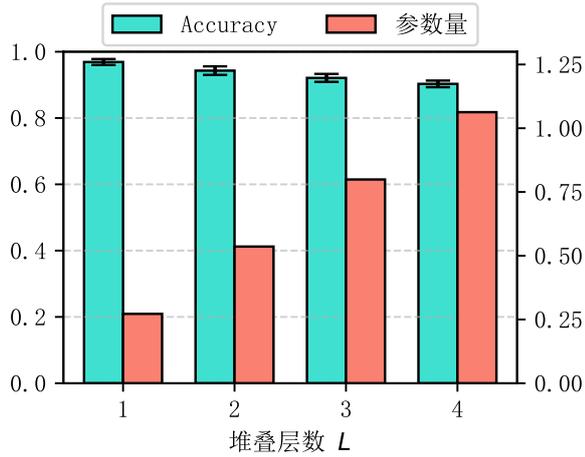


图5 不同层数 L 的性能分析

不同 EMGT 变体。为了验证各关键模块对整体 EMGT 的贡献,并在 BFG 案例的异步采样的两类不规则模式下进行实验,其中相同采样间隔设置为 2 步,不同采样间隔采用第一类组合,结果如表 2 所示。在完整模型的基础上,依次移除或替换部分组件,构建五个变体模型:

表2 各种 EMGT 变体在异步采样模式下的性能比较

方法	相同间隔	不同间隔	参数量(M)
w/o-EMIT	0.892±0.009	0.821±0.011	0.272
w/o-DSGC	0.873±0.008	0.816±0.010	0.272
w/o-AMRGC	0.831±0.015	0.782±0.018	0.406
EMGT-GCN	0.897±0.016	0.825±0.013	0.273
EMGT-GIN	0.913±0.011	0.857±0.015	0.406
EMGT	0.969±0.009	0.906±0.010	0.272

1) w/o-EMIT: 移除 EMIT, 将插补后的观测值作为数值通道。

2) w/o-DSGC: 移除 DSGC, 采用传统的 K 近邻算法构建图结构。

3) w/o-AMRGC: 移除 AMRGC, 采用自注意力机制对图像特征进行处理。

4) EMGT-GCN: 将 AMRGC 替换为图卷积神经网络 (GCN)^[24], 利用一阶谱卷积近似实现信息传播。

5) EMGT-GIN: 将 AMRGC 替换为图同构网络 (GIN)^[25], 通过求和聚合与 MLP 更新进行邻居特征融合。

在相同采样间隔和不同采样间隔下, w/o-EMIT 的平均精度分别为 0.892 和 0.821, 较完整模型分别降低 7.7 和 8.5 个百分点, 说明直接引入不规则信息对性能至关重要, 为后续图建模奠定了基础。采用传统 K 近邻构图的 w/o-DSGC, 平均精度下降至 0.873 和 0.816, 这表明动态扇形构图能够覆盖纵横及斜向依赖, 减少冗余节点, 有效提升图表示的表达能力。采用自注意力机制的 w/o-AMRGC, 较完整模型精度分别降低 13.8 和 12.4 个百分点, 这主要源于自注意力机制缺乏固有的归纳偏置, 在小规模工业数据集上易导致过拟合和性能下降。此外, 将 AMRGC 替换为传统图神经网络结构, EMGT-GCN 和 EMGT-GIN 的精度分别优于 w/o-AMRGC, 但仍低于完整模型。这表明, 所提出的基于相对差分特征的自适应图卷积通过逐维差值与注意力加权机制突出关键邻居, 有效缓解了传统图卷积的过平滑问题, 从而在不规则工业数据场景下展现出更强的适应性和表达能力。

在复杂度方面, 各模型的参数量差异主要由聚合机制引起。完整 EMGT 在仅 0.272M 参数量下即可达到最高精度, 展现了在准确性与效率之间的良好平衡。综上所述, 消融实验证明了 EMIT 保证了不规则模式的多元信息输入, DSGC 增强了跨变量依赖的图结构建模, 而 AMRGC 有效缓解过平滑并突出关键差异特征。三者协同作用使得 EMGT 能够在复杂工业不规则 MTS 上取得显著优势。

不同 EMIT 变体。为了验证 EMIT 中 RGB 三通道的贡献, 以及相比于现有图像编码方法的优势, 进行了 RGB 通道消融和与不同图像编码方法得比较实验。在保持后续 DSGC 与 AMRGC 不变的条件下, 改变 EMIT 编码方式或替换为现有的图像编码方法, 形成以下变体模型:

1) 递归图^[26]: 通过延迟嵌入构造相空间轨迹, 计算任意两状态的距离并以递归矩阵形式可视化。

2) 折线图栅格化^[27]: 保持真实时间轴, 把每个变量的观测点按时间戳画成折线, 再把多变量按子图拼贴成图像。

3) 灰度图^[28]: 通过插值离散到统一时间网格, 再经过周期分段将一维序列映射为二维结构化矩阵。

4) 角度场^[29]: 先将不规则时间序列对齐, 然后映射到极坐标角度空间, 并计算任意两时刻角度的和或构成二维格兰姆矩阵。

- 5) EMIT-R: 仅保留数值通道.
 6) EMIT-RG: 保留数值和缺失通道.
 7) EMIT-RB: 保留数值和时间间隔通道.

结果如表3所示,在异步采样设置下,仅用数值通道的EMIT-R分类精度分别为0.927和0.842.当加入缺失通道,EMIT-RG平均精度分别提升3.1和4.1个百分点,说明缺失分布增强了模型判别能力.当加入时间间隔通道EMIT-RB在不同间隔情况下平均精度提升3.3个百分点,说明时间间隔能够刻画跨变量异步导致的时间尺度差异.完整EMIT平均分类精度进一步达到0.969和0.906,这表明缺失与时间间隔信息对分类能力提升是有价值的.

表3 各种EMIT变体在异步采样模式下的性能比较

方法	相同间隔	不同间隔
递归图	0.925±0.012	0.832±0.005
折线图栅格化	0.956±0.006	0.869±0.011
灰度图	0.936±0.008	0.834±0.015
角度场	0.904±0.012	0.796±0.005
EMIT-R	0.927±0.010	0.842±0.012
EMIT-RG	0.958±0.009	0.883±0.011
EMIT-RB	0.951±0.011	0.875±0.013
EMIT	0.969±0.009	0.906±0.010

与现有图像编码方法相比,EMIT的优势在于直接利用不规则信息,而不是依赖更复杂的图像渲染.递归图基于延迟嵌入后状态点两两距离构造递归矩阵,异步采样会改变状态点密度与距离结构,使得不同间隔场景下精度下降至0.832.角度场通常需要先对齐再映射到极坐标并构造格兰姆矩阵,对齐或插值会弱化采样间隔差异,不同间隔下精度仅为0.796.灰度图将序列插值离散到统一网格并进行周期分段映射,插值过程同样会平滑并稀释不规则模式,不同间隔下为0.834.折线图栅格化虽然保留时间轴形态,但缺失与时间间隔往往只能通过纹理细节被动体现,模型往往难以区分真实动态变化与采样、缺失机制造成的表观变化,因此不同间隔下仍落后,为0.869.相较之下,EMIT将数值、缺失与采样间隔直接写入R、G、B三通道,使不规则性以可学习信号形式直接输入模型,从而在更复杂的不同间隔场景下表现更优,相比折线图栅格化提升3.7个百分点.

不同AMRGC变体.为验证AMRGC中三类相对特征(差值、绝对差值与交互项)的有效性,我们在BFG系统数据集上进行了特征级消融.如表4所示,对于完整的AMRGC移除任一项都会导致精度下降,且在更具挑战的“不同间隔”场景下降更明显.去差值精度下降最大,说明符号方向变化对不规则

采样下的工况区分更明显.在异步采样不同时间间隔情况下,单看数值大小容易混淆.绝对差值主要补充变化幅度信息,当存在噪声或局部尖峰时,提高模型对变化强弱的分辨能力,它与差值形成方向和强度的互补.交互项则反映了不同特征块之间的协同联动关系,当被去掉时,也会使精度下降.

表4 AMRGC内部相对特征的消融分析

AMRGC变体	相同间隔	不同间隔	参数量(M)
w/o-交互	0.964±0.009	0.900±0.011	0.272
w/o-绝对差值	0.962±0.010	0.895±0.012	0.272
w/o-差值	0.957±0.011	0.892±0.013	0.272
AMRGC	0.969±0.009	0.906±0.010	0.272

4.2 性能评估与比较

为了评估本文提出的EMGT在不规则工业时间序列分类中的性能,将其与11种经典模型进行全面比较,包括GRU-D^[9]、mTAND^[13]、IP-Net^[8]、SeFT^[11]、Raindrop^[12]、WarpFormer^[15]、MTSFormer^[16]、ContiFormer^[30]、WaveGNN^[31]、PatchTST^[32]和iTransformer^[33].其中PatchTST和iTransformer为用于MTS预测的SOTA方法,其余为专门针对不规则MTS分类的方法.

1) 工业数据集性能评估

整体性能评估.表5给出了在BFG案例上三种不规则模式下的分类精度.其中,异步采样模式下,相同采样间隔设置为2步,不同采样间隔采用第一类组合,随机缺失模式下缺失率设置为20%.传统序列模型GRU-D和mTAND虽能通过衰减或时间嵌入刻画不规则性,但当不规则来源由单变量时间偏移进一步扩展到跨变量异步与缺失机制扰动时,性能波动明显,反映出其对跨变量依赖与缺失模式信息的利用仍不充分.而依赖插补或集合化输入的方法IP-Net和SeFT往往会弱化甚至扭曲缺失与采样间隔所携带的判别线索,导致整体效果受限.图引导与对齐增强的模型Raindrop和WarpFormer能够在一定程度上提升不规则场景下的表现,但在高缺失或间隔差异更显著时,仍可能受限于较为粗粒度的聚合方式或对齐误差累积,难以保持稳定优势.EMGT在三种不规则模式下均取得最优性能,在相同采样间隔下达到0.969,并在随机缺失场景下仍保持0.897.其优势源于将观测值、缺失模式与时间间隔作为互补信息进行联合表征,使模型在判别时获得更充分的条件信息.在贝叶斯最小风险意义下,这种“更充分条件化”不会降低理论最优判别能力,反而有助于减少由采样机制变化引入的表示偏差.同时,

基于图结构的块级交互能更有效地传递跨变量耦合关系, 使关键依赖在异步与稀疏观测下仍可被捕获, 从而提升对缺失扰动与采样差异的鲁棒性与泛化性.

表5 三种不规则模式下的整体性能比较

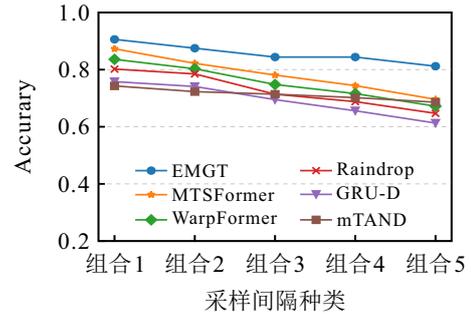
方法	相同采样间隔	不同采样间隔	随机缺失
GRU-D	0.842±0.018	0.758±0.021	0.775±0.020
mTAND	0.807±0.020	0.743±0.023	0.714±0.024
IP-Net	0.782±0.022	0.715±0.025	0.652±0.026
SeFT	0.723±0.025	0.664±0.027	0.608±0.028
Raindrop	0.883±0.015	0.802±0.018	0.784±0.017
WarpFormer	0.904±0.012	0.836±0.014	0.829±0.013
MTSFormer	0.938±0.010	0.873±0.011	0.857±0.010
EMGT	0.969±0.008	0.906±0.009	0.897±0.008

异步采样与随机缺失下多种不规则模式的性能评估. 图6进一步展示了在不同采样步长与缺失率条件下各模型的性能变化趋势. 在异步采样模式中(图6(a)、(b)), 随着采样步长逐渐增大, 所有模型的精度均呈下降趋势, 这是由于时序细节丢失和数据不确定性增加所致. 传统基于插补或序列建模的方法在大步长条件下性能下降尤为明显, 如 mTAND 在步长超过 2 时精度大幅下滑.

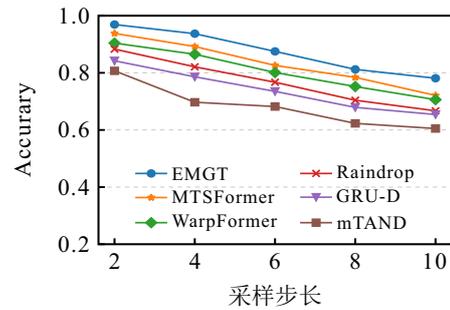
相比之下, Raindrop、WarpFormer 和 MTSFormer 能够部分缓解这一问题, 但仍出现性能衰减. EMGT 得益于 EMIT 提供的多源信息输入表示, 结合 DSGC 与 AMRGC 在邻域聚合中对关键差异特征的强化, 在不同步长下始终保持最高精度, 且下降幅度最小, 表明其不仅能刻画不规则动态, 还能在噪声干扰下提取稳定特征. 在随机缺失模式中(图6(c)), 缺失比例从 10% 逐步增加至 70% 时, 整体趋势同样表现为各模型精度随缺失率升高而降低, 且低于异步采样的模式. 其中, GRU-D 和 mTAND 能够通过衰减或时间注意力机制在低缺失率下保持一定性能, 但在缺失率超过 50% 后精度下降明显. WarpFormer 和 MTSFormer 由于具备多尺度建模能力, 整体表现相对稳定, 但仍存在一定性能衰减. 相比之下, EMGT 在缺失率逐渐增大的情况下依然保持较高精度, 尤其在缺失率达到 70% 时仍能维持显著优势.

2) 基准数据集性能评估

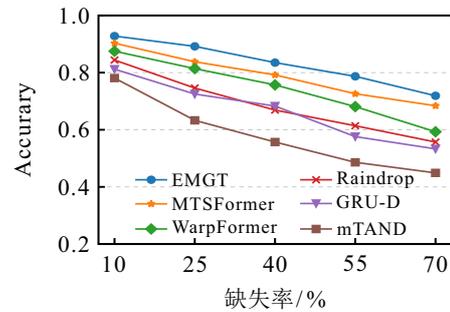
SGCC 数据集上性能评估. 表6给出了在 SGCC 数据集上 EMGT 与其他 SOTA 方法的性能比较结果. 其中 PatchTST 与 iTransformer 原本面向规则采样的时间序列预测任务. 为将其用于本文的不规则多变量分类对比, 首先将不规则观测通过插值的方式对齐到统一时间网格, 保持原编码器结构不变的前提下, 我们移除预测头并接入轻量分类头(池



(a) 不同间隔



(b) 相同间隔



(c) 随机缺失

图6 异步采样与随机缺失下多种不规则模式的性能评估 化+全连接) 输出类别概率, 从而实现公平可复现的基线对比. EMGT 在三项指标上均取得最优, 其中 F1 与 AUC 优势较为明显, 这说明本文方法并不是通过把阈值调得更激进提高 Recall, 对于 SGCC 这种缺失模式本身带有行为信息的数据, PatchTST 和 iTransformer 这类规则时间序列预测骨干在预处理时容易把缺失压平为数值噪声, 模型更偏向学习插补后的曲线形状. EMGT 能够直接利用观测值、缺失指示与时间间隔, 因此在 AUC 上更加稳定、在 F1 上更具优势. 此外, 与专门面向不规则的方法相比, EMGT 仍能取得更高的性能, EMGT 将不规则序列统一转为结构化表征, 并通过后续的关系建模捕获关键片段之间的依赖, 使得在天然缺失的窃电检测中更具判别能力.

P12、P19 和 PAM 数据集上性能评估. 表7给出了各方法在 P12、P19 与 PAM 三个数据集上的性能. 在 P12 上, 与较强的 MTSFormer, WaveGNN 相比, EMGT 的优势在 AUPRC 上更明显, 说明其对正类样本的识别更稳定. 在 P19 上, EMGT 相对

表6 不同方法在 SGCC 数据集上的性能比较

方法	AUROC	Recall	F1-Score
PatchTST	0.841±0.006	0.705±0.009	0.637±0.013
iTransformer	0.852±0.008	0.708±0.018	0.642±0.006
WarpFormer	0.859±0.011	0.704±0.013	0.654±0.011
ContiFormer	0.835±0.014	0.699±0.008	0.633±0.006
WaveGNN	0.847±0.005	0.703±0.017	0.639±0.017
MTSFormer	0.857±0.017	0.706±0.021	0.642±0.014
EMGT	0.876±0.011	0.712±0.014	0.658±0.012

MTSFormer 和 WaveGNN 虽然提升幅度相对较小, 但方差较小, 在高缺失、强异步的临床记录下判别更稳定. PAM 上各方法整体分数较高, EMGT 在接近饱和的区间仍有小幅提升, 说明模型对多传感器异步与缺失扰动的鲁棒性更强. 综上所述, EMGT 的收益主要来自对不规则信息的直接利用. 在 P12 和 P19 这种缺失极高、且本身携带临床流程信息的场景, 通过后续结构化关系学习能够捕获更加关键的特征.

表7 不同方法在 P12、P19 和 PAM 数据集上的性能比较

方法	P12		P19		PAM			
	AUROC	AUPRC	AUROC	AUPRC	Accuracy	Precision	Recall	F1-Score
GRU-D	0.819±0.021	0.461±0.047	0.839±0.017	0.469±0.021	0.833±0.016	0.846±0.012	0.852±0.016	0.848±0.012
mTAND	0.842±0.008	0.482±0.034	0.844±0.013	0.506±0.020	0.746±0.043	0.743±0.040	0.795±0.028	0.768±0.034
IP-Net	0.826±0.014	0.476±0.031	0.846±0.013	0.381±0.037	0.743±0.038	0.756±0.021	0.779±0.022	0.766±0.028
SeFT	0.739±0.025	0.311±0.041	0.812±0.023	0.419±0.031	0.671±0.022	0.700±0.024	0.682±0.015	0.685±0.018
Raindrop	0.828±0.017	0.440±0.030	0.870±0.023	0.518±0.055	0.885±0.015	0.899±0.015	0.899±0.006	0.899±0.010
WarpFormer	0.835±0.019	0.451±0.035	0.877±0.032	0.534±0.064	0.935±0.010	0.945±0.009	0.940±0.009	0.942±0.008
ContiFormer	0.812±0.008	0.439±0.030	0.792±0.023	0.358±0.023	0.890±0.010	0.900±0.008	0.910±0.009	0.902±0.008
WaveGNN	0.839±0.012	0.494±0.015	0.880±0.009	0.571±0.047	0.956±0.011	0.957±0.011	0.956±0.011	0.956±0.011
MTSFormer	0.849±0.014	0.511±0.037	0.888±0.015	0.577±0.044	0.968±0.009	0.973±0.008	0.969±0.006	0.971±0.007
EMGT	0.854±0.010	0.527±0.031	0.897±0.012	0.585±0.027	0.974±0.008	0.980±0.006	0.977±0.005	0.978±0.006

4.3 计算复杂度

为便于分析, 设不规则 MTS 经 EMIT 转换后得到大小为 $H \times W$ 的 RGB 图像, 图像块尺寸为 $h \times w$, 每个图像块通过一次线性映射得到维度为 D 的节点特征, 此时图中的节点数为 $n = \frac{HW}{hw}$. EMIT 主要包括三通道构造以及 patch 级线性映射, 其时间复杂度约为 $\mathcal{O}(HW + nD)$. 可以看出, EMIT 仅涉及通道级组合与一次线性映射, 其计算量远小于后续图卷积, 因此在整体复杂度中可以近似忽略. 在 DSGC 中, 每个节点在 90° 扇形区域内按沿向右、向下以及右下斜向三个方向离散枚举候选, 最大范围 K , 步长为 1, 因此单节点候选边数上界为 $3K$. 记候选对集合为 E , 则图中边数约为 $|E| \leq 3Kn$. 对每个候选对需要计算一次 D 维欧氏距离并进行阈值判定, 因此构图的时间复杂度为 $\mathcal{O}(|E|D) = \mathcal{O}(nKD)$. 在 AMRGC 中, 对每条边 (i, j) 需计算多尺度相对特征 $h_{ij} \in \mathbb{R}^{4D}$, 再通过轻量 MLP 得到注意力分数, 并对邻居特征进行加权聚合; 随后对每个节点执行一次线性变换和前馈网络更新. 单层 AMRGC 的时间复杂度可写为 $\mathcal{O}(|E|D^2 + nD^2) = \mathcal{O}(nKD^2)$, 堆叠 L 层后, 自适应多尺度图 Transformer 的总时间复杂度为

$\mathcal{O}(LnKD^2)$. 在 EMGT 中, 特征维度 D , 网络层数 L , 以及图像块大小 h, w 都是固定常数, 因此 n 与 HW 成正比, 整体时间复杂度关于 n 呈线性增长.

表 8 选取 BFG、P12、P19 作为代表性数据集统计了各方法在相同硬件条件下的运行效率与资源开销, 其中训练和测试耗时包含了数据加载过程. 计时前预热两次, 重复 5 次取平均, 显存占用为峰值. 总体来看, EMGT 在三个数据集上保持了相对适中的模型规模与计算开销. 以 BFG 为例, EMGT 的峰值显存仅为 520.342 MiB. 其训练和测试耗时分别为 2.480 和 0.941 s/epoch, 具备较高推理效率. 此外, EMGT 的参数量仅 0.272M、单样本 FLOPs 为 1.233M, 体现出轻量化优势. 在当前 batch 设置下, 完成一轮数据的前向推理小于 1 秒, 在实际应用中单个样本的推理时间会更小.

BFG 系统运行状态在线识别通常采用分钟级滚动更新, 即每 1 min 接收新数据并输出一状态标签. 因此, 工程上可采用的实时性标准为: 单次增量数据预处理加推理时延应小于 60s, 并为通信与上位机逻辑预留裕度. 采用表 8 的结果推算, EMGT 单个样本的推理耗时会远小于该阈值, 具备满足分钟级在线识别的充足裕度.

表8 各方法在相同硬件下的运行效率与资源开销对比

指标	数据集	GRU-D	mTAND	Raindrop	WarpFormer	ContiFormer	WaveGNN	MTSFormer	EMGT
峰值显存(MiB)	BFG	533.839	589.555	1094.819	1520.433	3094.245	568.842	547.102	520.342
	P12	1274.055	1355.621	8858.365	13453.078	7458.246	13133.448	10406.773	8533.176
	P19	775.515	992.383	2404.293	11084.037	4972.473	5267.339	4370.364	3583.612
参数量 (M)	BFG	0.216	0.089	0.027	0.041	0.267	0.823	1.344	0.272
	P12	0.263	0.089	0.094	0.064	0.279	2.101	13.246	0.423
	P19	0.259	0.089	0.089	0.063	0.273	2.003	13.189	0.415
FLOPs (M)	BFG	1.206	2.767	1.083	1.774	1.493	2.902	9.651	1.233
	P12	56.141	99.729	11.361	101.483	56.802	157.994	48.114	25.838
	P19	15.453	6.543	3.057	64.370	16.112	106.113	35.041	14.354
训练耗时(s/epoch)	BFG	8.383	3.246	8.675	3.360	2.456	3.784	6.584	2.480
	P12	33.734	61.842	22.151	23.628	7.124	63.386	15.125	7.453
	P19	17.982	4.792	10.088	5.319	4.031	35.625	8.806	5.411
测试耗时(s/epoch)	BFG	1.175	1.025	1.165	1.033	0.913	1.812	1.876	0.941
	P12	2.518	2.138	4.077	3.756	2.896	3.057	3.967	2.844
	P19	1.372	1.228	2.368	1.079	2.498	2.248	1.984	1.624

4.4 案例研究

为了直观反映注意力权重的分布模式, 我们对不同工况在去自环设置下的边注意力权重 α 分布进行了可视化. 图7(a)给出了 α 的累积分布函数对比, 三条曲线整体形态一致但在低权重区间仍存在可观察的分离, 工况间在 $\alpha \leq 0.2$ 附近的累计上升速度不同, 表明不同工况下弱关联边的占比与累积方式存在差异, 即模型对不同工况的局部依赖强度呈现不同的分散特性. 图7(b)进一步从密度曲线角度刻画了这种差异, 各工况在 $\alpha \leq 0.003$ 处均出现明显峰值, 说明大多数边被赋予较小权重, AMRGC 倾向于将

信息聚合集中到少量关键边上. 同时在 α 接近于1附近仍出现上扬, 意味着每个工况中都存在少量“主导边”承担主要信息传递, 但其峰值幅度与尾部形状在不同工况间并不完全一致, 反映了关键依赖关系的强弱分配随工况变化. 图7(c)按工况分别给出直方图计数, 直观呈现出低权重边占主导, 少量高权重关键边的长尾集中特征, 并且不同工况在中等权重段与高权重尾部的柱形分布存在差别, 说明模型在不同工况下选择的关键跨图像块连接并非固定不变. AMRGC 能够在动态构图基础上自适应地分配边级注意力, 形成抑制无效边突出关键边的机制.

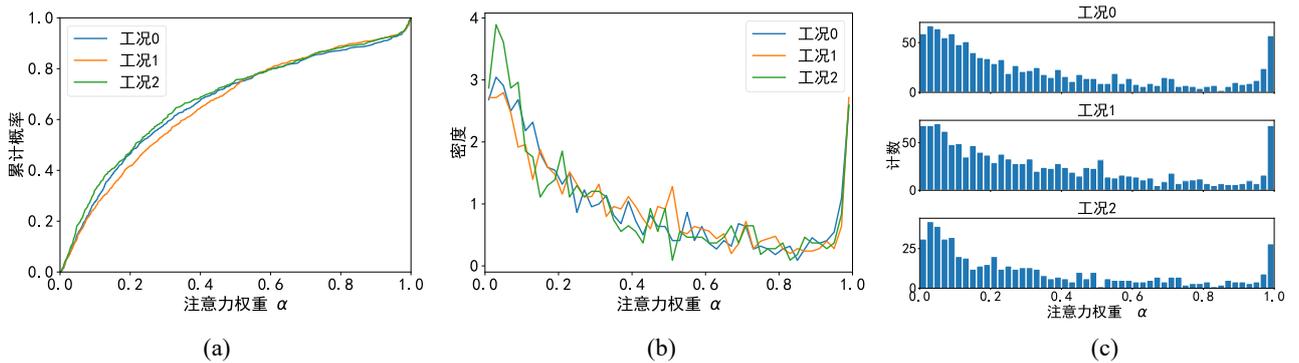


图7 不同工况下 AMRGC 边注意力权重分布对比

为了直观反映模型的决策依据, 以 BFG 一个典型样本为例, 完整展示了关键特征识别, 机理解释, 结构支撑和反事实验证的闭环过程. 图8(a)在数值通道 R 上标注了模型识别出的 Top-20 重要图像块, 这些高重要性区域在“变量 \times 时间”平面上呈现出明显的局部聚集与跨变量联动特征, 表明模型关注的并非单一时间点或单一变量, 而是能够反映工况差异的关键片段组合. 图8(b)进一步给出了 AMRGC 的相对特征贡献分解, 中心项贡献最高, 差值项与绝

对差值项贡献相近, 交互项贡献相对较小, 说明模型判别主要依赖局部状态本身并结合变量间变化幅度和方向差异进行补充, 而交互项用于刻画协同耦合关系, 从机制层面解释了 AMRGC 同时引入差值、绝对差值与交互项的必要性与互补性. 图8(c)展示了 DSGC 在图像块网格上构建的稀疏图结构, 连边主要发生在相邻或近邻图像块之间, 使得信息传播能够围绕图8(a)中的关键区域进行有效聚合, 避免了全连接带来的冗余传播与噪声干扰, 从而为关键

特征的可解释聚合提供结构先验. 最后, 图 8(d) 给出按重要性逐步遮挡图像块的删除曲线, 随着累计遮挡图像块数量增加, 预测类概率由接近 1 逐步下降至约 0.3 左右, 同时“预测-对比类边际”从明显为正持续下降并在遮挡约 30 个图像块后接近 0 乃至转

负. 这表明模型对该样本的判别证据确实集中于图 8(a) 所定位的关键图像块, 一旦这些区域被移除, 模型对原类别的支持减弱并向竞争类别转移. 上述闭环证据链通过在图像块粒度上给出可追溯和可解释的关键特征识别过程, 为模型决策提供了依据.

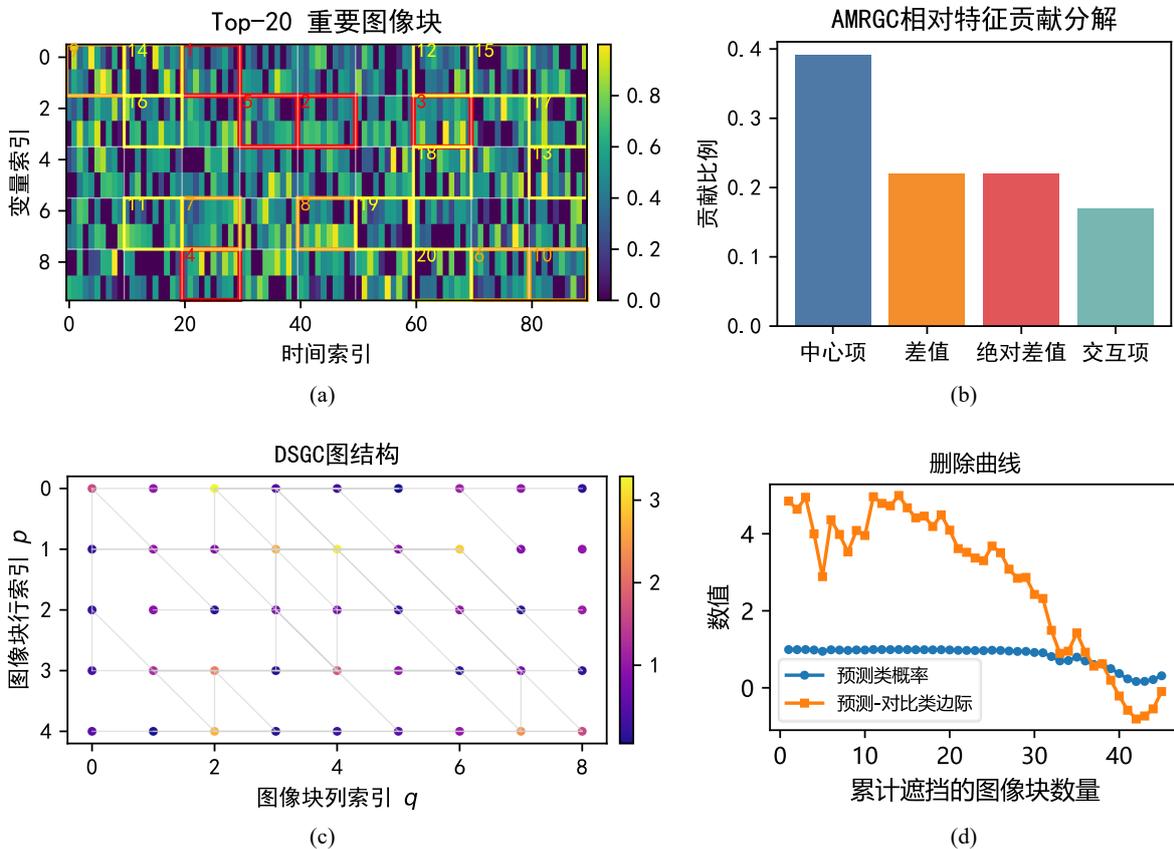


图8 关键特征识别与反事实验证

5 结论

针对工业不规则 MTS 分类中的插补失真与信息稀释问题, 本文提出增强型多尺度图 Transformer, 并构建由增强型多通道图像转换、动态扇形图构建和自适应多尺度相对图卷积组成的端到端流程, 实现无需插补的不规则序列分类. 在某钢铁企业高炉煤气数据及四个公开不规则基准上验证了方法在多种不规则模式下的有效性.

未来将进一步探索引入可解释性建模与跨场景迁移学习, 以提升方法在复杂工业系统中的应用价值. 可解释性用于揭示不规则 MTS 中的关键变量与时序模式并增强决策可信度, 迁移学习用于在不同工况与装置间共享与迁移知识并提升分类通用性.

参考文献 (References)

[1] Yuan Y G, Wei J N, Huang H S, et al. Review of resampling techniques for the treatment of imbalanced industrial data classification in equipment condition monitoring[J]. *Engineering Applications of Artificial*

Intelligence, 2023, 126: 106911.

[2] 刘颖, 李旭, 吕政, 等. 基于深度残差概率随机森林的时间序列分类方法[J]. *控制与决策*, 2024, 39(7): 2315-2324.

(Liu Y, Li X, Lv Z, et al. Time series classification using deep residual probability random forests[J]. *Control and Decision*, 2024, 39(7): 2315-2324.)

[3] Li Z Y, Kang S Y, Feng N H, et al. PSCFormer: A lightweight hybrid network for gas identification in electronic nose system[J]. *Pattern Recognition*, 2024, 145: 109912.

[4] Wei Y X, Peng J T, He T, et al. Compatible transformer for irregularly sampled multivariate time series[C]. *IEEE International Conference on Data Mining*. Shanghai, 2023: 1409-1414.

[5] 梁浩鹏, 曹洁, 赵小强. 基于并行双向时间卷积网络和双向长短期记忆网络的轴承剩余使用寿命预测方法[J]. *控制与决策*, 2024, 39(4): 1288-1296.

(Liang H P, Cao J, Zhao X Q. Remaining useful life prediction method for bearing based on parallel bidirectional temporal convolutional network and

- bidirectional long and short-term memory network[J]. *Control and Decision*, 2024, 39(4): 1288-1296.)
- [6] 高杨, 王新宇, 贺达, 等. 面向缺失多元时间序列的图神经网络异常检测算法[J]. *国防科技大学学报*, 2025, 47(3): 32-40.
(Gao Y, Wang X Y, He D, et al. Anomaly detection algorithm based on graph neural network for missing multivariate time series[J]. *Journal of National University of Defense Technology*, 2025, 47(3): 32-40.)
- [7] 李康, 李爽, 高小永, 等. 多变量时序标记 Transformer 及其在电潜泵故障诊断中的应用[J]. *控制与决策*, 2025, 40(4): 1145-1153.
(Li K, Li S, Gao X Y, et al. Multivariate time series tokenized Transformer and its application in fault diagnosis of electric submersible pump[J]. *Control and Decision*, 2025, 40(4): 1145-1153.)
- [8] Shukla S N, Marlin B M. Interpolation-prediction networks for irregularly sampled time series[C]. *Proceedings of the 7th International Conference on Learning Representations*. New Orleans, 2019: 557-570.
- [9] Che Z P, Purushotham S, Cho K, et al. Recurrent neural networks for multivariate time series with missing values[J]. *Scientific Reports*, 2018, 8: 6085.
- [10] Rubanova Y, Chen R T Q, Duvenaud D. Latent ODEs for irregularly-sampled time series[C]. *Conference on Neural Information Processing Systems*. Vancouver, 2019: 5320-5330.
- [11] Horn M, Moor M, Bock C, et al. Set functions for time series[C]. *International Conference on Machine Learning*. Online, 2020: 4353-4363.
- [12] Zhang X, Zeman M, Tsiligkaridis T, et al. Graph-guided network for irregularly sampled multivariate time series[C]. *International Conference on Learning Representations*. Online, 2022: 1-18.
- [13] Shukla S N, Marlin B M. Multi-time attention networks for irregularly sampled time series[C]. *International Conference on Learning Representations*. Online, 2021: 14897-14911.
- [14] Tipimeni S, Reddy C K. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series[J]. *ACM Transactions on Knowledge Discovery from Data*, 2022, 16(6): 1-17.
- [15] Zhang J W, Zheng S, Cao W, et al. Warpformer: A multi-scale modeling approach for irregular clinical time series[C]. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Long Beach, 2023: 3273-3285.
- [16] Zheng L N, Li Z Y, Dong C G, et al. Irregularity-informed time series analysis: Adaptive modelling of spatial and temporal dynamics[C]. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. Boise, 2024: 3405-3414.
- [17] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. *IEEE Transactions on Neural Networks*, 2008, 20(1): 61-80.
- [18] Gu Y, Jin F, Zhao J, et al. A hybrid lightweight transformer architecture based on fuzzy attention prototypes for multivariate time series classification[J]. *Information Sciences*, 2025, 703: 121942.
- [19] Jin F, Wang L Q, Zhao J, et al. Granular-causality-based byproduct energy scheduling for energy-intensive enterprise[J]. *IEEE Transactions on Automation Science and Engineering*, 2020, 17(4): 1662-1673.
- [20] Zheng Z B, Yang Y T, Niu X D, et al. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(4): 1606-1615.
- [21] Goldberger A L, Amaral L A N, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals[J]. *Circulation*, 2000, 101(23): e215-e220.
- [22] Reyna M A, Josef C, Seyedi S, et al. Early prediction of sepsis from clinical data: The PhysioNet/computing in cardiology challenge 2019[J]. *Computing in Cardiology Conference*. Singapore, 2019, 45: 1-4.
- [23] Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring[C]. *2012 16th International Symposium on Wearable Computers*. Newcastle, 2012: 108-109.
- [24] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[C]. *International Conference on Learning Representations*. Toulon, 2017: 2713-2726.
- [25] Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks?[C]. *International Conference on Learning Representations*. New Orleans, 2019: 9104-9120.
- [26] Hatami N, Gavet Y, Debayle J. Classification of time-series images using deep convolutional neural networks[C]. *Proceedings of the 10th International Conference on Machine Vision*. Vienna, 2017: 1069606.
- [27] Li Z, Li S, Yan X. Time series as images: Vision transformer for irregularly sampled time series[C]. *Advances in Neural Information Processing Systems*. New Orleans, 2023: 49187-49204.
- [28] Chen M, Shen L, LI Z, et al. VisionTS: Visual masked autoencoders are free-lunch zero-shot time series forecasters[C]. *Proceedings of the 42nd International Conference on Machine Learning*. Vancouver, 2025: 8979-9007.
- [29] Wang Z, Oates T. Imaging time-series to improve classification and imputation[C]. *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Buenos Aires, 2015: 3939-3945.
- [30] Chen Y, Ren K, Wang Y, et al. ContiFormer:

- Continuous-time transformer for irregular time series modeling[C]. Advances in Neural Information Processing Systems. New Orleans, 2023: 47143-47175.
- [31] Hajisafi A, Siampou M D, Azarjoo B, et al. WaveGNN: Integrating graph neural networks and Transformers for decay-aware classification of irregular clinical time-series[C]. IEEE International Conference on Big Data. 2025: 1-10.
- [32] Nie Y, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: Long-term forecasting with transformers[C]. International Conference on Learning Representations. Kigali, 2023.
- [33] Liu Y, Hu T, Zhang H, et al. iTransformer: Inverted transformers are effective for time series forecasting[C]. International Conference on Learning Representations.

Vienna, 2024: 4004-4028.

作者简介

顾岩 (1992-), 男, 博士生, 主要研究方向为工业能源优化调度、智能信息处理, E-mail: guyan9959@mail.dlut.edu.cn;

金锋 (1989-), 男, 副教授, 博士, 主要研究方向为工业能源预测与优化调度、智能感知与过程数据分析, E-mail: jin_feng@dlut.edu.cn;

赵珺 (1981-), 男, 教授, 博士, 主要研究方向为工业生产调度、计算机集成制造, E-mail: zhaoj@dlut.edu.cn;

王伟 (1955-), 男, 教授, 博士, 主要研究方向为流程工业自动化、人工智能技术及其应用, E-mail: wangwei@dlut.edu.cn.