



完备时空信息引导的工业过程内生变量缺失下的鲁棒KPI预测

王凯 张小刚 雷捷维 陈华

A Complete Spatio-temporal Information Guided Approach for Robust KPI Forecasting Under Endogenous-variable Missingness

WANG Kai, ZHANG Xiao-Gang, LEI Jie-Wei, CHEN Hua

在线阅读 View online: <https://doi.org/10.16383/j.aas.c250653>

您可能感兴趣的其他文章

面向认知表现预测的时空共变混合深度学习模型

Spatio-temporal Co-variant Hybrid Deep Learning Framework for Cognitive Performance Prediction

自动化学报. 2022, 48(12): 2931–2940 <https://doi.org/10.16383/j.aas.c220025>

基于跨时空稳定因果动态贝叶斯网络的工业过程安全控制

Industrial Process Safety Control Based on Spatio-temporal Stable Causal Dynamic Bayesian Network

自动化学报. 2025, 51(9): 2058–2071 <https://doi.org/10.16383/j.aas.c250108>

基于多变量时空融合网络的风机数据缺失值插补研究

Study of Missing Value Imputation in Wind Turbine Data Based on Multivariate Spatiotemporal Integration Network

自动化学报. 2024, 50(6): 1171–1184 <https://doi.org/10.16383/j.aas.c230534>

多聚点子空间下的时空信息融合及其在行为识别中的应用

Recognizing Action Using Multi-center Subspace Learning-based Spatial-temporal Information Fusion

自动化学报. 2022, 48(11): 2823–2835 <https://doi.org/10.16383/j.aas.c190327>

“结构内容”框架下融合时空特征的技术预测模型

A Technology Forecasting Model Integrating Spatiotemporal Features Under the “Structure-Content” Framework

自动化学报. 2025, 51(12): 2621–2632 <https://doi.org/10.16383/j.aas.c250094>

城市固体废物焚烧过程炉温的鲁棒加权异构特征集成预测模型

Robust Weighted Heterogeneous Feature Ensemble Prediction Model of Temperature in Municipal Solid Waste Incineration Process

自动化学报. 2024, 50(1): 121–131 <https://doi.org/10.16383/j.aas.c230042>

完备时空信息引导的工业过程内生变量缺失下的鲁棒 KPI 预测

王 凯¹ 张小刚² 雷捷维² 陈 华³

摘 要 关键性能指标 (KPI) 预测对工业过程优化和安全至关重要。然而, 在现实工业环境中传感器故障常导致推理阶段内生变量 (预测目标) 缺失, 引发信息不对称。现有方法在推理阶段因缺乏内生变量的历史自回归信息, 难以建立鲁棒时空特征映射, 严重影响多步预测性能。针对该挑战, 提出完备时空信息引导网络。该网络采用包含“完备变量引导”和“外生变量学习”的双流架构, 基于贝叶斯理论将内生变量缺失下的预测问题转换为特征对齐任务, 通过分布对齐约束使网络在变量缺失时仍能学习到逼近完备变量提取的时空表征; 同时, 提出多尺度时空聚合模块, 结合图结构学习与注意力机制动态建模变量间的耦合关系, 并精炼特征空间, 有效捕获与 KPI 相关的复杂时空关联。在电力变压器温度数据集和氧化铝回转窑烧结温度数据集上的实验表明, 在内生变量缺失下, 所提网络表现出良好的泛化能力和鲁棒的多步预测性能。

关键词 关键性能指标预测; 内生变量缺失; 特征对齐; 时空表征

引用格式 王凯, 张小刚, 雷捷维, 陈华. 完备时空信息引导的工业过程内生变量缺失下的鲁棒 KPI 预测. 自动化学报, 2026, 52(7): 1449–1462

DOI 10.16383/j.aas.c250653

CSTR 32138.14.j.aas.c250653

A Complete Spatio-temporal Information Guided Approach for Robust KPI Forecasting Under Endogenous-variable Missingness

WANG Kai¹ ZHANG Xiao-Gang² LEI Jie-Wei² CHEN Hua³

Abstract Forecasting of key performance indicator (KPI) is vital to industrial process optimization and safety. However, sensor failures often result in missing endogenous variables (the prediction target) during inference, causing information asymmetry. Existing methods struggle to build robust spatio-temporal mappings without historical autoregressive information during inference, severely impacting multi-step prediction performance. To address this challenge, the complete spatio-temporal information guided network (CSTG-Net) is proposed. Featuring a dual-stream architecture with “complete-variable guidance” and “exogenous-variable learning” branches, CSTG-Net reformulates prediction under endogenous-variable missingness as a feature alignment task based on Bayesian theory. Distribution alignment constraints enable the network to learn spatio-temporal representations approximating complete data even when variables are missing. Additionally, a multi-scale spatio-temporal aggregation module is proposed, combining graph structure learning and attention mechanisms to dynamically model variable couplings and refine the feature space, effectively capturing complex KPI-related correlations. Experiments on electricity transformer temperature and alumina rotary kiln sintering temperature datasets demonstrate that CSTG-Net maintains superior generalization and robust multi-step prediction performance under endogenous-variable missingness.

Keywords forecasting of key performance indicator; endogenous-variable missingness; feature alignment; spatio-temporal representation

Citation Wang Kai, Zhang Xiao-Gang, Lei Jie-Wei, Chen Hua. A complete spatio-temporal information guided approach for robust KPI forecasting under endogenous-variable missingness. *Acta Automatica Sinica*, 2026, 52(7): 1449–1462

收稿日期 2025-11-15 录用日期 2026-04-12

Manuscript received November 15, 2025; accepted April 12, 2026

国家自然科学基金 (62273139, U23A20385, 62171184) 资助
Supported by National Natural Science Foundation of China (62273139, U23A20385, 62171184)

本文责任编辑 那靖

Recommended by Associate Editor NA Jing

1. 湖南大学人工智能与机器人学院 长沙 410082 2. 湖南大学电气与信息工程学院 长沙 410082 3. 湖南大学计算机学院 长沙 410082

1. School of Artificial Intelligence and Robotics, Hunan University, Changsha 410082 2. College of Electrical and Informa-

在工业过程中, 关键性能指标 (key performance indicators, KPI) 是理解与优化生产系统的基础。基于分布式控制系统 (distributed control system, DCS) 中的多变量工业过程数据对 KPI 进行分析和预测, 可实现工艺过程的控制优化^[1]、设备的预测性维护^[2] 和异常工况的早期预警^[3]。因此, KPI

tion Engineering, Hunan University, Changsha 410082 3. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082

预测技术已广泛应用于化工^[4]、电力^[5]与金属冶炼^[6]等工业领域,在提升生产效率和保障安全运行等方面发挥着关键作用。

随着传感器和大数据分析技术的发展,基于数据驱动的 KPI 预测方法逐渐成为主流. 由于工业过程数据固有的非线性、动态和多变量强耦合等特点^[7],实现 KPI 的准确预测极具挑战. 现有方法可分为传统统计模型、经典机器学习方法与深度学习方法. 传统统计模型如自回归移动平均^[8]、高斯过程^[9]等,非线性拟合能力有限,难以适应工业过程的复杂特性,预测精度受限;机器学习方法如支持向量回归^[10]、随机森林^[11]等虽然增强了非线性拟合能力,但因其捕捉长程依赖和时变耦合关系能力有限,难以实现多步准确预测;近年来,深度学习凭借其强大的非线性特征提取能力,在多步 KPI 预测方面取得突出成就:以 Transformer 为代表的序列神经网络^[12-15]可有效建模长期依赖,基于分解的框架^[16-17]则增强了复杂时序模式的表示能力,动态时空图注意力网络^[18]与交叉注意力变体模型^[19-20]显式地捕捉变量间跨时空维度的交互关系. 尽管现有方法在理论与应用上取得了显著进展,但数据驱动模型的有效性严重依赖输入数据的质量. 然而,实际工业应用中的数据采集常受测量困难或传感器故障影响,出现信息粒度差异^[21]或数据缺失^[22],从而产生信息不对称问题,导致模型性能严重下降. 如何建立鲁棒的预测模型,使其在信息不对称的情况下也能保持较高的预测性能,是一个亟待解决的重要问题。

工业现场信息不对称的一种情况是信息部分缺失,可分为两种场景. 一种场景是变量数据存在数值粒度差异的不对称:采集的传感器数据由于测量困难、存储效率及其他一些限制,导致一些连续的信号被记录为离散值形式,使得输入信息由连续和离散两种混合变量组成,离散变量与连续变量之间存在数值粒度上的差异,造成变量间信息不对称. 针对这种情况,近年来研究者们有较为成熟的解决方案:如 MiTCD^[21]和 MiTSformer^[23]利用连续变量作为指导,结合对抗训练与自监督学习恢复离散变量的潜在连续性. 另一种场景是变量在时间维度有少量时刻数据缺失:由于噪声干扰、传输丢失或传感器短期故障,在部分变量序列中某些时刻或短时段出现数据丢失,致使缺失变量与完整变量在时间层面产生信息不对称. 现有研究通常利用缺失变量自身上下文信息对缺失部分进行插补或恢复其时空表征^[24],以缓解少量数据缺失导致的预测精度下降问题. 例如,庞昭辰等^[25]提出基于条件扩散模型的时间序列补全方法,通过条件扩散方法建模缺失

值的后验分布,实现对卫星遥测数据缺失值的精准插补;Dai 等^[26]通过掩码重建任务与马尔可夫动态变分自编码器,从非完整时序中提取概率化时空潜在表示;Li 等^[27]提出包含“历史重建”与“未来预测”两阶段的 Transformer 模型,通过自蒸馏机制将预测阶段学到的变量关系反馈至重建阶段,学习缺失数据与其他数据的时空表征. 上述信息部分缺失所导致的信息不对称,因仅存在信息缺失而变量本身未完全缺失,故可借助其他变量结合待补全变量未缺失的上下文,对缺失部分进行插补或学习其隐式表征,以减轻部分信息缺失对预测精度的影响。

工业现场信息不对称另一种更严重的情况是变量数目不对称. 这是由于传感器较长时间损坏,导致相应变量整段历史数据不可用,致使模型在推理与建模时可用的输入变量数目不相等. 与部分信息缺失不同,当整条变量信息完全缺失时,其自身的局部上下文无法被利用. 此时,上述基于上下文信息插补和寻找隐式表征的方法失效,插补及学习难度大,其研究相对缺乏. 一些研究者尝试采用整段填充或掩膜的方式使模型推理与建模时变量数目一致来简化问题,如 Chauhan 等^[28]通过序列的相似性计算在原训练数据中检索与缺失序列相似的片段,将其对缺失序列进行整段填充后再基于标准预测模型进行预测;Yu 等^[29]在训练时就对缺失变量进行掩膜处理,用图卷积与插值注意力从可用变量中恢复缺失变量隐式特征表示. 上述方法在一定程度上解决了当变量数目信息不对称时的 KPI 预测问题,但当面临如图 1 所示的内生变量本身(待预测 KPI)在推理阶段发生序列级缺失而仅存在外生变量(其他辅助变量)时,这些方法将面临严峻

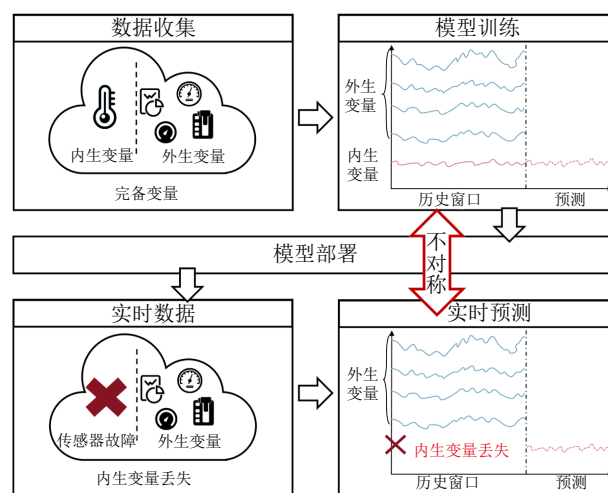


图 1 工业过程推理阶段内生变量缺失下的 KPI 预测

Fig. 1 KPI forecasting under endogenous-variable missingness at the inference stage in industrial processes

挑战.

这一挑战的根源在于变量作用存在差异——内生变量作为待预测目标, 其历史信息具有强自回归时序动态特征, 对 KPI 预测起到重要作用; 外生变量作为辅助变量, 主要提供辅助互补语境信息, 其与待预测目标的直接时序依赖通常较弱. 以上方法采取先对缺失序列填充或借助其他变量寻找隐式表达的方式来估计缺失内生变量信息, 其填充或估计误差将在后续预测步骤中被进一步放大, 且因内生变量历史信息整体缺失, 其自回归特征完全丢失, 采用辅助变量估计的时空依赖准确性低, 难以建立鲁棒的时空特征映射, 严重影响了最终的预测性能.

针对以上推理阶段内生变量缺失的变量信息不对称问题, 在其自回归信息缺失的条件下, 最大化利用训练阶段可用的完备变量信息, 将其高效迁移至外生变量表征空间, 进而实现准确且鲁棒的预测.

1) 当内生变量发生序列缺失时, 训练阶段 (变量完备) 与推理阶段 (内生变量缺失) 数据出现显著的分布偏移, 进而导致模型精度大幅下降. 如何克服这种分布偏移, 是变量信息不对称情况下的鲁棒预测模型首先要解决的问题.

2) 当内生变量缺失时, 其时间轨迹无法再为过程动态建模提供主要参考, 这不仅阻碍与内生变量相关的时间模式提取, 同时也破坏变量间的空间关联. 如何充分捕捉内生变量与外生变量间的复杂时空依赖关系, 有效识别并利用多尺度时序模式与变量结构关系, 使外生变量能够充分逼近缺失的内生变量动态, 是建立高效鲁棒预测模型的关键问题.

针对上述挑战, 本文提出一种端到端的完备时空信息引导网络 (complete spatio-temporal information guided network, CSTG-Net) 以解决工业过程内生变量缺失时导致变量信息不对称的 KPI 预测问题. 该方法借鉴特权信息思想^[30-31], 将内生变量作为特权信息参与训练, 在贝叶斯框架下将包含内生变量的完备变量提取的时空特征分布作为后验分布, 外生变量提取的特征作为先验分布, 从而将训练推理信息不对称问题建模为潜在空间时空分布对齐, 确保仅从外生变量中提取的时空表征能够有效逼近完整变量条件下的时空表征. 具体而言, CSTG-Net 采用双流变分架构, 包含完备变量引导分支与外生变量学习分支, 通过引入 Kullback-Leibler 散度 (Kullback-Leibler divergence, KLD) 约束对齐二者潜在特征分布, 使学习分支在内生变量缺失情况下仍能近似获得完备的时空表征. CSTG-Net 在训练阶段基于引导分支的时空特征进行预测, 在推理阶段则只采用学习分支进行预测. 为充

分捕捉内生变量与外生变量间的复杂时空依赖关系, 本文进一步设计多尺度时空聚合 (multi-scale spatio-temporal aggregation, MSTA) 模块以提取不同粒度的时空信息. MSTA 首先采用移动平均下采样方法得到多尺度时间序列; 继而在每一尺度上利用图结构学习构建变量间的图结构, 并借助图注意力机制获得该尺度的时空表征; 随后采用信息聚合机制对各尺度特征进行选择融合, 强化与 KPI 关联显著的模式; 最后通过多尺度特征的拼接, 形成多尺度 KPI 时空表征. 此外, 在双分支特征对齐过程中引入对抗训练机制以增强学习分支的表征学习能力. 本文主要贡献如下:

1) 针对 KPI 预测中内生变量缺失导致在变量信息不对称场景下的多步预测问题, 创新性地将其建模为贝叶斯框架下的特征对齐任务. 提出 CSTG-Net 模型, 其双流架构包含完备变量引导分支与外生变量学习分支, 通过 KLD 约束实现潜在时空表征分布对齐, 克服训练与推理情况下因变量信息不对称造成的分布偏移问题.

2) 面向工业过程中固有的复杂时空特性, 本文提出多尺度时空聚合模块以挖掘与 KPI 预测紧密关联的多粒度时空特征. 该模块通过图结构学习与图注意力机制动态建模变量间的耦合关系, 并引入信息聚合机制对特征空间进行压缩精炼, 从而有效捕获 KPI 相关的复杂时空依赖.

3) 基于氧化铝回转窑烧结温度数据集与四个公开的电力变压器温度 (electric transformer temperature, ETT) 数据集进行实验, 其结果验证了本文方法在不同数据集下的优越性能.

本文结构安排如下: 第 1 节对推理阶段内生变量缺失情况下的 KPI 多步预测问题进行形式化定义; 第 2 节详细阐述所提出的 CSTG-Net 模型; 第 3 节基于五个工业时序数据集对模型性能进行实验验证; 第 4 节总结全文.

1 推理阶段内生变量缺失 KPI 预测问题描述

本文重点研究推理阶段内生变量缺失条件下的工业过程 KPI 预测问题. 设 $\tilde{\mathbf{X}}_{1:T} = \{\mathbf{x}_{1:T}^1, \mathbf{x}_{1:T}^2, \dots, \mathbf{x}_{1:T}^C\} \in \mathbf{R}^{T \times C}$ 表示历史外生变量, 其中 T 为历史序列长度, C 为外生变量个数; $\mathbf{y}_{1:T} \in \mathbf{R}^{T \times 1}$ 为内生变量. 完备变量定义为 $\mathbf{X}_{1:T} = [\tilde{\mathbf{X}}_{1:T}, \mathbf{y}_{1:T}] \in \mathbf{R}^{T \times (C+1)}$.

在训练阶段, 内生变量与外生变量往往均可完整获取. 因此, 现有方法大多基于完备变量对未来 L 步的内生变量进行预测, 其形式如下所示:

$$\mathbf{y}_{T+1:T+L} = f(\mathbf{X}_{1:T}) \quad (1)$$

其中 $f(\cdot)$ 为工业时序预测模型。

然而, 当监测内生变量的传感器发生故障时, 数据完全缺失. 此时模型需仅依靠外生变量进行推理, 其预测目标可表述为:

$$\mathbf{y}_{T+1:T+L} = f(\tilde{\mathbf{X}}_{1:T}) \quad (2)$$

2 方法介绍

本节提出面向推理阶段内生变量缺失的 KPI 预测模型 CSTG-Net. 首先, 本文从理论上进行推导, 将该问题重构为特征对齐任务, 旨在仅从外生变量提取的时空表征逼近完整输入条件下的时空表征. 随后, 为实现该目标, 设计了如图 2 所示的双流架构, 包含完备变量引导分支与外生变量学习分支, 引导分支基于完备变量捕获时空表征, 学习分支仅以外生变量为输入. 最后, 二者通过 KLD 实现潜在表示对齐, 使学习分支与引导分支所学习的时空表征相接近.

2.1 理论分析

本文从理论层面对该问题进行建模, 将其转化为概率建模问题, 为表述简洁, 记历史外生变量为 $\tilde{\mathbf{X}}$, 历史完整变量为 $\mathbf{X}$, 未来的内生变量为 $\mathbf{y}$. 随后, 本文引入隐变量 $\mathbf{z}$, 并假设观测变量仅由隐变量决定:

$$p(\mathbf{y}|\tilde{\mathbf{X}}) = \int p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\tilde{\mathbf{X}})d\mathbf{z} \quad (3)$$

其中, $p(\cdot)$ 表示通用概率密度函数.

在实际工业场景中, 通常可从正常运行的传感

器获取变量完整的训练数据. 因此, 本文借鉴基于特权信息的博弈学习策略^[32]: 在学习阶段通过完备信息学习博弈策略, 在模型训练阶段引入完备变量作为归纳偏置. 基于贝叶斯理论, 将基于外生变量 $\tilde{\mathbf{X}}$ 的时空特征提取器建模为隐变量先验分布 $p(\mathbf{z}|\tilde{\mathbf{X}})$, 而基于完备变量 $\mathbf{X}$ 的时空特征提取器则建模为后验分布 $q(\mathbf{z}|\mathbf{X})$. 将 $q(\mathbf{z}|\mathbf{X})$ 引入 (3), 可得:

$$p(\mathbf{y}|\tilde{\mathbf{X}}) = \int \frac{q(\mathbf{z}|\mathbf{X})}{q(\mathbf{z}|\mathbf{X})} p(\mathbf{y}|\mathbf{z}) p(\mathbf{z}|\tilde{\mathbf{X}}) d\mathbf{z} \quad (4)$$

对式 (4) 两边取对数, 利用对数函数的凹性及 Jensen 不等式, 将边缘似然的对数转化为期望形式. 随后, 通过展开对数项并利用 KLD 的定义, 即可推导出证据下界 (ELBO), 即式 (5).

$$\begin{aligned} \log p(\mathbf{y}|\tilde{\mathbf{X}}) &= \log \int q(\mathbf{z}|\mathbf{X}) \frac{p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\tilde{\mathbf{X}})}{q(\mathbf{z}|\mathbf{X})} d\mathbf{z} \geq \\ E_{q(\mathbf{z}|\mathbf{X})} \left[\log \frac{p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\tilde{\mathbf{X}})}{q(\mathbf{z}|\mathbf{X})} \right] &= \\ E_{q(\mathbf{z}|\mathbf{X})} [\log p(\mathbf{y}|\mathbf{z})] - \text{KLD}(q(\mathbf{z}|\mathbf{X}) \parallel p(\mathbf{z}|\tilde{\mathbf{X}})) & \quad (5) \end{aligned}$$

基于式 (5), 推理阶段内生变量缺失下预测问题的目标转化为最大化 $E_{q(\mathbf{z}|\mathbf{X})} [\log p(\mathbf{y}|\mathbf{z})]$, 且同时最小化 $\text{KLD}(q(\mathbf{z}|\mathbf{X}) \parallel p(\mathbf{z}|\tilde{\mathbf{X}}))$, 即基于完备变量的内生变量预测任务、从完备变量中提取特征与仅从外生变量中提取特征的对齐任务.

2.2 CSTG-Net 架构

网络采用双流并行架构, 包含一个完备变量引导分支、一个外生变量学习分支及一个共享预测层. 两分支分别提取时空特征, 经变分网络与重参数化

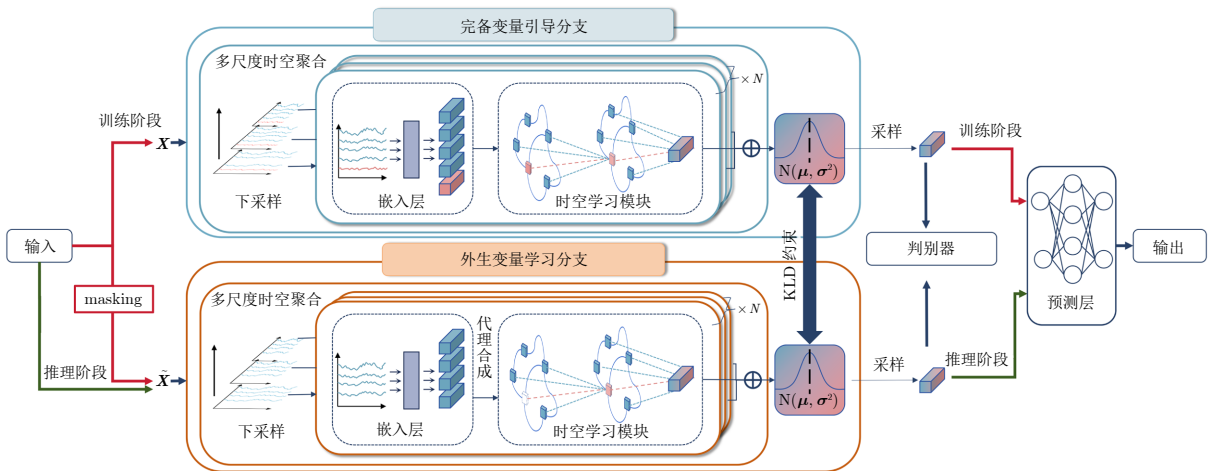


图 2 CSTG-Net 结构图

Fig.2 The structure diagram of CSTG-Net

采样获取隐变量表征. 由于先验分布与后验分布均不确定, 仅依赖对齐约束优化难以有效控制训练方向, CSTG-Net 进一步利用预测任务监督信号对引导分支增加显式约束, 这不仅防止引导分支为满足对齐约束而向低信息的学习分支退化, 也迫使学习分支在对齐损失约束下向时空信息丰富的引导分支逼近, 从而确定稳定的优化方向. 此外, 在理论框架约束下, 模型采用特权信息学习的训练推理不对称范式, 并引入外生-完备判别器做对抗训练, 实现了引导网络与学习网络的动态协同优化.

2.2.1 输入

模型在训练和推理时接受不同的输入: 训练阶段, 模型的输入为完备变量 $\mathbf{X}$, 通过掩码 (masking) 显式地分离出外生变量 $\tilde{\mathbf{X}}$. 引导分支的输入为完备变量 $\mathbf{X}$, 学习分支的输入为外生变量 $\tilde{\mathbf{X}}$. 推理阶段由于内生变量缺失, 模型仅使用 $\tilde{\mathbf{X}}$ 作为输入.

2.2.2 并行分支

引导分支与学习分支均包含 MSTA 和一个由两个多层感知机 (multi-layer perceptron, MLP) 组成的变分网络 Φ , 旨在求解时空表征分布. 首先, 各分支通过 MSTA($\cdot$) 提取聚合的时空信息:

$$\mathbf{F}_G = \text{MSTA}_G(\mathbf{X}) \quad (6)$$

$$\mathbf{F}_L = \text{MSTA}_L(\tilde{\mathbf{X}}) \quad (7)$$

其中, $\mathbf{F}_G$ 、 $\mathbf{F}_L$ 分别表示引导分支与学习分支的聚合时空表征.

进一步, 假设隐变量 $\mathbf{z}$ 服从高斯分布, 其均值向量 $\boldsymbol{\mu}$ 与对数方差向量 $\log \sigma^2$ 由变分网络 Φ 自聚合时空信息表征推断得出. 引导分支隐变量 $\mathbf{z}^G$ 继而由重参数化采样获得.

$$(\boldsymbol{\mu}_G, \log \sigma_G^2) = \Phi(\mathbf{F}_G) \quad (8)$$

$$\mathbf{z}^G \sim \mathcal{N}(\boldsymbol{\mu}_G, \sigma_G^2) \quad (9)$$

同理, 学习分支隐变量 $\mathbf{z}^L$ 也由变分网络与重参数化采样得来.

$$(\boldsymbol{\mu}_L, \log \sigma_L^2) = \Phi(\mathbf{F}_L) \quad (10)$$

$$\mathbf{z}^L \sim \mathcal{N}(\boldsymbol{\mu}_L, \sigma_L^2) \quad (11)$$

2.2.3 预测

训练阶段, 采用 MLP 捕获引导分支生成的潜在变量 $\mathbf{z}$ 到预测序列 $\hat{\mathbf{y}}$ 的映射; 推理阶段, 利用学习分支生成的潜在特征, 通过同一 MLP 进行预测. 此设计符合建模目标, 旨在使预测层从完备信息中学习最优映射关系.

$$\hat{\mathbf{y}} = \text{MLP}(\mathbf{z}) \quad (12)$$

2.2.4 对抗性外生-完备判别器

为防止后验坍塌并提升所学表征的鲁棒性, 提出一种对抗性外生-完备判别器, 其目标如式 (13) 所示.

$$\min_{\theta_L, \theta_G} \max_{\theta_{\text{Dis}}} (\mathbb{E} [\log \text{Dis}(\mathbf{z}^G)] + \mathbb{E} [\log(1 - \text{Dis}(\mathbf{z}^L))]) \quad (13)$$

其中, $\mathbf{z}^G$ 和 $\mathbf{z}^L$ 分别表示引导分支和学习分支采样得到的隐变量; θ_L 、 θ_G 和 θ_{Dis} 分别表示学习分支、引导分支以及判别器的网络参数; $\text{Dis}(\cdot)$ 为判别器.

在该博弈过程中, θ_{Dis} 的优化目标是最大化判别准确率, 即尽可能准确地将来自引导分支的 $\mathbf{z}^G$ 识别为“真”, 将来自学习分支的 $\mathbf{z}^L$ 识别为“假”; 与之相对, θ_L 和 θ_G 的优化目标则是最小化该判别概率, 即通过生成分布上难以区分的潜在特征来混淆判别器, 从而迫使学习分支的特征分布向包含完备信息的引导分支分布靠拢, 实现隐空间的对抗对齐.

2.3 多尺度时空聚合模块

2.3.1 下采样

工业过程变量内蕴含多尺度特性, 为分离并刻画其不同时间尺度下的动态趋势, 本文采用下采样构建多尺度表征, 进而实现不同尺度时空依赖关系的建模. 两个分支均通过滑动平均下采样将输入的多变量时间序列分解为 N 个多尺度时间序列.

$$\mathcal{X} = \text{DownSampling}(\mathbf{X}) \quad (14)$$

其中, $\text{DownSampling}(\cdot)$ 表示下采样操作; $\mathcal{X} = \{\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_{N-1}\}$, 其元素的下标表示下采样次数; $\mathcal{X}_n \in \mathbf{R}^{\lfloor T/2^n \rfloor \times M}$ 为第 n 次下采样序列. 最低层序列 $\mathcal{X}_0$ 为原始序列, 包含最精细的时间变化; 最高层序列 $\mathcal{X}_{N-1}$ 对应最宏观的模式, 表征时间序列趋势. 为统一步骤描述, 本节将输入序列的变量维度记为 M . 具体而言, 对于仅利用外生变量的学习分支, $M = C$, 其中 C 表示外生变量的数量; 对于利用完备变量的引导分支, $M = C + 1$.

2.3.2 嵌入

嵌入层将这些多尺度序列投影为变量维度的令牌 (token), $\mathbf{H}_n = \{\mathbf{h}_n^1, \mathbf{h}_n^2, \dots, \mathbf{h}_n^M\}$ 为第 n 次下采样的特征集合:

$$\mathbf{H}_n = \text{Embedding}(\mathcal{X}_n) \in \mathbf{R}^{M \times d} \quad (15)$$

其中 d 为嵌入维度; Embedding 表示嵌入函数, 由单层 MLP 构成, 将序列映射到嵌入维度 d .

针对学习分支中内生变量缺失的问题, 本文设计了一种代理合成机制以生成占位向量, 其结构如图 3 所示, 从而保证双流架构在拓扑结构上的统一. 该机制由前馈神经网络 (feed-forward neural net-

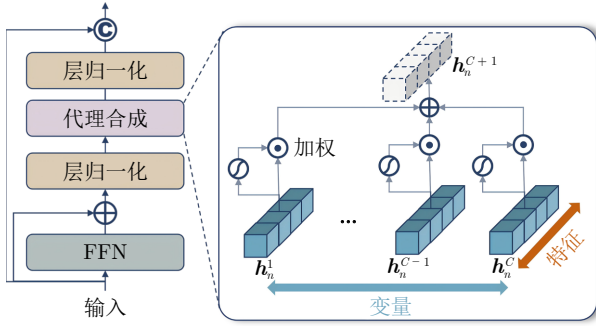


图3 代理合成机制结构

Fig.3 The structure of the proxy synthesis mechanism

work, FFN) 与代理合成层构成: FFN 通过全连接线性映射、非线性激活函数和归一化操作捕捉各变量 token 中的非线性时间依赖关系; 代理合成层则动态计算各变量的贡献权重, 进而加权聚合生成内生变量的代理向量。

$$\mathbf{G}_n = \sigma((\text{LayerNorm}(\text{FFN}(\mathbf{H}_n) + \mathbf{H}_n)) \mathbf{W}_g + \mathbf{b}_g) \quad (16)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 激活函数; $\text{LayerNorm}(\cdot)$ 表示层归一化; $\mathbf{W}_g \in \mathbf{R}^{d \times d}$ 为门控权重; $\mathbf{b}_g \in \mathbf{R}^d$ 为门控偏置; $\mathbf{G}_n \in \mathbf{R}^{C \times d}$ 是外生变量的系数矩阵。

$$\mathbf{V}_n = \mathbf{G}_n \odot \text{FFN}(\mathbf{H}_n) \quad (17)$$

$$\mathbf{h}_n^{C+1} = \text{LayerNorm} \left(\sum_{i=1}^C \mathbf{V}_n^i \right) \quad (18)$$

其中, $\mathbf{V}_n \in \mathbf{R}^{C \times d}$ 是门控后的张量; $\mathbf{h}_n^{C+1}$ 是内生变量的代理表征, 旨在融合筛选后的外生变量信息, 形成占位符与引导分支保持对称。

最终, 将生成的内生变量代理向量与外生变量的嵌入特征进行拼接, 形成完整的输入表示:

$$\mathbf{H}_n^{\text{out}} = \text{Concat}(\mathbf{H}_n, \mathbf{h}_n^{C+1}) \in \mathbf{R}^{(C+1) \times d} \quad (19)$$

其中, $\mathbf{H}_n^{\text{out}}$ 表示经代理合成后的嵌入表征; $\text{Concat}(\cdot)$ 表示拼接操作。

2.3.3 时空学习模块

各分支基于时空学习模块 (spatio-temporal learning module, STLM) 捕捉各尺度内的核心表征。由于工业过程具有高度的非线性和强耦合特性, 单一的时间维度分析难以全面刻画系统动态。特别是在内生变量缺失的情况下, 挖掘变量间的空间依赖关系成为恢复缺失信息的关键。因此, 如图4所示, STLM 被设计为通过图结构学习与注意力机制, 深度融合时空信息。

首先, 针对工业变量间拓扑结构往往未知或动

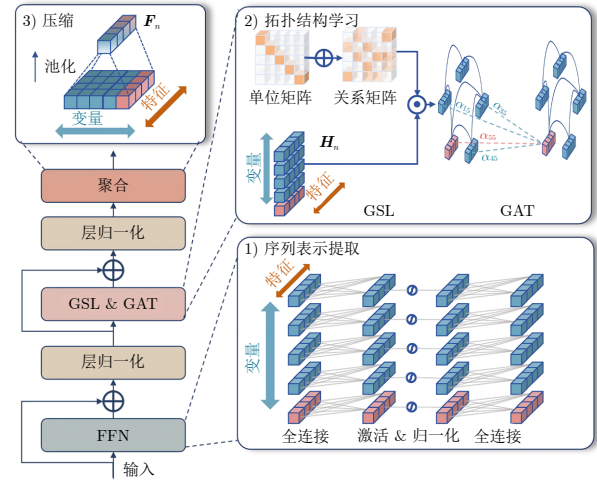


图4 STLM 结构

Fig.4 The structure of the STLM

态变化的难题, 引入图结构学习 (graph structure learning, GSL) 机制自适应学习变量间的拓扑结构。不同于依赖固定先验图的方法, GSL 通过数据驱动的方式自适应地学习变量间的潜在关联, 这通过一个可学习的节点嵌入矩阵 $\mathbf{E} \in \mathbf{R}^{(C+1) \times d_g}$ 实现, 其中 d_g 为节点关系嵌入维度。通过计算嵌入矩阵的内积生成邻接矩阵, 以度量变量间的关联强度。为在消息传递中保留节点自信息, 加入单位矩阵 $\mathbf{I}$ 。

$$\mathbf{A} = \mathbf{I} + \text{Softmax}(\text{ReLU}(\mathbf{E}\mathbf{E}^T)) \quad (20)$$

其中, $\mathbf{A}$ 表示学习到的邻接矩阵, 它为后续的空间信息聚合提供了拓扑基础; $\text{ReLU}(\cdot)$ 为激活函数。

随后, 利用图注意力网络 (graph attention network, GAT) 在学习到的拓扑结构上进行特征传播。受启发于复杂工业系统的多变量强耦合特性, 虽然内生变量的历史时序在推理阶段不可见, 但其物理状态往往被强相关的外生变量所蕴含。GAT 通过注意力机制动态聚合邻居节点的信息, 使得模型能够从外生变量的空间模式中推断出内生变量的演化趋势。

$$\alpha_{ij} = \frac{\exp(\text{ReLU}(\mathbf{a}^T \text{Concat}(\mathbf{W}\mathbf{h}_n^i, \mathbf{W}\mathbf{h}_n^j)))}{\sum_{k \in \mathcal{N}_i} \exp(\text{ReLU}(\mathbf{a}^T \text{Concat}(\mathbf{W}\mathbf{h}_n^i, \mathbf{W}\mathbf{h}_n^k)))} \quad (21)$$

其中, α_{ij} 表示变量 j 对变量 i 的注意力系数, 反映了变量间动态的空间依赖强度; $\mathbf{a}$ 表示可学习的注意力参数向量; $\mathbf{W}$ 表示可学习的权重矩阵; $\exp(\cdot)$ 表示指数函数; $\mathcal{N}_i$ 表示变量 i 的邻居集合。

最后, 基于注意力权重聚合邻居信息以更新特征表示 $\mathbf{H}_n'$, 完成空间信息与时间特征的深度融合。

$$\mathbf{h}_n^i = \text{ReLU}\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} \mathbf{h}_n^j\right) \quad (22)$$

其中, $\mathbf{h}_n^i \in \mathbf{H}_n'$, 表示尺度 n 下的变量 i 的特征表示.

在 STLMM 之后, 采用一个聚合层将时空信息压缩至单一令牌. 此设计旨在提取与内生变量预测高度相关的特征, 同时降低引导分支与学习分支对齐的复杂度. 受 SOFTS^[33] 启发, 本文采用随机池化聚合 $C + 1$ 个变量的表征, 从而获得核心表征.

$$\mathbf{F}_n = \text{StochPool}(\mathbf{H}_n') \in \mathbf{R}^d \quad (23)$$

其中, $\mathbf{F}_n$ 表示第 n 次下采样的最终聚合核心特征, $\text{StochPool}(\cdot)$ 表示随机池化.

2.3.4 拼接

为有效集成不同时间分辨率的信息, 将各尺度的核心时空表征拼接为一个时空表征 $\mathbf{F}_{\text{fused}}$.

$$\mathbf{F}_{\text{fused}} = \text{Concat}(\mathbf{F}_0, \dots, \mathbf{F}_{N-1}) \quad (24)$$

2.4 目标函数

由式 (5) 可知, 模型训练的核心目标是双流时空特征的一致性与内生变量的准确预测. 本文模型的损失函数 $\mathcal{L}$ 是预测损失 $\mathcal{L}_{\text{pred}}$ 、判别损失 $\mathcal{L}_{\text{Dis}}$ 和对齐损失 $\mathcal{L}_{\text{aligned}}$ 的加权和:

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda \cdot \mathcal{L}_{\text{Dis}} + \beta \cdot \mathcal{L}_{\text{aligned}} \quad (25)$$

其中 λ 为对抗性损失权重, β 为 KLD 权重.

$\mathcal{L}_{\text{pred}}$ 衡量预测的准确性. 训练中, 使用均方误差 (mean square error, MSE) 计算预测值 $\hat{y}$ 与真实值 y 间的差异:

$$\mathcal{L}_{\text{pred}} = \frac{1}{s} \sum_{i=1}^s (y_i - \hat{y}_i)^2 \quad (26)$$

其中 s 表示一个批次内所有预测样本点的总数.

$\mathcal{L}_{\text{pred}}$ 在联合优化中起到了关键的锚定作用. 由于在训练过程中, 引导分支是唯一负责生成预测结果的分支, 为了最小化预测误差, 它必须从完备数据中提取包含有效信息的时空特征. 这有效防止了引导分支为了降低对齐损失走捷径而向信息匮乏的学习分支坍塌.

判别损失 $\mathcal{L}_{\text{Dis}}$ 确保学习分支提取的表征以对抗方式展现出更鲁棒的时空动态特性:

$$\mathcal{L}_{\text{Dis}} = \mathbb{E} [\log(\text{Dis}(\mathbf{z}^G))] + \mathbb{E} [\log(1 - \text{Dis}(\mathbf{z}^L))] \quad (27)$$

$\mathcal{L}_{\text{aligned}}$ 是一个关键正则化项, 用于强制学习分支与引导分支间的一致性. 本文采用 KLD 最小化引导分支从完备数据推断的后验分布 $q(\mathbf{z}|\mathbf{X})$ 与学习分支仅从外生数据推断的分布 $p(\mathbf{z}|\tilde{\mathbf{X}})$ 之间的

差异.

$$\mathcal{L}_{\text{aligned}} = \text{KLD}(q(\mathbf{z}|\mathbf{X})||p(\mathbf{z}|\tilde{\mathbf{X}})) \quad (28)$$

尽管 $\mathcal{L}_{\text{aligned}}$ 同时参与双分支的梯度更新, 但受 $\mathcal{L}_{\text{pred}}$ 的约束, 引导分支被锚定在正确位置, 迫使学习分支在 $\mathcal{L}_{\text{aligned}}$ 的驱动下向引导分支单向逼近, 从而习得近似完备的时空特征.

CSTG-Net 的完整训练与推理过程如算法 1 所示.

算法 1. CSTG-Net 训练与推理流程

输入. 训练集 $\mathcal{D}_{\text{train}} = \{(\mathbf{X}, \mathbf{y})\}$, 验证集/测试集 $\mathcal{D}_{\text{infer}} = \{(\tilde{\mathbf{X}}, \mathbf{y})\}$, 引导分支 θ_G , 学习分支 θ_L , 预测层 θ_P , 判别器 θ_D .

输出. $\hat{\mathbf{y}}$

// — 阶段 1: 训练阶段 —

- 1) **while** 模型未收敛 **do**
- 2) 从 $\mathcal{D}_{\text{train}}$ 采样 $(\mathbf{X}_{\text{batch}}, \mathbf{y}_{\text{batch}})$
- 3) 通过掩码从 $\mathbf{X}_{\text{batch}}$ 提取 $\tilde{\mathbf{X}}_{\text{batch}}$
- // 完备变量引导分支
- 4) $\mathbf{F}_G \leftarrow \text{MSTAG}(\mathbf{X}_{\text{batch}}; \theta_G)$
- 5) $\mu_G, \sigma_G^2 \leftarrow \Phi(\mathbf{F}_G)$
- 6) $\mathbf{z}^G \sim \mathcal{N}(\mu_G, \sigma_G^2)$
- 7) $\hat{\mathbf{y}} \leftarrow \text{MLP}(\mathbf{z}^G; \theta_P)$
- // 仅外生变量学习分支
- 8) $\mathbf{F}_L \leftarrow \text{MSTAL}(\tilde{\mathbf{X}}_{\text{batch}}; \theta_L)$
- 9) $\mu_L, \sigma_L^2 \leftarrow \Phi(\mathbf{F}_L)$
- 10) $\mathbf{z}^L \sim \mathcal{N}(\mu_L, \sigma_L^2)$
- // 损失计算与权重更新
- 11) 根据 (25) 计算总损失 $\mathcal{L}$
- 12) 更新 $\theta_G, \theta_L, \theta_P, \theta_D$
- 13) **end while**

// — 阶段 2: 推理阶段 —

// 引导分支和判别器被停用

- 14) $\mathbf{F}_L \leftarrow \text{MSTAL}(\tilde{\mathbf{X}}; \theta_L)$
- 15) $\mu_L, \sigma_L^2 \leftarrow \Phi(\mathbf{F}_L)$
- 16) $\mathbf{z}^L \leftarrow \mu_L$
- 17) $\hat{\mathbf{y}} \leftarrow \text{MLP}(\mathbf{z}^L; \theta_P)$
- 18) **返回:** $\hat{\mathbf{y}}$

3 实验设计与结果分析

为验证所提出的 CSTG-Net 模型在工业过程内生变量缺失场景下时序预测任务中的有效性与泛化能力, 本文在源自金属冶炼领域的氧化铝回转窑烧结温度数据集与电力领域四个公开的电力变压器温度数据集^[34]上进行了广泛实验. 通过与多种先进基准模型进行对比, 以评估 CSTG-Net 在上述任务

中的性能优势. 本节首先详述所用数据集, 包括数据来源、样本规模、关键变量及划分方式; 随后, 介绍用于对比实验的基准模型; 接着说明实验的具体配置, 包括实验环境、关键超参数等; 最后展示并分析 CSTG-Net 在不同数据集上的实验结果, 以验证所提模型的有效性和泛化性.

3.1 数据集

氧化铝回转窑烧结温度数据集. 本数据集采集自金属冶炼领域的回转窑烧结工序, 数据来源于中国西部某铝厂的 DCS. 数据集采样频率为 2 min/次, 共计 15 659 条记录. 依据专家经验, 选取 7 个关键工艺变量, 其详情见表 1. 数据集按 7 : 1 : 2 的比例划分为训练集、验证集和测试集.

表 1 回转窑数据集描述
Table 1 Description of the rotary kiln dataset

变量	描述	单位	均值
x_1	窑头温度	°C	659.5
x_2	窑尾温度	°C	258.1
x_3	主电机电流	A	191.2
x_4	冷却机电机电流	A	287.7
x_5	给煤速率	rad/s	506.0
x_6	给料量	t/h	69.6
y	烧结温度	°C	950.6

电力变压器温度数据集. 本数据集为电力领域公开的数据集, 被广泛用于时序预测模型性能验证. 数据集包含四个基准数据集, 分别为 ETTh1、ETTh2、ETTm1 和 ETTm2. 其中, ETTm1 和 ETTm2 采集自中国同一省份内的两个不同地区, 时间跨度为 2 年, 采样间隔为 15 min; ETTh1 与 ETTh2 则基于相同数据源以小时为单位进行采样获得. 在每个子数据集中, 数据均按照 6 : 2 : 2 比例划分为训练集、验证集和测试集.

3.2 实验设计

为综合评估 CSTG-Net 在推理阶段内生变量缺失下的预测性能与泛化能力, 本文选取五种时间序列预测领域的代表性模型作为基准进行对比实验, 分别为长短期记忆 (long short-term memory, LSTM)^[35] 网络、Crossformer^[36]、iTransformer^[37]、CrossLinear^[38] 和 GinAR. 各模型的核心特性对比见表 2.

鉴于 LSTM、Crossformer、iTransformer 和 CrossLinear 等模型不具备直接处理整条序列缺失的能力, 本文采用业界传统解决方法, 即“插补 +

表 2 基线模型汇总
Table 2 Summary of baseline models

范式	模型	特性
循环神经网络	LSTM	天然的序列结构, 善于捕捉序列中的时间依赖
Transformer	Crossformer	两阶段注意力机制捕捉跨时间与跨变量的依赖关系
	iTransformer	将变量维度视作令牌, 高效提取多变量间的空间耦合关系
内外生变量分离建模	CrossLinear	显式分离内生与外生变量的表征学习路径, 并建模二者的交互
序列缺失预测网络	GinAR	基于插值注意力机制从可观测变量中重建缺失变量的表示

预测”的两阶段范式对模型进行增强. 具体而言, 首先引入 TimesNet (当前时序模型库 Time-Series-Library^[39] 中公开的最优插补模型) 作为内生变量恢复模块, 将其作为预处理步骤单独训练, 用于学习外生变量历史序列到内生变量历史序列的映射. 随后, 基准模型在训练阶段均使用完备数据 (包含真实内生变量), 确保其能学习到准确的时空特征并处于理论性能上限. 在推理阶段, TimesNet 被用于跨变量插补任务, 即以历史时刻的外生变量序列为输入, 以恢复同时间段的内生变量历史序列为输出目标; 随后再将恢复后的完整序列输入基准模型进行预测.

模型评价指标设置为平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean square error, RMSE). MAE 与 RMSE 均用于衡量预测值与真实值之间的偏差, 其值越低, 表明模型的预测精度越高, 性能越优.

3.3 参数设置与训练策略

本实验基于 Python 3.11 版本开发, 深度学习框架基于 PyTorch 2.0.1. 所有实验均在 Windows11 操作系统上完成, 具体硬件配置为: CPU: Intel® Core™ i5-12490F @ 3.00 GHz, GPU: NVIDIA GeForce RTX 5060. 在所有的实验中, 历史回看窗大小为 24, 预测长度为 {1, 3, 6, 12, 24}. 基线模型的超参数选择遵循其公开的训练脚本, CSTG-Net 的关键超参数是通过网格搜索 (grid search) 策略在验证集上优选确定. 同时, 对于优化器类型、Dropout 比率等通用参数, 参考时序模型库^[39] 中基准模型的官方配置, 详情如表 3 所示.

由于 CSTG-Net 引入了对抗训练机制, 为规避极大极小博弈可能引发的梯度不稳定及模型发散问题, 本文在训练过程中采用了分阶段的预热策略. 具体而言, 在训练前期, 暂不引入 KLD 约束, 仅优化预测损失与判别损失. 该设置旨在避免在特征提取能力的构建初期, 因过强的分布正则化约束而限

制了编码器对复杂时空动态的捕捉能力. 同时, 允许判别器预先确立特征边界, 为后续对抗训练夯实基础. 预热结束后, 引入 KLD 对潜在分布进行正则化约束, 联合优化引导分支与学习分支, 从而使模型在保持良好预测能力的同时实现分布对齐.

3.4 结果分析

3.4.1 有效性实验

为验证 CSTG-Net 的有效性, 本文首先在氧化铝回转窑烧结温度数据集上开展了实验. 所有模型均使用完整的数据集进行训练; 而在验证与测试阶段, 则对内生变量施加掩码处理, 以模拟推理时的缺失场景. 同时, 为确保对比公平性, 针对 LSTM、Crossformer、iTransformer 与 CrossLinear 四种基准模型, 引入 TimesNet 作为内生变量的重建网络. 实验结果如表 4 所示, 其中最优结果被加粗, 次优结果被标注下划线, 表中所列数值为 3 次不同随机种子下重复实验的平均结果.

从表 4 可知, CSTG-Net 在所有预测步长上的性能指标均显著优于其他对比模型. CSTG-Net 在

表 3 CSTG-Net 超参数设置
Table 3 Hyperparameter settings of the CSTG-Net

超参数	参数值	超参数	参数值
优化器	Adam	早停耐心值	15
学习率	0.001	预热轮数	30
批量大小	512	KLD 权重	5×10^{-4}
训练轮数	100	对抗损失权重	0.05
下采样次数	3	隐藏层维度	128

表 4 回转窑数据集不同模型在不同步长下的性能比较
Table 4 Performance comparison of different models in the rotary kiln dataset at various steps

模型	指标	预测步长				
		1	3	6	12	24
LSTM	MAE	35.571	35.583	35.592	35.609	35.649
	RMSE	45.766	45.782	45.798	45.828	45.884
Crossformer	MAE	<u>29.057</u>	<u>27.108</u>	29.141	28.749	30.058
	RMSE	<u>34.545</u>	35.413	34.772	34.969	37.993
iTransformer	MAE	35.574	35.586	35.585	35.609	35.641
	RMSE	45.768	45.783	45.789	45.825	45.874
CrossLinear	MAE	30.490	28.321	<u>26.666</u>	<u>28.433</u>	<u>29.411</u>
	RMSE	36.107	<u>34.390</u>	<u>34.301</u>	<u>34.485</u>	<u>36.164</u>
GinAR	MAE	36.748	37.237	36.857	37.537	36.521
	RMSE	47.000	47.455	46.938	47.685	46.695
CSTG-Net	MAE	26.434	26.993	26.317	26.957	27.657
	RMSE	34.431	32.323	32.290	33.163	33.536

内生变量缺失的推理场景下表现出的优越性, 主要归功于其所设计的由完整信息引导的双流架构. 同时, MSTA 模块通过多尺度分解及跨尺度空间关系提取与融合, 有效适应了工业过程的复杂时空特性, 显著提升了模型的时空建模能力.

Crossformer 和 CrossLinear 是表现次优的两个模型. Crossformer 在步长为 1 和 3 的短期预测上性能优于 CrossLinear. 而 CrossLinear 在中长期预测的 MAE 和 RMSE 上表现更好. 这表明, 无论是基于交叉注意力的变量关系捕捉, 还是显式分离内外生变量, 均有助于在一定程度上缓解因内生变量缺失导致的时间依赖模式失效问题.

LSTM 与 iTransformer 的实验结果高度一致, 其 MAE 和 RMSE 值在所有预测步长上几乎保持不变. 这说明, 无论是依赖自回归特性的 LSTM, 还是依赖变量维度 token 的 iTransformer, 均严重依赖序列自身的完整性. 当内生变量缺失时, 这两种模型的预测模式均告失效, 退化为输出一个静态的偏置结果, 未能捕获到有效的动态信息.

GinAR 模型虽然设计场景相似, 但在所有指标上表现最差. 分析其原因, 回转窑烧结过程具有显著的变量强耦合特性, 而 GinAR 依赖基于相似性计算的静态图结构, 导致模型无法动态适应当前工况. 此外, 在内生变量缺失场景中, GinAR 退化为简单的外生-内生变量映射. 由此可知, 当处于非完备变量时, 外生到内生的偏置映射会打破真正的空间关系, 诱导模型学习错误的时空关系. 因此, 完备的信息引导在内生变量缺失时是必要的.

图 5 直观地展示了不同模型在各预测步长下的 MAE 变化趋势. 由图可见, CSTG-Net 的 MAE 曲线在所有步长上均处于最低位置, 且最为平稳; 虽然 CrossLinear 和 Crossformer 在该数据集上次优, 但其误差波动较大; LSTM 和 iTransformer 高度重叠, 再次印证了它们在内生变量缺失时已退化,

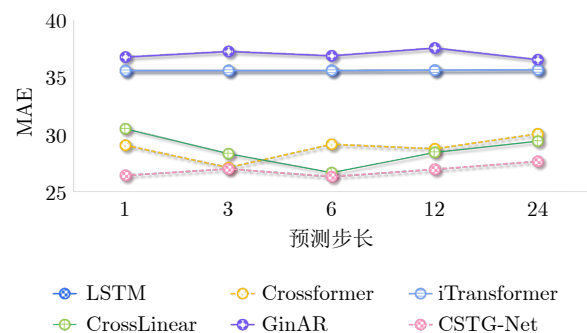


图 5 不同模型在多预测步长下的 MAE 变化

Fig. 5 Variation in MAE of different models across prediction steps

无法提供有效的动态预测; GinAR 的 MAE 值全程最高, 反映了其在处理多尺度、强耦合时空关系上的局限性.

3.4.2 泛化性实验

为了验证 CSTG-Net 在不同领域的泛化能力, 本文在电力领域公开的 ETT 数据集的四个子集 (ETTh1, ETTh2, ETTm1, ETTm2) 上进行了推理阶段内生变量缺失下的时间序列预测实验. 表 5 汇总了所有模型在四个数据集上不同预测步长的平均评估结果, 同样, 最优结果被加粗, 次优结果以下划线标注. 由表可知, CSTG-Net 在 ETT 数据集上均超越了其他对比模型, 取得最优. CSTG-Net 的 MAE, 在 ETTh1 数据集中较次优模型 Crossformer 降低 51.34%, 在 ETTh2 数据集中较次优 CrossLinear 降低 36.90%, 在 ETTm1 数据集中较次优降低 46.64%, 在 ETTm2 数据集中较次优降低 43.48%.

图 6 可视化了不同模型在 ETT 四个子集不同

预测步长下的 MAE. 在 ETT 数据集上, CSTG-Net 大幅领先, 充分证明了 CSTG-Net 在处理不同工业场景数据时强大的泛化能力. GinAR 的性能较在氧化铝回转窑烧结温度数据集中也取得了提升. 这是由于回转窑强耦合, 而 ETT 数据中的负载和油温之间的关联更为直接, 耦合程度稍弱, 因此 GinAR 能够通过插值注意力从可用变量中恢复缺失变量的表示, 故而获得了精度的提升. Crossformer 和 CrossLinear 在五个数据集上的表现也说明了内外生变量分离和精细时空建模能有效缓解变量整条缺失带来的影响. LSTM 与 iTransformer 的表现依旧最差, 再次验证了仅依赖自回归特性或利用变量维度 token 同时简化空间建模的操作无法处理序列级变量缺失下的时序预测任务.

图 7 与图 8 展示了在 ETTh1 和 ETTm2 上的预测绝对误差箱线图, 以此来多维度地展示模型性能. ETTh1 数据集的采样频率较低, 在此数据集上更能表现出不同模型对宏观变化的把控; ETTm2 数据集来源于不同城镇且采样频率较高, 在该数据集上能表现出不同模型在细粒度时间变化中的性能对比. 在 ETTh1 和 ETTm2 的箱线图中, CSTG-Net 的箱体是最窄且位置最低的, 其中位线同样是所有模型中最低的. 这与表 5 的结果相呼应, 表明 CSTG-Net 预测性能优且预测结果稳定性高. 而能够处理内生变量缺失的 GinAR 方法的中位线在 ETTh1 中较差, 但其异常点极少, 因此表明 GinAR 在处理这一问题时由于缺少内生变量历史信息引导, 其性能受限. 在 ETTm2 数据集上 GinAR 的中位线居于次优, 但箱体较大, 表明其误差存在波动, 而整体预测稳定, 没有产生明显异常结果. CrossLinear 和 Crossformer 在 ETTh1 上表现较好, 其中位线居于次优, 但都存在较多的异常预测误差. 这是由于其严重依赖内生变量历史信息, 当历史信息由插补给出进而破坏正确的时空依赖关系时, 其

表 5 不同模型在 ETT 数据集上的性能比较
Table 5 Performance comparison of different models on ETT datasets

模型	指标	ETTh1	ETTh2	ETTh1	ETTh2
LSTM	MAE	9.291	14.413	12.283	15.817
	RMSE	9.836	16.043	12.679	17.686
Crossformer	MAE	<u>6.486</u>	7.025	<u>6.162</u>	12.791
	RMSE	<u>7.344</u>	8.288	<u>7.039</u>	14.345
iTransformer	MAE	8.930	14.416	12.282	12.790
	RMSE	9.461	16.049	12.677	14.343
CrossLinear	MAE	7.049	<u>6.628</u>	7.032	14.126
	RMSE	8.089	<u>8.064</u>	7.937	15.794
GinAR	MAE	8.705	12.542	9.742	<u>10.693</u>
	RMSE	9.339	15.080	10.762	<u>13.376</u>
CSTG-Net	MAE	3.156	4.182	3.288	6.044
	RMSE	3.779	5.118	3.902	7.495

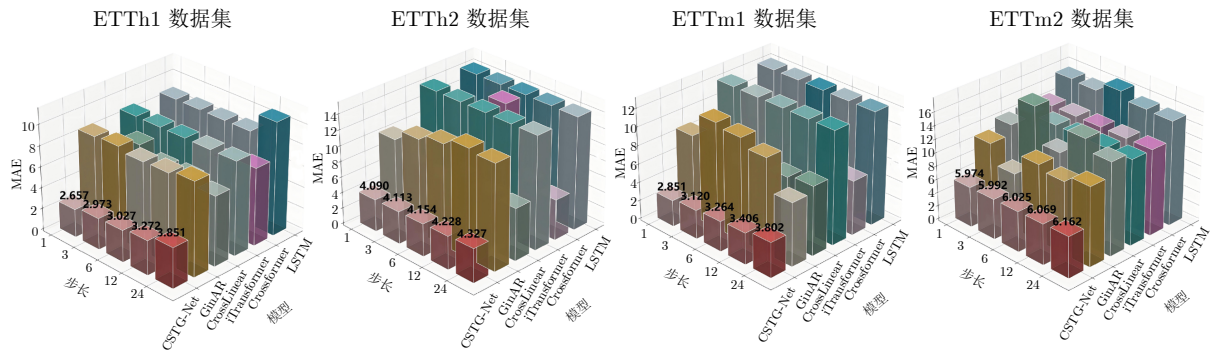


图 6 ETT 数据集的 MAE 柱状图
Fig.6 The bar chart of the MAE of the ETT datasets

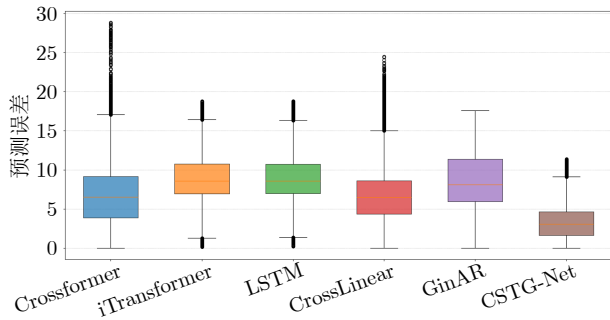


图7 ETTh1 的预测误差箱线图

Fig.7 The box plot of the prediction error for ETTh1

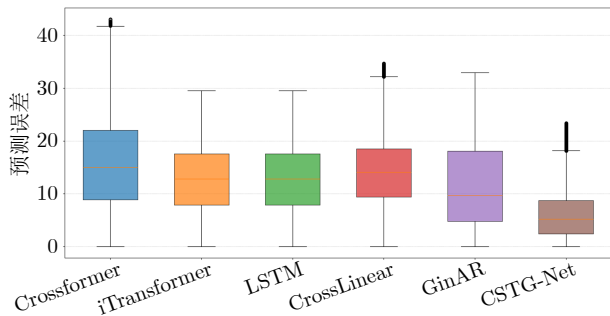


图8 ETTm2 的预测误差箱线图

Fig.8 The box plot of the prediction error for ETTm2

性能会随软测量的准确度上下波动. 而在小尺度的 ETTm2 上, CrossLinear 和 Crossformer 表现较差, 这进一步验证了二者对插补的依赖——工业过程变量往往在高频波动, 而大趋势变化缓慢, 因此小尺度的时序数据更难恢复. LSTM 和 iTransformer 在两个数据集上均具有几乎无差别的表现, 这进一步验证了在序列级变量缺失时, 高度依赖内生变量自身自回归特性的模型往往更容易出现性能崩溃.

3.4.3 消融实验

为进一步剖析 CSTG-Net 架构中各核心组件对整体预测性能的贡献, 并探究其内在作用机制, 本文选取工艺过程复杂、变量耦合程度高的氧化铝回转窑烧结温度数据集开展消融实验, 其结果如图 9 和图 10 所示. 实验设计了四种变体模型, 分别从时空特征提取、下采样与多尺度信息聚合、双流架构设计以及完备信息引导机制四个维度进行对比分析:

1) w/o STL: 剔除时空学习模块. 此变体旨在验证显式提取时空拓扑结构与动态演化特征对模型性能的增益.

2) w/o MSAM (multi-scale decomposition and aggregation module, MSAM): 剥离多尺度分解与聚合. 即去除序列的下采样分解与多尺度特征拼接操作. 此变体用于评估多粒度时空建模在捕捉

复杂工业过程动态中的重要性.

3) Guider-only: 仅保留完备变量引导分支. 在该设定下, 模型退化为传统的工业过程 KPI 预测模型, 仅在训练阶段利用完备数据, 无法显式处理推理阶段变量缺失造成的信息不对称问题. 此实验旨在验证显式构建非完备变量学习分支的必要性.

4) Learner-only: 仅保留外生变量学习分支. 此时模型退化为直接基于外生变量映射内生变量的单流网络, 缺乏完备信息的后验引导. 此实验旨在验证完备信息引导机制在缓解映射偏差方面的关键作用.

当移除 STL 后, CSTG-Net 退化为一种基础的双分支多尺度变分自编码器结构. 实验结果显示, 模型精度出现一定程度的下滑. 究其原因, 模型丢失了对复杂工业系统内部相似时间模式与潜在空间拓扑关联的显式建模能力, 即缺失了关键的时空归纳偏置. 但值得注意的是, 其性能降幅相对有限, 这表明双流架构下的完备信息引导机制在一定程度上弥补了特征提取能力的不足, 保证了基本的预测性能.

当移除 MSAM 后, 模型性能显著恶化. 单一尺度的特征表征显然难以有效支撑内生变量的精确预测. 这一结果揭示了在非完备信息下进行推断的本质规律: 从复杂系统中提取的信息层级越丰富、粒度越细致, 内生变量缺失所带来的不确定性影响便越小.

当移除学习分支, 仅保留引导分支时, 模型退化为变量信息不对称的 KPI 预测范式, 即训练推理存在信息不对称情况. 由于在推理阶段内生变量完全缺失, 模型丢失了关键的自回归历史信息, 导致预测性能大幅衰退. 尽管如此, 该变体在多步预测任务上仍优于部分基准模型, 这证明即使在内生变量缺失下, 模型也能利用学习到的完备变量间的耦合关系给出误差较小的预测结果. 但由于缺乏内生变量的自回归信息, 在缺乏专门设计的缺失适应机制下, 其长程依赖捕捉能力受限.

当移除引导分支, 仅保留学习分支时, 模型退化为仅依赖外生变量直接回归内生变量的软测量范式网络. 结果显示模型性能显著下降, 这有力地证实了仅依靠外生变量进行直接映射存在严重的映射偏差. 即便如此, 得益于多尺度时空聚合强大的特征提取能力, 该变体仍能从外生变量序列中挖掘出丰富的时空关联信息, 从而在一定程度上恢复了与内生变量演化相关的潜在特征, 避免了预测失效.

综上所述, 消融实验的结果系统地验证了在 CSTG-Net 中各组件的存在意义. 双流架构对于缓

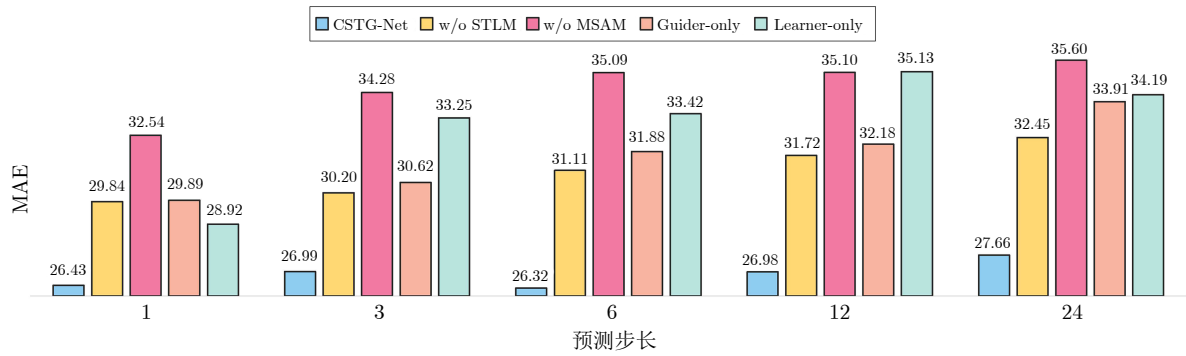


图9 消融实验 MAE 对比

Fig.9 Comparison of MAE in ablation experiments

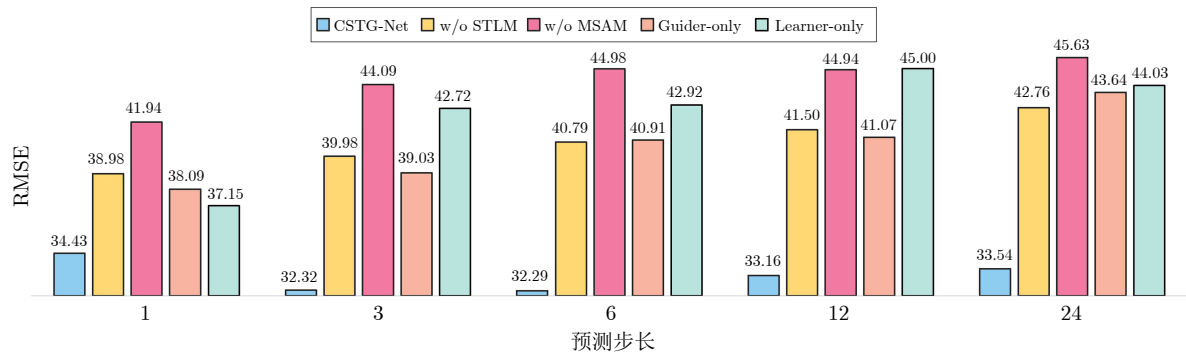


图10 消融实验 RMSE 对比

Fig.10 Comparison of RMSE in ablation experiments

解映射偏差至关重要,它通过完备变量引导学习过程,使网络能够提取更接近实际数据分布的特征;多尺度分析是解决内生变量缺失问题的基本方法,更丰富的信息层次结构有利于获取模型推理和预测所需的关键信息;显式空间建模是模型成功的基石,通过结合图结构学习和图注意力网络来显式地建模时间模式和空间关系相关的归纳偏置,它可以提升模型在复杂工业过程中提取时空特征的能力。

3.4.4 收敛性分析

如图11所示,为了验证训练策略有效性与训练过程的收敛性,本文基于公共基准数据集 ETTh1 展示了 CSTG-Net 在 24 步预测任务的对抗训练过程。图中虚线为对抗损失,表示学习分支学习的特征被判断为引导分支特征的性能;实线为判别损失,表示判别器区分两分支特征的能力。由图可见,损失曲线演化过程可分为预热与联合优化两个阶段。前 30 轮为预热阶段,此阶段未施加 KLD 约束,引导分支与学习分支提取的特征分布差异较大,判别器易于区分真实样本与生成样本,因此判别损失维持在较低水平。同时,学习分支在对抗初期尚不能有效欺骗判别器,导致对抗损失在较高值附近振荡。自第 30 轮起引入 KLD 约束,引导分支与学习分支

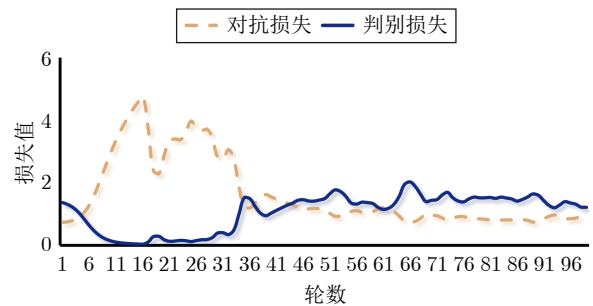


图11 对抗训练损失收敛曲线

Fig.11 Convergence curve of adversarial training loss

的特征分布被强制拉近,判别器的区分难度显著增大,判别损失迅速上升。得益于预热阶段已建立的特征基础,学习分支能够较快逼近目标分布,对抗损失随之下降。最终,两条损失曲线趋于平稳,表明学习分支与判别器在特征对齐与判别能力之间达到了纳什均衡,模型进入相对稳定的收敛状态。

尽管引入对抗训练在一定程度上增加了收敛的复杂性,但其为模型性能带来的增益是显著的。表6展示了在 ETTh1 数据集上的消融实验结果。实验数据表明,若去除对抗训练模块,模型在各预测步长上的 MAE 与 RMSE 均出现明显上升。这一结果

表 6 对抗训练在 ETTh1 数据集的消融实验

Table 6 Ablation experiment of adversarial training in the ETTh1 dataset

模型	指标	预测步长				
		1	3	6	12	24
CSTG-Net	MAE	2.657	2.973	3.027	3.272	3.851
	RMSE	3.288	3.584	3.640	3.872	4.509
去除对抗损失	MAE	3.603	3.623	4.934	5.311	6.437
	RMSE	4.165	4.180	5.575	5.933	7.028

证明, 本文采用的预热策略在有效化解训练不稳定性的同时, 实现了模型预测性能与收敛鲁棒性之间的平衡.

4 结束语

为有效解决工业过程中因传感器故障导致内生变量缺失引发训练推理变量信息不对称, 进而导致预测模型精度显著下降的问题, 本文提出一种完备时空信息引导的工业过程 KPI 预测方法 CSTG-Net. 首先, 该方法基于贝叶斯理论, 将内生变量缺失下的预测问题创新性地构建为特征对齐任务; 其次, 设计包含完备变量引导分支与外生变量学习分支的双流神经网络, 并藉由分布对齐约束, 使网络在内生变量缺失下仍能学习到接近完备的时空表征, 有效缓解了分布偏移带来的映射偏差; 此外, 本文构建了多尺度时空聚合模块, 该架构融合图结构学习、图注意力与多通道时空信息聚合机制, 充分捕捉工业时序中复杂的动态时空耦合; 最后, 为验证所提方法的有效性和优越性, 本文在某氧化铝回转窑烧结温度数据集和四个公开的电力变压器温度数据集上进行了广泛实验. 实验结果表明, 所提方法在内生变量缺失下表现出良好的性能和泛化能力, 与现有方法相比, 展示出更高的准确性, 为工业过程优化与运行安全等下游任务提供了有力支持.

参考文献

1 Zhang Hui, Yan Xing-Yu, Mao Jian-Xu, Biekezhathi Baheti, Du Rui, Wang Yao-Nan. A review of intelligent data collaborative inference techniques for source-grid-load systems. *Acta Automatica Sinica*, 2025, **51**(11): 2387–2411 (张辉, 颜星雨, 毛建旭, 别克扎提·巴合提, 杜瑞, 王耀南. 面向源网荷的智能化数据协同推断技术研究综述. *自动化学报*, 2025, **51**(11): 2387–2411)

2 Ayvaz S, Alpay K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Systems With Applications*, 2021, **173**: Article No. 114598

3 Li J B, Izakian H, Pedrycz W, Jamal I. Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing*, 2021, **100**: Article No. 106919

4 Pang H J, Ben Y W, Cao Y, Qu S, Hu C Z. Time series-based machine learning for forecasting multivariate water quality in full-scale drinking water treatment with various reagent dosages. *Water Research*, 2025, **268**: Article No. 122777

5 Huang X Q, Li Q, Tai Y H, Chen Z Q, Liu J, Shi J S, et al.

Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy*, 2022, **246**: Article No. 123403

6 Jiang Z H, Zhu J C, Pan D, Gui W H, Xu Z H. Soft sensors using heterogeneous image features for moisture detection of sintering mixture in the sintering process. *IEEE Transactions on Instrumentation and Measurement*, 2023, **72**: Article No. 2517012

7 Wang S R, Chen H, Xiong H, Wang K, Zhang X G. NAR broad learning system for dynamical systems prediction. *Neural Networks*, 2026, **198**: Article No. 108617

8 Sen P, Roy M, Pal P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy*, 2016, **116**: 1031–1038

9 Aigrain S, Foreman-Mackey D. Gaussian process regression for astronomical time series. *Annual Review of Astronomy and Astrophysics*, 2023, **61**: 329–371

10 Karmy J P, Maldonado S. Hierarchical time series forecasting via support vector regression in the European travel retail industry. *Expert Systems With Applications*, 2019, **137**: 59–73

11 Auret L, Aldrich C. Change point detection in time series data with random forests. *Control Engineering Practice*, 2010, **18**(8): 990–1002

12 Wang X M, Zhang C B, Liu Y Q, Liang X J, Yang C H, Gui W H. Advancing industrial process control with deep learning-enhanced model predictive control for nonlinear time-delay systems. *IEEE Transactions on Industrial Informatics*, 2025, **21**(9): 6823–6833

13 Zhang X G, Lei Y Y, Chen H, Zhang L, Zhou Y C. Multivariate time-series modeling for forecasting sintering temperature in rotary kilns using DCGNet. *IEEE Transactions on Industrial Informatics*, 2021, **17**(7): 4635–4645

14 Wang X M, Liu Y Q, Liang X J, Zhang C B, Yang C H, Gui W H. Learning an enhanced TCN-LSTM network for temperature process modeling in rotary kilns. *IEEE Transactions on Automation Science and Engineering*, 2025, **22**: 3056–3067

15 Shi F, Li B L, Zhang W D. Transformer with sparse mixture of experts for time-series data prediction in industrial IoT systems. *Engineering*, 2025, **17**(3): 241–258

16 Su Z Y, Zhang J, Yang Z H, Ma L H. A hybrid monthly electricity demand forecasting model combining an Hodrick-Prescott filter, recurrent neural networks, and autoregressive integrated moving average. *Energy and AI*, 2025, **22**: Article No. 100600

17 Tao X L, Liu H W, Zhao W B, Li W K, Nie Y Q, Fu J Q. GCRA-FWVAE: Anomaly detection for IIoT univariate time series using time-frequency domain analysis. *Digital Communications and Networks*, 2026, **12**(3): 405–416

18 Chen H, Jiang Y, Zhang X G, Zhou Y C, Wang L H, Wei J C. Spatio-temporal graph attention network for sintering temperature long-range forecasting in rotary kilns. *IEEE Transactions on Industrial Informatics*, 2023, **19**(2): 1923–1932

19 Zhang L, Ren G F, Li S L, Du J S, Xu D Y, Li Y H. A novel soft sensor approach for industrial quality prediction based TCN with spatial and temporal attention. *Chemometrics and Intelligent Laboratory Systems*, 2025, **257**: Article No. 105272

20 Wang J, Xie Y F, Xie S W, Chen X F. Dual cross-attention transformer networks for temporal predictive modeling of industrial process. *IEEE Transactions on Instrumentation and Measurement*, 2024, **73**: Article No. 2515211

21 Chen J W, Zhao C H. Addressing information asymmetry: Deep temporal causality discovery for mixed time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, **47**(7): 5723–5741

22 Kuan Y H, Narayanan V, Li J S. Iterative reservoir computing networks for reconstructing irregular time series. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, **36**(8): 14189–14200

23 Chen J W, Zhao C H. Addressing spatial-temporal heterogeneity: General mixed time series analysis via latent continuity re-

covery and alignment. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 569

- 24 Ma L, Wang M W, Peng K X. A spatiotemporal industrial soft sensor modeling scheme for quality prediction with missing data. *IEEE Transactions on Instrumentation and Measurement*, 2024, **73**: Article No. 3521010
- 25 Pang Zhao-Chen, Liu Ming, Zhang Li-Xian, Cao Xi-Bin, Duan Guang-Ren. Conditional diffusion model-based imputation method for missing satellite telemetry data. *Acta Automatica Sinica*, 2025, **51**(10): 2302–2312
(庞昭辰, 刘明, 张立宪, 曹喜滨, 段广仁. 基于条件扩散模型的卫星遥测数据缺失值插补方法. 自动化学报, 2025, **51**(10): 2302–2312)
- 26 Dai Q Y, Zhao C H, Huang B. M²D-VAE: Self-supervised probabilistic temporal-spatial latent representation learning for unsupervised industrial operational applications under missing value interference. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, **36**(7): 12214–12227
- 27 Li X, Li L K, Zhang K S, Chen X M, Feng T, Zhao Y, et al. Multivariate correlation self-distillation transformer for time series forecasting with incomplete data. *IEEE Transactions on Industrial Informatics*, 2025, **21**(6): 4734–4744
- 28 Chauhan J, Raghuveer A, Saket R, Nandy J, Ravindran B. Multi-variate time series forecasting on variable subsets. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington, USA: ACM, 2022. 76–86
- 29 Yu C Q, Wang F, Shao Z Z, Qian T W, Zhang Z, Wei W, et al. GinAR: An end-to-end multivariate time series forecasting model suitable for variable missing. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: ACM, 2024. 3989–4000
- 30 Vapnik V, Vashist A. A new learning paradigm: Learning using privileged information. *Neural Networks*, 2009, **22**(5–6): 544–557
- 31 Sun S L, Li M C, Wang S Y, Zhang C Y. Multi-step ahead tourism demand forecasting: The perspective of the learning using privileged information paradigm. *Expert Systems With Applications*, 2022, **210**: Article No. 118502
- 32 Han D Q, Kozumo T, Luo X F, Chen Z Y, Doya K, Yang Y Q, et al. Variational oracle guiding for reinforcement learning. In: Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2022.
- 33 Han L, Chen X Y, Ye H J, Zhan D C. SOFTS: Efficient multivariate time series forecasting with series-core fusion. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024. Article No. 2046
- 34 Zhou H Y, Zhang S H, Peng J Q, Zhang S, Li J X, Xiong H, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 11106–11115
- 35 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, **9**(8): 1735–1780
- 36 Zhang Y H, Yan J C. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.
- 37 Liu Y, Hu T G, Zhang H R, Wu H X, Wang S Y, Ma L T, et al. iTransformer: Inverted transformers are effective for time series forecasting. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.
- 38 Zhou P F, Liu Y L, Liang J L, Song Q, Li X Y. CrossLinear: Plug-and-play cross-correlation embedding for time series forecasting with exogenous variables. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2. Toronto, Canada: ACM, 2025. 4120–4131
- 39 Wang Y X, Wu H X, Dong J X, Liu Y, Wang C, Long M S, et al. Deep time series models: A comprehensive survey and benchmark. arXiv preprint arXiv: 2407.13278, 2024.



王 凯 湖南大学人工智能与机器人学院博士研究生. 2020 年获得南昌大学计算机技术专业硕士学位. 主要研究方向为复杂系统时空建模, 时序分析.
E-mail: wangkai1024@hnu.edu.cn

(WANG Kai Ph.D. candidate at the School of Artificial Intelligence and Robotics, Hunan University. He received his master degree in computer technology from Nanchang University in 2020. His research interests include complex system spatio-temporal modeling, time series analysis.)



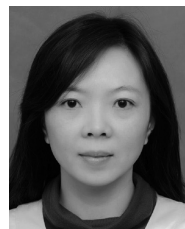
张小刚 湖南大学电气与信息工程学院教授. 2003 年获得湖南大学博士学位. 主要研究方向为工业控制系统的模式识别与数据挖掘. 本文通信作者.
E-mail: zhangxg@hnu.edu.cn

(ZHANG Xiao-Gang Professor at the College of Electrical and Information Engineering, Hunan University. He received his Ph.D. degree from Hunan University in 2003. His research interests include pattern recognition and data mining for industrial control systems. Corresponding author of this paper.)



雷捷维 湖南大学电气与信息工程学院博士研究生. 2020 年获得南昌大学计算机科学与技术专业硕士学位. 主要研究方向为强化学习, 时序分析.
E-mail: leijw@hnu.edu.cn

(LEI Jie-Wei Ph.D. candidate at the College of Electrical and Information Engineering, Hunan University. He received his master degree in computer science and technology from Nanchang University in 2020. His research interests include reinforcement learning and time series analysis.)



陈 华 湖南大学计算机学院副教授. 2014 年获得湖南大学博士学位. 主要研究方向为图像与视频处理, 计算机视觉和数据挖掘.
E-mail: chua@hnu.edu.cn

(CHEN Hua Associate professor at the College of Computer Science and Electronic Engineering, Hunan University. She received her Ph.D. degree from Hunan University in 2014. Her research interests include image and video processing, computer vision, and data mining.)