

DOI: 10.19650/j.cnki.cjsi.J2514702

基于持续蒸馏的磨煤机自适应状态监测研究*

徐 健^{1,2}, 牛玉广^{1,2}, 杜 鸣^{1,2}, 姚 珺^{1,2}, 朱国雄^{1,2}

(1. 华北电力大学新能源电力系统全国重点实验室 北京 102206; 2. 华北电力大学控制与计算机工程学院 北京 102206)

摘 要: 煤电机组磨煤机运行状态受煤质变化、负载波动、设备老化等因素影响, 容易产生动态偏移。现有监测方法普遍采用“离线建模、在线部署”方式, 适应性不足, 难以实现持续、精准的状态监测。知识蒸馏方法使用轻量化的学生模型继承复杂教师模型的优越性能, 便于实现模型快速更新与在线部署。因此, 提出了一种基于持续蒸馏的磨煤机自适应状态监测方法, 通过教师模型持续指导、学生模型在线更新的机制, 快速适应磨煤机运行状态的动态变化。针对磨煤机数据特性, 结合图卷积和时间卷积网络的特征提取优势, 提出图时卷积网络作为教师模型。构建基于复合损失函数的学生模型, 通过蒸馏损失继承教师模型知识, 使用监督损失保障监测的准确性。设计新旧参数融合策略, 基于实时数据周期性更新学生模型参数, 实现参数的迭代优化。基于某电厂运行数据验证, 结果表明: 提出方法在监测精度和自适应性方面均优于对比方法。在正常工况段, 基于持续蒸馏方式的预测残差标准差相较于离线建模方式平均降低了 8.45%, 显著提升了模型的稳定性; 在异常工况段, 提出方法在保持零误警率的前提下, 提前 116 h 捕捉到故障征兆并发出预警信号。综上所述, 所提方法能够显著提升设备运维的智能化水平, 具有广阔的工程应用前景。

关键词: 磨煤机; 状态监测; 知识蒸馏; 持续学习; 图卷积网络

中图分类号: TH17 TM621 **文献标识码:** A **国家标准学科分类代码:** 470.2010

Research on adaptive condition monitoring of coal mill based on continuous distillation

Xu Jian^{1,2}, Niu Yuguang^{1,2}, Du Ming^{1,2}, Yao Jun^{1,2}, Zhu Guoxiong^{1,2}

(1. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China; 2. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: The operation status of coal mills in coal-fired power plants is influenced by factors such as coal quality variations, load fluctuations, and equipment aging, which can lead to dynamic shifts. Existing monitoring methods typically rely on the “offline modeling, online deployment” approach, which is difficult to realize the adaptive, continuous and precise condition monitoring. It's known that knowledge distillation methods use lightweight student models to inherit the superior performance of complex teacher models, facilitating the rapid model updates and online deployment. Therefore, we propose an adaptive condition monitoring method for coal mills based on the continuous distillation. This method enables the quick adaptation to dynamic changes of coal mill during the operation process by continuously guiding the student model through the teacher model and updating the student model online. Considering the characteristics of coal mill data, we propose the graph temporal convolutional network as the teacher model by combing the feature extraction advantages of graph convolutional networks and temporal convolutional networks. The student model is constructed based on a composite loss function, inheriting the knowledge from the teacher model via distillation loss and ensuring the monitoring accuracy with supervised loss. The new parameter fusion strategy is designed to periodically update the student model's parameters based on real-time data, achieving the iterative optimization of the parameters. Validation with operational data of power plant shows that the proposed method outperforms comparison methods in both monitoring accuracy and adaptability. At the normal operating conditions, the standard deviation of prediction residuals under the continuous distillation approach is reduced by an average of 8.45% compared to the offline

收稿日期: 2025-11-21 Received Date: 2025-11-21

* 基金项目: 国家自然科学基金(52206009)、国家重点研发计划(2022YFB4100404)项目资助

modeling method, significantly enhancing the stability of model. While in the abnormal operating scenarios, the proposed method successfully captures the fault symptoms and issues early warning signals 116 hours in advance, while maintaining a zero false alarm rate. In conclusion, the proposed method can improve the intelligence level of equipment operation and maintenance, demonstrating the broad prospects for engineering applications.

Keywords: coal mill; condition monitoring; knowledge distillation; continual learning; graph convolutional network

0 引 言

在新型电力系统建设背景下,煤电机组正在从主力电源向灵活调节电源转型^[1]。近期发布的《新一代煤电升级专项行动实施方案(2025—2027年)》中指出:要夯实煤电兜底保障作用,积极推进煤电转型升级,并对煤电机组中关键设备/部件的智能运维水平提出明确要求。磨煤机作为煤电燃烧系统中的关键设备之一,其运行状态直接影响机组的燃烧效率,并与锅炉的负荷调节、燃烧稳定性密切相关^[2-3]。因此,对磨煤机的运行状态监测与故障预警十分必要。

煤电机组通常配备测点齐全的分布式控制系统,对运行过程中的各类参数进行数据采集和实时监控^[4]。近年来,许多学者通过结合运行数据与机器学习方法,实现对磨煤机的状态监测与早期故障预警。文献[5-8]中提出了多种创新的数据驱动模型,如:卷积神经网络-长短时记忆神经网络、改进鲸鱼算法优化 Transformer 网络、TimeGAN-自适应门递归单元、套索回归特征选择与双向长短期记忆网络。这些模型基于正常行为建模思想,通过“离线建模、在线部署”方式进行磨煤机状态监测和故障预警研究。离线建模阶段,使用历史数据对监测模型进行训练和优化;在线部署阶段,将训练好的模型应用于实时数据流中,监测磨煤机的运行状态并进行故障预警。

然而,上述研究均基于训练与测试数据同分布的假设。但在实际应用中,磨煤机的运行易受煤质变化、负载波动、设备老化等多种因素影响,导致其在线工作状态常常偏离训练工况。这种数据分布偏移引发的模型失配问题会显著影响监测精度和预警效果,使得无法准确识别出潜在的故障或异常。因此,提升模型在面对数据分布偏移或工况变化时的自适应能力,成为了实际应用中亟待解决的问题之一。

近期,部分学者尝试解决数据分布偏移引起的模型失配问题。文献[9]通过概念漂移监测与增量更新机制进行模型迭代优化,提升了波动场景下风电功率预测的准确性。文献[10]提出了一种自适应故障检测方法,通过识别电池组在变工况条件下正常行为的变化,从而有效减少模型的误报率。文献[11]提出了一种基于持续迁移学习系统的故障诊断方法,解决了非独立同分布连续流数据的诊断问题,提升了机械故障诊断精度。文

献[12]提出一种基于知识蒸馏的电力系统暂态稳定评估持续学习框架,自适应电网数据流的变化,从而提高评估精度与实时响应能力。磨煤机状态监测中同样存在这一问题,但目前尚未得到充分研究与讨论。

为增强磨煤机在数据分布偏移场景下的状态监测准确性,提出一种基于持续蒸馏的自适应状态监测方法。该方法基于持续学习和知识蒸馏思想,结合实时数据在线优化模型参数,以灵活应对磨煤机运行状态的动态变化。针对磨煤机中的多变量时序数据,结合图卷积和时间卷积网络特征提取优势,提出图时卷积网络作为教师模型。构建基于复合损失函数的学生模型,通过蒸馏损失继承教师模型知识,使用监督损失保障监测的准确性。在线监测阶段,通过新旧参数融合策略周期性更新学生模型参数,实现对数据分布偏移场景的快速适应,显著提升故障预警的灵敏度与准确性。通过某电厂的实际运行数据,对所提出方法的有效性进行了验证和讨论。

1 理论基础

1.1 知识蒸馏

知识蒸馏是一种结合知识迁移与模型压缩思想的深度学习学习方法,于2015年由Hinton等^[13]提出,已被广泛应用于计算机视觉与自然语言处理领域^[14]。其核心思想是将复杂教师模型中提取的丰富知识迁移至轻量化学生模型中,使得在保证较高精度的同时提升计算效率和泛化能力。如图1所示,知识蒸馏通过教师-学生结构实现知识迁移,使学生模型在学习真实数据的同时继承教师模型的知识,从而实现模型性能与计算效率的平衡。

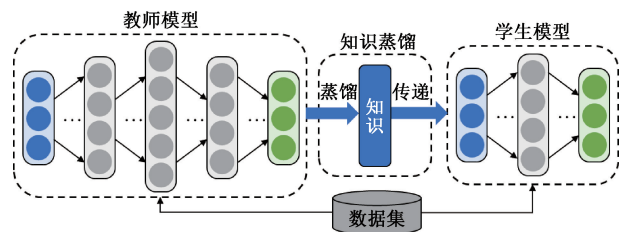


图1 知识蒸馏原理

Fig. 1 Knowledge distillation principle

在训练阶段,先使用结构复杂、性能较强的教师模型对历史数据进行学习,获取高精度的预测结果。随后,轻

量化学生模型以目标数据作为硬标签、以教师模型输出作为软标签,联合优化监督损失和蒸馏损失^[15]。对于多变量时序预测任务,模型训练过程中的复合损失函数如式(1)所示。

$$L_{KD} = (1 - \alpha) \|y - \hat{y}_s\|_2^2 + \alpha \|\hat{y}_t - \hat{y}_s\|_2^2 \quad (1)$$

式中: y 为真实观测值; \hat{y}_t 为教师模型预测结果; \hat{y}_s 为学生模型预测结果; $\alpha \in [0, 1]$ 为蒸馏平衡系数,用于调节监督损失和蒸馏损失的权重。

上述过程中,学生模型不仅能够学习到原始数据的规律,还能够吸收教师模型捕捉到的隐含特征。相比直接训练的模型,经过知识蒸馏的学生模型在保证精度的同时显著降低了计算量和存储需求,适合在资源受限环境中部署。因此,知识蒸馏不仅是模型压缩的有效方法,也为模型在线更新、工业实时推理提供了可行的解决方案。

1.2 图卷积网络

卷积神经网络(convolutional neural network, CNN)凭借其卓越的特征提取能力,被广泛应用于各类欧式结构数据的分析与建模。为将卷积思想进一步推广至具有非欧几里得结构特征的图数据,提出了图卷积网络(graph convolutional network, GCN)^[16]。如图2所示,GCN通过聚合目标节点自身特征及其邻居节点特征,能够自动捕捉图中节点间的关联信息。

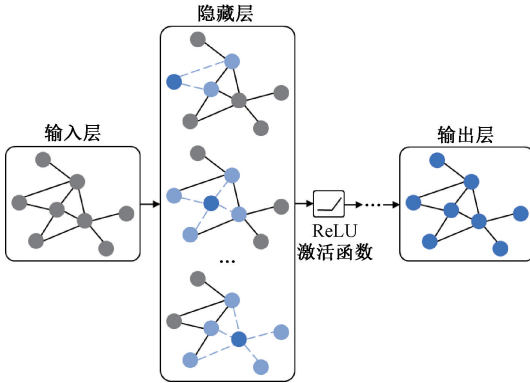


图2 图卷积网络结构

Fig. 2 Graph convolutional network structure

通过对图拉普拉斯算子的多项式进行近似一阶化简,Kipf等^[17]在谱域图卷积理论的基础上提出了简化版GCN,有效降低了模型计算复杂度。GCN的更新公式如式(2)所示。

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)} + \mathbf{b}^{(l)}) \quad (2)$$

式中: $\mathbf{H}^{(l+1)}$ 和 $\mathbf{H}^{(l)}$ 分别表示第 $l+1$ 层和第 l 层节点特征矩阵; $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ 表示在邻接矩阵 \mathbf{A} 上添加自连接后的矩阵; $\tilde{\mathbf{D}}$ 为对应的度矩阵; $\sigma(\cdot)$ 为非线性激活函数; $\mathbf{W}^{(l)}$ 和 $\mathbf{b}^{(l)}$ 分别为第 l 层可学习的参数矩阵和偏置。

相比传统CNN方法,GCN能够在特征学习过程中融合节点自身属性与邻域节点信息,实现多变量间的关联建模与特征提取。该结构不仅保留了局部特征的表达能力,还能通过图拓扑关系捕捉变量间的潜在依赖性,从而在复杂系统的多维状态表征中表现出更高的准确性与鲁棒性。

1.3 时间卷积网络

传统CNN和GCN在处理序列数据时存在局限,无法有效捕捉长时间跨度的动态特征。为此,时间卷积网络(temporal convolutional network, TCN)被提出,通过引入因果卷积和扩张卷积机制,在保持时间顺序一致性的同时大幅扩大感受野,实现对长序列动态特征的快速提取^[18]。

如图3所示,TCN以残差块为基本单元堆叠而成。每个残差块包括两个扩张卷积层,每层均由扩张因果卷积、权重归一化、ReLU激活函数、Dropout正则化组成,最后通过残差连接实现输入与输出特征的融合。

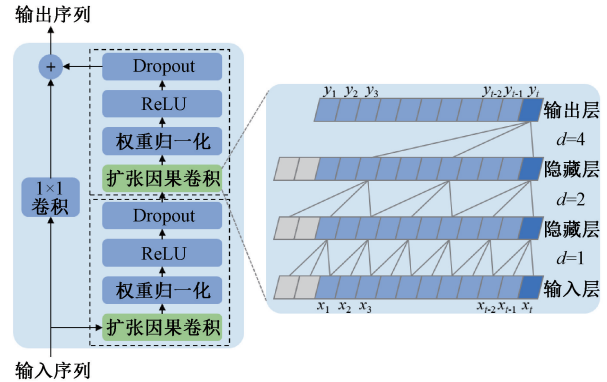


图3 时间卷积网络结构

Fig. 3 Temporal convolutional network structure

扩张卷积中,假设输入序列为 $x = [x_1, x_2, \dots, x_t]$,卷积核权重为 $w = [w_0, w_1, \dots, w_{k-1}]$,则在扩张系数 d 下的扩张卷积定义如式(3)^[19]所示。

$$(x * {}_d w)(t) = \sum_{i=0}^{k-1} w_i \cdot x_{t-d \cdot i} \quad (3)$$

式中: $*$ 表示卷积; k 为卷积核大小; d 决定感受野跨度。当 d 逐层按指数增长时,感受野随层数迅速扩大,使模型能够捕获长时依赖特征。

为保证时间因果性,TCN仅使用当前及历史信息进行卷积,即输出 y_t 只与当前和过去时刻的输入相关,从而避免未来的信息泄露到输入中,影响模型预测的准确性。因果表达如式(4)所示。

$$y_t = f(x_t, x_{t-1}, \dots, x_{t-(k-1)d}) \quad (4)$$

此外,为防止深层网络造成梯度消失问题,引入残差结构,每个残差块的输出如式(5)所示。

$$y = \text{ReLU}(x + F(x)) \quad (5)$$

式中: $F(\cdot)$ 表示两层扩张因果卷积及权重归一化和非线性变换的组合映射。

TCN 通过扩张卷积实现对大范围时序数据的建模,在保持时间因果性的同时有效捕获长时依赖特征。相比循环神经网络结构,TCN 具备并行计算、高稳定性和高训练效率等优势,能够在复杂时序任务中实现更快收敛与更强泛化性能。

2 基于持续蒸馏的自适应状态监测方法

2.1 图时卷积网络

为充分挖掘多变量时序数据中蕴藏的状态信息,结合上述 GCN 和 TCN 的特征提取优势,提出一种新颖的图时卷积网络 (graph temporal convolutional network, GTCN),该网络的结构图如图 4 所示。

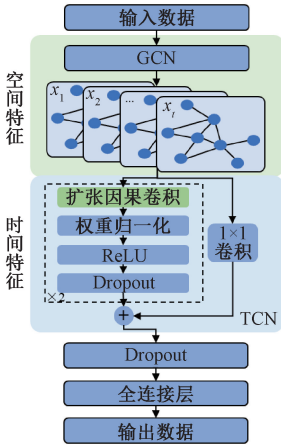


图 4 图时卷积网络结构

Fig. 4 Graph temporal convolutional network structure

如图 4 所示,多变量时序数据输入 GTCN 模型后,首先经过 GCN 模块,根据变量间的非欧几里得图结构,对每一时刻的数据进行关联建模与空间特征提取。随后,将提取到的空间特征送入 TCN 模块,在保证时间因果性的基础上,通过扩张卷积高效捕获多尺度时间依赖特征。为增强模型的泛化能力并抑制过拟合风险,在 TCN 后引入 Dropout 层进行随机失活处理。最后,通过全连接层实现特征融合,并对下一时刻的数据进行预测。

提出的 GTCN 模型主要有 3 点优势:1) 在特征提取方面,融合图卷积和时间卷积结构,充分挖掘多变量时序数据中的空间和时间特征,实现对设备运行状态的精细刻画;2) 在计算效率方面,模型采用卷积式并行计算框架,提升了训练和推理速度,满足在线监测对实时性的要求。3) GTCN 模型相比于时空图卷积网络具有更简单的架构,减少了计算复杂度,适合处理空间与时间依赖性相对独立的任务。这种简化的架构不仅使得 GTCN 在处理

大规模时序数据时具备更高的计算效率,还能有效降低过拟合的风险。在空间依赖性较为固定的场景中,如设备健康监测或故障预警,GTCN 模型能够提供更加快速且精准的预测。

2.2 持续蒸馏自适应机制

在设备运行过程中,受原料变化、负载波动、寿命损耗等因素影响,其在线工作状态常常偏离训练工况,产生的模型失配问题会影响监测精度和预警效果。为此,在知识蒸馏的基础上,提出一种持续蒸馏自适应机制。其中,教师模型提供稳定的知识指导,学生模型通过持续学习实现快速参数更新,在线适应数据分布偏移。图 5 为提出的持续蒸馏自适应机制原理图。

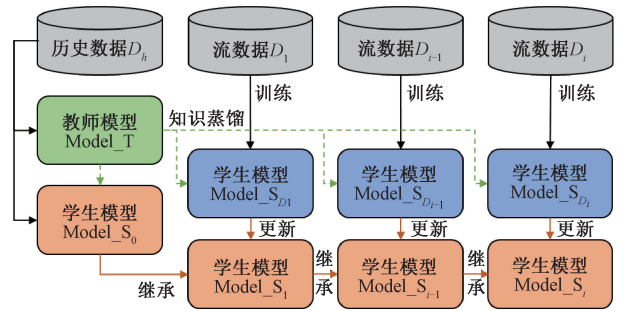


图 5 持续蒸馏自适应机制

Fig. 5 Adaptive mechanism of continuous distillation

如图 5 所示,持续蒸馏自适应机制通过固定教师模型与逐步更新的学生模型实现对数据分布偏移的在线适应。首先,基于历史数据 D_h 对教师模型 $Model_T$ 进行离线训练,充分学习数据中的特征规律,并在 $Model_T$ 指导下训练初始学生模型 $Model_S_0$ 。随后,将在线采集到的流数据 D_1, D_2, \dots, D_i 依次输入,用于学生模型自适应更新。在每个阶段的 D_i 中,学生模型 $Model_S_{D_i}$ 在 $Model_T$ 的指导下基于复合损失函数进行训练,并通过继承 $Model_S_{i-1}$ 的参数实现模型在线更新,提出的新旧参数融合策略如式(6)所示。

$$Model_S_i = (1 - \beta) \cdot Model_S_{i-1} + \beta \cdot Model_S_{D_i} \quad (6)$$

式中: $Model_S_i$ 为更新后的模型; $\beta \in [0, 1]$ 为自适应更新权重系数,用于控制学生模型在持续蒸馏过程中,对历史参数的依赖程度与对新数据特征的学习速率,以保证在线适应的平滑与稳定。

在持续蒸馏自适应机制中,学生模型的更新没有设置固定的迭代终止条件,而是随着在线数据的波动不断进行自适应调整。每次更新过程中,学生模型将继承上一阶段的参数,并在教师模型的指导下,融合新的数据特征进行训练。此外,模型的更新采用了自适应更新权重

系数,确保历史知识与新数据的平衡,从而避免模型过度依赖历史信息或出现过拟合现象。

通过持续蒸馏自适应机制,教师模型提供稳定可靠的知识约束,学生模型基于流数据进行持续学习与参数更新,从而实现对数据分布偏移的快速适应。该方法在保持模型稳定性的同时提升了模型的自适应能力,可在数据分布偏移时持续保持监测模型的高精度与鲁棒性。

2.3 自适应状态监测方法

在上述内容的基础上,基于正常行为建模思想,提出一种基于持续蒸馏的自适应状态监测方法,如图6所示。依据数据来源,将状态监测流程分为离线和在线两个阶段。其中,教师模型为提出的 GTCN,学生模型采用门控循环单元(gated recurrent unit, GRU)网络。

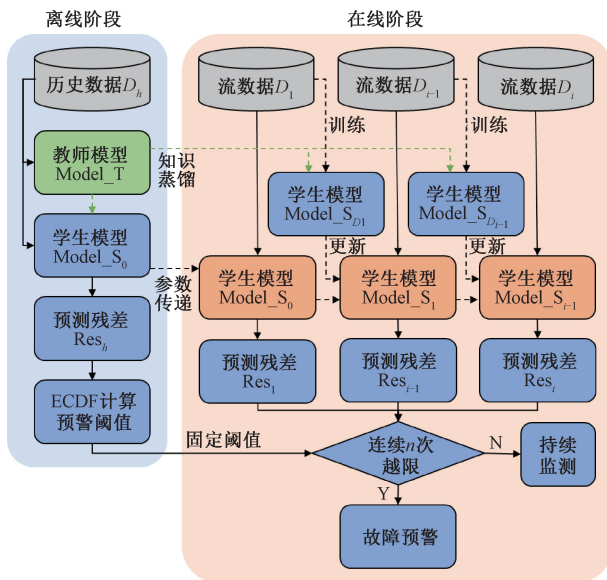


图6 基于持续蒸馏的自适应状态监测方法

Fig. 6 Adaptive condition monitoring method based on continuous distillation

ECDF是一种基于样本数据的非参数统计方法,用于刻画随机变量取值的累积分布特征^[20]。对于一组样本残差 $\{r_1, r_2, \dots, r_n\}$, 其ECDF定义如式(7)所示。

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(r_i \leq x) \quad (7)$$

式中: $I(\cdot)$ 为指示函数,当 $r_i \leq x$ 时取值为1,否则为0。该函数描述了样本中 $\leq x$ 的比例,直观反映残差分布的累积概率特性。本文中,选取 $\geq 99.9\%$ 分位点的残差值作为故障预警阈值,有效抑制随机波动对监测结果的影响,实现对设备异常状态的灵敏检测。

在线阶段,流数据 D_i 基于学生模型 $\text{Model}_{S_{i-1}}$ 实时计算预测残差 Res_i 。当残差连续 n 次越限时发出故障预警信号,提醒操作人员及时关注设备潜在异常。当流数

据 D_i 达到设定容量时,采用知识蒸馏方法训练学生模型 $\text{Model}_{S_{D_i}}$,并通过新旧参数融合策略更新学生模型 Model_{S_i} 。此外,为避免将异常状态数据引入训练样本中,在学生模型 Model_{S_i} 训练前,根据 Res_i 状态监测结果隔离异常样本数据。如此,确保学生模型在持续更新过程中对正常运行模式的准确学习,实现对设备正常状态的稳定跟踪与快速自适应。

在持续蒸馏自适应状态监测方法中,模型的大部分参数通过在线学习和自适应更新进行调整。然而,针对不同的应用场景,特别是数据波动较大或环境变化剧烈的情况下,部分超参数,如蒸馏平衡系数和自适应更新权重系数,可能需要人工干预和微调,以确保模型在这些特殊工况下的稳定性。

通过部署提出的基于持续蒸馏的自适应状态监测方法,模型能够在设备运行中保持高精度与高鲁棒性的状态感知能力,为数据分布偏移时设备的状态监测和早期故障预警提供有力支撑。

3 案例分析

以华东某煤电机组为研究对象,该机组主用磨煤机B于2023年11月17日03:09跳闸停机。从厂级分散控制系统中提取磨煤机B相关运行数据,时间跨度从2023年10月1日00:00至磨煤机跳闸停机时刻,时间间隔为1min。基于该数据集,验证提出方法在状态监测中的自适应能力与早期预警性能。

3.1 数据预处理

为模拟实际应用中的监测场景,将数据集划分为训练集和测试集两部分。具体而言,选取前14天数据作为训练集,用于模型训练、参数优化和阈值计算;后续34天数据作为测试集,模拟在线流数据,评估提出方法的状态监测效果。

数据预处理共包括数据清洗、变量选择、数据归一化、样本切分这4个步骤。首先进行数据清洗,剔除不完整的数据记录和离群野点,为后续研究提供可靠的数据基础。变量选择时,根据皮尔逊相关系数和现场运行人员经验,选择与磨煤机运行相关的状态变量,如表1所示。

各变量取值范围参差不齐,在模型训练时容易产生局部最优问题。为统一数据量纲,使各变量在相同尺度下发挥作用,采用0~1归一化方法。对于某个变量的原始数据值 x ,其归一化后的值 x' 计算公式如式(8)所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

式中: x_{\max} 和 x_{\min} 分别是该变量的最大值和最小值。

表1 磨煤机相关状态变量

Table 1 Coal mill related state variables

符号	变量名称/单位	取值范围
V1	给煤管入口给煤量/(t·h ⁻¹)	[31.10, 75.90]
V2	磨煤机电流/A	[26.32, 48.42]
V3	电机轴承温度/°C	[30.19, 50.10]
V4	入口一次风温度/°C	[139.21, 343.78]
V5	入口一次风流量/(m ³ ·s ⁻¹)	[95.61, 147.02]
V6	入口一次风压力/kPa	[4.62, 11.91]
V7	密封风与一次风压差/kPa	[2.43, 9.40]
V8	发电机功率/MW	[374.58, 996.47]

随后,对训练集和测试集数据进行样本切分。每个样本由连续17个时刻的多变量数据构成,其中前16个时刻的数据作为输入,最后时刻的数据作为目标输出。此外,根据专家在故障后的判定结果,以11月13日0时为界,将测试集划分为正常段和异常段。如表2所示,详细介绍了各样本集的时间范围、样本个数及其具体作用。

表2 样本切分结果

Table 2 Sample segmentation results

样本集	时间范围	样本数/个	作用
训练集	10:01 至 10:14	20 043	初始模型训练优化
测试集-正常段	10:15 至 11:12	40 286	模型自适应性评估
测试集-异常段	11:13 至 11:17	7 360	模型预警效果评估

样本切分后,基于主成分分析法对训练和测试样本进行降维,将高维特征映射为二维主成分^[21]。具体来说,以主成分1为横轴,主成分2为纵轴,训练和测试样本分布情况如图7所示。

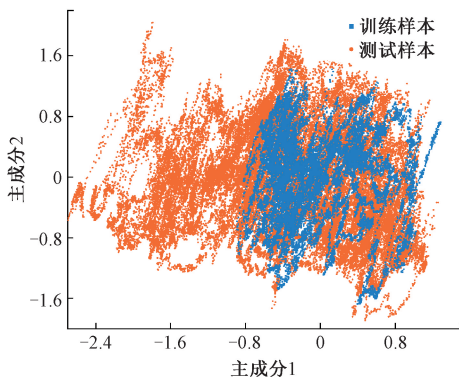


图7 基于主成分分析降维的训练和测试样本分布情况
Fig. 7 The distribution of training and testing samples with the dimensionality reduced principal component analysis

如图7所示,训练集和测试集的样本在主成分2上的分布大致相同,而在主成分1上则存在显著差异。当主成分1<-0.8时,几乎没有训练样本,而测试样本则较为集中。如果仅依赖当前训练集进行训练,模型的适应性较差,可能会导致模型在测试集中失配,表现为预测精度下降或误差增大。针对这一问题,需考虑通过自适应方法来提高模型的泛化能力,以确保其在测试集上的准确性和稳定性。

3.2 参数设置与指标确立

基于PyTorch深度学习框架,构建了教师模型GTCN和学生模型GRU。在GTCN模型中,邻接矩阵的构建采用了最大相关系数与Top-K最近邻方法^[22],结合实践经验,将K值设置为4。GTCN模型中,单层GCN用于聚合当前节点及其最相关的4个邻居节点的局部特征。时间特征的提取通过双隐藏层的TCN完成,其中隐藏层单元数设置为24,卷积核大小为3,步长为1,并采用因果填充方式,在处理时序信息时避免未来信息泄漏。此外,Dropout层的丢弃率设置为0.01,以防止过拟合。全连接层由双层网络组成,单元数分别为16和8。在GRU模型中,模型的输入层接收前16个时刻的多变量数据,输出为最后一个时刻的预测值。为提高模型的特征提取能力,GRU的隐藏层单元数设置为64。为防止过拟合,采用了Dropout层,丢弃率设置为0.05。GRU模型的输出层使用双层全连接网络,单元数设置为16和8,用于输出最终的预测结果。

在模型训练过程中,训练样本会被随机打乱,20%的样本用作验证集。为加速模型的收敛,batch_size设置为32,即每批次输入32个样本。在梯度下降优化过程中,采用Adam优化器,初始学习率为0.003。训练周期epochs设置为200,并采用提前停止策略,即当验证集损失在连续10个周期内未出现改善时,训练过程将自动终止,防止模型过拟合。超参数蒸馏平衡系数 α 和自适应更新权重系数 β 分别设置为0.6和0.3。

模型在线部署阶段,使用测试集样本模拟在线流数据,当流数据覆盖时间跨度为24h,生成一组模型自适应更新数据集 D_t 。选择24h作为边界的原因主要是电厂设备通常具有明显的日周期性,24h能够涵盖完整的工作周期,包括负荷波动和设备状态的变化。此外,24h作为数据流的时间窗口,能够平衡模型的实时性和数据的充分性,避免过短的时间跨度无法反映设备状态变化趋势,或过长的时间跨度导致模型更新滞后。在学生模型Model S_d 训练前,根据Res $_i$ 状态监测结果隔离异常样本。本文中,当预测残差连续10次越限时发出故障预警信号,对应的样本也被标记为异常样本。

指标计算方面,选择均方根误差(root mean square error, RMSE)衡量模型预测值与实际观测值之间的残

差,残差值 res 计算公式如式(9)所示。

$$res = \frac{1}{m} \sum_{j=1}^m \sqrt{(\hat{x}^j - x^j)^2} \quad (9)$$

式中: m 为样本变量数; \hat{x}^j 和 x^j 分别为变量 j 的预测值和实际值。

此外,通过以下4个指标定量计算模型的状态监测和故障预警效果,即:正常段残差均值(mean of normal stage residuals, MNR)、正常段残差标准差(standard deviation of normal stage residuals, SNR)、故障检测率(fault detection rate, FDR)和误警率(false alarm rate, FAR),计算公式如式(10)~(13)所示。

$$MNR = \frac{1}{n_{nor}} \sum_{i=1}^{n_{nor}} res_i \quad (10)$$

$$SNR = \sqrt{\frac{1}{n_{nor}} \sum_{i=1}^{n_{nor}} (res_i - \mu)^2} \quad (11)$$

$$FDR = \frac{alarm_{abn}}{n_{abn}} \times 100\% \quad (12)$$

$$FAR = \frac{alarm_{nor}}{n_{nor}} \times 100\% \quad (13)$$

式中: res_i 为第 i 个样本的残差值; n_{nor} 和 n_{abn} 分别为正常段和异常段的样本数; μ 为正常段样本残差均值; $alarm_{nor}$ 和 $alarm_{abn}$ 分别为正常段和异常段报警数。

其中, MNR 反映模型在正常状态下残差的平均值,用于衡量模型的稳定性与适应性; SNR 则度量残差的离散程度,较大的标准差表明模型结果存在显著的波动性; FDR 表示故障检测的准确性,是评估预警效果的关键;

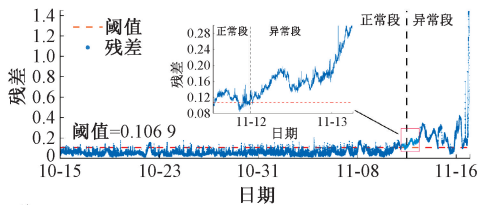
FAR 衡量错误报警频率,较低的误警率有助于提高预警系统的可靠性。

3.3 对比实验与结果讨论

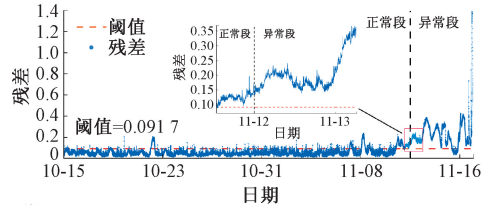
为验证提出方法的有效性,引入多种方法进行对比实验。根据训练方式的差异,将对比方法分为离线建模和持续蒸馏两大类。其中,离线建模指模型训练基于传统“离线建模,在线部署”方式,无教师模型指导;持续蒸馏指本文提出的训练方式,教师模型在线指导学生模型持续更新。

离线建模方式下,选用3种模型进行对比实验,分别为GRU、CNN-BiLSTM^[23]、GTCN模型。参数设置方面,GRU与GTCN的模型参数如3.2节所述;CNN-BiLSTM中,卷积核大小设置为 1×8 ,数量为24,BiLSTM隐藏层单元数设置为64,输出层使用双层全连接网络,单元数设置为16和8。持续蒸馏方式下,也选用3种模型进行对比实验。除提出的GTCN/GRU模型外,将提出方法中的教师模型和学生模型分别替换为CNN-BiLSTM和长短期记忆(long short-term memory, LSTM)网络模型。LSTM模型的参数设置与前文中的GRU模型一致。所有模型在训练过程中的batch_size、梯度下降优化器、学习率、epochs等参数均与提出方法一致。

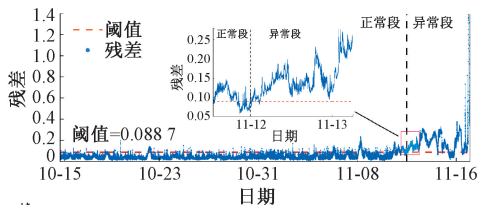
基于上述模型进行对比实验,磨煤机状态监测结果如图8所示。各子图中均有残差和故障预警两块区域,残差区域直观展示各样本的残差分布及趋势变化情况,故障预警区域通过1/0信号反映模型预警结果。根据专家故后诊断结果,以11月13日0时为界将测试集分为正常段和异常段。



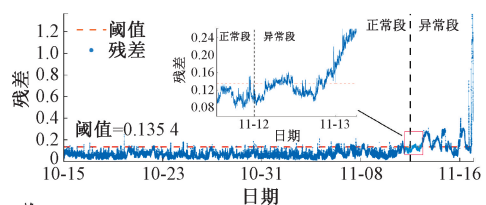
(a) 离线建模下GRU模型
(a) The GRU model under offline modeling



(b) 离线建模下CNN-BiLSTM模型
(b) The CNN-BiLSTM model under offline modeling



(c) 离线建模下GTCN模型
(c) The GTCN model under offline modeling



(d) 持续蒸馏下CNN-BiLSTM/GRU模型
(d) The CNN-BiLSTM / GRU model under continuous distillation

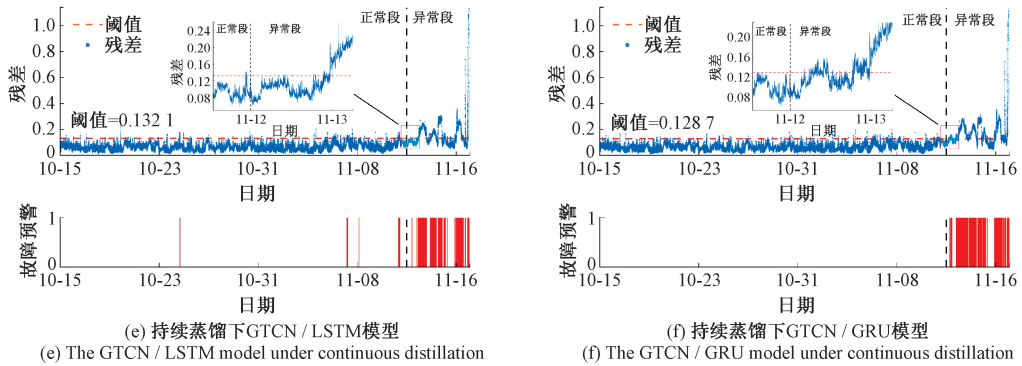


图 8 不同方式或模型下的磨煤机状态监测结果

Fig. 8 Condition monitoring results of the coal mill with different methods or models

离线建模方式下的状态监测结果如图 8(a)~(c) 所示。正常段初期残差平稳,无误报警出现,模型能够正常完成状态监测任务;正常段中期残差发生轻微波动,偶尔有误警出现;正常段后期逐渐偏离阈值,误警频繁,尤其在 11 月 11 日至 12 日,模型已经失去可靠性;异常段中残差几乎完全偏离阈值,持续发出故障预警信号。此外,模型 GRU、CNN-BiLSTM、GTCN 的预警阈值分别为 0.106 9、0.091 7、0.088 7,这表明模型 GTCN 拥有更强的学习能力,能够更精确地捕捉到数据中的复杂时空特征。

持续蒸馏方式下的状态监测结果如图 8(d)~(f) 所示。正常段初期残差平稳,无报警出现;正常段中后期残差轻微波动,偶尔有报警出现,误报警频率显著低于离线建模下的几种模型,在 11 月 11 日至 12 日对比效果更加显著;异常段残差波动剧烈,持续发出故障预警信号。通过在线更新学生模型参数,残差波动趋势得到有效修正,预警信号的触发也更加精准,提升了状态监测中模型的适应性与故障预警的准确性。此外,模型 CNN-BiLSTM/GRU、GTCN/LSTM、GTCN/GRU 的预警阈值分别为 0.135 4、0.132 1、0.128 7,整体高于离线建模下的几种模型。这是因为在教师模型的指导下,训练出的学生模型鲁棒性更强。值得注意的是,提出的 GTCN/GRU 模型在正常段无报警的情况下,在异常段 11 月 12 日 06:24 发出报警信号,相较磨煤机跳闸停机时刻提前了 116.75 h,能够为设备的维护和调整提供充足的响应时间。

此外,通过对比分析图 8(a)~(f) 中的局部放大图,可以看出:持续蒸馏方式下的状态监测残差在正常段内稳定性较高,几乎未出现越限现象,表明该方法能有效适应数据变化,维持监测结果的精度。相比之下,离线建模方式下的残差波动较大,出现了更多的越限情况,表明该方法在面对工况变化时的适应能力较弱,监测精度和稳定性较差。

从定性分析角度来看,持续蒸馏方式相比离线建模,在残差波动情况和误报警频率方面均有显著的提升。特

别是在正常段中期和后期,持续蒸馏方式表现出更高的稳定性与适应性,误报警频率明显低于离线建模方式。下面将从定量分析角度,更加具体地评估各模型的状态监测表现。

如表 3 所示,为不同训练方式下各模型的磨煤机状态监测定量分析结果。

表 3 磨煤机状态监测定量分析结果
Table 3 Quantitative analysis results for the condition monitoring of the coal mill

训练方式	教师模型/学生模型	MNR	SNR	FAR/%	FDR/%
离线建模	无/GRU	0.057 5	0.026 4	4.61	88.94
	无/CNN-BiLSTM	0.054 1	0.031 8	7.42	90.67
持续蒸馏	无/GTCN	0.046 3	0.024 6	4.44	85.63
	CNN-BiLSTM/GRU	0.065 2	0.026 9	1.39	60.64
	GTCN/LSTM	0.063 1	0.024 8	0.58	45.04
	GTCN/GRU	0.062 2	0.024 1	0	53.18

对于 MNR,离线建模方式下的指标数值显著低于持续蒸馏方式。这是因为,离线建模方式训练的模型通常具有较高的拟合度,能够较精确地反映训练集中的数据特征。而在持续蒸馏方式下,模型不仅通过监督损失进行训练,还引入了蒸馏损失,这种额外的损失项提高了模型的鲁棒性。模型在面对新数据时具备更好的适应能力,但其残差均值略高。在离线建模方式下,提出 GTCN 模型的 MNR 最低为 0.046 3,相较对比模型降低了 14.41%~19.47%,表明该模型的特征提取能力优于对比模型。对于 SNR,持续蒸馏方式下的模型相较离线建模方式平均降低了 8.45%。GTCN/GRU 模型最低为 0.024 1,相较对比模型降低了 2.82%~10.40%。通过在线自适应更新学生模型参数,提出模型能够更好地应对数据的动态变化,提高了状态监测中模型的稳定性、适应性和准确性。

关于指标 FAR 和 FDR,离线建模方式下的 FDR 和 FAR 数值均偏大,这是由于模型在监测过程中逐渐失配导致。具体来说,随着时间推移,磨煤机运行状态发生变化,模型未能及时更新,导致其对新数据的适应性变差,监测结果的可信度也迅速下降。在持续蒸馏方式中,这一现象得到改善。学生模型通过对实时数据的在线学习,不断调整参数,能够更好地适应设备状态的变化,因此 FAR 和 FDR 相对较低,监测结果更加准确。其中,提出的 GTCN/GRU 模型在无误报警(FAR 为 0)的情况下,FDR 为 53.18%。这表明提出方法能够使监测模型持续跟踪磨煤机运行状态的变化,提升设备故障预警结果的可靠性。

3.4 超参数实验与敏感性分析

超参数的合理设置对于模型的性能优化至关重要,通过实验与敏感性分析,可以深入理解不同超参数对模型性能的影响。因此,对提出方法中的两个关键超参数,蒸馏平衡系数 α 和自适应更新权重系数 β ,进行对比实验与分析讨论。实验基于持续蒸馏方式下的 GTCN/GRU 模型,参数设置情况与前文一致。

如图 9 所示,为超参数 α 的对比实验结果。作为蒸馏平衡系数,当 α 偏大时,学生模型在训练时的蒸馏损失起主导作用,倾向学习教师模型的映射规则;当 α 偏小时,学生模型在训练时的监督损失起主导作用,倾向直接拟合当前数据集。

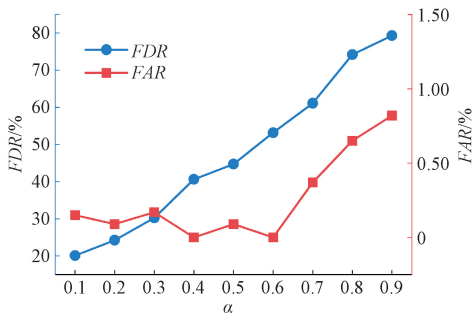


图 9 超参数 α 的对比实验结果

Fig. 9 The comparison experiment results of hyperparameter α

实验结果显示,随着 α 的增加,FDR 单调递增。这是由于监督损失比例降低导致模型精准度下降,残差整体偏高,逐渐偏离预警阈值。然而,指标 FAR 的情况则略有不同。当 α 为 0.1~0.3 时,FAR 整体偏低;当 α 为 0.4~0.6 时,FAR 趋近于 0;当 α 为 0.7~0.9 时,FAR 迅速升高。这是因为蒸馏损失偏小时,学生模型对数据特征学习不全面,产生少许误报警;而蒸馏损失偏大时,学生模型过度依赖教师模型的知识,失去了对实时数据特征的适应性,导致误警率迅速上升。综合考虑 FDR 和 FAR 的变化趋势,最终选择蒸馏平衡系数 α 为 0.6。在

无误报警的情况下确保 FDR 更高,平衡了模型的准确性与适应性。

如图 10 所示,为超参数 β 的对比实验结果。作为自适应更新权重系数,当 β 偏大时,旧学生模型的参数占主导作用;当 β 偏小时,在线训练的新学生模型的参数占主导作用。

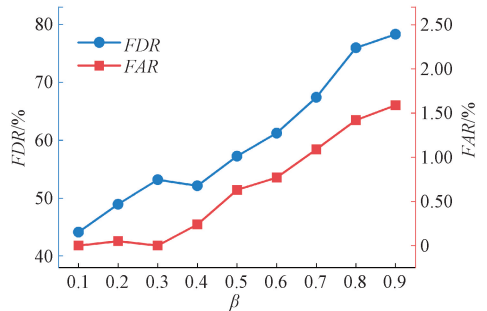


图 10 超参数 β 的对比实验结果

Fig. 10 The comparison experiment results of hyperparameter β

实验结果显示,随着 β 的增加,FDR 整体为单调递增趋势。这是由于当新学生模型占主导作用时,模型在训练过程中更加依赖实时数据,在对新环境适应性变强的同时,模型过拟合风险逐渐增加。此外,指标 FAR 的情况与 FDR 大致相同。当 β 为 0.1~0.3 时,FAR 趋近于 0;当 β 为 0.7~0.9 时,FAR 迅速升高,误警频繁。因此,在选择超参数 β 时需要综合考虑 FDR 和 FAR,避免过度拟合或过于保守的参数更新比例。最终,选择自适应更新权重系数 β 为 0.3,确保模型在适应性和稳定性之间达到平衡。

4 结 论

在数据分布偏移场景下,为提升磨煤机状态监测的准确性,提出了一种基于持续蒸馏的自适应方法。结合 GCN 和 TCN 的特征提取优势,提出 GTCN 模型,充分挖掘多变量时序数据中的时空特征。在线监测过程中,以提出的 GTCN 为教师模型,指导学生模型 GRU 基于复合损失函数训练,并通过新旧参数融合策略进行参数在线更新,在保证模型稳定性的前提下,快速适应新数据分布。以某煤电机组磨煤机运行数据为例,提出的持续蒸馏方式显著优于传统离线建模方式,正常段预测残差更加平稳,SNR 平均降低了 8.45%,在低 FAR 的前提下实现准确可靠的故障预警。离线建模方式下,提出 GTCN 模型的预警阈值和 MNR 均低于对比模型,表明其拥有更强的学习能力。持续蒸馏方式下,GTCN/GRU 模型的 SNR 最低为 0.024 1,相较于对比模型降低了 2.82%~10.40%,在无误报警的情况下提前 116 h 发出预警信号。

综上,提出方法为磨煤机在数据分布偏移下的自适应状态监测提供了可靠的技术支持,具有工程应用价值。

参考文献

- [1] 刘吉臻,王庆华,张效宁,等. 支撑我国能源转型的灵活燃煤发电新技术:燃煤耦合储能系统及智能控制系统[J]. 中国电机工程学报, 2024, 44(17): 6855-6883.
- LIU J ZH, WANG Q H, ZHANG X N, et al. Novel technologies of flexible coal-fired power generation to support China energy transition: Coal-fired coupled energy storage system and smart control system [J]. Proceedings of the CESS, 2024, 44(17): 6855-6883.
- [2] 常玉清,康孝云,王福利,等. 基于贝叶斯网络的磨煤机过程异常工况诊断模型实时更新方法[J]. 仪器仪表学报, 2021, 41(8): 52-61.
- CHANG Y Q, KANG X Y, WANG F L, et al. Real-time updating strategy for Bayesian network-based coal mill process abnormality diagnosis model [J]. Chinese Journal of Scientific Instrument, 2021, 41(8): 52-61.
- [3] HUANG C ZH, HE J X, ZHENG W, et al. A health monitoring method based on multivariate-time series adaptive gated recurrent unit transfer learning model for coal mill system [J]. Reliability Engineering & System Safety, 2025, 256: 110767.
- [4] 李晓彬,牛玉广,葛维春,等. 基于改进堆叠自编码网络的电站辅机故障预警[J]. 仪器仪表学报, 2019, 40(6): 39-47.
- LI X B, NIU Y G, GE W CH, et al. Early fault warning of power plant auxiliary engine based on improved stacked autoencoder network [J]. Chinese Journal of Scientific Instrument, 2019, 40(6): 39-47.
- [5] 杨婷婷,高乾,李浩千,等. 基于卷积神经网络-短时记忆神经网络的磨煤机故障预警[J]. 热力发电, 2022, 51(10): 122-129.
- YANG T T, GAO Q, LI H Q, et al. Coal mill fault early warning technology based on CNN-LSTM network [J]. Thermal Power Generation, 2022, 51(10): 122-129.
- [6] 罗毅,段明达. 基于IWOA-Transformer的磨煤机故障预警[J]. 动力工程学报, 2024, 44(6): 939-946.
- LUO Y, DUAN M D. Fault warning of coal mill based on IWOA-Transformer [J]. Journal of Chinese Society of Power Engineering, 2024, 44(6): 939-946.
- [7] 黄从智,何佳璇. 融合数据增强和深度迁移学习算法的磨煤机状态监测方法[J/OL]. 控制工程, 1-9 [2025-10-20].
- HUANG C ZH, HE J X. Condition monitoring method of coal mill combining data augmentation and deep transfer learning algorithm [J/OL]. Control Engineering of China, 1-9 [2025-10-20].
- [8] 罗云,李战国,付陇霞,等. 基于特征选择与BiLSTM多变量回归预测的磨煤机故障预警研究[J]. 动力工程学报, 2025, 45(5): 724-732.
- LUO Y, LI ZH G, FU L X, et al. Research on fault warning of coal mill based on feature selection and multivariate regression prediction with BiLSTM [J]. Journal of Chinese Society of Power Engineering, 2025, 45(5): 724-732.
- [9] 潘春阳,文书礼,朱森,等. 基于概念漂移监测与增量更新机制的超短期风电功率在线预测[J]. 中国电机工程学报, 2025, 45(6): 2133-2145.
- PAN CH Y, WEN SH L, ZHU M, et al. Online ultra-short-term wind power forecasting based on concept drift detection and incremental updating mechanism [J]. Proceedings of the CESS, 2025, 45(6): 2133-2145.
- [10] PENG X M, DUAN SH M, SANKAVARAM C, et al. Unsupervised adaptive fleet battery pack fault detection with concept drift under evolving environment [J]. IEEE Transactions on Automation Science and Engineering, 2024, 21(3): 2276-2288.
- [11] 石明宽,丁传仓,王锐,等. 面向变工况下工业流数据故障诊断的持续迁移学习系统[J]. 仪器仪表学报, 2024, 45(4): 10-16.
- SHI M K, DING CH C, WANG R, et al. Continuous transfer learning system for fault diagnosis of industrial stream data under variable operating conditions [J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 10-16.
- [12] 崔俊杰,秦文萍,朱志龙,等. 基于知识蒸馏的电力系统暂态稳定评估持续学习框架[J/OL]. 高电压技术, 1-15 [2025-10-20].
- CUI J J, QIN W P, ZHU ZH L, et al. A continuous learning framework for transient stability assessment of power systems based on knowledge distillation [J/OL]. High Voltage Engineering, 1-15 [2025-10-20].
- [13] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network [J]. ArXiv preprint arXiv: 1503.02531, 2015.
- [14] TIAN Y J, PEI SH CH, ZHANG X L, et al. Knowledge distillation on graphs: A survey [J]. ACM Computing Surveys, 2025, 57(8): 3711121.
- [15] 梁宵,陈莹. 基于知识蒸馏的空间通道双自编码器无监督异常检测[J]. 电子测量与仪器学报, 2025, 39(3): 146-158.
- LIANG X, CHEN Y. Knowledge distillation based spatial channel dual autoencoders for unsupervised anomaly detection [J]. Journal of Electronic Measurement and

- Instrumentation, 2025, 39(3): 146-158.
- [16] 谭会生, 杨威, 严舒琪. 交通速度预测时空图卷积网络及其 FPGA 实现研究[J]. 电子测量技术, 2024, 47(18): 108-119.
TAN H SH, YANG W, YAN SH Q. Research on spatio-temporal graph convolutional network for traffic speed prediction and their FPGA implementation[J]. Electronic Measurement Technology, 2024, 47(18): 108-119.
- [17] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [C]. Proceedings of the International Conference on Learning Representations, 2017: 114.
- [18] 高学金, 马东阳, 韩华云, 等. 基于 DAE 和 TCN 的复杂工业过程故障预测[J]. 仪器仪表学报, 2021, 42(6): 140-151.
GAO X J, MA D Y, HAN H Y, et al. Fault prediction of complex industrial process based on DAE and TCN[J]. Chinese Journal of Scientific Instrument, 2021, 42(6): 140-151.
- [19] 江灏, 叶铭新, 林蔚青, 等. 基于时空动态检测的核电厂堆外中子探测器故障检测方法[J]. 仪器仪表学报, 2024, 45(9): 131-144.
JIANG H, YE M X, LIN W Q, et al. Fault detection for ex-core neutron detectors in nuclear power plants using the spatial-temporal dynamic detection model [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 131-144.
- [20] 刘长良, 徐健, 王梓齐. 基于时空自编码网络的风电齿轮箱状态监测[J]. 信息与控制, 2023, 52(1): 84-92.
LIU CH L, XU J, WANG Z Q. Condition monitoring of wind turbine gearbox based on spatial-temporal autoencoder network[J]. Information and Control, 2023, 52(1): 84-92.
- [21] 常至轩, 徐鑫涛, 武二永, 等. 基于超声平面波成像的钢轨螺孔裂纹检测方法研究[J]. 仪器仪表学报, 2024, 45(6): 144-155.
CHANG ZH X, XU X T, WU ER Y, et al. Research on

the inspection method of railway bolt holes cracks using ultrasonic plane wave imaging [J]. Chinese Journal of Scientific Instrument, 2024, 45(6): 144-155.

- [22] 牛玉广, 徐健, 杜鸣, 等. 基于时空图注意网络的磨煤机故障预警[J]. 动力工程学报, 2025, 45(9): 1519-1526.
NIU Y G, XU J, DU M, et al. Fault warning for coal mill based on spatio-temporal graph attention network[J]. Journal of Chinese Society of Power Engineering, 2025, 45(9): 1519-1526.
- [23] MANUEL M, MERCEDES G M, MANUEL N. Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model[J]. Engineering Applications of Artificial Intelligence, 2023, 121: 106041.

作者简介



徐健, 2019 年于华北电力大学获得学士学位, 现为华北电力大学博士研究生, 主要研究方向为电站设备状态监测与故障预警, 以及人工智能在工程领域中的应用研究。

E-mail:ncepu_zdsxj@163.com

Xu Jian received his B. Sc. degree from North China Electric Power University in 2019. He is currently pursuing his Ph. D. degree at North China Electric Power University. His main research interests include power plant equipment condition monitoring and fault early warning, as well as the application of artificial intelligence in engineering.



杜鸣 (通信作者), 2021 年于华北电力大学获得博士学位, 现为华北电力大学讲师、硕士生导师, 主要研究方向为燃煤机组灵活智能发电技术、网源协同控制技术。

E-mail:zjajdming@163.com

Du Ming (Corresponding author) received his Ph. D. degree from North China Electric Power University in 2021. He is currently a lecturer and master's supervisor at North China Electric Power University. His main research interests include flexible and intelligent power generation technology for coal-fired units, as well as grid-source coordination control technology.