

基于自监督学习的辊压机轴承故障诊断方法

束庆宇

安徽理工大学 人工智能学院, 安徽 淮南 232000

摘要:目的 为了提高水泥生产效率,实现水泥辊压机轴承实时故障诊断,解决工程现场难以获得大量标签故障数据用于训练的问题。方法 提出一种使用连续小波卷积层的 Transformer 模型,通过使用小波核替换传统卷积核,作为特征提取网络第一层,提高网络对于机械故障特征的学习能力;再通过设计一种掩码训练的自监督学习训练,使网络从大量无标签数据中学习有用的相关特征,并通过设计一种对比损失迁移学习策略对模型进行训练,从相关公开数据中学习领域通用的判别特征;最后通过微调,实现水泥辊压机故障诊断网络的训练。结果 实验验证:本模型可以有效提高诊断进度,并且采用自监督学习与迁移学习训练策略后,在只有训练集 1%数据的情况下,模型微调仍可达到 78%的准确率;在只有 10%训练集数据的情况下,就可以接近监督训练使用 100%训练集数据下的性能。结论 本方法可以有效解决工程实际中无法采集大量标签故障数据的问题,提高模型诊断精度,在少量的目标数据下,保证模型的诊断精度。

关键词:轴承故障诊断;水泥辊压机;小样本;自监督学习

中图分类号:TQ172.6;TP18 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2025.0006.010

A Fault Diagnosis Method for Roller Press Bearings Based on Self-supervised Learning

SHU Qingyu

School of Artificial Intelligence, Anhui University of Science and Technology, Huainan 232000, Anhui, China

Abstract: **Objective** This study aims to improve the efficiency of cement production, realize real-time fault diagnosis of cement roller press bearings, and solve the problem that it is difficult to obtain a large amount of labeled fault data for training in the engineering field. **Methods** A Transformer model using continuous wavelet convolution layers is proposed. The model enhances the network's learning ability for mechanical fault features by using wavelet kernels instead of traditional convolutional kernels in the first layer of the feature extraction network. It then utilizes mask-training self-supervised learning to enable the network to learn useful relevant features from a large amount of unlabeled data. Additionally, the model is trained using a designed contrastive loss transfer learning strategy to learn domain-general discriminative features from relevant public data. Finally, through fine-tuning, the model completes training of the cement roller press fault diagnosis network. **Results** Experimental validation shows that this model effectively accelerates diagnostic progress. With the self-supervised learning and transfer learning training strategies, the model achieves an accuracy of 78% in fine-tuning with only 1% of the training set data, approaching the performance of supervised training using 100% of the data with only 10% of the training set data. **Conclusion** This method effectively addresses the challenge of acquiring large amounts of labeled fault data in engineering practice, improving diagnostic accuracy and maintaining diagnostic precision with limited target data.

Keywords: bearing fault diagnosis; cement roller press; small sample; self-supervised learning

收稿日期:2023-11-13 修回日期:2024-01-10 文章编号:1672-058X(2025)06-0072-06

基金项目:安徽省自然科学基金(2108085UD07)。

作者简介:束庆宇(2000—),男,安徽省六安市舒城县人,硕士研究生,从事时序预测、故障诊断研究。

引用格式:束庆宇.基于自监督学习的辊压机轴承故障诊断方法[J].重庆工商大学学报(自然科学版),2025,42(6):72-77.

SHU Qingyu. A fault diagnosis method for roller press bearings based on self-supervised learning[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(6): 72-77.

投稿地址 <http://journal.ctbu.edu.cn/zr/ch/index.aspx>

1 引言

水泥辊压机作为水泥生产线上的重要设备,其运行状态与水泥生产效率紧密相关^[1-2]。轴承作为水泥辊压机中易发生故障,且无法直接监测状态的零部件,获取辊压机轴承的实时故障状态能有效避免检修带来的过修、欠修以及耽误生产等问题,大大提高水泥生产效率,降低检修成本。本文所用任务数据采集于水泥生产线,生产线轴承振动监测布置以后,存在机器不能带故障工作,采集大量数据容易,但确定数据标签难,依旧依赖于定期巡检的问题^[3]。所以大量采集数据中,有标签的只包含较小部分。

在传统故障诊断方法中,通常依赖于故障特征的人为选取^[4-5]。Hu等^[6]提出一种采用经验模式分解(Empirical Mode Decomposition, EMD)方法,从原始信号中获得固有模式分量,并将其转换为模糊特征向量,从而检测出机械故障。其缺点是算法的性能极大依赖于选取特征的优劣,即使使用人为选取的特征来训练模型,最终分类性能仍有待提高。为了克服这一问题,具有强大特征提取能力的深度学习方法被学者们运用于故障诊断领域,有效地学习到了隐藏在诊断数据中的基本特征,提高了诊断精度^[7-9]。为了解决深度神经网络提取特征的可解释性问题,以及在训练样本较少时网络难以学习到与故障相关的影响分量问题,有学者提出将连续小波卷积层作为深度网络的第一层,以帮助神经网络发现有一定物理意义的卷积核^[10-11]。Li^[10]等提出一种新的小波驱动深度神经网络,称为Wavelet Kernel Net(WKN),其中设计了连续小波卷积(CWConv)层来代替标准CNN的第一卷积层。这使得第一个CWConv层能够发现更有意义的内核。经实验验证:连续小波卷积层在公开的轴承振动数据集上取得了很好的效果,有效提升了诊断精度。但其有监督的训练方法,使得在工程实际应用场景下,使用连续小波卷积层并不能解决标签数据缺乏带来的问题,诊断精度无法保证。

自监督学习和迁移学习一直是深度学习领域解决目标数据缺乏的重要研究方向。近年来,自监督学习在解决大规模语言模型训练中,可获得大量文本数据,它在难以针对目标任务大量制作标签的问题中取得了显著成果,因此,也成为当前的研究热点^[12-15]。自监督学习通过设计借口任务来利用大规模无标签数据,从中学习到对目标任务相关的特征。而迁移学习是一种领域相关性,它将源域学习到的模型和知识用于目标域,主要思路为通过收集相关领域数据并通过策略使

网络学习领域通用的特征,以解决目标域数据不足,无法学习到可靠特征的问题。

因此,为了充分利用大量无标签数据,提升在训练标签有限下的诊断精度,本文提出一种使用自监督学习训练结合迁移学习策略的模型。首先,将原始诊断数据输入连续小波卷积层,得到能够反映原始数据局部状态的一组向量编码;再考虑故障状态的反映往往体现在故障冲击局部与整体的关系。Transformer的自注意力机制能很好地描述各局部状态向量与其他局部状态向量的相关性,但其解码器部分考虑到时序性,读取信息时对未来数据进行遮挡,所以本文采用Transformer编码器部分作为特征提取网络,最终经线性层完成故障分类。在自监督训练时,采用掩码预测的借口任务,通过遮盖部分输入局部状态向量,再用对应位置输出向量来预测对应位置的原始振动数据,使得网络学习到当前位置状态与其他位置之间的关系。之后本文设计了一种对比损失,使源域数据不同来源但同标签的样本通过特征提取网络输出的特征向量相似度尽可能高,不同标签样本的相似度尽可能低,来使特征提取网络学到领域通用的特征。最后在目标域数据,即有标签的少量辊压机轴承振动数据上进行微调训练,实现诊断模型。

本文提出一种结合连续小波卷积层的Transformer模型,结合自监督学习训练与迁移学习策略对模型进行训练,从无标签样本以及相关公开数据中学习领域通用的判别特征。本方法可以有效解决工程实际中无法采集大量标签故障数据的问题,提高模型诊断精度,在少量的目标数据下,保证模型的诊断精度。对于存在类似问题的工程故障领域实时诊断问题的解决具有一定的参考意义。

2 基于对比学习的故障诊断迁移学习方法

本节详细介绍基于自监督学习的故障诊断迁移学习模型的组成部分。对于故障数据,通过自适应小波卷积层,得到每个时刻的状态向量输入Transformer编码器部分进行特征提取;再通过随机掩码遮盖Transformer部分位置输入,预测遮盖部分对应原始故障数据,使编码器学习各个时刻状态之间的关联性;再通过对比学习方法,使得不同域下同标签数据的编码器输出向量相似性最大化,不同标签数据的编码器输出向量相似性最小化,以学习领域不变的特征;最后通过在少量目标域数据上微调,实现故障诊断模型。

2.1 连续小波卷积层

在诊断数据为一种时频域变换方法时,小波变换

既能够反应信号的频率变化信息,又能反映频率的时域变化信息。为了更高效捕捉不同频率上的故障特征,本文构建自适应小波层替换卷积神经网络的第一层,将一组 Laplace、Morlet、Mexh 小波基作为卷积核,与输入信号进行卷积操作。将小波基的尺度因子和平移因子作为自适应小波层需要训练优化的参数。

构建自适应小波层卷积核,其表达式如式(1)~式(3)。

Mexh 小波卷积核:

$$k = \left(1 - \left(\frac{t-u}{s}\right)^2\right) * e^{-\frac{\left(\frac{t-u}{s}\right)^2}{2}} \quad (1)$$

Laplace 小波卷积核:

$$k = 0.08 * e^{\frac{-0.03}{\sqrt{1-0.03^2}}} * \left(2\pi f * \left(\left(\frac{t-u}{s}\right) - 0.1\right)\right) \times \left(-\sin\left(2\pi f * \left(\left(\frac{t-u}{s}\right) - 0.1\right)\right)\right) \quad (2)$$

Morlet 小波卷积核:

$$k = \pi^{0.25} * e^{-\frac{\left(\frac{t-u}{s}\right)^2}{2}} * \cos\left(2\pi \frac{t-u}{s}\right) \quad (3)$$

式(1)、式(2)和式(3)中, s 为小波基的尺寸因子, u 为平移因子。

将 s, u 作为可学习的参数,使卷积核保持小波基的形状,经过训练得到自适应的小波基参数 s, u 。

得到诊断数据每个时刻不同小波基、不同尺度小波运算的输出作为当前时刻的状态向量,反映了当前时刻在不同频率对于不同变换的状态。连续小波卷积层原理如图 1 所示。

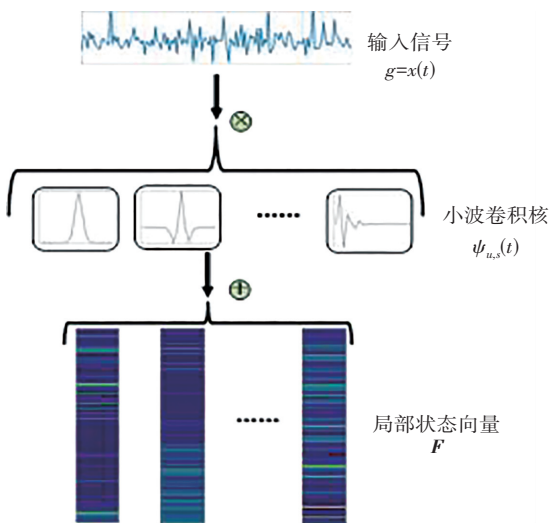


图 1 连续小波卷积层原理图

Fig. 1 Schematic diagram of continuous wavelet convolution layer

图 1 中卷积核的总体表达形式为

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \quad (4)$$

式(4)中,卷积核与输入信号 $g=x(t)$ 做卷积运算,即内积,得到局部状态矩阵为

$$F = \langle g, \psi_{u,s}(t) \rangle = \frac{1}{\sqrt{s}} \int x(t) \psi\left(\frac{t-u}{s}\right) \quad (5)$$

式(5)中, u, s 为连续小波卷积层可迭代参数。

本文引入自适应小波卷积层,通过内积提取机械信号的特征,在保留小波波形的约束性上自适应地得到合适的小波,可以极大程度提高故障特征提取的准确性和效率。对于自适应小波层不同种类,参数小波所提取的特征矩阵使用注意力机制进行筛选,针对不同数据不同故障使特征提取侧重于相对应的小波,提高诊断模型的泛化性和诊断性能。

2.2 自监督训练

自监督学习的目的是通过借口任务实现预训练,充分利用大量无标签数据,使模型在借口任务训练中能够学到一些通用特征,从而通过微调预训练模型来减少下游任务所需数据量与训练时间。

为了充分利用现场采集的大量无标签数据,提升模型性能,本文通过掩码预测的自监督训练任务,学习各时刻状态的关联性。随机对部分位置的 Transformer 编码器输入向量进行遮盖,再将相应位置的输出经全连接网络来预测对应位置的原始振动数据。将现场采集的全部数据,以及所收集的公开数据进行自监督学习,得到预训练模型。模型结构及自监督训练策略如图 2 所示。

图 2 中, x 为原始振动数据,经连续小波卷积层及位置编码后进行随机遮掩操作,mask 为遮掩位置。经 Transformer 编码层提取特征后,通过遮掩位置所对应的输出向量对遮掩位置的原始振动数据进行预测。通过对局部特征向量进行遮盖并预测遮盖位置原始数据的方法,充分利用工程中所采集的大量无标签数据,从中学习目标任务相关特征。

通过自监督学习的方法可以充分训练大量现场无标签数据。通过这种掩码预测的方式,使模型在训练中可以学习到有关目标域数据,即水泥辊压机振动采集数据的相关特征,有助于提升后面进行水泥辊压机轴承故障诊断上的正确率。

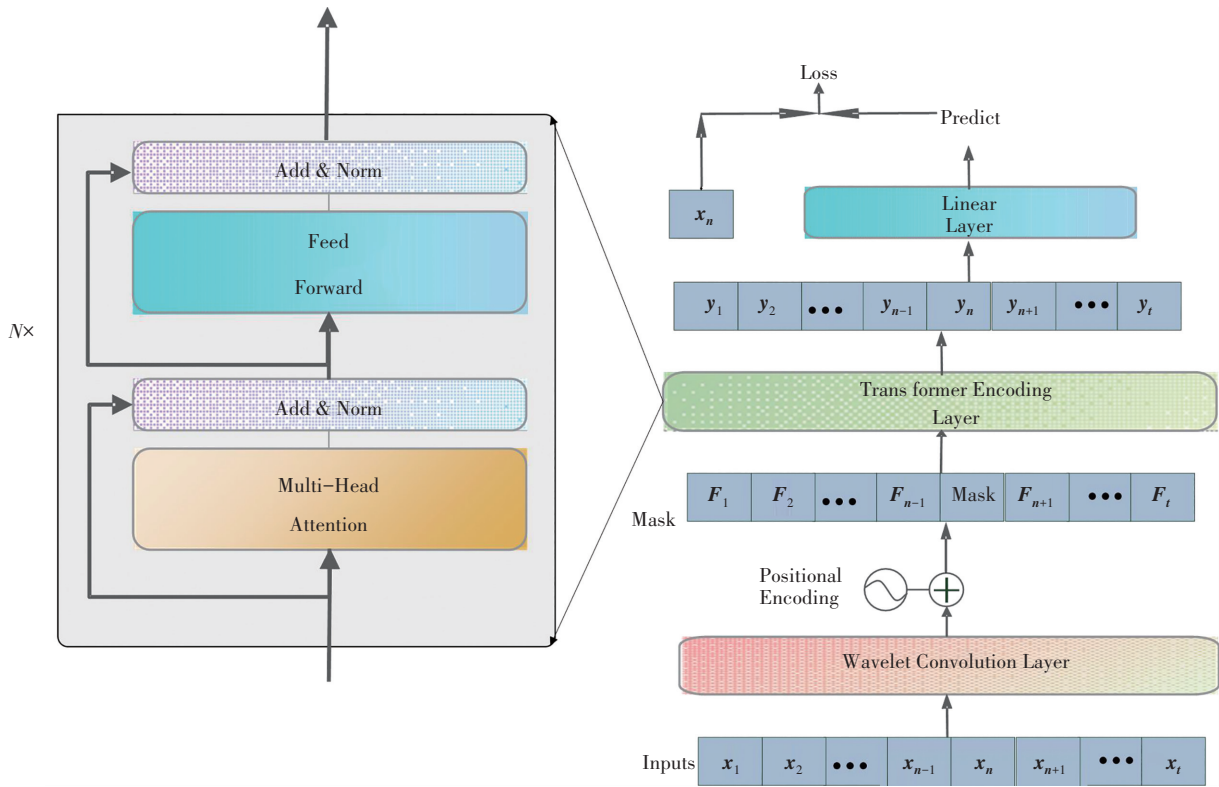


图 2 网络结构图及自监督学习策略

Fig. 2 Network structure diagram and self-supervised learning strategy

2.3 迁移学习

迁移学习是一种领域相关性,将源域学习到的模型和知识用于目标域的一种方法,主要为了解决目标域数据不足,无标签的情况。

为了解决水泥辊压机标签数据量不足的问题,本文收集大量公共数据作为源域数据,辊压机轴承振动数据作为目标域数据。利用数据领域的相通,通过迁移学习方法使模型学到领域通用的特征,再利用有标签的目标域数据对网络进行微调,实现辊压机轴承故障诊断模型。

为了使模型从不同来源的数据中学到领域通用的特征,本文通过点积衡量样本特征向量的相似度。将不同数据集的源域数据以及目标域样本送入预训练模型进行特征提取,设计对比损失,使各个数据集同标签样本的特征向量点积最大化,而最小化同一批次内与其他样本的点积。据此,本文设计损失如下:

$$L = -\frac{1}{K} \sum_{k=1}^K \log \frac{\sum_{n \in N_k} \exp((y_k)^T y_n)}{\sum_{n \in M_k} \exp((y_k)^T y_n)} \quad (6)$$

式(6)中, K 为一个训练批次内样本数量, y 为样本数据经 Transformer 编码器提取的特征向量, N_k 为和 y_k 同一

批次相同标签样本的特征向量集合, M_k 为同一批次内所有剩余样本特征向量的集合。

本文通过设计一种对比损失的方法在大量源域数据上最大化不同领域间相同标签数据特征向量的相似度,最小化不同标签数据特征向量的相似度,使特征提取网络学到领域通用的特征,可适用于目标域数据;再将迁移学习后的网络在水泥辊压机轴承振动采集标签数据上进行训练微调,即可通过少量标签数据完成水泥辊压机轴承故障诊断模型的训练。

3 仿真实验与结果分析

3.1 数据集

目标数据集来源于水泥生产线辊压机轴承座所采集的振动数据,由于数据标签的测定影响生产线效率,所以采集数据包含大量无标签数据和部分有标签数据。

3.2 实验细节

本文使用 PyTorch 建立模型,并在 NVIDIA RTX 3060 GPU 上进行训练。首先使用所有源域数据与大量无标签目标域数据进行自监督训练,防止测试结果存在数据泄露问题;再用目标域数据,通过对比损失对预训练模型进行训练;最后将有标签的源域数据分成

60%、20%、20%进行训练微调、验证和测试。目标任务训练将在验证精度连续不再下降时结束,取验证进度最高的模型进行测试。实验用 5 个不同的种子重复了 5 次,本文将取实验数据的平均数和标准差来描述数据结果。

3.3 实验结果分析

为了验证本文提出模型的有效性以及自监督学习预训练模型的效果,将 CNN 模型、Transformer 模型、LSTM 模型和本文模型进行比较。遵循线性评估方案,对比模型与本文设计模型采用同样的自监督预训练方式与迁移学习策略,最后在冻结模型的基础上训练针对目标任务的线性分类器。表 1 为本文方法与对比方法的线性评估结果,表中 ST-Transformer 为本文所提出模型。结果显示:在同样标签数据的训练与测试情况下,本文所提出的模型取得了最佳整体性能。这表明本文通过小波卷积层将原始振动数据转换为局部状态向量,再通过 Transformer 模型编码器学习各局部向量与其他向量的整体关系,模型结构可以更有效地分辨出故障状态。

表 1 模型对比实验结果

Table 1 Results of model comparison experiments

模 型	诊断精度
LSTM	82.63±3.5
CNN	93.30±1.3
Transformer	91.41±2.3
ST-Transformer	97.83±1.7

为了验证本文提出模型每个改进方法的有效性,进行了消融实验,对不同的模型变体进行比较。具体如下:首先,振动信号通过常规卷积提取特征送入 Transformer 编码器,对没有自监督预训练以及迁移学习下的模型进行训练,这个变体称为“CNN+Transformer”;第二,将“CNN+Transformer”常规卷积部分的卷积核替换为小波核,可训练参数为小波的尺度因子与平移因子,这个变体称为“WCNN+Transformer”;第三,对“WCNN+Transformer”只进行自监督学习预训练,再在目标数据上微调,这个变体称为“SL-Transformer”;第四,对“WCNN+Transformer”只进行迁移学习训练,再在目标数据上微调训练,这个变体称为“TL-Transformer”;最终,本文对“WCNN+Transformer”先进进行子监督学习预训练,再进行迁移学习训练,最后在目

标数据上微调这个变体称为“ST-Transformer”。表 2 显示了在水泥辊压机轴承数据上,验证消融实验的结果。很明显,使用小波卷积比传统卷积提取局部状态的效果更好,使用自监督学习与迁移学习组合能产生最好的性能。

表 2 消融实验结果

Table 2 Results of ablation experiment

模 型	诊断精度
CNN+Transformer	78.70±2.9
WCNN+Transformer	81.53±0.9
SL-Transformer	90.24±1.1
TL-Transformer	94.13±0.5
ST-Transformer	97.83±1.7

为了验证本文在有限标签下对训练效果的提升,通过实验使用 1%、5%、10%、50%、75% 和全部的训练集数据样本对本文模型进行微调,来研究 ST-Transformer 模型在不同标签量下的性能。图 3 显示了上述设置下对目标数据进行微调的 ST-Transformer 和监督训练的结果。可以看出:在有限标签样本下,监督训练的表现很差,而通过预训练并迁移学习后模型微调的结果要明显优于监督训练;在只有训练集 1% 数据的情况下,ST-Transformer 微调仍可达到 78% 的准确率;在只有 10% 训练集数据的情况下,就可以接近监督训练使用 100% 训练集数据下的性能,这证明 ST-Transformer 模型在小数据样本性能提升方面的有效性。

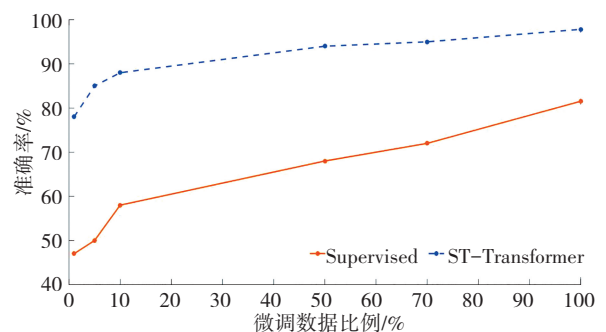


图 3 小样本微调对比实验结果

Fig. 3 Results of small sample fine-tuning comparison experiment

4 结 论

提出一种适用于缺乏数据标签的故障诊断算法,通过小波卷积将原始每一时刻的数据转换为描述局部状态的特征向量;再在自监督学习的预训练中通过大

量无标签数据以及领域相关数据,使模型学到各局部状态向量之间的内在联系,得到一个较好的预训练模型;后续通过对比损失的迁移学习策略,使模型在大量源域数据中,提取出领域通用的特征。在此基础上,针对目标任务训练微调一个线性分类器。在只使用训练集10%数据的情况下,模型性能便可以接近使用全部标签数据有监督训练的模型。此外,使用自监督预训练方案,使模型最终性能的提升得益于所采集的大量无标签数据。通过自监督与迁移学习结合的方法,有效解决了轴承故障诊断工程应用中数据标签采集困难,标签数据量少的问题。

标签数据收集困难存在于大部分大型工业旋转设备中,其成本高、负荷大、工况多变,使得无法通过实验室数据采集获得相应的训练数据。在生产线上布置传感器采集数据,运行过程中所获得的大量数据均无法获得标签。文中所提出的自监督学习与迁移学习方案,可以很好地适用于相关工业场景。

参考文献(References):

- [1] 周智勇. 辊压机轴承损坏原因分析及预防措施[J]. 水泥, 2023(4): 55-58.
ZHOU Zhi-yong. Cause analysis and preventive measures of roller bearing damage in roller press[J]. Cement, 2023(4): 55-58.
- [2] 代斌. 水泥厂辊压机故障处理分析[J]. 机械工程与自动化, 2021(5): 206-207.
DAI Bin. Analysis of failure treatment of roller press in cement plant [J]. Mechanical Engineering & Automation, 2021(5): 206-207.
- [3] 李巍华, 何琛, 陈祝云, 等. 基于对称式对比学习的齿轮箱无监督故障诊断方法[J]. 仪器仪表学报, 2022, 43(3): 121-131.
LI Wei-hua, HE Chen, CHEN Zhu-yun, et al. Unsupervised fault diagnosis of gearbox based on symmetrical contrast learning[J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 121-131.
- [4] SI Y, WANG Y, ZHOU D. Key-performance-indicator-related process monitoring based on improved kernel partial least squares[J]. IEEE Transactions on Industrial Electronics, 2021, 68(3): 2626-2636.
- [5] MA X, SI Y, YUAN Z, et al. Multistep dynamic slow feature analysis for industrial process monitoring[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9535-9548.
- [6] HU Q, HE Z J, ZI Y, et al. Intelligent fault diagnosis in power plant using empirical mode decomposition, fuzzy feature extraction and support vector machines[J]. Key Engineering Materials, 2005, 293-294: 373-382.
- [7] XU Y, DU B, ZHANG L. Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: attacks and defenses [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2): 1604-1617.
- [8] DU B, RU L, WU C, et al. Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 9976-9992.
- [9] LIU Y, DU B, TU W, et al. Log Det metric-based domain adaptation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(11): 4673-4687.
- [10] LI T, ZHAO Z, SUN C, et al. Wavelet Kernel Net: An interpretable deep neural network for industrial intelligent diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(4): 2302-2312.
- [11] ZHAO M, KANG M, TANG B, et al. Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis[J]. IEEE Transactions on Industrial Electronics, 2019, 66(6): 4696-4706.
- [12] QI G J, LUO J. Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2168-2187.
- [13] JING L, TIAN Y. Self-supervised visual feature learning with deep neural networks: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(11): 4037-4058.
- [14] JAISWAL A, BABU A R, ZADEH M Z, et al. A survey on contrastive self-supervised learning[J]. Technologies, 2020, 9(1): 2-10.
- [15] 王颖, 朱南阳, 谢浩川, 等. 基于对比学习辅助训练的超短期风功率预测方法[J]. 仪器仪表学报, 2023, 44(3): 89-97.
WANG Ying, ZHU Nan-yang, XIE Hao-chuan, et al. Ultra-short-term wind power forecasting based on contrastive learning-assisted training[J]. Chinese Journal of Scientific Instrument, 2023, 44(3): 89-97.

责任编辑:李翠薇