

DOI:10.3969/j.issn.1001-4551.2024.05.012

基于 SimAM 注意力机制的轴承故障迁移诊断模型*

包从望^{1,2}, 朱广勇¹, 邹旺¹, 郭灏¹

(1. 六盘水师范学院 矿业与机械工程学院, 贵州 六盘水 553000; 2. 中国矿业大学 机电工程学院, 江苏 徐州 221116)

摘要: 针对轴承故障在跨工况迁移诊断时, 其域不变特征难以提取, 易出现模型过拟合这一问题, 提出了一种基于无参数注意力模块 (SimAM) 的轴承故障迁移诊断方法。首先, 以一维卷积神经网络作为基本框架, 利用自适应批量归一化 (AdaBN) 对各输出层进行了归一化处理, 经两层卷积层和两层池化层后, 对输出特征进行了随机节点失活操作; 然后, 利用改进后的参数化修正线性单元 (PReLU) 激活函数自适应提取负值输入权重系数, 分别以交叉熵损失函数监督训练有标签的源域数据, 以均方对数误差 (MSLE) 作为损失函数训练无标签的目标数据; 最后, 利用自制实验台数据和凯斯西储轴承公开数据对模型进行了验证, 分别以不同的单一工况作为源域, 其余工况作为目标域进行了迁移诊断任务研究。研究结果表明: 基于 SimAM 的轴承故障迁移诊断方具有较好的域不变特征提取的性能, 且所提特征具有较好的聚类效果; 自制实验台中的平均迁移精度在 89.1% 以上, 最高均值可达 97.85%, CWRU 数据集中的平均迁移精度达 98.68%。该成果可为后续轴承故障由实验向工业现场的迁移诊断奠定基础。

关键词: 轴承故障诊断; 迁移学习; 无参数注意力机制; 自适应批量归一化; 参数化修正线性单元; 均方对数误差; 卷积神经网络

中图分类号: TH133.3; TP277

文献标识码: A

文章编号: 1001-4551(2024)05-0862-09

Rearing fault transfer diagnosis model based on SimAM attention mechanism

BAO Congwang^{1,2}, ZHU Guangyong¹, ZOU Wang¹, GUO Hao¹

(1. School of Mining and Mechanical Engineering, Liupanshui Normal University, Liupanshui 553000, China;

2. School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Aiming at the problem that domain invariant features are difficult to extract and model overfitting is easy to occur in bearing fault migration diagnosis during cross-working condition migration diagnosis, a bearing fault migration diagnosis method based on a simple parameter-free attention module (SimAM) was proposed. Firstly, the one-dimensional convolutional neural network was used as the basic framework, and adaptive batch normalization (AdaBN) was used to normalize each output layer. After two convolutional layers and two pooling layers, the output features were deactivated by random nodes. Then, the improved parametric rectified linear unit (PReLU) activation function was used to adaptively extract the negative input weight coefficient, and the cross-entropy loss function was used to monitor the trained labeled source domain data and the mean squared logarithmic error (MSLE) was used as the loss function to train the unlabeled target data. Finally, the model was verified by the self-made experimental bench data and the open data of Case Western Reserve bearing. Different single working conditions were taken as the source domain, and the other working conditions were taken as the target domain to carry out the migration diagnosis task. The experimental results show that the proposed method has good domain invariant feature extraction performance, and the proposed features have good clustering effect. The average migration accuracy of the self-made experimental bench was above 89.1%, the highest mean was up to 97.85%, and the average migration accuracy of CWRU dataset was up to 98.68%. The results can lay a foundation for subsequent bearing fault transfer diagnosis from experimental data to industrial sites.

Key words: bearing fault diagnosis; transfer learning; simple parameter-free attention module (SimAM); adaptive batch normalization (AdaBN); parametric rectified linear unit (PReLU); mean squared logarithmic error (MSLE); convolutional neural network

收稿日期: 2023-11-24

基金项目: 贵州省教育厅基金资助项目 (黔教合 KY 字 [2020] 117); 六盘水市科技计划项目 (52020-2022-PT-02, 52020-2019-05-12); 六盘水师范学院基金资助项目 (LPSSYylzy2205)

作者简介: 包从望 (1989-), 男, 贵州六盘水人, 硕士研究生, 副教授, 主要从事机械装备故障诊断与设备运维方面的研究。E-mail: bcongw@163.com

0 引言

滚动轴承作为重要的工业零部件,被广泛用于机械领域的齿轮箱、减速器、发电机、风机等旋转机械中^[1]。

因传动部件的工况复杂多变,常常使滚动轴承产生轴承内圈、轴承外圈、滚动体等故障,严重的还会进一步造成机器停机,给企业带来经济损失,甚至还可能造成人员伤亡。因此,研究针对滚动轴承故障的智能诊断方法具有经济效益和安全生产方面的双重意义^[2-3]。

常用的轴承故障诊断方法包括解析法、信号处理和数据驱动三种方式。三种方法中的共同点在于提取故障的特征值作为模式识别的基础。其中,解析法对轴承运行状态建模的要求较高,边界条件选择不当极有可能导致故障诊断失效;信号处理方法可实现对轴承运行状态的实时监测,但选择特征值困难,且该方法的泛化性能较差。

随着人工智能技术的发展,基于数据驱动的智能故障诊断算法可根据采集信号自动提取轴承的故障特征,并实现故障模式的识别与分类。该技术在复杂机械系统的滚动轴承故障监测方面具有明显优势,是目前轴承故障诊断的研究热点^[4]。轴承的振动信号蕴含了丰富的轴承故障信息,可实现轴承的端到端故障诊断,常被用作轴承故障运行状态评估的监测信号^[5]。根据特征提取方式的不同,可将数据驱动方式的故障诊断方法分为基于人工特征提取的机器学习方法和无人工特征提取的深度学习^[6]。当轴承运行工况为恒定工况时,基于深度学习的故障诊断模型具有良好的诊断性能。

然而,由于转速和载荷等因素的影响,机械设备大多处于变工况状态,同一个故障类型下采集到的振动信号幅值和特征频率分布不一致;此外,不同故障类型下采集到的振动信号分布可能接近,这给故障诊断带来了严峻的挑战。为实现各个工况下的轴承故障诊断,最直接的方法是采集所有工况下的振动数据,对模型进行训练。但这显然不具备可操作性^[7]。

因此,将某一特定工况下的故障诊断模型推广至其余工况,即研究变工况下的轴承故障迁移诊断方法,有利于提高诊断模型的精度和稳定性^[8]。

变工况下的轴承故障诊断方法可根据不同工况下的特征分布对齐,利用已标记的工况数据训练诊断模型,并将其用于未知工况领域的故障诊断。在迁移学习模型中,源域数据采用不同工况下的共享特征,或者

将不同工况下的分布映射到新的特征空间,以减小不同域之间的分布差异,从而实现目标域的故障诊断^[9]。

目前,已有大量关于轴承故障诊断的迁移学习算法研究。ZHANG Wei 等人^[10]建立了首层宽核深度卷积神经网络(deep convolutional neural networks with wide first-layer kernels, WDCNN)方法,实现了不同领域振动信号的对齐分布。GUO Liang 等人^[11]基于卷积神经网络,以最大均值差异作为模型优化准则,有效降低了源域数据和目标域间的边缘分布。ZHAO Xiao-li 等人^[12]设计了一种类内和类间卷积神经网络,从而改善了不同域间的分布差异,实现了从源域工况向目标域工况的迁移诊断。

迁移学习常用的激活函数包含 Sigmoid、Tanh 和 ReLU 三种,当输入值的绝对值较大时, Sigmoid、Tanh 两种激活函数的输出值均趋于平缓,这将导致网络梯度消失,使模型无法训练更新参数^[13]。相比之下,当 ReLU 函数输入大于 0 时,其梯度为常数,可有效地减轻梯度消失问题。然而,当输入小于 0 时,输出均为 0,因此负值特征难以保留。为此,研究者提出了 Leaky ReLU,当输入小于 0 时,将赋予一个非零斜率,但实践证明该斜率系数难以确定;为确定该斜率系数, PReLU 可根据神经网络训练自适应获得该系数。然而引入神经网络无疑给整个诊断模型带来训练负担,且易出现过拟合现象^[14-16]。

为解决滚动轴承故障诊断在跨工况域中存在的问题,笔者提出一种基于 SimAM 注意力机制的轴承故障诊断方法,根据 SimAM 算法的引入,自适应调整激活函数 PReLU 中的斜率系数。

首先,将振动信号输入一维卷积神经网络中,并采用 AdaBN 方法对每层卷积输出做归一化处理;然后,以改进后的 PReLU 作为激活函数,以全面提取故障特征;并采用交叉熵作为损失函数,对源域数据进行监督训练,同时以 MSLE 作为目标域的损失函数,进行无监督训练;最终,在变工况条件下,利用实验数据集验证所提故障诊断方法的有效性。

1 问题描述及理论基础

1.1 问题描述

滚动轴承的振动信号中包含了丰富的故障信息。在对轴承进行故障诊断时,其信号极易受设备载荷、转速等运行工况的影响,导致同类故障的特征分布存在较大差异,甚至可能出现不同故障类别特征分布接

近的问题,导致轴承故障诊断失效^[17]。

为保证故障诊断模型的有效性,可以将有限标记样本作为源域,将实际工况运行条件下的无标签样本作为目标域,分别用数据集 R^s 、 R^t 表示;用迁移学习的思想,根据提取 R^s 、 R^t 中的公共特征,以减小两者的分布差异为目标,从而提高故障诊断模型泛化能力^[18]。

迁移模型如图 1 所示。

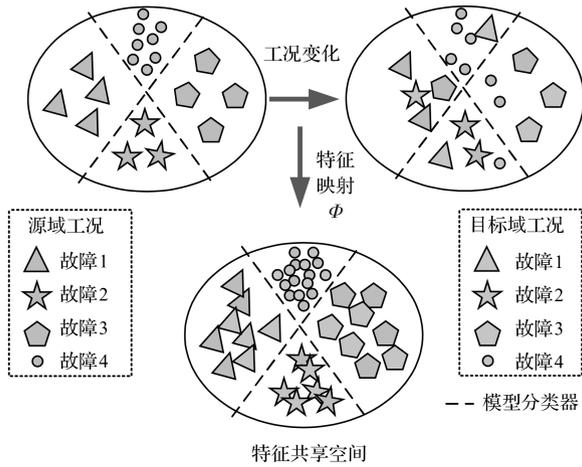


图 1 轴承故障迁移诊断简图

Fig. 1 Bearing fault migration diagnosis diagram

若以 X^s 、 X^t 分别表示源域和目标域的特征空间,对应边缘分布为 $L(X^s)$ 、 $L(X^t)$,记 $R^s = \{X^s, L(X^s)\}$ 、 $R^t = \{X^t, L(X^t)\}$;受工况的影响, $L(X^s) \neq L(X^t)$,因此无法将源域中训练的诊断模型直接用于目标域。笔者需构建特征映射函数 Φ ,将 X^s 、 X^t 映射到公共空间,使 $L(\Phi(X^s)) \approx L(\Phi(X^t))$ 。

因此,跨工况下的轴承故障诊断问题转变为求两域间的独立同分布问题。

1.2 改进的 PReLU 激活函数

ReLU 具有分段线性函数的特点,且当输入值无穷大时,其梯度值不会消失。因此,ReLU 被广泛用作深度学习的激活函数。

ReLU 计算表达式如下:

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

式中: x 为特征值。

显然,当输入特征值小于 0 时直接置 0。为改善这一影响,提出的表达式如下:

$$\text{PReLU}(x) = \begin{cases} x, & x \geq 0 \\ kx, & x < 0 \end{cases} \quad (2)$$

式中: k 为负值特征的斜率。

当特征值小于 0 时,以斜率 k 表示负值特征的输出系数。由此避免了 ReLU 中易出现神经元死亡的问题。

问题。

针对 k 有以下说明:当 $k=0$ 时,激活函数回归为 ReLU;当 k 为可训练的参数时,则激活函数变为 PReLU。

为实现模型轻量化,此处引入 SimAM 注意力机制以实现 k 值的求取。

1.3 SimAM 注意力机制

为提高模型的识别精度,同时保留激活函数中的负神经元,笔者引入 SimAM 注意力机制,自适应计算系数 k ,在减轻模型训练负担的同时提高特征表达;以能量函数表达各个神经元的活跃程度,以更高的权重赋予目标神经元,信息匮乏的神经元将被抑制。

能量函数的表达式如下:

$$e_t(w_t, b_t, y_t, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1}(y_0 - \hat{x}_i)^2 \quad (3)$$

式中: t 为目标神经元; x_i 为其余神经元; w_t, b_t 为权值和偏置; M 为神经元数; y_t, y_0 为指导搜索的最佳值。

其中: $\hat{t} = w_t t + b_t$, $\hat{x}_i = w_t x_i + b_t$ 。

为权值引入正则系数,则能量公式变为:

$$e_t(w_t, b_t, y_t, x_i) = \frac{1}{M-1}(-1 - (w_t x_i + b_t))^2 + (1 - (w_t x_i + b_t))^2 + \lambda w_t^2 \quad (4)$$

式中: λ 为正则化系数。

由式(4)求解 w_t, b_t 如下:

$$\begin{cases} w_t = \frac{2(t - \mu_t)}{(t - \mu_t)^2 + 2\sigma_t^2 + 2\lambda} \\ b_t = -\frac{1}{2}(t + \mu_t)w_t \end{cases} \quad (5)$$

式中: μ_t 为均值; σ_t 为方差。

最终简化计算最小能量值如下:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \mu_t)^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (6)$$

式中: $\hat{\sigma}$ 为协方差值。

根据式(6)每个特征向量的元素,求解均值和方差如下:

$$\begin{cases} \mu_t = \frac{1}{M} \sum_{i=1}^M x_i \\ \hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \mu_t)^2 \end{cases} \quad (7)$$

最小能量值的计算式表明:目标神经元与其余神经元的可分性,与能量值之间为反比关系;当能量值越小时,目标神经元对应的可区分性越大。因此,可将注意力参数以 $1/e_t^*$ 表示,去掉常数项,并进一步化简,得到的表达式如下:

$$z = 4(\hat{\sigma}^2 + \lambda)(t - \mu_t)^2 \quad (8)$$

式中: z 为简化后的能量值。

利用 Sigmoid 对注意力参数值进行调整,将其限定在一定范围内,并将其作为 PReLU 函数的斜率系数。表达式如下:

$$k = \text{sigmod}(z) \quad (9)$$

当特征输入为正值时,PReLU 函数的梯度为 1;当输入为负值时,其梯度值为 k 。因此,可在不增加模型

复杂度的情况下,确定 PReLU 的参数值。

2 基于 SimAM 注意力机制的轴承故障诊断模型

在跨工况域下的滚动轴承故障诊断中,为避免轴承的振动信号被破坏,影响故障诊断的性能,笔者建立了基于 SimAM 注意力机制的轴承故障诊断模型,如图 2 所示。

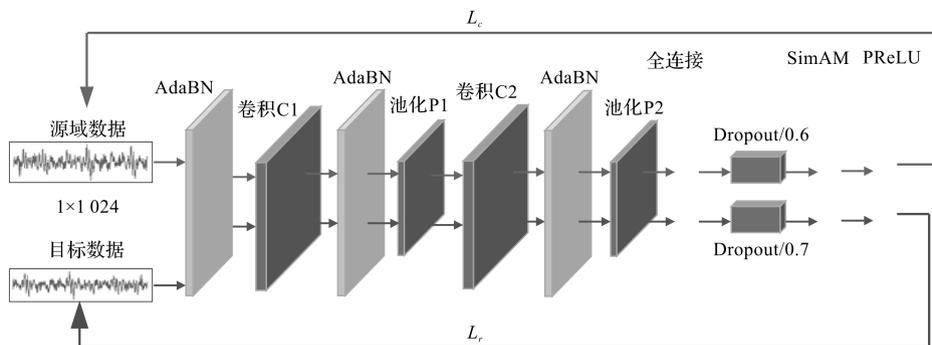


图 2 轴承故障诊断模型框架

Fig. 2 Bearing fault diagnosis model framework

图 2 中,模型以 CNN 作为基本诊断框架,直接将 1 024 个数据点组成的一维原始振动信号,作为诊断模型的输入,从而提取域不变特征。特征提取结构由三层 AdaBN、两层卷积、两层池化和一层 PReLU 组成。原始信号经 AdaBN 归一化后直接输入到第一卷积层中,其卷积核大小为 64×1 ,从而初步提取轴承故障诊断的域不变特征,同时去除无用的特征。

为进一步降低工况变化引起的数据分布差异,使模型具备良好的训练效果,提高跨工况域下的轴承故障诊断泛化性能,需在每层卷积层后均设置 AdaBN 和最大池化层。

AdaBN 可根据输入数据的不同,自适应地调整均值与方差,从而增强网络的自适应能力,其表达式如下:

$$\mu_t = \frac{1}{N} \sum_{i=1}^N x_i \quad (10)$$

$$\sigma_t^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_t)^2 \quad (11)$$

$$\hat{x}_t = x_t - \mu_t \sqrt{\sigma_t^2 + \varepsilon} \quad (12)$$

$$y_t = \gamma_s \hat{x}_t + \beta_s \quad (13)$$

式中: x_t 为目标域中的输入特征; y_t 为目标域中的输出特征; ε 为一个接近 0 的常数; γ_s, β_s 为可训练的参数,分别用于缩放和移动输出特征。

所构建的诊断模型参数如表 1 所示。

表 1 模型结构

Table 1 Model structure

参数名称/单位	结构参数	输出
C1 卷积核个数	64	$1 \times 1\ 024 \times 64$
C1 卷积核尺寸	1×4	$1 \times 1\ 024 \times 64$
C1 步长	1	$1 \times 1\ 024 \times 64$
P1 方式	最大池化 2×2	$1 \times 512 \times 64$
P1 步长	2	$1 \times 512 \times 64$
C2 卷积核个数	32	$1 \times 512 \times 32$
C2 卷积核尺寸	1×4	$1 \times 512 \times 32$
C2 步长	1	$1 \times 512 \times 32$
P2 方式	最大池化 2×2	$1 \times 256 \times 32$
P2 步长	2	$1 \times 256 \times 32$
全连接	8 192	$1 \times 8\ 192$
Dropout/%	源域 60	—
	目标域 70	—
激活函数	改进 PReLU	故障类型数 N

表 1 中,为使 ReLU 激活函数保留负值特征,采用基于 SimAM 无参数注意力机制的 PReLU 自适应提取负值特征。同时,为避免模型过拟合,基于 Dropout 技术,分别对源域数据和目标域数据做 0.6、0.7 的随机节点失活操作。

为充分识别跨工况下轴承的不同健康状态,可将模型的输入数据分为带标签的源域数据和无标签的目标域数据,并以交叉熵损失函数训练源域数据。其表达式如下:

$$L_c = \frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{c=1}^C y_i^c \log \hat{y}_i^c \quad (14)$$

式中: n_s 为源域的样本数; C 为故障状态数; y_i, \hat{y}_i 为真实值和预测值。

以 MSLE 作为目标域的损失函数, 其表达式如下:

$$L_M = \frac{1}{n_t} \sum_{i=1}^{n_t} (\log(y_i + 1) - \log(\hat{y}_i + 1))^2 \quad (15)$$

式中: n_t 为目标域的样本数。

综合两种损失函数, 目标损失函数表达式如下:

$$L = \alpha L_c + \beta L_M \quad (16)$$

式中: α, β 为两个损失函数的权重, 经不同权重下模型识别效果的对比实验, 最终取 $\alpha = 0.6, \beta = 0.8$; L 为综合损失函数。

3 实验分析

3.1 实验装置描述

为验证轴承故障诊断模型在变工况环境下的性能, 笔者采用自制实验台和凯斯西储大学轴承数据集 (简称 CWRU), 对轴承故障诊断模型的有效性进行验证。

自制实验台如图 3 所示。

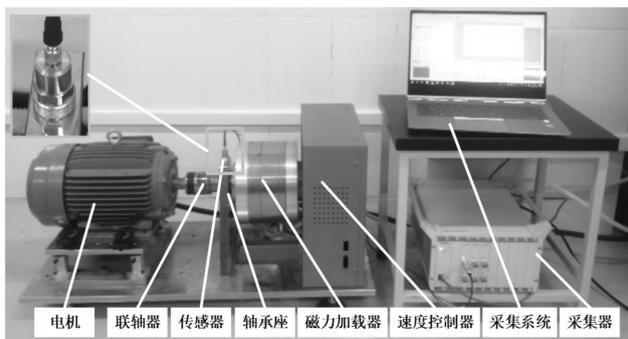


图 3 自制轴承故障实验台

Fig. 3 Self-made bearing failure test bench

自制实验台包括驱动电机 (转速范围为 $0 \sim 3\,000$ r/min)、磁力加载器 (加载范围为 $0 \sim 15$ N·m); 轴承型号为 NU202, 可模拟轴承正常状态、内圈故障、外圈故障、滚动体故障 (分别用 N、IF、OF、RF 表示); 故障尺寸深度为 0.5 mm、直径为 0.4 mm。

自制实验台工况说明如表 2 所示。

表 2 中, 速度工况分为 $1\,000$ r/min、 $1\,500$ r/min、 $1\,800$ r/min, 载荷工况为 5 N·m、 10 N·m、 15 N·m, 每种工况下均包含 N、IF、OF、RF 四种健康状况, 因此, 笔者共构建了不同转速和载荷下的 9 种工况。

在每个工况下, 笔者都对轴承的健康状态采集 $1\,000$ 个样本, 每个样本均由 $1\,024$ 个数据点组成, 9 种工况下共计 $36\,000$ 个样本。

表 2 自制实验台实验工况

Table 2 Self-made experimental bench test conditions

工况	转速/(r/min)	载荷/(N·m)	故障类型
Z1	1 000	5	N/IF/OF/RF
Z2	1 000	10	N/IF/OF/RF
Z3	1 000	15	N/IF/OF/RF
Z4	1 500	5	N/IF/OF/RF
Z5	1 500	10	N/IF/OF/RF
Z6	1 500	15	N/IF/OF/RF
Z7	1 800	5	N/IF/OF/RF
Z8	1 800	10	N/IF/OF/RF
Z9	1 800	15	N/IF/OF/RF

为进一步验证模型的泛化性能, 采用 CWRU 数据的工况说明如表 3 所示。

表 3 CWRU 实验工况

Table 3 CWRU test condition

工况	转速/(r/min)	载荷/hp	故障类型
W1	1 797	0	N/IF/OF/RF
W2	1 772	1	N/IF/OF/RF
W3	1 750	2	N/IF/OF/RF
W4	1 750	3	N/IF/OF/RF

取驱动端轴承故障数据集为验证数据, 每种故障均包含 17.78 cm、 35.56 cm、 53.34 cm 三种故障尺寸, 共有不同转速和载荷下的 4 种工况; 每种工况下选取 $1\,000$ 个数据样本, 以数据平滑叠加取样的方式, 每个样本取 $1\,024$ 个数据点, 4 种工况下共计 $16\,000$ 个样本。

根据实验数据, 笔者将所提方法在 Python3.7 软件中, 基于 TensorFlow2.1 框架编程运行。其中, 实验平台为 Dell G3 Intel Core i7-10750H CPU, 内存为 16 G, GPU 型号为 NVIDIA GeForce RTX 2060。

此处采用 Adam 优化器, 以最小化损失函数为目标, 以实现模型参数的自适应优化。

初始学习率 $l_0 = 0.01$, 学习率的更新表达式如下:

$$l_r = l_0 \times \gamma^{epoch} \quad (17)$$

式中: l_r 为更新后的学习率; γ 为学习率更新的乘法因子, 取值 0.1 ; $epoch$ 为这个数据集训练网络的次数。

3.2 自制实验台实验对比

自制实验台中, 笔者分别以表 2 中 Z1 ~ Z9 工况下的一种工况数据为源域数据, 其余 8 种工况为目标域数据, 对模型进行训练。

在不同工况下, 迁移诊断任务的识别准确率如表 4 所示。

表4 自制实验台不同工况迁移的识别率

Table 4 The recognition rate of the self-made test bench under different working conditions (%)

T \ S	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	均值
Z1	—	73.26 ± 1.02	85.17 ± 1.41	99.12 ± 0.77	94.43 ± 2.19	92.14 ± 0.52	98.31 ± 0.14	78.15 ± 4.38	92.22 ± 4.13	89.10 ± 1.82
Z2	84.34 ± 2.21	—	93.11 ± 2.11	88.12 ± 1.55	100 ± 0.0	93.21 ± 0.88	94.17 ± 1.33	98.66 ± 1.11	90.19 ± 0.52	92.73 ± 1.21
Z3	85.19 ± 2.88	91.1 ± 0.15	—	74.35 ± 1.21	98.55 ± 1.76	100.0 ± 0.0	98.47 ± 0.17	98.34 ± 0.47	97.17 ± 2.69	92.90 ± 1.17
Z4	100 ± 0.0	81.27 ± 1.74	92.87 ± 1.94	—	98.96 ± 0.27	96.37 ± 0.29	98.62 ± 0.95	73.61 ± 8.69	89.93 ± 0.35	91.45 ± 1.78
Z5	86.78 ± 2.91	100 ± 0.0	98.32 ± 0.26	100 ± 0.0	—	98.53 ± 0.65	100 ± 0.0	100 ± 0.0	99.16 ± 0.30	97.85 ± 0.52
Z6	90.64 ± 0.24	86.42 ± 0.13	100 ± 0.0	88.39 ± 3.08	99.58 ± 0.33	—	98.24 ± 0.81	100 ± 0.0	97.11 ± 3.39	95.05 ± 1.00
Z7	99.11 ± 0.81	78.22 ± 1.69	77.24 ± 5.44	99.08 ± 0.09	97.31 ± 0.33	98.56 ± 0.76	—	94.44 ± 3.69	95.93 ± 0.92	92.49 ± 1.72
Z8	80.16 ± 4.24	98.35 ± 2.11	99.04 ± 0.05	75.52 ± 5.98	98.83 ± 0.76	100 ± 0.0	95.53 ± 2.23	—	95.09 ± 3.91	92.82 ± 2.41
Z9	79.21 ± 0.36	91.56 ± 0.92	92.88 ± 0.89	87.46 ± 1.88	100 ± 0.0	100 ± 0.0	99.56 ± 0.85	100 ± 0.0	—	93.83 ± 0.61

注:S为源域;T为目标域。

由表4可知:所提方法可获得较高的准确率,当载荷变化时,部分迁移任务的识别率可达100%。当Z5作为源域工况时,跨工况的识别效果最好,平均识别率可达97.85%。

为进一步分析所提方法的有效性,笔者以Z5作为源域,分别构建模型1至模型4,以进行对比分析。为防止模型的随机性,每种模型下获得的识别率为重复10次实验的平均值。

模型1。所有输入均不包含归一化,即在网络模型中移除 AdaBN 模块;

模型2。未引入注意力机制,即移除 SimAM 模块,将全连接后的数据经随机节点失活后直接输入到激活函数中,激活函数 PReLU 的负值特征梯度值 k 取 0.01;

模型3。将激活函数 PReLU 替换为 Sigmoid 函数;

模型4。以激活函数 ReLU 直接替换 PReLU。

几种模型的结果如表5所示。

表5 几种方法在自制实验台上的对比实验

Table 5 Comparative experiments of several methods on self-made bench (%)

模型	识别率
模型1	81.13 ± 3.77
模型2	84.56 ± 2.99
模型3	77.62 ± 4.86
模型4	69.33 ± 7.11
该文模型	98.79 ± 1.01

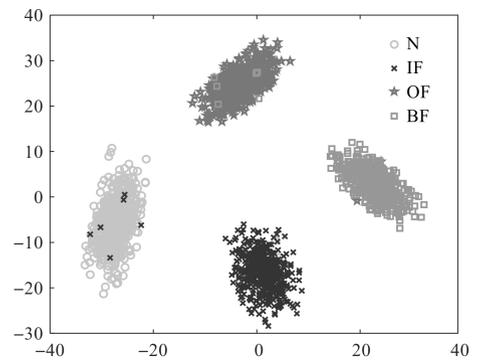
由表5可知:结果进一步验证了所提方法的优越性。当对模型去除归一化后,模型的识别率为81.13%,表明归一化可提高模型的泛化性能,从而改善识别精度。

此外,激活函数 PReLU 在负值处的斜率系数 k 为固定值 0.01 时的识别率为 84.56%,远低于引入注意力机制后的识别率,表明引入注意力机制后能有效提取工况变化时的域不变特征。

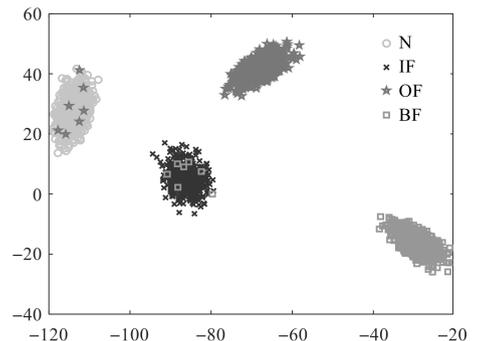
当激活函数替换为 Sigmoid 时,识别率仅为

77.62%,迁移效果并不理想。若以原始的 ReLU 作为激活函数,识别率仅为 69.33%,负迁移现象较为严重,可能出现过拟合现象。

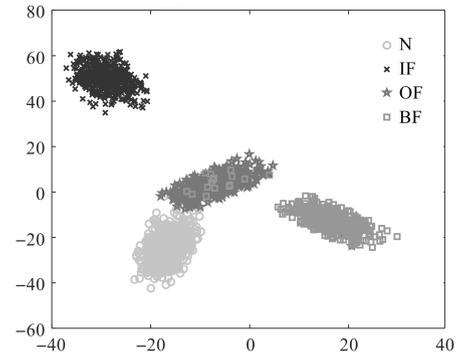
为更好地说明各种模型所提取的特征,笔者根据 t-SEN 降维在二维空间中实现分类特征的可视化,如图4所示。



(a) 模型1特征可视化



(b) 模型2特征可视化



(c) 模型3特征可视化

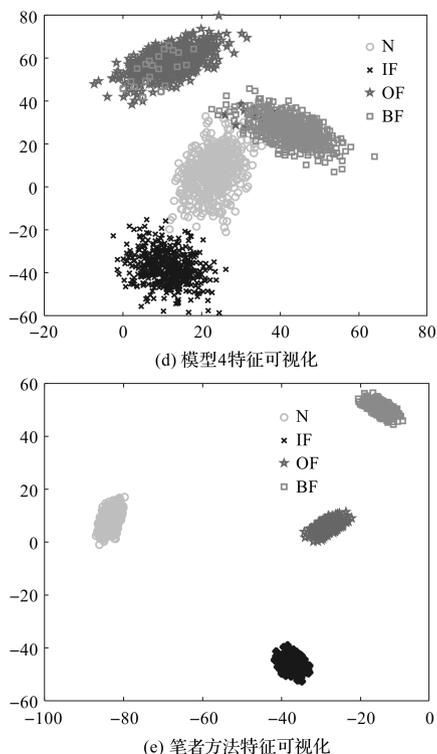


图 4 自制实验台特征可视化

Fig. 4 Feature visualization of self-made experimental bench

以 Z5 到 Z9 的迁移任务为例,从图 4 中可看出模型 1 至模型 4 的几种方法中,特征存在明显的重叠现象,因此无法有效实现分类。比较模型 1 和模型 2,可知模型中引入 AdaBN 归一化的必要性;若模型去除归一化,则特征聚类效果变差。

此外,观察模型 4 可发现,特征的混叠现象更为严重,表明当直接以 ReLU 作为激活函数时,域不变特征的提取效果较差,模型的迁移效果不佳。

由图 4 可知:笔者提出的基于 SimAM 注意力机制的轴承故障迁移诊断方法能有效实现跨工况域下的特征分离。该方法不仅对特征有较好的可分性,而且还具备较好的聚类效果。

3.3 CWRU 实验对比

为验证方法的普适性,笔者依托 CWRU 轴承公开数据集对基于 SimAM 的方法(运用 3.2 节中的几种模型对每种迁移方式)进行了验证。W1 ~ W4 四种工况下共有 12 种迁移任务。

CWRU 数据集不同方法的诊断结果如表 6 所示。

表 6 CWRU 数据集不同方法的诊断结果

Table 6 Diagnostic results of different methods in the CWRU dataset

(%)

方法 迁移	模型 1	模型 2	模型 3	模型 4	笔者方法
W1-W2	82.34 ± 0.33	87.13 ± 0.77	77.34 ± 0.64	70.11 ± 1.83	99.07 ± 0.12
W1-W3	80.89 ± 0.22	82.21 ± 0.75	76.53 ± 0.87	68.33 ± 0.48	100 ± 0.0
W1-W4	80.11 ± 0.35	80.88 ± 1.22	78.56 ± 1.22	69.09 ± 5.55	98.79 ± 0.29
W2-W1	84.21 ± 0.46	88.65 ± 1.01	79.80 ± 0.98	69.27 ± 5.79	99.89 ± 0.11
W2-W3	84.49 ± 0.31	88.73 ± 0.32	80.11 ± 0.19	70.38 ± 0.78	100 ± 0
W2-W4	89.34 ± 0.87	86.33 ± 0.15	80.99 ± 0.35	69.33 ± 2.88	98.89 ± 0.14
W3-W1	83.56 ± 0.43	85.99 ± 0.89	79.36 ± 0.77	70.52 ± 1.93	97.18 ± 0.98
W3-W2	86.39 ± 0.32	87.44 ± 0.77	78.54 ± 0.44	69.91 ± 3.77	98.12 ± 0.99
W3-W4	86.97 ± 0.23	87.75 ± 0.62	80.22 ± 0.98	70.79 ± 0.58	100 ± 0.0
W4-W1	85.82 ± 0.66	83.54 ± 0.76	80.01 ± 1.699	67.54 ± 1.96	94.55 ± 0.77
W4-W2	80.64 ± 0.58	82.39 ± 0.68	75.22 ± 0.69	70.98 ± 0.53	98.13 ± 0.96
W4-W3	85.49 ± 0.83	86.49 ± 1.15	74.33 ± 0.26	71.81 ± 2.31	100 ± 0.0
均值	84.15 ± 0.47	85.71 ± 0.82	78.62 ± 0.73	69.83 ± 2.34	98.68 ± 0.40

由表 6 可知:当激活函数直接替换为 ReLU 时,识别效果最差,平均识别率仅为 69.83% (由于激活函数省去了负值特征,导致域不变特征提取效果不佳,从而导致识别率较低);

当激活函数替换为 Sigmoid 后,识别效果略有提升,但依然不到 80%,表明特征值较大时容易出现梯度消失。

在上面几种方法中,笔者所提方法的识别效果可

达 98.68%;当源域为 W4,目标域为 W1 时,迁移效果偏差,仅为 94.55%,证明当转速和载荷跨度都较大时,工况域间的迁移效果变差。

4 结束语

为解决跨工况域时,轴承故障的域不变特征难以提取这一问题,笔者提出了一种基于 SimAM 注意力机制的轴承故障迁移诊断模型。

以 CNN 作为基本网络构架,并引入 AdaBN 对各层输入进行了归一化处理,同时对传统的 PReLU 激活函数进行了改进,并以交叉熵和 MSLE 作为损失函数,采用 SimAM 实现了对损失函数参数的优化目的;同时,利用 CWRU 数据集和自制实验台所得数据,对基于 SimAM 注意力机制的轴承故障迁移诊断方法进行了验证。

研究结论总结如下:

1) 针对变工况下滚动轴承故障迁移诊断效果差的问题,提出了基于 SimAM 注意力机制的滚动轴承故障迁移诊断方法;解决了激活函数 PReLU 负值输入时,斜率系数难以求解的问题,有效避免了网络过拟合的风险;

2) 以交叉熵损失函数监督训练带标签的源域数据,以 MSLE 作为目标域的损失函数进行了无监督训练,并利用 AdaBN 作为每层卷积输出的归一化手段,实验证明 AdaBN 有效地提高了模型的泛化性能;

3) 根据自制实验台和 CWRU 两个数据集对所提方法进行了验证,结果表明所提方法能有效提取跨工况域下的域不变特征,有效实现跨速度、跨载荷工况下的故障迁移诊断。

目前,笔者仅使用两种实验数据集对上述故障诊断模型进行了验证。下一步,笔者将利用工业现场实验数据,对模型开展实际工程的应用研究,同时对轴承故障诊断的跨设备迁移进行研究,为轴承故障由实验向工业现场的迁移诊断提供参考。

参考文献 (References):

- [1] CHEN Yong-yi, ZHANG Dan. Dual-path mixed-domain residual threshold networks for bearing fault diagnosis[J]. IEEE. Transactions on Industrial Electronics, 2022 (12): 13462-13472.
- [2] 牟红霞. 轴承变工况故障的域自适应迁移深度学习诊断[J]. 机械设计与制造, 2024, 62(2): 364-368.
MOU Hong-xia. Bearing fault diagnosis under variable conditions based on domain adaptive migration deep learning[J]. Machinery Design & Manufacture, 2024, 62(2): 364-368.
- [3] 刘亚. 数据驱动的滚动轴承故障诊断与健康状态评估[D]. 济南: 山东大学控制科学与工程学院, 2019.

- LIU Ya. Fault Diagnosis and Health Assessment of Rolling Bearings Based on Data-driven Model[D]. Jinan: School of Control Science and Engineering, Shandong University, 2019.
- [4] 刘博, 李晨, 阎彦, 等. 电机驱动系统故障诊断技术综述[J]. 中国电机工程学报, 2023, 43(14): 5619-5634.
LIU Bo, LI Chen, YAN Yan, et al. Review of fault diagnosis techniques for motor drive systems[J]. Proceedings of the CSEE., 2023, 43(14): 5619-5634.
- [5] 邓巍, 韩斌. 风力发电机轴承故障包络谱扫频分析诊断[J]. 热力发电, 2019, 48(7): 126-130.
DENG Wei, HAN Bin. Envelope spectrum sweeping analysis method for bearing fault diagnosis of wind turbine[J]. Thermal Power Generation, 2019, 48(7): 126-130.
- [6] WANG Huan, LIU Zhi-liang, PENG Dan-dan, et al. Feature-level attention-guided multitask CNN for fault diagnosis and working conditions identification of rolling bearing[J]. IEEE. Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4757-4769.
- [7] 程亮, 董子健, 王树民, 等. 基于改进一维卷积神经网络的滚动轴承故障诊断分析[J]. 机械设计与研究, 2023, 39(3): 126-130.
CHEN Liang, DONG Zi-jian, WANG Shu-min, et al. Research on fault diagnosis of rolling bearing based on improved one-dimensional convolutional neural network[J]. Machine Design & Research, 2023, 39(3): 126-130.
- [8] FENG Yong, CHEN Jing-long, HE Shui-long, et al. Globally localized multisource domain adaptation for cross-domain fault diagnosis with category shift[J]. IEEE. Transactions on Neural Networks and Learning Systems, 2021, 34(6): 3082-3096.
- [9] CHEN Yong-yi, ZHANG Dan, YAN Ru-qiang. Domain adaptation networks with parameter-free adaptively rectified linear units for fault diagnosis under variable operating conditions[J]. IEEE. Transactions on Neural Networks and Learning Systems, 2023(8): 1-14.
- [10] ZHANG Wei, PENG Gao-liang, LI Chuan-hao, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425.
- [11] GUO Liang, LEI Ya-guo, XING Sai-bo, et al. Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data[J]. IEEE. Transactions on Industrial Electronics, 2019, 66(9): 7316-7325.

(下转第 893 页)

本文引用格式:

包从望, 朱广勇, 邹旺, 等. 基于 SimAM 注意力机制的轴承故障迁移诊断模型[J]. 机电工程, 2024, 41(5): 862-869, 893.

BAO Congwang, ZHU Guangyong, ZOU Wang, et al. Rearing fault transfer diagnosis model based on SimAM attention mechanism[J]. Journal of Mechanical & Electrical Engineering, 2024, 41(5): 862-869, 893.

《机电工程》杂志: <http://www.meem.com.cn>

- Tool & Hydraulics,2020,48(9):117-119.
- [18] 徐礼林. 轴向柱塞泵配流机构过流面积分析及计算[J]. 机床与液压,2011,39(24):54-56.
XU Li-lin. Analysis and calculation of flow port areas in axial piston pump[J]. Machine Tool & Hydraulics,2011,39(24):54-56.
- [19] 那成烈. 轴向柱塞泵可压缩流体配流原理[M]. 北京:兵器工业出版社,2003.
NA Cheng-lie. Principle of Compressible Fluid Flow Distribution in Axial Piston Pump [M]. Beijing: The Publishing House of Ordnance Industry,2003.
- [20] 高伟,张立茹,姚慧龙,等. 基于动网格技术水平轴风力机叶片及尾迹流场旋涡特性[J]. 排灌机械工程学报,2023,41(2):172-178.
GAO Wei, ZHANG Li-ru, YAO Hui-long, et al. Vortex characteristics of horizontal axis wind turbine blades and wake flow field based on dynamic grid technology [J].
- Journal of Drainage and Irrigation Machinery Engineering, 2023,41(2):172-178.
- [21] 李树勋,刘斌才,吴翰林,等. 基于动网格的轴流式止回阀关闭过程动态特性分析[J]. 流体机械,2023,51(10):69-75.
LI Shu-xun, LIU Bin-cai, WU Han-lin, et al. Dynamic characteristics analysis of axial flow check valve closing process based on dynamic grid[J]. Fluid Machinery,2023,51(10):69-75.
- [22] 董亮,张立新,刘厚林,等. 车用凸轮式氢气循环泵内非定常流动特性[J]. 排灌机械工程学报,2022,40(5):433-439.
DONG Liang, ZHANG Li-xin, LIU Hou-lin, et al. Unsteady flow characteristics inside a cam type hydrogen circulation pump for vehicles [J]. Journal of Drainage and Irrigation Machinery Engineering, 2022, 40 (5) : 433-439.

本文引用格式:

郭志敏,戴海曙,翟江,等. 基于3D空间点云模型的轴向柱塞泵配流盘过流面积计算方法[J]. 机电工程,2024,41(5):886-893.

GUO Zhimin, DAI Haishu, ZHAI Jiang, et al. Calculation method of flow area of valve plate of axial piston pump based on 3D spatial point cloud model[J]. Journal of Mechanical & Electrical Engineering, 2024,41(5):886-893. 《机电工程》杂志;http://www.meem.com.cn

(上接第869页)

- [12] ZHAO Xiao-li, YAO Jian-yong, DENG Wen-xiang, et al. Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network [J]. IEEE. Transactions on Neural Networks and Learning Systems,2022,34(9):6339-6353.
- [13] 袁红春,张文凤. 融合 SimAM 注意力机制和双向 ConvLSTM 的异常检测方法[J]. 小型微型计算机系统, 2023,44(8):1777-1784.
YUAN Hong-chun, ZHANG Wen-feng. Anomaly detection method fused with SimAM attention mechanism and bidirectional convlstm [J]. Journal of Chinese Computer Systems,2023,44(8):1777-1784.
- [14] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qi, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [J]. IEEE. Computer Society,2015(12):1026-1034.
- [15] THILO S, VASILY T. Beyond ImageNet: deep learning in industrial practice [J]. Applied Data Science, 2019 (6) : 205-232.
- [16] PARK H, KANG Y, KIM J. Enhancing structure-property relationships in porous materials through transfer learning and cross-material few-shot learning [J]. ACS Applied Materials & Interfaces,2023(15):56375-56385.
- [17] WANG Yu, LIU Ruo-nan, LIN Di, et al. Coarse-to-fine: progressive knowledge transfer-based multitask convolutional neural network for intelligent large-scale fault diagnosis[J]. IEEE. Transactions on Neural Networks and Learning Systems,2021(8):761-774.
- [18] 杜义浩,常超群,杜正,等. 基于改进迁移学习的运动想象分类识别算法[J]. 计量学报,2023,44(11):1740-1748.
DU Yi-hao, CHANG Chao-qun, DU Zheng, et al. Classification and recognition algorithm of motion imagination based on improved transfer learning [J]. Acta Metrologica Sinica,2023,44(11):1740-1748.