

基于锐度感知增强卷积神经网络的变工况机械故障诊断

范家源, 徐德胜, 罗灵鲲, 胡士强

引用本文

范家源, 徐德胜, 罗灵鲲, 胡士强. [基于锐度感知增强卷积神经网络的变工况机械故障诊断](#)[J]. 计算机科学, 2024, 51(10): 320-329.

FAN Jiayuan, XU Desheng, LUO Lingkun, HU Shiqiang. [Mechanical Fault Diagnosis Under Variable Working Conditions Based on Sharpness Awareness Reinforced Convolutional Neural Network](#) [J]. Computer Science, 2024, 51(10): 320-329.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于注意力机制的CNN和BiGRU的加密流量分类](#)

Encrypted Traffic Classification of CNN and BiGRU Based on Self-attention
计算机科学, 2024, 51(8): 396-402. <https://doi.org/10.11896/jsjcx.230500032>

[基于熵值过滤和类质心优化的无监督域适应](#)

Unsupervised Domain Adaptation Based on Entropy Filtering and Class Centroid Optimization
计算机科学, 2024, 51(7): 345-353. <https://doi.org/10.11896/jsjcx.230500144>

[基于TCN-A模型的高效查询负载预测算法](#)

Efficient Query Workload Prediction Algorithm Based on TCN-A
计算机科学, 2024, 51(7): 71-79. <https://doi.org/10.11896/jsjcx.231100200>

[结合图卷积神经网络和集成方法的推荐系统恶意攻击检测](#)

Malicious Attack Detection in Recommendation Systems Combining Graph Convolutional Neural Networks and Ensemble Methods
计算机科学, 2024, 51(6A): 230700003-9. <https://doi.org/10.11896/jsjcx.230700003>

[融合三维人脸动态信息和光流信息的人脸表情识别](#)

Facial Expression Recognition Integrating 3D Facial Dynamic Information and Optical Flow Information
计算机科学, 2024, 51(6A): 230700210-7. <https://doi.org/10.11896/jsjcx.230700210>

基于锐度感知增强卷积神经网络的变工况机械故障诊断

范家源¹ 徐德胜² 罗灵鲲¹ 胡士强¹

1 上海交通大学航空航天学院 上海 200240

2 大型客机集成技术与模拟飞行全国重点实验室 上海 201210

(jiayuanf@sjtu.edu.cn)

摘要 传统的深度迁移学习网络从有标签的源域故障数据中学习并迁移知识,完成无标签目标域上的诊断任务,有效解决了机械故障诊断中变工况场景面临的数据特征空间非对称问题。然而其知识迁移模块加剧了深度学习网络结构的复杂性,导致其损失函数的地貌特征比浅层网络复杂得多,优化难度更高。传统方法无法感知损失函数的地貌特征,容易使模型参数陷入参数泛化间隙大的局部最小值处,导致模型泛化性差,诊断精度降低。为了应对这一挑战,研究提出锐度感知增强的卷积神经网络(Sharpness Awareness Reinforced Convolutional Neural Network, SA-CNN),通过感知一定范围内模型损失函数的锐度,联合优化损失函数与其地貌特征的平坦程度,约束模型参数向损失函数锐度降低的方向收敛,进而提升模型的泛化性能。经典机械故障诊断数据集上的实验结果表明,相比传统的深度迁移模型,所提方法在变工况场景下进行跨域机械故障诊断时性能提升显著。

关键词 轴承故障诊断;损失函数地貌分析;迁移学习;卷积神经网络

中图分类号 TP181

Mechanical Fault Diagnosis Under Variable Working Conditions Based on Sharpness Awareness Reinforced Convolutional Neural Network

FAN Jiayuan¹, XU Desheng², LUO Lingkun¹ and HU Shiqiang¹

1 School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China

2 State Key Laboratory of Airliner Integration Technology and Flight Simulation, Shanghai 201210, China

Abstract Traditional deep transfer learning networks have effectively addressed the challenges arising from the asymmetry introduced by cross-domain data distributions in variable operational scenarios. It is achieved by leveraging knowledge learned from labeled fault data and applying it to the task of diagnosing unlabeled fault data collected under varying conditions. However, the inclusion of knowledge transfer modules has added complexity to the deep network's structure, resulting in a more intricate loss landscape. This, in turn, presents challenges for optimization. Traditional methods often struggle to navigate the sharpness of this loss landscape, leading to the model's parameters getting stuck in local minima characterized by high sharpness. This hinders model generalization and reduces accuracy. To tackle this challenge, this paper proposes the sharpness awareness reinforced convolutional neural network (SA-CNN). This approach involves a joint optimization of the loss function and its flatness by assessing sharpness within a specified range. This process steers the fault diagnosis model parameters away from regions of high sharpness, ultimately improving model generalization. Extensive experiments on established mechanical fault diagnosis datasets demonstrate that, compared to traditional deep transfer learning-based fault diagnosis models, the proposed SA-CNN significantly enhances the performance of bearing fault diagnosis under varying working conditions.

Keywords Bearing fault diagnosis, Loss function landscape analysis, Transfer learning, Convolutional neural network

1 引言

机械设备的稳定和安全运行,始终是工业界最关注的问题之一。随着科学和技术水平不断发展,机械系统日趋复杂;伴随复杂系统的综合化和智能化程度不断提高,其维护保障

成本和难度急剧提升^[1]。为了减少不必要的停机检查和维修,防止突发故障导致人员伤亡^[2],机械设备的故障诊断方法受到越来越多的关注^[3]。其中,基于深度学习的模型优化技术作为一种数据驱动的方法,可以避免对复杂机械进行建模,具有强大的特征挖掘和模式识别能力。与此同时,工业生产

到稿日期:2023-09-25 返修日期:2024-03-12

基金项目:国家自然科学基金(61773262,62006152);中国航空科学基金(2022Z071057002,20142057006)

This work was supported by the National Natural Science Foundation of China (61773262,62006152) and China Aviation Science Foundation (2022Z071057002,20142057006).

通信作者:胡士强(squhu@sjtu.edu.cn)

系统的智能化和集成化水平不断提高,各种传感器的应用和海量的历史数据也为基于深度学习的方法提供了数据基础。近年来,基于深度学习的故障诊断方法成为了应用和研究的热点,在有监督故障诊断领域取得了长足的进步^[4-6]。但是这些深度学习算法通常假设训练样本与测试样本独立同分布^[7],同时极度依赖海量有标签的高质量训练数据来进行模型训练。然而,机械故障诊断的实际应用场景往往难以满足上述高质量先验标注的数据保障^[7]。首先,由于设备的工况复杂多变,其数据分布可能随工况变化而改变,无法满足上述较强的假设先验条件^[8];其次,进行数据标注的人力与时间成本很高,无法对全部工况下的故障数据进行标注,导致大部分工况下的数据无监督。上述由设备变工况带来的问题会导致算法的诊断精度下降,限制深度学习算法在实际工程中的应用。针对上述问题,近期的热点研究聚焦于深度迁移方法,旨在将在已有标记数据上所学习的知识和模型迁移到未标记且概率分布有差异的数据上进行预测和分类^[9-11]。其中,深度迁移优化技术作为机器学习领域的一个重要研究分支,可以有效处理传统深度学习所遇到的域偏移问题并做出改善和提升,极大地促进了变工况场景下智能故障诊断方法的发展。

基于深度迁移学习的故障诊断算法采用域自适应技术结合有标签的故障数据集(源域)完成决策模型的训练,拟在无标签的其他工况下精确辨识故障数据(目标域)。然而,上述深度迁移模型很大程度上依赖于迁移模型的损失设计与网络结构的优化增强^[12]。现有的研究策略一定程度上缓解了跨域任务面临的数据特征空间非对称性,但也因增加模型复杂度而加剧了源域大样本模型训练下的模型过度拟合^[13],进而导致测试数据集(目标域)上的性能衰退。针对该现象的理论分析为深度神经网络的优化模型呈现强非凸非光滑性,导致优化的函数空间频繁出现局部极小值。这些局部极小值在欧氏空间上近似,但对应的模型参数在目标域上的性能不同,即泛化性能不同^[14]。图1结合二维损失函数地貌示意图展示了大小相同的损失函数极小值拥有不同泛化性能的结果可视化。损失函数地貌本质上是经验损失函数的结构,之所以称为地貌,是强调函数在参数空间中具有像地貌一般高低起伏的特性,这种特性对理解损失函数诱导的训练过程有重要的意义。可以发现,模型的训练损失作为其泛化损失的估计值,与泛化损失之间存在间隙(泛化间隙,Generalization Gap),间隙越大则代表模型越拟合于训练数据。通过观察图1中2个训练损失函数的极小值点可以发现,虽然2个极值点的训练损失相同,但是其泛化损失不同(由红色箭头代表),平坦的极小值点(图1中**b**点)具有更小的泛化损失。因此,为了优化泛化性能良好的跨域诊断模型,需要在保证最小化损失函数的基础上,尽可能平滑损失函数的地貌特征以减小泛化间隙。然而,现有的深度迁移技术缺乏对模型锐度的感知,因此无法选取泛化性能强的参数空间(图1中**b**点)进行优化,导致最终的模型优化易陷入泛化性能差的参数空间(图1中**a**点)。

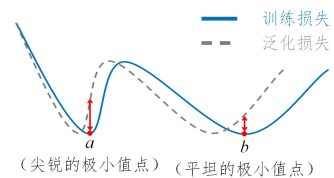


图1 二维损失函数地貌示意图(电子版为彩图)

Fig.1 Illustration of the landscape of 2D loss function

综上所述,基于深度迁移模型的设备变工况场景机械故障诊断技术存在如下挑战:深度迁移学习模块的嵌入加剧了传统深度模型的复杂性,导致模型的泛化性能下降,进而无法保障故障诊断任务的高效完成。该挑战详细讨论如下:深度学习网络自身存在参数量庞大、网络结构复杂的特点,而新增的域自适应模块在解决跨域优化问题的同时增加了模型的复杂度,进而导致训练后的网络泛化性差,无法保障目标域故障的精准预测。

针对上述挑战,本文提出了锐度感知增强的卷积神经网络(SA-CNN)。该算法旨在解决传统深度迁移算法无法感知不同损失函数极小值与模型泛化性能强弱的联系,导致模型易收敛到泛化能力较弱的局部极值参数空间,进而无法保障满足任务需求的故障预测精度的问题。本研究将锐度感知最小化方法(Sharpness Awareness Minimization, SAM)^[15]引入深度迁移学习的任务场景,并通过对网络损失函数锐度的感知与优化,有效提升迁移模型的泛化能力,保障诊断精度。本文的主要贡献概括如下:

- 1) 讨论了当前深度迁移学习算法在变工况机械故障诊断场景中缺乏对损失函数地貌锐度的感知而导致性能下降的问题。
- 2) 提出 SA-CNN 算法,有效解决了在变工况场景下深度迁移故障诊断模型中模型复杂度增加导致参数泛化性下降的问题。
- 3) 利用 5 个流行的深度迁移学习算法基准数据集,基于两个公开轴承数据集,设计了 14 个迁移任务,进行了广泛的实验。实验证明所提方法在跨域故障诊断任务中具有较强的竞争力。

2 相关工作

本文将从以下 3 个方面对相关的工作进行介绍:1)深度迁移学习方法在故障诊断领域中的应用现状;2)关于损失函数地貌分析的相关研究;3)锐度感知最小化方法的应用现状。

2.1 基于深度迁移学习的故障诊断

深度迁移学习是将在已有标记数据上所学习的知识和模型迁移到未标记且概率分布有差异的数据上进行预测和分类的一种机器学习方法^[11]。迁移学习作为机器学习领域的一个重要分支,可以解决传统机器学习方法依赖独立同分布假设先验条件的问题,进而高效解决更普遍的数据分布非对称问题。近些年,深度迁移学习较为热门的应用场景包括计算机视觉^[9]、文本分类^[16]、医疗健康和自然语言处理等。由于迁移学习在这些领域表现突出,因此不少研究者尝试将其

应用到故障诊断领域,并取得了不错的成果。其中,LYU等^[8]针对飞行器不同舵机轴承在不同工况下出现域偏移的问题,引入多源域深度迁移学习的思想,利用多核 MMD 距离度量减小源域与目标域之间的偏移,并结合特定域分类器降低不同域对目标样本在类边界附近的分类损失。Kavianpour等^[17]针对变工况和缺少标记的齿轮箱故障诊断问题,提出一种混合卷积神经网络和域适应网络的算法,通过卷积神经网络提取特征后,利用域对抗学习和深度域适应 2 个模块减少域间差异。Wang 等^[18]针对工程实际中往往缺乏标签和故障样本的问题,提出将专家知识和域适应相结合,实现无监督故障诊断。该研究首先考虑专家系统合成的先验数据集,然后利用不平衡鲁棒域适应方法将合成数据集与真实数据集进行域对齐。

综上,深度迁移学习技术已经被广泛应用于故障诊断工作中,并保障了变工况场景下故障诊断算法对数据分布有差异的目标域数据的有效辨识。但上述深度迁移模型基于已有的深度学习骨干网络进行结构和参数的再优化,在保障跨域特征辨识的同时也增加了模型参数优化的难度。该类模型的损失函数易产生更多局部极小值^[14],导致利用传统优化方法如随机梯度下降法^[19](Stochastic Gradient Descent, SGD)或自适应矩估计法^[20](Adaptive Moment Estimation, Adam)等无法收敛到具有良好泛化性能的模型参数中,进而无法保障跨域故障诊断算法的高诊断准确率。针对该问题,本文提出 SA-CNN 算法,通过捕获损失函数的锐度感知模型的泛化性,并在模型训练过程中进行优化,最终得到泛化性良好的深度迁移故障诊断模型。

2.2 损失函数地貌

损失函数地貌(Loss-landscape)是经验损失函数的结构,其几何形状与模型泛化性之间的相关性受到了广泛的研究。学者通过理论证明和广泛实验验证了拥有更平坦损失函数地貌的模型参数拥有更好的泛化性能。Dziugaite 等^[21]通过直接优化一个基于 PAC-Bayes 的双隐层神经网络泛化误差界,得到了深度神经网络的一个非空泛化界,对损失函数地貌的锐度(损失函数地貌的平坦程度)与模型泛化之间的联系进行了解释。Behnam 等^[22]概述了度量深度网络泛化性的方法,实验发现锐度相关的度量能够捕捉到深度神经网络的大部分泛化行为。目前,经验损失函数地貌分析已成为揭示深度网络泛化性能的有效途径之一,也是机器学习领域的重要研究方向。本研究尝试探索损失函数地貌分析在无监督域自适应的机械故障诊断任务中的贡献,将模型参数收敛位置邻域半径中的最大损失值与当前损失值的差看作模型损失函数地貌的锐度的度量方式,并与损失函数共同优化,以保障模型向具有更小泛化误差的损失函数极小值处收敛,进而保证模型的泛化性能,提升诊断准确率。

2.3 锐度感知最小化算法

受到锐度与模型泛化性能之间联系的启发,锐度感知最小化算法^[15]通过联合优化损失函数与锐度,使模型参数收敛到相对平坦的损失函数最小值处来获得更好的模型泛化

性能。目前,已有学者尝试结合锐度感知最小化(SAM)的优化方法提升不同领域深度学习模型的泛化能力。例如,在机器人视觉领域中,Chen 等^[23]通过实验证明 SAM 能够对 Transformer 和 MLP-Mixer 等图像识别模型进行更好的训练,以取得更好的泛化性能。除了针对单层网络泛化性能进行的相关工作,Momin 等^[24]探索了网络结构更加复杂的双层网络 MAML 的优化问题,在不增加计算开销的基础上优化了 MAML 的泛化性。

已有的基于 SAM 展开的模型泛化能力增强的研究工作已在很多领域取得了显著的成果^[23,25],但是在变工况机械故障诊断领域,该类方法的有效性仍待进一步探索。本研究尝试利用损失函数地貌的平坦程度与泛化性能之间的联系解释变工况场景下深度迁移的机械故障诊断算法性能下降的机理,并结合实际故障诊断数据集论证所提方法在无监督跨域机械故障诊断领域的贡献。结合深度卷积神经骨干网络,嵌入锐度感知最小化方法,提出的变工况场景故障诊断网络在目标域故障数据与源域故障数据存在分布差异时能够有效保障诊断精度。

3 基于 SA-CNN 跨域故障诊断算法

针对变工况场景下深度迁移学习算法无法感知不同极小值与模型泛化性能强弱的联系,导致模型易优化到泛化能力较弱的局部极值参数空间,进而无法保障故障预测精度的问题,提出一种锐度感知增强的卷积神经网络(SA-CNN),其利用损失函数的锐度感知模型的泛化性能与损失函数进行联合优化,有效提升了网络进行跨域故障诊断的性能。所提方法的整体框架如图 2 所示。模型训练包含 2 个关键步骤:步骤②利用最佳扰动的估计值虚拟更新模型参数,获得感知锐度损失的虚拟模型;步骤④利用锐度损失真实更新模型参数,获得强泛化性的诊断模型。

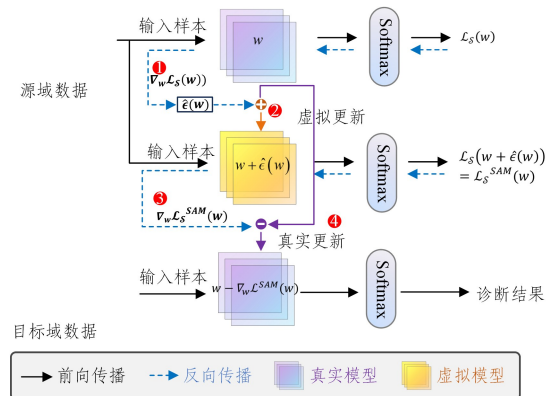


图 2 SA-CNN 网络的整体框架

Fig. 2 Framework of SA-CNN

本章中首先对算法场景进行定义,然后介绍锐度感知最小化方法与所提出的框架。

3.1 场景定义

假设场景中,源域数据是某工况下具有标签的数据,定义如下:

$$D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}, x_i^s \in X_s, y_i^s \in Y_s \quad (1)$$

其中, D_s 代表源域; X_s 为所有源域样本的集合, 包含 n_s 个样本; x_i^s 为集合中第 i 个样本, 标签为 y_i^s 。目标域数据是与源域不同的工况下无标签的数据, 定义如下:

$$D_t = \{x_i^t\}_{i=1}^{n_t}, x_i^t \in X_t \quad (2)$$

其中, D_t 代表目标域; X_t 为所有目标域样本的集合, 包含 n_t 个样本, x_i^t 为集合中第 i 个样本, 标签未知。

任务场景中, 希望找到一个故障诊断模型(见式(3)), 在正确诊断源域数据的基础上, 能够对目标域无标签的故障数据进行高质量判别。

$$\hat{y} = f(x; w) \quad (3)$$

其中, w 代表深度神经网络的参数, \hat{y} 代表诊断结果。因此, 可以定义基于源域数据的训练损失函数。

$$\mathcal{L}_s(w) \triangleq \frac{1}{n} \sum_{i=1}^{n_s} \mathcal{L}(f(x_i^s; w), y_i^s) = \frac{1}{n} \sum_{i=1}^{n_s} \mathcal{L}(\hat{y}_i^s, y_i^s) \quad (4)$$

其中, $\mathcal{L}(\cdot)$ 代表交叉熵损失函数。源域与目标域的数据遵循不同的分布, 分别以分布 \mathcal{S} 与分布 \mathcal{T} 进行表示。假设存在分布 \mathcal{D} 可以描述包含多种工况(本研究为两种工况)的故障数据的分布。任务的目标是找到泛化良好的诊断模型, 在遵循分布 D 的变工况数据下保证诊断精度。本文使用目标域数据的诊断结果计算模型的准确率, 该准确率能够突出反映模型的泛化性能, 准确率越高, 模型泛化性能越好。最终诊断模型对目标域故障数据的诊断结果表示为:

$$\hat{y}_i^t = f(x_i^t; w) \quad (5)$$

3.2 锐度感知增强的卷积神经网络

深度学习网络通过基于训练数据的经验损失进行训练, 旨在得到具有更低泛化损失的网络参数。当前方法通常将经验损失作为泛化损失的估计值, 结合 SGD 和 Adam 等方法最小化经验损失来训练网络参数。然而, 由于深度迁移网络的模型过度参数化, 其经验损失函数的优化空间存在多个局部极小值, 这些极小值在欧氏空间上近似, 但是对应模型参数的泛化性能存在显著不同。在跨域故障诊断任务中, 基于深度迁移学习的方法因无法感知不同极小值与模型泛化性能强弱的联系, 故其在目标域的诊断任务中的性能受到影响。

锐度感知最小化算法^[15]是利用损失函数地貌的锐度感知模型参数的泛化性能, 通过同时最小化损失函数值和地貌锐度来提升模型性能。算法基于以下定理^[15]。

定理 1 对于任意 $\rho > 0$, 在基于分布 \mathcal{D} 得到的遵循分布 \mathcal{S} 的训练集上, 不等式有很高的概率成立。

$$\mathcal{L}_s(w) \leq \max_{\|\epsilon\|_2 \leq \rho} \mathcal{L}_s(w + \epsilon) + h\left(\frac{\|\nabla_w \mathcal{L}_s(w)\|_2}{\rho}\right) \quad (6)$$

其中, w 为模型参数, h 为一个单调递增的函数。因此, 可以设计经验损失函数 $\mathcal{L}_s^{\text{SA-CNN}}(w)$ 作为泛化损失 $\mathcal{L}_s(w)$ 的估计值。

$$\begin{cases} \mathcal{L}_s^{\text{SA-CNN}}(w) \triangleq \mathcal{L}_s^{\text{SAM}}(w) + \lambda \|w\|_2^2 \\ \mathcal{L}_s^{\text{SAM}}(w) \triangleq \max_{\|\epsilon\|_2 \leq \rho} \mathcal{L}_s(w + \epsilon) \end{cases} \quad (7)$$

其中, ρ 为超参数; 超参数项 $\lambda \|w\|_2^2$ 代替定理 1 中的函数 h , 为模型参数的正则化项。可以将 $\mathcal{L}_s^{\text{SA-CNN}}(w)$ 改写为以下形式:

$$\begin{aligned} \mathcal{L}_s^{\text{SA-CNN}}(w) &\triangleq [\mathcal{L}_s^{\text{SAM}}(w) - \mathcal{L}_s(w)] + \mathcal{L}_s(w) + \lambda \|w\|_2^2 \\ &\triangleq \left[\max_{\|\epsilon\|_2 \leq \rho} \mathcal{L}_s(w + \epsilon) - \mathcal{L}_s(w) \right] + \mathcal{L}_s(w) + \\ &\quad \lambda \|w\|_2^2 \end{aligned} \quad (8)$$

其中, 方括号中的项度量了参数 w 在一定范围内扰动时损失函数增加的最大值, 反映了模型在参数 w 处损失函数地貌的平坦程度; 而方括号外的两项则是原卷积神经网络的损失函数与正则化项, 反映了网络在源域上的分类精度。因此, 通过最小化损失函数 $\mathcal{L}_s^{\text{SA-CNN}}(w)$ 可以实现卷积神经网络的交叉熵损失函数与其锐度的联合优化, 保障网络的诊断精度和损失函数地貌的平坦程度。

为了最小化 $\mathcal{L}_s^{\text{SA-CNN}}(w)$, 需要确定在设定的范围 $\|w\|_2 \leq \rho$ 内最优的扰动 ϵ , 最大化损失函数 $\mathcal{L}_s(w + \epsilon)$ 。根据 $\epsilon = 0$ 处的泰勒展开式, 可得到所求最优扰动 $\epsilon^*(w)$ 的近似表示:

$$\begin{aligned} \epsilon^*(w) &\triangleq \arg \max_{\|\epsilon\|_2 \leq \rho} \mathcal{L}_s(w + \epsilon) \\ &\approx \arg \max_{\|\epsilon\|_2 \leq \rho} \mathcal{L}_s(w) + \epsilon^T \nabla_w \mathcal{L}_s(w) \\ &= \arg \max_{\|\epsilon\|_2 \leq \rho} \epsilon^T \nabla_w \mathcal{L}_s(w) \end{aligned} \quad (9)$$

此时, 该问题可以看作求解向量 ϵ^T 与 $\nabla_w \mathcal{L}_s(w)$ 点乘的极值问题。

$$\epsilon^*(w) \approx \hat{\epsilon}(w) = \rho \frac{\nabla_w \mathcal{L}_s(w)}{\|\nabla_w \mathcal{L}_s(w)\|_2} \quad (10)$$

其中, ρ 为超参数, $\nabla_w \mathcal{L}_s(w)$ 可通过网络对损失函数的反向传播得到。将结果代入式(7)中并求导, 可得:

$$\nabla_w \mathcal{L}_s^{\text{SAM}}(w) \approx \nabla_w \mathcal{L}_s(w + \hat{\epsilon}(w)) \quad (11)$$

于是, 所提出的 SA-CNN 网络的参数优化问题可以描述为:

$$w_{t+1} = w_t - \alpha \nabla_w \mathcal{L}_s(w + \hat{\epsilon}(w)) \quad (12)$$

其中, α 为学习率。

SC-ANN 网络的整体框架如图 2 所示。首先, 基于源域故障数据进行前向传播得到损失 $\mathcal{L}_s(w)$, 通过反向传播得到梯度 $\nabla_w \mathcal{L}_s(w)$, 利用式(10)计算最优扰动的估计值 $\hat{\epsilon}(w)$ 。然后, 基于扰动值, 对模型参数进行虚拟更新(指更新后的模型只用于计算梯度 $\nabla_w \mathcal{L}_s(w + \hat{\epsilon}(w))$)。最后, 根据式(12)对原模型进行真实更新, 完成模型训练。训练完成的模型将用于诊断符合目标域分布 \mathcal{T} 的故障。

本文所提算法的伪代码如算法 1 所示。

算法 1 锐度感知增强的卷积神经网络算法

输入: 训练集 $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$, 损失函数 $\mathcal{L}: X \times Y \rightarrow \mathbb{R}_+$, 步长 $\eta > 0$, 超参数 $\rho > 0$, 训练 Epoch $T > 0$

输出: 锐度感知增强的卷积神经网络

1. 初始化: 网络参数 $w_0, t = 0$

2. While $t < T$ do

2.1. 采样 $B = \{(x_1^s, y_1^s), \dots, (x_b^s, y_b^s)\}$

2.2. 计算损失 $\mathcal{L}_s(w) = \frac{1}{b} \sum_{i=1}^b \mathcal{L}(y_i^s, \hat{y}_i^s)$

2.3. 计算梯度 $\nabla_w \mathcal{L}_s(w)$

2.4. 计算扰动 $\hat{\epsilon}(w) = \rho \frac{\nabla_w \mathcal{L}_s(w)}{\|\nabla_w \mathcal{L}_s(w)\|_2}$

2.5. 计算 $\nabla_w \mathcal{L}_s^{\text{SAM}}(w) \approx \nabla_w \mathcal{L}_s(w + \hat{\epsilon}(w))$

2.6. 更新参数: $w_{t+1} = w_t - \eta \nabla_w \mathcal{L}_s^{\text{SAM}}(w)$

2.7. $t=t+1$

3. 结束

4. 返回模型参数 w_t

4 模型性能评估实验与分析

4.1 实验数据集

选用凯斯西储大学轴承数据集(CWRU)和东南大学轴承数据集(SEU)对本文方法的有效性进行验证。遵循先前部分研究^[8,17-18]和综述性文献^[7]的数据准备工作,根据轴承负载与转速的不同对工况进行划分并用大写字母表示,如表1和表2所列。本实验中,基于不同的工况构建了14个跨域诊断任务(CWRU数据集12个,SEU数据集2个),箭头“ \rightarrow ”表示领域迁移。例如,“ $A \rightarrow B$ ”代表该跨域诊断任务中A工况下的故障数据为有标签的源域数据,而B工况下的故障数据为无标签的目标域数据。跨域故障诊断算法需要通过源域和目标域中的样本学习域共有知识,最终在目标域上验证诊断的准确率。值得注意的是,作为对比方法之一的骨干网络CNN与本文提出的SA-CNN只利用源域的数据进行学习,这实际上是一个领域泛化问题,难度更高。

表1 凯斯西储大学轴承数据集内容

Table 1 CWRU dataset

工况代号	负载/HP	转速/rpm
A	0	1797
B	1	1772
C	2	1750
D	3	1730

表2 东南大学轴承数据集内容

Table 2 SEU dataset

工况代号	负载/v	转速/Hz
A	0	20
B	2	30

凯斯西储大学轴承数据集^[26]由凯斯西储大学轴承数据中心提供。数据集包括:在4种不同电机工况(转速-负载)下,以12kHz或48kHz采集的正常轴承和单点缺陷损坏轴承的振动信号。实验中,轴承滚动元件、内环和外环上分别引入了故障直径为0.007Inch,0.014Inch和0.021Inch的单点故障,并收集其在不同工况下的运行数据。1797rpm-0HP转速-负载工况下的具体数据如表3所列,其中包括1个健康状态和9个故障状态。

表3 1797rpm-0HP下CWRU数据集部分内容

Table 3 Part of CWRU dataset at 1797rpm-0HP

故障模式	故障直径/Inch	标签 i
健康状态(NA)	0.000	0
内环故障1(IF)	0.007	1
内环故障2(IF)	0.014	2
内环故障3(IF)	0.021	3
滚动元件故障1(BF)	0.007	4
滚动元件故障2(BF)	0.014	5
滚动元件故障3(BF)	0.021	6
外环故障1(OF)	0.007	7
外环故障2(OF)	0.014	8
外环故障3(OF)	0.021	9

东南大学(SEU)轴承数据集^[27]由中国东南大学提供。

实验基于动力传动系统动态模拟器进行,总共收集了8个通道的数据,选取其中两个通道的数据,包含轴承数据集和齿轮数据集两个子数据集,每个子数据集包含2个转速-负载工况(20Hz-0v和30Hz-2v)下1个健康状态和4个故障状态的运行数据。20Hz-0v工况下数据集的内容如表4所列。

表4 20Hz-0v下SEU数据集部分内容

Table 4 Part of SEU dataset at 20Hz-0v

故障模式	标签 i	故障模式/Inch	标签 i
健康齿轮	0	表面故障	5
健康轴承	1	内环故障	6
齿轮剥蚀	2	外环故障	7
齿缺失	3	内环+外环故障	8
齿根故障	4	滚动元件故障	9

4.2 实验环境与参数设置

实验基于Pytorch的框架实现。实验中,输入的振动信号被均匀采样为 1×1024 的形式,选择Adam优化器进行参数优化。为了降低训练中随机性带来的影响,取5次重复实验的均值作为最终结果。

研究选取卷积神经网络作为骨干网络。卷积神经网络是最常见的骨干网络之一,能够提取输入信号中的特征信息,随着网络层数增多,其抽象能力越丰富,但是过拟合风险也越高。因此,在跨域故障诊断任务中,单纯使用卷积神经网络往往无法达到令人满意的性能。为了体现所提方法在跨域场景下进行故障诊断的优越性,本研究选择卷积神经网络作为骨干网络。网络的结构与具体参数如图3所示。骨干网络结合锐度感知最小化方法成为SA-CNN算法。本文选择了5种流行的迁移学习算法与其进行对比,包括:多核最大均值差异法(Multi-Kernal Maximum Mean Discrepancy, MK-MMD)^[28]、联合最大均值差异法(Joint Maximum Mean Discrepancy, JMMD)^[29]、域对抗神经网络(Domain Adversarial Neural Network, DANN)^[30]、相关对齐法(Correlation Alignment, CORAL)^[31]和条件域对抗神经网络(Conditional Adversarial Domain Adaptation Network, CDAN)^[32]。

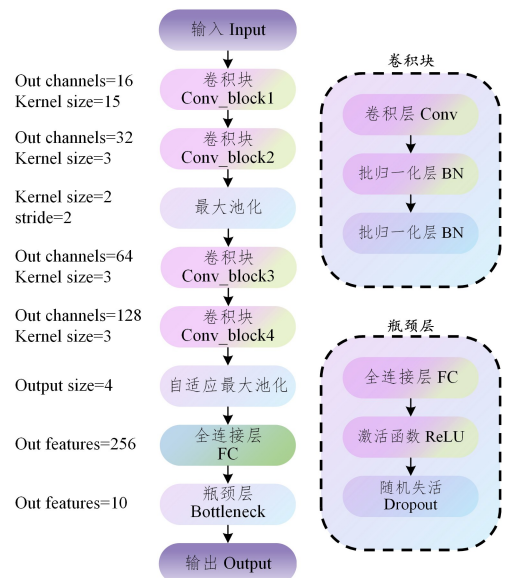


图3 一维卷积神经网络的结构与参数选取

Fig. 3 Structure and parameter selection of 1D-CNN

深度学习算法中的超参数是训练前需要人为设置的参数,不同的超参数设置会对训练结果产生较大的影响。因此在实验部分首先分析算法中涉及的超参数,并通过实验分析其选择范围与适应性,最终确定合适的超参数值。首先,整理实验涉及算法中对实验结果影响较大的超参数,如表5所列。其中,批次大小(Batch Size)、学习率(Learning Rate)和轮次数(Epoch)是各类深度学习算法中常见且重要的超参数,而迁移模块的损失权重是深度迁移学习算法中重要的超参数。因此,分别选取上述参数作为讨论的对象,并基于SEU数据集的两个跨域故障诊断任务分别进行实验与分析。

表5 实验涉及算法中部分重要的超参数

Table 5 Important hyperparameters of algorithms in experiment

	批次大小	学习率	轮次数	损失权重
JMMD	✓	✓	✓	✓
MK-MMD	✓	✓	✓	✓
CDAN	✓	✓	✓	✓
DANN	✓	✓	✓	✓
CORAL	✓	✓	✓	✓
CNN	✓	✓	✓	✓
SA-CNN	✓	✓	✓	✓

为了研究批次大小对本研究涉及算法的影响,依次改变批次的大小设置并进行实验,结果如表6所列。每次实验中准确率最高的结果用加粗突出显示。实验结果显示,适应不同算法的最优批次大小不同,但是整体上随着批次大小的增加,各迁移学习算法的准确率呈现先上升后下降的趋势。这是由于越小的批次中数据噪声的占比越高,网络学习的效果就越差,因此适当增加批次大小可以提升模型的诊断准确率。但是随着批次的增大,模型需要更多的训练时间,同时也会增加过拟合的风险,导致跨域诊断性能下降。由于大部分算法在批次为32~128时达到性能最优,综合考虑训练时间和算法性能,在后续实验中将批次大小设定为64。

表6 不同批次大小下算法诊断准确率的对比

Table 6 Accuracy comparison of algorithm with different batch size settings

		Batch size					
		16	32	64	128	256	512
MK-MMD	A→B	0.3372	0.3578	0.3255	0.3651	0.3211	0.3461
	B→A	0.5352	0.5073	0.4839	0.5411	0.4839	0.4428
JMMD	A→B	0.3402	0.3959	0.3842	0.3534	0.3314	0.3313
	B→A	0.5191	0.4677	0.4882	0.5337	0.5117	0.4472
DANN	A→B	0.3724	0.4018	0.3563	0.3974	0.4011	0.3709
	B→A	0.5264	0.5396	0.4663	0.5073	0.5059	0.4853
CORAL	A→B	0.3167	0.4032	0.3871	0.3504	0.3827	0.3974
	B→A	0.5337	0.5015	0.5308	0.5073	0.5001	0.5161
CDAN	A→B	0.3519	0.3534	0.3812	0.3651	0.3079	0.3802
	B→A	0.4985	0.5308	0.5015	0.5161	0.5303	0.4839
CNN	A→B	0.3709	0.4179	0.4428	0.4408	0.3812	0.3915
	B→A	0.4191	0.4091	0.4824	0.4369	0.4311	0.5234
SA-CNN	A→B	0.4275	0.4277	0.4319	0.4345	0.4399	0.4502
	B→A	0.3987	0.5203	0.5423	0.5539	0.5601	0.5512

为了研究训练轮次对算法的影响并设定合适的训练轮次,依次增加训练轮次进行重复实验,实验结果如表7所列。实验结果与上文针对批次大小进行的相关实验结果类似,各迁移学习算法的准确率随着训练轮次的增加呈现先上升后下降的趋势。这是由于太小的训练轮次使得模型无法很好地对

训练数据进行学习,产生欠拟合。但是随着训练轮次的增加,模型同样承担了更多的训练时间与过拟合的风险,导致诊断性能下降。实验结果显示,在训练轮次达到300时,大部分的算法达到了最优的性能,因此在后续的实验中将训练轮次设定为300。

表7 不同训练轮次下算法诊断准确率的对比

Table 7 Accuracy comparison of algorithm with different epoch settings

		Epoch			
		75	150	300	600
MK-MMD	A→B	0.3106	0.3123	0.4707	0.4604
	B→A	0.4399	0.4223	0.5528	0.5001
JMMD	A→B	0.3182	0.3842	0.4326	0.4252
	B→A	0.4487	0.4296	0.5381	0.5279
DANN	A→B	0.3079	0.3695	0.3931	0.4293
	B→A	0.3754	0.4589	0.5305	0.5267
CORAL	A→B	0.3196	0.2874	0.4736	0.4648
	B→A	0.3094	0.4707	0.4971	0.5543
CDAN	A→B	0.3226	0.3241	0.4956	0.4897
	B→A	0.3901	0.4853	0.5205	0.5204
CNN	A→B	0.2478	0.3255	0.4457	0.4971
	B→A	0.3768	0.4443	0.5117	0.5367
SA-CNN	A→B	0.5011	0.5321	0.5432	0.5431
	B→A	0.4987	0.5012	0.5632	0.5631

为了研究学习率对算法的影响并选取合适的学习率,依次改变算法中的学习率进行实验,实验结果如表8所列。学习率的大小代表了深度学习网络在训练过程中模型参数向梯度方向更新的幅度。当学习率较小时,网络参数的更新速度较慢,有时甚至难以在设定的训练轮次内收敛。随着学习率的增大,网络参数的更新幅度将增加。但是学习率过大,可能导致模型参数在损失函数最小值附近震荡,难以收敛。实验结果显示,当学习率过高或过低(如0.1或0.0001)时,最终模型的诊断性能均较差。综合考虑各算法在不同学习率下的表现,后续实验选取0.001作为学习率。

表8 不同学习率下算法的诊断准确率对比

Table 8 Accuracy comparison of algorithm with different learning rate

		Learning rate			
		0.1	0.01	0.001	0.0001
MK-MMD	A→B	0.3915	0.2683	0.4707	0.2918
	B→A	0.2831	0.5806	0.5528	0.2815
JMMD	A→B	0.2815	0.3959	0.4326	0.2683
	B→A	0.2874	0.4135	0.5381	0.3241
DANN	A→B	0.2831	0.4428	0.3921	0.2874
	B→A	0.4648	0.5147	0.5315	0.2624
CORAL	A→B	0.3489	0.2829	0.4736	0.2991
	B→A	0.3754	0.5235	0.4971	0.2859
CDAN	A→B	0.3504	0.3944	0.4956	0.2922
	B→A	0.3006	0.5204	0.5705	0.2703
CNN	A→B	0.2171	0.3358	0.4457	0.2346
	B→A	0.1965	0.4501	0.5117	0.2889
SA-CNN	A→B	0.3202	0.4632	0.5432	0.5031
	B→A	0.3601	0.5321	0.5632	0.5002

在传统深度迁移学习中,损失函数的形式可以概括为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_c + \lambda \mathcal{L}_{\text{transfer}} \quad (13)$$

其中, \mathcal{L}_c 为交叉熵损失函数, $\mathcal{L}_{\text{transfer}}$ 为迁移模块产生的损失函数; λ 为损失权重,代表迁移损失在总损失中的权重。在不同的迁移学习算法中, $\mathcal{L}_{\text{transfer}}$ 代表了不同的损失函数。例如,其

在基于特征映射的迁移学习方法(JMMD, MK-MMD 和 CORAL)中代表不同程度源域与目标域差异的函数,而在基于生成对抗的迁移学习方法(CDAN 和 DANN)中则代表不同的对抗损失函数。

为了研究迁移损失权重的变化对实验涉及算法的影响,依次改变权值进行实验,结果如表 9 所列。实验结果显示,大部分算法的性能在损失权重为 1 时达到最佳,同时相比其他超参数,损失权重对最终模型诊断性能的影响较小且规律性较差。因此,将损失函数权重设定为 1。

表 9 不同损失权重下算法诊断准确率的对比

Table 9 Accuracy comparison of algorithm with different loss weight

		Tradeoff parameter					
		0.1	0.2	0.4	0.6	0.8	1.0
MK-MMD	A→B	0.3269	0.3826	0.3358	0.3856	0.3519	0.4707
	B→A	0.4575	0.4941	0.5132	0.4619	0.5455	0.5528
JMMD	A→B	0.3636	0.3241	0.3901	0.3358	0.3754	0.4326
	B→A	0.4413	0.5499	0.5044	0.5219	0.5587	0.5381
DANN	A→B	0.4018	0.3871	0.3563	0.3123	0.3827	0.3931
	B→A	0.5264	0.5279	0.4809	0.4633	0.4487	0.5305
CORAL	A→B	0.3827	0.4062	0.4633	0.4194	0.3152	0.4736
	B→A	0.5117	0.4633	0.4604	0.5513	0.5102	0.4971
CDAN	A→B	0.3929	0.3196	0.3299	0.3255	0.3006	0.4956
	B→A	0.4912	0.5307	0.4326	0.5073	0.5001	0.5308

本节说明了实验环境与相关设置,讨论了实验涉及算法中部分超参数的选取范围,并最终通过实验选取了适应性较好的超参数值,以保证后续实验的公平性与可重复性。这些参数分别是:批次大小 64、训练轮次 300、学习率 0.01、损失权重 1。

4.3 与流行深度迁移学习框架的准确率对比

首先,基于 CWRU 数据集进行实验。表 10 中比较了变工况任务场景中各个方法的性能,每个跨域任务中性能最优的结果加粗显示。实验共测试了 7 种不同的方法,包括流行的深度迁移学习方法 MK-MMD, JMMD, DANN, CORAL, CDAN, CNN 和本文提出的 SA-CNN。从实验结果可以看出,SA-CNN 通过锐度感知最小化方法提升了网络的泛化性,在大部分迁移任务中超过了当前流行的基于深度迁移学习的故障诊断方法。例如,DANN 方法在所有任务上取得了均值为 0.975 的准确率,比 SA-CNN 低 0.025。

表 10 基于 CWRU 数据集的模型诊断准确率对比

Table 10 Accuracy comparison of model diagnosis on CWRU dataset

	MK-MMD	JMMD	DANN	CORAL	CDAN	CNN	SA-CNN
A→B	0.996	0.977	0.988	0.981	0.991	0.988	1.000
A→C	0.969	0.946	0.998	0.933	0.983	0.987	1.000
A→D	0.877	0.872	0.991	0.866	0.994	0.926	1.000
B→A	1.000	0.989	0.994	0.992	1.000	0.961	0.999
B→C	1.000	0.998	1.000	0.999	0.981	0.950	0.998
B→D	0.935	0.957	0.914	0.894	0.987	0.872	0.999
C→A	0.973	0.937	0.959	0.932	0.987	0.848	1.000
C→B	0.968	0.988	0.976	0.971	0.982	0.921	1.000
C→D	0.991	0.977	0.984	0.960	0.971	0.924	1.000
D→A	0.905	0.928	0.965	0.810	0.952	0.874	1.000
D→B	0.931	0.934	0.961	0.834	0.901	0.864	1.000
D→C	0.989	0.902	0.974	0.893	0.943	0.901	1.000
AVG	0.961	0.950	0.975	0.922	0.973	0.918	1.000

为了进一步验证算法的有效性,本研究在更具有挑战性

的 SEU 数据集上进行了实验,实验结果如图 4 所示。可以看出,SA-CNN 在两个跨域故障诊断任务中均取得了最优的性能表现。

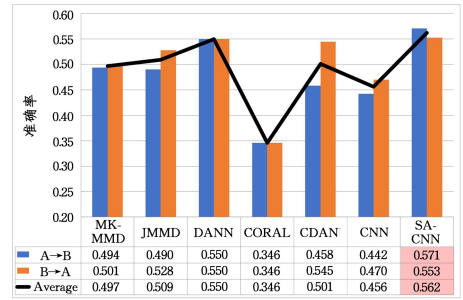


图 4 基于 SEU 数据集的模型诊断准确率对比 (电子版为彩图)

Fig. 4 Accuracy comparison of model diagnosis on SEU dataset

进一步分析实验结果可以发现,在部分跨域诊断任务中,个别流行的深度迁移学习网络在变工况场景下的诊断性能低于不考虑域偏移现象的 CNN。例如 CORAL 方法在两个迁移任务中的诊断精度均低于 CNN,在综述性文献[7]中也能观察到类似的现象。上述实验结果说明如果无法进行良好的参数优化,深度迁移学习模型的泛化性会受到严重影响,导致精度下降。上述实验结果验证了本文方法的重要性。

4.4 超参数分析

本节研究超参数 ρ 对模型性能的影响。图 5 展示了不同超参数设定下所提算法 SA-CNN 在 SEU 数据集上的诊断结果。可以发现,所提方法在较广的超参数选择范围内都能产生比原始 CNN 与大部分其他深度迁移学习算法更好的性能,减少了对该超参数进行微调的需要。这也表明,代表扰动半径大小的超参数 ρ 能够在较大的范围内避免参数落入锐度较高的局部极小值内。

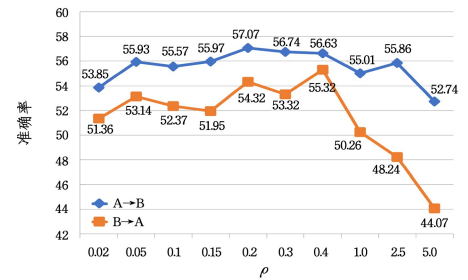


图 5 评估不同的 ρ 值对模型性能的影响

Fig5 Impact of different ρ on model performance

4.5 特征可视化实验

为了进一步研究锐度感知最小化方法对模型泛化性的提升,从凯斯西储大学轴承数据集中选取任务 C→A 进行可视化实验。从目标域 A 中随机选取 200 个故障样本并在迭代过程中提取它们的特征信息,使用 t-SNE^[33] 绘制了特征信息的二维特征表示,如图 6 所示。图中每个点代表一个故障样本,不同颜色的点代表不同的健康状态。

图 6(a)和图 6(b)分别是 JMMD 与 MK-MMD 方法的特征可视化图。JMMD 与 MK-MMD 方法基于最大均值差异(Maximum Mean Discrepancies, MMD)进行跨域数据特征对齐,但是由于目标域故障数据没有标签,需要基于伪标签计算

MMD距离。如果训练过程中网络过于拟合源域数据,可能混淆目标域中部分故障类别,导致伪标签的错误标注,产生负迁移,损失模型的诊断精度。由可视化结果可以发现,对于迁移任务中两组易混淆类别的特征,两个方法均无法实现良好的分类,最终导致性能下降。图 6(c)和图 6(d)分别是 CDAN 与 DANN 方法的特征可视化图。DANN 基于生成对抗网络的框架,利用域鉴别器减少特征提取器产生的特征分布差异,CDAN 在 DANN 的基础上添加额外的条件信息,捕获复杂的多模态结构。由可视化结果可以看出,基于生成对抗框架的方法的表现优于基于 MMD 度量的方法,而由于条件信息的引入,CDAN 实现了比 DANN 更好的泛化性能。但是基于生成对抗的跨域算法优化难度高,很难达到最优性能。图 6(e)是 CORAL 方法的特征可视化图,该方法通过简单地对齐源域与目标域数据特征分布的二阶统计量实现域对齐,但是由于其单纯使用二阶统计量度量差异,因此性能较差。由图中可视化结果可见,该方法在目标域中混淆了 3 个类别的特征。图 6(f)和图 6(g)中分别是 CNN 与 SA-CNN 方法的特征可视化图。CNN 由于只是简单地拟合了源域的数据,因此难以应对跨域故障诊断任务中的域偏移现象。而 SA-CNN 通过锐度感知最小化方法约束网络损失函数地貌的平坦程度,防止网络过度拟合源域,提升模型的泛化性能,从而增强模型在目标域上的表现。观察可视化结果可以发现,相较于原网络与其他方法,图 6(g)中具有相同颜色的点实现了紧凑的聚集,

而不同颜色的点实现了清晰且明确的分离。

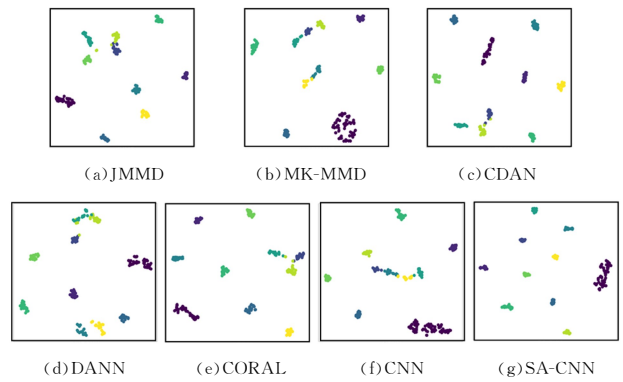


图 6 实验涉及方法的特征可视化图

Fig6 Feature visualization of the methods used in experiment

综上,从可视化结果可以看出,SA-CNN 方法的跨域诊断性能优于其他流行的深度迁移学习方法。基于源域数据,对模型损失函数锐度的优化,使模型学习到了具有良好泛化性能的网络参数,能够实现对目标域数据的良好分类。

4.6 损失函数地貌可视化实验

为了进一步验证传统深度迁移学习算法中知识迁移模块对损失函数地貌的影响以及 SA-CNN 算法对损失函数地貌锐度的优化,基于 CWRU 数据集进行了算法损失函数地貌的可视化实验^[34]。MK-MM, CNN 与 SA-CNN 算法在任务 A→D 中的损失函数地貌等高线图如图 7 所示。

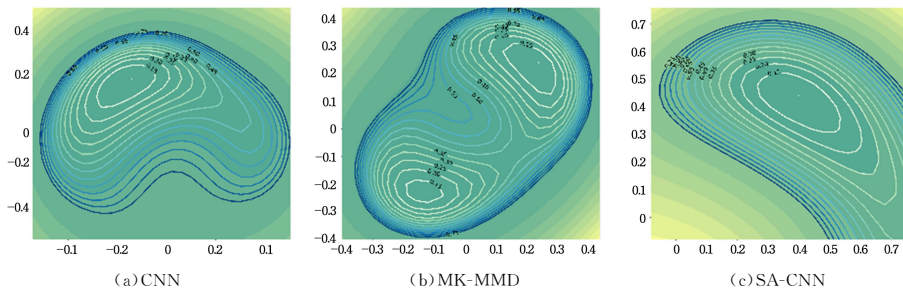


图 7 实验涉及部分方法的训练损失函数地貌可视化图

Fig. 7 Loss function landscape visualization of methods use in in experiment

可以发现,图 7(a)中 CNN 算法的函数地貌较为简单,但是局部极小值处的上方等高线分布密集,地貌较为陡峭;图 7(b)中增加了基于多核最大均值差异的知识迁移模块后, MK-MMD 算法的损失函数地貌明显复杂化,在较小的模型参数摄动范围内具有两个局部极小值,并且其外围的等高线较为密集,地貌整体上较为陡峭;图 7(c)中 SA-CNN 算法得到了最为简单且平滑的损失函数地貌,等高线的分布较为规律且稀疏。结合表 10 中任务 A→D 的诊断结果可以发现,由于损失函数地貌的复杂与高锐度的特征, MK-MMD 方法表现出了低于 CNN 方法的诊断精度,而 SA-CNN 通过对损失函数锐度的优化获得了最高的诊断精度。

以上对损失函数地貌的可视化分析,验证了现有深度迁移学习方法加剧损失函数地貌复杂性与锐度的问题,同时突出了所提算法对损失函数地貌锐度的优化效果。

4.7 计算成本与适用性分析

为了对所提算法的时间复杂性和成本进行分析,本文对

实验中所涉及算法在训练过程中所需要的时间成本进行了统计,并以骨干网络 CNN 为基准进行了归一化处理,如表 11 所列。

表 11 基于 CWRU 数据集任务 A→B 的训练时间成本分析
Table 11 Analysis of training time of task A→B on CWRU dataset

算法	训练时间*
CNN	×1
MK-MMD	×1.5
JMMD	×1.6
DANN	×1.65
CORAL	×1
CDAN	×1.64
SA-CNN	×1.78

注: * 训练时间以 CNN 的训练时间为基准进行归一化处理。

可以发现,由于锐度感知最小化算法在每次训练过程中需要额外进行一次梯度计算,因此所提算法在训练的过程中需要更多的时间成本。但是,传统深度迁移学习算法中 JMMD 和 MK-MMD 算法中需要额外的计算量对最大均值差异

进行计算,生成对抗网络如 CDAN 和 DANN 需要通过特征提取器和领域判别器的对抗训练学习领域不变特征,这些都增加了跨域故障诊断算法训练过程中的复杂性和时间成本。因此,相较于部分传统深度迁移学习算法,本文所提方法虽然在训练过程中增加了较少的额外计算量,但是在跨域机械故障诊断任务中获得了稳定且显著的性能提升。

在机械故障诊断领域中,训练数据与实际应用时的输入数据总是存在或大或小的分布差异,但是在部分工作场景中可以忽略该差异的影响,并做出数据独立同分布的假设检验。本文所提方法通过对锐度的感知来增强算法的泛化性能,弥补了分布差异导致的网络泛化性能下降。因此,不论数据分布差异大或小,所提方法均能在一定程度上提升网络的泛化性能,进而提高诊断精度。但是由于锐度感知最小化算法需要额外的时间成本,因此需要综合考虑训练成本、模型泛化性和诊断精度要求决定是否使用本文所提方法。

结束语 针对现有基于深度迁移模型的变工况机械故障诊断技术存在的无法感知损失函数地貌锐度、模型泛化性差、诊断精度低的问题,本研究提出锐度感知增强的卷积神经网络架构。其中,为解决深度迁移模型结构复杂、参数量庞大导致模型参数陷入泛化性差的局部极值参数空间的问题,本研究引入锐度感知最小化算法,利用损失函数的锐度感知模型的泛化性能并进行优化,进而有效提升模型性能。在实验环节中,所提方法与 5 个被广泛采用的深度迁移学习方法在凯斯西储大学轴承数据集和东南大学轴承数据集上进行对比,实验结果有效论证了所提算法相较于对比迁移算法的优势,其在变工况机械故障诊断任务中呈现出较高的泛化性能。后期研究将尝试结合锐度感知与基于元学习的跨域故障诊断算法,解决元学习框架普遍存在的复杂双层网络结构导致的优化难度高的问题,有效助力变工况环境下故障的理解与诊断。

参 考 文 献

- [1] ZHANG Y P, LIU B G, LIU G, et al. Equipment fault diagnosis technology based on data-driven [J]. *Mechanical Engineering and Automation*, 2022(2): 130-132.
- [2] JIAO J, ZHAO M, LIN J, et al. Deep Coupled Dense Convolutional Network with Complementary Data for Intelligent Fault Diagnosis [J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(12): 9858-9867.
- [3] XIAO Q H. Survey of mechanical fault diagnosis methods based on machine learning theory [J]. *Modern manufacturing engineering*, 2021(7): 148-161.
- [4] XU P. Research on bearing fault diagnosis method based on neural network [D]. Yancheng Institute of Technology, 2023.
- [5] SUN S. Research on real-time fault diagnosis of motor Bearing based on Deep Learning [D]. Daqing: Northeast Petroleum University, 2023.
- [6] MA B T. Abnormal detection of industrial robot joint health state based on current signal [D]. Shanghai: Donghua University, 2023.
- [7] ZHAO Z, ZHANG Q, YU X, et al. Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-28.
- [8] LYU C H, CHENG J J, HU Y G, et al. Servo Online Fault Diagnosis Based on Multi-source Domain Deep Transfer Learning [J]. *Journal of Ordnance and Equipment Engineering*, 2022, 43(9): 60-67.
- [9] XU S M, LUO L K, HU J L, et al. Semantic driven attention network with attribute learning for unsupervised person re-identification [J]. *Knowledge-Based Systems*, 2022, 252(27): 10935. 1-10935. 13.
- [10] LUO L K, CHEN L, HU S, et al. Discriminative and Geometry-Aware Unsupervised Domain Adaptation [J]. *IEEE Transactions on Cybernetics*, 2020, 50(9): 3914-3927.
- [11] LUO L K. Transfer Learning and Interactive Image Segmentation [D]. Shanghai: Shanghai Jiao Tong University, 2019.
- [12] TAN M, QUOC V L. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [C]// *Proceedings of International Conference on Machine Learning*. PMLR, 2019: 6105-6114.
- [13] ZHANG C, BENGIO S, HARDT M R, et al. Understanding deep learning requires rethinking generalization [J]. *Association for Computing Machinery*, 2021, 64(3): 107-115.
- [14] KESKAR N S, MUDIGERE D, NOCEDAL J, et al. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima [C]// *Proceedings of International Conference on Learning Representations*. Palais des Congrès Neptune: ICLR, 2017: 1-16.
- [15] FORET P, KLEINER A, MOBAHI H, et al. Sharpness-aware minimization for efficiently improving generalization [C]// *Proceedings of International Conference on Learning Representations*. 2021: 1-18.
- [16] WANG J. Research on Text Representation and Classification with Deep Learning [D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [17] KAVIANPOUR M, GHORVEI M, KAVIANPOUR P, et al. An Intelligent Gearbox Fault Diagnosis under Different Operating Conditions using Adversarial Domain Adaptation [C]// *Proceedings of International Conference on Control, Instrumentation and Automation*. Tehran: ICCIA, 2022: 1-6.
- [18] WANG Q, TAAL C, FINK O. Integrating Expert Knowledge with Domain Adaptation for Unsupervised Fault Diagnosis [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-12.
- [19] NESTEROV Y E. A method for solving the convex programming problem with convergence rate $O(1/k^2)$ [J]. *Dokl. Akad. Nauk SSSR*, 1983, 269: 543-547.
- [20] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization [C]// *Proceedings of 3rd International Conference on Learning Representations*. 2015: 1-15.
- [21] DZIUGAITE G, ROY D M. Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data [C]// *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*. 2017: 1-14.

- [22] NEYSHABUR B, MCALLESTER D, SREBRO N, et al. Exploring generalization in deep learning[C]// Proceedings of 31st Conference on Neural Information Processing Systems, 2019: 5947-5956.
- [23] CHEN X, HSIEH C, GONG B. When vision transformers outperform resnets without pretraining or strong data augmentations[C]// Proceedings of International Conference on Learning Representations, 2022: 1-20.
- [24] ABBA S, MOMI N, QUAN X, et al. Sharp-MAML: Sharpness-Aware Model-Agnostic Meta Learning[C]// Proceedings of International Conference on Machine Learning, 2022: 10-32.
- [25] BAHRI D, MOBAHI H, YI T. Sharpness-aware minimization improves language model generalization[C]// Annual Meeting of the Association for Computational Linguistics, 2022: 1-12.
- [26] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study [J]. Mechanical Systems and Signal Processing, 2015, 64: 100-131.
- [27] SHAO S Y, STEPHEN M M. Highly accurate machine fault diagnosis using deep transfer learning [J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2446-2455.
- [28] ZHANG Y, REN Z, ZHOU S. A new deep convolutional domain adaptation network for bearing fault diagnosis under different working conditions [J]. Shock and Vibration, 2020, 2020: 1-14.
- [29] JIAO J, ZHAO M, LIN J, et al. Residual joint adaptation adversarial network for intelligent transfer fault diagnosis [J]. Mechanical Systems and Signal Processing, 2020, 145: 106962.
- [30] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030.
- [31] SUN B, SAENKO K. Deep coral: Correlation alignment for deep domain adaptation[C]// Proceedings of European Conference on Computer Vision, Switzerland: Springer International Publishing, 2016: 443-445.
- [32] LONG M, CAO Z, WANG J, et al. Conditional adversarial domain adaptation[C]// Proceedings of Advances in Neural Information Processing. Burlington: Morgan Kaufmann, 2018: 1640-1650.
- [33] RAUBER P E, FALCAO A X, TELEA A C. Visualizing time-dependent data using dynamic t-SNE[C]// Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers, 2016: 1-5.
- [34] LI H, XU Z, TAYLOR G, et al. Visualizing the loss landscape of neural nets[C]// Advances in Neural Information Processing Systems. Burlington: Morgan Kaufmann, 2018.



FAN Jiayuan, born in 1998, postgraduate. His main research interests include deep learning and mechanical fault diagnosis.



HU Shiqiang, born in 1969, Ph.D, professor, Ph.D supervisor. His main research interests include pattern recognition and application and optimization of machine learning on aviation tasks.

(责任编辑:柯颖)