智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

不平衡小样本基于局部域对抗适应网络的发动机振动预测模型

季友昌, 袁伟伟, 毛善斌, 任春红, 关东海

引用本文:
季友昌,袁伟伟,毛善斌,任春红,关东海. 不平衡小样本基于局部域对抗适应网络的发动机振动预测模型[J]. 智能系统学报, 2023, 18(5): 1005–1016.
JI Youchang, YUAN Weiwei, MAO Shanbin, et al. Partial domain adversarial adaptation networks for imbalanced small samples in aeroengine vibration prediction[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(5): 1005–1016.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202210030

DOI: 10.11992/tis.202210030
网络出版地址: https://kns.cnki.net/kcms2/detail/23.1538.TP.20230608.1024.002.html

不平衡小样本基于局部域对抗适应网络的发动机振动预测模型

季友昌[1], 袁伟伟[1], 毛善斌[2], 任春红[2], 关东海[1]
(1. 南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106; 2. 北京动力机械研究所, 北京 100074)

摘　要: 在发动机振动预测中, 实际装配数据样本量小且类别不平衡, 难以直接建立有效的预测模型。迁移学习方法能够通过迁移源域知识来提高目标域模型性能, 为此, 本文提出了基于局部域对抗适应网络的发动机振动预测模型。将领域按标签分为多个局部域, 通过多个局部域对抗适应网络将目标域样本映射到源域, 保证小样本中的少数类得到正确迁移。并通过伪标签来解决目标样本的域转换, 使用源域分类器给出可靠的预测结果。本文在多个真实数据集上验证了所提方法的有效性和泛化性, 与其他方法相比, 振动预测准确率能够平均提升 15% 左右。

关键词: 不平衡数据; 域适应; 对抗学习; 振动预测; 异构迁移学习; 目标域; 小样本学习; 特征空间
中图分类号: TP391.41; TP18　　**文献标志码**: A　　**文章编号**: 1673−4785(2023)05−1005−12

中文引用格式: 季友昌, 袁伟伟, 毛善斌, 等. 不平衡小样本基于局部域对抗适应网络的发动机振动预测模型 [J]. 智能系统学报, 2023, 18(5): 1005–1016.
英文引用格式: JI Youchang, YUAN Weiwei, MAO Shanbin, et al. Partial domain adversarial adaptation networks for imbalanced small samples in aeroengine vibration prediction[J]. CAAI transactions on intelligent systems, 2023, 18(5): 1005–1016.

Partial domain adversarial adaptation networks for imbalanced small samples in aeroengine vibration prediction

JI Youchang[1], YUAN Weiwei[1], MAO Shanbin[2], REN Chunhong[2], GUAN Donghai[1]
(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; 2. Beijing Power Machinery Institute, Beijing 100074, China)

Abstract: In aeroengine vibration prediction, due to the imbalanced small samples of the actual assembly data, directly establishing an effective prediction model is difficult. Transfer learning can improve model performance in the target domain by transferring the knowledge of the source domain. Therefore, this study proposes an aeroengine vibration prediction model based on a partial domain adversarial adaptation network. Each domain is divided into multiple local domains according to labels. Through multiple local domain adversarial adaptation networks, the samples in the target domain can be mapped onto the source domain so that the minority class samples can be transferred correctly. The pseudo label is used to solve the domain transformation of the samples in the target domain, and the classifier of the source domain is used to provide a reliable prediction result. In this study, the validity and generalization of the proposed method are verified on several real datasets. Compared with other methods, the area under the curve and F1 of the vibration prediction model can be improved by approximately 15% on average.

Keywords: imbalanced data; domain adaptation; adversarial learning; vibration prediction; heterogeneous transfer learning; target domain; few-shot learning; feature space

收稿日期: 2022−10−24.　**网络出版日期**: 2023−06−09.
基金项目: 基础科研项目(JCKY2020204C009).
通信作者: 关东海. E-mail: dhguan@nuaa.edu.cn.

在发动机制造中, 发动机装配完成后需要进行试车测试, 若出现振动超差, 则需要拆解发动

机，更换零部件或者调整装配操作，再重新装配测试。而影响发动机振动水平的因素众多，依靠经验进行调试需要花费大量人力，严重影响发动机的生产进度。随着人工智能技术的发展，研究人员尝试开发智能算法对发动机振动水平进行预测，期望算法能够对装配过程给予指导。但由于复杂的装配操作和高昂的数据获取成本，能采集到的数据量小，并且振动合格的样本数量远超振动超差的样本，因此该问题属于不平衡小样本预测问题。

现有的发动机振动预测方法可分为：基于传统机器学习算法[1]，以及基于复杂神经网络的深度学习算法，例如借助 LSTM（long short-term memory）和 RNN（recurrent neural network）来预测涡轮发动机振动水平[2]。但由于实际应用场景中训练样本的不平衡且数量少的特点，直接使用传统机器学习或深度学习算法容易产生标签偏差和过拟合等问题。

迁移学习技术是解决小样本预测问题的关键技术之一。它能够利用源域数据中的知识来提高模型在目标域的性能，减少目标域对样本的依赖[3]。为解决目标域和源域数据分布上的差异，提出了如基于子结构迁移的跨域行为识别框架[4]、基于类质心匹配与局部流形学习的域自适应方法[5]、基于鲁棒专家模型的连续性领域自适应[6] 等方法。但由于不同型号发动机的装配参数不同，即源域和目标域的特征空间不同，在进行域自适应前需要先对齐特征空间，因此该问题属于异构迁移学习问题。

异构迁移学习的主要思路是通过将源域和目标域映射同一个特征空间，在该空间两域数据分布接近，从而实现异构数据的迁移[7]。随着深度学习蓬勃发展，越来越多的学者通过复用深度网络模型实现领域迁移[8-11]，而自从生成对抗网络[12]的提出，其对抗的思想也被运用到迁移学习中，衍生了许多对抗迁移学习方法[13-16]。然而，由于采集到的实际装配数据呈现出样本量小且类别不平衡的特点，现有的异构迁移学习方法很难训练一个具有标签自然不平衡的域不变特征的分类器，少数类样本被错误迁移，导致最终的分类模型可能也会出现标签偏差问题。此外，由于样本量小，采用层数过多、复杂的神经网络会出现过拟合问题。

为解决现有工作的问题，本文提出了基于局部域对抗适应网络的发动机振动预测模型（engine vibration prediction model based on partial domain adversarial adaptation network，EVP-PDAA）。EVP-PDAA 将领域按标签分为多个局部域，建立多个局部域对抗适应网络将目标域样本映射到相应标签的局部源域，保证少数类样本也能得到合理的迁移。由于样本量小，为保证局部域对抗适应网络训练的稳定性和域转换的正确性，将推土机距离（earth mover distance，EMD）作为网络的优化目标，并在网络参数更新时使用梯度惩罚策略。进行振动预测时，利用伪标签来解决待预测目标样本的局部域对抗适应网络选择问题，使用源域分类器给出可靠的预测结果，矫正伪标签可能出现的错误。实验结果表明，本文所提出的方法在面对不平衡小样本时的表现优于其他迁移学习方法，实现了迁移其他型号发动机的数据来提高目标发动机的振动预测效果。

1　相关工作

本研究采用基于对抗思想的异构迁移学习方法迁移源域知识。因此在相关工作中，分别对异构迁移学习方法、深度神经网络迁移方法和深度对抗网络迁移方法进行介绍。

异构迁移学习方法可以被主要分为两类：基于对称特征变换的方法和基于非对称特征变换的方法。基于对称特征变换的方法，即将源域和目标域转换到一个公共子空间，在这个空间里，源域和目标域的数据分布较之前更接近。比如，Duan 等[17] 提出了异构特征增强方法（heterogeneous feature augmentation，HFA），该方法使用两个变换矩阵将源域和目标域映射到公共子空间，并将两个变换矩阵合并，以 SVM 的结构风险函数最小化对合并后矩阵进行优化求解。基于非对称特征变换的方法，即将源域特征空间转换到目标域特征空间或将目标域特征空间转换到源域特征空间。Sukhija 等[18] 提出了基于随机森林的有监督异构领域自适应（supervised heterogeneous domain adaptation via random forests，SHDA-RF）。SHDA-RF 以目标域和源域的共享标签分布作为特征变换的核心，通过随机森林来定义共享标签分布和特征之间的关系，从而得到源域和目标域特征空间之间的关系。Feuz 等[19] 提出了一个特征空间重映射（feature-space remapping，FSR）方法。FSR 定义了目标域和源域的元特征，并以此构建了目标域和源域特征的相似度矩阵，最后通过特征映射关系将目标域样本映射到源域样本空间。

随着深度学习的发展，越来越多的学者将深度网络应用于迁移学习。Ferhat 等[20] 通过集成预

训练的 Transformer 模型来检测恶意软件。Zhang 等[21] 通过预训练和微调来进行小样本意图检测。

自从 Goodfellow 提出生成对抗网络[12] 以来，有诸多学者尝试将网络对抗的思想应用到迁移学习中，并提出了各种深度对抗网络方法。例如 Hong 等提出了一个基于交叉模态肝分割、联合对抗学习和自学习的域自适应框架[22]。

但现有研究在处理不平衡小样本的分类预测问题时，少数类样本在迁移过程中会出现错误迁移的情况，训练得到的分类模型会出现如图 1 所示的标签偏差现象。采用层数过多、复杂的神经网络会出现过拟合问题。

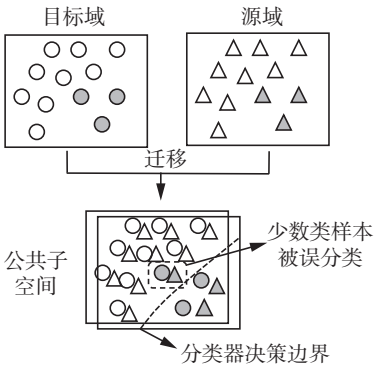


图 1　标签偏差现象

Fig. 1　Label bias phenomenon

2　问题定义

发动机振动预测，即根据装配参数，预测装配后的振动水平是否合格。实际采集到的数据样本总量不超过 150，振动超差样本比例不超过 20%，难以直接使用机器学习或深度学习技术建立有效的预测模型，为此，将迁移其他型号发动机的知识来辅助目标发动机的振动预测。为将问题形式化，本文作出以下几个定义。

定义 1　源域：源域$\mathcal{D}_s$由特征集$X_s=(x_s^1,x_s^2,\cdots,x_s^N)$和标签集$Y_s=(y_s^1,y_s^2,\cdots,y_s^N)$组成，$(x_s^i,y_s^i)$代表$\mathcal{D}_s$的第 i 条样本，N 为源域样本数量。

定义 2　目标域：目标域$\mathcal{D}_t$由特征集$X_t=(x_t^1,x_t^2,\cdots,x_t^N)$和标签集$Y_t=(y_t^1,y_t^2,\cdots,y_t^M)$组成，$(x_t^i,y_t^i)$代表$\mathcal{D}_t$的第 i 条样本，M 为目标域样本数量。

定义 3　局部源域：局部源域$\mathcal{D}_s^\gamma$为标签为γ的源域样本所组成的集合：$\mathcal{D}_s^\gamma=\{x_s^i\,|\,y_s^i=\gamma,i=1,2,\cdots,N\}$。

定义 4　局部目标域：局部目标域$\mathcal{D}_t^\gamma$为标签为γ的目标域样本所组成的集合：$\mathcal{D}_t^\gamma=\{x_t^i\,|\,y_t^i=\gamma,i=1,2,\cdots,M\}$。

定义 5　生成器集合：生成器集合$G_{\text{en}}=\{G_i|i=0,1,\cdots,L-1\}$，其中$G_i$为将$\mathcal{D}_t^i$映射到$\mathcal{D}_s^i$的局部域对抗适应网络中的生成器。

3　基于局部域对抗适应网络的发动机振动预测模型

本文针对发动机振动预测中不平衡小样本的问题，提出了基于局部域对抗适应网络的发动机振动预测模型。该方法的核心思想是将领域按标签分为多个局部域，建立多个局部域对抗适应网络将各个局部目标域样本映射到相应的局部源域，振动预测时使用伪标签解决映射函数选择问题，并使用源域分类器给出预测的可靠结果，纠正伪标签可能出现的错误。EVP-PDAA 包含两个阶段：一是局部域对抗适应网络的建立；二是振动预测机制。基于局部域对抗适应网络的发动机振动预测模型的框架如图 2 所示。

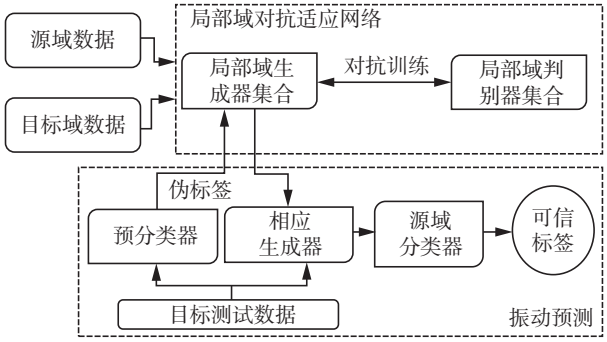


图 2　EVP-PDAA 的框架

Fig. 2　Structure of EVP-PDAA

3.1　局部域对抗适应网络

局部域对抗适应网络由一个生成器 G 和一个判别器 D 组成。生成器学习局部目标域到局部源域样本空间的映射关系，以局部目标域样本为输入，输出为转换到局部源域样本空间的样本。判别器学习判断数据是转换后的目标域数据还是源域数据，以局部源域样本和生成器的输出为输入，输出为输入样本与局部源域的接近程度。通过生成器和判别器的对抗训练，使生成器能够准确地将局部目标域样本映射到相应局部源域。

若使用二元交叉熵作为损失函数，生成器和判别器的损失函数分别为

$$L_D=-\mathrm{E}_{x\sim P_r}[\log(D(x))]-\mathrm{E}_{x\sim P_g}[\log(1-D(x))] \tag{1}$$

$$L_G=\mathrm{E}_{x\sim P_g}[\log(1-D(x))] \tag{2}$$

式中：E为期望函数；P_r和P_g分别是局部源域样本所服从的分布和由生成器转换后的局部目标域样本所服从的分布；$G(\cdot)$和$D(\cdot)$分别是生成器和判别器网络的可微分函数。

但以二元交叉熵作为损失函数可能会出现判别器训练得越好，最小化式 (2) 就会越近似于最小化P_r和P_g的 JS 散度，但若P_r和P_g没有重叠或重

叠部分可以忽略时(可能性很大), P_r和P_g的 JS 散度越接近于一固定常数 lg2, 进而面临梯度消失问题。此外, 还有可能导致生成器生成样本多样性不够等问题[23]。

即使P_r和P_g没有重叠, EMD 仍能反映它们的远近, 从而提供有意义的梯度。因此, 将 EMD 定义为生成器的损失函数, 可以有效地将生成器生成的样本分布向局部源域样本分布靠拢。EMD 的定义为[23]。

$$\mathcal{D}(\rho_1,\rho_2)=\inf_{\gamma\in\prod(\rho_1,\rho_2)}E_{(x,y)\sim\gamma}[\|x-y\|] \tag{3}$$

其中$\prod(\rho_1,\rho_2)$为分布ρ_1和分布ρ_2所有可能的联合分布所组成的集合。

虽然 EMD 定义中的$\inf_{\gamma\in\prod(P_1,P_2)}$无法直接求解, 但当函数$f$满足 Lipschitz 连续, 即满足$|f(x_1)-f(x_2)|\leqslant K|x_1-x_2|(K\epsilon\mathrm{R})$时, EMD 可表示为[23]。

$$K\cdot\mathcal{D}(\rho_1,\rho_2)=\sup_{\|f\|_L\leqslant K}\mathrm{E}_{x\sim\rho_1}[f(x)]-E_{x\sim\rho_2}[f(x)] \tag{4}$$

使用带参数ω的神经网络来定义一系列可能的函数f_ω, 式 (4) 就可以近似转化为

$$K\cdot\mathcal{D}(\rho_1,\rho_2)\approx\max_{\omega}\{\mathrm{E}_{x\sim\rho_1}[f_\omega(x)]-\mathrm{E}_{x\sim\rho_2}[f_\omega(x)]\} \tag{5}$$

为使f_ω满足 Lipschitz 条件, 可以采用权重裁剪策略, 即限制参数ω的变化范围不超过某个特定范围$[-c,c]$, c为固定常数。但权重裁剪策略可能会导致判别器学习成为一种简单的函数映射, 或出现梯度消失或爆炸等问题[23]。

因此, 本文使用梯度惩罚策略[24], 加入一个正则项, 将梯度的 L2 范数约束在 1 附近, 使判别器的参数不超过某个常数, 满足 Lipschitz 条件。此时, 局部域对抗适应网络的目标函数为

$$\max\left\{-\mathrm{E}_{\tilde{x}\sim P_g}[D(\tilde{x})]+\mathrm{E}_{x\sim P_r}[D(x)]+\lambda\cdot\mathrm{E}_{\hat{x}}\left[(\|\nabla_{\hat{x}}D(\hat{x})\|_2-1)^2\right]\right\} \tag{6}$$

其中随机样本$\hat{x}$的计算方法为

$$\hat{x}=\varepsilon x_s+(1-\varepsilon)G(x_t) \tag{7}$$

其中ε为 0 到 1 之间的随机数。

综上所述, 生成器和判别器的损失函数分别为

$$L_G=-\mathrm{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] \tag{8}$$

$$L_D=\mathrm{E}_{\tilde{x}\sim P_g}[D(\tilde{x})]-\mathrm{E}_{x\sim P_r}[D(x)]+\lambda\cdot\mathrm{E}_{\hat{x}}\left[(\|\nabla_{\hat{x}}D(\hat{x})\|_2-1)^2\right] \tag{9}$$

生成器和判别器均由输入层、隐藏层和输出层组成。由于数据量小, 为避免出现过拟合的情况, 隐藏层的网络不宜设计得过于复杂。此外, 由于 batch normalization(BN)是对一个批次的样本进行归一化, 在判别器网络中加入 BN 层会使得每个样本的梯度计算出错, 因此在判别器中不加入 BN 层。局部域对抗适应网络的网络结构如图 3 所示, 图中生成器和判别器的隐层层数为 1。

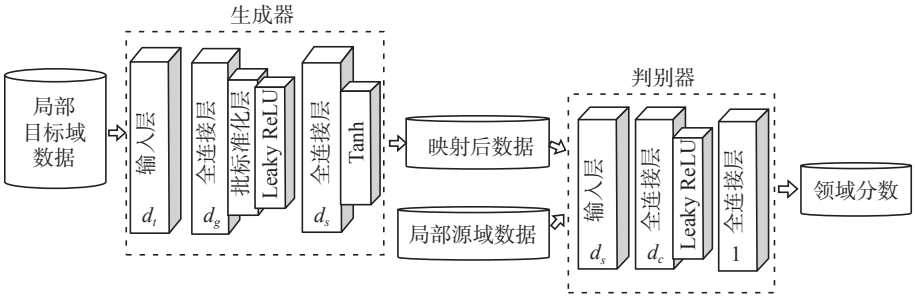


图 3　局部域对抗适应网络的结构

Fig. 3　Structure of PDAA

局部域对抗适应网络的训练伪代码如算法 1 所示。

算法 1　局部域对抗适应网络训练伪代码

输入　局部目标域训练数据X_t, 对应的局部源域数据X_s, 梯度惩罚项系数λ, 判别器和生成器训练次数比例n_d, 训练轮次 T

输出　生成器

1) For $t=1,2,\cdots,T$:

2) 随机从X_t中选择一个样本: x_t

3) 随机从X_s中选择一个样本: x_s

4) For $i=1,2,\cdots,n_d$:

5) 根据式 (5), 生成随机样本

6) 根据式 (7), 计算判别器损失

7) 使用 Adam 优化算法更新判别器参数

8) End For

9) 根据式 (8), 计算生成器损失

10) 使用 Adam 优化算法更新生成器参数

11) End For

3.2　振动预测机制

在 EVP-PDAA 中存在 L 个局部域对抗适应网络, 使用网络中的生成器可以将目标域中各类样本映射到对应局部源域。但当面对没有标签的目标样本时, 很难决定使用哪一个生成器进行领域转换, 无法对振动水平进行预测。为此, EVP-PDAA

设计了一个基于伪标签的振动预测机制。

EVP-PDAA 在有限的目标域训练样本上建立一个预分类器, 预分类器对待预测的目标样本x_t进行预测并给出标签, 将该标签记为伪标签y_p, 伪标签的值域同为目标域标签空间。与最终标签相比, 伪标签y_p的可靠性不足, 需要借助后续机制进行验证或进行修改。

随后, 根据伪标签在生成器集合 Generators 中选择相应生成器G_{y_p}, 利用G_{y_p}即可将目标样本转换到源域样本空间。得到转换到源域样本空间下的目标样本$G_{y_p}(x_t)$后, 利用在源域样本空间中训练的源域分类器, 对转换后的样本再次进行预测, 给出预测的最终结果。当预分类器出现误分类, 导致后续选择了错误的生成器G_{y_p}进行域转换, 但由于G_{y_p}学习的是$\mathcal{D}_t^{y_p}$到$\mathcal{D}_s^{y_p}$的映射函数, 不属于$\mathcal{D}_t^{y_p}$的目标样本经过转换后较$\mathcal{D}_s^{y_p}$必然存在一定差异, 而当差异超过了源域分类器对$\mathcal{D}_s^{y_p}$的决策边界且位于真实标签的样本空间时, 源域分类器就能够矫正预分类器的错误, 给出正确的标签。

EVP-PDAA 的振动预测机制伪代码如算法 2 所示。

算法 2　振动预测机制伪代码

输入　待分类的目标样本D_{test}, 目标域数据$\mathcal{D}_t$, 源域数据$\mathcal{D}_s$, 生成器集合 Generators

输出　目标样本的标签Y_{test}

1) 在 Generators 中选择相应的生成器, 将$\mathcal{D}_t$转换到源域样本空间, 得到转换后的数据$\mathcal{D}_{\text{t_trans}}$

2) 使用$\mathcal{D}_{\text{t_trans}}$和$\mathcal{D}_s$训练源域分类器

3) 使用$\mathcal{D}_t$训练预分类器

4) 预分类器对D_{test}进行预测, 给出伪标签

5) 根据伪标签, 在 Generators 中选择相应的生成器对D_{test}进行域转换, 得到$D_{\text{test_trans}}$

6) 源域分类器对$D_{\text{test_trans}}$进行预测, 给出预测的最终结果Y_{test}

4　实验结果及分析

4.1　实验设置

本文在 3 个源域数据集 DR、EP-1、EP-2 和 3 个目标域数据集 SR-1、SR-2、SR-3 上进行了实验。单轴发动机的实际装配数据中包含 248 个装配参数, 使用不同的特征选择方法进行 3 轮特征选择, 分别筛选出 9、12、23 个关键装配参数, 即 SR-1、SR-2 和 SR-3。DR 是从双轴发动机的实际装配过程中采集而来, 双轴发动机具有与单轴发动机振动相关的共性关键特征。EP-1 和 EP-2 是从单轴发动机的实验平台采集而来, 该实验平台是对单轴发动机的简化模拟, 源域发动机和目标域发动机的装配工艺对发动机振动水平的影响具有一定共性关系, 但很难通过机理分析得出关系的具体表现形式。各数据集的统计信息如表 1 所示。

表 1　数据集的统计信息

Table 1　Statistic results of datasets

数据集	特征数量	样本总量	正样本数量	负样本数量	正样本比例/%
SR-1	9	85	15	70	17.64
SR-2	12	85	15	70	17.64
SR-3	23	85	15	70	17.64
DR	9	131	16	115	12.21
EP-1	5	320	124	196	38.75
EP-2	8	640	212	428	33.13

本文选取了 5 种对比方法, 分别是: 1) RF-T: 不使用迁移学习方法, 直接使用目标域数据建立分类器; 2) TCA(transfer component analysis)[25]: 通过最小化源域和目标域边缘概率分布的距离解决两域数据分布差异; 3) CORAL(correlation alignment)[25]: 通过对齐源域和目标域协方差解决数据分布差异; 4) FSR; 5) SHDA-RF。由于 CORAL 和 TCA 要求目标域和源域的特征空间相同, 因此在迁移之前, 先使用 UMR[26] 来统一目标域和源域的特征空间。所有方法中的分类器均采用随机森林(random forests, RF)算法。

本文使用 AUC 和 F_1 来评价各方法所建立的振动预测模型性能, 其中 F_1 用于衡量模型对少数类的预测性能, AUC 用于评价模型的整体性能。为避免随机因子对实验造成影响, 每组实验都重复 20 次, 使用 Wilcoxon 符号秩检验[27] 判断两个方法的实验结果是否具有统计意义上的不同, 置信水平设为 95%。此外, 本文使用 Cohen's d 效应量来量化两个方法的差异, Cohen's d 值和效应等级的对应关系如表 2 所示[22]。

表 2　Cohen's d 效应量等级

Table 2　The effectiveness levels of Cohen's d

Cohen's d	[0, 0.2)	[0.2, 0.5)	[0.5, 0.8)	[0.8, +∞)
效应等级	Negligible(N)	Small(S)	Medium(M)	Large(L)

为使得网络结构不会过于复杂, 在实验中, EVP-PDAA 局部域对抗适应网络中的生成器和判别器全连接层的神经元数量设为 32, 隐层层数为 1, 预分类器和源域分类器均使用 RF。根据交叉验证结果, 将网络的学习率设为 0.0005、判别器和生成器训练次数比例设为 5、梯度惩罚项系数设

为 10、训练轮次为 200。

4.2　实验结果分析

从表 3 和表 4 可以看出，在面对不同的源域和目标域，EVP-PDAA 的 AUC 和 F_1 较未使用任何迁移学习方法的 RF-T 都具有显著的提升，AUC 平均提升 9%，F_1 平均提升 25%。说明了 EVP-PDAA 通过多个局部域对抗适应网络隐性地表示了源域发动机同目标域发动机的装配工艺对振动水平的共性影响关系，进而建立起目标域到源域的映射。并得益于源域分类器优异的分类性能，纠正了预分类器所给伪标签的错误，充分运用了源域的知识，进行了有效的迁移，具有较强泛化性。

表 3　各方法的 AUC 指标

Table 3　The AUC of different methods　%

目标域	源域	RF-T	CORAL	FSR	SHDA-RF	TCA	EVP-PDAA
SR-1	DR	76.50	64.30	44.23	78.35	37.02	**82.40**
	EP-1		82.92	58.33	77.20	70.33	**83.63**
	EP-2		80.52	59.52	76.92	29.72	**88.35**
SR-2	DR	72.62	74.08	74.12	73.97	71.18	**76.74**
	EP-1		75.94	46.74	75.28	63.23	**82.99**
	EP-2		73.27	44.49	76.07	60.36	**78.05**
SR-3	DR	60.15	65.19	71.49	58.90	63.57	**74.72**
	EP-1		**77.04**	47.14	54.17	58.11	71.27
	EP-2		64.58	39.29	56.25	55.00	**74.44**

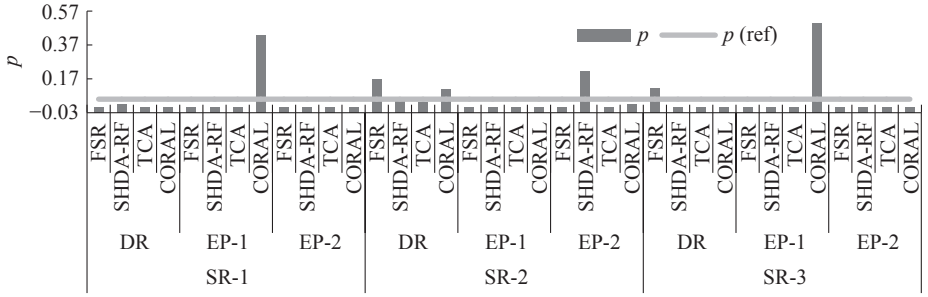
表 4　各方法的 F_1 指标

Table 4　F_1 of different methods　%

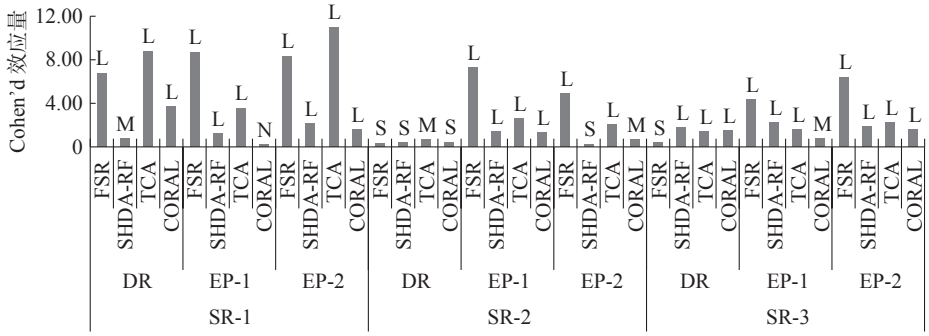
目标域	源域	RF-T	CORAL	FSR	SHDA-RF	TCA	EVP-PDAA
SR1	DR	24.75	1.25	30.00	25.86	0	**49.96**
	EP-1		**56.96**	18.18	29.44	14.28	50.43
	EP-2		1.43	18.18	30.53	0	**54.66**
SR-2	DR	20.00	25.85	**37.11**	20.45	0	36.06
	EP-1		27.35	19.91	39.16	6.61	**60.51**
	EP-2		38.88	19.00	28.69	1.25	**50.23**
SR-3	DR	0	0	30.35	0	0	**30.75**
	EP-1		0	0	0	0	**40.14**
	EP-2		0	0	0	0	**39.49**

在大部分迁移场景下，CORAL 建立的振动预测模型的 AUC 和 F_1 较 RF-T 都能够有一定提升，但在某些场景下，CORAL 会出现负迁移，预测性能不升反降，说明 CORAL 的泛用性较差。而其他迁移学习方法建立的预测模型在大部分情况下的性能均劣于 RF-T，出现了严重的负迁移，说明这些迁移学习方法在面对不平衡小样本不能进行合理的迁移。EVP-PDAA 的表现显著超过其他迁移学习方法，AUC 能够平均提升 26%，F_1 能够平均提升 30%。

从图 4 和图 5 可以看出，EVP-PDAA 和其他迁移学习方法建立的预测模型在性能上都具有统计意义的不同。AUC 和 F_1 指标的效应量等级几乎都为 L，且效应量值远大于 L 的阈值，说明较其他迁移学习方法，EVP-PDAA 能够更有效地迁移其他型号发动机的知识，建立性能更强的振动预测模型。



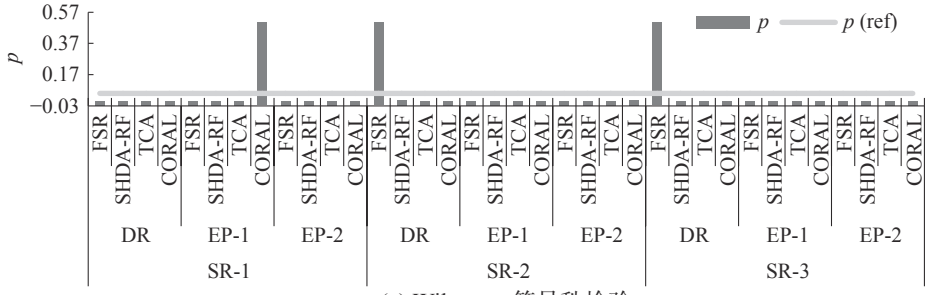
(a) Wilcoxon 符号秩检验



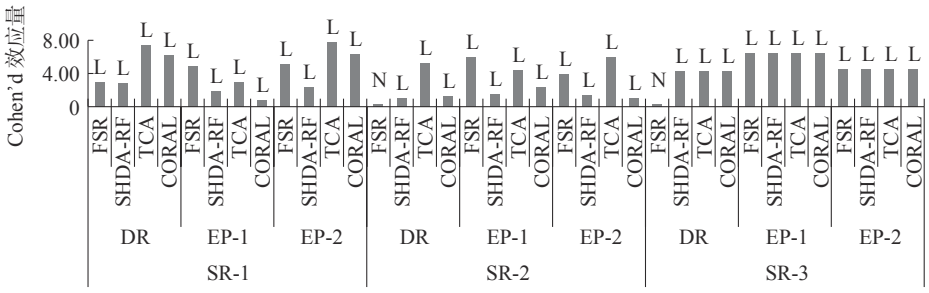
(b) Cohen’s d 效应量校验

图 4　EVP-PDAA 和其他迁移学习方法的 AUC 统计分析

Fig. 4　Statistical analysis of AUC between EVP-PDAA and other transfer learning methods



(a) Wilcoxon 符号秩检验



(b) Cohen’s d 效应量校验

图 5　EVP-PDAA 和其他迁移学习方法的 F_1 统计分析

Fig. 5　Statistical analysis of F_1 between EVP-PDAA and other transfer learning methods

此外, 对于 EVP-PDAA, 迁移 EP-1 和 EP-2 性能要优于迁移 DR 的性能。这是由于 EP-1 和 EP-2 的源域分类器性能要优于 DR, 表 5 给出了 RF 在源域上的分类性能, EP-1 和 EP-2 的源域分类器对伪标签的错误矫正能力更强, 所建立的振动预测模型的性能更强。也进一步说明了, 当 EVP-PDAA 通过多个局部域对抗适应网络建立起目标域到源域的映射, 且能在源域上建立一个分类性能很强的分类器时, 可以实现有效的迁移, 较好地解决了不平衡小样本带来的迁移难等问题。EVP-PDAA 不适用于在源域样本空间中无法建立一个具备优异分类性能的分类器的情况。

表 5　RF 在源域上的性能

Table 5　The performance of RF in source domain　%

指标	DR	EP-1	EP-2
AUC/	87.14	99.24	99.19
F_1/	88.89	92.95	93.84

4.3　预分类器的有效性分析

为对预分类器进行有效性分析, 预分类器使用 6 种不同类型的分类算法, 分别是人工神经网络(artificial neural network, ANN)、贝叶斯分类器(Bayes)、决策树(decision tree, DT)、逻辑回归(logistic regression, LR)、RF、支持向量机(support

vector machine, SVM）。其中, ANN 的网络结构为：输入层、包含 64 个神经元的全连接层、LeakyReLU 激活函数、输出层, 使用二元交叉熵作为损失函数, Adam 作为优化器。Bayes 使用高斯朴素贝叶斯, LR 使用 L2 作为正则项。EVP-PDAA 采用不同预分类器的 AUC 和 F_1 平均值如图 6 和图 7 所示。

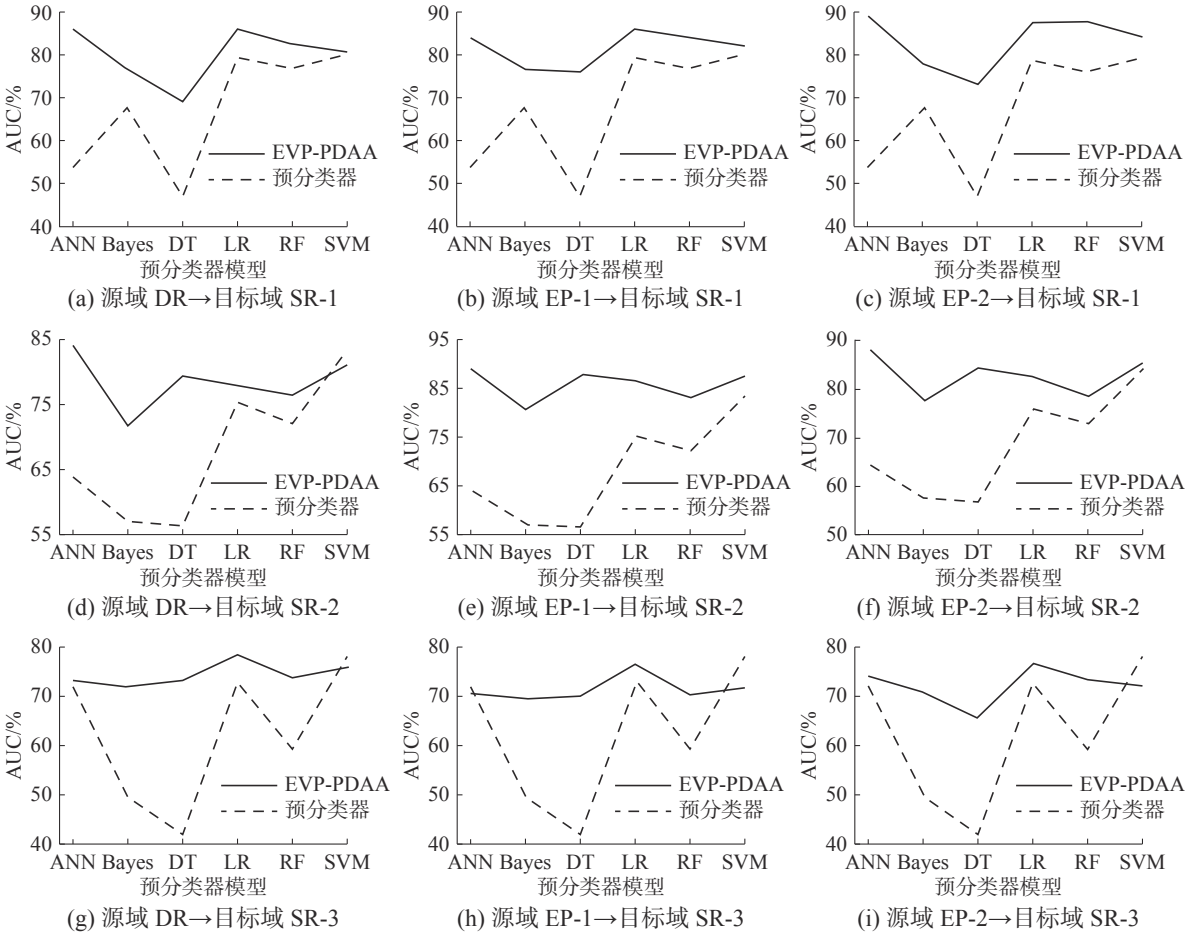
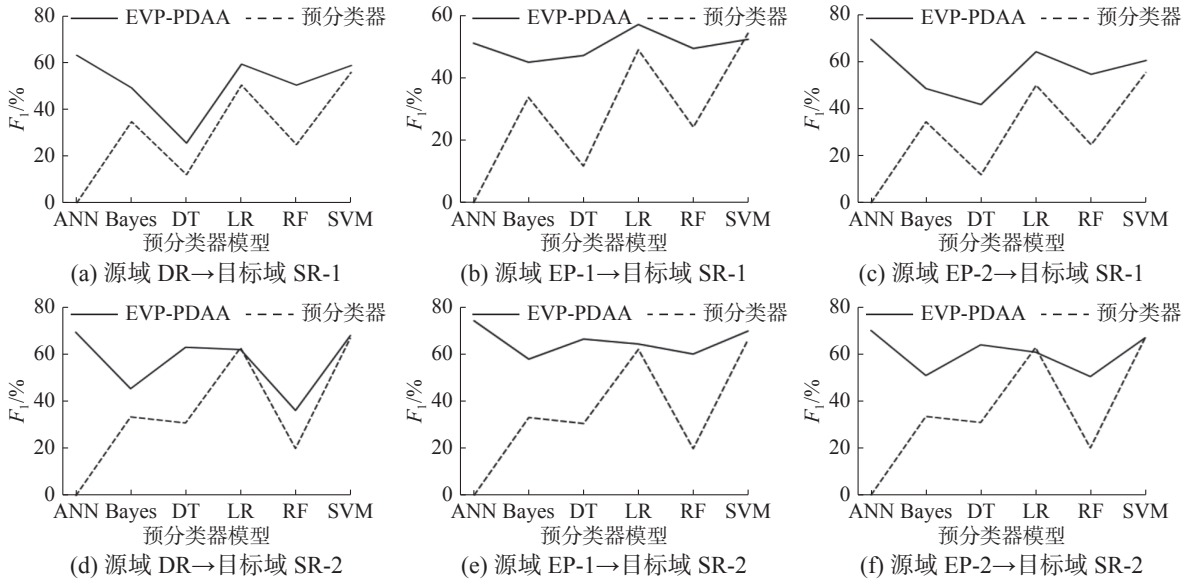
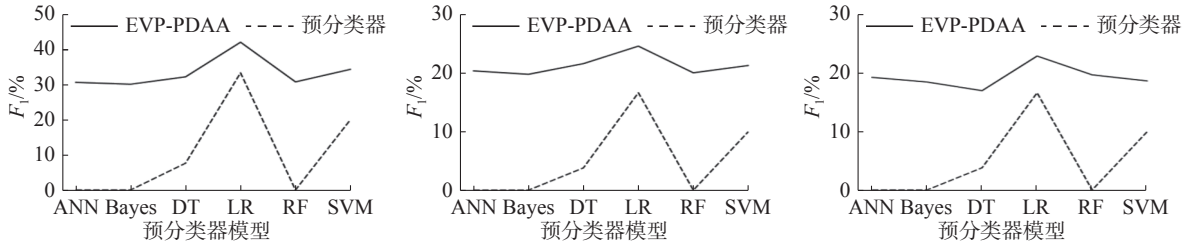


图 6　EVP-PDAA 和预分类器的 AUC 对比分析

Fig. 6　Comparison of AUC between EVP-PDAA and pre-classifier





(g) 源域 DR→目标域 SR-3　(h) 源域 EP-1→目标域 SR-3　(i) 源域 EP-2→目标域 SR-3

图 7　EVP-PDAA 和预分类器的 F_1 对比分析

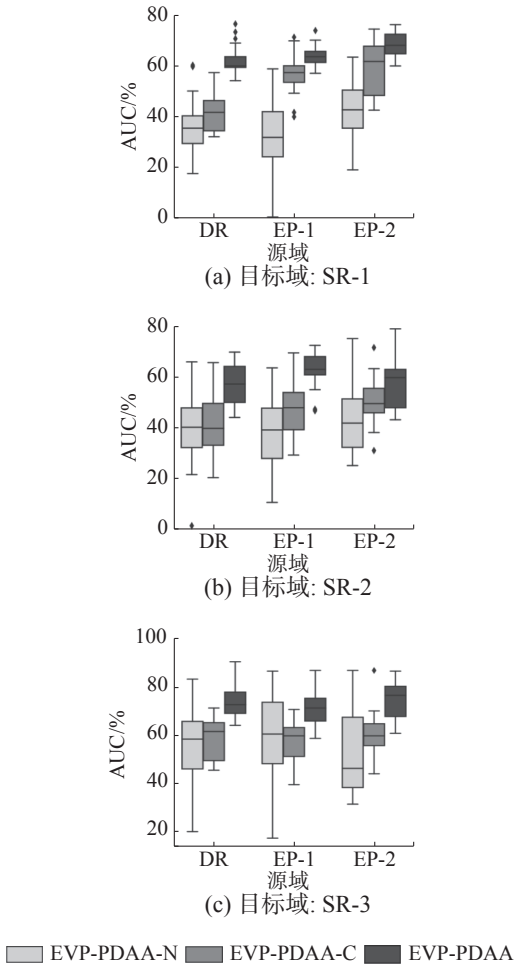
Fig. 7　Comparison of F_1 between EVP-PDAA and pre-classifier

从实验结果上看, 无论预分类器采用何种类型的分类算法, EVP-PDAA 的 AUC 和 F_1 均优于预分类器。当预分类器的指标较低时, EVP-PDAA 的提升幅度非常明显, EVP-PDAA 的性能与预分类器的性能基本呈正相关的关系, 预分类器的性能越强, EVP-PDAA 的性能越强。这是因为预分类器给出的伪标签越准确, 执行错误转换越少, 需要标签矫正的次数越少。此外, 从实验结果中, 我们发现 SVM 和 LR 较其他分类算法的性能更强, 这可能是因为目标域数据存在一个线性平面能够相对较好地划分多数类和少数类的决策边界。ANN、Bayes 和 DT 识别少数类的能力较差, 这可能是因为模型出现了过拟合, 决策边界更偏向多数类。

4.4　局部域对抗适应网络结构的有效性分析

为对局部域对抗适应网络结构进行有效性分析, 本文对另外两种网络结构进行了实验, 分别是使用梯度裁剪策略来保证 Lipschitz 连续, 记采用这个网络结构的 EVP-PDAA 为 EVP-PDAA-C; 以及使用二元交叉熵作为生成器和判别器的损失函数, 记为 EVP-PDAA-N。

PDAA、EVP-PDAA-N 和 EVP-PDAA-C 的预分类器均以 RF 为预分类器, 图 8 和图 9 分别显示了各方法的 AUC 和 F_1 值。从实验结果中, 我们可以发现采用 EMD 距离作为目标函数并使用梯度惩罚策略的 PDAA 在 AUC 和 F_1 指标上均优于 EVP-PDAA-C 和 EVP-PDAA-N, 而 EVP-PDAA-C 的性能要优于 EVP-PDAA-N。这是因为当 EVP-PDAA-N 的判别器训练得过好时, 生成器的损失函数会出现梯度消失, 并且生成器损失函数的梯度不够稳定, 容易出现模型崩溃等问题。采用 EMD 作为目标函数的 EVP-PDAA-C 虽然能够解决生成器损失函数梯度消失的问题, 保证网络训练的稳定性, 但梯度裁剪策略直接将梯度暴力地限制在一个常数空间, 样本生成能力差和目标函数不能收敛的问题仍会存在。当局部域对抗适应网络中生成器训练得越好, 建立的目标域到源域的映射函数越准确, EVP-PDAA 的性能也会越强。



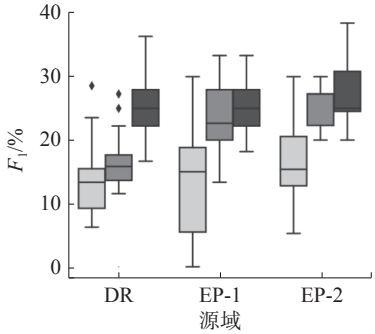
(a) 目标域: SR-1

(b) 目标域: SR-2

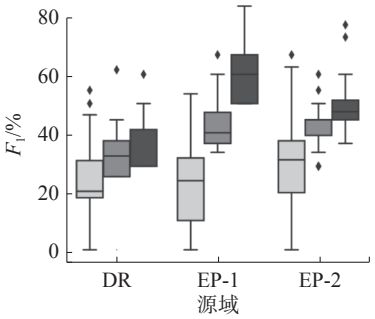
(c) 目标域: SR-3

图 8　EVP-PDAA、EVP-PDAA-C、EVP-PDAA -N 的 AUC 箱型图

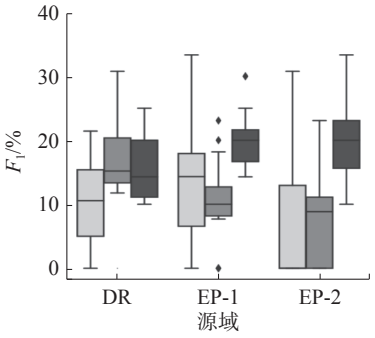
Fig. 8　AUC box plot of EVP-PDAA、EVP-PDAA-C、EVP-PDAA -N



(a) 目标域: SR-1



(b) 目标域: SR-2



(c) 目标域: SR-3

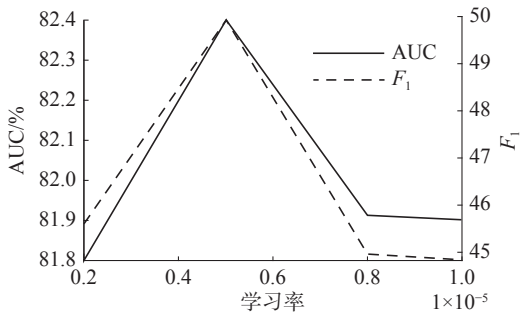
图 9　EVP-PDAA、EVP-PDAA-C、EVP-PDAA -N 的 F_1 箱型图

Fig. 9　F_1 box plot of EVP-PDAA、EVP-PDAA-C、EVP-PDAA -N

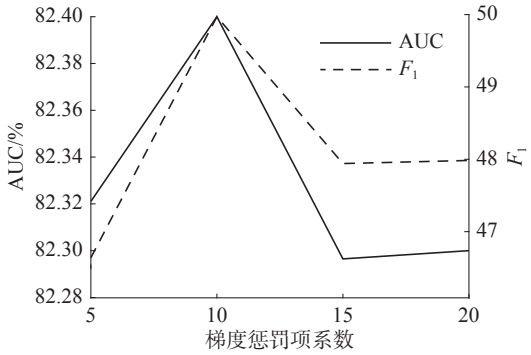
4.5　超参数的有效性分析

为分析各训练超参数对 EVP-PDAA 性能的影响，本文对学习率、梯度惩罚项系数λ、判别器和生成器训练次数比例n_d、训练轮次 T 分别进行了实验。在实验中，源域数据集为 DR，目标域数据集为 SR-1，使用 RF 作为预分类器。分析某一超参数时，其他超参数均固定。为避免随机因子造成影响，每组实验都重复 20 次，取平均值作为对比。

图 10 给出了 EVP-PDAA 在不同学习率和λ下，AUC 和 F_1 指标的变化情况。从图中可以看出，当学习率取 0.000 5 时，EVP-PDAA 的性能最佳。不同梯度惩罚项系数λ的取值为 10 时，EVP-PDAA 性能最强。



(a) 不同学习率

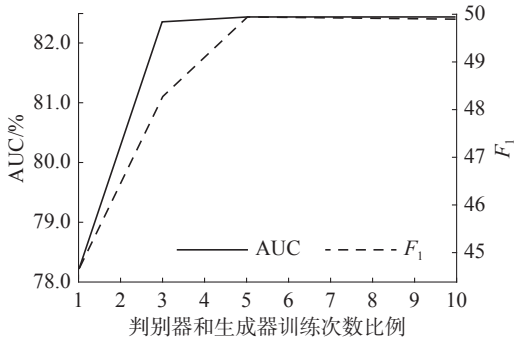


(b) 不同梯度惩罚项系数

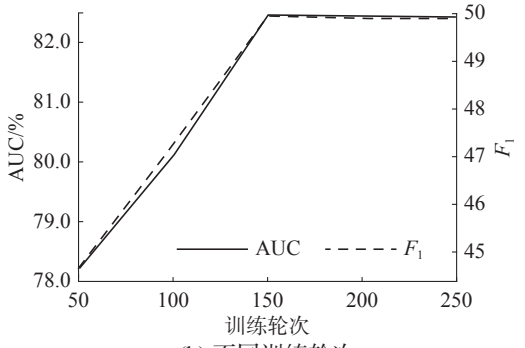
图 10　EVP-PDAA 在不同学习率和梯度惩罚项系数 λ 下 AUC 和 F_1 指标的变化情况

Fig. 10　AUC and F_1 of EVP-PDAA under different learning rate and gradient penalty coefficient λ

图 11 给出了 EVP-PDAA 在n_d、T 不同的取值下，AUC 和 F_1 指标的变化情况。从图中可以看出，当n_d在到达 5 时，判别器已能够得到较好地训练，EVP-PDAA 性能已基本最优，而当n_d小于 5 时，由于判别器训练得不够充分，导致模型性能受到较大影响。此外，当 T 到达 200 时，网络已基本训练完成，此时模型性能已趋于稳定。



(a) 不同判别器和生成器训练次数比例



(b) 不同训练轮次

图 11　EVP-PDAA 在不同n_d、T 下 AUC 和 F_1 指标的变化情况

Fig. 11　AUC and F_1 of EVP-PDAA under different n_d and T

5　结束语

实际装配过程中采集到的发动机装配数据呈现样本量小且类别不平衡的特点，直接建立的发

动机振动预测模型性能不佳。借助迁移学习技术, 可以将其他型号的发动机装配数据和实验平台的装配数据迁移到目标发动机, 辅助目标发动机的振动预测。但现有迁移学习方法在面对不平衡小样本并不能进行合理有效的迁移, 因此, 我们提出了基于局部域对抗适应网络的发动机振动预测模型。将领域按标签分为多个局部域, 通过多个局部域对抗适应网络将目标域样本映射到源域, 保证小样本中的少数类得到正确的迁移。通过伪标签来解决目标样本的域转换, 并使用标签矫正机制给出可靠的预测结果。

参考文献:

[1] BÖYÜKDIPI Ö, TÜCCAR G, SOYHAN H S. Experimental investigation and artificial neural networks (ANNs) based prediction of engine vibration of a diesel engine fueled with sunflower biodiesel-NH3 mixtures[J]. Fuel, 2021, 304: 121462.

[2] ELSAID A, EL JAMIY F, HIGGINS J, et al. Optimizing long short-term memory recurrent neural networks using ant colony optimization to predict turbine engine vibration[J]. Applied soft computing, 2018, 73: 969–991.

[3] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述 [J]. 软件学报, 2021, 32(2): 349–369.
ZHAO Kailin, JIN Xiaolong, WANG Yuanzhuo. Survey on few-shot learning[J]. Journal of software, 2021, 32(2): 349–369.

[4] LU Wang, CHEN Yiqiang, WANG Jindong, et al. Cross-domain activity recognition via substructural optimal transport[J]. Neurocomputing, 2021, 454: 65–75.

[5] TIAN Lei, TANG Yongqiang, HU Liangchen, et al. Domain adaptation by class centroid matching and local manifold self-learning[J]. IEEE transactions on image processing, 2020, 29: 9703–9718.

[6] TASKESEN B, YUE M C, BLANCHET J, et al. Sequential domain adaptation by synthesizing distributionally robust experts[C]//International Conference on Machine Learning. [S.l.]: PMLR, 2021: 10162−10172.

[7] 朱应钊. 异构迁移学习研究综述 [J]. 电信科学, 2020, 36(3): 100–110.
ZHU Yingzhao. Review on heterogeneous transfer learning[J]. Telecommunications science, 2020, 36(3): 100–110.

[8] 李荣军, 郭秀焱, 杨静远. 面向鲁棒口语理解的声学组块混淆语言模型微调算法 [J]. 智能系统学报, 2023, 18(1): 131–137.
LI Rongjun, GUO Xiuyan, YANG Jingyuan. A fine-tuning algorithm for acoustic text chunk confusion language model orienting to understand robust spoken language[J]. CAAI transactions on intelligent systems, 2023, 18(1): 131–137.

[9] LIU Bingyan, CAI Yifeng, GUO Yao, et al. TransTailor: pruning the pre-trained model for improved transfer learning[EB/OL]. (2021−03−02)[2021−06−06]. https://doi.org/10.48550/arXiv.2103.01542.

[10] 孔伶旭, 吴海锋, 曾玉, 等. 迁移学习特征提取的 rs-fMRI 早期轻度认知障碍分类 [J]. 智能系统学报, 2021, 16(4): 662–672.
KONG Lingxu, WU Haifeng, ZENG Yu, et al. Transfer learning-based feature extraction method for the classification of rs-fMRI early mild cognitive impairment[J]. CAAI transactions on intelligent systems, 2021, 16(4): 662–672.

[11] OUYANG L, KEY A. maximum mean discrepancy for generalization in the presence of distribution and missingness shift[C]//NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications. [S.l.]: IEEE, 2021.

[12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139–144.

[13] BALDEON C M G, LAI-YUEN S K. C-MADA: unsupervised cross-modality adversarial domain adaptation framework for medical image segmentation[C]//SPIE Medical Imaging Proc SPIE 12032, medical imaging 2022: image processing. San Diego: [s.n.], 2022: 971−978.

[14] 钱亚冠, 马骏, 何念念, 等. 面向边缘智能的两阶段对抗知识迁移方法 [J]. 软件学报, 2022, 33(12): 4504–4516.
QIAN Yaguan, MA Jun, HE Niannian, et al. Two-stage adversarial knowledge transfer for edge intelligence[J]. Journal of software, 2022, 33(12): 4504–4516.

[15] SUN Mingfei, MA Xiaojuan. Adversarial imitation learning from incomplete demonstrations[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 3513.

[16] ROBBIANO L, UR RAHMAN M R, GALASSO F, et al. Adversarial branch architecture search for unsupervised domain adaptation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022: 1008−1018.

[17] DUAN Lixin, XU Dong, TSANG I W. Learning with augmented features for heterogeneous domain adaptation[C]//Proceedings of the 29th International Coference on International Conference on Machine Learning. New York: ACM, 2012: 667−674.

[18] SUKHIJA S, KRISHNAN N C. Supervised heterogeneous feature transfer via random forests[J]. Artificial intelligence, 2019, 268: 30–53.

[19] FEUZ K D, COOK D J. Transfer learning across feature-rich heterogeneous feature spaces via feature-space remapping (FSR)[J]. ACM transactions on intelligent systems and technology, 2015, 6(1): 1–27.

[20] DEMIRKıRAN F, ÇAYıR A, ÜNAL U, et al. An ensemble of pre-trained transformer models for imbalanced multiclass malware classification[J]. Computers & security, 2022, 121: 102846.

[21] ZHANG Jianguo, BUI T, YOON S, et al. Few-shot intent detection via contrastive pre-training and fine-tuning[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 1906−1912.

[22] HONG Jin, YU S C H, CHEN Weitian. Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning[J]. Applied soft computing, 2022, 121: 108729.

[23] ADLER J, LUNZ S. Banach wasserstein gan[J]. Advances in neural information processing systems, 2018, 31.

[24] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 5769−5779.

[25] LONG Mingsheng, WANG Jianmin, DING Guiguang, et al. Transfer feature learning with joint distribution adaptation[C]//2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2014: 2200−2207.

[26] SUN Baochen, FENG Jiashi, SAENKO K. Correlation alignment for unsupervised domain adaptation[M]// Csurka G. Domain Adaptation in Computer Vision Applications. Cham: Springer, 2017: 153−171.

[27] GONG Lina, JIANG Shujuan, JIANG Li. Conditional domain adversarial adaptation for heterogeneous defect prediction[J]. IEEE access, 2020, 8: 150738–150749.

作者简介：

季友昌，硕士研究生，主要研究方向为机器学习。

袁伟伟，教授，博士，主要研究方向为机器学习、人机协同。主持完成国家自然科学基金 2 项，参与重点研发计划 2 项。发表学术论文 100 余篇。

关东海，副教授，博士，主要研究方向为数据挖掘、知识推理。主持完成国家自然科学基金 2 项，参与重点研发计划 2 项、重大科技专项 1 项。发表学术论文 100 余篇。