

# 改进均衡分布适配的滚动轴承寿命阶段识别

吴昊年<sup>1</sup>, 陈仁祥<sup>1,2</sup>, 胡小林<sup>3</sup>, 张霞<sup>1</sup>, 张焱<sup>4</sup>, 唐林林<sup>1</sup>

(1. 重庆交通大学交通工程应用机器人重庆市工程实验室, 重庆 400074; 2. 重庆大学机械传动国家重点实验室, 重庆 400030; 3. 重庆工业大数据创新中心有限公司, 重庆 400056; 4. 重庆邮电大学自动化学院, 重庆 400065)

**摘要:** 针对不同工况下训练样本与测试样本分布差异导致滚动轴承寿命阶段无法被有效识别的问题, 提出改进均衡分布适配的滚动轴承寿命阶段识别方法。采用无重复均匀随机抽样对源域类间样本进行多次均匀随机抽样, 得到源域多样本训练集, 以减小源域内部样本选择对目标域预测标签的影响; 在再生核希尔伯特空间上利用平衡因子  $\mu$  动态调节边缘分布和条件分布所占权值, 并通过迭代的方式不断优化目标域伪标签以减小两域的最大均值差异; 利用源域多样本数据集各自的映射矩阵构造多个分类器, 经过一致性判别得到目标域样本最终识别结果。在两组滚动轴承寿命阶段数据集上进行实验验证, 证明了所提方法的可行性和有效性。

**关键词:** 故障诊断; 滚动轴承; 寿命阶段识别; 条件概率分布; 边缘分布

**中图分类号:** TH165+.3; TH133.33; TN911.7 **文献标志码:** A **文章编号:** 1004-4523(2021)01-0194-08

**DOI:** 10.16385/j.cnki.issn.1004-4523.2021.01.022

## 引言

滚动轴承作为旋转机械关键零部件之一, 对其寿命阶段识别可以监测其性能衰退过程, 有效防止重大安全事故的发生。为准确识别滚动轴承寿命阶段, 国内外已开展了相关研究。如: Yu 等<sup>[1]</sup>提出多域特征融合和降维学习的滚动轴承退化状态识别方法; 陈仁祥等<sup>[2]</sup>提出基于振动敏感时频特征的航天轴承寿命状态识别方法; 王冰等<sup>[3]</sup>提出了基于模糊聚类的退化状态识别方法。以上方法对同种工况下滚动轴承寿命阶段识别效果明显。但在不同工况下, 可获得的训练样本数目有限, 且训练数据与测试数据不满足独立同分布条件, 降低了传统机器学习寿命识别模型的泛化能力, 甚至使得模型不适用。

近年来, 迁移学习以其跨领域、跨任务学习的优势在各领域得到了广泛应用<sup>[4-5]</sup>。在机械研究领域, 解决寿命阶段识别问题时, 往往借鉴故障诊断方法。沈飞等<sup>[6]</sup>将奇异值分解与迁移学习用于不同工况下的电机轴承故障的识别; 段礼祥等<sup>[7]</sup>将迁移成分分析应用于不同工况下的齿轮箱故障诊断; 康守强等<sup>[8]</sup>利用多核半监督迁移成分分析方法解决了变工况下滚动轴承故障诊断问题。上述方法关注于最小

化域之间的边缘分布差异以达到较好的适配效果。然而在实际工程中, 外在摩擦力、温度、工况条件等因素的变化, 使得采集到的标记数据和目标域轴承寿命阶段数据分布特性差异较大, 将两种概率分布的重要性同等对待, 往往导致对滚动轴承寿命阶段识别效果不佳。Wang 等<sup>[9]</sup>针对实际应用中边缘分布适配和条件分布适配并不是同等重要的问题, 提出了均衡分布适配方法 (Balanced Distribution Adaptation, BDA), 通过平衡因子适配两域分布取得了不错效果。然而 BDA 方法构建的单一弱分类器具有局限性, 无法较好完成不同工况条件下寿命阶段样本的识别任务。

为解决上述问题, 本文提出改进均衡分布适配的滚动轴承寿命阶段识别方法。首先, 通过无重复均匀随机抽样, 对源域类间样本进行多次等量随机抽样得到源域多样本训练集, 分别在目标域上预测其伪标签, 避免了源域样本的选择对目标域预测标签的影响; 随后, 将源域多样本训练集与目标域测试样本集共同映射到再生核希尔伯特空间中, 利用平衡因子  $\mu$  动态调节边缘分布和条件分布所占权值, 适配轴承不同寿命阶段数据两域分布差异; 最后, 通过多个分类器识别目标域寿命阶段数据, 经一致性判别获得最终识别结果。在两组滚动轴承寿命阶段

**收稿日期:** 2019-05-21; **修订日期:** 2019-12-23

**基金项目:** 国家自然科学基金资助项目 (51975079, 51705056); 重庆市教委科学技术研究项目 (KJQN201900721); 机械传动国家重点实验室开放基金资助项目 (SKLMT-KFKT-201710); 重庆市技术创新与应用示范项目 (cstc2018jscx-msybX0012); 交通工程应用机器人重庆市工程实验室开放基金资助项目 (CELTEAR-KFKT-202002); 重庆交通大学硕士研究生科研创新项目 (2018S0138)

数据集上进行验证,证明了本文方法能有效识别不同寿命阶段样本,提高了识别准确率。

## 1 均衡分布适配方法的原理

假设源域为  $D_s = \{x_i, y_i\}_{i=1}^n$ ,  $x_i$  为源域特征样本集,  $y_i$  为样本标签向量。没有标记的目标域  $D_t = \{x_j\}_{j=n+1}^{n+m}$ , 假定源域与目标域的特征空间相同, 即  $\mathcal{X}_s = \mathcal{X}_t$ , 并且它们的类别空间也相同, 即  $\mathcal{Y}_s = \mathcal{Y}_t$ 。当条件分布不相同, 即  $P(y_s|x_s) \neq P(y_t|x_t)$  的情况下, 源域与目标域的边缘分布也不同,  $P_s(x_s) \neq P_t(x_t)$ 。

均衡分布适配方法通过最小化域间的边缘分布和条件分布差异来解决迁移学习问题。利用有标记的数据  $D_s$  来预测  $D_t$  的标签  $y_t \in \mathcal{Y}_t$ 。通过引入平衡因子  $\mu$ , 可以根据具体的任务自适应调节两个分布各自的重要性, 其表达式为

$$D(D_s, D_t) \approx (1 - \mu)(P(x_s), P(x_t)) + \mu(P(y_s|x_s), P(y_t|x_t)) \quad (1)$$

式中 平衡因子  $\mu \in [0, 1]$  决定了边缘分布和条件分布哪个应该被优先考虑。当  $\mu \rightarrow 0$  时, 说明源域与目标域数据本身存在较大的差异, 应优先考虑数据边缘分布。当  $\mu \rightarrow 1$  时, 说明两域有较高的相似性, 应该先考虑条件分布差异。BDA 在两域间寻找一个变换矩阵  $A$ , 使得经过变换后  $P(A^T x_s)$  和  $P(A^T x_t)$  的距离与  $P(y_s|A^T x_s)$  和  $P(y_t|A^T x_t)$  的距离最小。利用最大均值差异 MMD<sup>[10]</sup> 距离, 将公式(1)统一为

$$D(D_s, D_t) \approx (1 - \mu) \left\| \frac{1}{n} \sum_{i=1}^n x_{si} - \frac{1}{m} \sum_{j=1}^m x_{tj} \right\|_{\mathcal{H}}^2 + \mu \sum_{c=1}^c \left\| \frac{1}{n_c} \sum_{x_{si} \in D_s^{(c)}} x_{si} - \frac{1}{m_c} \sum_{x_{tj} \in D_t^{(c)}} x_{tj} \right\|_{\mathcal{H}}^2 \quad (2)$$

式中  $\mathcal{H}$  为再生核希尔伯特空间 (Reproducing Kernel Hilbert Space, RKHS),  $n$  和  $m$  分别代表源域和目标域样本的个数。  $n_c$  和  $m_c$  分别表示源域和目标域中来自第  $c$  类的样本个数。

当边缘分布较接近时, 迁移学习性能高度依赖于条件分布<sup>[9]</sup>。由于 BDA 算法目标域没有类别向量  $y_t$ , 无法直接求其条件分布, 故根据贝叶斯公式忽略  $P(y_t)$ , 用  $P(x_t|y_t)$  来近似  $P(y_t|x_t)$ 。从技术角度可以直接利用  $(x_t|y_t)$  来训练一个简单的分类器, 在  $x_t$  上直接预测目标域伪标签  $\hat{y}_t$ , 并通过迭代的方式优化得到可靠的目标域标签。利用矩阵技巧和正则化, 式(2)可以形式化为

$$\min \text{tr} \left( A^T X \left[ (1 - \mu) M_0 + \mu \sum_{c=1}^c M_c \right] X^T A \right) + \lambda \|A\|_F^2 \quad (3)$$

s.t.  $A^T X H X^T A = I$

式中  $\lambda$  为正则化项  $\|\cdot\|_F^2$  的 Frobenius 系数。  $X$  表示由  $x_s$  和  $x_t$  组成的矩阵,  $A$  表示变换矩阵,  $I \in \mathbb{R}^{(n+m) \times (n+m)}$  为单位矩阵,  $H$  为中心矩阵即  $H = I - (1/n)\mathbf{1}$ ,  $\mathbf{1}$  是全为 1 的列向量。  $M_0$  和  $M_c$  为 MMD 矩阵, 表达式为

$$(M_0)_{ij} = \begin{cases} \frac{1}{n^2}, & x_i, x_j \in D_s \\ \frac{1}{m^2}, & x_i, x_j \in D_t \\ -\frac{1}{mn}, & \text{otherwise} \end{cases} \quad (4)$$

$$(M_c)_{ij} = \begin{cases} \frac{1}{n^2}, & x_i, x_j \in D_s \\ \frac{1}{m^2}, & x_i, x_j \in D_t \\ -\frac{1}{mn}, & \begin{cases} x_i \in D_{si}^{(c)}, x_j \in D_{tj}^{(c)} \\ x_i \in D_{ti}^{(c)}, x_j \in D_{sj}^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

式(3)中涉及两个约束: 第一个约束确保转换后的数据 ( $A^T x$ ) 保留原始数据的内部属性; 第二个约束表示平衡因子  $\mu$  的取值范围。引入拉格朗日乘子  $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_d)$ , 并将公式(4)和(5)联立得到式(3)的拉格朗日函数为

$$L = \text{tr} \left( A^T X \left[ (1 - \mu) M_0 + \mu \sum_{c=1}^c M_c \right] X^T A \right) + \lambda \|A\|_F^2 + \text{tr} \left( (I - A^T X H X^T A) \Phi \right) \quad (6)$$

作为一个广义特征分解问题, 只需要其偏导  $\partial L / \partial A = 0$  即可得到优化

$$\left\{ X \left[ (1 - \mu) M_0 + \mu \sum_{c=1}^c M_c \right] X^T + \lambda I \right\} A = X H X^T A \Phi \quad (7)$$

解之即可得到最佳映射变换矩阵  $A$ 。值得注意的是伪标签的预测直接影响了整个求解过程。

## 2 改进均衡分布适配的滚动轴承寿命阶段识别

### 2.1 改进均衡分布适配方法的原理

为避免复杂数据下 BDA 方法易受源域内部样本选择的影响和单一弱分类器的局限, 采用无重复均匀随机抽样与多分类器同时识别为切入点, 对

BDA进行改进。具体步骤如下:

设 $X_s$ 包括 $M$ 个类别,每一类样本表示为 $\varphi_i$ , $\forall i \in \mathbf{A} = \{1, \dots, M\}$ 的源域为已标记数据集,其中 $\varphi_1$ 代表源域第一类样本。 $X_t$ 为目标域无标记数据集,将两个数据集提取所需特征。

1)对 $X_s$ 每类样本进行无重复均匀随机抽样,每类样本的抽样数均为 $n$ , $n \leq \min(\varphi_1, \varphi_2, \dots, \varphi_M)$ 。共得到 $M \times n$ 个训练样本,组成源域单样本训练集 $x_i$ ;

2)重复步骤1) $k$ 次,得到 $k$ 个源域单样本训练集 $\{x_{s1}, x_{s2}, \dots, x_{sk}\}$ ,称为源域多样本训练集;

3)将步骤2)中 $k$ 个源域多样本训练集与目标域测试样本集共同映射到 $\mathcal{X}$ 空间中。直接在目标域上,训练 $k$ 个弱分类器分别预测得到目标域测试集伪标签 $\hat{y}_i$ ;

4)同时动态调节平衡因子 $\mu$ 适配 $P(A^T x_{si})$ 和 $P(A^T x_t)$ 与 $P(y_s | A^T x_{si})$ 和 $P(y_t | A^T x_t)$ 的最小均值差异,提高数据类内紧凑性和类间区分性,以获得较好的伪标签;

5)设 $f_i(\cdot)$ 为训练产生的一个分类器,其中 $i=1, \dots, k$ 。存在 $\omega_j(j=1, \dots, M)$ 是第 $j$ 类的标记, $f_i(j)$ 表示第 $i$ 个分类器对第 $j$ 类样本的预测。每一类得票

数目

$$A_j = \{t | f_t(j) = \varphi_i\} \quad (8)$$

式中 $A_j(j=1, \dots, M)$ ,包含 $M$ 个类别,每个类别的表达式如式(8)所示。

6)将步骤5)中 $k$ 个分类器识别结果,经过一致性判别公式

$$f_E(x_{sk}) = \operatorname{argmax}(A_j) \quad (9)$$

得到最终识别结果。

## 2.2 改进均衡分布适配的滚动轴承寿命阶段识别流程

根据上文论述,为实现改进均衡分布适配的滚动轴承寿命阶段识别,其实现流程图如图1所示。该算法实现主要包括:

1)样本特征提取。由于单域特征评估效果不足,本文提取多域特征构建高维特征集。包括16维时域特征和12维频域特征,8维db3小波3层小波包能量特征,8维db3小波3层小波包相对能量特征,8维db3小波3层小波包能量谱熵以及振动信号幅值谱熵、倒谱熵、自相关谱熵和奇异值谱熵等共56维特征。

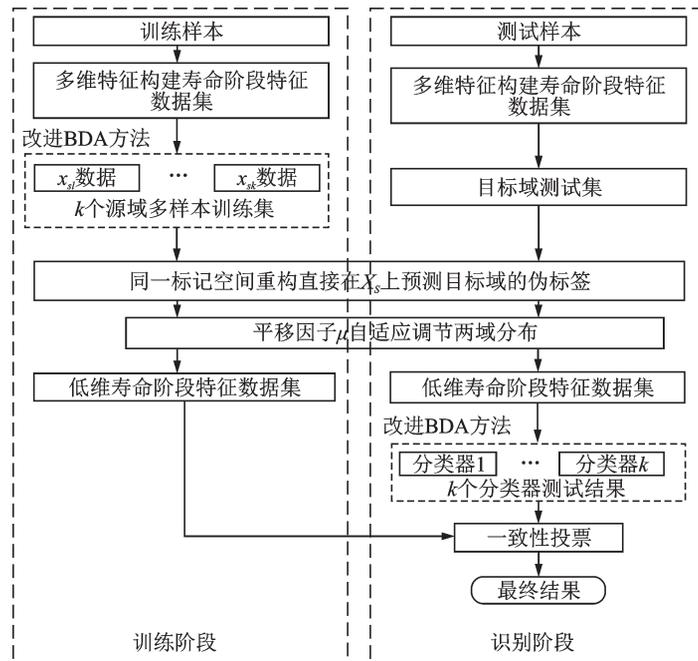


图1 寿命阶段识别流程图

Fig. 1 Flow chart of life stage identification

2)源域多样本训练集与目标域训练集构建。对于源域标记数据,采用无重复均匀随机抽样从 $M$ 种寿命阶段样本中抽取 $n$ 个,得到一个源域单样本训练集,训练集大小为 $M \times n$ 。重复上述过程 $k$ 次,得到源域多样本训练集。其中 $k$ 个单样本训练集,每

个待识别阶段均为 $M$ 。对于目标域未标记数据,从中每次每类抽取等量样本构建测试集,待识别阶段为 $M$ 种。

3)同一标记空间重构。将源域多样本训练集与目标域测试集共同映射至 $\mathcal{X}$ 空间中, $k$ 个源域多样

本训练集各自为待识别目标域数据集赋予伪标签。通过平衡因子 $\mu$ 动态调节两域分布,执行两域类内、类间知识迁移不断优化伪标签。

4)输出识别结果。构造多个分类器分别识别重构后目标域寿命阶段特征,通过一致性判别方法输出最终识别结果。由于KNN分类器具有计算简单,易于实现,无需参数估计和训练的优点。本文选择KNN分类器用于伪标签的预测和最终结果的分类。

### 3 全寿命周期数据实验验证

#### 3.1 实验设备及参数设置

实验采用 PRONOSTIA 实验台采集的 IEEE PHM2012 Data Challenge<sup>[11]</sup>加速寿命实验振动信号数据进行实验验证和分析。此数据采样频率为 25.6 kHz,采样间隔为 10 s,每个样本采样时间为 0.1 s。PRONOSTIA 实验平台如图 2 所示。

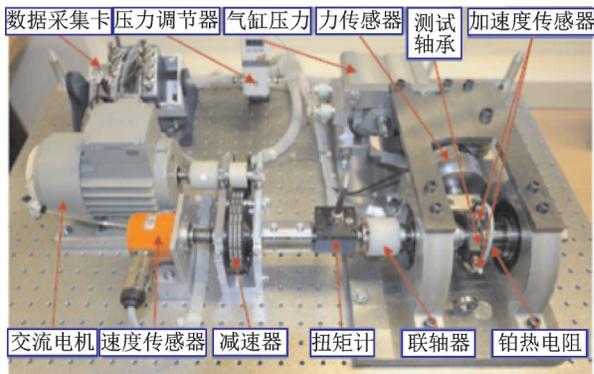


图 2 PRONOSTIA 实验平台示意图

Fig.2 Schematic diagram of PRONOSTIA experimental platform

该数据集包含多个工况条件下的全寿命周期实验数据,选择 3 种工况条件下的滚动轴承全寿命周期振动信号数据,数据工况信息如表 1 所示。

表 1 工况信息

Tab.1 Working condition information

工况	转速/(r·min <sup>-1</sup> )	载荷/N	样本数	寿命阶段
A	1650	4200	911	3
B	1500	5000	1637	3
C	1800	4000	2259	3

滚动轴承从全新装配到完全失效的整个寿命周期共经历 3 种寿命阶段:磨合期、有效工作期和衰退期。实验设置 3 种工况条件下数据样本集:1)工况 A 为 1650 r/min,4200 N 数据样本集;2)工况 B 为 1500 r/min,5000 N 数据样本集;3)工况 C 为 1800 r/

min,4000 N 数据样本集。为了更好地完成寿命阶段识别实验,利用文献[12]的方法划分不同工况的 3 种寿命阶段时,截取阶段明显的寿命阶段样本,把不明确样本进行了少量剔除。具体每个寿命阶段样本数分布如表 2 所示。

表 2 全寿命周期数据各寿命阶段样本数

Tab.2 Sample number of life cycle data in each life stage

工况	磨合期/个	有效工作期/个	衰退期/个
A	60	800	37
B	50	1100	100
C	150	1375	180

#### 3.2 改进均衡分布适配方法实验分析

本实验主要验证不同工况下源域少标记且目标域完全没有标记时,改进均衡分布适配方法对异分布寿命阶段数据识别效果。训练样本采用无重复随机抽样从 B 工况 3 个寿命阶段中各抽取 10,20 和 30 个,获得 3 个寿命阶段共 30,60,90 个为一个源域样本集(为使实验符合实际工程中训练数据获取困难的情况,抽取的源域单样本数量占整个 B 工况源域样本数量的 2.4%,4.8%,7.2%),抽取 1000 次得到 1000 个源域多样本数据集。

设置 A 工况与 C 工况样本为测试样本,数量分别为 30,60,90。提取 56 维特征构建高维特征集,对此轴承数据进行不同工况下寿命阶段识别结果如表 3 所示。

表 3 不同工况下全寿命周期数据寿命阶段识别结果

Tab.3 Life phase recognition of life cycle data under different working conditions

测试	测试样本数	B 工况		
		30	60	90
A 工况	30	90.46%	96.67%	96.67%
	60	77.29%	95.56%	96.67%
	90	78.89%	91.11%	97.78%
C 工况	30	76.67%	90.00%	87.08%
	60	74.77%	83.33%	88.89%
	90	78.89%	89.96%	86.67%
平均		79.49%	91.10%	92.29%

从表 3 中可以看出,本方法在不同工况下滚动轴承寿命识别上表现良好。随着训练样本数量的增多,识别率也呈缓慢增加的趋势。不同转速与负载的工况下平均识别率在 90% 左右,样本数量在 60 左右已经能达到很好的效果。

为验证改进均衡适配算法较原方法识别精度明显提高。选择B工况训练60个样本,A工况测试60个样本,将整个寿命阶段样本作为源域单样本BDA,无重复均匀随机抽样得到改进均衡分布适配方法进行实验对比。为直观对比结果将源域多样本数据集得到的诊断结果直接求平均值,称为源域多样本平均。对比结果如图3所示。

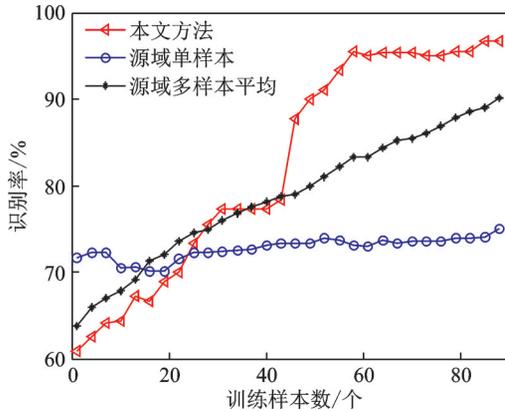


图3 样本集成对比图

Fig. 3 Sample integration comparison chart

由图3可知,由于目标域没有标签,只能通过源域数据直接预测其伪标签,所以源域内部样本的选择直接影响识别结果。而单分类器识别能力有限,样本数量对源域单样本识别影响不大。而本文方法反映了识别率与训练样本数目呈正相关的规律,凭借改进均衡分布适配方法动态调节两域分布重要性的优势,更好地最小化了域间分布差异,提升了识别率,平均达到92.29%;与源域多样本平均对比可得,随着不同源域样本在各自分类器上预测目标域伪标签,样本数量与准确率呈正相关规律;且多分类器集成后使最终识别精度大大提升。

### 3.3 改进均衡分布适配的迭代次数 $N$ 与 $\mu$ 的选择

改进均衡分布适配方法涉及两个主要参数:1)迭代次数 $N$ ;2)平衡因子 $\mu$ 。文献[9]对这两个参数的选择方法已经进行了讨论。 $N$ 值的设置不宜过大,选择适当的值既可以节约运算时间又能保证识别的准确率。经过交叉验证实验,本文取 $N=20$ 。平衡因子 $\mu \in [0, 1]$ 决定了边缘分布和条件分布哪个应该被优先考虑,其取值直接决定识别精度,图4详细展示了平衡因子适配效果(以B工况到A工况的迁移为例)。

根据图4(a)可知,B工况样本边缘概率分布与A工况相差较小,并不是影响适配的主因。故针对本组数据,条件分布的重要性要高于边缘分布,应设置 $\mu > 0.5$ ,适配结果如图4(b)所示。结合图4(d)验

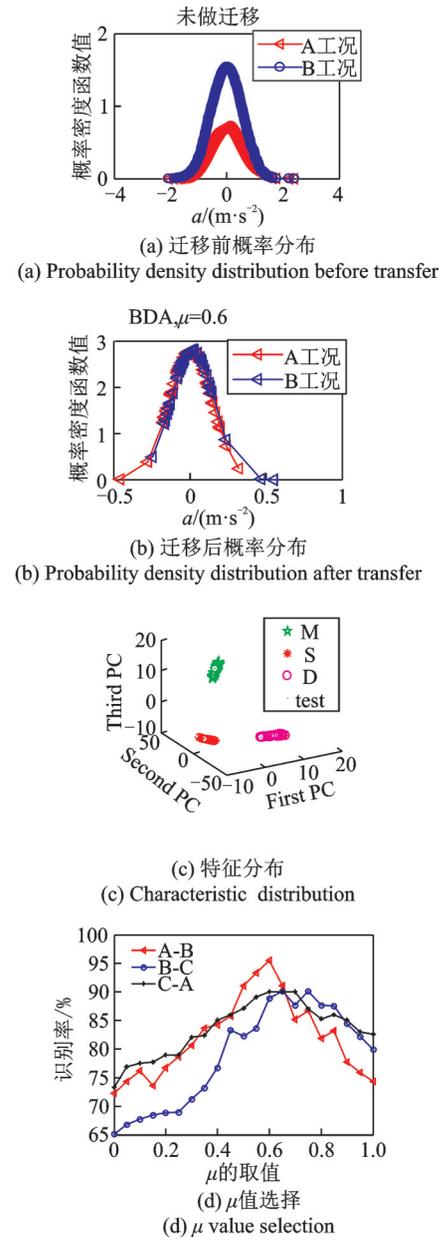


图4 适配效果图

Fig. 4 Adaptation Effect renderings

证得知 $\mu$ 值在0.5-0.7之间时,识别率均达到90%以上,在0.6时取得最优值,识别准确率达到97.78%。最终分类结果通过t-SNE可视化如图4(c),其中M,S和D分别代表3种寿命阶段,test代表测试样本。所提识别方法明显区分了3种寿命阶段样本。

### 3.4 与传统机器学习降维方法对比

为了验证改进均衡分布适配方法降维后可以很好地保留数据样本本身属性,将本文方法与传统降维方法进行对比。结合文献[13-14],本实验中SVM均采用高斯核函数,宽度为1;PCA与KPCA均采用径向基核函数,核参数 $\gamma=15$ ,BDA采用迭代次数 $N=20$ ,正则化参数 $\lambda=0.01$ ,平衡因子 $\mu=0.6$ 。

由图5可以看出,训练样本与测试样本维数从1

增加到 11 附近时,本文方法与 KPCA,PCA 结合 SVM 的 3 种方法的准确率不断提高。当维数超过 11 后,传统机器学习方法的识别率不高且伴随波动,而改进 BDA 方法的准确率总体保持平稳并略有提高,均明显高于 KPCA 和 PCA 两种方法。

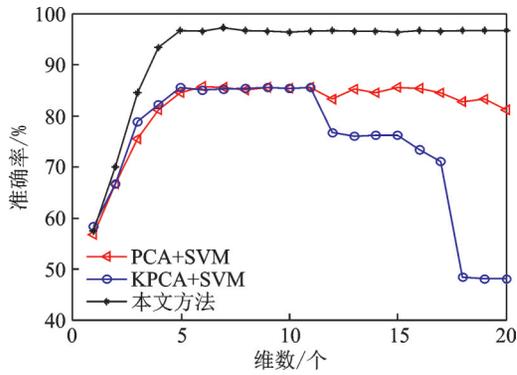


图 5 映射后的特征维数与准确率关系图

Fig. 5 Mapped characteristic dimension and accuracy diagram

造成这样的原因是 PCA、KPCA 作为传统机器学习的降维方法,它们在降维过程中不需要领域知识,忽略了领域间的差异。通过将两域中的所有样本映射到一个子空间来进行全局特征变换,没有考虑寿命阶段类别的内部关联性。这种全局特征变换方式,只能在一般超平面上学习两域异分布数据(超平面是松散的),导致其无法类内相关联,各寿命阶段样本内部不具关联性。

### 3.5 与其他迁移学习算法对比

为验证不同工况下,均衡分布适配在迁移上的优势,将改进均衡分布适配方法与迁移成分分析方法<sup>[10]</sup>(Transfer Component Analysis, TCA)、联合分布适配方法<sup>[15]</sup>(Joint Distribution Adaptation, JDA)、测地线流式核方法<sup>[16]</sup>(Geodesic Flow Kernel, GFK)等迁移学习方法对比。源域仍采用工况 B 的训练集,目标域采用工况 A 和 C 的测试集,对比结果如表 4 所示。

由表 4 可见, TCA 通过适配边缘分布,将全局

表 4 不同迁移学习算法对比测试结果

Tab. 4 Comparing test results of different transfer learning algorithms

不同方法	B-A	B-C	C-A	平均
TCA-SVM	62.36%	51.19%	53.33%	55.62%
JDA	62.64%	58.52%	65.56%	62.25%
BDA	74.67%	68.58%	71.11%	71.46%
GFK-SVM	89.79%	66.67%	68.47%	74.97%
本文方法	96.67%	88.89%	90.00%	91.85%

域特征进行变换,不能很好处理不同寿命阶段间数据的区分性。JDA 虽然考虑到两种分布对数据样本的影响,却没有根据具体数据衡量两种分布各自的重要性,泛化能力较差。GFK 将原始特征变换到流形空间,最近距离的选择高度依赖于流形核的构建,对复杂分布适应性较弱。本文方法通过自适应的适配两域权值,有效地解决了不同工况下滚动轴承寿命阶段识别的问题,平均识别精度达到 91.85%。特别是在 BDA 方法中,当  $\mu=0$  时, BDA 方法退化为 TCA; 当  $\mu=0.5$  时, BDA 退化为 JDA。这两种算法都可以看作是 BDA 的特殊情况。

## 4 角接触球轴承实验验证

### 4.1 角接触球轴承寿命状态识别

为验证本文构建的模型在不同型号轴承上的效果,现采用自测 C36018 型角接触球轴承,节径 15 mm,包含 7 个滚动体,接触角度  $15^\circ$ 。在相同转速、不同负载下运行相同圈数,分别以  $L1, L2$  和  $L3$  表示,在运行圈数和转速相同的条件下,负载越大,寿命损耗越多,故 3 种不同负载对应 3 个不同寿命阶段,实际寿命损耗为  $L1 < L2 < L3$ 。具体寿命状态信息如表 5 所示。

表 5 寿命状态数据信息

Tab. 5 Life state data information

寿命阶段	工作状态	运行圈数
$L1$	4 kg 负载	$1.44 \times 10^7$
$L2$	5 kg 负载	$1.44 \times 10^7$
$L3$	6 kg 负载	$1.44 \times 10^7$

采用不同工况条件对处于不同寿命阶段的滚动轴承进行振动信号采集,采集时运行转速分别为 500, 1000 和 1500 r/min,加载负荷均为 1 kg,采样频率均为 25.6 kHz,采样长度均为 102400,每种寿命阶段采样 2 次,对每种工况下各寿命阶段数据以 2048 为分析点数。1 kg 载荷、500 r/min 转速记为 D 工况,以 1 kg 载荷、1000 r/min 转速的采集工况数据记为 E 工况,以 1 kg 载荷、1500 r/min 记为 F 工况,各工况  $L1, L2$  和  $L3$  寿命阶段样本各 100 个。

设置 D 工况数据为源域数据, F、E 工况数据为目标域数据。从源域每类样本中分别抽取 1, 2, 3 个即训练样本 3, 6, 9 个(所占比例为整个源域寿命阶段样本数的 1%, 2%, 3%)。目标域每类样本中抽取 10, 20, 30 和 40 个,测试样本数分别为 30, 60, 90 和 120 个构成测试样本集。源域多样本数为 100 个。

表 6 不同工况下角接触球轴承寿命阶段识别结果

Tab. 6 Identification results of life stages of angular contact ball bearings under different working conditions

测试	测试 样本数	D 工况		
		3	6	9
E 工况	30	93.33%	90.00%	93.33%
	60	95.56%	95.00%	98.33%
	90	94.44%	91.11%	96.67%
	120	90.83%	94.17%	97.50%
F 工况	30	83.33%	96.67%	93.33%
	60	90.00%	93.33%	96.00%
	90	91.11%	92.50%	94.44%
	120	89.95%	98.33%	95.83%
平均		91.06%	93.89%	95.67%

采用本文方法对角接触球轴承不同工况下滚动轴承寿命识别,设置  $\mu=0.3$ ,  $N=20$ , 识别结果如表 6 所示。

从表 6 可以看出,本文方法在角接触球轴承不同工况下寿命阶段识别上表现良好。很好地适应了目标域数据完全无标记的适配问题,针对不同工况下分布差异较大的数据,平均识别精度最高可达 95.67%。

#### 4.2 与其他迁移学习算法对比

改进均衡分布适配方法与 TCA, JDA 和 GFK 等迁移学习方法对比。选择源域 3 样本为训练集(所占比例为整个源域寿命阶段样本数的 1%),目标域 60(所占比例为整个目标域寿命阶段样本数 60%)样本为测试集对比识别结果,得到如表 7 所示。

表 7 角接触球轴承数据下不同迁移学习算法对比结果

Tab. 7 Comparisons between different transfer learning algorithms and angular contact ball bearings

不同方法	D-E	D-F	F-E	平均
TCA-SVM	75.00%	71.67%	67.78%	71.48%
JDA	67.78%	65.00%	63.33%	65.37%
BDA	81.67%	78.33%	76.67%	78.89%
GFK-SVM	75.00%	83.33%	89.56%	82.63%
本文方法	95.56%	90.00%	98.33%	94.63%

本节实验利用不同工况下角接触球轴承的数据证明了改进均衡分布适配方法的可行性。无论是在与传统机器学习还是迁移学习方法的对比中都具有明显的优势。实验结果表明改进均衡分布适配方法很好的完成了滚动轴承寿命阶段识别的任务,识别

准确率达到 94% 左右。

## 5 结 论

1)改进均衡分布适配方法,采用无重复均匀随机抽样对源域类间样本进行多次均匀随机抽样得到多个训练样本,充分发掘了类间样本的潜在信息,减少了迭代造成的误差,更好地为无标记的目标域空间提供了可适配的伪标签;

2)在再生核希尔伯特空间,调节平衡因子  $\mu$  以适配两域边缘分布与条件分布所占权值,伴随定量随机抽样提高了数据类内紧凑性和类间区分性,大大提升了滚动轴承寿命阶段识别的精确率;

3)改进均衡分布适配方法与其他领域适应性方法对比识别结果。说明源域与目标域的边缘概率与条件概率在适配过程中各自的重要性不能被同等看待。改进均衡分布适配方法对不同工况下滚动轴承寿命阶段的识别效果更佳。

#### 参考文献:

- [1] Yu H, Li H R, Tian Z K, et al. Rolling bearing degradation state identification based on LPP optimized by GA [J]. International Journal of Rotating Machinery, 2016, 2016:9281098.
- [2] 陈仁祥,黄鑫,杨黎霞,等.加噪样本扩展深度稀疏自编码神经网络的滚动轴承寿命阶段识别[J].振动工程学报,2017,30(5):874-882.  
CHEN Renxiang, Huang Xin, Yang Lixia, et al. Bearing life state recognition using deep sparse auto-encoder neural network with noise adding sample expansion [J]. Journal of Vibration Engineering, 2017, 30 (5) : 874-882.
- [3] 王冰,王微,胡雄,等.基于GG模糊聚类的退化状态识别方法[J].仪器仪表学报,2018,39(3):21-28.  
WANG Bing, Wang Wei, Hu Xiong, et al. Degradation condition recognition method based on Gath-Geva fuzzy clustering [J]. Chinese Journal of Scientific Instrument, 2018, 39(3): 21-28.
- [4] Pan S J, Yang Q. A Survey on transfer learning [J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359.
- [5] 庄福振,罗平,何清,等.迁移学习研究进展[J].软件学报,2015,26(1):26-39.  
Zhuang Fuzhen, Luo Ping, He Qing, et al. Research progress of migration learning [J]. Journal of Software, 2015, 26 (1): 26-39.
- [6] 沈飞,陈超,严如强.奇异值分解与迁移学习在

- 电机故障诊断中的应用[J]. 振动工程学报, 2017, 30(1):118-126.
- SHEN Fei, CHEN Chao, YAN Ruqiang. Application of singular value decomposition and transfer learning in motor fault diagnosis[J]. Journal of Vibration Engineering, 2017, 30(1): 118-126.
- [7] 段礼祥, 谢骏遥, 王 凯, 等. 基于不同工况下辅助数据集的齿轮箱故障诊断[J]. 振动与冲击, 2017, 36(10):104-108.
- DUAN Lixiang, XIE Junyao, WANG Kai, et al. Gear-box fault diagnosis based on auxiliary data sets under different working conditions[J]. Journal of Vibration and Shock, 2017, 36(10): 104-108.
- [8] 康守强, 胡明武, 王玉静, 等. 基于特征迁移学习的变工况下滚动轴承故障诊断方法[J]. 中国电机工程学报, 2019, 39(3): 764-772.
- KANG Shouqiang, HU Mingwu, WANG Yujing, et al. Fault diagnosis method of rolling bearing under variable conditions based on feature transfer learning [J]. Proceedings of the CSEE, 2019, 39(3): 764-772.
- [9] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning [C]. IEEE International Conference on Data Mining, 2017.
- [10] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis.[J]. IEEE Transactions on Neural Networks, 2011, 22(2):199-210.
- [11] Li X, Lu W F, Zhai L, et al. Predictive modeling for life cycle reliability analysis and machine health condition prediction in remanufacturing[M]. London: Springer, 2014.
- [12] 阙子俊, 金晓航, 孙 毅. 基于UKF的轴承剩余寿命预测方法研究[J]. 仪器仪表学报, 2016, 37(9):2036-2043.
- Que Zijun, JIN Xiaohang, SUN Yi. Remaining useful life prediction for bearings with the unscented Kalman filter-based approach [J]. Chinese Journal of Scientific Instrument, 2016, 37(9): 2036-2043.
- [13] Guo S, Deng F, Jie C, et al. Sensor multi-fault diagnosis with improved support vector machines [J]. IEEE Transactions on Automation Science & Engineering, 2017, 14(2):1053-1063.
- [14] Deng X, Tian X, Chen S, et al. Deep learning based nonlinear principal component analysis for industrial process fault detection [C]. International Joint Conference on Neural Networks (IJCNN 2017). IEEE, 2017.
- [15] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation [C]. IEEE International Conference on Computer Vision, Sydney, NSW, 2013: 2200-2207.
- [16] Gong B, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation [C]. IEEE Conference on Computer Vision & Pattern Recognition, 2015.

## Life stage identification of rolling bearing based on improved BDA method

WU Hao-nian<sup>1</sup>, CHEN Ren-xiang<sup>1,2</sup>, HU Xiao-lin<sup>3</sup>, ZHANG Xia<sup>1</sup>, ZHANG Yan<sup>4</sup>, TANG Lin-lin<sup>1</sup>

(1. Chongqing Engineering Laboratory for Transportation Engineering Application Robot, Chongqing Jiaotong University, Chongqing 400074, China; 2. The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400030, China; 3. Chongqing Innovation Center of Industrial Big-Data Co. Ltd., Chongqing 400056, China; 4. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** In view of the problem that the distribution differences between training samples and test samples under different working conditions cannot effectively identify the life stage of rolling bearings, an improved method for identifying the life stage of rolling bearings based on balanced distribution is proposed. Firstly, non-repetitive uniform random sampling is used to conduct multiple uniform random sampling of inter-class samples in source domain, the training set of multi-sample in source domain is obtained to reduce the influence of sample selection in source domain on target domain prediction label. Furthermore, the weights of edge distribution and conditional distribution are dynamically adjusted in reproducing kernel Hilbert space by using equilibrium factor  $\mu$ , the weights of edge distribution and conditional distribution are continuously optimized by iteration. In order to reduce the maximum mean difference between the two domains, pseudo-labels in the target domain are transformed into pseudo-labels. Finally, multiple classifiers are constructed by using the mapping matrices of the source domain data sets, the final recognition results of the target domain samples are obtained by consistency discrimination. Experiments on two sets of data sets of rolling bearing life stages show that the proposed method is feasible and effective.

**Key words:** fault diagnosis; rolling bearing; life state identification; conditional probability distribution; marginal distribution

作者简介: 吴昊年(1993-),男,硕士。电话:(023)62539903;E-mail:296018167@qq.com

通讯作者: 陈仁祥(1983-),男,博士,教授,博士生导师。电话:(023)62539903;E-mail:manlou.yue@126.com