

2025 年中国研究生数学建模竞赛 E 题

高速列车轴承智能故障诊断问题

高速列车因其安全高效、便捷舒适、绿色低碳等优点，已经成为中国客运的主流运输工具。轴承作为高速列车走行系统的关键旋转部件，长期处于高转速、交变载荷等复杂恶劣的工况环境中，具有故障率高、易损坏等特点，是高速列车走行系统设备故障的主要源头。一旦轴承发生故障，轻则导致列车延误和晚点，重则会诱发脱轨等恶性事故，危及生命安全。

当前高铁轴承状态监测方法主要依赖于专家经验构建的特征指标或传统信号处理技术实现。随着轨道交通系统向高密度、强耦合、智能化的方向快速发展，现有方法在诊断精度、泛化能力和实时性方面难以满足复杂运营场景下的精准诊断需求。近年来，随着大数据、人工智能技术的快速发展，数据驱动的智能故障诊断方法在列车装备运维领域受到了广泛关注。基于海量运营数据构建的深度学习模型，能够实现更高精度的故障识别、更强的工况适应性和更高效的实时诊断能力。然而，在实际应用场景中，受复杂运行环境与多变工况的影响，传感器采集的原始振动信号易受到背景噪声、干扰源响应等诸多成分影响，这不仅削弱了故障特征的显著性，也直接制约了深度学习模型的识别准确率。更关键的是，为保障列车安全运营，关键设备的异常状态通常会被及时处理，导致在途列车的真实故障数据极度稀缺，进而造成训练数据分布严重失衡，极大地限制了深度学习模型向工程实践的转化。与实际运行数据相对的是，在台架实验环境下采集的轴承数据不仅数量丰富、标签完备，而且其故障演化机理也与在途列车轴承相似。在此背景下，迁移学习技术为解决上述问题提供了良好思路。

迁移学习技术是一种新兴机器学习方法，其核心思想是将在一个任务或领域中学到的知识迁移到另一个相关但不同的任务或领域，以提升目标任务的模型性能。因此，通过解析轴承故障机理对既有试验台数据展开分析，提取代表性的轴承故障特征以构建源域模型，再结合迁移学习技术将诊断知识迁移至实际运营列车数据，可有效缓解样本不平衡问题。轴承故障特征通常可以从时域、频域、时频域[1]和二维图像[2]等多个维度进行分析。

根据上述思路，本题给出 161 个具有代表性的轴承试验台架振动数据文件作为问题研究的源域数据集，另外给出来自实际的 16 个轴承故障文件作为目标域数据集，这些数据文件的相关信息见附件 1。请完成以下几项任务：

1、数据分析与故障特征提取：请考虑目标域的迁移任务，从提供的源域数据中筛选部分数据组成数据集。结合轴承故障机理，选择合适的方法或指标对有代表性的源域数据进行特征分析，并对整体数据集进行特征提取，用于后续诊断任务。（源域数据可选用其他公开轴承数据集，但应标明出处）。

2、源域故障诊断：在任务 1 提取的故障特征基础上，划分源域训练集与测试集。设计合适的诊断模型实现源域诊断任务，并对诊断结果进行评价。

3、迁移诊断：在任务 2 设计的诊断模型基础上，充分考虑源域与目标域的共性与差异特征，设计合适的迁移学习方法，构建目标域诊断模型，对目标域未知标签的数据进行分类和标定，给出迁移结果的可视化展示和分析，并给出数据对应的标签。

4、迁移诊断的可解释性：可解释性是机器学习领域的重要研究方向之一。由于机器学习模型的“黑箱”问题，其迁移和诊断过程难以被观测和理解，这可能造成使用者对模型结果的不信任或盲目信任，进而影响诊断模型的应用。迁移诊断可解释性研究的核心目标是解决迁移学习模型在跨工况、跨设备故障诊断中的透明性问题，提高诊断人员对迁移过程和诊断模型输出的理解和信任度。请考虑任务 3 中模型的结构设计、迁移过程和决策过程，结合轴承故障特点与故障机理，对迁移诊断的事前/迁移过程/事后（任选一点或多点）可解释性进行分析。

名词解释：

- 1) **事前可解释性：**模型本身具有透明的结构和决策逻辑，即“自解释”能力，一般无需额外的解释技术即可直接理解其工作原理。
- 2) **迁移过程可解释性：**揭示知识从源域到目标域的迁移路径和共享特征情况。分析知识是否真的实现了迁移，说明模型如何适应新工况/设备。可通过结构设计或外部工具进行分析。
- 3) **事后可解释性：**通过外部工具或算法反推模型的决策依据，以人类可理解的方式展现模型的决策过程和机制。

附件 1：数据集介绍

1. 源域数据集（轴承试验台架振动数据，存在轻微噪声影响）

试验台架主体为电动机，其主要包括驱动端、风扇端和基座三部分。数据通过安装在电动机壳体驱动端、风扇端和基座上的加速度传感器采集，驱动端和风扇端各存在一个试验轴承。轴承由人为破坏生成单点故障（即只有单个部位存在单一缺损），不存在复合故障，且每组数据仅有一个轴承损坏。故障的直径有四种大小，具体情况在后文中详述。

不同位置传感器采集的振动信号能够反映不同的物理特征，从而为不同诊断任务提供数据支持。

驱动端数据（DE）：驱动端直接连接电机转轴，从此处采集的振动信号主要受转子旋转和传动系统的激励影响，能够清晰捕获驱动端轴承的振动信号，以及风扇端轴承经过转轴传递后的振动信息。

风扇端数据 (FE)： 风扇端即电机风扇端，从此处采集的振动信号主要受叶片旋转和风扇系统的激励影响，能够清晰捕获风扇端轴承的振动信号，以及驱动端轴承经过转轴传递后的振动信息。

基座数据 (BA)： 基座即电机底座，从此处采集的振动信号受到电机整体结构和运行状态的影响，能够反映整个电机系统的振动信息。由于多层结构传递，振动信号故障特征高度衰减，通常用于辅助分析。

一般来说，距离故障轴承较近的采集位置振动传递路径短，故障信号的幅值更明显（如驱动端采样驱动端故障），远端则由于传递路径较长而产生衰减（如驱动端采样风扇端故障）。

1) 试验平台轴承信息

- 待检测轴承支撑电动机转轴；
- 驱动端轴承为 SKF6205，采样频率为 12KHz 和 48KHz；
- 风扇端轴承为 SKF6203，采样频率为 12KHz。

轴承尺寸参数如表 1 所示。

表 1 源域数据集轴承尺寸参数

轴承类型	滚动体数 n	滚动体直径 d	轴承节径 D
SKF6205 (DE)	9	0.3126 英寸	1.537 英寸
SKF6203 (FE)	9	0.2656 英寸	1.122 英寸

2) 数据格式和变量名称

数据文件为 MATLAB 的 .mat 格式，每个文件包含的数据种类不一致。变量名解释如下：

- DE: drive end accelerometer data 驱动端加速度数据；
- FE: fan end accelerometer data 风扇端加速度数据；
- BA: base accelerometer data 基座加速度数据；
- time: time series data 时间序列数据；
- RPM: rpm during testing 转/每分钟，除以 60 为旋转频率。

3) 轴承工作状态类别

在该数据集中，轴承有 4 种工作状态：外圈故障（OR）、内圈故障（IR）、滚动体故障（B）以及正常工作状态（N）。其中外圈故障样本 77 个，内圈故障样本 40 个，滚动体故障样本 40 个，正常样本 4 个。轴承外圈故障有 3 种故障尺寸：0.007、0.014、0.021 英寸；轴承内圈故障和滚动体故障都有 4 种故障尺寸：0.007、0.014、0.021、0.028 英寸。由于外圈故障位置固定（不随轴承旋转而移动），因此通过三个点位全面采集振动信号：3 点钟（Orthogonal）、6 点钟（Centered）、12 点钟（Opposite）。轴承的载荷有四种情况：0、1、2、3 马力。以 12kHz 下命名为“B007_0”的数据文件为例，其表示 0 载荷下 0.007 英寸的滚动体故障（B）。其导入 MATLAB 的数据组成包括“X118_BA_time”，“X118_DE_time”，“X118_FE_time”和“X118RPM”四组，其中 X118 表示数据编号，BA/DE/FE 为基座/驱动端/风扇端采样信号，RPM 为该数据的轴承转速。

如上所述，该数据集包含足够类别的通用轴承振动数据，可以根据目标需求，筛选合适的数据进行分析和特征提取，作为源域数据集对模型进行训练。

2. 目标域数据集（列车轴承故障数据集）

该数据集包含列车滚动轴承外圈（OR）、内圈（IR）、滚动体（B）故障和正常状态（N）下的振动信号数据，采集时间为 8 秒，采样频率为 32kHz，列车速度约 90km/h（轴承转速约 600 rpm）。数据文件以英文字母 A~P 编号命名，各数据所属工作状态未知。

3、源域和目标域数据集链接：

<https://pan.baidu.com/s/1H8nHXyMTv085jGRaiBOr0Q?pwd=anlh>

提取码: anlh

附件 2：概念介绍

1. 轴承故障诊断与振动监测数据

列车轴承的工作环境往往较为恶劣，在长时间运行过程中，轴承部件可能因过热、润滑不良、腐蚀等多种因素发生损坏，进而影响列车正常运行。轴承故障诊断是指通过监测轴承运行状态下的各类信号（其中振动加速度信号为主要监测数据），运用数据挖掘、信号处理及机器学习等技术手段，判定轴承是否存在故障的技术。从本质上讲，基于机器学习的故障诊断问题属于模式识别问题——通过信号处理等技术提取故障特征，再借助机器学习模型实现精准分类，是该领域常见的实现路径。其核心目标在于实现对故障的提前预警与精准诊断，从而避免设备突发性失效，为列车的安全运行提供可靠保障。

轴承结构主要包括内圈、外圈、滚动体和保持架四部分，其中典型故障多发生在内圈、外圈和滚动体这三个核心承载部件。当轴承出现局部缺陷时，滚动体

在接触并通过缺陷点的瞬间，会产生突变的冲击脉冲，而在轴承周期性运转的过程中，这种脉冲力会持续作用，进而形成周期性的冲击分量。在传感器采集的振动信号中，由缺陷引发的周期冲击分量呈现为一系列类周期性的振动冲击及衰减响应，示意图如图 1 所示。

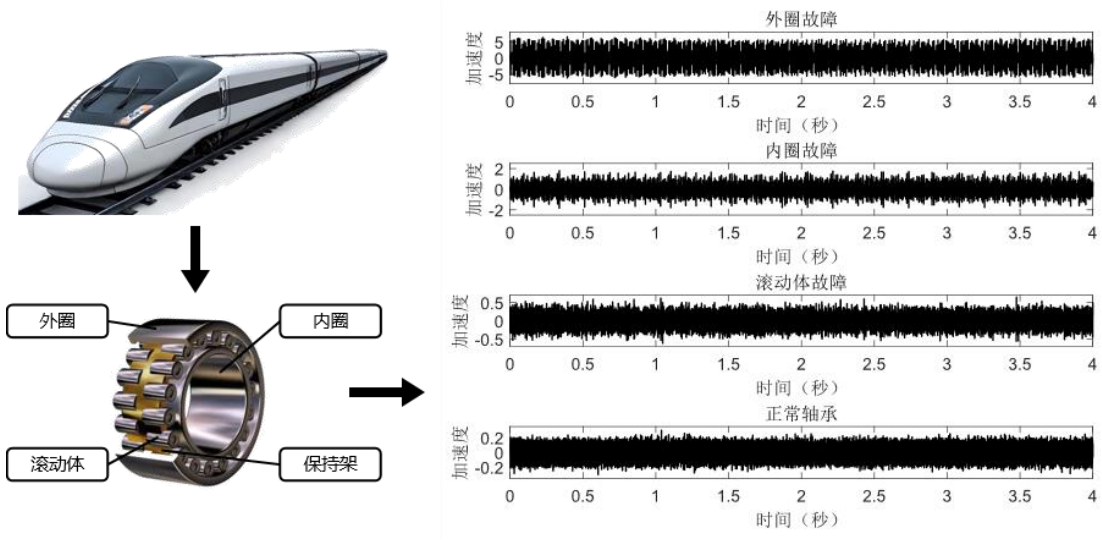


图 1 列车轴承结构及轴承典型故障数据示意

这类轴承故障的频率特性通常由故障特征频率来描述，该频率可通过轴承转速、滚动体数量、滚动体直径等固有参数计算得出，具体计算公式如表 2 所列。

表 2 轴承故障特征频率

名称	特征频率
外圈故障特征频率 (Ball Pass Frequency on the Outer Race, BPFO)	$f_r \cdot \frac{N_d}{2} \cdot (1 - \frac{d}{D})$
内圈故障特征频率 (Ball Pass Frequency on the Inner Race, BPFI)	$f_r \cdot \frac{N_d}{2} \cdot (1 + \frac{d}{D})$
滚动体故障特征频率 (Ball Spin Frequency, BSF)	$f_r \cdot \frac{D}{d} [1 - (\frac{d}{D})^2]$

表中， $f_r = \frac{n}{60}$ 为轴承转频， n 为轴承内圈转速（单位：rpm）， d 为滚动体直

径， D 为轴承节径（通常可由 $\frac{\text{内径}+\text{外径}}{2}$ 估算）， N_d 为滚动体个数，轴承结构示意图如图 2 所示。

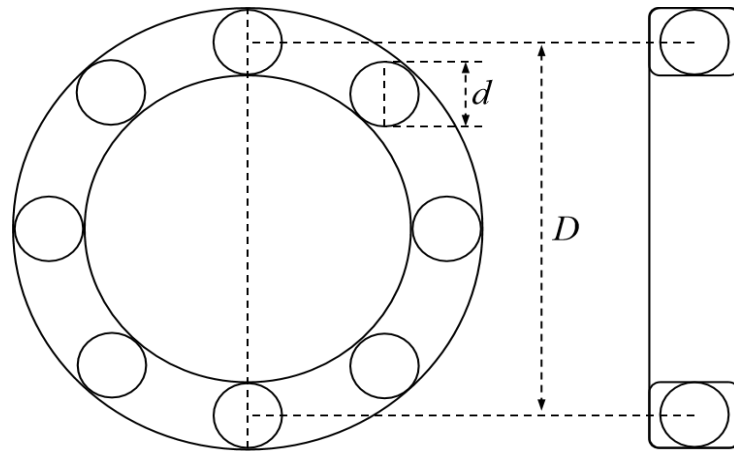


图 2 轴承结构和参数示意图

常见的轴承数据集中，一般涵盖内圈故障、外圈故障和滚动体故障三种典型故障类型，为故障诊断算法的训练与验证提供了基础数据支撑。

由于结构功能与受力特点存在差异，不同部位发生故障时，其振动信号呈现出不同的特征，具体表现如下：

1) 外圈故障

从时域上分析，轴承外圈故障信号如图 3 所示。

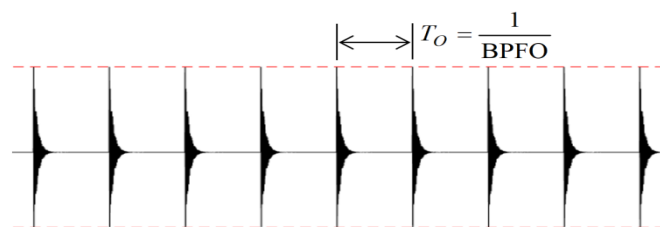


图 3 外圈故障示意图

图中 T_o 为故障周期（相邻冲击的时间间隔），BPFO 为外圈故障特征频率，二者互为倒数。由于外圈通常固定不动，在载荷方向不变的情况下，外圈缺陷每次碰撞到滚动体的强度近似相同，因而理想情况下在时域上呈现等周期等幅值的冲击分量。

2) 内圈故障

由于内圈随轴旋转，当载荷方向不变时，故障点将周期性通过载荷区，导致每次碰撞的强度不同。具体表现为振动信号的波形受到转频 f_r 的调制，具有变化的幅值。如图 4 所示。

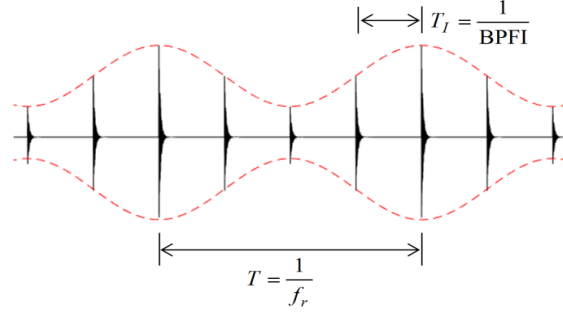


图 4 内圈故障示意图

图中 T_I 为故障周期，BPFI 为内圈故障特征频率， T 表示轴承的旋转周期，为转频的倒数。

3) 滚动体故障

与内圈故障相似，在载荷方向不变时，故障点和载荷的相对位置随滚动体的旋转而发生周期性变化，表现为振动信号波形受到滚动体公转频率的调制。同时，滚动体缺陷将随着滚动体自旋与内外圈发生交替碰撞，如图 5 所示。

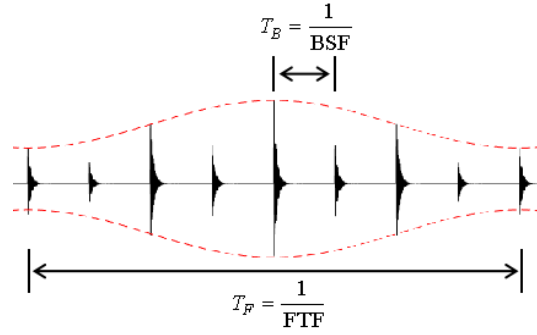


图 5 滚动体故障示意图

图中 T_B 为故障周期，BSF 为滚动体故障特征频率，二者互为倒数； T_F 为滚动体公转周期，FTF 为滚动体公转频率，一般可通过 $\text{FTF} = \frac{1}{2} \cdot f_r [1 - \frac{d}{D}]$ 计算，二者互为倒数。

一般来说，结合轴承故障机理，通过统计学方法和信号处理技术，能够有效降低背景噪声影响，有助于在采样信号中提取故障特征，从而提高深度学习模型的学习效率和分类准确性。

2. 迁移学习

迁移学习 (Transfer Learning) 是一种机器学习方法，其利用在源域 (源任务) 学习到的知识来改善目标域 (目标任务) 的学习效果，以减少目标域所需的训练数据量，并提高学习效率。迁移学习包括两个重要概念：

1) 领域(域): 模型学习的主体, 主要由数据和其概率分布组成。本题包括两个领域, 即轴承试验台数据及其分布(源域)和实际运营列车数据及其分布(目标域)。

2) 任务: 任务是模型在领域内要完成的具体目标, 其数学本质是通过构建一个预测函数来实现输入空间到标签空间的映射关系。其中, 输入空间表示所有可能的输入样本组成的集合, 标签空间表示模型需要预测的输出值(标签)的集合。本题中, 具体任务包括轴承试验台数据的分类任务(源任务)和实际运营列车数据的分类任务(目标任务)。

具体来说, 给定一个源域 D_S 和一个目标域 D_T , 以及对应的任务 T_S 和 T_T , 迁移学习的目标是找到一个映射函数 $h \in H$ (H 为假设空间), 使得 h 在目标域 D_T 上的预测误差最小化。常见的迁移学习方法包括基于特征的迁移、基于模型的迁移、基于关系的迁移和基于样本的迁移四种类型。具体介绍如下:

1) 基于特征的迁移: 将源域和目标域数据映射到统一的特征空间, 减少领域间分布差异, 进而实现迁移。常用方法包括特征转换[3]、对抗训练[4]等。

2) 基于模型的迁移: 复用源域模型的参数或结构, 使其适配目标任务。常用方法包括参数共享和微调[5]等。

3) 基于关系的迁移: 将源域中数据之间的关系模式迁移到目标域, 如相似性、依赖关系等。常用方法包括关系类比[6]、马尔可夫逻辑网络[7]等。

4) 基于样本的迁移: 通过调整源域样本的权重实现迁移。与目标域相似的样本赋予高权重, 使其发挥更大作用, 不相似的样本则相反。常用方法包括重要性加权、实例选择[8]等。

迁移学习的优势在于, 其能够显著减少目标域的数据和标注需求, 并能够提高模型训练速度和泛化能力, 以及解决无历史数据时的冷启动问题等。由于本题的目标域无标签, 请选择适用的迁移学习方法完成任务。

提示:

1) 论文需要提交对应的源代码和数据集;

2) 若选择其他源域数据, 需在报告中给出该数据集的简要介绍、相关链接和所选数据情况。

参考文献

- [1] 陈强强, 戴邵武, 戴洪德, 等. 滚动轴承故障诊断方法综述[J]. 仪表技术, 2019(9): 1-4+42. DOI: 10.19432/j.cnki.issn1006-2394.2019.09.001.
- [2] 张安安, 黄晋英, 冀树伟, 等. 基于卷积神经网络图像分类的轴承故障模式识别[J]. 振动与冲击, 2020, 39(4): 165-171. DOI: 10.13465/j.cnki.jvs.2020.04.021.
- [3] 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+955. DOI: 10.13334/j.0258-8013.pcsee.180130.
- [4] 洪晓翠, 段礼祥, 徐继威, 等. 变工况下轴承故障的残差对抗网络诊断方法[J]. 石油机械, 2022, 50(5): 32-42. DOI: 10.16082/j.cnki.issn.1001-4578.2022.05.005.
- [5] Zhang, R., Tao, H., Wu, L., & Guan, Y. (2017). Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. IEEE Access, 5, 14347-14357. <https://doi.org/10.1109/ACCESS.2017.2720965>
- [6] 郭敏, 陈鹏, 周超, 等. 基于坐标注意力关系网络的小样本轴承故障诊断[J]. 四川大学学报(自然科学版), 2024, 61(4): 338-346. DOI: 10.19907/j.0490-6756.2024.047001.
- [7] Davis, J., & Domingos, P. (2009). Deep transfer via second-order Markov logic. Proceedings of the 26th Annual International Conference on Machine Learning (pp. 217-224). ACM.
- [8] 戴文渊. 基于实例和特征的迁移学习算法研究[D]. 上海: 上海交通大学, 2008.