

# 硕士学位论文

## 基于动力学仿真和迁移学习的滚动轴承故障 诊断方法研究

### INVESTIGATION OF DYNAMIC SIMULATION AND TRANSFER LEARNING BASED FAULT DIAGNOSIS METHOD FOR ROLLING ELEMENT BEARINGS

董韵佳

哈尔滨工业大学

2019 年 6 月



Classified Index: TP277

U.D.C: 621

Dissertation for the Master Degree in Engineering

**INVESTIGATION OF DYNAMIC SIMULATION  
AND TRANSFER LEARNING BASED FAULT  
DIAGNOSIS METHOD FOR ROLLING ELEMENT  
BEARINGS**

<b>Candidate:</b>	Dong Yunjia
<b>Supervisor:</b>	Prof. Xu Minqiang
<b>Academic Degree Applied for:</b>	Master of Engineering
<b>Speciality:</b>	Mechanics
<b>Affiliation:</b>	School of Astronautics
<b>Date of Defence:</b>	June, 2019
<b>Degree-Conferring-Institution:</b>	Harbin Institute of Technology

## 摘 要

滚动轴承是旋转机械中最重要的部件之一。滚动轴承发生故障会影响机械设备的正常运转，严重时会引起设备的失效，造成巨大的时间损失、财产损失甚至人员伤亡。为了提高机械设备的安全性和可靠性，研究滚动轴承的故障诊断技术十分必要。近年来，随着人工智能技术的蓬勃发展，滚动轴承的智能诊断技术受到广泛关注。许多智能诊断方法在实验室验证时，在利用特定设备的大量历史数据的前提下获得了很好的诊断效果。然而，在实际工程中，很难事先获取待诊断设备的充足的故障历史数据，在少量可获取的数据上构建诊断模型会造成诊断模型泛化能力的下降。本文针对滚动轴承故障诊断中的小样本问题，结合卷积神经网络，提出基于动力学仿真和迁移学习的智能诊断方法，该方法以从滚动轴承动力学模型获取的仿真数据中迁移诊断知识的方式，帮助实际诊断场景中的轴承诊断任务，为解决小样本条件下的故障诊断问题提供了新的思路。

首先，基于动力学建立滚动轴承故障仿真模型，获取不同健康状态下滚动轴承的振动仿真数据。滚动轴承的动力学模型能够从理论上揭示滚动轴承的振动特性，由动力学模型获取的仿真数据与实际数据应具有相似但不同的特征分布，通过滚动轴承动力学模型生成仿真信号是一种获取各种型号轴承在各种工况下的振动数据的有效方式。

其次，分析仿真数据与实际数据的共性与差异性，在此基础上，提出小样本条件下，基于卷积神经网络，将仿真数据中的诊断知识迁移到实际数据中的滚动轴承智能诊断方法框架。该方法的基本思想是用大量仿真时域信号预训练卷积神经网络，随后将网络参数迁移到实际诊断网络中并用可获取的少量实际数据微调。

最后，将本文所提出的方法应用在两个实际诊断问题中，并设计多个对比实验，验证方法的优越性。另外，基于这两个诊断实例，对诊断方法的诊断机理进行详细分析，说明本文方法可以学习到更多有效的故障特征并缩小特征分布差异，使得小样本条件下的实际问题诊断性能有了显著提高。

**关键词：**滚动轴承；智能故障诊断；迁移学习；深度学习；动力学仿真；  
卷积神经网络

## Abstract

Rolling element bearings are one of the most critical components in rotating machinery. Bearing faults would affect the equipment operation and cause the whole rotary machines failures, leading to the huge property damage, time-consuming and casualties. In order to enhance the security and reliability of machinery equipment, it is essential to focus on the condition monitoring and fault diagnosis of rolling element bearings. Intelligent fault diagnosis of rolling element bearings gains increasing attention in recent years due to the promising development of artificial intelligent technology. Many intelligent diagnosis methods work well based on massive historical data for a certain equipment. However, it is hard to get sufficient fault data in advance in real diagnosis scenario and diagnosis model established on such small dataset suffers from serious loss of generalization. Focus on the small sample problem in bearing fault diagnosis, this paper proposes a new intelligent fault diagnosis method that transfers the diagnosis knowledge from simulation data to real scenario based on convolutional neural network (CNN). In the proposed method, the diagnosis knowledge of simulation data is used to help real scenario diagnosis. The research in this paper provides a feasible approach to solve lack-data problem in fault diagnosis.

First of all, dynamic model of rolling element bearings is established to generate massive simulation vibration data. Dynamic modeling of bearing is an effective and cheap way to acquire bearing data with different bearing type and working condition, which can reveal the characteristics of the bearing vibration signals and simulation data should share related but different feature distributions with real data.

Then, the similarity and discrepancy between simulation data and real data are analyzed. After that, the frame of the deep transfer fault diagnosis method based on CNN that transfers diagnosis knowledge from simulation model to real scenario is proposed. The basic idea of the method is pre-training a basic CNN on raw simulation data, and then the CNN parameters learned in pre-training are transferred to real diagnosis tasks with fine-tuning the whole CNN by little real data.

Finally, the proposed method is utilized on two diagnosis cases, the superior performance of the proposed method is verified by comparing with different diagnosis experiments. Furthermore, the inner mechanism of the method is analyzed in detail,

which shows that the proposed method can learn more transferable features and reduce the feature distribution discrepancy, contributing to enhancing the fault identification performance significantly.

**Key words:** rolling element bearings, intelligent fault diagnosis, transfer learning, deep learning, dynamic simulation, convolutional neural network

# 目 录

摘 要 .....	I
Abstract .....	II
第 1 章 绪 论 .....	1
1.1 课题背景及研究的目的和意义 .....	1
1.2 国内外研究现状 .....	3
1.2.1 滚动轴承故障仿真模型研究现状 .....	3
1.2.2 基于数据驱动的智能故障诊断方法研究现状 .....	4
1.2.3 国内外研究现状分析 .....	8
1.3 本文的主要研究内容 .....	9
第 2 章 基于动力学的滚动轴承故障仿真模型 .....	11
2.1 引言 .....	11
2.2 滚动轴承动力学模型 .....	11
2.2.1 滚动轴承动力学方程 .....	11
2.2.2 Hertz 接触刚度计算 .....	13
2.2.3 滚动轴承接触变形计算 .....	15
2.3 滚动轴承滚道局部故障建模 .....	16
2.4 滚动体随机滑动因素的引入 .....	18
2.5 滚动轴承动力学仿真模型的验证与分析 .....	19
2.5.1 西储大学轴承实验 .....	19
2.5.2 仿真模型参数设置 .....	20
2.5.3 正常状态下仿真模型的验证与分析 .....	21
2.5.4 故障状态下仿真模型的验证与分析 .....	22
2.6 本章小结 .....	23
第 3 章 基于动力学仿真和迁移学习的滚动轴承智能故障诊断方法 .....	24
3.1 引言 .....	24
3.2 仿真数据与实际数据的共性及差异性分析 .....	24
3.2.1 动力学角度的共性及差异性 .....	24
3.2.2 概率分布的共性及差异性 .....	25
3.2.3 频率分布的共性及差异性 .....	27
3.2.4 仿真数据与实际数据的共性及差异性总结 .....	36

3.3 基于动力学仿真和迁移学习的滚动轴承故障诊断框架 .....	37
3.3.1 卷积神经网络概述 .....	37
3.3.2 故障诊断问题设定 .....	40
3.3.3 基于动力学仿真和迁移学习的滚动轴承故障诊断流程 .....	41
3.4 本章小结 .....	43
<b>第 4 章 基于动力学仿真和迁移学习的滚动轴承故障诊断方法应用与诊断机理分析 .....</b>	<b>44</b>
4.1 引言 .....	44
4.2 源域和目标域对象描述及数据集构建 .....	44
4.2.1 目标域对象描述及数据集构建 .....	45
4.2.2 源域对象描述及数据集构建 .....	45
4.3 故障诊断测试验证场景及主要参数设定 .....	48
4.3.1 对比任务设定 .....	48
4.3.2 CNN 参数设定 .....	49
4.3.3 CNN 训练过程 .....	50
4.3.4 诊断模型性能评估指标 .....	50
4.4 故障诊断结果及分析 .....	51
4.4.1 SKF-6205 轴承诊断结果 .....	51
4.4.2 SKF-6203 轴承诊断结果 .....	53
4.5 基于动力学仿真和深度迁移学习的故障诊断方法机理分析 .....	55
4.5.1 混淆矩阵分析 .....	55
4.5.2 卷积核迁移机理分析 .....	57
4.5.3 全连接层迁移机理分析 .....	61
4.5.4 诊断方法的诊断机理总结 .....	63
4.6 本章小结 .....	63
<b>结 论 .....</b>	<b>65</b>
<b>参考文献 .....</b>	<b>67</b>
攻读硕士学位期间发表的学术论文 .....	73
哈尔滨工业大学学位论文原创性声明和使用权限 .....	74
<b>致 谢 .....</b>	<b>75</b>



# 第 1 章 绪 论

## 1.1 课题背景及研究的目的和意义

随着科学技术的不断发展,现代机械设备的复杂化和综合化程度不断提高,智能化及自动化水平日益增强,这些变化为设备的健康监测和维修管理提出了更高的要求<sup>[1-3]</sup>。滚动轴承是机械设备中的关键部件,具有易于润滑、运行精度高、摩擦阻力小等特点<sup>[4]</sup>,常用于高速、轻载情况下,具有支撑轴及承受载荷的功能,广泛应用于船舶、航空航天等多个工业领域。作为轴的支撑部件,滚动轴承的正常运转对保证机械设备的平稳运行起到至关重要的作用。但由于滚动轴承的抗冲击性能较差,当机械系统尤其是旋转机械发生故障时,通常与滚动轴承有密切联系。因此,研究滚动轴承的故障诊断技术,对提高机械设备的安全性和可靠性,保证工业生产的效率,提高经济效益具有重要的现实意义。

故障诊断技术主要包括三个层面的任务,一是检测设备是否异常,即是否发生故障;二是故障隔离,即确定故障发生的位置;三是故障类型的识别。实现故障诊断的过程实际上从测量空间或特征空间向故障模式空间映射的模式识别过程<sup>[2]</sup>。传统的识别过程很大程度上依赖人类专家的先验知识<sup>[3]</sup>,如基于模型的诊断方法,需要根据故障机理建立系统的物理模型,进一步通过实际输出与期望输出的偏差实现故障诊断。这种诊断的准确性取决于系统建模的准确性,而对于复杂系统,想要建立准确的物理模型是非常困难的。再如基于信号处理的诊断方法,一般通过信号在正常情况与故障情况下相关图谱或相关信号特征量体现出的故障征兆实现故障诊断。同样这种方法适合于信号波形较简单,特征较明确的系统,对于信号成分多样,信噪比低,难以提取有效信号特征的复杂系统,利用该方法识别也很困难。以上两种方法均需要人类专家利用丰富的经验与领域知识进行人工识别,面对越来越复杂的系统和冗余性越来越强的信息,仅用手动识别已经难以满足要求,因此研究智能故障方法尤为重要。

智能故障诊断方法是基于知识的数据驱动诊断方法,它不再过度依赖显式的系统物理模型或信号特征等人类专家知识,而更多依赖于从系统大量的历史数据中提取的隐性知识<sup>[3]</sup>。基于知识的数据驱动诊断方法包括定性方法和定量方法,典型的定性方法是故障诊断专家系统,而典型的定量方法是基于机器学习的智能诊断方法。智能诊断方法的实施主要包括三步<sup>[2,5]</sup>:数据获取,特征提取及选择和故障识别。在传统的基于机器学习的故障诊断方法中,特征提取过程通常仍由人

类专家完成，而特征选择和故障识别可由机器学习方法实现，后两步在一定程度上实现了诊断的自动化，而在特征提取方面，当数据量很大或数据所包含的信息十分复杂时，仅依赖人类专家提取特征不仅耗费人力和时间，还易造成信息的丢失。为了同时实现自动化的特征提取和故障识别，基于深度学习的故障诊断方法受到广泛关注。基于深度学习的诊断方法利用深度神经网络实现层次化的非线性特征学习，特征提取过程和故障识别过程在一个诊断模型中自动化地完成，是一种更高层次的智能诊断方法。然而，传统的机器学习方法和深度学习方法的前提假设是训练数据与测试数据具有相同的概率分布，为了满足这个要求，这些智能故障诊断方法在操作时，通常将采集到的同一设备同工况下的数据按一定比例分成训练数据和测试数据，在训练数据上训练得到故障诊断模型，再在测试集上测试，验证诊断模型的可行性。这样的实施方式虽然可以获得很好的识别率，但与实际的诊断问题设定不甚相符。第一，实际的诊断问题应是将待诊断设备新采集到的数据输入到诊断模型中，输出诊断结果，因此理论上，在训练诊断模型的数据中必须包含待诊断设备所发生故障的所有可能类型，但实际上很难获取大量的待诊断设备在特定工况下各种类型的故障样本，若诊断模型仅在所能获取的少量样本上训练，则很容易发生过拟合现象，泛化能力较差，即模型在训练数据上有效，但在测试数据上性能显著降低；第二，这种数据集的划分只能证明诊断模型在该数据集所能体现的特定设备特定工况特定健康状态下的可行性，对于同类其他设备的其他工况，模型可能并不具有泛化能力，这可能是目前传统智能诊断方法在实际中难以推广的重要原因之一。

基于以上背景，本课题拟以滚动轴承为对象，针对实际诊断问题中样本量很少，不足以训练稳定的故障识别模型的小样本问题，通过动力学仿真的方式模拟轴承的故障振动信号，得到更丰富的故障数据，再基于深度迁移学习方法，利用卷积神经网络和参数迁移策略，从仿真时域信号中自动化地学习故障诊断知识，并将其迁移到实际的诊断问题中，实现小样本条件下对实际故障轴承的有效诊断。本文的研究首先通过结合动力学仿真模型和迁移学习的思想，一定程度上解决了传统数据驱动诊断方法尤其是深度学习对大量有标签样本的依赖，为解决小样本情况下的轴承诊断问题提供了新的可行途径；其次知识迁移策略在诊断领域的应用对实现智能诊断方法的实际应用，同时对增强设备的安全性和可靠性，提高经济效益具有积极的意义。

## 1.2 国内外研究现状

### 1.2.1 滚动轴承故障仿真模型研究现状

滚动轴承主要由内圈、外圈、滚动体和保持架四部分构成，其工作时，承载区内各个接触点的载荷大小不同。对于每个接触点，当滚动体经过时，便承受一次载荷，且每次承受的载荷大小固定，即受到稳定的脉动循环载荷作用。这种长期的交变载荷会导致轴承产生疲劳裂纹，当裂纹进一步扩展，会导致内外圈产生剥落坑，剥落坑在轴承运转中引发的冲击载荷使轴承振动增强，噪声加剧，这也是导致轴承失效的最主要原因<sup>[4]</sup>。为了从理论上描述这种故障引发的冲击对轴承振动信号的影响，很多研究者建立了不同的故障仿真模型，以得到轴承在故障冲击下的振动响应。最简单的轴承故障仿真模型由 McFadden 和 Smith 建立，模型中将滚动轴承系统简化为一个单自由度弹簧—阻尼系统，将故障冲击看成一系列周期脉冲信号，基于此模型描述了当滚动轴承具有内圈故障时的振动信号指数衰减的故障信号形式，并描述其解调谱特征<sup>[6]</sup>；基于这种信号特性仿真的思想，Aktürk 等将轴承滚道的表面波纹度缺陷抽象成具有波动性的周期激励，探究其对轴承振动信号的影响<sup>[7]</sup>；文献[8]指出，实际中故障冲击的周期并不是恒定的，因此在周期中加入随机因素，并得到相应的故障信号；文献[9]研究了不同故障冲击形式下（包括矩形冲击、三角冲击及半正弦冲击）的轴承故障信号。这些将故障简化为冲击信号，进行信号仿真的方法，可以定性直观模拟轴承故障信号特性，但将轴承直接看成一维的单自由度系统，忽略内外圈及各个部件之间的接触关系，显得过于简化。

另一种建模方式是从动力学角度出发，基于 Hertz 理论，考虑轴承各个部件之间的接触关系及故障模型，将各个部件简化成一系列集中质量，进行轴承的动力学建模。这种建模方法首先建立轴承在正常状态下各个部件的运动微分方程，在此基础上，注入故障模型，实现故障条件下的仿真。Sassi 等<sup>[10]</sup>将轴承-轴承座的振动问题简化为一维三自由度线性有阻尼受迫振动问题，三个自由度分别为外圈、内圈及一个滚动体，轴承座简化为刚体，润滑油膜简化为弹簧-阻尼器，在径向载荷作用下建立轴承的三自由度运动微分方程，在此基础上，仍用周期性脉冲模拟单点故障，施加在具有故障的部件上，利用数值方法求解方程，得到径向载荷及故障冲击载荷共同作用时的动力学响应，实现故障仿真；Sawalhi 等<sup>[11]</sup>将轴承—轴承座简化为二维五自由度系统，考虑故障的几何形式，建立轴承在二维平

面下的仿真模型；Ahmadi 等<sup>[12]</sup>考虑了每个滚动体的自由度，利用非线性多体动力学的方法建立了具有  $2N_b + 6$ （ $N_b$  为滚动体个数）个自由度的滚动轴承的故障模型。基于动力学方程的数值仿真建模方法灵活性较强，复杂度根据模型对实际系统的描述方式与简化程度不同有很大差异，主要体现在以下几方面：（1）所考虑的故障几何模型复杂度；（2）将动力学问题简化为一维问题，二维平面问题还是三维问题；（3）所考虑的系统非线性因素（如 Hertz 接触、时变刚度、游隙和润滑油膜等）等。

近年来，有限元及多体动力学相关仿真软件的发展为实现更复杂更精细的轴承建模提供了新的手段，许多研究者利用 LS-DYNA、ADAMS 等参数化建模软件实现对滚动轴承的故障仿真<sup>[12-15]</sup>，研究滚动轴承更精细的动力学行为，如保持架行为<sup>[13]</sup>、共振特性分析<sup>[14]</sup>等。有限元起源于静力分析，当分析的侧重点在于轴承内部的应力变化时，利用有限元仿真更合适；利用 ADAMS 可以将轴承放在整个机械系统中，实现更复杂的故障模拟<sup>[15]</sup>。

对滚动轴承故障仿真模型的研究十分广泛<sup>[16]</sup>，用什么手段仿真取决于仿真的目的，需要根据具体问题具体分析，选择最合适的仿真方式。

## 1.2.2 基于数据驱动的智能故障诊断方法研究现状

基于数据驱动的故障诊断方法包括定性方法和定量方法，其中定性方法主要指基于符号智能的方法，包括专家系统，有向图方法等<sup>[3]</sup>。本部分主要介绍定量方法。针对定量方法，主要包括两方面的研究，一是研究有利于故障识别的特征提取方法，如基于熵的时域特征提取方法<sup>[17]</sup>，基于谱峭度（Spectral Kurtosis, SK）的频域特征提取方法<sup>[18]</sup>和基于经验模式分解（Empirical Mode Decomposition, EMD）<sup>[19]</sup>、小波分解等的时频域特征提取方法等。二是研究故障识别方法<sup>[20]</sup>。本节针对故障识别部分，从传统的机器学习方法，深度学习方法和迁移学习方法三个层面介绍智能诊断方法的研究现状。

### 1.2.2.1 基于机器学习的传统智能诊断方法研究现状

基于机器学习的传统智能诊断方法主要包括支持向量机（Support Vector Machine, SVM）方法、人工神经网络（Artificial Neural Network, ANN）方法、 $k$ 近邻（ $k$ -Nearest Neighbor, KNN）方法等<sup>[20]</sup>，应用这些方法时多需要对传感器采集到的信号进行预处理，提取信号特征，并利用最大相关最小冗余（minimum Redundancy Maximum Relevance, mRMR）等特征选择方法进行特征优选，进一步利用这些机器学习算法实现故障识别。

KNN 方法的基本思想是对于给定的测试样本,按照某种距离度量方式,寻找与其最接近的  $k$  个训练样本,根据这些训练样本以某种准则判断该测试样本所属的类别。Pandya 等将经 HHT (Hilbert-Huang Transform) 方法提取到的信号特征作为 KNN 方法的输入,实现了对滚动轴承故障识别<sup>[21]</sup>;Yaqub 等基于高阶累积量和小波变换提取信号特征,利用 KNN 作为分类器实现了对滚动轴承的早期故障检测<sup>[22]</sup>;另外 KNN 方法常与降维方法一起使用,如 Li 将 WPT (Wavelet Package Transform) 得到的故障特征向量利用局部线性嵌入 (Locally Linear Embedding, LLE) 方法降维,再利用 KNN 方法分类,实现了对齿轮的故障识别<sup>[23]</sup>;Safizadeh 利用 PCA 对特征降维,再利用 KNN 分类实现对滚动轴承的故障识别<sup>[24]</sup>。

SVM 方法的基本思想是基于结构风险最小化理论找到划分不同类别的最优超平面。核函数的选择及超参数优化是影响 SVM 性能的关键。针对故障识别问题,郭磊等将 Mexican hat 小波函数作为核函数,将小波支持向量机 (Wavelet Support Vector Machine, WSVM) 应用到滚动轴承故障诊断中,具有较好的识别性能<sup>[20]</sup>;基于多核的支持向量机也被用于故障诊断中<sup>[25]</sup>;Chen 利用改进的粒子群优化算法对最小二乘支持向量机进行优化,用于滚动轴承的故障诊断中<sup>[26,27]</sup>。

ANN 方法是一种常用的模式识别方法,常见的 ANN 结构包括多层感知机 (Multi-Layer Perceptron, MLP)、径向基网络 (Radial Basis Function Network, RBF) 及概率神经网络 (Probabilistic Neural Network, PNN),结合不同的特征提取方法,ANN 已广泛应用于故障诊断领域中,如 Rafiee 将振动信号的小波包系数的标准差作为特征,利用 MLP 实现了对齿轮箱的故障诊断<sup>[28]</sup>;Lei 等利用 WPT, EMD 进行特征提取,利用 RBF 网络实现了对滚动轴承轻微磨损故障的诊断<sup>[29]</sup>;Yaghobi 将离散小波变换和 PNN 结合实现了发电机的故障诊断<sup>[30]</sup>。

#### 1.2.2.2 基于深度学习的智能诊断方法研究现状

为了实现从复杂信号中自动化地学习信号特征并实现故障识别,基于深度学习的故障诊断方法受到广泛关注<sup>[31]</sup>。基于深度学习的智能诊断方法可在一个深度网络中自动化地完成故障特征提取和故障识别两个过程,相比于基于机器学习的传统智能诊断方法,是一种更高层次的智能诊断方式。这类研究主要集中在自动编码器 (auto-encoder, AE) 及其变体方法、受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 构建的深度信念网络 (Deep Belief Networks, DBN) 和深度玻尔兹曼机 (Deep Boltzmann Machines, DBM) 方法、稀疏滤波方法 (Sparse Filtering, SF)、卷积神经网络 (Conventional Neural Networks, CNN) 方法等<sup>[20,32]</sup>。

Lei 等利用 SAE (Stacked Auto-Encoders) 直接从原始振动信号中学习特征,并使用 softmax 分类器进行分类,实现了滚动轴承的故障诊断<sup>[33]</sup>;Gan 等利用 DBN

构建两层诊断网络，并从振动信号的小波能量特征出发，实现滚动轴承的故障诊断，其中第一层用以识别故障类型，第二层用以识别故障程度<sup>[34]</sup>；Li 等提出了 MDSVC（Multimodal Deep Support Vector Classification）模型，基于齿轮箱振动信号的时域、频域及时频域特征，利用三个 DBM 分别对三个层面的特征进一步抽象，最终将抽象后的特征利用 SVC（Support Vector Classification）实现特征融合及故障的识别<sup>[35]</sup>；Lei 等利用 SF 从原始时域信号提取特征，利用 softmax 分类器分类，构建了两阶段的故障诊断模型，实现了滚动轴承的故障诊断<sup>[36]</sup>；Zhang 等基于 CNN 实现了滚动轴承在噪声环境和变工况下的故障诊断<sup>[37]</sup>；Pan 等基于 CNN 提出一种名为 LiftingNet 的诊断网络，在噪声环境下对滚动轴承进行故障诊断取得很好的诊断效果<sup>[38]</sup>；Guo 等人提出自适应学习率的方式训练 2 层 CNN 网络，第一层网络识别故障类型，第二层网络用于识别故障程度，实现了滚动轴承故障的层次化诊断<sup>[39]</sup>；Yang 等结合层次符号分析（hierarchical symbolic analysis, HSA）方法和 CNN，利用 HSA 初步提取特征，进一步用简单的网络结构完成自动化的特征学习和故障识别过程<sup>[40]</sup>。这些故障诊断方法基于特定设备的历史数据实现了滚动轴承故障的有效识别，显示出深度学习方法在解决轴承诊断问题上的具有很大潜力。

#### 1.2.2.3 基于迁移学习的智能诊断方法研究现状

传统的机器学习方法以及深度学习方法的基本假设是要求训练数据与测试数据服从相同的概率分布，在很多实际问题中，这个假设常不能被满足，使得这些方法的性能显著下降<sup>[41]</sup>。迁移学习（Transfer Learning）方法通过实现不同领域之间的知识迁移可以缓解这个问题，在计算机视觉，自然语言处理等领域已广泛应用<sup>[42]</sup>。近几年，一些研究者尝试用迁移学习的方法解决故障诊断问题。考虑不同故障诊断问题设定，基于迁移学习的故障诊断方法近年来受到广泛关注，这些问题设定一般为考虑同一设备的不同工况下的迁移，如文献[41,43-49]，或考虑同类型设备之间的迁移，如文献[50]。另外，Zheng 等提出了一种多源域泛化的故障诊断方法，能够实现同类型不同设备或不同工况下诊断知识的迁移<sup>[51]</sup>。这些诊断方法将所能获取的实际数据作为源域，实现了目标域任务的有效诊断。下面从浅层迁移和深度迁移两个方面介绍迁移学习在故障诊断问题中的研究现状。

##### （1）基于浅层迁移的智能诊断方法研究现状

这里浅层迁移（Shallow Transfer）指在进行识别前的特征提取部分，特征由基于先验知识的手动提取方式得到，而在识别部分，采用迁移方法<sup>[52]</sup>。典型的浅层迁移方法如基于实例迁移的 TrAdaBoost 方法及基于特征迁移的迁移成分分析（Transfer Component Analysis, TCA）方法，另外还有 SSTCA（Semi-Supervised

TCA) 方法、GFK (Geodesic Flow Kernel) 方法等。

TrAdaBoost 方法的思想是将源域中的样本重新赋予权重以帮助目标域识别<sup>[54]</sup>。沈飞等将该方法用在滚动轴承故障诊断问题中, 目标域样本较少时, 使用 TrAdaBoost 方法能帮助目标域较好地完成任务<sup>[54]</sup>。TCA 方法的基本思想是将源域和目标域映射到一个低维的特征子空间中, 在这个子空间中源域和目标域的分布差异最小<sup>[55]</sup>。谢骏遥等人将 TCA 方法结合 SVM 应用到变工况的齿轮故障识别中, 并讨论了不同核函数对识别性能的影响, 相比于传统的机器学习方法, 识别率有了明显的提升<sup>[56]</sup>。SSTCA 方法在 TCA 方法的基础上, 考虑标签与数据之间的依赖关系, 利用 Hilbert-Schmidt 独立性准则 (Hilbert-Schmidt Independence Criterion, HSIC) 度量样本与标签之间的依赖程度, 并利用局部保持项提高数据类内的聚集程度<sup>[55]</sup>。康守强等将 SSTCA 方法与 SVM 结合, 应用在滚动轴承的故障诊断中, 并采用多核函数提高对数据的表示能力, 实现了变工况下轴承的故障诊断<sup>[57]</sup>。

## (2) 基于微调策略的深度迁移智能诊断方法研究现状

浅层迁移方法可以在一定程度上改善由于分布不同造成的机器学习算法性能下降的问题, 但诊断过程不是从原始信号出发, 而是依赖于人工提取的某些数据特征, 这造成了信息的丢失, 使得迁移能力有限<sup>[52]</sup>。考虑到深度学习有较强的挖掘数据深层抽象特征的能力, 可能能够学习到更有利于迁移的特征, 人们尝试将深度网络应用到迁移学习中, 深度迁移方法 (Deep Transfer Learning) 在近几年成为迁移学习领域中重要的研究方向。

在深度迁移方法中, 深度域适应方法 (Deep Domain Adaptation) 受到广泛关注, 其中, 基于微调策略的网络参数迁移是一类常见的深度域适应方法, 其基本思想是利用源域数据预训练一个基础网络, 重用特定层的网络参数, 基于目标域数据构建一个新的网络。在图像识别领域, Yosinski 等通过实验的方式, 说明基于微调策略, 利用深度网络学习到的可迁移特征可以提高模型的泛化能力<sup>[58]</sup>; Chu 等提出, 当可用于微调的数据很少时, 应尽可能多地迁移预训练网络参数<sup>[59]</sup>; 在故障诊断问题中, Cao 等将用于图像处理的 AlexNet (CNN) 网络前七层固定, 把信号的时域图作为输入, 微调最后一层权重, 实现了对齿轮故障识别<sup>[60]</sup>; Zhang 等迁移源域网络参数, 并对于不同的目标域数据分布调整网络结构, 实现了变工况下的滚动轴承故障诊断<sup>[45]</sup>; Zhong 等基于正常数据训练 CNN 网络, 实现了燃气轮机的故障特征映射, 并结合 SVM 实现其故障诊断<sup>[61]</sup>; Hasan 等提出基于声发射信号的声谱图训练 CNN 网络, 并迁移网络学习到的特征, 实现了滚动轴承的变工况诊断<sup>[44]</sup>; Xu 等应用图像处理领域的网络 LeNet-5, 构造了一种基于 CNN

的在线诊断方法。该方法迁移离线 CNN 网络的浅层特征到在线 CNN 网络，并微调网络参数，实现滚动轴承的在线诊断<sup>[49]</sup>。

这些故障诊断方法的源域数据组织形式多将一维的振动信号转化为 2 维的图片，并采用与目标域工况（转速）不同设备相同的数据，相比于这种源域组织形式，本文中提出利用滚动轴承故障仿真信号作为源域，从仿真模型中获得的振动信号能从本质上揭示故障特性，同时通过调整仿真参数，能方便地获取与待诊断设备转速相同轴承型号相同的数据，基于这样的源域可能能够产生更有利于诊断的故障诊断知识。

### 1.2.3 国内外研究现状分析

#### 1.2.3.1 关于滚动轴承故障仿真模型

滚动轴承的故障仿真模型主要集中在直接对故障信号仿真、基于动力学的数值仿真以及利用商用软件进行参数化建模的仿真三种方式。故障信号仿真直接从振动信号应具有的形式出发，是一种比较简单的定性仿真方式；基于动力学的数值仿真方法根据对实际问题的简化程度不同，模型复杂度相差较大，使用起来较为灵活；基于商用软件建模方式主要包括有限元及多体动力学两类。不同的仿真方式侧重点不同，实际应用时应考虑问题的需求选择合适的仿真方式。

#### 1.2.3.2 关于基于数据驱动的智能故障诊断方法

基于数据驱动的故障诊断方法可以总结为以下几个方面：

##### （1）基于机器学习的传统智能诊断方法

该类方法通过结合一些特征提取和特征选择的方法在故障识别任务中取得了较好的效果，但其要求训练数据及测试数据满足同分布条件，因此在完成识别任务时总是将同一数据集的一部分作为训练集构建模型，另一部分作为测试集验证模型的可行性，这种数据集的构造方式在实际问题中是极难实现的。

##### （2）基于深度学习的智能诊断方法

该类方法将特征提取和故障识别过程融合在一个网络中，与传统的机器学习方法相比，诊断过程体现了更高的自动化和智能化水平。其主要问题是网络的超参数调整依赖经验，目前还没有明确的规则；且由于深度网络是一个黑箱模型，对于其提取到的特征，物理意义不够明确。另外，虽然深度学习诊断方法在公开数据集或一些故障模拟实验的数据中取得了较好的效果，但我们所掌握的数据相对于实际情况中的数据来说是极少的，这也是深度学习方法在故障诊断领域无法像图像识别、语音识别领域一样能够真正用于解决实际问题的主要原因。



### (3) 基于迁移学习的智能诊断方法

浅层迁移方法在一定程度上解决了由于数据分布不同造成的机器学习诊断方法性能下降的问题，但由于依赖人工提取的特征，迁移能力有限；深度迁移方法利用深度学习抽象特征的能力，使迁移的效果有了更大程度的提高。对于故障识别问题，目前多停留在不同工况之间的迁移，对于同一类型的不同设备、不同工况等更一般化的、数据分布差异更大的诊断问题设定有待进一步研究和扩展，进而为实现智能诊断方法从实验室阶段到实际应用阶段的转变提供可能。

## 1.3 本文的主要研究内容

本文的主要研究内容是针对实际故障诊断问题中的小样本特性，即只有少量的有标签样本并且不足以训练稳定诊断模型的情形，以滚动轴承为对象，通过动力学仿真的方式获得大量的故障模拟数据，在此基础上设计从大量仿真数据中迁移诊断知识到实际诊断场景的滚动轴承故障诊断方法，将诊断方法应用于实际诊断问题中，验证诊断方法的有效性并对方法的诊断机理进行分析，为解决实际诊断场景中的小样本诊断问题提供新思路。

本文的具体研究内容包括以下几个方面：

#### (1) 基于动力学的滚动轴承故障仿真模型研究

该部分建立滚动轴承故障仿真模型，以获得能够反映轴承在不同健康状态下（正常状态、内圈故障及外圈故障）振动特性的振动信号，为产生更丰富的滚动轴承故障仿真数据奠定基础。首先建立正常状态下滚动轴承数学模型，随后引入滚动轴承滚道局部故障，得到滚动轴承故障仿真模型，最后将仿真模型生成的仿真信号与实际信号对比，验证模型的合理性。

#### (2) 构建基于动力学仿真和迁移学习的滚动轴承智能故障诊断方法框架

该部分从动力学、时域信号概率分布与信号频域上的频率分布三个角度对仿真数据与实际数据的共性与差异性进行分析，并在此基础上，利用卷积神经网络和深度迁移策略，构建小样本条件下，将仿真数据中的诊断知识迁移到实际诊断问题中的滚动轴承智能诊断方法框架，明确了诊断问题设定并给出诊断方法的详细流程。

#### (3) 基于动力学仿真和迁移学习的滚动轴承故障诊断方法应用与机理分析

该部分基于两个诊断实例，实现方法的实际应用，并设计多个对比实验，以进一步验证诊断方法在解决小样本故障诊断问题上的优越性。另外对仿真数据和迁移策略在诊断过程中起到的作用进行详细分析，得到诊断方法在实际诊断问题

中的诊断机理。

本文的组织结构及研究思路如图 1-1 所示。

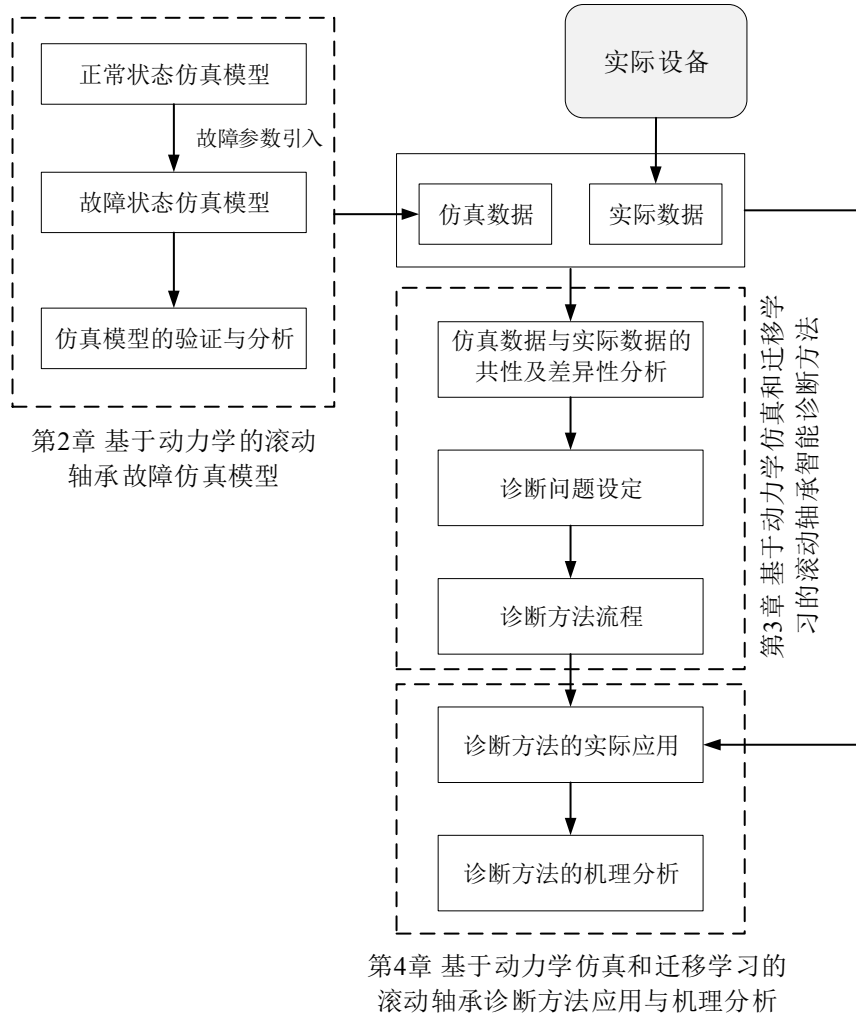


图 1-1 本文的组织结构及研究思路

## 第 2 章 基于动力学的滚动轴承故障仿真模型

### 2.1 引言

为了解决实际故障诊断问题中缺少待诊断设备的大量故障历史数据的问题，在本文中拟采用动力学仿真的方式获取更丰富的故障数据，利用这些数据辅助实际小样本问题的诊断。这里建立动力学仿真模型的目的，是要在故障条件下生成能够反映滚动轴承动力学系统共性故障特性的振动数据，动力学模型可不具备精细描述滚动轴承的动力学行为（如描述保持架行为或发生故障冲击时冲击过程的细节）的能力。基于上述建模目的，本章建立四自由度滚动轴承故障仿真模型，首先建立正常状态下的动力学模型；其次引入轴承故障的几何参数，建立滚动轴承故障仿真模型；最后从时域波形和频谱两个角度对模型的合理性进行验证。

### 2.2 滚动轴承动力学模型

#### 2.2.1 滚动轴承动力学方程

采用集中质量的动力学建模方式，仅考虑轴和内圈的水平位移  $x_{in}$  和竖直位移  $y_{in}$ ，外圈和轴承基座的水平位移  $x_{out}$  和竖直位移  $y_{out}$ ，将滚动轴承运动系统看成四自由度系统，建模过程对整个轴承运动系统做如下简化：

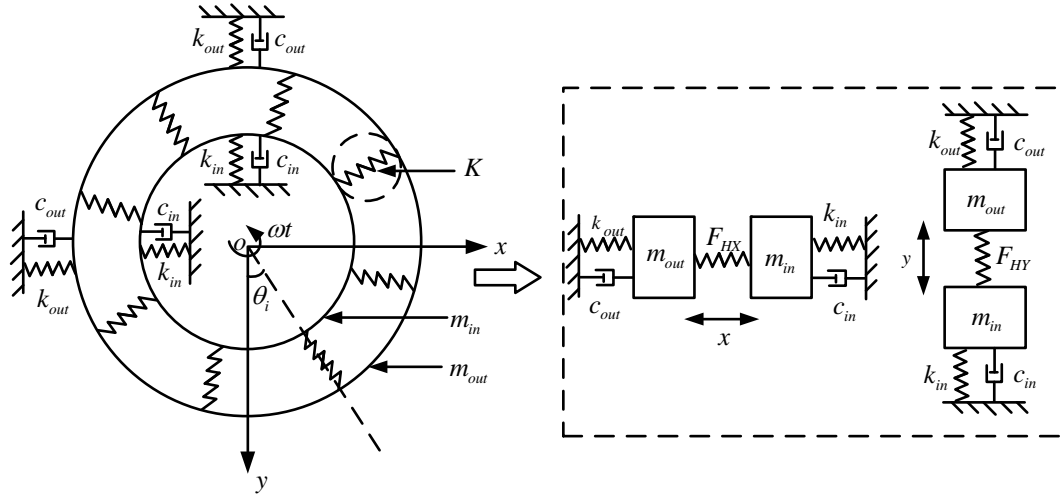
（1）不考虑弹性流体润滑的刚度和阻尼，系统中的非线性因素包括内外圈与滚动体接触的非线性接触力，由于滚动体周期性承载引起的时变刚度和滚动体与滚道之间的径向游隙。

（2）整个运动在  $xoy$  平面内发生，内圈与轴一起绕  $z$  轴以常角速度  $\omega$  转动，外圈不转动；

（3）内外圈滚道和滚动体之间的接触简化为 Hertz 接触；

（4）不考虑滚动体和保持架的质量和转动惯量，将滚动体简化成提供弹性恢复力的非线性弹簧，且相对于  $z$  轴均匀分布；

基于上述假设，滚动轴承系统的四自由度动力学仿真模型示意图如图 2-1a) 所示，该系统可进一步看成由 2 个方向的二自由度系统构成，如图 2-1b) 所示。



a) 滚动轴承四自由度动力学系统

b) 滚动轴承 2 个二自由度子系统

图 2-1 滚动轴承四自由度系统示意图

在动力学建模过程中考虑制造或安装误差等因素造成的离心力，得到系统的动力学方程为：

$$\begin{cases} m_{in} \ddot{x}_{in} + c_{in} \dot{x}_{in} + k_{in} x_{in} = -F_{HX} + em_{in} \omega^2 \cos \omega t \\ m_{in} \ddot{y}_{in} + c_{in} \dot{y}_{in} + k_{in} y_{in} = m_{in} g - F_{HY} + em_{in} \omega^2 \sin \omega t \\ m_{out} \ddot{x}_{out} + c_{out} \dot{x}_{out} + k_{out} x_{out} = F_{HX} \\ m_{out} \ddot{y}_{out} + c_{out} \dot{y}_{out} + k_{out} y_{out} = F_{HY} + m_{out} g \end{cases} \quad (2-1)$$

式中  $m_{in}$  ——内圈与轴的质量之和 (kg)；

$m_{out}$  ——外圈与轴承基座的质量之和 (kg)；

$c_{in}$  ——轴与轴固定处的连接阻尼 (Ns/m)；

$c_{out}$  ——轴承基座与地面基础固定处的连接阻尼 (Ns/m)；

$k_{in}$  ——轴与轴固定处的等效连接刚度 (N/m)；

$k_{out}$  ——轴承基座与地面固定处的等效连接刚度 (N/m)；

$F_{HX}$  ,  $F_{HY}$  ——由于内外圈与滚动体接触而产生弹性恢复力在  $x$  ,  $y$  方向的分量，与内外圈的相对位移和接触刚度有关 (N)；

$e$  ——偏心距 (m)；

$\omega$  ——内圈转动的角频率 (rad/s)，设转频为  $f_r$ ，则  $\omega = 2\pi f_r$ ；

$g$  ——重力加速度 (m/s<sup>2</sup>)；

$t$  ——时间 (s)。

在对方程求解前，需计算非线性弹性恢复力  $F_H$ 。基于 Hertz 接触理论， $F_H$  可由下式计算：

$$F_H = K\delta^n \quad (2-2)$$

式中  $K$ ——载荷-变形系数，与相互接触的弹性体的材料和几何参数有关，或称等效接触刚度 ( $\text{N/m}^n$ )；

$\delta$ ——接触变形 ( $\text{m}$ )；

$n$ ——载荷-变形指数，对球轴承， $n = 1.5$ 。

以下两小节将对等效接触刚度和接触变形的计算进行详细介绍。

### 2.2.2 Hertz 接触刚度计算

为方便叙述，图 2-2 给出了滚动轴承相关几何参数的示意图。滚动轴承滚动体与滚道间的接触可看成线弹性体的 Hertz 接触问题，等效接触刚度基于 Hertz 接触理论计算，Hertz 接触理论示意图如图 2-3 所示。

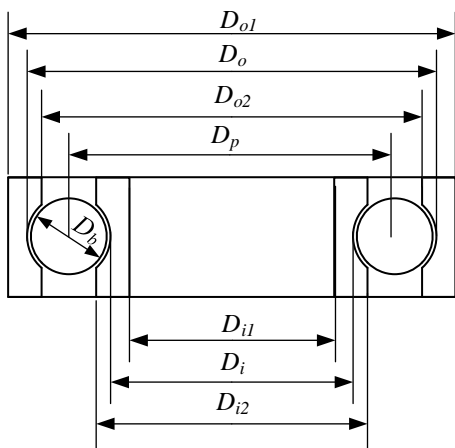


图 2-2 滚动轴承几何参数

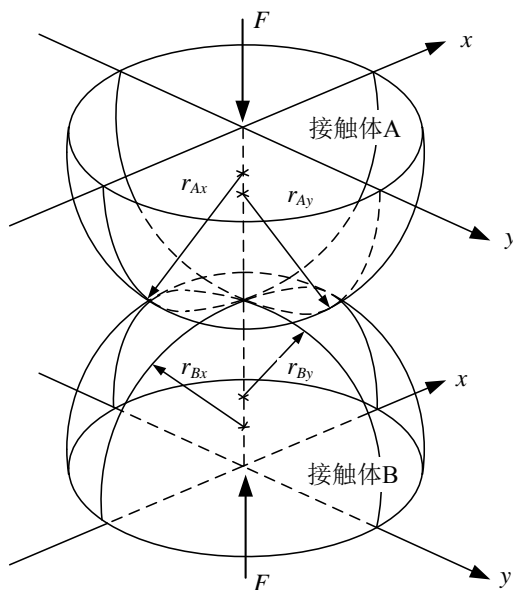


图 2-3 Hertz 接触示意图

对滚动轴承中滚动体与内、外滚道的接触，将滚动体定义为接触体 A，内滚道（外滚道）为接触体 B；定义过滚动体与内滚道（外滚道）相互接触面的法线，且与轴承径向相互平行的平面为  $X$  平面，与轴承轴向相平行的平面为  $Y$  平面，则滚动体与内圈滚道接触的主半径为：

$$r_{Axin} = r_{Ayin} = r_b, r_{Bxin} = \frac{D_i}{2}, r_{Byin} = r_i \quad (2-3)$$

式中  $D_i$ ——内圈滚道直径 ( $\text{m}$ )；

$r_i$ ——内圈沟槽半径 ( $\text{m}$ )。

同理，滚动体（接触体 A）与外圈滚道（接触体 B）接触曲面的主半径为：

$$r_{Axout} = r_{Aout} = r_b, \quad r_{Bxout} = \frac{D_o}{2}, \quad r_{Byout} = r_o \quad (2-4)$$

式中  $D_o$ ——外圈滚道直径（m）；

$r_o$ ——外圈沟道半径（m）。

考虑实际情况，内、外圈沟槽半径略大于滚动体半径， $r_i$ ， $r_o$ 可由以下两式计算：

$$r_i = k_i D_b \quad (2-5)$$

$$r_o = k_o D_b \quad (2-6)$$

式中  $k_i$ ——内圈深沟曲率系数，本文设置  $k_i = 0.515^{[62]}$ ；

$k_o$ ——外圈深沟曲率系数，本文设置  $k_o = 0.525^{[62]}$ ；

另外，考虑实际轴承的套圈挡边高度，内、外圈滚道直径由下两式计算：

$$D_i = D_{i2} - K_d D_b \quad (2-7)$$

$$D_o = D_{o2} + K_d D_b \quad (2-8)$$

式中  $K_d$ ——套圈挡边高度系数，本文设置  $K_d = 0.35^{[62]}$ 。

则内圈（外圈）与滚动体的接触刚度  $K_{in}(K_{out})$ 可由下式计算<sup>[63]</sup>：

$$K_{in} = \left( \frac{\pi^2 H_{in}^2 E_{in}^2 \psi_{in}}{4.5 \Gamma_{in}^3 (1/R_{in})} \right)^{0.5} \quad (2-9)$$

$$K_{out} = \left( \frac{\pi^2 H_{out}^2 E_{out}^2 \psi_{out}}{4.5 \Gamma_{out}^3 (1/R_{out})} \right)^{0.5} \quad (2-10)$$

式（2-9）至式（2-10）中， $E_{in}, E_{out}$ 为接触物体的等效弹性模量，设两接触体的弹性模量和泊松比分别为  $E_1, E_2, \mu_1, \mu_2$ ，则等效弹性模量  $E$  由下式计算：

$$\frac{2}{E} = \frac{1 - \mu_1^2}{E_1} + \frac{1 - \mu_2^2}{E_2} \quad (2-11)$$

式（2-9）至式（2-10）中的其他参数<sup>[63]</sup>：

$$H_{in} = 1.0339 \left( \frac{R_{yin}}{R_{xin}} \right)^{0.636}, \quad H_{out} = 1.0339 \left( \frac{R_{yout}}{R_{xout}} \right)^{0.636} \quad (2-12)$$

$$\psi_{in} = 1.0003 + \frac{0.5968}{R_{yin} / R_{xin}}, \psi_{out} = 1.0003 + \frac{0.5968}{R_{yout} / R_{xout}} \quad (2-13)$$

$$\Gamma_{in} = 1.5277 + 0.6023 \ln \frac{R_{yin}}{R_{xin}}, \Gamma_{out} = 1.5277 + 0.6023 \ln \frac{R_{yout}}{R_{xout}} \quad (2-14)$$

$$\frac{1}{R_{in}} = \frac{1}{R_{xin}} + \frac{1}{R_{yin}}, \frac{1}{R_{xin}} = \frac{1}{r_{Axin}} + \frac{1}{r_{Bxin}}, \frac{1}{R_{yin}} = \frac{1}{r_{Ayin}} - \frac{1}{r_{Byin}} \quad (2-15)$$

$$\frac{1}{R_{out}} = \frac{1}{R_{xout}} + \frac{1}{R_{yout}}, \frac{1}{R_{xout}} = \frac{1}{r_{Axout}} - \frac{1}{r_{Bxout}}, \frac{1}{R_{yout}} = \frac{1}{r_{Ayout}} - \frac{1}{r_{Byout}} \quad (2-16)$$

等效接触刚度  $K$  为:

$$\frac{1}{K} = \left( \frac{1}{K_{in}^n} + \frac{1}{K_{out}^n} \right)^{\frac{1}{n}} \quad (2-17)$$

对于深沟球轴承,  $n = 1.5$ , 对圆柱滚子轴承,  $n = \frac{10}{9}$ 。

### 2.2.3 滚动轴承接触变形计算

滚动体与内外圈接触产生的接触变形与内外圈的相对位移和径向游隙有关, 在第  $i$  个滚动体处, 由内外圈的相对位移造成的弹性变形  $\delta_i$  为:

$$\delta_i = (x_{in} - x_{out}) \sin \theta_i + (y_{in} - y_{out}) \cos \theta_i - c_r \quad (2-18)$$

式中  $\theta_i$  ——第  $i$  个滚动体在  $t$  时刻的角位置 (rad);

$c_r$  ——轴承的径向游隙 (m)。

设  $\theta_1$  初始位置为 0, 滚动体个数为  $N$ , 则:

$$\theta_i = \frac{2\pi(i-1)}{N} + \omega_c t \quad (2-19)$$

式中  $\omega_c$  ——保持架转动频率 (rad/s)。

设  $D_b$  为滚动体直径,  $D_p$  为轴承节径, 则:

$$\omega_c = \frac{1}{2} \left( 1 - \frac{D_b}{D_p} \cos \alpha \right) \omega \quad (2-20)$$

式中  $\alpha$  ——接触角 ( $^{\circ}$ )。

只有当内外圈的径向相对位移大于游隙，即弹性变形大于 0 时，滚动体才能在内外圈之间起到非线性弹簧的作用，产生非线性弹性力，由 Hertz 接触理论，第  $i$  个滚动体所引起的弹性恢复力  $F_{H_i}$  为：

$$F_{H_i} = KH(\delta_i)\delta_i^{1.5} \quad (2-21)$$

式中  $K$  ——滚动体与内外圈的等效接触刚度 ( $\text{N/m}^{1.5}$ )。

$H(\delta_i)$  ——定义为：

$$H(\delta_i) = \begin{cases} 1, & \delta_i > 0 \\ 0, & \delta_i \leq 0 \end{cases} \quad (2-22)$$

得到接触变形后，代入式 (2-21)，进一步可得弹性恢复力在  $x$  方向及  $y$  方向的分力分别为：

$$F_{HX} = \sum_{i=1}^N KH(\delta_i)\delta_i^{1.5} \sin \theta_i \quad (2-23)$$

$$F_{HY} = \sum_{i=1}^N KH(\delta_i)\delta_i^{1.5} \cos \theta_i \quad (2-24)$$

## 2.3 滚动轴承滚道局部故障建模

对于滚动轴承滚道的局部故障一般有两种建模方式，一种是基于冲击力模型的故障建模，直接将故障的影响转化为动力学方程中的周期冲击激励项；另一种是基于几何模型的故障建模，将故障简化为具有一定几何形状的凹坑，本质上是通过游隙的瞬时变化来表征故障，建模时能够在仿真参数上更直观的反映故障的不同程度。本文将滚动轴承滚道上的局部故障在几何上简化为一矩形凹坑，以外圈局部故障为例，其故障位置处的几何关系示意图如图 2-4 所示。



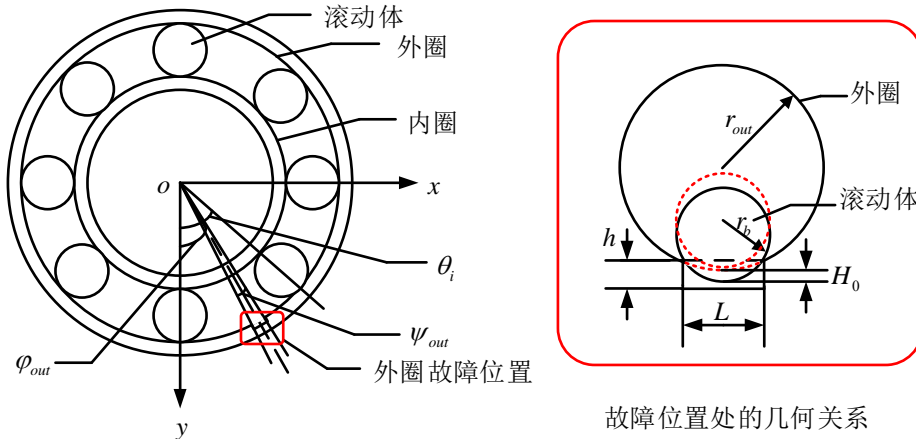


图 2-4 外滚道局部故障位置处几何关系示意图

当滚动体进入故障区域时，滚动体与滚道接触瞬时消失，当其离开故障区域时，滚动体与滚道接触恢复。接触消失和恢复来源于局部故障造成的  $\delta_i$  的变化。具体地，当引入故障后，滚动体  $i$  处的接触变形  $\delta_i$  变为：

$$\delta_i = (x_{in} - x_{out}) \sin \theta_i + (y_{in} - y_{out}) \cos \theta_i - c_r - H \quad (2-25)$$

其中  $H$  为引入故障后，在滚动轴承的整个运行周期中陷入故障区域的距离，由于滚动体内外圈的运动形式不同，因此对于发生在不同滚道的故障， $H$  的计算方法不同：

(1) 当外圈发生故障时， $H$  由下式计算：

$$H = \begin{cases} H_0, & \cos(\theta_i - \varphi_{out}) \geq \cos\left(\frac{\psi_{out}}{2}\right) \\ 0, & \text{else} \end{cases} \quad (2-26)$$

式中  $\varphi_{out}$  ——外圈故障的角位置，对外圈故障来说是一不随时间变化的量 (rad)；

$\psi_{out}$  ——外圈故障所对应的圆心角 (rad)，设  $D_o$  为外圈直径，由于故障尺寸相比于外滚道直径很小，因此  $\psi_{out}$  可近似写成：

$$\psi_{out} \approx \frac{2L}{D_o} \quad (2-27)$$

在式 (2-26) 中， $H_0$  为滚动体进入故障区域后接触变形的变化量，实际上是外圈与滚动体圆心距的变化量。根据图 2-2 中的几何关系，当外圈发生故障时， $H_0$  可由下式计算：

$$H_0 = \begin{cases} r_b - \sqrt{r_b^2 - \frac{L^2}{4}} - \left( \frac{D_o}{2} - \sqrt{\frac{D_o^2}{4} - \frac{L^2}{4}} \right), & 8r_b h \geq 4h^2 + L^2 \\ h, & 8r_b h < 4h^2 + L^2 \end{cases} \quad (2-28)$$

(2) 当内圈发生故障时,  $H$  由下式计算:

$$H = \begin{cases} H_0, & \cos(\theta_i - \varphi_{in}) \geq \cos\left(\frac{\psi_{in}}{2}\right) \\ 0, & \text{else} \end{cases} \quad (2-29)$$

式中  $\varphi_{in}$ ——内圈故障的角位置, 与外圈故障不同, 内圈故障角位置随轴一起以角速度  $\omega$  转动, 即:

$$\varphi_{in} = \omega t \quad (2-30)$$

$\psi_{in}$ ——外圈故障所对应的圆心角 (rad), 设  $D_i$  为内圈直径, 由于故障尺寸相比于内滚道直径很小, 因此  $\psi_{in}$  可近似写成:

$$\psi_{in} = \frac{2L}{D_i} \quad (2-31)$$

当内圈故障发生时, 与外圈故障类似,  $H_0$  为内圈与滚动体圆心距的变化量,  $H_0$  可由下式计算:

$$H_0 = \begin{cases} r_b - \sqrt{r_b^2 - \frac{L^2}{4}} + \left( \frac{D_i}{2} - \sqrt{\frac{D_i^2}{4} - \frac{L^2}{4}} \right), & 8r_b h \geq 4h^2 + L^2 \\ h, & 8r_b h < 4h^2 + L^2 \end{cases} \quad (2-32)$$

将引入故障参数后的  $\delta_i$  代入动力学方程 (2-1) 中, 得到具有局部故障时的滚动轴承故障仿真模型。

## 2.4 滚动体随机滑动因素的引入

在滚动轴承实际运行的过程中, 由于每个滚动体所受到的载荷会存在周期性波动, 滚动体在滚道中的运动不会一直保持纯滚动, 会产生一定的随机滑动。当故障发生时, 这种滑动可能会比正常状态下的滑动更剧烈。在本文中, 参考文献 [10], 在仿真模型中引入这种滑动对系统的影响。具体地, 引入随机滑动后, 滚动体  $i$  在  $t$  时刻的位置可写成:

$$\theta_i = \frac{2\pi(i-1)}{N} + \omega_c t + \theta_0 + (0.5 - rand) \times \theta_r \quad (2-33)$$

式中  $\theta_r$  ——由于滑动带来的滚动体相位角偏差。

至此建立了滚动轴承四自由度故障仿真模型，利用 Matlab/Simulink 搭建动力学仿真模型，采用变步长 ode45 方法对动力学方程求解，可得滚动轴承各种健康状态下的振动响应。

## 2.5 滚动轴承动力学仿真模型的验证与分析

为了验证上述滚动轴承仿真模型的合理性，本节将由仿真模型产生的振动加速度信号与实验中采集的实际振动加速度信号从时域和频域两个角度对比。从时域的角度，主要验证故障仿真信号是否与实际信号有相似的冲击波形，从频域的角度，主要验证是否在频谱或包络谱中具有特征频率。

### 2.5.1 西储大学轴承实验

验证所用的实际信号采集自凯斯西储大学（Case Western Reserve University, CWRU）的轴承实验。实验台如图 2-5 所示。实验台包含一个 2hp 的电机，一个扭矩传感器，一个测功率器和控制电机。测试轴承为电机轴的支撑轴承。驱动端轴承型号为 SKF-6205-RS，风扇端轴承型号为 SKF-6203-RS。两轴承的几何参数及故障频率如表 2-1 所示。

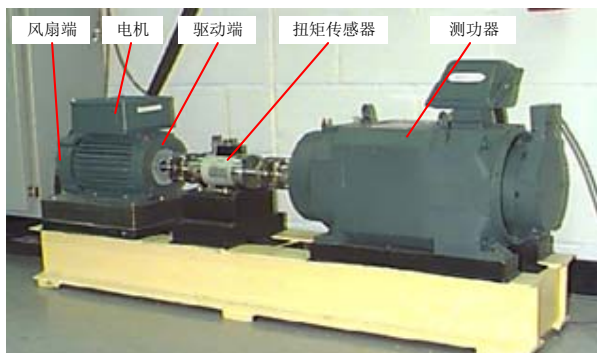


图 2-5 凯斯西储大学轴承实验台

为了模拟轴承故障，实验中利用电火花的方式向轴承的内滚道及外滚道注入尺度为 0.18mm、0.36mm 及 0.53mm 的凹坑，表征三种不同程度的故障。实验中加速度传感器安装在待测轴承的轴承座上，方向为竖直方向，采集了轴承在不同健康状态下，4 种不同负载/转速（0hp/1797rpm、1hp/1772rpm、2hp/1750rpm、3hp/1720rpm）的振动加速度信号，采样频率为 12kHz 或 48kHz。

表 2-1 轴承几何参数及故障频率

轴承型号	SKF-6203-RS	SKF-6205-RS
滚动体数 $N$	8	9
滚动体直径 $D_b$ (mm)	6.746	7.940
节径 $D_p$ (mm)	28.500	39.040
接触角 $\alpha$ (°)	0	0
内圈故障特征频率 $f_{in}$	$4.947 f_r$	$5.149 f_r$
外圈故障特征频率 $f_o$	$3.053 f_r$	$3.585 f_r$

本文选择 SKF-6203-RS 轴承在负载/转速为 0hp/1797rpm, 采样频率为 12kHz, 正常状态以及内外圈具有 0.36mm 单点故障时的信号, 作为实际验证信号, 说明模型的合理性。

## 2.5.2 仿真模型参数设置

仿真模型的参数设置如表 2-2 所示, 为使仿真信号与实际信号采集的位置大体一致, 选择外圈竖直方向的振动加速度响应作为待验证的仿真振动信号, 仿真模型的采样频率与实验的采样频率相同, 为 12kHz。轴承的几何参数与 SKF-6203-RS 轴承一致。接触刚度按照 2.2.2 节计算, 轴承材料参数设为:  $E_1 = E_2 = 207\text{GPa}$ ,  $\mu_1 = \mu_2 = 0.3$ 。

表 2-2 滚动轴承仿真模型参数

仿真参数名称及符号 (单位)	参数数值	仿真参数名称及符号 (单位)	参数数值
转频 $f_r$ (Hz)	1797/60	重力加速度 $g$ (m/s <sup>2</sup> )	9.8
内圈-轴质量 $m_{in}$ (kg)	50	故障宽度 $L$ (mm)	0.36
轴-连接刚度 $k_{in}$ (N/m) [10]	$7.42 \times 10^7$	故障深度 $h$ (mm)	1
轴-连接阻尼 $c_{in}$ (Ns/m) [10]	1376.8	径向游隙 $c_r$ (m)	$2 \times 10^{-6}$
外圈-轴承座质量 $m_{out}$ (kg)	5	偏心距 $e$ (m)	$50 \times 10^{-6}$
轴承座-基础连接刚度 $k_{out}$ (N/m) [10]	$1.51 \times 10^7$	滚动体数量 $N$	8
轴承座-基础连接阻尼 $c_{out}$ (Ns/m) [10]	2210.7	滚动体直径 $D_b$ (mm)	6.746
接触刚度 $K$ (N/m <sup>1.5</sup> )	$8.753 \times 10^9$	轴承节径 $D_p$ (mm)	28.500

需要说明的是, 由于模型简化等因素, 仿真模型与实际系统必然存在一定差异。但是, 本文仿真的主要目的是能够获得描述某一型号轴承在特定转速下, 具有故障特性的振动信号, 并能以此为依据将仿真信号中的诊断知识迁移到实际问题中, 而并不需要精确地模拟出与某一特定机械设备完全一致的信号。因此与设

备相关的动力学参数，包括  $m_{in}$ 、 $k_{in}$ 、 $c_{in}$ 、 $m_{out}$ 、 $k_{out}$  及  $c_{out}$  等可根据经验设定。另外，当进行仿真信号与实际信号的对比时，幅值的绝对大小不是判断仿真合理性的主要因素，因此，将仿真和实际的时域信号幅值均线性映射至  $[-1, 1]$  的区间，仅定性地关注时域的冲击波形特性以及频域特性。

### 2.5.3 正常状态下仿真模型的验证与分析

正常状态下，取时域信号长度为 0.5s，频谱的频率范围为 0-500Hz，归一化的仿真信号与实际信号的时域波形和频谱对比如图 2-6 所示，正常状态下，轴承振动响应的典型特征是出现由于时变接触刚度造成的 VC (Varing Compliance) 振动，VC 振动的特征频率为：

$$f_{VC} = \frac{f_r}{2} \left( 1 - \frac{D_b}{D_p} \cos \alpha \right) \quad (2-34)$$

由图 2-6 可以看出，在仿真信号的时域波形中，能明显得到由于离心力造成的周期性波动，而在实际信号中信号的周期性被噪声湮没。在仿真信号频谱中，可得到由于时变接触刚度激励产生的 VC 频率，其分量较小。受复杂结构和环境因素的影响，实际信号的频谱中频率成分丰富，包含转频及其谐波成分，几乎无法找到相应的 VC 频率。

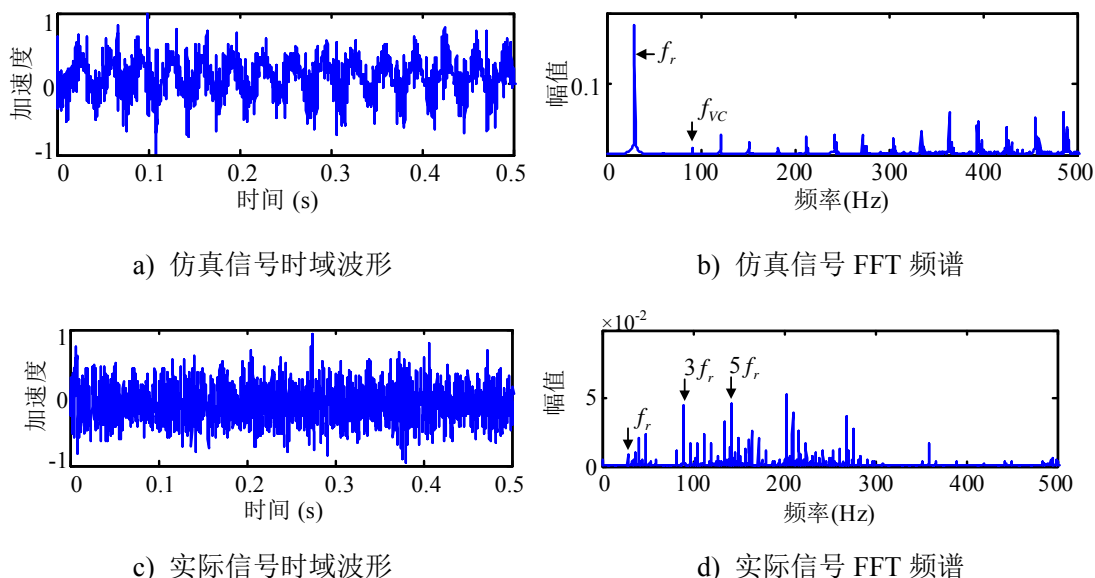


图 2-6 正常状态下仿真信号与实际信号对比

## 2.5.4 故障状态下仿真模型的验证与分析

内圈故障状态下，取时域信号长度为 0.5s，频谱的频率范围为 0-500Hz，归一化的时间波形和信号包络谱对比如图 2-7 所示，为了方便对比时域冲击特性，人为将冲击的相位对齐。图 2-7 中， $f_{in}$  表示内圈故障频率， $f_r$  表示转频。当轴承内圈发生故障时，内圈故障位置随轴一起转动，滚动体经过一次故障位置就会造成一次冲击，且由于滚动体经过故障位置时，所处的承载位置不同，脉冲力的大小和方向不同，因此造成的每个冲击的幅值有一定差异。当滚动体与故障的接触位置承载较小时，冲击很小，在时间波形上难以看出，这种差异与转频有关，反应到包络谱上，是出现以故障频率为中心频率，转频为边频的边频带。由图 2-7 可以看出，内圈故障的仿真信号与时域信号具有相似的冲击波形，在包络谱中，均有明显的故障频率、以故障频率为中心频率，转频为边频的边频带以及转频频率，这说明了内圈故障建模的合理性。

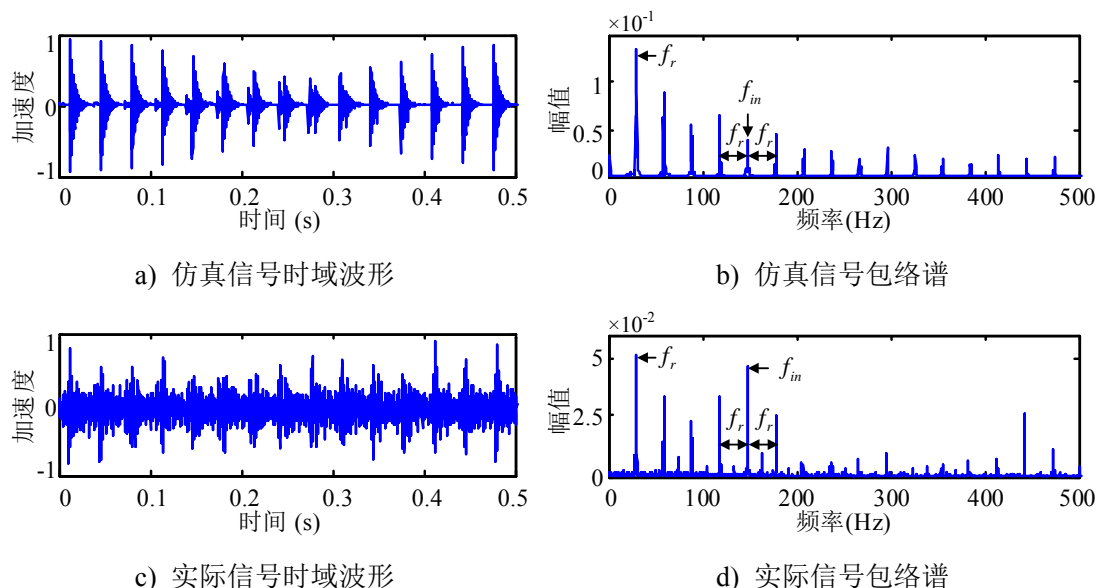


图 2-7 内圈故障状态下仿真信号与实际信号对比

外圈故障下，归一化的时间波形和信号包络谱对比如图 2-8 所示。图中， $f_o$  表示外圈故障频率， $f_r$  表示转频。可以看出仿真信号与实际信号的时域波形具有相似的周期性冲击特性，仿真与实际信号中冲击间隔相同。包络谱具有相似的频率特性，说明了外圈故障建模的合理性。与内圈故障不同，外圈发生故障时，故障位置不变，因此理论上，滚动体经过故障位置产生的冲击峰值不变化。由于仿真和实际系统中轴会存在一定的偏心，具有离心力，因此在包络谱中也出现了以故障频率为中心频率，转频为边频的边频带。

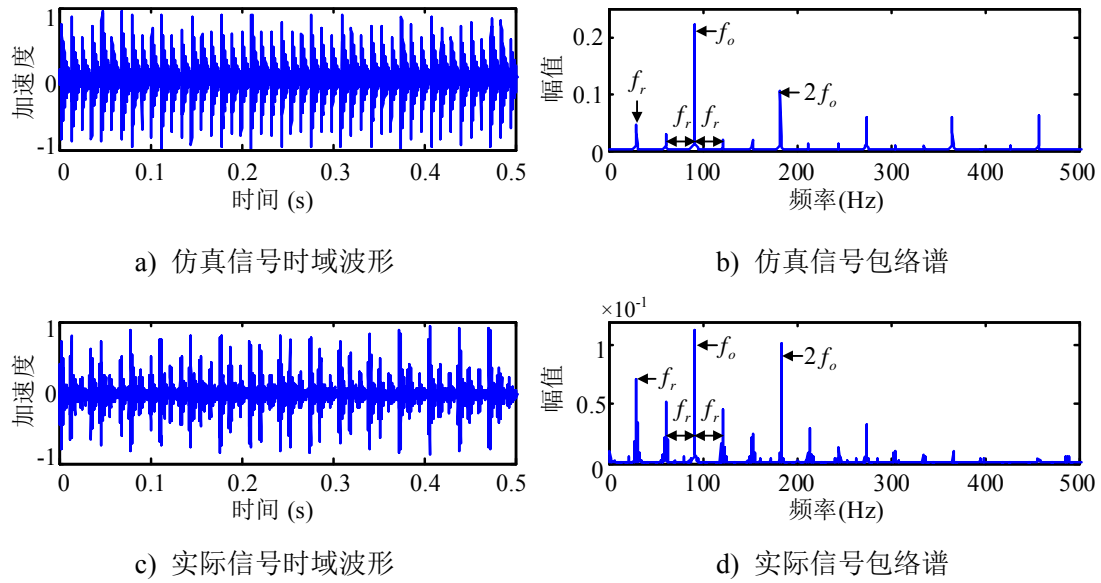


图 2-8 外圈故障状态下仿真信号与实际信号对比

## 2.6 本章小结

本章建立了滚动轴承的四自由度动力学模型，以获得能够反映滚动轴承在正常状态及内、外圈局部故障状态下滚动轴承振动特性的仿真信号。首先，对实际系统简化，建立了正常状态下的滚动轴承动力学模型；在此基础上，考虑故障的几何参数及滚动体的随机滑动因素，得到故障状态下的滚动轴承动力学模型；最后将仿真振动加速度信号与实际系统采集的振动加速度信号从时域和频域两个角度对比，验证了仿真信号能够反映典型的正常状态特性以及故障特性，说明了仿真模型的合理性，为从仿真模型中迁移诊断知识，在一定程度上解决实际故障诊断任务中缺少数据的难题奠定基础。

## 第3章 基于动力学仿真和迁移学习的滚动轴承智能故障诊断方法

### 3.1 引言

第2章建立了滚动轴承的故障仿真模型，并通过与实际信号对比，从时域波形和包络谱两个角度说明了仿真模型的合理性。本章以该仿真模型为基础，基于卷积神经网络和深度迁移策略，提出一种小样本条件下，利用仿真数据辅助实际问题诊断的滚动轴承智能故障诊断方法。首先，进一步分析仿真数据与实际数据在动力学、概率分布与频率分布上的共性与差异性，说明利用迁移策略将仿真模型的诊断知识迁移到实际诊断问题中的合理性与必要性。其次，提出小样本条件下，基于深度迁移学习的滚动轴承故障诊断方法的诊断框架，对卷积神经网络的相关理论进行介绍，随后明确了故障诊断问题的设定并给出诊断方法的详细流程。

### 3.2 仿真数据与实际数据的共性及差异性分析

从第2章的分析可知，通过滚动轴承动力学模型获得的仿真振动加速度信号与实际系统中采集的振动加速度信号，在故障状态下具有相似的冲击特性和包络特性，这是仿真数据和实际数据在信号特征上的共性，也是认为仿真信号中的诊断知识是可以应用到实际诊断问题中的基础。然而，仿真模型对实际系统进行了较大程度的简化，同时仿真模型中的参数设置很难与实际系统完全相同，这导致仿真数据与实际数据在很多方面存在差异性，使得仿真模型的诊断知识不能直接应用，需要引入迁移策略以更好地解决这种跨域学习的问题。本节从动力学、时域信号概率分布和频域中的频率成分三个角度分析，进一步说明仿真数据与实际数据之间的共性和差异性。

#### 3.2.1 动力学角度的共性及差异性

从物理的角度对滚动轴承动力学系统分析，滚动轴承动力学系统是一个典型的非线性参激振动系统，其中非线性的主要原因之一是滚动体与滚道之间的接触为非线性接触，体现在实际系统对应的动力学方程中，是方程中包含位移的1.5次方项。而系统中的参数激励主要为轴承运行中由于承载的滚动体数目和位置不断变化产生的时变刚度激励。除了时变刚度激励外，轴承系统中还包含由于制造、



安装误差等原因造成的微小偏心激励，以及轴承发生局部故障后产生的冲击激励等。作为一个非线性系统，其典型特点是可能产生超谐共振、亚谐共振和组合共振现象。当轴承具有故障时，故障冲击这种宽频激励可激出接近系统固有频率的更高次的谐波共振，造成系统不同激励频率的组合及倍频现象。仿真模型是根据滚动轴承系统的这些动力学特点建立的，因此这些特性在仿真系统和实际系统中是共享的，是仿真模型与实际系统的共性特征。

从仿真模型建立过程以及仿真信号和实际信号获取等角度来看，仿真模型与实际系统的信号表征形式在很多方面存在差异性。第一，在建立仿真模型的过程中，很难考虑到实际设备的所有细节，通过一定的简化条件，在建模时仅考虑所关心的关键部件，如在本文建立的模型中忽略了润滑油膜、电机及传感器等多个次要部分，仅保留了轴承内外圈相关的四个自由度。第二，在仿真模型中很难考虑由于环境等造成的不确定性因素与随机因素。第三，实际设备的信号均由传感器采集得到，传感器的安装位置、漂移以及信号在设备中复杂的传递特性都会影响最终的信号表征形式，而从仿真模型中获得的振动信号仅仅是通过动力学方程解出的振动加速度响应。这些建模过程中的简化与信号获取过程的差异使得仿真模型与实际系统的动力学响应的信号表征形式具有一定差异性。

### 3.2.2 概率分布的共性及差异性

传统的机器学习方法和深度学习方法的前提假设是训练数据与测试数据具有相同的概率分布，将这些方法用在故障诊断问题中，方法有效的前提为训练数据与测试数据来自同设备同工况，一旦这个条件不满足，诊断模型性能会有显著下降。仿真模型与实际系统也可以看做是具有相同型号相同转速轴承的两个不同的设备，因此，它们的振动信号在概率分布上应存在共性和差异性。

本节对仿真时域加速度信号与实际时域加速度信号的概率密度分布进行分析。这里的仿真对象仍为西储大学 SKF-6203 轴承，仿真参数与第二章相同，在内、外圈故障状态下设置与实际轴承相同的 3 种故障程度（0.18mm、0.36mm 和 0.53mm），取正常信号和故障信号的数据长度为 1200，即 0.1s。在计算概率密度前，对仿真信号和实际信号进行 0 均值 1 方差的标准化处理。不同健康状态下，仿真与实际时域信号的概率密度分布分别如图 3-1 和图 3-2 所示。由图 3-1 和图 3-2 可知，仿真信号和实际信号的概率分布有以下两点共性：（1）仿真信号与实际信号的正常状态下的概率分布相比于故障状态下的概率分布均更平缓，这是由于故障发生后，故障冲击会使得信号概率密度函数峰顶高度增高，而故障冲击是

在仿真模型和实际系统中均存在的特性。(2) 在仿真信号和实际信号的概率分布中，相同故障状态下不同程度的故障信号概率分布很接近，而不同故障类型的概率分布相差很大。然而，在仿真信号中，内圈故障信号的概率分布的尖峰程度高于外圈故障，而在实际信号中，外圈故障信号的尖峰程度高于内圈故障，这是仿真信号与实际信号概率分布的差异。

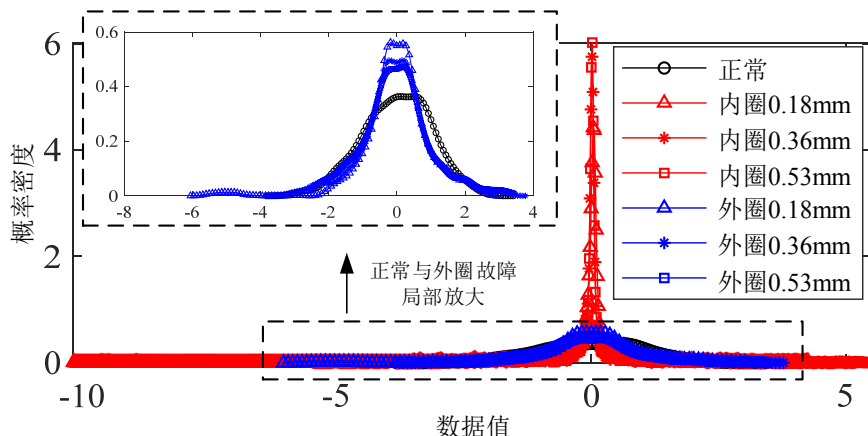


图 3-1 仿真时域信号不同健康状态下的概率密度分布

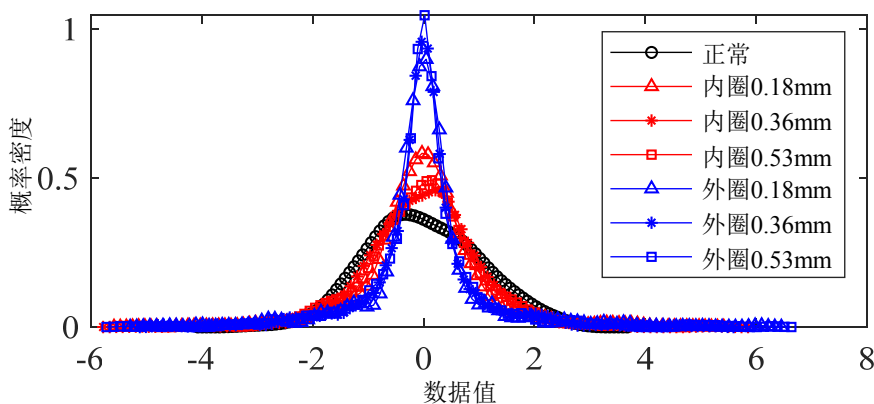


图 3-2 实际时域信号不同健康状态下的概率密度分布

为了比较相同健康状态下，仿真信号与实际信号的概率分布差异，在图 3-3 中给出了正常、内圈故障与外圈故障条件下两者信号概率分布的对比图。由上述分析得知，同故障类型下不同故障程度的信号，其概率分布类似，因此这里仅以故障程度为 0.36mm 为例，展示故障信号的概率分布。可以看出，正常状态下，仿真信号和实际信号的概率分布很相似，而故障状态下，仿真信号与实际信号的分布差异较大。这进一步说明了，虽然仿真信号和实际信号具有相似的冲击波形和故障特征频率特性，但其在概率分布上仍具有显著的差异性。因此，若仅在仿真数据上训练智能诊断模型，不进行迁移操作，直接在实际数据上诊断，不满足

传统机器学习方法和深度学习方法要求的训练数据与测试数据概率分布相同的条件，可能无法实现诊断。

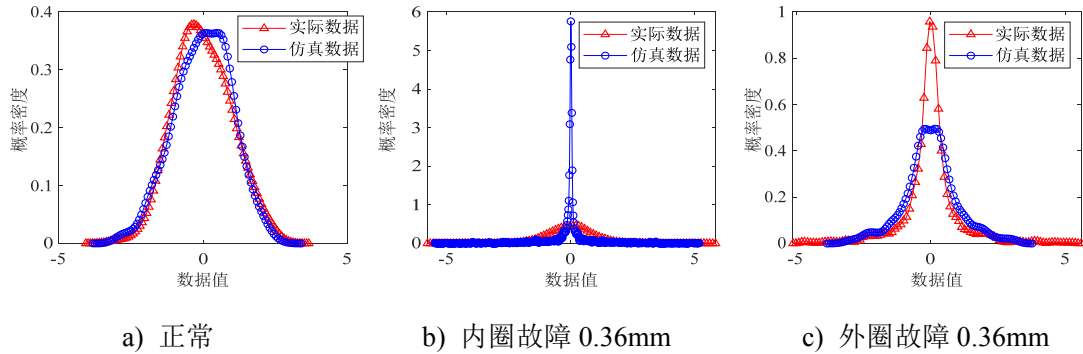


图 3-3 相同健康状态下仿真数据和实际数据概率分布对比

### 3.2.3 频率分布的共性及差异性

当轴承发生局部故障时，故障产生的冲击激励会激起系统中的高频成分，由于仿真模型对实际系统进行了较大程度的简化，仿真信号中的高频成分与实际系统中的高频成分应有一定差异。本文拟采用卷积神经网络构建故障诊断模型，在卷积神经网络中，卷积核是提取信号特征的重要部分，卷积核对信号特征的提取过程可以看成是滤波的过程，而在传统的信号处理领域，滤波的过程可以看做是两个信号在频域上的相乘，因此，仿真数据与实际数据在频率成分之间的差异性可能是一个影响知识迁移的重要因素。本节首先基于仿真模型，分析影响振动信号频率分布的滚动轴承动力学系统的物理参数，随后说明仿真数据与实际数据在频率分布上的共性与差异性。

#### 3.2.3.1 影响频率分布的系统物理参数分析

当故障发生时，振动信号中会出现固有频率调制故障频率的现象，即频谱上出现以固有频率为中心，以故障频率为边频的边频带，因此固有频率的大小对信号的频率分布起重要作用。研究影响频率分布的因素，可以归结为研究影响固有频率大小的因素。固有频率与系统的结构参数和物理参数有关，对于一个实际系统，我们很难改变其结构属性和物理参数以探究哪些参数会对固有频率产生什么样的影响。因此，本节仍从仿真模型出发，首先对其固有频率进行估计，随后探究影响轴承振动响应的频率分布的物理参数。

##### (1) 滚动轴承动力学系统的固有频率估计

当滚动轴承发生故障时，故障冲击这种宽频激励会激起系统中的固有频率。本节基于仿真模型，对滚动轴承动力学系统固有频率进行估计。

对一个线性系统，其一般形式的动力学方程为：

$$\mathbf{M}\ddot{\mathbf{X}} + \mathbf{C}\dot{\mathbf{X}} + \mathbf{K}\mathbf{X} = \mathbf{F} \quad (3-1)$$

式中  $\mathbf{M}, \mathbf{K}, \mathbf{C}$  ——系统的质量阵、刚度阵和阻尼阵；

$\mathbf{F}$  ——系统激励构成的向量；

$\mathbf{X}$  ——系统动力学响应构成的向量。

该系统的固有频率通常由频率方程计算：

$$|\mathbf{K} - \omega^2 \mathbf{M}| = 0 \quad (3-2)$$

对于滚动轴承四自由度非线性系统，估计其固有频率采用线性化刚度的处理方式<sup>[64]</sup>，计算步骤如下：

第一步，基于动力学方程求系统响应： $x_{in}, y_{in}, x_{out}, y_{out}$ ；

第二步，将响应代入非线性弹性恢复力表达式（2-23）及（2-24）中计算  $F_{HX}, F_{HY}$ ；

第三步，按照下式计算仅考虑弹性恢复力时的等效线性化刚度：

$$K_{xx\_xin} = \frac{1}{N_{fr}} \sum_{j=1}^{N_{fr}} \frac{\partial F_{HX}(j)}{\partial x_{in}(j)} = \sum_{j=1}^n \sum_{i=1}^N 1.5KH(\delta_{i,j}) \delta_{i,j}^{0.5} \sin^2 \theta_{i,j} \quad (3-3)$$

$$K_{xx\_xout} = -\frac{1}{N_{fr}} \sum_{j=1}^{N_{fr}} \frac{\partial F_{HX}(j)}{\partial x_{out}(j)} = \sum_{j=1}^n \sum_{i=1}^N 1.5KH(\delta_{i,j}) \delta_{i,j}^{0.5} \sin^2 \theta_{i,j} \quad (3-4)$$

$$K_{yy\_yin} = \frac{1}{N_{fr}} \sum_{j=1}^{N_{fr}} \frac{\partial F_{HY}(j)}{\partial y_{in}(j)} = \sum_{j=1}^n \sum_{i=1}^N 1.5KH(\delta_{i,j}) \delta_{i,j}^{0.5} \cos^2 \theta_{i,j} \quad (3-5)$$

$$K_{yy\_yout} = -\frac{1}{N_{fr}} \sum_{j=1}^{N_{fr}} \frac{\partial F_{HY}(j)}{\partial y_{out}(j)} = \sum_{j=1}^n \sum_{i=1}^N 1.5KH(\delta_{i,j}) \delta_{i,j}^{0.5} \cos^2 \theta_{i,j} \quad (3-6)$$

式中  $N_{fr}$  ——参与计算等效线性化刚度的样本点数。

第四步，考虑轴和基座的连接刚度，可得该系统的刚度阵为（交叉刚度项很小，不考虑）：

$$\mathbf{K} = \begin{pmatrix} K_{xx\_xin} + k_{in} & 0 & 0 & 0 \\ 0 & K_{yy\_yin} + k_{in} & 0 & 0 \\ 0 & 0 & K_{xx\_xout} + k_{out} & 0 \\ 0 & 0 & 0 & K_{yy\_yout} + k_{out} \end{pmatrix} \quad (3-7)$$

四自由度系统的质量阵为：

$$\mathbf{M} = \begin{pmatrix} m_{in} & 0 & 0 & 0 \\ 0 & m_{in} & 0 & 0 \\ 0 & 0 & m_{out} & 0 \\ 0 & 0 & 0 & m_{out} \end{pmatrix} \quad (3-8)$$

将质量阵和线性化刚度阵代入频率方程，得到系统的角频率形式的固有频率 $\omega$ ，对于四自由度系统，应具有四阶固有频率，分别对应内圈-轴- $x$ ，内圈-轴- $y$ ，外圈-基座- $x$  及外圈-基座- $y$  四个自由度。以（Hz）为单位的系统各自由度的固有频率 $f_i$ 为：

$$f_i = \frac{\omega_i}{2\pi}, i=1,2,3,4 \quad (3-9)$$

按照上述计算方式，计算第二章仿真对象的固有频率，结果如表 3-1 所示。可以看出，基于不同健康状态的信号估计的系统的固有频率，差别不大。从物理上分析，滚动轴承的局部故障几何尺寸很小，对整个设备的物理参数影响很小。而固有频率只跟结构的物理属性有关，因此轴承发生局部故障并不会对固有频率产生较大影响。

表 3-1 固有频率估计值（Hz）

	内圈-轴- $x$ (IR- $x$ )	内圈-轴- $y$ (IR- $y$ )	外圈-基座- $x$ (OR- $x$ )	外圈-基座- $y$ (OR- $y$ )
基于正常响应估计结果	210.3	237.1	378.0	512.6
基于内圈故障响应估计结果	210.3	236.4	378.0	509.4
基于外圈故障响应估计结果	210.3	234.9	377.8	502.4

图 3-4 至图 3-6 给出了不同健康状态下，内外圈  $y$  方向的振动响应的频谱。可以看出，正常状态下，由于系统中只有时变刚度激励，没有冲击激励，因此，只有系统中与时变刚度激励接近的较低阶的固有频率被激励出来。而故障状态下，故障冲击可激励出系统更高阶的固有频率。

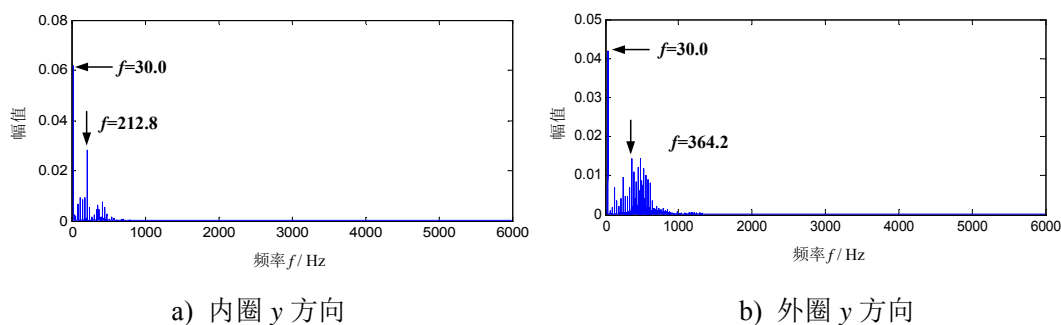


图 3-4 正常状态下信号振动加速度响应频谱

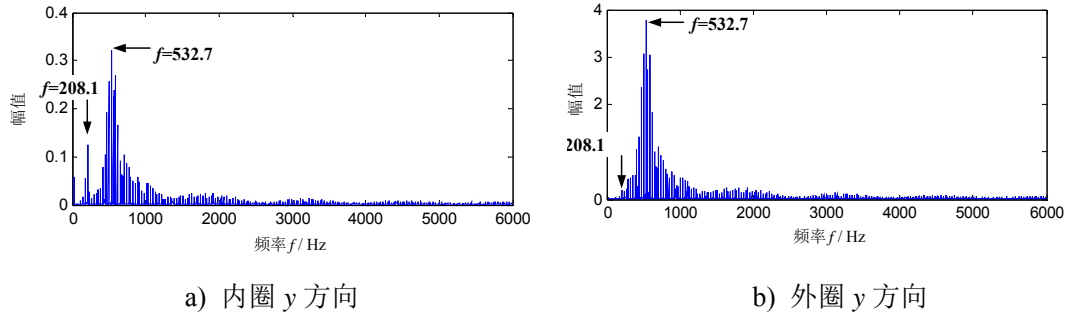


图 3-5 内圈故障下信号振动加速度响应频谱

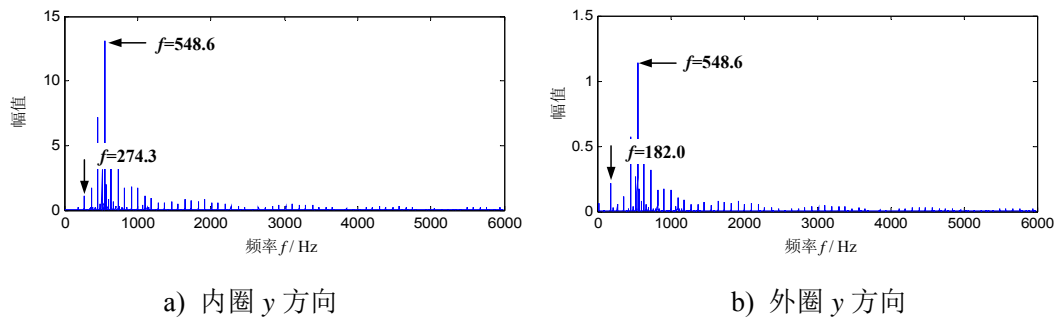


图 3-6 外圈故障下信号振动加速度响应频谱

## (2) 影响固有频率的系统物理参数分析

固有频率反映了系统的固有属性，系统属性的变化可归结为系统的物理参数上的变化。本节对滚动轴承动力学的物理参数进行分析，研究哪些物理参数会影响系统的固有频率。

滚动轴承动力学系统的物理参数可以分成两类，一类是与设备属性相关的参数，另一类是仅反映轴承自身属性的参数，具体如表 3-2 所示。

表 3-2 滚动轴承动力学系统的物理参数分类

与设备属性相关的参数	反映轴承自身属性的参数
转频 $f_r$ (Hz)	接触刚度 $K$ (N/m <sup>1.5</sup> )
内圈-轴质量 $m_{in}$ (kg)	径向游隙 $c_r$ (m)
轴-连接刚度 $k_{in}$ (N/m)	轴承型号
轴-连接阻尼 $c_{in}$ (Ns/m)	—
外圈-基座质量 $m_{out}$ (kg)	—
轴承座-基础连接刚度 $k_{out}$ (N/m)	—
轴承座-基础连接阻尼 $c_{out}$ (Ns/m)	—
偏心距 $e$ (m)	—

为了分析这些参数对滚动轴承固有频率的影响, 本文采取单一变量的方式, 在研究某一参数对固有频率的影响时, 固定其他参数不变, 设定待分析的参数在一定合理区间内取不同值, 并计算不同参数下的固有频率数值, 得到固有频率随该参数变化的变化规律。

需要说明的是, 虽然对于特定型号的轴承, 按照 2.2.2 节的方式可以估计出其接触刚度  $K$ , 但由于轴承部件装配、制造等原因, 对于同一个轴承, 实际接触刚度的值也会在一定范围内变化, 因此我们也把接触刚度作为一个可变的物理参数。另外, 对于不同的轴承型号, 我们在仿真时按照 2.2.2 节的方式估计其接触刚度, 因此轴承型号不同隐含了两个条件, 一是接触刚度不同, 二是轴承的几何参数不同。

所研究的 6 种不同型号轴承的几何参数<sup>[65]</sup>和接触刚度如表 3-3 所示。各个参数的基准值及变化范围设定如表 3-4 所示。

表 3-3 不同型号轴承的几何参数及接触刚度

轴承型号	节径 $D_p$ (mm)	滚动体直径 $D_b$ (mm)	滚动体数量 $N_b$	接触刚度 $K$ ( $\times 10^9 \text{N/m}^{1.5}$ )
6203	28.500	6.747	8	8.75
6205	33.500	7.938	9	9.50
6211	38.500	7.938	10	12.74
6304	77.500	14.288	7	10.41
6305	36.000	9.525	7	11.43
6403	43.500	11.500	6	12.00

表 3-4 滚动轴承动力学系统的物理参数及其变化范围设定

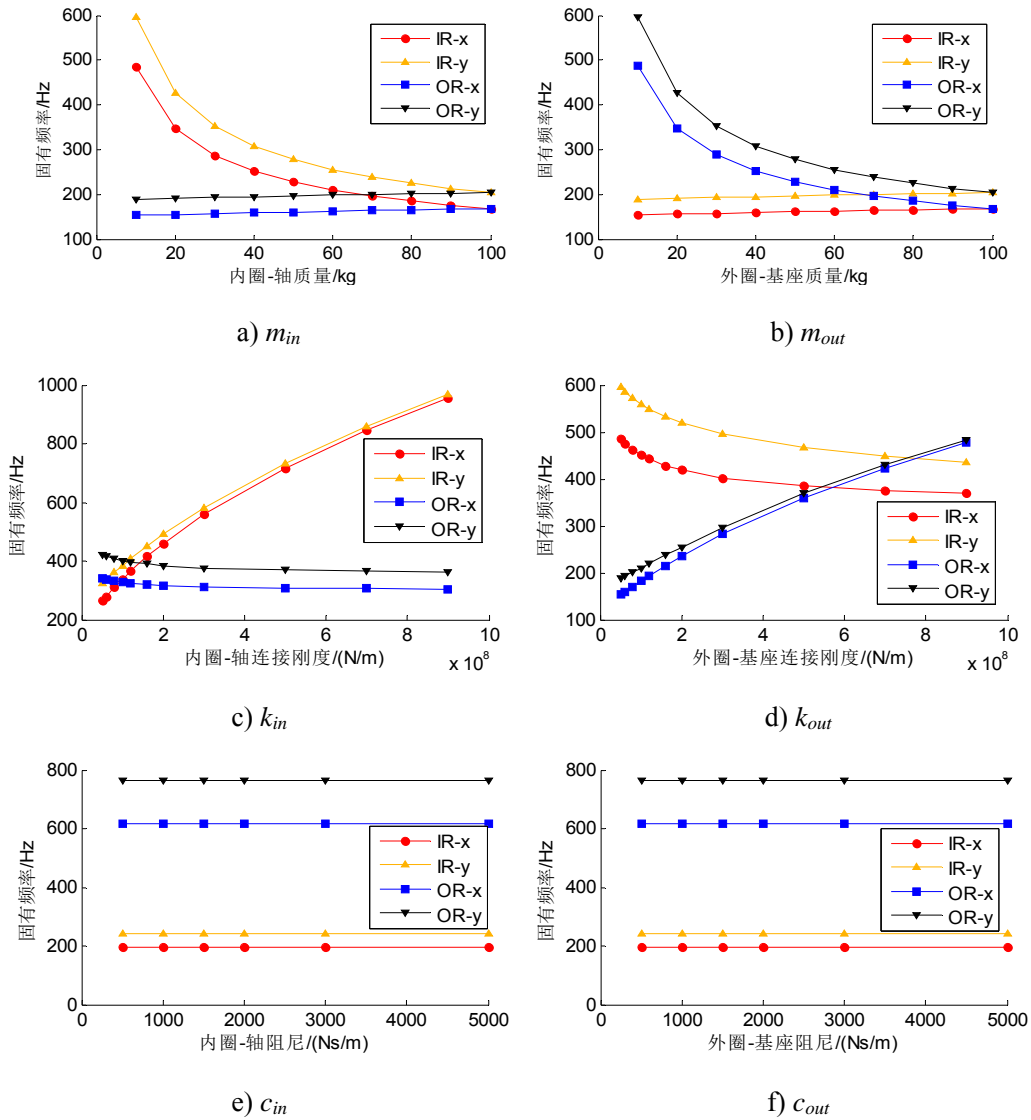
待测 参数	$m_{in}$ (kg)	$m_{out}$ (kg)	$k_{in}$ ( $\times 10^7$ N/m)	$k_{out}$ ( $\times 10^7$ N/m)	$c_{in}$ (Ns/m)	$c_{out}$ (Ns/m)	$f_r$ (Hz)	$K$ ( $\times 10^9$ N/m <sup>1.5</sup> )	$c_r$ ( $\mu$ m)	$e$ ( $\mu$ m)	轴承型号
$m_{in}$ (kg)	[10, 100]	100	5	5	1000	2000	1797/60	14	2	50	6203
$m_{out}$ (kg)	100	[10, 100]	5	5	1000	2000	1797/60	14	2	50	6203
$k_{in}$ ( $\times 10^7$ N/m)	25	15	[5, 90]	5	1000	2000	1797/60	14	2	10	6203
$k_{out}$ ( $\times 10^7$ N/m)	10	100	5	[5, 90]	1000	2000	1797/60	14	2	10	6203
$c_{in}$ (Ns/m)	50	5	5	5	[5, 50] $\times 10^2$	2000	1797/60	14	2	10	6203
$c_{out}$ (Ns/m)	50	5	5	5	1000	[5, 50] $\times 10^2$	1797/60	14	2	10	6203
$f_r$ (Hz)	10	20	5	5	1000	2000	[900, 5000]/60	14	2	10	6203
$K$ ( $\times 10^9$ N/m <sup>1.5</sup> )	50	5	5	5	1000	2000	1797/60	[8, 20]	2	10	6203
$c_r$ ( $\mu$ m)	50	5	5	5	1000	2000	1797/60	14	[1, 10]	10	6203
$e$ ( $\mu$ m)	50	5	5	5	1000	2000	1797/60	14	2	[1, 50]	6203
轴承型号	50	5	5	5	1000	2000	1797/60	14	2	10	6203, 6205, 6211, 6304, 6305, 6403



经计算,系统各自由度固有频率随设备属性相关参数变化曲线如图 3-7 所示。由图 3-7 可得以下两点结论:

(1) 质量 ( $m_{in}$ ,  $m_{out}$ ) 和刚度 ( $k_{in}$ ,  $k_{out}$ ) 的变化对系统固有频率的影响较大,而阻尼 ( $c_{in}$ ,  $c_{out}$ )、偏心距  $e$  和转频  $f_r$  对固有频率几乎没有影响。 $m_{in}$  ( $m_{out}$ ) 增大时,内圈(外圈)自由度的固有频率减小,外圈(内圈)自由度固有频率略有增大; $k_{in}$  ( $k_{out}$ ) 增大时,内圈(外圈)自由度的固有频率增大,外圈(内圈)自由度固有频率略有减小;

(2) 与轴承内圈相关的动力学参数变化时,对内圈的两个自由度频率影响较大;与轴承外圈相关的动力学参数变化时,对外圈的两个自由度影响较大。



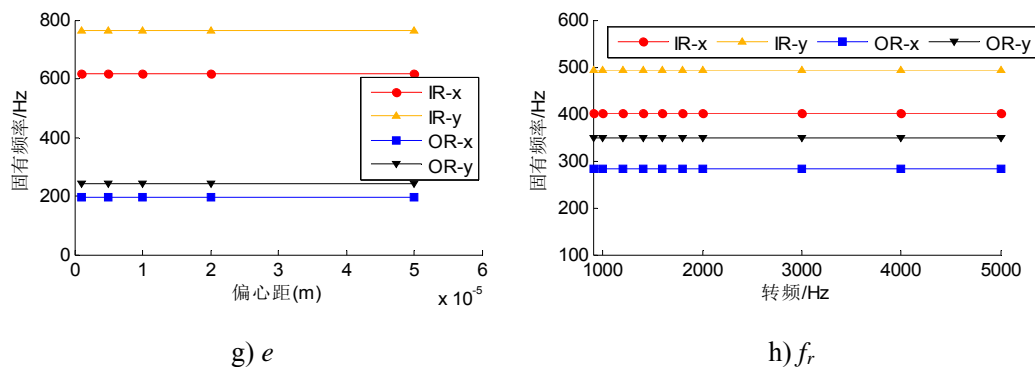


图 3-7 系统各自由度固有频率随设备属性相关参数变化曲线

系统各自由度固有频率随轴承自身属性参数变化曲线如图 3-8 所示。由图 3-8 可得以下两点结论：

(1) 与设备相关参数变化造成内外圈自由度固有频率反向变化不同，轴承自身属性变化会对各自由度的频率产生同向的影响。当径向游隙  $c_r$  增加时，系统各自由度的固有频率均下降，从物理上分析，游隙增加，轴承部件之间的装配更松，可以看成是一种刚度的降低，因此固有频率会有所下降；当接触刚度  $K$  增加时，系统各自由度的固有频率均增加。

(2) 不同轴承型号的固有频率有差异，说明轴承型号是造成系统之间差异性的因素之一。

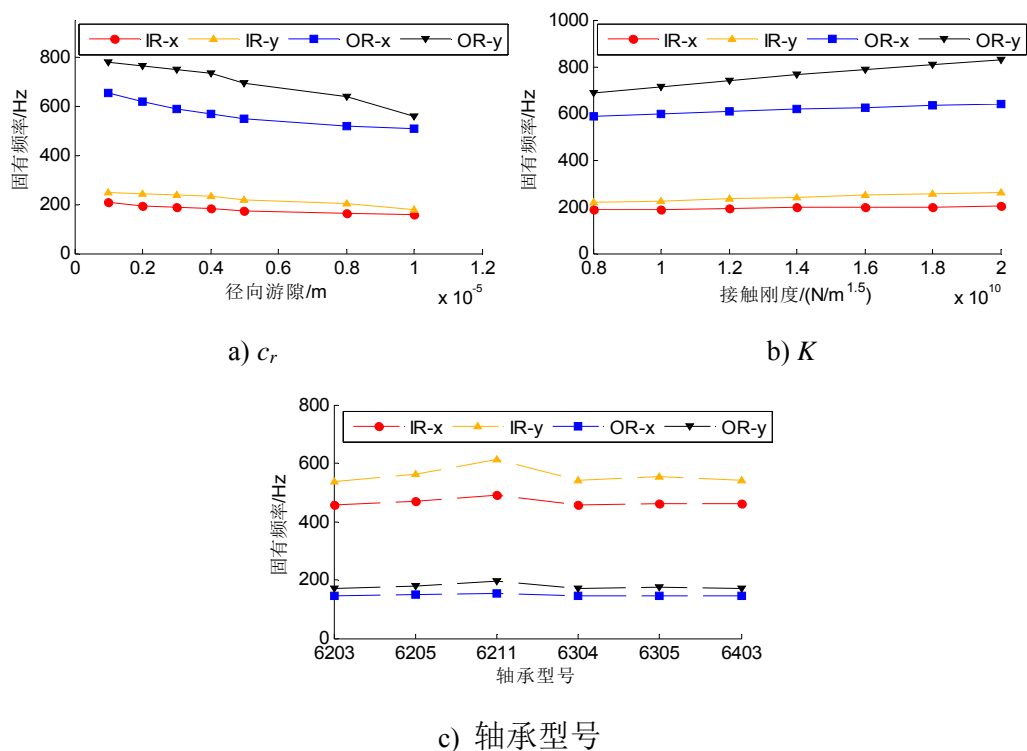


图 3-8 系统各自由度固有频率随轴承自身属性参数变化曲线

表 3-5 总结了各个参数对系统各自由度固有频率的影响。通过本节分析，在一定程度上明确了影响系统固有频率的物理参数，得到了影响系统频率分布的原因。

表 3-5 滚动轴承系统物理参数对系统各自由度固有频率的影响

参数变化	内圈-x	内圈-y	外圈-x	外圈-y
$m_{in}$ 增大	减小	减小	略有增大	略有增大
$m_{out}$ 增大	略有增大	略有增大	减小	减小
$k_{in}$ 增大	增大	增大	略有减小	略有增大
$k_{out}$ 增大	略有减小	略有减小	增大	增大
$c_{in}$ 增大		无影响		
$c_{out}$ 增大		无影响		
$e$ 增大		无影响		
$f_r$ 增大		无影响		
$c_r$ 增大		减小		
$K$ 增大		增大		
轴承型号变化		有影响		

### 3.2.3.2 仿真数据与实际数据频率分布的共性与差异性

从滚动轴承的故障机理上分析，当故障发生时，振动信号中会出现固有频率调制故障频率的现象，即频谱上出现以固有频率为中心，以故障频率为边频的边频带，这是仿真数据与实际数据在频率分布上的共性。而仿真数据与实际数据频率分布的差异性首先来源于固有频率数值上的差异性。由上节的分析可知，系统的多个物理参数会对系统的固有频率产生不同程度的影响。虽然仿真模型与实际系统的轴承型号与转速相同，但其他的参数可能会有一定程度的差异，这些差异会使得仿真模型和实际系统的固有频率不同，进而造成信号的频率分布上的差异。

另外，以 SKF-6203 和 SKF-6205 轴承具有内圈 0.36mm 故障为例，外圈 y 方向振动加速度信号的仿真频谱和实际频谱图如图 3-9 和图 3-10 所示。可以看出，在仿真信号频谱中只有 1 个或 2 个频率集中分布的区域，而实际信号中有多个频率集中分布的区域。这是由于，在仿真模型中仅考虑了与内圈、外圈有关的四个自由度，而实际系统中的自由度更多，结构更复杂，可能会产生高阶的固有频率

值，即实际系统中可能会出现更多的集中分布的频带，而不仅限于仿真数据中的中低频带。这是仿真数据与实际数据在频率分布上又一差异。

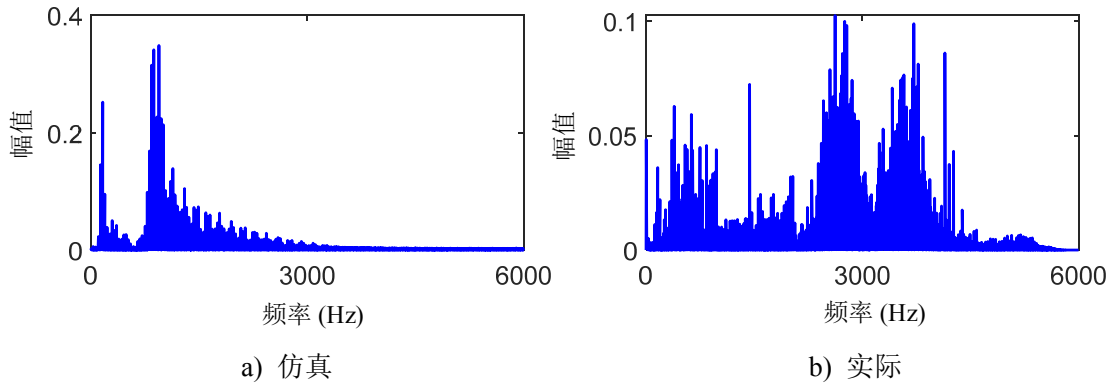


图 3-9 SKF-6205 轴承外圈 y 方向振动加速度信号频谱

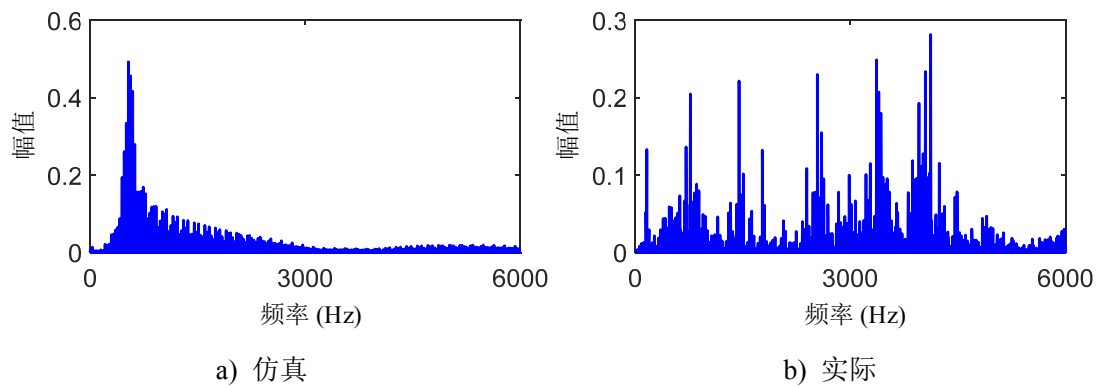


图 3-10 SKF-6203 轴承外圈 y 方向振动加速度信号频谱

### 3.2.4 仿真数据与实际数据的共性及差异性总结

将 3.2.1 节至 3.2.3 节仿真数据与实际数据的共性及差异性进行简要总结，如表 3-6 所示。仿真数据与实际数据的这些共性，说明仿真数据中的诊断知识是可以迁移到实际诊断问题中的；而它们之间的差异性，则说明不加处理地直接利用仿真数据中的知识，即直接按照传统的基于数据驱动的诊断方法的处理方式，在仿真数据上训练诊断模型，直接将其在实际数据上测试，可能不能起到对实际诊断问题的辅助作用，因此考虑迁移策略，在一定程度上消除这种差异性是有必要的。

表 3-6 仿真数据与实际数据的共性及差异性总结

分析角度	共性	差异性
动力学	(1) 均为非线性参激系统； (2) 具有时变刚度激励、故障冲击激励和参数激励。	(1) 与仿真模型相比，实际系统的结构更复杂； (2) 与仿真模型相比，实际系统包含更多不确定性因素和随机因素； (3) 仿真模型与实际系统获取信号的方式不同。
概率分布	(1) 正常状态下的信号概率分布比故障状态下平缓； (2) 相同故障状态下不同故障信号概率分布接近，不同故障类型概率分布较大； (3) 正常状态的概率分布形状接近。	故障状态下的概率分布形状差异较大。
频率分布	当存在故障冲击时，频谱出现以固有频率为中心的集中分布频带。	(1) 固有频率的数值不同； (2) 当存在故障冲击时，频谱中集中分布的频带的个数不同。

### 3.3 基于动力学仿真和迁移学习的滚动轴承故障诊断框架

上一节揭示了仿真数据与实际数据在动力学、概率分布和频率分布上的共性和差异性，这些共性和差异性使得仿真数据中的知识可以但不能直接用于实际问题的诊断中，需要考虑适当的迁移策略在利用诊断知识时保留某些共性，同时在一定程度上消除差异性。本节提出基于动力学仿真和迁移学习滚动轴承故障诊断方法，该方法利用卷积神经网络（CNN）和网络参数迁移策略，基于 CNN 从时域信号直接学习振动信号特征，并利用网络参数迁移及微调的方式将仿真数据的知识迁移到实际数据的诊断中，以实现小样本条件下滚动轴承的智能故障诊断。

#### 3.3.1 卷积神经网络概述

卷积神经网络是一类包含卷积运算的前馈神经网络<sup>[65]</sup>，其结构中主要包含输入层、卷积层、激活层、池化层、全连接层及输出层，其结构示意图如图 3-11 所示。在卷积神经网络中有三种典型操作：卷积操作、非线性激活操作和池化操作。

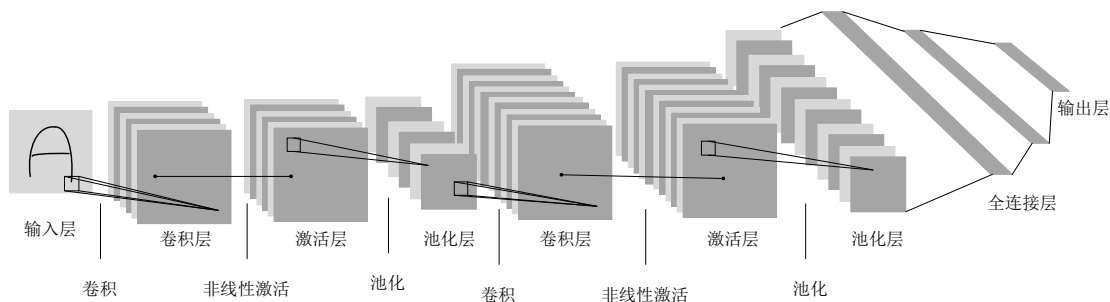


图 3-11 CNN 结构示意图

### (1) 卷积操作

在卷积层中，通过输入与核函数的卷积运算获得网络输入的特征映射，每个卷积核以相同的权重和偏置提取信号相同的局部特征<sup>[40,55]</sup>。卷积操作类似于信号处理中的滤波操作，每个卷积核可以看成是一个滤波器。二维卷积计算的示意图如图 3-12 所示。在一个卷积层中，通常可设置多个不同的卷积核从不同的角度提取信号的局部特征。其表达式如下：

$$\mathbf{x}_j^l = \mathbf{x}^{l-1} * \mathbf{w}_j^l + \mathbf{b}_j^l \quad (3-10)$$

式中  $\mathbf{w}_j^l$ ,  $\mathbf{b}_j^l$ ——第  $l$  层的第  $j$  个卷积核的权重矩阵和偏置向量；

$\mathbf{x}_j^l$ ——第  $l$  层经过第  $j$  个卷积核运算后的输出；

$\mathbf{x}^{l-1}$ ——第  $l-1$  层的输出。

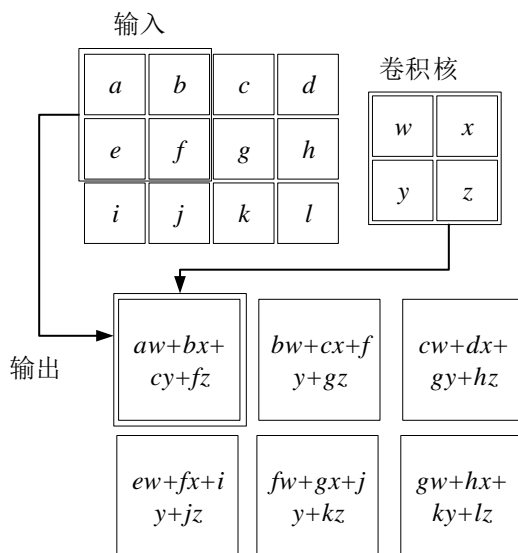


图 3-12 二维卷积计算示意图

在卷积神经网络中，采用离散卷积的运算方式，在实际应用中可能会用到其变体，而不是标准的离散卷积<sup>[65]</sup>。对于一个一维序列  $x(n)$ ，其与一维序列  $h(n)$  的

离散卷积运算的标准表达式如下：

$$y(n) = x(n) * h(n) = \sum_{i=0}^{N-1} x(i) h(i-n) \quad (3-11)$$

### (2) 非线性激活操作

非线性激活操作利用非线性映射函数增强卷积神经网络的特征表示能力。在卷积神经网络中，典型的激活函数为修正线性单元（Rectified Linear Unit, ReLU），ReLU 函数的表达式如下：

$$\mathbf{x}^l = \text{ReLU}(\mathbf{x}^{l-1}) = \max\{\mathbf{x}^{l-1}, 0\} \quad (3-12)$$

### (3) 池化操作

池化操作可以看做信号处理中的降采样过程，通过池化的方式降低特征维数，进而减少网络所需学习的参数，提高计算效率。在卷积神经网络中，常见的两种池化方式为平均池化（mean pooling）和最大池化（max pooling）。平均池化和最大池化示意图如图 3-13 所示。

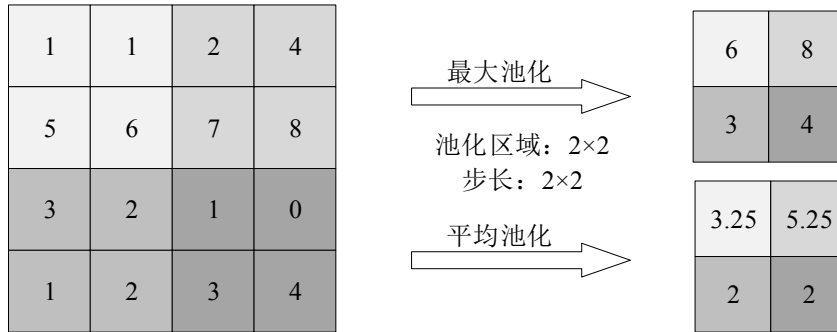


图 3-13 池化操作示意图

平均池化以所选取的池化区域的平均值来表征整个池化区域的特性。其表达式如下：

$$x_r^l = \text{mean}_{(r-1)W+1 \leq i \leq rW} \{\mathbf{x}^{l-1}(i)\} \quad (3-13)$$

式中  $x_r^l$  ——第  $l$  层第  $r$  个池化区域的输出；

$W$  ——池化区域的宽度。

最大池化以所选取的池化区域的最大值来表征整个池化区域的特性。其表达式如下：

$$x_r^l = \max_{(r-1)W+1 \leq i \leq rW} \{\mathbf{x}^{l-1}(i)\} \quad (3-14)$$

#### (4) 全连接层

卷积神经网络的全连接层部分可以看成是一个多层前馈全连接神经网络，通过全连接层的操作，可将基于卷积、非线性激活和池化操作后得到的信号特征映射到标签空间，最终实现对输入的分类。这个映射的过程可以看成是提取与信号标签更加相关的更深层次特征的过程。

与前馈神经网络相同，卷积神经网络的训练采用反向传播算法。在训练过程中可利用某些优化算法及正则化策略来提高网络模型的训练速度与泛化能力，例如 Adam 优化、随机梯度下降等优化算法，Dropout、early stopping、权重衰减等正则化策略。这些内容不是本文研究的重点，因此不做展开介绍，可参考文献[65]。

### 3.3.2 故障诊断问题设定

本文所要解决的问题是故障诊断中，实际故障数据难以获取的问题，解决的途径是从仿真模型中迁移诊断知识到实际设备的诊断中，这是一个迁移学习问题。为了方便叙述，首先引入迁移学习的相关定义和符号描述。

#### 3.3.2.1 迁移学习相关概念

对于一个有标签的数据集  $\{X, Y\} = \{x_i, y_i\}_{i=1}^n$ ，其中， $x_i$  为数据集中的第  $i$  个样本， $y_i$  为其标签，定义  $\mathcal{X}$  为描述  $X$  的特征空间，即  $X = \{x_i\}_{i=1}^n \in \mathcal{X}$ ， $\mathcal{P}(X)$  为  $X$  的边缘概率分布， $\mathcal{Y}$  为描述标签  $y$  的标签空间，即  $Y = \{y_i\}_{i=1}^n \in \mathcal{Y}$ ， $f(X) = \mathcal{P}(Y|X)$  为条件概率分布。在迁移学习中，定义域和任务如下<sup>[67]</sup>：

**定义 1 (域  $\mathcal{D}$ )**：域由样本的特征空间及其边缘概率分布描述，即  $\mathcal{D} = \{\mathcal{X}, \mathcal{P}(X)\}$ 。

**定义 2 (任务  $\mathcal{T}$ )**：任务由标签的标签空间及目标预测函数描述，即  $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ 。

在任务  $\mathcal{T}$  中，目标预测函数  $f(\cdot)$  基于训练数据  $\{x_i, y_i\} (x_i \in X, y_i \in \mathcal{Y})$  学习得到，用于预测一个新样本  $x$  对应的标签，从概率的角度，可将其描述为： $f(\cdot) = \mathcal{P}(y|x)$ 。

基于上述两个定义，用  $\mathcal{D}^s$  表示源域， $D_s = \{(x_{s1}, y_{s1}), \dots, (x_{sn_s}, y_{sn_s})\}$  为源域数据，其中  $x_{si} \in \mathcal{X}_s$  为样本， $y_{si} \in \mathcal{Y}_s$  为样本标签。 $\mathcal{T}^s$  表示源域学习任务； $\mathcal{D}^t$  表示目标域， $D_t = \{(x_{t1}, y_{t1}), \dots, (x_{tm_t}, y_{tm_t})\}$  为目标域数据， $\mathcal{T}^t$  表示目标域任务。定义迁移学习问题如下<sup>[67]</sup>：

**定义 3 (迁移学习问题)**：给定源域  $\mathcal{D}^s$  和源域学习任务  $\mathcal{T}^s$ ，目标域  $\mathcal{D}^t$  和目标域学习任务  $\mathcal{T}^t$ ，迁移学习问题是在  $\mathcal{D}^s \neq \mathcal{D}^t$  ( $\mathcal{X}^s \neq \mathcal{X}^t$  或  $\mathcal{P}_s(X) \neq \mathcal{P}_t(X)$ ) 或



$T^s \neq T^t$  ( $\mathcal{Y}_s \neq \mathcal{Y}_t$  或  $\mathcal{P}(Y_s|X_s) \neq \mathcal{P}(Y_t|X_t)$ ) 的条件下, 利用源域和源域学习任务中的知识帮助目标域更好地完成目标域任务, 即获得更好的  $\mathcal{P}(Y|X)$ 。

若源域和目标域相同, 源域任务和目标域任务相同, 这样的问题退化成传统的机器学习问题。

### 3.3.2.2 小样本条件下故障诊断问题设定

本文中, 期望把从仿真数据中学习到的诊断知识迁移到实际问题中, 在实际待诊断设备仅包含少量样本, 无法构建具备泛化能力的诊断模型时, 利用仿真数据和实际数据的共性特征帮助实际问题实现诊断。因此, 将仿真数据作为源域, 实际数据作为目标域, 源域和目标域不同, 但源域任务和目标域任务相同, 即:  $\mathcal{X}^s = \mathcal{X}^t$ ,  $\mathcal{P}(X^s) \neq \mathcal{P}(X^t)$ ,  $T^s = T^t$ 。所讨论的诊断问题具体设定如下:

- (1) 仿真模型与实际待诊断设备的轴承型号, 所包含的故障类别及工况 (本文指内圈转速) 相同, 仿真信号和实际信号的采样频率相同。
- (2) 对于仿真模型, 可获取各种健康状态下的大量有标签数据。
- (3) 对于实际待诊断设备, 仅可获取很少量的有标签样本。

### 3.3.3 基于动力学仿真和迁移学习的滚动轴承故障诊断流程

基于上述诊断问题的设定, 本文提出小样本条件下, 将仿真模型的诊断知识迁移到实际诊断问题中的深度迁移诊断方法。该方法基于大量仿真数据预训练 CNN 诊断模型, 迁移预训练诊断模型中的网络参数到实际诊断网络中, 并用可获取的少量实际样本微调整个网络。另外, 该方法中滚动轴承的信号特征直接由 CNN 提取, 即 CNN 的输入为时域信号, 而不是经过人工提取后的特征或频谱, 特征提取和故障识别在一个网络中完成。

具体地, 对于一个滚动轴承诊断问题, 该方法的具体操作步骤如下:

#### 第一步: 数据获取

从滚动轴承的仿真模型中获得大量不同健康状态下的仿真振动加速信号, 并收集待诊断设备的实测振动加速度信号。

在该步骤中, 待诊断轴承的轴承几何参数, 故障类别, 轴的转速及轴承信号的采样频率应为已知, 以满足问题设定中的 (1)。另外, 为了在一定程度上提高诊断模型的泛化能力, 仿真数据中对于每种故障模式应该在合理范围内包含多种不同的故障尺寸, 本文认为, 故障尺寸的下限应为故障特征不能完全被噪声湮没; 上限为, 当考虑仿真模型对应的实际轴承时, 在该故障尺寸下, 经验地认为轴承能够运转而不是完全失效。

## 第二步：数据预处理

在这个故障诊断问题中，期望以时域振动加速度信号作为网络输入，通过 CNN 网络学习到不同健康状态下信号的不同动态特性，由于幅值造成的信号差异性应该被消除。因此，对所有样本进行均值为 0 方差为 1 的标准化处理，即：

$$x' = \frac{x - \bar{x}}{\sigma} \quad (3-15)$$

式中  $x$ ——原始样本；

$\bar{x}, \sigma$ ——样本的均值和方差；

$x'$ ——归一化后的样本。

## 第三步：预训练 CNN 诊断模型

在保持类别平衡的条件下，将经过预处理后的仿真样本随机划分为训练样本和测试样本，用大量训练样本作为网络输入，预训练 CNN 诊断模型，预训练诊断模型的性能在测试样本上测试。

在该步骤中，所选取的 CNN 网络结构和参数应能保证预训练模型能够对仿真数据实现准确并稳定的识别，这是可以进行微调操作，实现知识迁移的前提。

## 第四步：迁移 CNN 网络参数

将预训练 CNN 网络的所有权重和偏置，作为诊断实际设备的 CNN 网络的权重和偏置。

## 第五步：微调预训练的诊断模型

在保持类别平衡的条件下，将经过预处理后实际样本随机划分为训练样本和测试样本，其中训练样本数量很少，以模拟实际过程中可以获取的少量的实际待诊断设备的样本。将少量训练样本作为实际诊断 CNN 网络的输入，微调整个网络，微调后的网络为最终的诊断模型，诊断模型的性能在测试样本上测试。

第四步和第五步是该方法实现知识迁移的关键步骤。在 CNN 网络中，权重和偏置为通过反向传播算法优化得到的网络参数，可以认为它们能够反映数据集诊断知识。由于实际诊断情形中可获取的数据量很少，在如此少的样本上 CNN 难以学到有效的诊断知识。因此，为了尽可能多地利用从仿真模型中学习到的诊断知识，该方法中迁移了预训练 CNN 诊断模型中的所有学习到的参数，包括卷积层和全连接层的所有权重和偏置。

基于上述步骤，该诊断方法的流程图如图 3-14 所示。

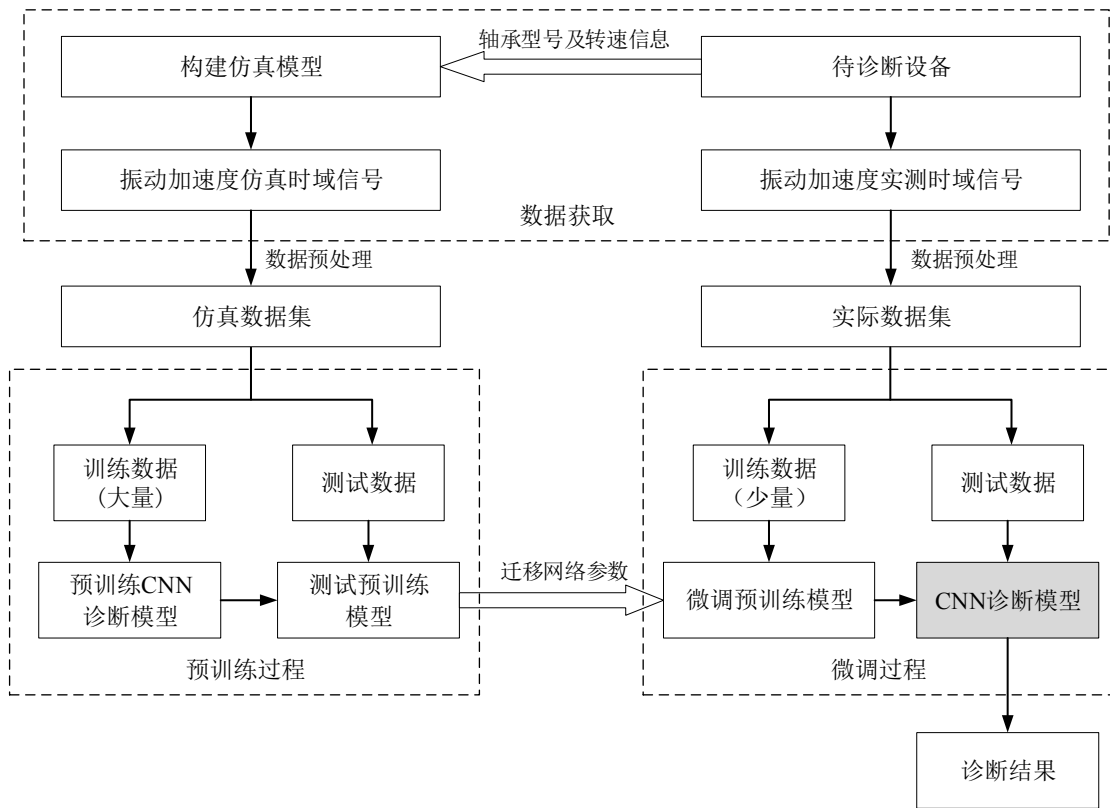


图 3-14 基于动力学仿真和迁移学习的滚动轴承故障诊断方法流程图

### 3.4 本章小结

为了解决实际可获取样本量很少时的故障诊断问题，本章提出了一种基于动力学仿真和迁移学习的滚动轴承故障诊断框架，首先，从动力学特性，时域信号的概率分布以及振动响应的频率分布三个角度分析了仿真数据与实际数据的共性与差异性，在此基础上，提出了一种基于基于动力学仿真和迁移学习的滚动轴承故障诊断方法，明确了诊断问题的设定，并给出了方法的详细流程。该方法利用卷积神经网络并结合参数迁移策略，其基本思想是在仿真数据上预训练 CNN 诊断模型，将预训练模型的参数迁移到实际诊断网络中，在实际所能获取的数据量很少时，提高实际问题的故障诊断能力。该方法直接从时域信号出发，利用 CNN 自动化地提取特征，并利用参数迁移策略，将仿真数据中的诊断知识迁移到实际数据中，以解决小样本条件下滚动轴承的故障诊断问题。

## 第 4 章 基于动力学仿真和迁移学习的滚动轴承故障 诊断方法应用与诊断机理分析

### 4.1 引言

在第 3 章给出了小样本条件下，将仿真模型的诊断知识迁移到实际问题诊断中的诊断方法基本问题设定和流程，本章应用该方法对西储大学轴承实验中的两个实际轴承进行小样本条件下的故障诊断。首先对故障诊断场景中的待诊断对象及数据集和所构造的源域对象及数据集进行描述，随后从是否将仿真数据作为源域进行知识迁移及是否利用迁移策略等角度设计多个对比诊断任务，并给出评估诊断模型性能的指标，在此基础上，进行诊断实验并给出诊断结果，验证本文所提方法在小样本故障诊断问题上的优越性。另外，深度神经网络是一个黑箱模型，对基于深度神经网络的智能诊断方法的诊断机理目前没有较好的解释。本章从多个角度对方法取得优越的识别性能的原因进行分析，以解释本文所提出的诊断方法诊断机理。首先，通过混淆矩阵分析方法对不同健康状态的识别性能差异，说明本文方法是在何种健康状态的识别中起作用，进而提高识别性能的；其次，通过分析卷积核形状及其频谱，说明迁移卷积核在方法中起到的作用；随后，通过分析全连接层输出的特征分布，说明迁移全连接层在方法中起到的作用；最后，结合上述分析结果及滚动轴承振动信号特性，总结了所提出诊断方法的诊断机理。

### 4.2 源域和目标域对象描述及数据集构建

本章利用第 3 章的方法对西储大学轴承实验中的两个轴承进行小样本条件下的故障诊断，因此，本节首先对两个待诊断轴承对象（即目标域对象）及基于其振动信号构建的目标域数据集进行描述，包括其工况、健康状态及样本构成等方面。随后，按照第 3 章的诊断方法流程，对基于待诊断对象构建的仿真源域对象进行描述，包括仿真参数的设定及仿真源域的数据集组成等方面。另外，为了验证本文方法中将与待诊断轴承同型号同转速的仿真模型作为源域对象的优越性，另取与本文待诊断对象设备不同、型号不同、工况不同的 MFPT 轴承实验中的轴承对象作为源域对象，同样对该对象及基于该对象构建的源域数据集进行描述，包括其工况、健康状态及样本构成等方面。

#### 4.2.1 目标域对象描述及数据集构建

本章拟将西储大学轴承实验中，驱动端型号为 SKF-6205 和风扇端型号为 SKF-6203 的两个轴承作为待诊断轴承。西储大学滚动轴承故障模拟实验的详细介绍参见第 2.5.1 节。待诊断轴承的振动加速度信号采样频率为 12kHz，工况为 0hp/1797rpm。驱动端及风扇端轴承的几何参数、轴承型号和故障特征频率如表 2-1 所示。待识别的健康状态包括正常状态（N），内圈故障状态（IR）和外圈故障状态（OR）。每种故障状态包含三种故障严重程度，尺度分别为 0.18mm、0.36mm 和 0.53mm。

对原始时域信号有重叠地截取长度为 1200 的样本，构成实际待诊断场景的数据集。每个健康状态包含 600 个样本（600 个正常样本和  $200 \times 3$  个故障样本）。该数据集定义为实际待诊断数据集（**R**），共包含 1800 个样本。

#### 4.2.2 源域对象描述及数据集构建

源域对象为待诊断对象诊断知识的来源，本文方法中将与待诊断设备同型号同工况的轴承仿真模型作为源域对象，在 4.2.2.1 节对仿真源域对象及仿真源域数据集进行详细的描述。另外，为了说明按本文方式生成的仿真数据作为源域数据的优越性，在 4.2.2.2 节基于 MFPT（Society For Machinery Failure Prevention Technology）实验轴承数据构造了一个与待诊断轴承设备不同、型号不同、工况不同的源域数据集，并对该数据集进行描述。

##### 4.2.2.1 仿真模型参数设定及数据集构建

为了实现知识迁移，本文所提出的基于知识迁移的滚动轴承故障诊断方法将从基于待诊断设备构建的仿真模型中获得的振动数据作为源域数据。为了满足 3.3.2 节中的本文方法所要求的诊断问题设定，仿真模型参数的设定具有如下限制：

（1）在仿真模型中，滚动轴承的几何参数和工况条件（转速）与实际待诊断轴承相同，均为 1797rpm。

（2）与实际待诊断轴承相同，仿真模型中也设定三种健康状态：正常状态、内圈故障和外圈故障。由于在实际诊断时，仅通过振动信号无法获得局部故障尺寸，因此，根据 3.3.3 节中的故障程度设定原则，对于每种故障状态，在仿真模型中设置与实际待诊断设备不同的 5 种故障程度：0.2mm、0.4mm、0.6mm、0.8mm 及 1mm。

(3) 从仿真模型采集的信号为外圈竖直方向的振动加速度信号，与实验中采集信号的位置相近，采样频率为 12kHz，与实际数据采样频率相同。

对待诊断设备的相关动力学参数很难进行精确测量，尤其是阻尼和刚度，且第 2 章已经说明本文的仿真也不必精确地模拟一个待诊断设备。因此，与动力学相关的参数（包括  $m_{in}$ 、 $k_{in}$ 、 $c_{in}$ 、 $m_{out}$ 、 $k_{out}$  及  $c_{out}$ ）仍按照经验设定，只要在数量级上不偏离实际导致仿真模型无法表征故障信号的振动特性即可。具体地，SKF-6205 和 SKF-6203 轴承的仿真参数如表 4-1 所示。

表 4-1 待诊断轴承的仿真模型参数设定

仿真参数名称及符号（单位）	SKF-6205 仿真参数数值	SKF-6203 仿真参数数值
转频 $f_r$ (Hz)	1797/60	
内圈-轴质量 $m_{in}$ (kg)	60	5
轴-连接刚度 $k_{in}$ (N/m)	$7.42 \times 10^7$	
轴-连接阻尼 $c_{in}$ (Ns/m)	1376.8	
外圈-轴承座质量 $m_{out}$ (kg)	5	50
轴承座-基础连接刚度 $k_{out}$ (N/m)	$1.51 \times 10^7$	
轴承座-基础连接阻尼 $c_{out}$ (Ns/m)	2210.7	
接触刚度 $K$ (N/m <sup>1.5</sup> )	$9.495 \times 10^9$	$8.753 \times 10^9$
重力加速度 $g$ (m/s <sup>2</sup> )	9.8	
故障宽度 $L$ (mm)	0.2/0.4/0.6/0.8/1	
故障深度 $h$ (mm)	1	
径向游隙 $c_r$ (m)	$2 \times 10^{-6}$	
偏心距 $e$ (m)	$50 \times 10^{-6}$	
滚动体数量 $N$	9	8
滚动体直径 $D_b$ (mm)	7.938	6.746
轴承节径 (mm)	28.500	38.500

与实际待诊断数据集相同，对原始仿真时域信号有重叠地截取长度为 1200 的样本。仿真数据集中每个健康状态包含 600 个样本（正常样本 600 个，每种故

障样本  $120 \times 5$  个), 该数据集定义为仿真数据集 (S), 共包含 1800 个样本。

#### 4.2.2.2 MFPT 实验轴承及数据集构建

若不用仿真模型产生数据, 在实际工程中, 我们更容易获得的数据是来自不同设备不同工况下的实际轴承数据, 这些数据理论上也可作为源域数据。为了说明基于同型号同工况条件下获得的轴承仿真数据相比于这些不同设备不同工况下的实际数据, 更加适合作为源域, 本文引入 MFPT 实验轴承数据作为对比。

MFPT 轴承故障实验的实验滚动轴承故障件如图 4-1 所示。几何参数如表 4-2 所示。实验中轴的转频为 25Hz, 可知 MFPT 实验轴承与待诊断的两个轴承所在设备不同、型号与工况也不同。

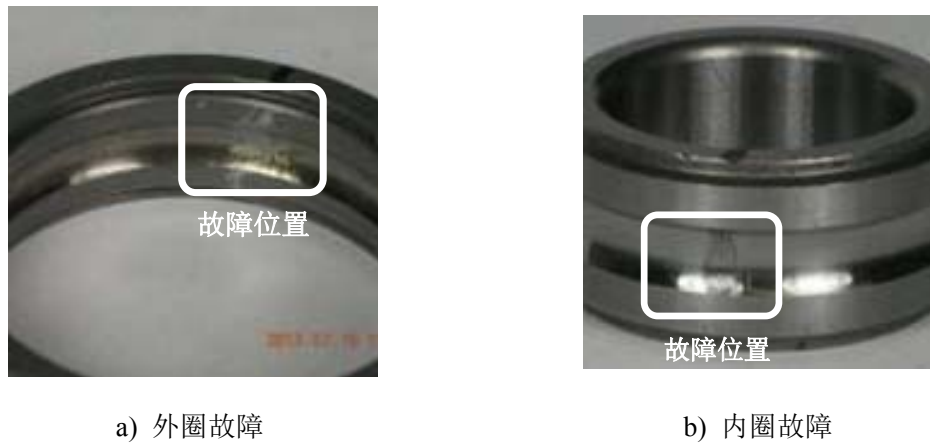


图 4-1 MFPT 故障模拟实验轴承对象

表 4-2 MFPT 实验轴承几何参数

接触角 ( $^{\circ}$ )	滚动体个数	滚动体直径 (mm)	节径 (mm)
0	8	5.97	31.62

选取原始数据集的部分数据构成本文的数据集, 该数据集中仍包含正常状态、内圈故障和外圈故障三种健康状态的数据, 其中, 正常状态包含三个载荷为 270lbs 的数据, 内圈故障和外圈故障数据包含载荷为 200lbs、250lbs 和 300lbs 的故障数据。正常数据的采样频率为 97656Hz, 故障数据采样频率为 48828Hz, 为了满足源域数据与目标域数据采样频率相同的条件, 对所有数据进行采样频率为 12kHz 的重采样操作。

对所选取的数据有重叠地截取长度为 1200 的样本, MFPT 源域数据数据集中每种健康状态包含 600 个样本 (正常、内圈故障与外圈故障样本均为  $3 \times 200$  个), 该数据集定义为 MFPT 源域数据集 (R1), 共包含 1800 个样本, 每个样本的数据点数为 1200。

所构造的目标域对象、源域对象的数据集详细信息如表 4-3 所示。

表 4-3 源域和目标域数据集详细信息

域	对象	数据集简称	转速	标签	健康状态	样本数量
目标域	SKF-6205 SKF-6203	R	1797rpm	1	正常状态 (N)	600
				2	内圈故障 (IR)	200×3=600
				3	外圈故障 (OR)	200×3=600
源域	仿真模型	S	1797rpm	1	正常状态 (N)	600
				2	内圈故障 (IR)	120×5=600
				3	外圈故障 (OR)	120×5=600
	MFPT 实验轴承	R1	1500rpm	1	正常状态 (N)	200×3=600
				2	内圈故障 (IR)	200×3=600
				3	外圈故障 (OR)	200×3=600

### 4.3 故障诊断测试验证场景及主要参数设定

本节对故障诊断测试验证场景及 CNN 主要参数进行设定, 包括为了验证本文方法有效性和优越性而设计的一系列对比任务, CNN 网络的参数及训练过程描述, 以及评价所构建的诊断模型所利用的诊断模型评估指标。

#### 4.3.1 对比任务设定

本文所提出的基于动力学仿真和迁移学习的故障诊断方法的主要特点是将仿真数据作为源域, 并利用网络参数迁移策略实现仿真模型到实际对象的诊断知识的迁移, 以解决实际诊断问题中的小样本难题。为了验证本文方法在解决小样本问题上的有效性, 说明本文方法将仿真数据作为源域以及利用参数迁移策略的优越性, 本文设置了如下 5 类 6 种不同的诊断任务:

##### (1) S→S

在该诊断任务中, 随机从仿真数据集中选取 95% 的轴承仿真样本构成训练集训练 CNN, 剩余 5% 的样本作为测试集用于测试。该任务的设定是保证预训练的 CNN 有足够准确及稳定的识别率, 这是保证预训练网络已从大量的仿真数据中学习到的有效的故障诊断知识的重要条件, 这样的网络才有可能用于对实际问题的辅助诊断中。

##### (2) R→R

在该诊断任务中, 随机从实际待诊断数据集中选取  $a\%$  的轴承实际样本训练 CNN, 剩余  $(1-a\%)$  的样本用于测试网络性能。该任务用于模拟只有少量的实际



样本可以获取时的实际诊断场景。CNN 网络的参数（权重和偏置）随机初始化，在该任务中没有预训练过程。

### （3） $S \rightarrow R$

在该诊断任务中，随机从仿真数据集中选取 95% 的轴承仿真样本构成训练集训练 CNN，剩余 5% 的样本作为测试集用于测试预训练的网络。随后，将预训练的网络直接用于实际任务的诊断中，在该诊断任务中，没有用实际数据微调预训练网络的过程。

### （4） $S \rightarrow R_{ft1}$ 和 $S \rightarrow R_{ft}$

在这两个诊断任务中，预训练过程与任务  $S \rightarrow R$  操作相同，在预训练后，随机选取  $\alpha\%$  的实际样本微调预训练的 CNN 网络，剩余实际样本测试网络性能。在任务  $S \rightarrow R_{ft1}$  中，仅迁移预训练 CNN 的网络卷积核的权重和偏置，在  $S \rightarrow R_{ft}$  中，迁移所有层的网络参数，对应本文所提方法。

### （5） $R1 \rightarrow R_{ft}$

在该诊断任务中，不用仿真数据作为源域，而采用 MFPT 源域数据集 R1 作为源域，进行预训练过程（ $R1 \rightarrow R1$ ）。其余操作与  $S \rightarrow R_{ft}$  相同。MFPT 轴承与待诊断轴承设备、型号及工况不同，在实际诊断任务中，通常更容易获取这样的数据。该任务模拟用这样的数据学习诊断知识，验证其是否能够帮助实现待诊断设备的故障识别，进一步说明将仿真数据作为源域的优越性。

在上述设定的诊断任务中， $R \rightarrow R$ 、 $S \rightarrow R$ 、 $S \rightarrow R_{ft1}$ 、 $S \rightarrow R_{ft}$  和  $R1 \rightarrow R_{ft}$  为诊断实际问题的不同诊断任务，而  $S \rightarrow S$ （ $R1 \rightarrow R1$ ）可以看成是一个保证预训练网络是具备迁移能力的基线任务。若  $S \rightarrow S$ （ $R1 \rightarrow R1$ ）不具备好的识别性能，则意味着它没有学习到可能有效的、用于迁移的诊断知识，在这样的条件下进行迁移是没有意义的。

## 4.3.2 CNN 参数设定

对于 4.3.1 中设定的所有诊断任务，采用相同的 CNN 网络结构。本文的故障诊断问题是一个较为简单的三分类问题，为了更好地解释在故障诊断问题中，CNN 网络学习故障特征的机制，CNN 网络的结构应尽可能简单，本文采用的 CNN 网络共 7 层，其中有 1 层卷积层，该卷积层具有 2 个卷积核。CNN 网络的具体结构如表 4-4 所示。

表 4-4 CNN 网络结构

层数	类型	名称	尺寸	步长	通道数	输出大小
1	输入层	input	/	/	1	1200
2	卷积层	CONV1	$256 \times 1$	$1 \times 1$	2	$945 \times 2$
3	RELU 层	RELU1	/	/	2	$945 \times 2$
4	池化层	POOL1	$16 \times 1$	$1 \times 1$	2	$930 \times 2$
5	全连接层	FC1	1860	/	1	3
6	Softmax 层	Softmax	3	/	1	3
7	分类层	CL	3	/	1	3

### 4.3.3 CNN 训练过程

在不同的诊断任务中，CNN 的训练细节如下：对于  $S \rightarrow S$  ( $R1 \rightarrow R1$ ) 训练过程及  $S \rightarrow R$ 、 $S \rightarrow R\text{-ft1}$ 、 $S \rightarrow R\text{-ft}$  和  $R1 \rightarrow R\text{-ft}$  的预训练过程，CNN 的所有权重和偏置随机初始化。对于  $S \rightarrow S$  ( $R1 \rightarrow R1$ ) 训练过程及  $S \rightarrow R$ 、 $S \rightarrow R\text{-ft1}$ 、 $S \rightarrow R\text{-ft}$  和  $R1 \rightarrow R\text{-ft}$  的预训练过程，训练时的最大迭代次数 (maximum training epoch) 为 30，每执行一次反向传播算法所利用的小样本批量 (mini-batch) 为 64。对于  $R \rightarrow R$  的训练过程及  $S \rightarrow R\text{-ft1}$ 、 $S \rightarrow R\text{-ft}$  和  $R1 \rightarrow R\text{-ft}$  的微调过程，训练时的最大迭代次数为 1000，mini-batch 的大小与  $a$  的值有关，在本文为模拟小样本诊断场景，设置  $a$  的值为 1-7 的整数。不同  $a$  值下上述任务中的 mini-batch 的大小如表 4-5 所示。所有任务的所有训练过程学习率为 0.001。

表 4-5 不同  $a$  值下的 mini-batch

$a$	1	2	3	4	5	6	7
mini-batch	18	36	54	36	45	54	63

为了在一定程度上消除随机因素的影响，更科学地度量不同诊断任务条件下的诊断模型性能，每个  $a$  值下每个任务执行 20 次重复实验，每次实验在保持类别平衡的条件下重新随机划分训练样本和测试样本。

### 4.3.4 诊断模型性能评估指标

本文用网络在测试数据上的识别率来评估不同任务下诊断模型的性能。识别率为测试集中所有预测标签与真实标签相同的样本数与测试时样本总数的比，若用  $x$  表示测试集  $D_{test}$  中的样本， $y_i$  表示测试样本的实际标签， $\hat{y}_i$  表示测试样本基

于诊断模型识别的预测标签，则识别率  $Acc$  表达式如下<sup>[51,68]</sup>：

$$Acc = \frac{|x: x \in D_{test} \cap \tilde{y}_i = y_i|}{|x: x \in D_{test}|} \quad (4-1)$$

为了度量并对比不同源域对实际诊断问题的性能提升程度，定义  $ACC.I$  和  $STD.I$  两个指标如下：

$$ACC.I = Acc_{D_s \rightarrow R\_ft} - Acc_{R \rightarrow R} \quad (4-2)$$

$$STD.I = std(Acc_{R \rightarrow R}) - std(Acc_{D_s \rightarrow R\_ft}) \quad (4-3)$$

式中  $std(\cdot)$ —— $(\cdot)$  的标准差。

$D_s$ ——不同的源域数据， $D_s \in \{S, R1\}$ ；

指标  $ACC.I$  反映了两个诊断任务识别准确性的差异，而指标  $STD.I$  反映了两个诊断任务在多次重复实验下识别性能稳定性的差异。两个指标值越大，说明提升效果越明显。

## 4.4 故障诊断结果及分析

基于 4.2 节中构造的源域和目标域数据集，按照 4.3 节中给出的对比任务设定及 CNN 网络参数和训练细节进行实验，得到两个待诊断轴承 SKF-6205 和 SKF-6203 在不同诊断任务下的故障诊断结果。本节给出两个轴承的故障诊断结果并对结果进行分析。

### 4.4.1 SKF-6205 轴承诊断结果

不同诊断任务下，SKF-6205 轴承的识别结果如表 4-6 和图 4-2 所示。在表 4-6 中对比了 6 种不同诊断任务在不同  $a$  值下的诊断识别率（需要说明的是，对于任务  $S \rightarrow S$  ( $R1 \rightarrow R1$ )，其识别结果与  $a$  值无关）。从表中可以看出，由于大量的仿真数据用于训练 CNN，任务  $S \rightarrow S$  的识别率可以达到 100%，这保证了将从这样的网络中获得的诊断知识迁移到实际问题中是可能有效的。在图 4-2 中，给出了不同诊断实际问题的任务 ( $S \rightarrow R$ 、 $R \rightarrow R$ 、 $S \rightarrow R\_ft1$ 、 $R1 \rightarrow R\_ft$  及  $S \rightarrow R\_ft$ ) 中，诊断性能的直观对比，由表 4-6 和图 4-2 可以看出，在绝大多数  $a$  值下，本文方法获得了最高的识别率。

表 4-6 SKF-6205 轴承识别结果

$a\%$	1%	2%	3%	4%	5%	6%	7%
S→S	100±0 (与 $a$ 无关)						
R→R	75.1±9.8	84.7±8.6	91.3±5.0	95.0±2.6	95.5±2.9	97.8±1.8	98.7±1.2
S→R	44.4±19.9	41.6±20.8	50.0±18.7	42.4±21.1	40.1±21.8	44.3±20.7	42.4±19.8
S→R_ft1	72.5±12.3	86.1±9.8	<b>95.2±3.4</b>	94.8±4.8	97.5±2.9	97.0±3.1	<b>99.2±0.9</b>
S→R_ft	<b>89.4±7.1</b>	<b>91.7±6.6</b>	<b>95.2±4.0</b>	<b>95.8±3.6</b>	<b>98.0±2.0</b>	<b>98.3±1.5</b>	98.0±1.7
R1→R1	100±0 (与 $a$ 无关)						
R1→R_ft	81.4±6.6	85.3±6.3	90.1±3.5	91.8±3.1	93.1±2.9	94.5±2.5	95.5±2.2

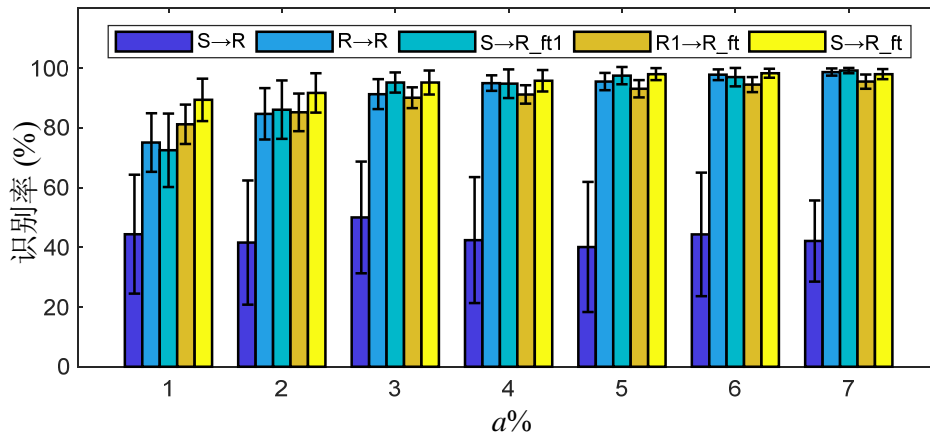


图 4-2 不同任务下实际诊断任务的识别结果 (SKF-6205)

图 4-3a)和 b)分别给出了任务 S→R\_ft 相比于 R→R 及任务 R1→R\_ft 相比于 R→R 的识别性能提升情况。由图 4-3a)可知,当  $a$  值很小时,本文方法能更加显著地提升识别性能。如当  $a=1$  时,识别率提升了 14.3%,20 次重复实验的识别率标准差下降了 2.7%。当  $a$  增加时,任务 S→R\_ft 的识别性能接近但仍然好于任务 R→R。这些结果说明基于仿真数据进行预训练得到的 CNN 网络学习到了有利于实际问题诊断的特征,这些特征帮助实际小样本诊断问题提高的诊断性能。而由图 4-3b)可知,任务 R1→R\_ft 仅在  $a=1,2$  时对识别率有一定提升, $a=1,2,3$  时对识别的稳定性有一定提升,其他条件下不能提升小样本条件下的识别性能。说明在该实例中,相比于不同设备不同型号不同工况下的实际轴承数据,同型号同工况下的仿真数据更适合作为源域帮助实际小样本诊断问题提高诊断性能。

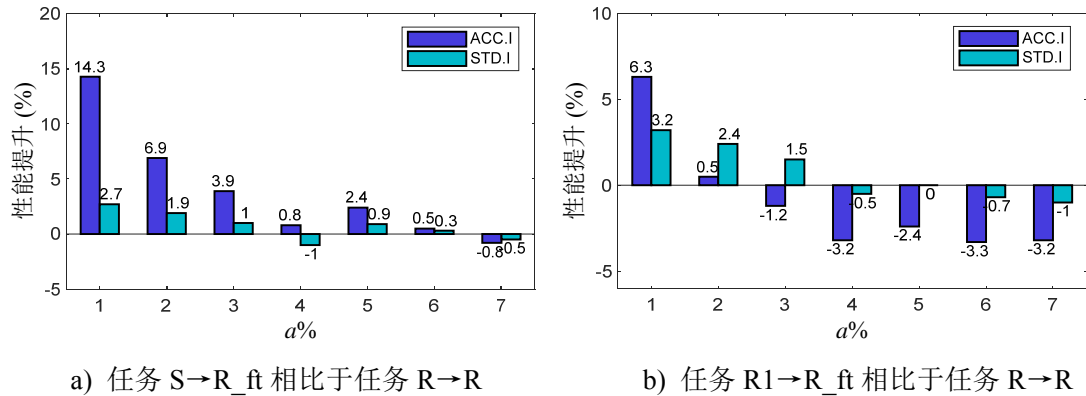


图 4-3 诊断性能提升结果 (SKF-6205)

#### 4.4.2 SKF-6203 轴承诊断结果

不同诊断任务下, SKF-6203 轴承的诊断结果如表 4-7 和图 4-4 所示。可以看出, 利用仿真数据预训练并利用实际数据微调在绝大多数  $a$  值中能够获得最优的识别性能, 该结果与 SKF-6205 轴承结果相似, 这说明通过仿真数据的预训练, CNN 网络学习到了一些能够帮助实际问题诊断的有效特征。在该诊断实例中, 任务  $R \rightarrow R$  中当  $a=3,4,5,6,7$  时, 出现了识别率上限超过 100% 的现象, 本文认为此现象是合理的, 因为训练样本过少以至于识别性能非常不稳定。一旦加入了预训练过程, 即在任务  $S \rightarrow R_{ft}$  及任务  $R1 \rightarrow R_{ft}$  中, 此现象有了较大程度的缓解。

表 4-7 SKF-6203 轴承识别结果

$a\%$	1%	2%	3%	4%	5%	6%	7%
$S \rightarrow S$	100 $\pm$ 0 (与 $a$ 无关)						
$R \rightarrow R$	72.0 $\pm$ 7.5	89.5 $\pm$ 8.3	95.2 $\pm$ 5.8	96.8 $\pm$ 5.7	98.6 $\pm$ 2.1	98.2 $\pm$ 3.5	<b>99.6 <math>\pm</math> 0.6</b>
$S \rightarrow R$	31.1 $\pm$ 6.4	23.7 $\pm$ 13.4	27.8 $\pm$ 10.5	24.3 $\pm$ 10.7	20.6 $\pm$ 13.4	27.5 $\pm$ 9.0	27.0 $\pm$ 11.8
$S \rightarrow R_{ft1}$	69.3 $\pm$ 8.6	82.6 $\pm$ 8.4	94.2 $\pm$ 4.1	95.3 $\pm$ 6.8	96.5 $\pm$ 5.6	98.5 $\pm$ 1.5	99.4 $\pm$ 0.7
$S \rightarrow R_{ft}$	<b>91.3 <math>\pm</math> 5.8</b>	<b>95.3 <math>\pm</math> 4.5</b>	<b>97.7 <math>\pm</math> 2.5</b>	<b>98.6 <math>\pm</math> 1.7</b>	<b>99.4 <math>\pm</math> 0.5</b>	<b>99.6 <math>\pm</math> 0.5</b>	99.5 $\pm$ 0.7
$R1 \rightarrow R1$	100 $\pm$ 0 (与 $a$ 无关)						
$R1 \rightarrow R_{ft}$	80.4 $\pm$ 6.4	87.5 $\pm$ 7.5	90.4 $\pm$ 4.8	95.8 $\pm$ 2.0	94.4 $\pm$ 2.3	96.7 $\pm$ 1.7	97.5 $\pm$ 2.0

与 SKF-6205 轴承的诊断实例类似, SKF-6203 轴承的诊断任务  $S \rightarrow R_{ft}$  相比于  $R \rightarrow R$  及任务  $R1 \rightarrow R_{ft}$  相比于  $R \rightarrow R$  的识别性能提升情况如图 4-5a)和图 4-5b)所示。结合上一个实例可以看出, 当  $a$  值越小时, 任务  $S \rightarrow R_{ft}$  相比于  $R \rightarrow R$  性能提升的越明显。在 SKF-6205 轴承的诊断中, 当  $a=1,2,3$  时, 识别率的提升超过了 3%, 识别率的标准差减少超过 1%。当  $a=4,5,6$  时, 性能提升不如前者明显。在 SKF-6203 轴承的诊断中,  $a=1,2,3,4$  时, 识别率的提升超过 2%, 或识别率的

标准差减少超过 3%，当  $a$  值增大时，性能提升并不明显。在两个诊断实例中，当  $a$  值增大到 7% 时， $R \rightarrow R$  任务的识别率已经达到较高水平，这时本文方法不再显示出优越性，两种诊断方式性能十分接近。

由图 4-5b) 可知，在对 SKF-6203 的诊断实例中，将 MFPT 数据作为源域，在  $a=1,2,3,4,6$  时，诊断模型的稳定性有了一定提升，说明加入源域数据能在一定程度上缓解小样本条件下识别稳定性差的缺陷。但用 MFPT 数据作为源域的任务  $R1 \rightarrow R\_ft$  在两个诊断实例中的识别率均不如本文方法。任务  $R1 \rightarrow R\_ft$  在 SKF-6205 诊断实例中仅在  $a=1,2$  时识别率高于任务  $R \rightarrow R$ ，在 SKF-6203 的诊断实例中仅在  $a=1$  时识别率高于任务  $R \rightarrow R$ ，这说明 MFPT 数据与待诊断设备的数据之间的差异性过大，出现了负迁移现象。这些现象说明在小样本诊断问题中，同型号同转速的仿真数据与实际数据更相近，相比于不同设备不同型号不同转速下的实际数据，更适合作为待诊断实际设备的有效诊断知识的来源，帮助其实现更有效的诊断。

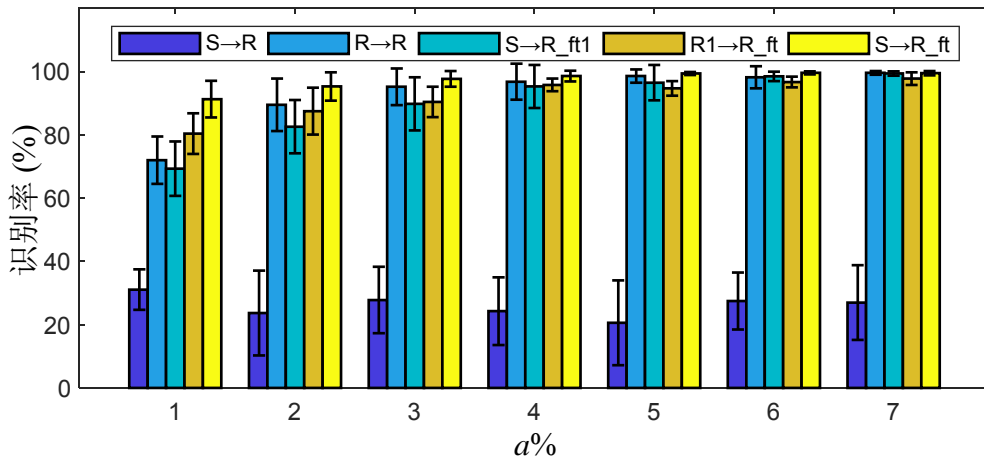
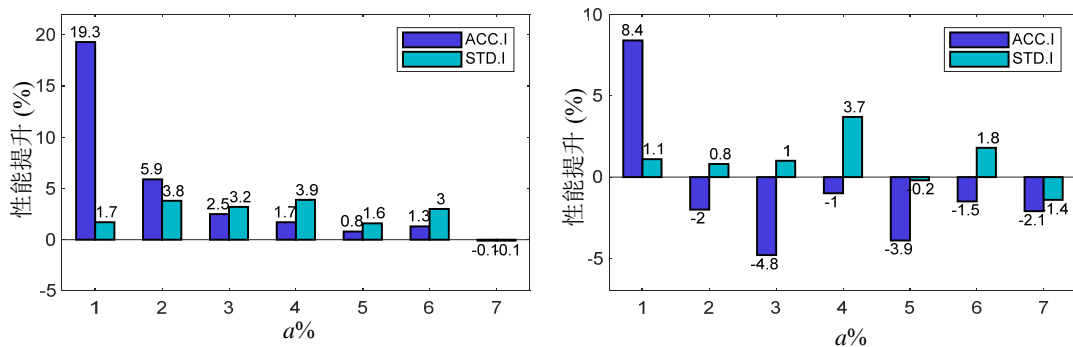


图 4-4 不同任务下实际诊断任务的识别结果 (SKF-6203)



a) 任务  $S \rightarrow R\_ft$  相比于任务  $R \rightarrow R$

b) 任务  $R1 \rightarrow R\_ft$  相比于任务  $R \rightarrow R$

图 4-5 诊断性能提升结果 (SKF-6203)

上述结果验证了所提方法在滚动轴承故障诊断的有效性，通过仿真数据预训练 CNN 并利用少量实际数据微调预训练的 CNN 网络，能够有效地解决小样本条件下滚动轴承的故障诊断问题。

## 4.5 基于动力学仿真和深度迁移学习的故障诊断方法机理分析

在本节，基于上述诊断实例，分析了本文所提故障诊断方法的诊断机理。首先，在 4.5.1 节分析了混淆矩阵，以揭示本文方法是在哪个部分的识别中对整个识别性能的提升起到重要作用的；在 4.5.2 和 4.5.3 节，分别分析了仿真数据和微调迁移策略在方法中的重要作用。最后，基于上述分析，并结合仿真数据和实际数据的特点，在 4.5.4 节给出了本文方法的诊断机理。

### 4.5.1 混淆矩阵分析

任务  $S \rightarrow R\_ft$  的总体识别性能高于其他诊断任务，说明其一定在某个或某些类别的分类上的识别效果好于其他任务。在模式识别任务中，混淆矩阵  $C = [C_{ij}]_{c \times c}$  能够表征分类器对每一类样本的识别情况，矩阵的行表示样本所属的真实类别，矩阵的列表示诊断模型预测的样本类别，矩阵中的每一个元素  $C_{ij}$  表示分类器将第  $j$  类样本分到第  $i$  类样本数占第  $j$  类样本总数的百分比。图 4-5 和图 4-6 给出了两个诊断实例中任务  $R \rightarrow R$ 、 $S \rightarrow R$ 、 $S \rightarrow R\_ft1$  和  $S \rightarrow R\_ft$  的混淆矩阵。受篇幅限制，仅列出  $a=2$  和  $a=5$  的情况。混淆矩阵中的值基于各个任务中 20 次重复实验后的平均性能计算。

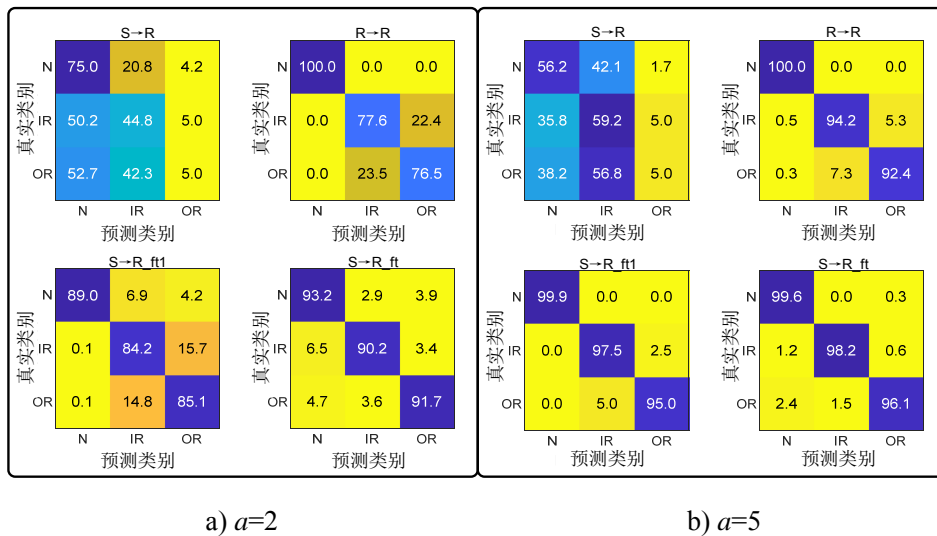


图 4-6 SKF-6205 轴承诊断结果的混淆矩阵

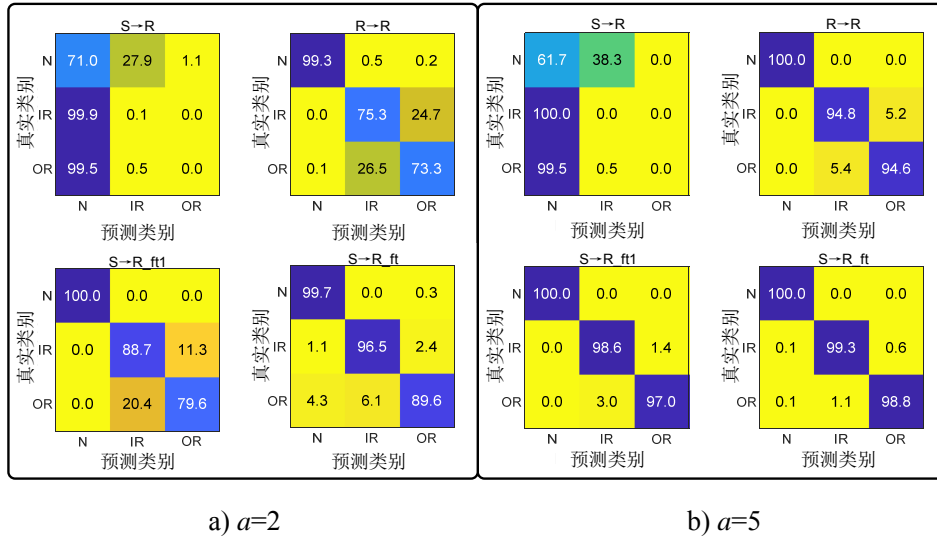


图 4-7 SKF-6203 轴承诊断结果的混淆矩阵

从图 4-6 和图 4-7 中可以看出，在所有任务中，正常状态的识别率最高，说明在故障诊断问题中，区分正常和异常（即故障检测）相比于识别出特定的故障类型（即故障识别）来说是更容易的。

将任务  $S \rightarrow R_{ft1}$  和任务  $S \rightarrow R_{ft}$  同任务  $R \rightarrow R$  的混淆矩阵结果对比，可以看出，由于仿真数据的辅助作用，任务  $S \rightarrow R_{ft1}$  和任务  $S \rightarrow R_{ft}$  的故障类型识别精度高于  $R \rightarrow R$ 。例如在 SKF-6205 轴承的诊断中，当  $a=2$  时，外圈故障（OR）的识别错误率由 23.5%（ $R \rightarrow R$ ）下降到 14.9%（ $S \rightarrow R_{ft1}$ ）和 8.3%（ $S \rightarrow R_{ft}$ ）。

将任务  $S \rightarrow R_{ft1}$  和任务  $S \rightarrow R_{ft}$  同任务  $S \rightarrow R$  的混淆矩阵结果对比，可以看出微调策略增强了所有健康状态的识别能力。在 SKF-6205 轴承的  $S \rightarrow R$  诊断任务中，外圈故障几乎完全不能被识别出来，在 SKF-6203 轴承的  $S \rightarrow R$  诊断任务中，外圈故障和内圈故障都被错分至正常状态。这个现象说明若不采用迁移策略，由于分仿真数据和实际数据的差异性，识别性能较差。在引入微调迁移策略后，每个健康状态的识别率都有了明显的提高。例如，在 SKF-6203 轴承的诊断中，当  $a=5$  时，内圈故障的识别率由 0 提高至 98.6%（ $S \rightarrow R_{ft1}$ ）和 99.3%（ $S \rightarrow R_{ft}$ ）。

上述现象证明了仿真数据和参数迁移策略在本文方法中的有效性。使用本文方法进行实际问题诊断的任务  $S \rightarrow R_{ft}$ ，通过仿真数据的辅助并结合参数迁移策略，相比于任务  $R \rightarrow R$  和任务  $S \rightarrow R_{ft1}$ ，能更准确地识别出不同故障类别，相比于任务  $S \rightarrow R$ ，能更准确地识别出所有的健康状态，这是实际诊断任务性能得到有效提升的直接原因。



## 4.5.2 卷积核迁移机理分析

### 4.5.2.1 CNN 特定层的输出分析

在分析迁移卷积核的作用之前，首先分析 CNN 特定层的输出（卷积层、ReLU 激活层和池化层）以说明 CNN 中的典型操作是如何影响原始信号的。当轴承故障发生时，由于故障冲击造成的幅值调制现象是滚动轴承故障信号的典型特性，该特性无论是在仿真数据中还是在实际数据中都会出现。由于在实际信号中，有很多随机噪声，分析实际信号很难获得较好的可视化结果。因此，本节仅展示仿真信号的分析结果，以说明 CNN 的各种操作如何影响这种具有调制特性的信号的。

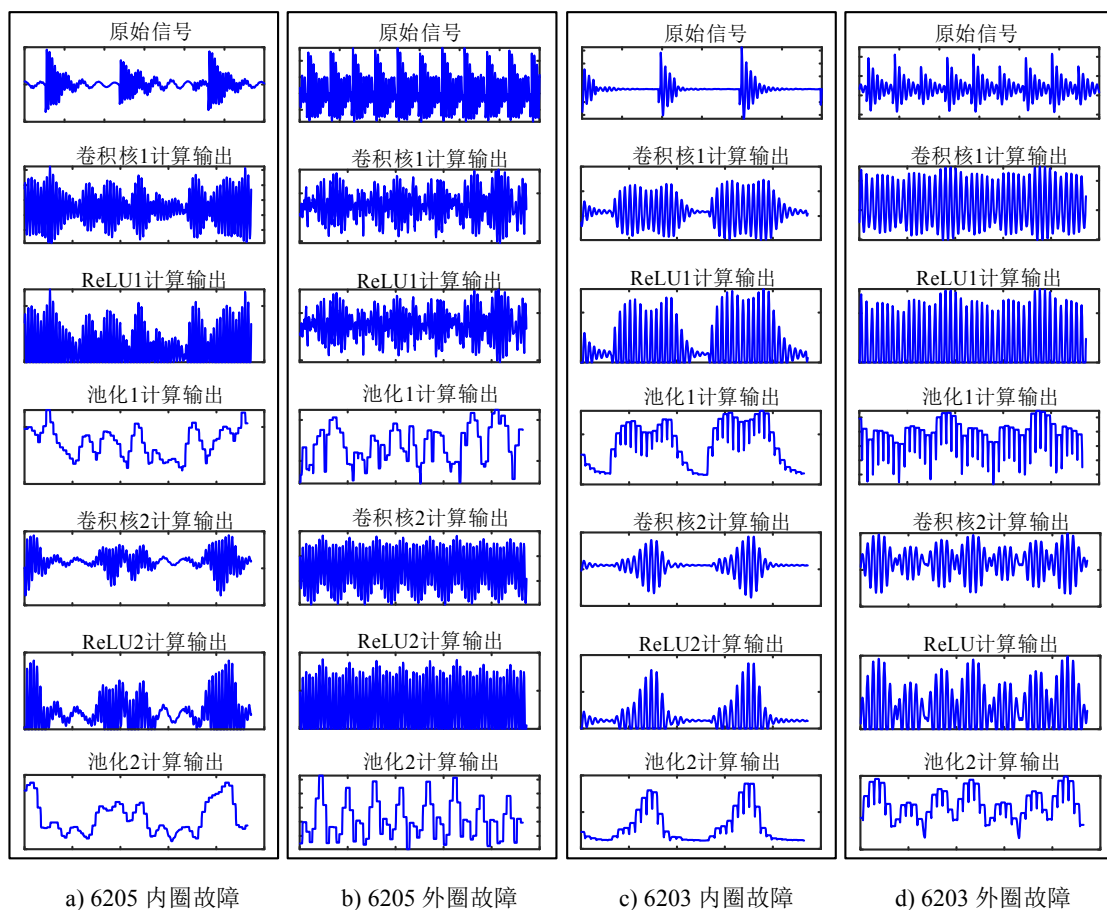


图 4-8 CNN 网络的特定层输出

图 4-8 展示了 SKF-6205 轴承和 SKF-6203 轴承内、外圈故障仿真信号在经过卷积、非线性激活和池化三种操作后的结果。本文中 CNN 网络的卷积层有 2 个卷积核，信号通过 2 个卷积核会产生 2 个输出，对应图中的卷积核 1 计算输出和卷积核 2 计算输出；这两个输出分别通过 ReLU 激活层和池化层会分别产生各自

的输出，对应 ReLU1 计算输出、ReLU2 计算输出、池化 1 计算输出和池化 2 计算输出。图 4-8a)和 c)对应内圈故障仿真信号，图 4-8b)和 d)对应外圈故障仿真信号。CNN 网络从  $a=4$  时 20 次重复实验中随机选取，原始信号从仿真数据集中随机选取。

由图 4-8 可以看出，其中的一个卷积核保持了原始信号的包络形状并有使包络形状变成“轴对称”的趋势。此现象在 6203 轴承的卷积核 2 计算输出中十分明显，在 6205 轴承的卷积核 2 计算输出中，此现象也存在。在经过这个卷积核的操作后，ReLU 操作除去了值小于 0 的信号部分，由于原始信号是关于均值 0 对称的，因此 ReLU 操作简化了信号信息，同时不会影响信号的可分性。最后进行的池化操作，可以看成是一个信号平滑的过程，得到了 ReLU 计算输出的大致的包络形状。池化计算输出将进一步指导 CNN 的其他层进行信号分类。对于两个诊断实例中的另一个卷积核，即 6205 卷积核 2 和 6203 卷积核 1，它的作用看起来是放大了信号中的高频成分，对于 ReLU 和池化操作，很难直观地看出其作用。

#### 4.5.2.2 卷积核形状与频谱分析

通过对图 4-8 的分析，可以看出 CNN 中的卷积核从不同的角度提取了信号中的特征。CNN 中的卷积核可以看成是信号处理中的滤波器，不同之处在于 CNN 中的卷积核是通过反向传播算法学习得到的，而信号处理中的滤波器是人为设定的。可以认为，CNN 学习到的诊断知识一部分蕴含在卷积核中。本节通过分析任务  $R \rightarrow R$  和任务  $S \rightarrow R_{ft}$  中学习到的卷积核，说明迁移卷积核的作用。

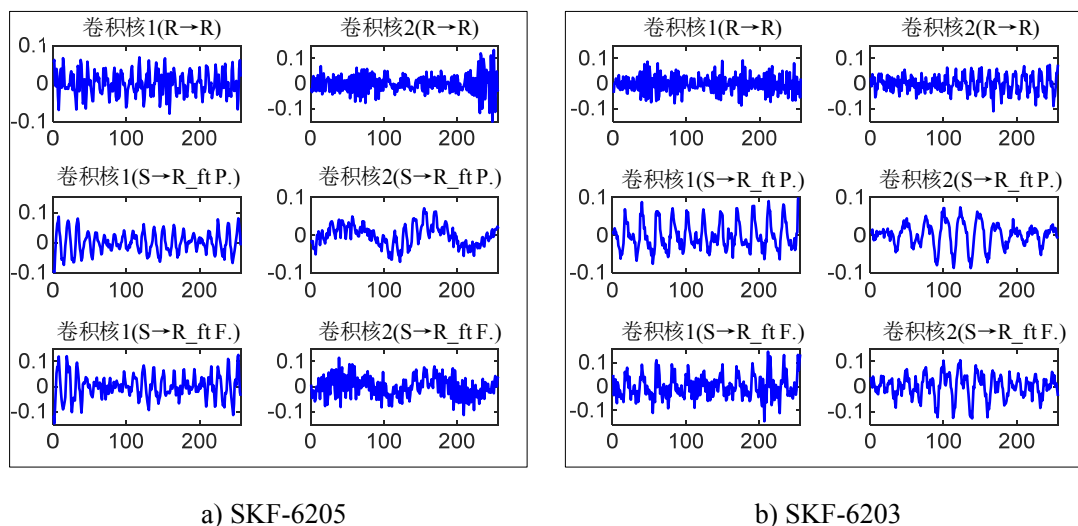


图 4-9 CNN 网络学习到的 2 个卷积核形状（注：图中， $R \rightarrow R$  表示任务  $R \rightarrow R$ ， $S \rightarrow R_{ft} P.$  表示任务  $S \rightarrow R_{ft}$  的预训练过程， $S \rightarrow R_{ft} F.$  表示任务  $S \rightarrow R_{ft}$  的微调过程。）

首先，分析 CNN 学习到的卷积核的形状。两个诊断实例中卷积核的形状如

图 4-9 所示。需要说明的是，在每个诊断实例的不同次重复实验中，卷积核 1 和卷积核 2 的形状可能会交换，但它们的基本形状均如图 4-9 中所示。因此，这里仅展示每个诊断实例中的一种典型情况，所选择的网络与 4.5.2.1 节中所选取的网络相同。从图中可以看出，预训练网络卷积核的基本形状在微调过程中得以保持，在微调后，卷积核中多出了一些其他成分，结合任务  $R \rightarrow R$  中学习到的卷积核形状来看，这些成分反映了待诊断实际数据中的知识。

为了进一步验证上述结论，图 4-10 给出了卷积核的频谱。可以看出在预训练网络中学习到的频率成分（实线箭头标识）在微调得到的卷积核中仍然存在。对比  $R \rightarrow R$  卷积核的频谱和  $S \rightarrow R\_ft F.$  卷积核的频谱，可以看出在任务  $R \rightarrow R$  中学习到的频率成分（虚线箭头标识）出现在了任务  $S \rightarrow R\_ft$  微调过程的卷积核中，这说明微调过程在保持预训练过程学习到的特征的同时，也学习到了实际轴承信号的特征。

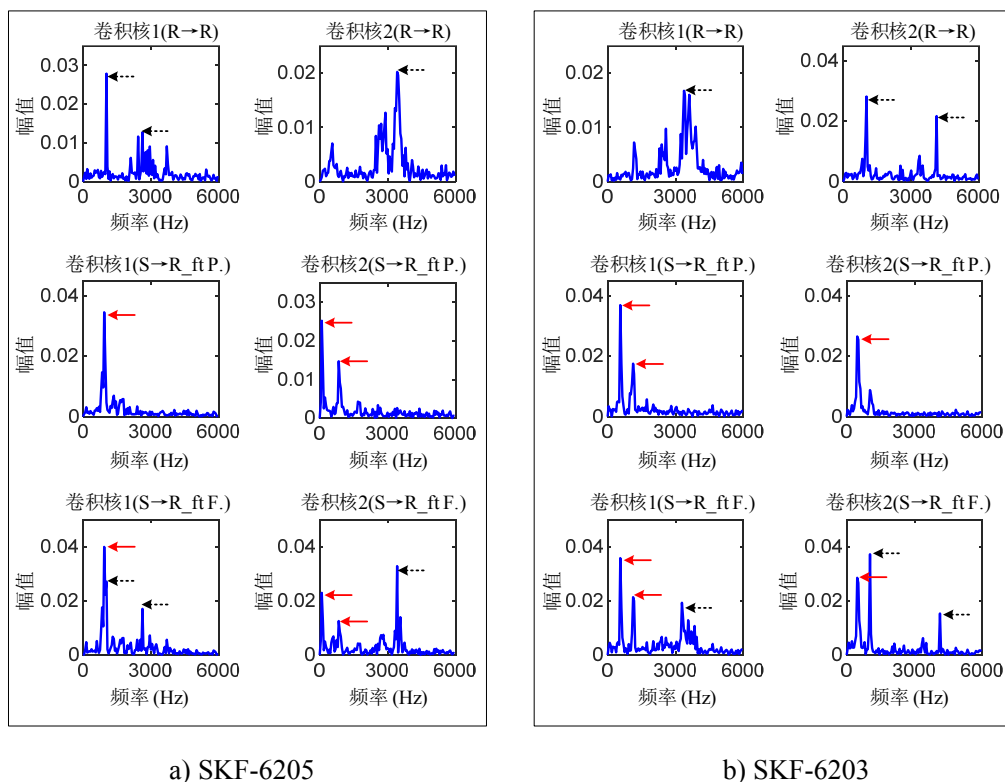


图 4-10 CNN 网络学习到的 2 个卷积核频谱

另外，由图 4-10 中的  $R \rightarrow R$  卷积核和  $S \rightarrow R\_ft P.$  卷积核，可以看出在这个诊断问题中，CNN 的两个卷积核具有如下的分工：一个卷积核主要学习信号中的较高频特征，另一个学习信号中的较低频特征。另外，在任务  $R \rightarrow R$  中，所学习到的特征多为信号的中高频特征，而  $S \rightarrow R\_ft$  学习到信号的低频和中频特征。在图

4-11 中展示了仿真和实际原始信号在三种健康状态下的频谱，原始信号从数据集中随机抽取，可以看出，CNN 卷积核频谱中的主要频率成分与原始信号中的主要频率成分有一致性。

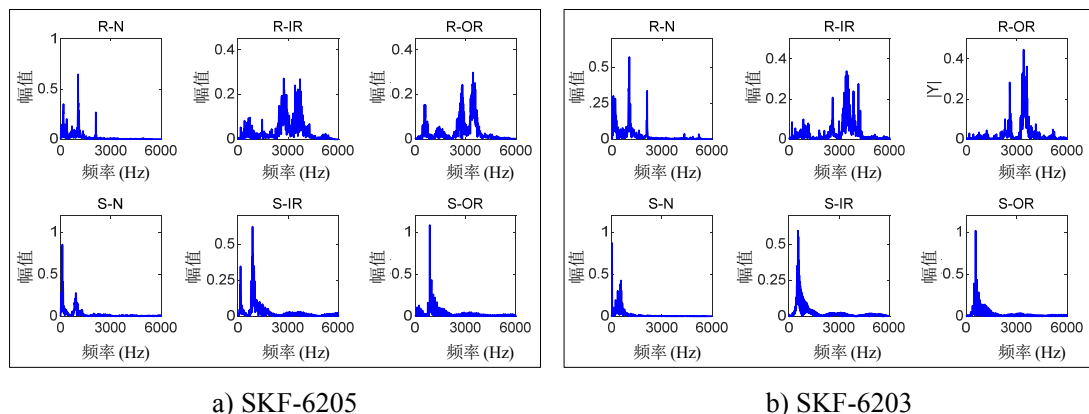


图 4-11 不同健康状态样本的频谱（注：S 表示仿真信号，R 表示实际信号，N 表示正常状态，IR 表示内圈故障，OR 表示外圈故障）

由上述分析可知，卷积核的学习过程可以看成是学习信号频谱的过程，由此可在一定程度上解释将 MFPT 实验轴承作为源域得到的识别结果不如仿真信号作为源域的原因。图 4-12 给出了 MFPT 轴承数据三种健康状态样本的频谱，与 4-11 对比可以发现，与 SKF-6205 和 SKF-6203 轴承类似，频谱中的高频成分仍占主导，但直观来看，与这两个轴承的频率分布有显著的差异性。结合第 3 章的分析可知，这种差异性来自于设备自身属性造成的固有频率特性差异。在 MFPT 数据集上预训练，也会学习到信号的这些的高频特性，再经过微调，与从待诊断设备中学习的高频特征与源域中学习的高频特征相互冲突干扰，使得诊断性能不仅不会提升，反而有所下降。

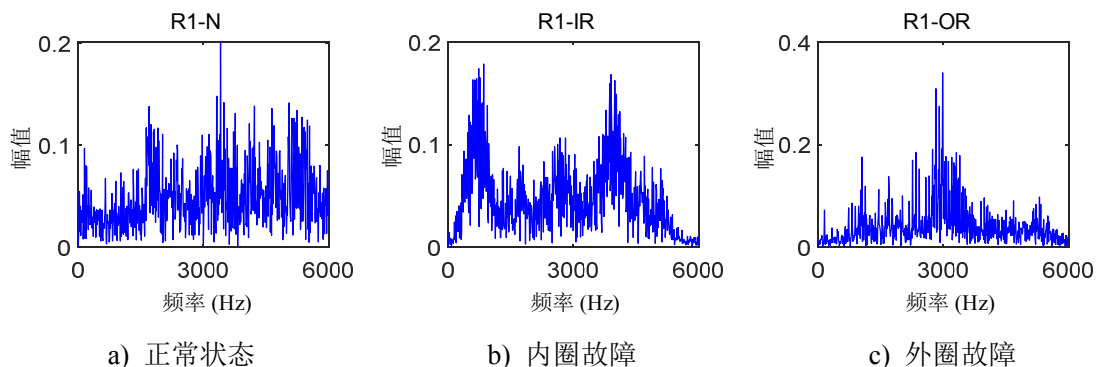


图 4-12 MFPT 轴承不同健康状态样本的频谱

### 4.5.3 全连接层迁移机理分析

#### 4.5.3.1 基于 PCA 的特征分布可视化分析

本节首先引入主元分析（principal component analysis, PCA）方法将 CNN 学习到的高维特征映射到二维空间，以更加直观的理解本文方法的迁移过程。以  $a=3$  为例，两个诊断实例的任务  $S \rightarrow R$ 、任务  $S \rightarrow R_{ft1}$  和任务  $S \rightarrow R_{ft}$  中，CNN 全连接层从仿真样本和实际样本中提取到的特征映射到二维空间后的特征分布如图 4-13 和图 4-14 所示。所考虑的三个诊断任务的差异在于，任务  $S \rightarrow R$  不进行迁移，任务  $S \rightarrow R_{ft1}$  仅迁移卷积层，任务  $S \rightarrow R_{ft}$  迁移卷积层和全连接层。由图 4-13 和图 4-14 可以看出，相比于任务  $S \rightarrow R$  和任务  $S \rightarrow R_{ft1}$ ，在任务  $S \rightarrow R_{ft}$  中，对应相同类别的源域（仿真）样本和目标域（实际）样本在特征空间中更加靠近，这说明迁移全连接层可能能够更好地缩小两个域之间的分布差异。

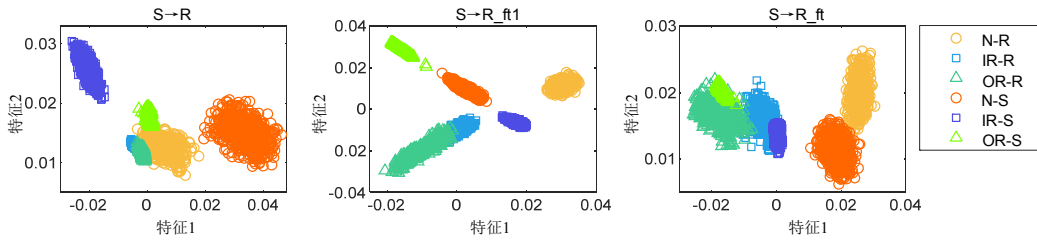


图 4-13 全连接层学习到的特征的二维可视化结果（SKF-6205）

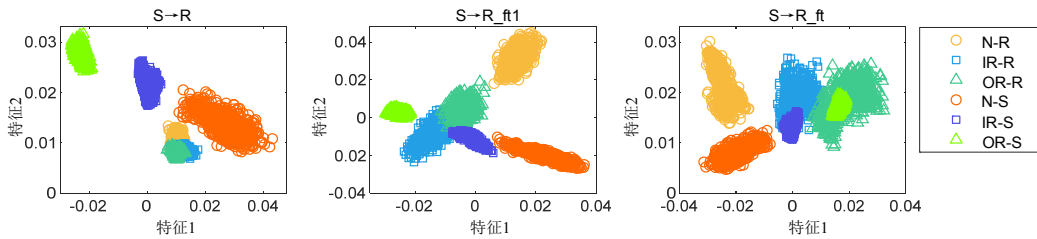


图 4-14 全连接层学习到的特征的二维可视化结果（SKF-6203）

#### 4.5.3.2 基于 MMD 距离的分布差异分析

为了更加全面地分析迁移全连接层的迁移机理，引入最大均值差异（Maximum mean discrepancy, MMD）距离来定量地度量不同域之间的分布差异<sup>[69]</sup>。MMD 距离越小，说明两个域之间的特征分布越靠近，在源域上训练得到的识别模型在考虑迁移策略的基础上，能够更好地帮助完成目标域任务。考虑上述三个任务，由全连接层学习到的源域和目标域的特征分布的 MMD 距离由下式计算<sup>[68]</sup>：

$$MMD^2(X_s, X_t) = \sum_{C=1}^{C^{s,t}} \left\| \frac{1}{n_{s,C}} \sum_{i=1}^{n_{s,C}} \phi(\mathbf{x}_{si}^C) - \frac{1}{n_{t,C}} \sum_{j=1}^{n_{t,C}} \phi(\mathbf{x}_{sj}^C) \right\|_{\mathcal{H}}^2 \quad (4-4)$$

式中  $C^{s,t}$  ——类别数，本文  $C^{s,t} = 3$ ；

$\phi(\cdot)$  ——  $\phi: \mathcal{X} \rightarrow \mathcal{H}$  从空间  $\mathcal{X}$  映射到再生核希尔伯特空间 (reproduce kernel Hilbert space, RKHS)  $\mathcal{H}$  的变换；

本文采用线性核函数，对两个诊断实例的不同  $a$  值计算任务  $S \rightarrow R$ 、 $S \rightarrow R\_ft1$  和任务  $S \rightarrow R\_ft$  中 20 次重复实验的 MMD 距离平均值，计算结果如表 4-5 和表 4-6 所示。图 4-15 及图 4-16 更加直观地对比了三种不同任务下的 MMD 距离，可以看出，在两个诊断实例中，任务  $S \rightarrow R\_ft$  在所有  $a$  值下的 MMD 距离最小，说明迁移卷积层和全连接层相比于不迁移（对应任务  $S \rightarrow R$ ）和仅迁移卷积层（对应任务  $S \rightarrow R\_ft1$ ）能显著地缩小源域和目标域分布差异，虽然任务  $S \rightarrow R\_ft1$  中迁移了卷积核参数，但在大多数情况下反而增加了特征分布间的 MMD 距离，这可能是造成任务  $S \rightarrow R\_ft1$  有时性能较差的原因。以上分析说明了本文方法中迁移全连接层是必要的，其主要作用是缩小仿真数据和实际数据之间的特征分布差异，从而获得更好的诊断结果。

表 4-8 不同  $a$  值下的 MMD 距离 (SKF-6205)

	$S \rightarrow R$	$S \rightarrow R\_ft1$	$S \rightarrow R\_ft$
1	55.67	60.28	21.37
2	54.76	55.78	19.70
3	53.06	81.55	19.61
4	55.76	63.92	18.50
5	55.30	64.70	21.28
6	55.61	61.20	20.36
7	53.82	69.90	17.16

表 4-9 不同  $a$  值下的 MMD 距离 (SKF-6203)

	$S \rightarrow R$	$S \rightarrow R\_ft1$	$S \rightarrow R\_ft$
1	67.44	54.29	40.99
2	62.68	55.62	49.61
3	62.01	60.49	50.26
4	66.28	68.36	57.48
5	64.20	72.24	45.54
6	58.44	74.86	54.58
7	63.25	64.97	44.27

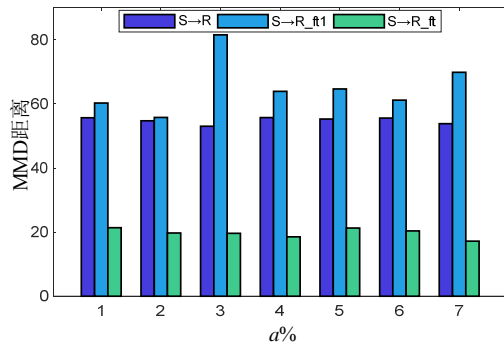


图 4-15 三种任务的 MMD 距离对比  
(SKF-6205)

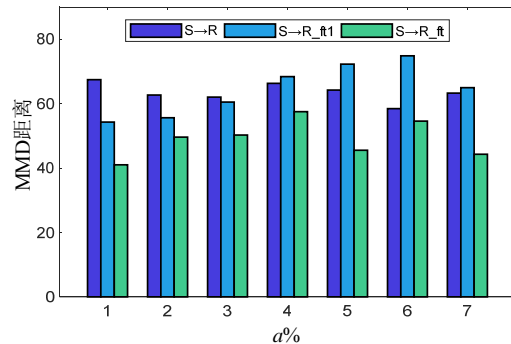


图 4-16 三种任务的 MMD 距离对比  
(SKF-6203)

#### 4.5.4 诊断方法的诊断机理总结

结合 4.5.1 节至 4.5.3 节对混淆矩阵、迁移卷积核和迁移全连接层的机理分析，并考虑轴承振动信号特性，给出小样本条件下，本文提出的诊断方法帮助实际诊断问题提升诊断性能的诊断机理如下：

(1) 在仿真模型中，仅考虑了滚动轴承相关的部件，忽略了复杂设备和多种随机因素的影响，仿真数据构成的样本可以看成是从理论上反映故障特征的样本。这些特征在预训练的过程中被 CNN 学习到。在实际诊断场景中，当故障发生时，由于故障冲击造成的更高频的频域特征出现，这些特征不仅与轴承相关部件有关，还与实际待诊断设备的复杂的固有属性及其他随机因素（如环境噪声）有关，虽然在实际信号中也具有相对低频的频率成分，这些频率成分的能量很低。因为 CNN 趋向于学习更加有区分性的特征来分类信号，所以在任务  $R \rightarrow R$  和任务  $S \rightarrow R_{ft}$  的微调过程中，CNN 更多地学习到的是在实际信号中占有主要地位的高频特征。

(2) 与采集自不同设备的振动信号在特征分布上存在差异类似，仿真信号和实际信号也存在于具有一定共性和差异性的特征空间中。因此，若仅将基于仿真数据训练的 CNN 网络用于实际场景的诊断，而不考虑任何迁移策略，诊断性能会非常差，对应任务  $S \rightarrow R$  的诊断结果。如果仅在极少量的可获取的实际场景的样本上训练 CNN，许多信号中的低频信息会丢失，而且由于数据量过少，整个实际待诊断数据集的特征分布很难在不同次随机选取的实际样本上被完整且一致地描述，这使得任务  $R \rightarrow R$  中的 CNN 网络泛化能力较差（识别率较低，识别率的标准差较大）。当 CNN 在仿真数据集上预训练，并利用可获取的少量实际样本微调后，从仿真数据和实际数据中提取的特征同时起作用，帮助实际诊断问题在小样本条件下获得更高的识别性能。

(3) 虽然迁移卷积核能够学习到更丰富有效的故障特征，但由 4.5.3 节的分析可以看出，迁移卷积核并不能缩小两个域之间的差异性。在本文所提出的方法中，迁移全连接层可以有效地缩小仿真域和实际域之间的分布差异，是实现实际诊断问题识别性能显著提升的重要原因。

#### 4.6 本章小结

本章应用第 3 章所提出的小样本条件下基于动力学仿真和深度迁移学习的滚动轴承诊断方法，将仿真数据的诊断知识迁移到实际问题中，实现了对两个不同



滚动轴承的有效诊断。通过将基于本文方法的诊断任务与所构造的多种其他诊断任务对比，验证了本文所提方法在实际诊断问题中的优越性。在此基础上，基于这两个诊断实例，从混淆矩阵、卷积核迁移机理和全连接层迁移机理三个角度分析了本文所提诊断方法有效的内在原因，并结合滚动轴承仿真信号和实际信号的特性，给出了本文方法对实际诊断场景中小样本问题的诊断机理。



## 结 论

本文针对故障诊断问题中,待诊断设备大量故障数据难以获得的小样本问题,以滚动轴承为研究对象,提出了一种基于动力学仿真和迁移学习的滚动轴承故障诊断方法。该方法利用卷积神经网络并结合参数迁移策略,将由滚动轴承动力学仿真模型生成的仿真数据中的诊断知识迁移到实际诊断场景中,帮助实际问题实现有效诊断。本文的主要研究工作总结如下:

(1)建立了滚动轴承四自由度动力学模型,考虑滚道的单点故障并引入滚动体的随机滑动,对滚动轴承正常状态,内圈故障及外圈故障进行仿真,得到仿真振动加速度仿真信号。通过将仿真信号和实际信号从时域和频域两个角度对比,说明仿真信号与实际信号具有相似的振动特性,即正常状态下 VC 振动特性,故障状态下的时域冲击特性与包络谱中故障频率特性,验证了仿真模型的合理性,为从仿真模型中生成大量故障模拟数据并实现从仿真模型迁移诊断知识到实际场景中奠定基础。

(2)详细分析了仿真数据和实际数据在动力学特性、时域信号概率分布及频域上的频率分布中存在的共性与差异性,并在此基础上提出了基于动力学仿真和迁移学习的滚动轴承智能诊断框架,该诊断方法为了解决滚动轴承故障诊断中的小样本问题,利用 CNN 预训练诊断模型,从大量仿真时域信号中自动化地学习信号特征,基于参数迁移策略,将预训练诊断模型中的网络参数迁移到实际诊断场景中,并利用少量可获取的实际样本微调整个网络,以实现实际问题的故障识别。

(3)实现了基于动力学仿真和迁移学习的滚动轴承智能故障诊断方法的两个实际应用,并构造多个对比诊断任务,验证了本文提出方法在小样本诊断问题上的优越性。另外,详细分析了诊断方法的诊断机理,说明通过仿真数据的预训练和实际数据的微调,诊断模型学习到了更丰富有效的故障特征,通过迁移全连接层,缩小了仿真数据和实际数据的特征分布差异,这些是使得诊断方法在实际小样本诊断问题中取得优越性能的主要原因。

然而,本研究还有一些不足之处,后续值得研究的工作包括以下几方面:

(1)基于其他迁移策略的小样本条件下故障诊断该方法研究。在本文提出的方法中,利用迁移学习的概念,通过网络参数迁移的方式实现了仿真模型到实际设备的知识迁移,没有用到更深层次的迁移理论,如域适应、域泛化、源域的选择问题、基于对抗思想的迁移等,若利用其他迁移策略能否构建更有效的小样本

条件下的故障诊断方法，是值得研究的问题。

(2) 当前诊断方法对其他对象的适用性研究。除了滚动轴承，针对其他机械设备，如齿轮箱，采用相同的思路，本文方法是否同样有效。

(3) 其他获得丰富故障数据的方法研究。本文采用动力学仿真的方式获取大量故障仿真数据，若采用其他方式，如基于生成式对抗网络，自动化地从随机噪声中学习出具有故障特征表现形式的样本，是否可行。

## 参考文献

- [1] Jardine A K S, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J]. Mechanical Systems and Signal Processing, 2006, 20(7): 1483–1510.
- [2] Gao Z, Cecati C, Ding S X. A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3768–3774.
- [3] Dai X, Gao Z. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis[J]. IEEE Transactions on Industrial Informatics, 2013, 9(4): 2226–2238.
- [4] 宋宝玉. 机械设计基础[M]. 第4版 哈尔滨: 哈尔滨工业大学出版社, 2011: 231-237.
- [5] Li Y, Yang Y, Wang X, et al. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine[J]. Journal of Sound and Vibration, 2018, 428: 72–86.
- [6] McFadden P D, Smith J D. Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing[J]. Journal of Sound and Vibration, 1984, 96(1): 69-82.
- [7] Aktürk N. The effect of waviness on vibrations associated with ball bearings[J]. Journal of Tribology, 1999, 121(4): 667–677.
- [8] Brie D. Modelling of the spalled rolling element bearing vibration signal: an overview and some new results[J]. Mechanical Systems and Signal Processing, 2000, 14(3): 353–369.
- [9] Tandon N, Choudhury A. An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect[J]. Journal of Sound and Vibration, 1997, 205(3): 275–292.
- [10] Sassi S, Badri B, Thomas M. A Numerical model to predict damaged bearing vibrations[J]. Journal of Vibration and Control, 2007, 13(11): 1603–1628.
- [11] Sawalhi N, Randall R B. Simulating gear and bearing interactions in the presence of faults. part I. the combined gear bearing dynamic model and the simulation of

- localized bearing faults[J]. *Mechanical Systems and Signal Processing*, 2008, 22(8): 1924–1951.
- [12] Moazen Ahmadi A, Petersen D, Howard C. A nonlinear dynamic vibration model of defective bearings - the importance of modelling the finite size of rolling elements[J]. *Mechanical Systems and Signal Processing*, 2015, 52–53(1): 309–326.
- [13] Singh S, Howard C Q, Hansen C H, et al. Analytical validation of an explicit finite element model of a rolling element bearing with a localized line spall[J]. *Journal of Sound and Vibration*, 2018, 416: 94–110.
- [14] Sakaguchi T, Harada K. Dynamic analysis of cage behavior in a tapered roller bearing[J]. *Journal of Tribology*. 2006,128: 604–611.
- [15] Mishra C, Samantaray A K, Chakraborty G. Ball bearing defect models: A study of simulated and experimental fault signatures[J]. *Journal of Sound and Vibration*, 2017, 400: 86–112.
- [16] Cao H, Niu L, Xi S, et al. Mechanical model development of rolling bearing-rotor systems: A review[J]. *Mechanical Systems and Signal Processing*, 2018, 102: 37–58.
- [17] Li Y, Wang X, Liu Z, et al. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: a review[J]. *IEEE Access*, 2018(6): 66723–66741.
- [18] Wang Y, Xiang J, Markert R, et al. Spectral kurtosis for fault detection , diagnosis and prognostics of rotating machines : a review with applications, *Mechanical Systems and Signal Processing*, 2016, 66–67: 679–698.
- [19] Li Y., Xu M., Liang X, et al. Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings[J]. *IEEE Transacitons. on Industrial. Electronics*, 2017, 64(8): 6506–6517.
- [20] Liu R, Yang B, Zio E, et al. Artificial intelligence for fault diagnosis of rotating machinery: A Review[J]. *Mechanical Systems and Signal Processing*, 2018, 108: 33–47.
- [21] Pandya D H, Upadhyay S H, Harsha S P. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN[J]. *Expert Systems with Applications*, 2013, 40(10): 4137-4145.

- [22] Yaquab M F, Gondal I, Kamruzzaman J. Inchoate fault detection framework: adaptive selection of wavelet nodes and cumulant orders[J]. IEEE Transactions on Instrumentation and Measurement, 2012, 61(3): 685-695.
- [23] Li Z X, Yan X P, Tian Z, et al. Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signal for gearing multi fault diagnosis[J]. Measurement, 2013, 46(1): 259-271.
- [24] Safizadeh M S, Latifi S K. Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and local cell[J]. Information Fusion, 2014, 18: 1-8.
- [25] 郭磊, 陈进, 朱义等. 小波支持向量机在滚动轴承故障诊断中的应用[J]. 上海交通大学学报, 2009, 43(4): 678-682.
- [26] Chen F F, Tang B P, Song T, et al. Multi-fault diagnosis study on rolling bearing based on multi-kernel support vector machine with chaotic particle swarm optimization[J]. Measurement, 2014, 47: 576-590.
- [27] Xu H B, Chen G H. An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO[J]. Mechanical Systems and Signal Processing, 2013, 35(1-2): 167-175.
- [28] Rafiee J, Arvani F, Harifi A, et al. Intelligent condition monitoring of a gearbox using artificial neural network[J]. Mechanical Systems and Signal Processing, 2007, 21(4): 1746-1754.
- [29] Lei Y G, He Z J, Zi Y Y. Application of an intelligent classification method to mechanical fault diagnosis[J]. Expert Systems with Applications, 2009, 36(6): 9941-9948.
- [30] Yaghobi H, Mashhadi H R, Ansari K. Artificial neural network approach for locating internal fault in salient-pole synchronous generator[J]. Expert Systems with Applications, 2011 38(10): 13328-133241.
- [31] Chen X F, Wang S B, Qiao B J. Basic research on machinery fault diagnostics: past present and future trends[J]. Frontier of Mechanical Engineering, 2018, 13(2): 264-291.
- [32] Khan S, Yairi T. A review on the application of deep learning in system health management[J]. Mechanical Systems and Signal Processing, 2018, 107: 241-265.

- [33]Jia F, Lei Y G, Lin J, et al. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical Systems and Signal Processing, 2016, 72-73: 303-315.
- [34]Gan M, Wang, C, Zhu C A. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearing[J]. Mechanical Systems and Signal Processing, 2016, 72-73: 92-104.
- [35]Li C, Sanchez R V, Zurita G, et al. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis[J]. Neurocomputing, 2015, 168: 119–127.
- [36]Lei Y G, Jia F, Lin J, et al. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5): 3137-3147.
- [37]Zhang W, Li C H, Peng G L, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load[J]. Mechanical Systems and Signal Processing, 2018, 100: 439-453.
- [38]Pan J, Zi Y, Chen J, et al. LiftingNet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification[J]. IEEE Transactions on Industrial Electronics, 2018, 65(6): 4973–4982.
- [39]Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis[J]. Measurement, 2016, 93: 490–502.
- [40]Yang Y, Zheng H, Li Y, et al. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network[J]. ISA Transactions, 2019.
- [41]Wen L, Gao L Li X Y. A new deep transfer learning based on sparse auto-encoder for fault diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017: 1-9.
- [42]龙明盛. 迁移学习问题与方法研究[D]. 北京: 清华大学学位论文, 2014.
- [43]Guo L, Lei Y, Xing S, 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.

- [44] Hasan J., Islam M.M.M., Kim J., Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions[J]. Measurement, 2019, 138: 620–631.
- [45] Zhang R, Tao H, Wu L. Transfer learning with neural networks for bearing fault diagnosis in changing working conditions[J]. IEEE Access, 2017, 5: 14347–14357.
- [46] Zhang B, Li W, Li X. Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks[J]. IEEE Access, 2018, 6: 66367–66384.
- [47] Qian W, Li S, Wang J. A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions[J]. IEEE Access, 2018, 6: 69907–69917.
- [48] Li X, Zhang W, Ding Q, et al. Multi-Layer domain adaptation method for rolling bearing fault diagnosis[J]. Signal Processing, 2019, 157: 180–197.
- [49] Xu G, Liu M, Jiang Z, et al. Online fault diagnosis method based on transfer convolutional neural networks[J]. IEEE Transactions on Instrumentation and Measurement, 2019: pp.1-12.
- [50] Yang B, Lei Y, Jia F, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings[J]. Mechanical Systems and Signal Processing, 2019, 122: 692–706.
- [51] Zheng H, Wang R, Yang Y, et al. Intelligent fault identification based on multi-source domain generalization towards actual diagnosis scenario[J]. IEEE Transactions on Industrial Electronics, 2019.
- [52] Wang M, Deng W. Deep visual domain adaptation: A Survey[J]. Neurocomputing, 2018, 312: 135–153.
- [53] Dai W Y, Yang Q, Xue G R, et al. Boosting for Transfer Learning[C]. Proceedings of the 24th International Conference on Machine Learning. ACM, 2007: 193-200.
- [54] 沈飞, 陈超, 严如强. 奇异值分解与迁移学习在故障诊断中的应用[J]. 振动工程学报, 2017, 30(1): 118-126.
- [55] Pan S J, Tsang W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.
- [56] 谢骏遥, 王金江, 赵锐等. 迁移因子分析在齿轮箱变工况故障诊断中的应用[J]. 电子测量与仪器学报, 2016, 30(4): 534-541.

- [57]康守强, 胡明武, 王玉静等. 基于特征迁移的变工况下滚动轴承故障诊断方法[J]. 中国机电工程学报, 2019, 39(03): 764-772+955.
- [58]Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks[C]. Montreal: The 28th Conference on Neural Information Processing Systems, 2014.
- [59]Chu B, Madhavan V, Beijbom O, et al. Best practices for fine-tuning visual classifiers to new domains[C]. Amsterdam: The 14th European Conference on Computer Vision, 2016: pp. 435–442.
- [60]Cao P, Zhang S, Tang J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning[J]. IEEE Access. 2018, 6: 26241–26253.
- [61]Zhong S, Fu S, Lin L, et al. A novel gas turbine fault diagnosis method based on transfer learning with CNN[J]. Measurement, 2019, 137: 435–453.
- [62]杨晓蔚. 国内外深沟球轴承设计方法的对比[J]. 轴承, 2010(02): 59–61.
- [63]Brewer D E, Hamrock B J, Simplified solution for point contact deformation between two elastic solids[J]. Journal of Lubrication Technology, 1977, 99(4): 485–487.
- [64]崔江. 双转子-滚动轴承系统的动力学特性研究[D]. 哈尔滨: 哈尔滨工业大学硕士学位论文, 2017.
- [65]成大先. 机械设计手册: 单行本. 轴承[M]. 北京: 化学工业出版社, 2017.
- [66]Goodfellow I, Bengio Y. 深度学习[M]. 北京: 人民邮电出版社, 2017.
- [67]Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359.
- [68]Lu W N, Liang B, Cheng Y, et al. Deep Model Based Domain Adaptation for Fault Diagnosis[J]. IEEE Transactions on Industrial Electronics, 2017, 64(3): 2296–2305.
- [69]Gretton A, Borgwardt K M, Rasch M J et al. A kernel method for the two-sample-problem[J]. 2007. Available: arXiv preprint arXiv: 0805.2368.



## 攻读硕士学位期间发表的学术论文

## 哈尔滨工业大学学位论文原创性声明和使用权限

### 学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于动力学仿真和迁移学习的滚动轴承故障诊断方法研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：董韵佳

日期：2019年6月22日

### 学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：董韵佳

日期：2019年6月22日

导师签名：徐永强

日期：2019年6月22日

## 致 谢

两年的硕士生活即将结束，这两年中，我经历了学业上取得成绩的喜悦与失败时的迷茫，也在生活中感受过细水长流的幸福与一些不可控的痛苦。在向硕士生涯告别之际，太多情绪难以言表，剩下的唯有感谢。

首先，感谢我的导师徐敏强老师。徐老师严谨的学术精神和勤奋的工作态度深深影响着我。每次与徐老师交流，徐老师总能揭示问题的本质所在，令我受益匪浅。在组会上，徐老师为我的研究不断提出新的意见和建议，使我的研究更严谨更深入。在此衷心地感谢徐老师的悉心指导，祝愿徐老师身体健康，生活幸福。另外，还要感谢王日新老师和李玉庆老师，两位老师在日常生活和学习中都给予了我很大的帮助，给了我许多锻炼自己的机会，并对本文的研究提出的许多宝贵意见，在此对两位老师表示真诚的谢意。

我还要感谢课题组的师兄师姐，与你们一起吃饭是我一天中最开心的时光。感谢郑怀亮师兄对我的一切指导与帮助，在我失败时给我中肯的意见，在我成功时鼓励我，同时师兄对待学术认真严谨，是我学习的榜样与力量；感谢尹建程师兄在动力学建模上对我的帮助；感谢魏禹师姐对我生活上的关心与包容，让我在人生的艰难阶段倍感温暖。感谢雷明佳，冯小恩，董诗音，江飞龙在工作和生活中对我的帮助，感谢课题组师弟师妹们的陪伴。

感谢我的朋友石乃文，我们一起度过了四年本科和两年硕士，有很多美好的回忆，希望以后的日子我们能共同进步，友谊长存。

最后，我要感谢我的父母和亲人，感谢爸爸妈妈的养育之恩，你们无条件的支持与付出让我有了今天，我会继续努力，做更好的自己，遇见更好的未来，让爸爸和天堂的妈妈放心。