DOI: 10.3901/JME.2021.15.080

基于改进后半监督深度信念网络的多工况 轴承故障诊断研究*

叶楠^{1,2} 常佩泽^{1,2} 张露予^{1,3} 王嘉^{1,2}

- (1. 河北工业大学省部共建电工装备可靠性与智能化国家重点实验室 天津 300401;
2. 河北工业大学机械工程学院 天津 300401;
3. 河北工业大学先进材料测试与分析中心 天津 300401)

摘要: 随着工业自动化的发展, 各类旋转机械在多工序、多任务约束下, 其轴承转速与载荷均会发生变化, 导致针对单一工况的故障诊断方法的准确率大打折扣。因此, 提出一种多工况约束下轴承故障诊断方法, 该方法基于半监督深度信念网络(Semi-supervised deep belief network, SSDBN), 利用少量标记数据即可完成故障分类、判断, 提高了跨工况故障诊断的准确率。首先将源域与目标域设置为负载相同、转速与损伤尺寸不同的数据集, 通过小波包分解对源域和目标域信号进行重构; 在此基础上, 应用最大平均差算法(Maximum mean discrepancy, MMD)作为源域和目标域数据特征分布差异评价指标, 筛选出分布差异较小的特征样本数据; 利用改进后的半监督深度信念网络训练较少的标签数据和大量无标签数据, 提高待测数据的分类精度; 最后, 以凯斯西储大学轴承数据集为例进行试验, 验证相同负载不同转速与损伤尺寸工况下模型的诊断精度, 以及不同负载、转速、损伤尺寸工况下的模型诊断精度。结果表明该方法能够提高轴承在多工况约束下的故障诊断准确率, 减少故障诊断的误报率, 并能降低训练模型过程中的梯度消失现象, 提高故障分类成功的概率。另外, 提出的方法无需考虑各种特征的灵敏度, 亦无须依赖专家知识与经验, 具有较强的普适性与兼容性, 有利于及时发现并替换损坏轴承, 保证机械设备安全、可靠地运行。

关键词: 轴承; 半监督深度信念网络; 最大平均差算法; 故障诊断

中图分类号: TH133; TN911

Research on Multi-condition Bearing Fault Diagnosis Based on Improved Semi-supervised Deep Belief Network

YE Nan^{1,2} CHANG Peize^{1,2} ZHANG Luyu^{1,3} WANG Jia^{1,2}

- (1. State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300401;
2. School of Mechanical Engineering, Hebei University of Technology, Tianjin 300401;
3. Advanced Materials Testing and Analysis Center, Hebei University of Technology, Tianjin 300401)

Abstract: With the development of industrial automation, the processes in the production of rotating systems are becoming more and more sophisticated and complex. When multiple processes are executed, the internal bearing speed and load are constantly changing. For this type of multi-working conditions, the original fault diagnosis method trained in a single working condition is no longer applicable. Therefore, a bearing fault diagnosis method for bearings under multiple operating conditions is proposed. This method is based on a semi-supervised deep belief network (Semi-supervised deep belief network, SSDBN). Using a small amount of labelled data to conduct the fault classification and judgment, and it can improve the accuracy of fault diagnosis under varying working

* 国家自然科学基金(72001069, 52075146)、河北省优秀青年基金(E2021202094)、河北省自然科学基金创新群体(E2020202142)、国家重点研发计划(2017YFB1301300)和国防科技大学装备综合保障技术重点实验室基金资助项目。20200811 收到初稿, 20210322 收到修改稿

conditions. First, set the source and target domains to the same load, different speed and damage size data set, and process the source and target domain signals by wavelet packet decomposition. Then the maximum mean discrepancy algorithm ((Maximum mean discrepancy, MMD) is used as the evaluation indicator of the difference between the distribution of data features in the source and target domains, and the smaller distribution difference is selected. Feature sample data, make the model easier to compare features, improve the classification accuracy of the target domain data. Use the improved semi-supervised deep belief network to train less labelled data and a large number of unlabelled data, improve the classification accuracy of the data to be tested. Take the bearing data set of Western Reserve University as an example to verify the diagnostic accuracy of the model under the same load at different speeds and damage size conditions, as well as the diagnostic accuracy of the model under different load, speed and damage size conditions. The results show that the method can improve the accuracy of the fault diagnosis of the bearing under the constraints of multiple operating conditions, reduce the false alarm rate of the fault diagnosis, and reduce the gradient disappearance during the training model, and increase the probability of successful failure classification. In addition, the proposed method does not need to be considered the sensitivity of various features does not need to rely on expert knowledge and experience. It has strong universality and compatibility, which is helpful to find and replace damaged bearings in time and ensure the safe and reliable operation of machinery and equipment.

Key words: bearing; semi-supervised deep belief network; maximum mean difference algorithm; fault diagnosis

0 前言

轴承广泛应用于风电机组、航天飞行器、汽车、工业机器人等旋转机械设备中。由于这些机械设备经常在复杂的工况环境下(如高温、高湿、变速和变负载^[1])服役,轴承容易出现润滑失效,或混入微小磨粒、产生微小裂纹等,发生诸如点蚀、磨损、胶合、断裂、塑性变形等故障,甚至导致机械系统整体瘫痪,造成巨大的经济损失^[2-3]。另外,由于生产过程集成化、智能化发展,单一设备经常面临工况、载荷同时发生变化的情况。因此,如何提升变工况下轴承的故障诊断,关系到整个设备及生产线的稳定运行。

传统轴承故障诊断主要从信号特征提取角度分析轴承的故障特征频率,例如张小龙等^[4]使用固有时间尺度分解-形态滤波与 Teager 能量谱提取滚动轴承故障特征,降低了原始信号的背景噪音,筛选出了效果较好的故障特征值;李宏坤等^[5]为了进一步挖掘低信噪比情况下的轴承故障特征信息,结合粒子滤波与谱峭度分析,去除干扰成分,成功提取了较为明显的故障特征频率;李志农等^[6]利用小波分析与核典型相关分析重构了轴承的原始振动信号,分离出了带有较为明显故障特征频率的信号。上述研究主要从去除噪声干扰这一目的出发,提取轴承的特征故障频率,但这种方法需要大量的标记信号样本,同时还需要人工选取特征。另外上述研究都从单一工况验证方法的有效性,缺乏对变工况条件下的故障诊断探索研究。

智能故障诊断方法多以神经网络作为基础,不需要人工选择特征,无需辨别特征的灵敏度,致力于快速有效地实现“端到端”学习。常用的神经网络主要包括人工神经网络^[7-8] (Artificial neural network,ANN)、自动编码器^[9] (Auto-encoder,AE)、受限玻尔兹曼机^[10] (Restricted Boltzmann machine,RBM)和卷积神经网络^[11] (Convolutional neural network,CNN),它们对于人类诊断知识的依赖性较小,能够处理较大批量信号数据,在早期故障不明显的应用下应用十分广泛。智能故障诊断方法分为无监督、有监督、半监督三类。无监督诊断模型使用未标记数据来提取特征,凭借相似性比较分类,致使该类方法对噪声的敏感性较强,分类识别精度达不到理想效果;有监督诊断模型致力于使用标记数据提取特征,虽然提升了模型的分类识别精度,但这通常需要大量昂贵的标记数据,并且容易因随机初始化权重引起模型陷入局部最小值的现象;半监督诊断模型所需的标签数据很少,并能在大量廉价无标签数据的帮助下辅助分类,提升模型分类精度,在充斥着大量无标签数据的大数据时代,半监督诊断模型的应用会越来越广。

近年来,深度信念网络是最为流行的模型之一,其学习方式有助于克服随机初始化权重引起的局部最小值现象,在轴承故障诊断方面得到了较广泛的应用。如杨宇等^[12]利用自适应深度信念网络对滚动轴承故障分类,通过熵误差的稳定状况与阈值状况选择保留网络结构或者增加网络深度,最终得到结果的精度很高;沈长青等^[13]引入了一种独立自适应学习率优化深度信念网络,从网络优化算法角度改

变, 替换了传统的随机梯度下降法, 从而使整体算法可以自适应地处理数据; 赵光权等^[14]使用深度信念网络深度挖掘故障特征, 在负载、损伤尺寸变化的情况下证明了其所提方法的有效性; 其他学者也在应用深度神经网络对轴承故障诊断领域进行了一系列深入探索^[15-17]。然而, 以上研究均未涉及变工况条件(负载、转速等均发生变化)下的轴承故障诊断研究, 实际应用中可能出现迁移能力不足, 导致故障分类错误, 并且深度信念网络偏重于无监督诊断, 分类结果的稳定性得不到保证。

为了解决上述问题, 提出一种基于改进后的半监督深度信念网络(Improved semi-supervised deep belief network, ISSDBN)的轴承故障诊断方法, 该方法具有较好的迁移学习能力, 可用于解决多工况下轴承故障识别精度低的问题。首先利用小波包分解对原始振动信号预处理, 提高重构信号的分辨率; 在此基础上, 提取一些时域特征并应用最大平均差算法作为评价指标, 筛选出源域与目标域特征分布差异较小的特征样本, 便于模型进行特征对比学习, 提高模型迁移学习能力与诊断精度; 利用半监督深度信念网络对特征样本进行故障诊断, 用少量标记数据与大量未标记数据拟合模型, 使模型通过自我学习提高待测数据分类精确性。另外, 使用改进后的 sigmoid 激活函数(Improved logistic sigmoid, Isigmoid), 通过分段函数与阈值的限制, 减少梯度消失现象, 加快收敛速度, 提高模型诊断精度; 最后用美国凯斯西储大学(Case Western Reserve University, CWRU)的轴承数据集验证了本文所提方法在多工况下轴承故障诊断的有效性。

本文分为四部分, 第 1 节主要阐述多工况下轴承识别精度低的原因及要解决的问题; 第 2 节介绍了所提方法的主要流程及组成; 第 3 节利用 CWRU 的轴承数据集对所提方法的有效性进行验证; 第 4 节对本文中的研究进行总结与展望。

1 问题提出

1.1 诊断方法迁移能力不足问题

以制造业中应用广泛的工业机器人为例, 工业机器人因为其灵活高效的属性, 广泛应用于汽车、铸造、食品、医药、电子等行业, 承担焊接、搬运、组装等各种高负载任务。由于这些工作内容涉及的运动姿态更改频繁, 导致应力及转速不断变化, 因此对工业机器人的健康信息监测及故障诊断须适应其灵活多变的工况条件。例如打磨机器人在执行任

务过程中, 要求末端在工件表面匀速直线运行, 为达到这一要求, 工业机器人中各关节处的轴承转速与负载必须时刻变化, 而不是在固定速度与负载下运行。当完成一道工序后, 工业机器人将会回到预设原点, 这一过程中, 各关节轴承的转速与负载仍在继续变化。由此可见, 打磨机器人在作业过程中, 各关节转速与所受负载总是发生高频率变换, 这对于泛化性能差、迁移能力低的非智能轴承故障诊断方法来说是极大的挑战。此外, 精密数控机床在加工工件过程中往往需要更换不同刀具并使用不同转速, 这时主轴轴承上的转速与所受载荷均在不断变化, 迁移能力较差的诊断模型难以适应多变的工况, 导致诊断精度较低。类似的问题也存在于各类自动化程度较高、需要多个工序才能制造出工件的旋转机械中。由于单一工况训练出的智能诊断方法迁移能力不足, 需要一种迁移能力较强, 适合多工况条件的诊断方法, 以提升诊断精度, 降低误报率。

此外, 大多数故障诊断方法均假设训练模型数据与测试数据符合相同的分布^[18], 但实际的测试过程中两者的分布往往是不同的, 原有的模型无法从有较大分布差异的数据中学习关联特征, 从而造成一些故障诊断模型诊断准确率下降。

通常把用于训练模型的数据称为源域, 用于测试模型但相关的未标记的数据集称为目标域, 现有研究将这种问题归结为迁移学习问题^[19]或域移位问题^[20], 其原理如图 1 所示。

图 1 域移位现象

解决上述问题需要用到迁移学习, 它是利用已经习得的知识对不同但相关的问题进行求解的一种新的机器学习方法^[18], 许多学者已经将迁移学习的思想运用到多工况下轴承的故障诊断领域, 例如 AN 等^[21]基于多核方法与最大平均差算法对轴承进行故障诊断, 在两种不同的数据集中分别验证负载变化与转速变化对模型的影响, 最后得到兼容性较好的模型; TONG 等^[22]将神经网络与最大平均差算法

(Maximum mean discrepancy, MMD) 结合, 并对数据用短时傅里叶变换做预处理。以上方法证明迁移学习能够应用于多工况轴承故障诊断问题, 但在故障诊断准确度方面仍有待提升。本文将最大平均差算法引入半监督深度神经网络算法, 可有效缩减源域与目标域的分布差异, 提高跨域诊断的精度。

1.2 标签数据困境问题

旋转机械的轴承故障标记数据稀少, 会使训练集中样本不足, 训练出的模型精确度不高, 对有效识别轴承故障带来了极大挑战。

这一问题可从数据价值与数据形式两方面进行阐述, 首先是数据价值, 通常情况下轴承封闭在旋转机械的内部, 传感器难以监测到轴承实际的振动信号, 因此, 需要搭建专门的轴承振动试验平台来模拟旋转机械的实际工况。高精度平台价格昂贵, 同时会消耗大量人力物力和时间成本, 这导致标记数据稀少而昂贵。另一方面是数据形式, 旋转机械实际工作环境下的工况可是看作是连续型数据, 而试验台采集得到的标记数据是离散型数据, 利用少量离散数据推测拟合连续数据的变量时间关系, 很容易产生欠拟合现象, 使模型预测失真。理论上讲, 采集的离散数据越多, 越接近连续数据, 然而实际采集足够多的试验数据很难完成。同时由于传感器安装精度、其他设备噪声等因素的影响使得高质量标签数据愈发稀少。因此, 如何从较为稀少数据中挖掘更有价值的故障信息成为目前故障诊断要解决的主要问题。

2 改进后的半监督深度信念网络

针对以上提出的问题, 提出一种将 MMD 应用于改进后的半监督深度信念网络的方法, 改进后的方法具有一定程度的迁移学习能力, 可以提升多工况轴承故障诊断精度。另外提出的方法充分利用少量标记数据, 使用大量无标记数据辅助提升模型精度, 同时 Isigmoid 激活函数的使用减少了模型梯度消失的现象, 使方法更加稳定。

改进后的半监督深度信念网络如图 2 所示。

S 与 T 分别代表源域特征样本与目标域特征样本, 黑色圆圈部分代表隐层与可视层的神经网络主体, 灰色圆圈部分为监督层。方法中改进的神经网络共有四层神经元, 组成 3 个半监督受限玻尔兹曼机(Semi-supervised restricted Boltzmann machine, SSRBM)。灰色监督层与黑色网络主体同时训练, 并由监督层辅助判断网络主体生成权重与偏差是否

正确。当训练完成后, 监督层与网络主体均有反向传播算法进行参数微调。

图 2 ISSDBN 结构示意图

在保证原始信号完整性的前提下, 本文使用小波包分解重构信号, 使信号的分辨率提高, 且为了增强信号对故障类型的敏感度, 对重构信号进行特征提取, 以此构成源域与目标域特征样本。另外, 为了提高模型的迁移能力, 使用 MMD 来度量源域与目标域样本的分布差异, 只有低于阈值的样本允许输入到半监督深度信念网络中, 这样有效隔离了分布差异较大的样本, 便于模型进行特征对比, 提高诊断精度。

半监督深度信念网络已被证实在 mnist 手写数据集中的识别效果较好^[23], 网络由 SSRBM 堆叠而成, 其增加的监督层可通过隐藏层共享和权重绑定将认知模型与生成模型集成在一起, 充分利用少量的标记数据与大量的未标记数据, 通过自我学习完成目标域数据的识别, 有效提高诊断的准确度。

受限玻尔兹曼机的堆叠实际上是 sigmoid 函数的嵌套相乘, 数值点容易落在函数趋近于 0 的位置上产生连乘效应, 进而引起梯度消失的现象。本文提出的 ISSDBN 方法使用 Isigmoid 激活函数, 利用分段函数与阈值将 sigmoid 中趋近于 0 的部分数值提升, 进而降低梯度消失现象发生的概率。ISSDBN 的学习过程包括两部分: 半监督预训练和监督微调。具体的预训练过程如下: (1) 最左侧的 SSRBM 的可见层和监督层分别接收样本数据和监督数据以开始训练; (2) 左侧 SSRBM 的输出充当相邻右侧 SSRBM 的输入; (3) 重复过程(1)和(2)直到训练至所设置层数。经过半监督的预训练后, ISSDBN 的参数被确定在一个较小且效果更好的范围内, 然后应用反向

传播算法来微调这些参数。

图 3 为提出方法基本流程, 包括信号获取、数据预处理、特征提取、训练过程、微调过程和诊断结果六个阶段。

图 3 ISSDBN 诊断流程图

具体步骤如下。

(1) 利用小波包分解对原始振动信号分解重构, 对重构信号提取时域特征, 对提取的特征分组, 标记十类不同状态的数据。

(2) 训练集中未标记的数据随机输入。

(3) 与标签数据组比较 MMD 估计值后, 超过阈值的返回步骤(2), 低于阈值的进行下一步。

(4) 确认 ISSDBN 诊断模型的隐层 l , 激活 Isigmoid 函数。

(5) 初始化 ISSDBN 中 SSRBM 的权重与偏置。

(6) 逐步训练第 i 个 SSRBM, 若 $i \leq l$ 则返回步骤(5), 否则, 程序继续进行。

(7) 利用反向传播算法微调模型, 结束模型的训练并输出诊断结果。

2.1 MMD 算法

在域自适应方面, MMD 算法^[24]已经广泛用于测量源域与目标域之间标准分布距离度量, 在本文中, 将其用于筛选出特征分布近似的样本。给定两个分布 s 和 t , 它们之间的 MMD 定义为

$$L_M(s, t) = \sup_{\|\phi\| \leq 1} \left\| E_{x^s \sim s} [\phi(x^s)] - E_{x^t \sim t} [\phi(x^t)] \right\|_H^2 \quad (1)$$

式中, E 为对分配的期望, ϕ 表示将原始数据映射

到再生核希尔伯特空间(RKHS)的函数。当 s 与 t 分布相同时, $MMD^2(s, t) = 0$, 与此映射关联的内核函数为 $k(x^s, x^t) = \langle \phi(x^s), \phi(x^t) \rangle$ 。

$$K = \left\{ k = \sum_{u=1}^m \beta_u k_u : \sum_{u=1}^m \beta_u = 1, \beta_u \geq 0, \forall u \right\} \quad (2)$$

式中, K 为 u 个不同核宽度的高斯核函数的凸集合, β_u 为第 u 个核的系数, 常用的核是高斯核 $k_u(x^s, x^t) = \exp(-\|x^s - x^t\|^2 / \gamma)$, $D_s = \{x_i^s\}_{i=1}^M$ 与 $D_t = \{x_i^t\}_{i=1}^M$ 分别表示分布 s 与分布 t , MMD 的经验估计如下

$$L_M(D_s, D_t) = \left(\frac{1}{M} \sum_{i=1}^M \phi(x_i^s) - \frac{1}{N} \sum_{i=1}^N \phi(x_i^t) \right)_H^2 \quad (3)$$

变形后的公式(3)更适合以小批量的形式进行梯度计算

$$L_M(D_s, D_t) = \frac{1}{M(M-1)} \sum_{i \neq j}^M k(x_i^s, x_j^s) + \frac{1}{N(N-1)} \sum_{i \neq j}^N k(x_i^t, x_j^t) - \frac{1}{MN} \sum_{i,j=1}^{M,N} k(x_i^s, x_j^t) \quad (4)$$

MMD 能够对网络进行优化, 根据域不变特征, 即在源域和目标域据有相同或相似的分布, 做出分类决策, 从而达到将源数据训练的模型应用于目标数据的目的^[25]。

2.2 Isigmoid 激活函数

为了减少 sigmoid 函数对深度信念网络整体造成的梯度消失问题, Isigmoid^[26]函数结合 sigmoid 和 LReLU 函数的部分特点, 其等式如下所示

$$f_l(x) = \begin{cases} \alpha(x-a) + \text{Sigmoid}(a) & x \geq a \\ \text{Sigmoid}(a) & -a < x < a \\ \alpha(x+a) + \text{Sigmoid}(a) & x \leq -a \end{cases} \quad (5)$$

$$f'_l(x) = \begin{cases} \alpha & |x| \geq a \\ \text{Sigmoid}'(x) & |x| < a \end{cases} \quad (6)$$

式中, a 是阈值, α 是斜率, 两者需要满足如下等式

$$\alpha > \alpha_{\min} = \frac{\exp(-a)}{(1 + \exp(-a))^2} \quad (7)$$

式中, $\alpha_{\min}$ 是使 Isigmoid 工作的最小斜率。

2.3 半监督受限玻尔兹曼机(SSRBM)

半监督学习通过部分标记与未标记的混合样本

来预测未标记样本的标签, SSRBM 也是基于这种思想被设计使用^[23]。SSRBM 的结构如图 4 所示。

图 4 SSRBM 的结构图

其中, h 表示隐层, v 表示显层, u 表示监督层, w 是隐层与显层之间的权重, p 是监督层与隐层之间的权重。SSRBM 的能量函数定义为

$$E(v, u, h / \psi) = -\sum_{i=1}^m a_i v_i - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m \sum_{j=1}^n a_i w_{ij} h_j - \lambda \sum_{k=1}^m c_k u_k - \lambda \sum_{k=1}^m \sum_{j=1}^n u_k p_{kj} h_j \quad (8)$$

式中, $\psi = (w_{ij}, p_{kj}, a_i, c_k, b_j)$, λ 为控制有监督学习与无监督学习比例的权重参数, SSRBM 条件概率分布为

$$P(h_j / v, u, \psi) = \sigma \left(\sum_{i=1}^m v_i w_{ij} + \lambda \sum_k u_k + b_j \right) \quad (9)$$

$$P(v_j / h, \psi) = \sigma \left(\sum_{j=1}^n h_j w_{ij} + a_i \right) \quad (10)$$

$$P(u_k / h, \psi) = \sigma \left(\lambda \sum_{j=1}^n h_j p_{kj} + \lambda c_k \right) \quad (11)$$

$\psi = (w_{ij}, p_{kj}, a_i, c_k, b_j)$ 可以更新为

$$\psi^{\tau+1} = \psi^{\tau} + \eta \frac{\partial \log P(v, u, h)}{\partial \psi} \quad (12)$$

其中 τ 为迭代次数, η 为学习率。半监督深度信念网络由 SSRBM 顺序堆叠而成, 第一个可视层可被视为输入层, 其余各层均由图 4 所示 SSRBM 构成。

3 案例分析

本节应用 CWRU 的滚动轴承数据集对提出的方法进行验证^[28]。该数据集包含的故障类型较多, 且在不同转速、负载、损伤尺寸下都有对应的故障轴承振动信号, 选择不同的数据可构成不同工况的数据集, 可模拟工业机器人等旋转机械轴承所处的

真实工况, 同时可用于测试所提模型的诊断精度。

案例设置两次测试, 第一组测试内容为转速、损伤尺寸变化下轴承的故障诊断; 第二组测试为负载、转速、损伤尺寸均变化下轴承的故障诊断。下文将说明包括数据集设置、参数设置、特征提取、信号预处理在内的前期准备以及两次测试的结果。

3.1 数据说明及预处理

CWRU 的轴承试验设备由 2 HP 的电机、扭矩传感器、功率测试计等组成, 驱动端轴承型号为 SKF6205, 试验台结构如图 5 所示^[28]。

图 5 CWRU 轴承试验台

试验利用电火花加工方法分别在轴承内圈、滚动体、外圈上制造单点损伤。本文使用的数据类型为损伤直径为 0.178 mm、0.356 mm、0.533 mm, 负载为 0 HP、1 HP、2 HP, 采样频率为 12 kHz, 健康状态为正常状态、内圈损伤、滚动体损伤、外圈损伤, 轴承转速分别为 1 797 r/min、1 772 r/min 以及 1 750 r/min 下的数据, 具体数据集如表 1 所示

表 1 轴承数据集设置

数据代号	负载/HP	轴承状态	损伤尺寸/mm
A	0	正常、内圈故障、滚动体故障、外圈故障	0、0.178
			0.356、0.533
B	1	正常、内圈故障、滚动体故障、外圈故障	0、0.178
			0.356、0.533
C	2	正常、内圈故障、滚动体故障、外圈故障	0、0.178
			0.356、0.533

案例设置了两组对比测试: 测试 1 分别对 A、B、C 数据集进行单数据集识别, 用于检验在变化速度与损伤尺寸条件下模型的准确性。三个数据集中每个数据集均包含 10 类不同状态数据, 标签为 1~10, 分别对应正常状态、三种不同损伤尺寸与三种不同故障位置, 详细标签划分如表 2 所示。每个数据集样本数为 18 000, 分别用 MMD 筛选出 50 组数据, 其中随机取 10 组为标记数据, 其余 30 组为未标记数据, 这 40 组数据作为训练集, 另外 10 组数据作为测试集。测试 2 对 A、B、C 数据集进行跨数据集识别, 以 A 中 10 组标记数据、30 组未标记数据为训练集, B 中 10 组未标记数据为测试集, 再以同样的方法检验 B 和 C, 验证模型在负载、转

速、损伤尺寸均变化时模型识别故障的准确率。

表 2 轴承标签设置

标签	损伤尺寸/mm	故障位置
1	0	无
2	0.178	内圈
3	0.356	滚动体
4	0.533	外圈
5	0.356	内圈
6	0.178	滚动体
7	0.356	外圈
8	0.533	内圈
9	0.533	滚动体
10	0.178	外圈

两次测试中半监督深度信念网络中学习率设置为 0.1, 学习率根据经验设置, 训练迭代次数为 10, 反向微调迭代次数为 1 000, 隐层节点数均为[36 20 15 10], 根据样本与输出确定了初始与结束时的隐层节点数, 分别为 36 与 10, 根据经验前一层输入节点大于后一层节点数, 剩余两层节点数利用多次试验进行调节。准确度数值比较见表 3。

表 3 不同隐层节点对应模型准确率

节点数	准确率 (%)	节点数	准确率 (%)	节点数	准确率 (%)
[35,12]	92.31	[30,12]	93.29	[20,12]	96.21
[35,15]	94.21	[30,15]	95.26	[20,15]	97.57
[35,17]	91.28	[30,17]	94.77	[20,17]	97.25
[35,20]	93.89	[30,20]	95.12	[20,20]	96.33

经过多次试验对比发现[36 20 15 10]对应的准确率最高, 因此选用该数量的隐层节点。另外, 通过多次试验获得 Isigmoid 函数中阈值为 5, 斜率为 0.2。表 4 列出一些阈值与斜率准确率比较, 每组试验进行 5 次, 对测试结果求平均值。

为了使结果更方便观察, 表 4 中的 10 次准确率数据变化如图 6 所示。

表 4 Isigmoid 阈值与斜率组合准确率比较

阈值	斜率	准确率 (%)
8	0.01	96.92
8	0.05	97.17
8	0.1	97.10
8	0.2	97.19
8	0.5	96.95
7	0.2	97.39
6	0.2	97.38
5	0.2	97.49
4	0.2	96.61
3	0.2	97.22

图 6 Isigmoid 不同阈值与斜率面准确率对比图

根据图 6 中放大的部位显示, 在最终收敛的准确率方面, 斜率在 0.2 时准确率相对其他数据较高, 在迭代效率方面, 基本优于其他数据。根据整体准确率走向综合选择 $\alpha = 5$, $\alpha = 0.2$ 。

首先对数据进行小波包处理, 小波包分解可以提高各频段信号分辨率, 并能有效保留有价值信号, 过滤噪声, 本文选用 Morlet 小波函数, 分解层数为 5 层。其次, 提取信号特征量进行故障诊断。为了便于研究, 本文仅选用时域特征量。时域信息对噪声的反应更为直接, 且时域特征对噪声更加敏感, 适合处理早期微弱故障^[27], 在后续处理过程中容易去除噪声的干扰。具体定义如表 5 所示。

表 5 时域特征表

名称	定义	名称	定义
均值	$\frac{1}{N} \sum_{i=1}^N x_i$	峰值	$\max x_i $
均方值	$\frac{1}{N} \sum_{i=1}^N x_i^2$	峰峰值	$\max x_i - \min x_i $
均方根值	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	方差	$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$
平均幅值	$\frac{1}{N} \sum_{i=1}^N x_i $	标准差	$\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$
方根幅值	$\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$	歪度值	$\frac{1}{N} \sum_{i=1}^N x_i^3$
脉冲因子	$\frac{\max x_i }{\frac{1}{N} \sum_{i=1}^N x_i }$	峭度值	$\frac{1}{N} \sum_{i=1}^N x_i ^4$
偏度因子	$\frac{\frac{1}{N} \sum_{i=1}^N x_i ^3}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \right)^3}$	峭度因子	$\frac{\frac{1}{N} \sum_{i=1}^N x_i ^4}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \right)^4}$
峰值因子	$\frac{\max x_i }{\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}}$	裕度因子	$\frac{\max x_i }{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2}$
波形因子	$\frac{\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}}{\frac{1}{N} \sum_{i=1}^N x_i }$	波形熵	$\frac{1}{N} \sum_{i=1}^N x_i \log x_i$

表 4 中 x 表示为样本的数值, N 代表样本总体

数量, $i=1,2,3,\cdots,n$ 。特征量共 36 维, 每类特征选取 2 次, 这样无需考虑特征的灵敏度, 无需靠专家知识与经验来挑选特征。

3.2 测试 1

分别对 A、B、C 数据集进行小波包分解进而筛选数据, 分成 A→A、B→B、C→C 三种诊断模式。训练过程中, 对比 DBN 与 ISSDBN 在 3 个玻尔兹曼机的重构误差大小, 图 7 表示三个由 sigmoid 激活的 RBM 与三个 Isigmoid 激活的 SSRBM 重构误差对比。从图中可以看出 Isigmoid 的重构误差要低于 sigmoid, 可以证明 Isigmoid 能有效降低误差。

图 7 迭代误差对比图

本文选择常用的智能诊断方法主成分分析 (Principal component analysis, PCA)、堆叠自动编码器 (Stacked auto-encoder, SAE) [29-30]、深度信念网络 (Deep belief network, DBN) [31-32] 与提出的 ISSDBN 方法进行识别准确性对比。SAE 的节点设置为 [500 200 100 10 100 200 500], 学习率为 0.01, DBN 的超参数设置与 ISSDBN 相同。

为详细描述 ISSDBN 识别准确率与负载的关系, 表 6 列出 5 次试验中第一组测试下对应的准确率。通过该表可以看出, 除了第 2 次试验外, 其他试验中随着负载的增大, ISSDBN 模型的识别精度随之下降, 表明工况变化会影响模型的识别精度。

表 6 ISSDBN 数据集识别准确率对比 %

试验序号	A	B	C
1	99.62	98.89	96.83
2	98.65	99.26	97.86
3	98.76	98.45	97.96
4	99.47	98.99	97.81
5	99.55	98.96	96.94

表 7 为第一组测试的识别精度。由该表可以看出, 所提方法准确率均高于其他三种方法。

表 7 不同方法各数据集识别总准确率对比 %

数据集	A	B	C
PCA	31.99	30.26	30.17
SAE	78.02	75.23	72.36
DBN	80.27	81.97	79.64
ISSDBN	99.21	98.91	97.48

图 8 为各方法识别精确度对比的柱状图, 能够更直观地展现出所提方法针对多工况条件下的诊断准确率优于另外三种方法。

图 8 各方法精确度对比柱状图

图 9 为四种方法迭代效率对比。从图中可以看出, 所提方法在 300 次迭代时准确率已趋于平缓, 其他三种方法在 700 次迭代或更多次迭代后才逐渐收敛, 证明所提方法形成的函数收敛能力较强。

图 9 测试 1 迭代效率对比

3.3 测试 2

本例测试数据分为 A→B、B→C、C→A 三种诊断模式, 负载、转速、损伤尺寸三者同时变化, 以检验模型的适应能力, 参数设置与测试 1 相同。

表 8 ISSDBN 数据 MMD 值

MMD	A	B	C
A	0	0.157 4	0.179 2
B	0.157 4	0	0.152 3
C	0.179 2	0.152 3	0

表 9 ISSDBN 测试 2 准确率分布 %

准确率	A	B	C
A	99.21	89.60	88.15
B	89.04	98.91	89.89
C	86.97	90.15	97.48

表 8 为试验过程中样本进行 MMD 筛选后对应样本的数值。表 9 为 ISSDBN 对应的测试 2 准确率分布, 与表 8 中的 MMD 值相互对应。结合表 8 与表 9 可以看出: 随着 MMD 值的上升, 准确率随之

下降, 负载相差越大, MMD 值越高, 同时准确率越低。在本模型的试验中, 当两域 MMD 值超过 0.5 时, 准确率大幅度下降, 因此可根据不同需求在 (0,0.5) 范围内选取阈值。

表 10 为不同方法的故障识别准确率。可以看出, 在负载、转速、损伤尺寸同时变化时, 四种方法的准确度均大幅度下降, 但本文所提模型的准确度仍然高于其他三种方法。另外, 在测试 C→A 时的精度要低于 A→B、B→C 两种情况, 这表明训练数据与测试数据负载相差越大模型精度越低。

表 10 不同方法在各数据集的识别准确率对比 %

数据集	A→B	B→C	C→A
PCA	17.60	17.22	15.36
SAE	43.25	40.72	36.16
DBN	65.91	67.54	59.87
ISSDBN	89.32	90.02	87.56

图 10 为各方法精确度对比的柱状图, 可更直观对比各种方法在负载、转速、损失尺寸均变化时的识别精度。由图 10 可以看出所提方法故障识别准确率高于其他三种方法。

图 10 各方法精确度对比柱状图

图 11 为四种方法的迭代效率。由图 11 可知, ISSDBN 在迭代 500 次左右趋于稳定, 而其他三种方法稳定时的迭代次数均高于 600 次, 对比第一组测试的结果可以发现 ISSDBN 的性能相对下降, 但相较其他三种方法下降幅度最小, 并且在准确度方面高于对比的三种方法, 在训练后形成的网络函数收敛能力表现出优势。

图 11 测试 2 迭代效率对比

4 结论

利用改进后的半监督深度信念网络, 解决了多工况条件下难以有效识别故障的问题, 得出以下结论。

(1) 提出的模型减少了梯度消失的现象, 在与普通深度信念网络对比时有更小的迭代误差与更快的收敛速度, 提高了模型整体的准确度;

(2) 模型使用的标签数据较少, 并有效利用了无标签数据辅助提升诊断准确度, 对于目前轴承故障标记数据较少的困境提出了一个新的解决方案, 节约了大量用于标记数据的成本;

(3) 模型对转速和损伤尺寸均有变化的轴承故障诊断准确率在 97% 以上, 并对转速、损伤尺寸与负载三者均变化的轴承诊断准确率也达到了 87% 以上, 证明提出的模型具备了一定程度的迁移学习能力, 能够在复杂多变的工况约束下进行有效的故障诊断。

后续研究将针对更多工况搭配模式开展, 以及各工况变化对方法稳定性的影响。

参 考 文 献

- [1] YAN Xiaonan, JIA Minping, ZHANG Wan, et al. Fault diagnosis of rolling element bearing using a new optimal scale morphology analysis method[J]. Isa Transactions, 2018, 73: 165-180.
- [2] JIA Feng, LEI Yaguo, GUO Liang, et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines[J]. Neurocomputing, 2018, 272(10): 619-628.
- [3] LU Siliang, HE Qingbo, ZHAO Jiwen, et al. Bearing fault diagnosis of a permanent magnet synchronous motor via a fast and online order analysis method in an embedded system[J]. Mechanical Systems and Signal Processing, 2017, 113(12): 36-49.
- [4] 张小龙, 张氢, 秦仙蓉, 等. 基于 ITD-形态滤波和 Teager 能量谱的轴承故障诊断[J]. 仪器仪表学报, 2016, 37(4): 788-795.
ZHANG Xiaolong, ZHANG Qing, QIN Xianrong, et al. Fault diagnosis method for rolling bearing based on ITD-morphological filter and Teager energy spectrum[J]. Chinese Journal of Scientific Instrument, 2016, 37(4): 788-795.
- [5] 李宏坤, 杨蕊, 任远杰, 等. 利用粒子滤波与谱峭度的

- 滚动轴承故障诊断[J]. 机械工程学报, 2017, 53(3): 63-72.
- LI Hongkun, YANG Rui, REN Yuanjie, et al. Rolling element bearing diagnosis using particle filter and kurtogram[J]. Journal of Mechanical Engineering, 2017, 53(3): 63-72.
- [6] 李志农, 张芬, 何旭平. 基于小波-KCCA 的非线性欠定盲分离方法研究[J]. 仪器仪表学报, 2014, 35(3): 601-606.
- LI Zhinong, ZHANG Fen, HE Xuping. Study on underdetermined blind source separation method of nonlinear mixture based on wavelet and Kernel Canonical Correlation Analysis[J]. Chinese Journal of Scientific Instrument, 2014, 35(3): 601-606.
- [7] MCCULLOCH W, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. Bulletin Mathematical Biology, 1943, 5(4): 115-133.
- [8] SAMUEL A. Some studies in machine learning using the game of checkers. II: recent progress[J]. Ibm Journal of Research and Development, 1967, 11(6): 601-617.
- [9] RUMELHART D, HINTON G, WILLIAMS R, et al. Learning representations by back-propagating errors[J]. Nature, 1988, 323(6088): 696-699.
- [10] FISCHER A, IGEL C. An introduction to restricted Boltzmann machines[C]//Iberoamerican Congress on Pattern Recognition, 2012: 14-36.
- [11] LÉCUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[C]// Proc. IEEE, 1998, 86(11): 2278-2324.
- [12] 杨宇, 罗鹏, 甘磊, 等. SADB N 及其在滚动轴承故障分类识别中的应用[J]. 振动与冲击, 2019, 38(15): 11-16, 26.
- YANG Yu, LUO Peng, GAN Lei, et al. SADB N and its application in rolling bearing fault identification and classification[J]. Journal of Vibration and Shock, 2019, 38(15): 11-16, 26.
- [13] 沈长青, 汤盛浩, 江星星, 等. 独立自适应学习率优化深度信念网络在轴承故障诊断中的应用研究[J]. 机械工程学报, 2019, 55(7): 81-88.
- SHEN Changqing, TANG Shenghao, JIANG Xingxing, et al. Bearings fault diagnosis based on improved deep belief network by self-individual adaptive learning rate[J]. Journal of Mechanical Engineering, 2019, 55(7): 81-88.
- [14] 赵光权, 葛强强, 刘小勇, 等. 基于 DBN 的故障特征提取及诊断方法研究[J]. 仪器仪表学报, 2016, 37(9): 1946-1953.
- ZHAO Guangquan, GE Qiangqiang, LIU Xiaoyong, et al. Fault feature extraction and diagnosis method based on deep belief network[J]. Chinese Journal of Scientific Instrument, 2016, 37(9): 1946-1953.
- [15] CHEN Zhiqiang, DENG Shengcai, CHEN Xudong, et al. Deep neural networks-based rolling bearing fault diagnosis[J]. Microelectronics Reliability, 2017, 75(75): 327-333.
- [16] SHAO Haidong, JIANG Hongkai, WANG Fuan, et al. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet[J]. Isa Transactions, 2017, 69: 187-201.
- [17] SHAO Haidong, JIANG Hongkai, ZHANG Haizhou, et al. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing[J]. Mechanical Systems and Signal Processing, 2018, 100: 743-765.
- [18] 雷亚国, 杨彬, 杜兆钧, 等. 大数据下机械装备故障的深度迁移诊断方法[J]. 机械工程学报, 2019, 55(7): 1-8.
- LEI Yaguo, YANG Bin, DU Zhaojun, et al. Deep transfer diagnosis method for machinery in big data era [J]. Journal of Mechanical Engineering, 2019, 55(7): 1-8.
- [19] ZHANG Zhongwei, CHEN Huaihai, LI Shunming, et al. A novel geodesic flow kernel based domain adaptation approach for intelligent fault diagnosis under varying working condition[J]. Neurocomputing, 2020, 376: 54-64.
- [20] LI Xiang, JIA Xiaodong, ZHANG Wei, et al. Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation[J]. Neurocomputing, 2020, 383: 235-247.
- [21] AN Zenghui, LI Shunming, WANG Jinrui, et al. Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method[J]. Neurocomputing, 2019, 352: 42-53.
- [22] TONG Zhe, LI Wei, ZHANG Bo, et al. Bearing fault diagnosis based on domain adaptation using transferable features under different working conditions[J]. Shock and Vibration, 2018, 2018(2018): 1-12.
- [23] WANG Gongming, QIAO Junfei, LI Xiaoli, et al. Improved classification with semi-supervised deep belief network[J]. IFAC-Papers on Line, 2017, 50(1): 4174-4179.
- [24] GRETTON A, BORGWARDT K, RASCH M, et al. A kernel method for the two-sample-problem[C]//Neural Information Processing Systems, 2006: 513-520.

- [25] WANG Mei, DENG Weihong. Deep face recognition with clustering based domain adaptation[J]. Neurocomputing, 2020, 393: 1-14.
- [26] QIN Yi, WANG Xin, ZOU Jingqiang. The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines[J]. IEEE Transactions on Industrial Electronics, 2019, 66(5): 3814-3824.
- [27] 张云强, 张培林, 王怀光, 等. 基于双时域微弱故障特征增强的轴承早期故障智能识别[J]. 机械工程学报, 2016, 52(21): 96-103.
- ZHANG Yunqiang, ZHANG Peilin, WANG Huaiguang, et al. Rolling bearing early fault intelligence recognition based on weak fault feature enhancement in time-time domain[J]. Journal of Mechanical Engineering, 2016, 52(21): 96-103.
- [28] LOPARO K. Bearing data center, case western reserve university[EB/OL]. 2013 [2020-05-25]. <http://www.eecs.case.edu/laboratory/bearing>.
- [29] RANZATO M, BOUREAU Y, LECUN Y, et al. Sparse feature learning for deep belief networks[C]//Neural Information Processing Systems, 2007: 1185-1192.
- [30] SALAKHUTDINOV R, HINTON G. Deep boltzmann machines[C]//International Conference on Artificial Intelligence and Statistics, 2009: 448-455.
- [31] HINTON G, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks[J]. Science 2006, 313(5786): 504-507.
- [32] HINTON G, OSINDERO S, TEH Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2014, 18(7): 1527 - 1554.
-
- 作者简介: 叶楠, 男, 1986 年出生, 博士, 讲师。主要研究方向为结构动态力学行为与工业机器人可靠性。
E-mail: yenan@hebut.edu.cn
- 常佩泽, 男, 1996 年出生, 硕士研究生。主要研究方向为工业机器人故障诊断。
E-mail: cpz18142200240@163.com
- 张露予, 女, 1989 年出生, 博士, 讲师。主要研究方向为工业机器人故障诊断及可靠性预测。
E-mail: lyzhang@hebut.edu.cn
- 王嘉(通信作者), 女, 1988 年出生, 博士, 副教授。主要研究方向为退化建模及可靠性评估。
E-mail: jwangno1@163.com