中国科学: 技术科学 2024 年 第 54 卷 第 6 期: 1105 ~ 1119
SCIENTIA SINICA Technologica techcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文

具有隐私保护功能的半监督小样本齿轮箱故障诊断

何心[1], 段亚穷[2,4], 王子栋[3], 张永[2*]

1. 华中科技大学机械科学与工程学院, 武汉 430074;
2. 武汉科技大学信息科学与工程学院, 武汉 430081;
3. Department of Computer Science, Brunel University London, London UB83PH, United Kingdom;
4. 元始智能科技有限公司, 南通 226000
* E-mail: zhangyong77@wust.edu.cn

收稿日期: 2023-11-02; 接受日期: 2024-01-31; 网络版发表日期: 2024-06-12
国家自然科学基金项目(批准号: 62273264)和广西重点研发计划项目(编号: 桂科AB22035023)资助

摘要 齿轮箱作为机械传动系统的关键组成部分, 面对有限标记数据和数据隐私保护的需求, 传统故障诊断方法存在数据泄露和模型精度不高等问题. 因此, 本研究提出了一种基于联邦学习的隐私保护框架, 采用半监督原型网络与对比学习相结合的小样本齿轮箱故障诊断方法. 首先, 构建了DeceFL联邦学习框架, 在每个客户端上使用有限数量的标记样本构建正负样本对. 同时, 采用对比学习的预训练方法, 为自动编码器提供了初始化参数. 随后, 将自动编码器作为原型网络的特征映射函数, 使用有限数量的标记样本计算类别原型. 最后, 通过原型细化方法对原型进行微调, 减少异常数据的干扰, 获得更加稳定和准确的原型. 经过实际齿轮箱数据的验证, 结果表明, 本研究提出的隐私保护框架下的半监督小样本故障诊断方法能够仅使用极少数量的样本即实现更为出色的故障识别精度. 这一研究为应对现实工业应用中的数据挑战提供了创新性的方法, 推动齿轮箱故障诊断领域的进一步研究.

关键词 齿轮箱, 故障诊断, 隐私保护, 小样本, 半监督学习

1 引言

齿轮箱作为传动装置的关键部件[1,2], 承载着改变动力方向、速度和扭矩的重要功能, 被广泛应用于汽车[3,4]、工业机械[5,6]、风力涡轮机[7,8]等领域. 作为传动系统故障频率最高的设备, 齿轮箱一旦发生故障, 传动装置将被迫停机甚至导致安全事故. 因此, 齿轮箱的故障诊断技术是保持机械系统的可靠性、稳定性和安全性, 延长设备寿命, 提高效率, 并降低维护成本的关键技术. 同时随着数据信息保密意识的不断加强[9,10], 国防军工企业以及工业界, 纷纷加强了对数据隐私的保护, 强调数据不离开工厂的原则, 造成了数据孤岛问题的出现, 导致高质量多样性数据的需求紧张.

为了解决以上问题, 小样本故障诊断[11,12]与隐私保护[13,14]受到了越来越多研究者的重视, 研究人员致力于在有限数据情况下实现精准故障诊断. 小样本算法分为基于对比学习和基于度量学习, 针对这两种思想分别提出了多种不同思路的算法. Huo等人[15]提出

引用格式: 何心, 段亚穷, 王子栋, 等. 具有隐私保护功能的半监督小样本齿轮箱故障诊断. 中国科学: 技术科学, 2024, 54: 1105–1119
He X, Duan Y Q, Wang Z D, et al. Semi-supervised small-sample gearbox fault diagnosis with privacy protection (in Chinese). Sci Sin Tech, 2024, 54: 1105–1119, doi: 10.1360/SST-2023-0351

© 2024 《中国科学》杂志社

了一种基于原型网络的多任务模型，在训练过程中综合考虑相对距离损失，减小类内距离、增大类间距离，显著提升诊断模型性能. 但当数据集噪声较大时, 容易导致原型偏离从而降低准确性. Peng等人[16]提出了一种基于SoftBrown采样的负增强监督对比学习的故障诊断算法, 能够通过采样模拟未知的故障, 增强了不平衡故障检测准确率. Zhang等人[17]提出了一种领域差异引导的对比学习算法，该方法通过在不同的工作条件下构造样本对，学习领域不变特征，提取故障差异特征，利用样本相似性进行对比学习，提高了诊断精度. 相同故障之间的分布并不是完全一致，当同一故障分布不同时很难捕捉领域不变特征. Chen等人[18]采用故障仿真模型中蕴含的丰富机理知识辅助实测数据，解决了目标域故障数据少标签甚至无标签场景的跨域诊断任务.

为了解决隐私保护和多用户建模问题，近年联邦学习方法被提出，通过模型参数的信息，隐式地对多用户数据进行信息融合，且本地数据不被其他用户和中央服务器所访问，形成了隐私保护的协同训练. Shao等人[19]提出了一种基于差分加权联邦平均的联邦迁移学习框架，协同训练诊断模型. Ma等人[20]在联邦学习中采用扩展Kalman作为模型更新算法来解决流量预测中潜在的攻击. 然而, 这些方法在隐私保护过程中严重依赖于中央服务器. 此外, 由于数据加密而导致的全局模型精度下降以及计算和通信开销的增加等问题也不容忽视. Warnat-Herresthal[21]提出了一种群学习方法，该方法在每次迭代中根据特定的规则选择一个变化的领导者, 从而实现去中心化. 然而, 该网络只是转移了中央服务器, 而没有实现完全的去中心化. 同时, 该框架没有解决网络的异构问题. 为了解决上述问题，本文采用去中心化联邦学习[22]DeceFL，利用客户端及其邻居之间的本地信息传输，代表了一种完全分散的学习框架. 它进一步证明，每个客户端都能以零性能差距达到全局最小值，并实现相同的收敛性. 它采用时变拓扑实现客户端之间的通信，消除了中心服务器，减少了传统联邦学习所需的通信带宽，有效地解决了集中联邦学习问题.

综上所述，近些年围绕隐私保护和齿轮箱小样本故障诊断的方法都取得了很大的发展，极大地丰富了齿轮箱故障处理技术, 拓展了应用场景. 然而, 现有的方法仍存在一些局限性. 现有的小样本故障诊断的研究多集中于如何建立度量，以增加类间距离，减少类内距离，从而忽略了低维特征表示的重要性，好的低维映射能够提供丰富的样本语义信息，使得度量更加准确. 受到上述研究的启发，本文提出了一种基于联邦学习框架的半监督小样本故障诊断方法，称为SSMDAE-PN. 该方法通过建立去中心化的联邦学习解决了数据远离工厂可能导致的隐私泄露问题. 此外, 本文提出了一种改进型原型网络，以提高原型的计算准确性，从而提高了诊断的准确性. 该模型包括两个主要阶段: 首先, 进行对比学习预训练阶段, 使用正负样本对自动编码器进行训练，以构建一个预训练参数模型. 然后，利用预训练参数的自动编码器来计算有标签数据的类别原型，并通过估算无标签数据的类别贡献和样本权重来进一步细化特征原型. 这一方法增强了模型提取差异特征的能力，同时通过对比学习预训练降低了过拟合的风险. 通过样本权重和类别贡献的优化, 能够减轻异常数据对原型的影响. 最后, 在东南大学齿轮箱数据集上进行了验证，并将其性能与现有方法进行了比较. 结果表明，这一方法的性能明显优于现有常见方法(图1).

本文的主要贡献如下: (1) 提出了一种混合注意机制的自编码器特征提取器，提高特征提取和选择能力; (2) 提出了一种不需要额外扩展的预训练框架, 以应对有限标记数据下的模型过拟合问题; (3) 为了降低异常数据对原型的影响并提高原型的准确性，提出了一种半监督原型网络，其具备联合样本权重和类别贡献计算的特性; (4) 为了应对齿轮箱模型训练中的隐私泄露问题，构建了DeceFL框架. 该框架的设计目标在于增强数据隐私保护、有效节省通信资源，并克服齿轮箱数据孤岛问题.

2 理论基础

2.1 原型网络

设具有N个标记样本的支持集为S=$(x_1, y_1,), \ldots, (x_N, y_N)$，其中$x_i \in \mathbb{R}^D$是一个维度为$D$的特性向量，$y_i \in 1, \ldots, K$是样本的对应标签，$S_k$代表的是有$k$个标签的集合(图2).

原型网络[23]通过计算一个M维度的特征表示$c_k \in \mathbb{R}^M$或者学习一个每一类的特征变换函数

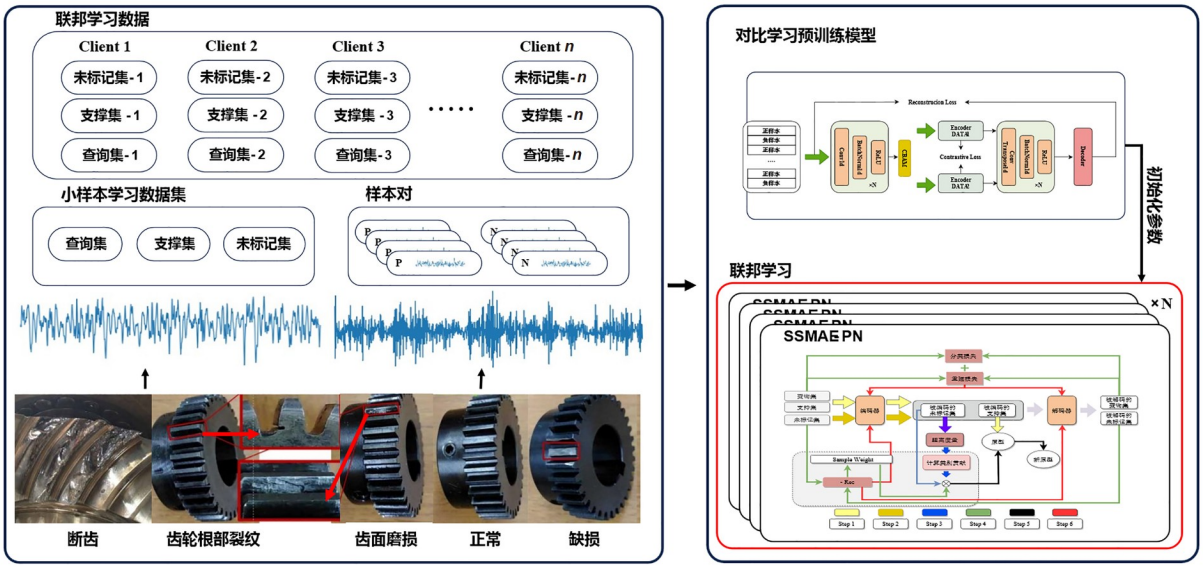


图 1　(网络版彩图)诊断总体架构图

Figure 1　(Color online) Diagnostic general architecture diagram.

$f_0:\mathbb{R}^D\to\mathbb{R}^M$, 其中$\theta$为可学习参数. 每个原型是属于其类的特征嵌入支持点的平均向量:

$$c_k=\frac{1}{|S_k|}\sum_{(x_i,y_i)}f_\theta(x_i). \tag{1}$$

基于距离函数$d:\mathbb{R}^M\times\mathbb{R}^M\to[0,+\infty)$, 原型网络生成查询点$x$的类别分布, 该分布基于特征嵌入空间中原型之间的距离进行Softmax操作.

$$p_\theta(y=k\mid x)=\frac{\exp\left(-d\left(f_\theta(x),c_k\right)\right)}{\sum_{k'}\exp\left(-d\left(f_\theta(x),c_{k'}\right)\right)}. \tag{2}$$

采用随机梯度下降和最小负对数概率函数进行模型训练, 训练集的形成过程: 首先从训练数据中随机选择一个类别的子集, 然后在该类别子集中随机选择一部分作为支持集, 将其余的部分作为查询点.

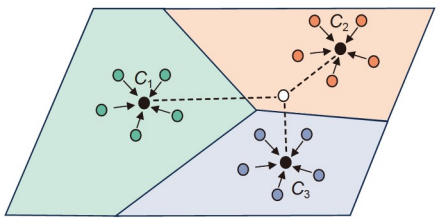


图 2　(网络版彩图)原型网络

Figure 2　(Color online) Prototype network.

$$\mathcal{L}\left(S_k,S_Q,V,c_k\right)=\frac{1}{N_C N_Q}\Big[d\left(f_\theta(x),c_k\right)+\log\sum_{k'}\exp\left(-d\left(f_\theta(x),c_k\right)\right)\Big], \tag{3}$$

其中, N_C为训练集中的类别数目, N_S为支持集中每个类别的个数, N_Q为查询集中每个类别的个数, 训练集为$D=\{(x_1,y_1),\ldots,(x_N,y_N)\}$, 则$D_{V_k}$为$D$的子集, $S_k=\mathrm{Sample}\left(D_{V_k},N_S\right)$, $Q_k=\mathrm{Sample}\left(D_{V_k}/S_k,N_Q\right)$, Sample(*S*,*N*)表示从集合$S$中均匀随机选取的不替换的$N$个元素的集合.

在小样本分类任务中, 分类器必须推广到训练集中没有出现的新类且每个新类只有少量样本. 原型网络学习一个度量空间, 在这个度量空间中, 可通过计算到每个类的原型表示的距离来进行分类. 与其他方法相比, 原型网络反映了一种更简单的归纳偏差, 在这个有限的数据区域是有益的, 并取得了很好的效果.

2.2　混合注意力模型

注意力机制是一种广泛应用于机器学习和人工智能领域的技术, 用于模拟人类的视觉和感知能力. 它允许模型在处理数据时有选择地关注和加权不同部分的信息, 从而有效地学习复杂的输入数据. 混合注意力[24](Convolutional block attention module, CBAM)是

一种用于前馈卷积神经网络的简单而有效的注意力模块. 给定一个中间特征图, CBAM模块会沿着两个独立的维度(通道和空间)依次推断注意力图, 然后将注意力图与输入特征图相乘以进行自适应特征优化. 其原理如图3所示.

通道注意力模块用于增强模型对不同通道或特征图上的特征选择能力. 它通过计算每个通道的重要性权重, 使网络更关注包含显著信息的通道, 同时抑制不太重要的通道. 有助于提高特征的多样性和可区分性. 空间注意力模块用于增强对不同空间位置上的特征选择能力. 它通过计算不同位置的重要性权重, 使模型能够更好地关注序列中的相关区域, 同时抑制无关区域, 有助于提高模型的位置感知能力.

$$Att_c = \sigma(\mathrm{MLP}(\mathrm{Avgpool}(F)) + \mathrm{MLP}(\mathrm{Maxpool}(F))), \tag{4}$$

$$Att_s = \sigma(\mathrm{Conv}([\mathrm{AvgPool}(\mathrm{F}); \mathrm{MaxPool}(\mathrm{F})])). \tag{5}$$

3 半监督故障诊断方法

3.1 基于对比学习的预训练模型

对比学习的预训练模型流程如图4所示, 首先从样本数据中获取预训练数据集, 随后将该数据集输入到经过改进的自动编码器进行预训练.

改进后的自动编码器由两个关键部分构成: 自动编码器[24,25]和注意力模块[6,26]. 为了增强其特征选择能力, 在隐含层输出之前引入卷积注意力模块. CBAM包括两个重要部分, 即通道注意力和空间注意力. 通道注意力模块主要用于捕获特征图通道之间的相互依赖关系, 而空间注意力模块则主要用于捕获不同位置之间的空间依赖关系. 其工作原理涉及逐渐计算通道维度和空间维度上的注意图, 随后将这些注意图与输入特征图相乘, 以进行自适应特征提取[23,27]. 以下是包含CBAM的自动编码器的计算过程描述:

$$f_{\mathrm{Encoder}}(x) = \sigma(BN(\mathrm{Conv1}d(x))), \tag{6}$$

$$f_{\mathrm{Decoder}}(x) = \sigma(BN(\mathrm{ConvTranspose1}d(x))), \tag{7}$$

$$F(x_s) = f_{\mathrm{CBAM}}(f_{\mathrm{Encoder}}(x_s)), \tag{8}$$

$$G(x_e) = f_{\mathrm{Decoder}}(x_e), \tag{9}$$

其中, x_s为输入原始样本, σ是激活函数. Conv1d为一维卷积模块, BN为批归一化层. 编码器的输出表示为x_e, $G(\cdot)$表示译码映射函数, ConvTranspose1d表示一维反卷积模块, 解码器输出表示为x_d. 自动编码器的重构损失计算过程如下:

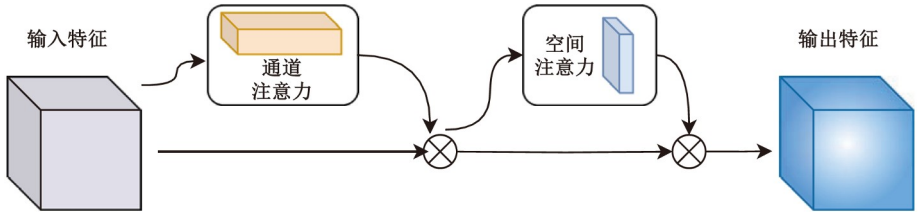


图 3 (网络版彩图)混合注意力模型

Figure 3 (Color online) Mixed attention module.

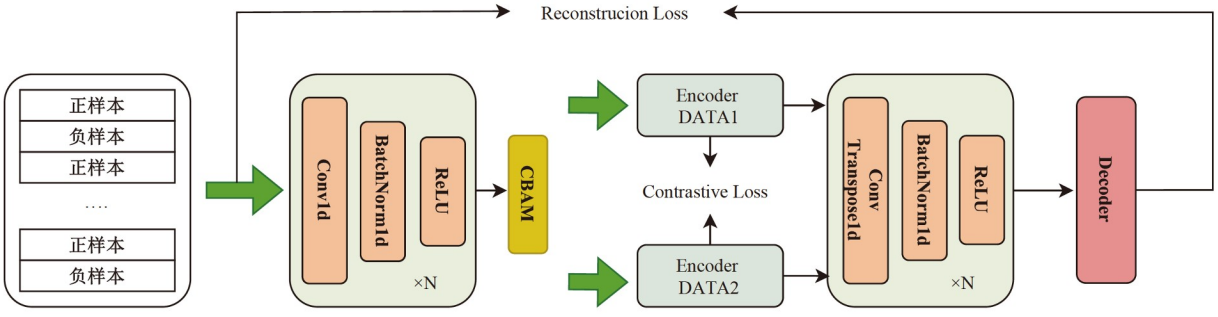


图 4 (网络版彩图)对比学习预训练模型

Figure 4 (Color online) Comparison learning pre-training model.

$$L\left(x_i \mid G(F(x_i))\right)=\frac{1}{n}\sum_{i=1}^{N}\left(x_i-G(F(x_i))\right)^2, \tag{10}$$

其中，$L\left(x_i \mid G(F(x_i))\right)$用于度量模型的性能，它评估了模型输入$\mathcal{X}_s$经过编码器和解码器处理后是否能够正确的重构，此过程有助于确保自动编码器能够有效地学习有用的特征并保持数据的信息.

预训练[28~31]方法是将构造好的样本对$\{x_i,x_j\}$输入到自动编码器中，通过编码器输出一对嵌入特征$\{z_i,z_j\}$. 有助于通过自动编码器将特征数据映射为低维度高浓度的特征表示向量，为诊断任务提供更好的初始化和特征表达.

嵌入特征的计算公式为

$$\{z_i,z_j\}=F\left(\{x_i,x_j\}\right), \tag{11}$$

其中，F代表自动编码器将高维特征空间映射为低维特征空间表示. 特征在经过编码器后，欧式距离能够快速准确地描述样本在低维特征中的距离，故采用欧式距离构建度量函数，用于度量两个嵌入特征之间的相似性. 其目标是最小化正样本对之间的距离，同时最大化负样本对之间的距离，这是通过最小化嵌入特征的欧氏距离来实现的. 正、负样本对的损失函数描述如下:

$$\mathrm{Loss}^{+}\left(X_{ij}\right)=\left\{\mathrm{d}\left(F(x_i),F(x_j)\right)\right\}^2, \tag{12}$$

$$\mathrm{Loss}^{-}\left(X_{ij}\right)=\max\left(0,-\mathrm{d}\left(F(x_i),F(x_j)\right)\right), \tag{13}$$

其中，Loss^{+}用于度量同一类别或相似样本对之间的嵌入特征距离，希望最小化这些距离，以确保相似的样本在嵌入空间中更加接近. Loss^{-}用于度量不同类别或不相似样本对之间的嵌入特征距离. 希望最大化这些距离，以确保不相似的样本在嵌入空间中距离更远.

综合公式(12)和(13)，对比学习预训练最终损失函数记为L_{ctr}，其数学表达如下:

$$L_{ctr}=\sum_{i=0}^{N_+}\mathrm{Loss}^{+}(X_i)+\sum_{j=0}^{N_-}\mathrm{Loss}^{-}\left(X_j\right), \tag{14}$$

其中，N_+为正样本对数，$N_$为负样本对数.

3.2　改进的原型网络

在训练过程中，每个任务包括支持集D_s、查询集D_q和无标记集D_u. 当支持集D_s中的n个样本无法充分代表每个类别的分布时，所计算的原型无法准确反映类别中心. 因此，在训练期间引入了未标记的数据，以进一步计算类别中心. 然而，未标记数据集中存在异常样本，过于依赖所有数据可能导致原型产生偏差，并加大了正常和故障数据的分类误差. 为了解决原型计算过程中异常样本对未标记数据的干扰，提出了一种改进原型网络方法(SSMAE-PN)，它利用自编码器的重构误差来估计样本权重，那些与标记数据有较大差异的样本将被分配较低的权重.

SSMAE-PN使用经过改进的自动编码器作为特征映射函数，借助原型网络[32]进行原型计算，以及距离度量模块进行故障识别，其结构如图5所示.

原型网络的第一步是使用编码器从输入数据中提取嵌入特征，然后计算每个类别特征的平均值，从而得到各个类别的原型. 这一步骤的目标是为每个类别生成原型，这些原型将在后续的故障识别步骤中用于计算与输入数据之间的距离，以确定数据属于哪个类别. 这有助于更准确地识别故障并进行分类. 这一过程对应图5中的Step 1，其原型计算公式如下:

$$p_c=\frac{1}{N_s^c}\sum_{i=0}^{N_s^c}F(x_s), \tag{15}$$

其中，C表示故障类别，N_s^c表示支持集中故障类别c的样本个数. 未标记样本对各分类贡献通过计算未标记样本的嵌入特征和解码器输出向量(图5中的Step 2)，然后利用距离度量函数$d(\bullet)$，计算嵌入特征与原型之间的距离. 最后，利用Softmax对$C(x_u,C)$进行归一化确定贡献度，并根据各类别的相对概率对其赋值(图5 Step 3). 这一过程有助于确定未标记样本对每个类别的贡献度，并根据相对的概率值为其分配权重. 这些权重可以在故障识别和分类中起到关键作用，帮助模型更准确地确定未标记样本的类别，其计算方式可以表示为

$$C_{x_{u,c}}=\frac{\exp\left(-d(F(x_{u,i}),p_c)\right)}{\sum_{j=0}^{N_c}\exp\left(-d(F(x_{u,i}),p_j)\right)}. \tag{16}$$

对于无标记样本权值，首先利用自动编码器计算无标记样本的重构误差. 这个误差反映了异常程度，即样本偏离正常模式的程度，误差越大表示异常程度

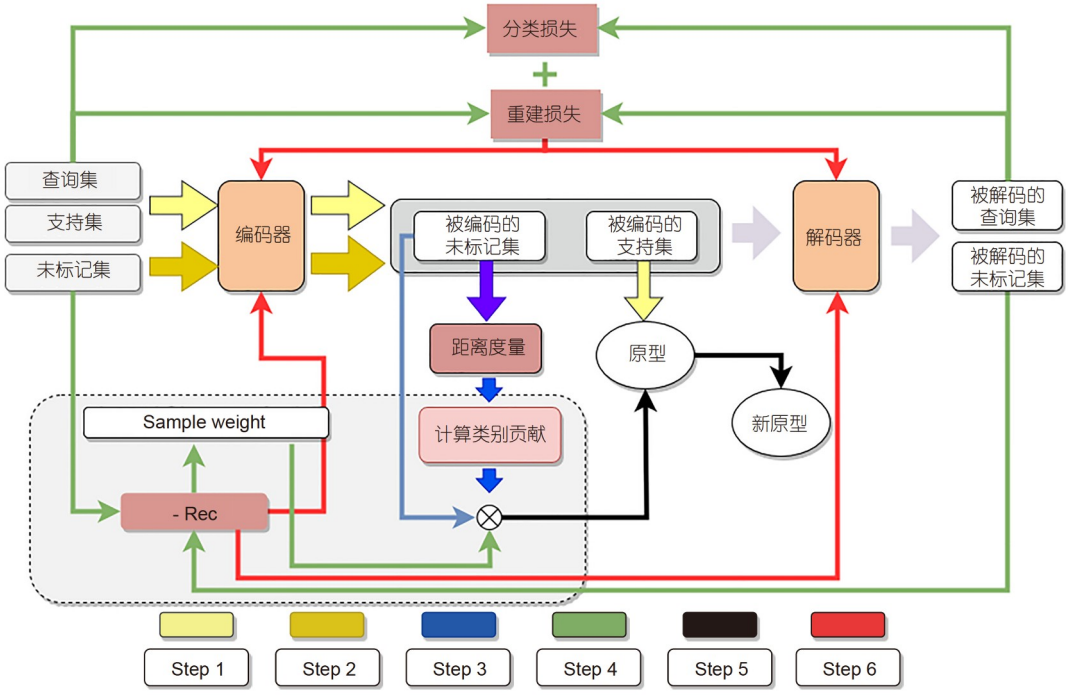


图 5 (网络版彩图) SSMAE-PN网络结构

Figure 5 (Color online) SSMAE-PN network.

越高(见图5中的Step 4). 通常情况下, 权重与重构误差成反比, 这意味着重构误差越大的样本被分配的权重越小. 这一步骤的目标是根据每个无标记样本的异常程度确定其权重, 从而在后续的计算中减小异常数据对原型的影响. 从而可以提高模型的稳定性和性能, 其计算公式如下:

$$L_{rc}(x_u) = MSE\left(x_u, G(F(x_u))\right), \tag{17}$$

$$w_{x_u} = \exp\left(-\text{loss}_{rc}(x_u)\right). \tag{18}$$

原型的细化过程是将未标记样本的权重与类别贡献相结合, 记为$w(x_u, p_c)$, 这种组合同时考虑了样本的类别贡献和权重. 这种细化过程有助于进一步优化原型, 并降低异常样本对模型的不良影响, 从而提高故障识别的准确性, 计算公式如下:

$$w_{x_u,c} = C_{x_u,c} * w_{x_u}. \tag{19}$$

然后通过查询集的重建损失和分类损失对带混合注意力的自编码器进行参数更新(图5中的Step 6), 可将其等同于预训练模型中的微调. 通过进一步更新模型参数, 增强模型对于数据的拟合能力和信息提取能力.

利用未标记数据的嵌入特征对类别原型进行微调, 然后利用原型计算提供的样本x_s属于c类的概率, 计算公式如下:

$$p_c^{\text{new}} = \frac{\sum_{j=0}^{N_s} F(x_{s,j}) + \sum_{i=0}^{N_u} F(x_{u,i}) * w_{x_{u,i},c}}{\left(N_s + \sum_{i=0}^{N_u} W_{x_{u,i},c}\right)}, \tag{20}$$

$$p\left(y = c \mid x_s\right) = \frac{\exp\left(-d(F(x_s), p_c)\right)}{\sum_{j=0}^{N_c} \exp\left(-d(F(x_s), p_j)\right)}. \tag{21}$$

以上改进原型网络利用查询集计算分类损失, 通过降低负对数概率最小化分类损失, 损失函数计算公式如下:

$$L_{kl} = \sum_{i=0}^{N} -\log\left(P\left(y = y_i \mid x_i\right)\right). \tag{22}$$

基于以上分析, 改进原型网络模型(SSMAE-PN)模型参数的计算过程如算法1所示.

算法1: SSMAE-PN的训练损失计算.
RANDOMSAMPLE(D)表示从集合D中均匀随机选取的两个集合, 不替换.

输入: 无标记的数据集D_u, 标记数据集D_l, 标签y
输出: 经过训练的自动编码器和原型网.

```
for i ← to N_1 do
  for j ← to N_1 do
    if y_i == y_j then
      add(X={X = {x_i, x_j}, Y = 1) to the D_p;
    else
      add(X = {x_i, x_j}, Y = 0 to the D_p;
    end
  end
end
D_s, D_q ← Randomsample(D_l)
for epoch_1 ←1 toEpoch_1 do
  z_p ← D_p
  loss_rc ← z_p
  loss_ctr ← z_p
  使用SGD更新自编码器的参数φ
end
for epoch_2 ← 1to Epoch_2 do
  p_k ← D_s
  C_{D_u,k} ← D_u
  w_{D_u} ← D_u
  p_k^new ← C_{D_u,k} and w_{D_u}
  for epoch_3 ← 1 to Epoch_3 do
    loss_clf ← D_q
    loss_rc ← D_q
    使用SGD更新自编码器的参数φ
  end
end
```

3.3　去中心化联邦学习DeceFL

在基于联邦学习的隐私保护技术领域, 客户端在与服务器的协同合作下积极参与模型训练过程. 这些客户端之间的数据共享需求被有效化解, 从而显著降低了传统机器学习中的隐私泄露风险. 本节采用DeceFL, 该框架采用了增强的时变方案, 旨在保持模型收敛性的同时, 有效实现完全去中心化的模型训练. 图6展示了DeceFL隐私保护功能与小样本故障诊断的协同集成. 图$G=(N,E,W)$用来模拟客户端之间的通信, 其中N表示客户端集合, E表示两个客户端通信通道集合, W为邻接矩阵, 其元素W_{ij}表示客户端i和客户端j之间是否存在通信. 当$W_{ij}>0$时, 两个客户端之间存在通信, 当$W_{ij}=0$时, 两个客户端之间不存在通信. 如果$W_{ij}>0$, 则称客户端j为客户端i的邻居.

DeceFL框架成功地克服了对中央服务器的依赖, 每个客户端都在本地维护了全局权重变量ω, 对于所有客户端weight $=\{\omega_1,\omega_2,...,\omega_i\}$, DeceFL不再依赖于中央服务器来更新全局模型, 但是每个客户端通过本地收集的邻居的权重来实现全局权重的更新. 权重更新遵循以下规则:

$$w_i(t+1)=\sum_{j=1}^{I}W_{ij}w_j(t)-a_t\tilde{g}_t(w_i(t)), \tag{23}$$

其中, I是客户端数量, a_t是学习率, $g_i(W_i(t))$是客户端I上的梯度.

基于DeceFL的隐私保护框架具备多项优势. 首先, 在迭代训练过程中, 由于不再依赖于中央服务器来接收所有其他客户端的信息, 成功地避免了数据泄露的风险. 其次, 客户端之间的连接采用拓扑结构, 这意味着任何网络连接结构都能够完成最终的训练任务, 使得客户端能够在训练迭代中随时加入或退出, 从而增加了培训的灵活性. 最后, 每次迭代的模型更新仅依赖于客户端与其邻居之间的本地信息传输, 有效节省了通信资源. 本文通过建立五台客户端之间的去中心化联邦学习对模型进行训练, 设置训练轮数为100轮, 规定每五轮客户端与相邻的客户端之间进行模型聚合, 更新自己的本地模型, 重复这个步骤并不断与全网的其他客户端交换模型、聚合和更新, 直到模型达到预设的训练轮数. 训练结束后, 每个客户端都保存了一个性能近似本地模型, 用于本地部署和推理. 综上所述, 基于DeceFL的小样本故障诊断过程如算法2所示.

算法2　基于小样本的故障诊断隐私保护

输入:$n_1,n_2,...,n_I$表示客户端节点集合, I表示客户端个数. F_ψ^I表示每个客户端上的SSMAE-PN网络模型. $t\in\{1,2,3,\ldots,T\}$代表最多可迭代次数. a_t代表学习率, W为邻接矩阵.
输出: 训练完成的F_ψ^I

1 初始化每一个客户端上的F_ψ^I
2 **While** $t<T$:
3　训练 F_ψ^I
4　更新权重 $w_i(t+1)=\sum_{j=1}^{I}W_{ij}w_j(t)-a_t\tilde{g}_t(w_i(t))$
5　IF $t=T$:
6　　Return F_ψ^I
7　**else**:
8　　Return $t=t+1$

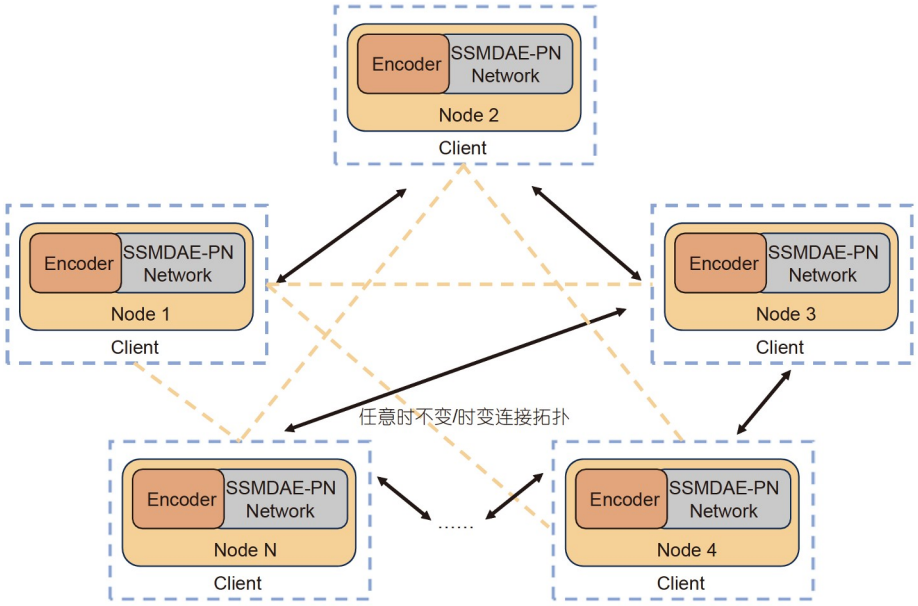


图 6 (网络版彩图) DeceFL和小样本故障诊断模型图
Figure 6 (Color online) Integration diagram of DeceFL and small-sample fault diagnosis.

4 实验分析

4.1 齿轮箱数据集

本研究采用了东南大学(SEU)齿轮箱数据集作为实验数据来源. 如图7所示, 该数据集基于动力传动系统动态模拟器(DDS)从两种不同的运行条件下获得, 这两种运行条件包括速度负载配置分别为20 Hz-0 V和30 Hz-2 V. 在不同的工况下, 故障类型可分为四类, 详见图8. 故障标签设置为0-正常、1-断齿、2-齿轮根部断裂、3-齿面磨损、4-缺损. 每个数据文件包含8列信号. 本实验中, 将每个样本选取1024个采样点, 其中每种故障类型选取1023个样本, 该数据集共包括5115个样本. 在实验中, 从20 Hz-0 V工况中挑选了1000个样本, 并保留了其中一部分作为标记数据, 其余的数据被分成一个测试集和一个未标记的数据集.

4.2 预训练数据构建

基于对比学习的预训练方法充分利用了已有的标记数据, 而无需生成额外的数据. 首先, 收集了所有已标记的样本, 将其表示为D_{labeled}, 其中$y \in [1,2,3,...,N]$. 然后, 构建样本对, 这些样本对由已标记数据构成, 记为$X_{ij}=x_i, x_j$, 其中x_i和x_j是任意的标记数据, 其标签被记为y_i和y_j, 且$y_i \neq y_j$. 如果一个样本对被标记为正样本, 则其标签为$y_i=1$, 相反, 如果被标记为负样本, 则标签为$y_i=-1$. 通过这一过程, 构建了用于预训练的数据集.

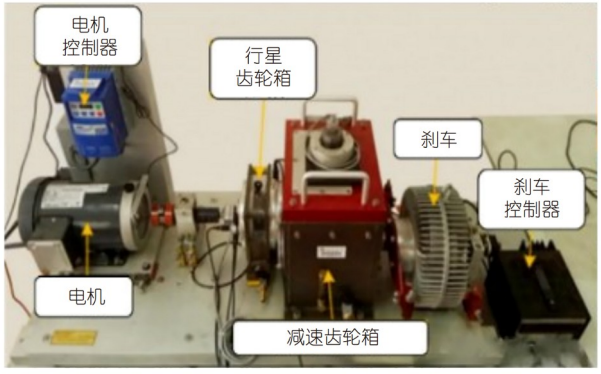


图 7 (网络版彩图)传动系动力学模拟器
Figure 7 (Color online) Drive train dynamics simulator.



图 8 (网络版彩图)齿轮故障类型
Figure 8 (Color online) Types of gear faults.

4.3　小样本学习任务数据构建

半监督学习方法利用无标记数据来提高模型的分类精度, 该方法将每个任务视为模型训练的基本单元, 每个任务由三个部分组成: 支持集(D_s)、查询集(D_q)和无标记集(D_u). 具体而言, 从已标记的数据集中随机抽取$N_c \times N_s$个样本构建支持集, 其中N_c代表抽取数据的类别个数, N_s代表每个类别中抽取的数量. 查询集则由已标记数据集中剩余的样本组成, 包含$N_c \times N_s$个样本组. 最后, 无标记集由未被标记的数据组成.

4.4　联邦学习数据构建

本研究采用了多个客户端联合进行的联邦学习, 以实现参数的更新, 同时确保了数据的隐私性. 所获得的小样本学习任务数据根据客户端的数量均匀划分为n份, 这些份数据分别用作不同客户端的训练和测试数据.

4.5　SSMAE-PN与其他模型诊断效果对比分析

为了评估SSMAE-PN的有效性, 将其与其他相关方法的分析结果进行比较. 准确率(ACC)为性能评估指标, 其计算公式如下:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}. \tag{24}$$

为了验证本文方法的优势, 将SSMAE-PN与其他四种方法进行了比较, 包括小样本学习模型: Siamese-Net[33]、MatchNet[34]、RelationNet[35]和ProtoNet[36]; 传统机器学习算法: 支持向量机, 决策树; 深度学习网络: STFT-CNN[37]、CWT-CNN[38]、CNN-LSTM[39]、CEEMD-1DCNN[40]. 所有小样本学习方法均在相同的数据集条件下进行评估, 以确保实验的公平性. 需要注意的是, 其他几种方法采用与本文提出的方法相同的特征提取器(超参数如表1所示). 传统机器学习和深度学习网络都为监督学习, 需要大量数据进行拟合, 使用1-shot或者5-shot是不公平的对比, 所以本文提供了10-shot结果和使用所有数据进行训练的效果. 诊断效果如表2所示. 表2中way代表类别数目, shot代表样本数量, ±号后的数字代表波动百分比. 实验结果表明, 在不同的实验条件下, 本文所提出的方法均优于其他四种小样本学习方法. 该方法在总体平均准确度方面表现出色, 明显超越了其他方法. 与深度学习算法和机器学习算法对比表明, 在样本数较少时, 本文方法具有明显的优势, 其准确率大幅度超越. 同时观察到, 因为每次训练进行随机采样, 深度学习算法十分的不稳定. 当增加深度学习训练样本数时, 本文方法能够达到与深度学习算法准确率相当.

表 1　超参数设置

Table 1　Hyperparameter settings

参数	取值
样本长度	1024
特征数量	8
Conv_1卷积核数量	16
Conv_1卷积核大小	64
Conv_1卷积步长	24
最大池化层	2
批归一化_1	16
Conv_2卷积核数量	32
Conv_2卷积核大小	8
Conv_2卷积步长	3
批归一化_2	32
ConvT_1卷积核数量	16
ConvT_1卷积核大小	8
ConvT_1卷积步长	3
批归一化_2	16
ConvT_2卷积核数量	8
ConvT_2卷积核大小	64
ConvT_2卷积步长	24
批归一化_2	8

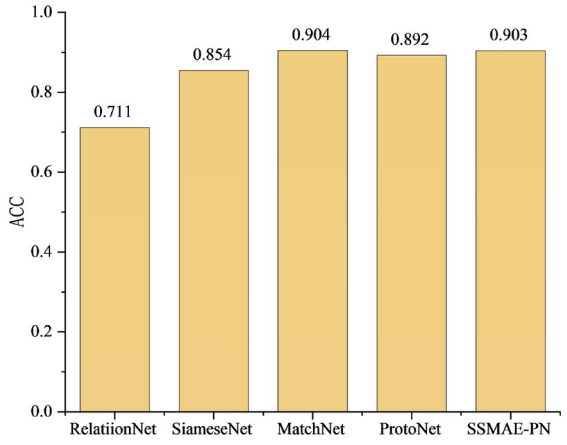


图 9　(网络版彩图)不同诊断模型效果对比柱状图

Figure 9　(Color online) Histogram comparing the effects of different diagnostic models.

此外, 图9呈现了本文方法与其他故障诊断的分析结果, 进一步证实了其优越性, 特别是在具有挑战性的少样本情境下. 模型训练期间准确性和损失的变化趋势如图10所示. 从图10(a)中可以观察到, 在最初的50次迭代中, 单一样本情况下的准确度表现出明显波动, 但随后逐渐趋于稳定, 最终稳定在0.9左右. 图10(b)揭示, 模型在大约40次迭代后达到了稳定状态. 图10(c)中的结果是基于10个样本的情况, 它显示出模型在大约20次迭代后趋于稳定, 表现出色. 值得强调的是, 对于5次和10次的实验结果, 性能更为出色, 这是因为每个类别拥有更多样本. 总之, 当每个类别包含更多样本时, 对比学习预训练有助于更全面地捕获故障信息, 从而获得更准确和更具代表性的模型表示. 此外, 随着样本数量的增加, 原型网络的收敛速度加快, 进一步提高了训练效率和性能. 图11和12具体展示了每个样本的诊断结果和诊断准确率, 图11分别介绍了类别为5, 样本个数为1、5、10情况下的每个类别的诊断准确率, 其中在样本个数为1、5时, 齿面磨损误诊率较高. 根据图12每个样本的诊断结果可知, 当样本个数为1时, 齿面磨损易被诊断为正常情况, 随着样本增加其诊断精度不断提高.

4.6 消融实验

本部分的目的在于探究预训练框架、MAE模块和原型改进算法对SSMAE-PN故障诊断性能的影响. 在实验中比较了四种不同网络结构的性能表现: (1) SSPN, 其中CNN替代了MAE模块; (2) SSAE-PN, 消除了MAE中的混合注意力模块; (3) MAE-PN, 采用原型网络但没有半监督原型优化算法; (4) SSMAE-PN, 其中特征提取模块无需经过预训练. 这四种方法的性

表 2 与其他算法进行比较

Table 2 Compares with other algorithms

Methods	5-way 1-shot	5-way 5-shot	5-way 10-shot	all
支持向量机	/	/	0.647±3.27	0.809±4.84
决策树	/	/	0.553±2.58	0.784±5.13
STFT-CNN	/	/	0.714±6.24	0.945±2.47
CNN-LSTM	/	/	0.431±8.33	0.963±0.87
CWT-CNN	/	/	0.684±10.76	0.974±1.32
CEEMD-1DCNN	/	/	0.732±5.75	**0.986**±1.21
RelatiionNet	0.711±3.31	0.762±4.09	0.771±2.01	/
SiameseNet	0.854±2.16	0.894±1.36	0.952±2.41	/
MatchNet	0.904±1.56	0.928±3.20	0.963±1.31	/
ProtoNet	0.892±2.14	0.919±1.05	0.943±0.91	/
SSMAE-PN	**0.903±1.51**	**0.951±2.01**	**0.988±1.01**	/

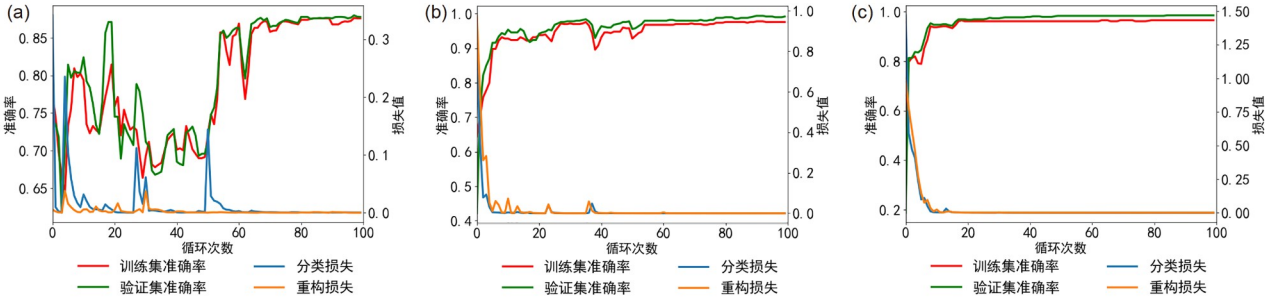


图 10 (网络版彩图)准确性和损失曲线. (a) 5-way 1-shot; (b) 5-way 5-shot; (c) 5-way 10-shot

Figure 10 (Color online) Accuracy and loss curves. (a) 5-way 1-shot; (b) 5-way 5-shot; (c) 5-way 10-shot.

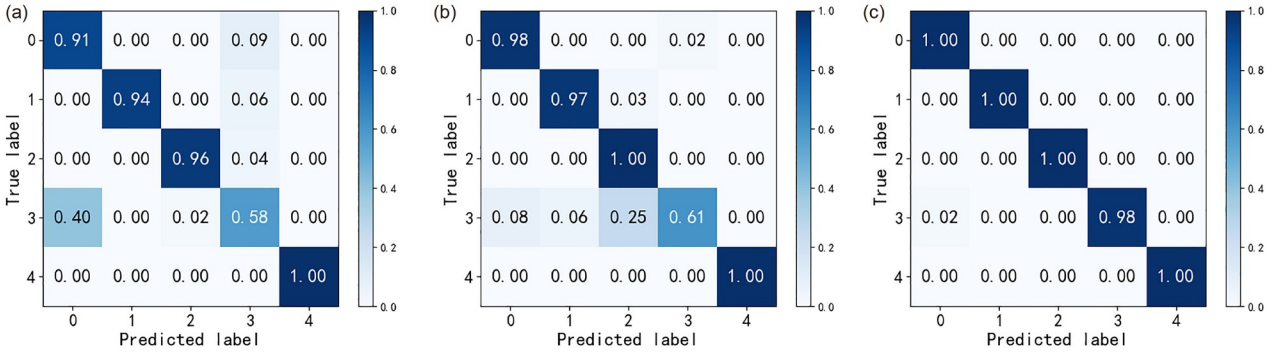


图 11　(网络版彩图)本文提出算法结果的混淆矩阵. (a) 5-way 1-shot; (b) 5-way 5-shot; (c) 5-way 10-shot

Figure 11　(Color online) The confusion matrix of the algorithm results proposed by this study. (a) 5-way 1-shot; (b) 5-way 5-shot; (c) 5-way 10-shot.

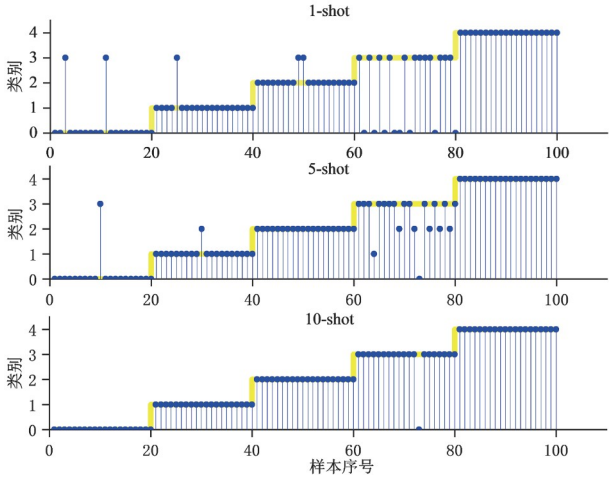


图 12　(网络版彩图)样本验证结果

Figure 12　(Color online) Sample verification results.

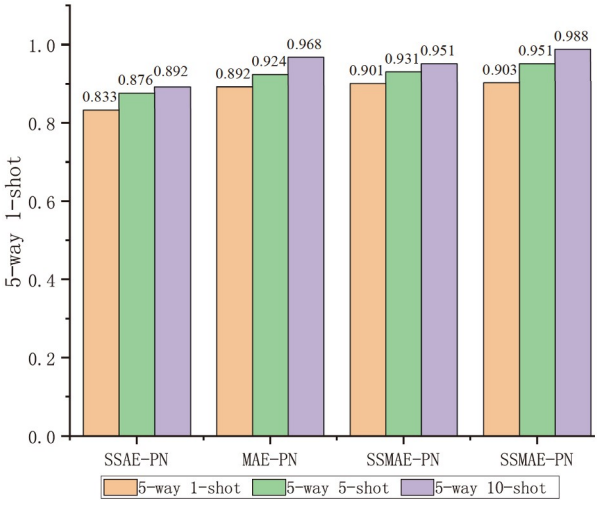


图 13　(网络版彩图)消融实验柱状图

Figure 13　(Color online) Histogram of ablation experiments.

能结果见表3和图13.

从表3中可以观察到，与未经过预训练的SSMAE-PN相比，经过预训练的SSMAE-PN(Pre-trained)的准确率提高了约2.60%. 这一结果证实了使用对比学习进行的预训练可以优化特征映射并提高模型性能，该方法的准确率比SSAE-PN提高了约9.03%，从而验证了改进后的自动编码器在特征提取方面的能力以及注意力机制在特征选择方面的有效性. 此外, 该方法的准确率也比MAE-PN高出约1.08%, 凸显了原型细化对模型性能的积极影响. 总之, 通过消融实验获得的准确性结

表 3　消融实验结果

Table 3　Ablation test results

Methods	Feature Extractor	Semi-Supervised	Pre-train	5-way1-shot	5-way 5-shot	5-way 10-shot
SSAE-PN	AE	√	√	0.833	0.876	0.892
MAE-PN	MAE	×-	√	0.892	0.924	0.968
SSMAE-PN	MAE	√	×	0.901	0.931	0.951
SSMAE-PN (pre-train)	MAE	√	√	**0.903**	**0.951**	**0.988**

果为不同模块对该方法性能的影响提供了重要依据. 该方法取得了最高的准确性, 验证了MAE模块、预训练框架和原型优化的有效性. 为了更深入地理解注意力机制对模型的贡献, 以便在视觉上进行解释, 并更好地理解模型的决策过程, 本文采用Grad-Cam来可视化模型的空间注意力. 图14中, 带有颜色的点表示在齿轮箱故障诊断中具有较大贡献的区域, 颜色深浅表示其贡献程度, 颜色越黄表示贡献度越高. 可以观察到, 第0维和第7维的点被标记得最多, 表明它们对齿轮箱故障诊断的贡献度较大.

5 总结

本文提出了具有隐私保护框架的半监督故障诊断模型, 首先提出了一种对比学习进行模型预训练方式, 为特征提取器提供初始化参数, 其次通过带有混合注意力的自动编码器, 该模块能够捕捉空间关联性和通道关联性, 增强齿轮箱故障相关特征提取效果, 同时抑制不相关特征, 并输出低维度表征信息, 最后采用度量学习进行样本故障诊断. 该方法侧重于细化类别原型, 同时利用未标注数据的样本权重和类别贡献, 有效地减轻异常数据带来的干扰, 为证明模型有效性本文在齿轮箱故障数据上进行了实验验证. 消融实验表明(以5-way 1-shot为例), 加入对比学习预训练模型诊断精度提升0.2%(90.1%→90.3%), 加入混合注意力模块模型诊断精度提升7%(83.3%→90.3%), 加入半监督网络模型诊断精度提升1.1%(90.3%→89.2%), 模型最终诊断精度为90.3%. 本文研究并验证了同种工况下的小样本故障诊断的有效性, 但在实际工程领域工况随着机械设备的磨损或外部条件的变化而产生变化, 不同时间段的数据通常呈现不同的特征分布. 未来研究将侧重于模型的跨工况的突发故障诊断任务, 以求实现跨工况零样本故障诊断, 进一步拓展其在工程领域的实际应用.

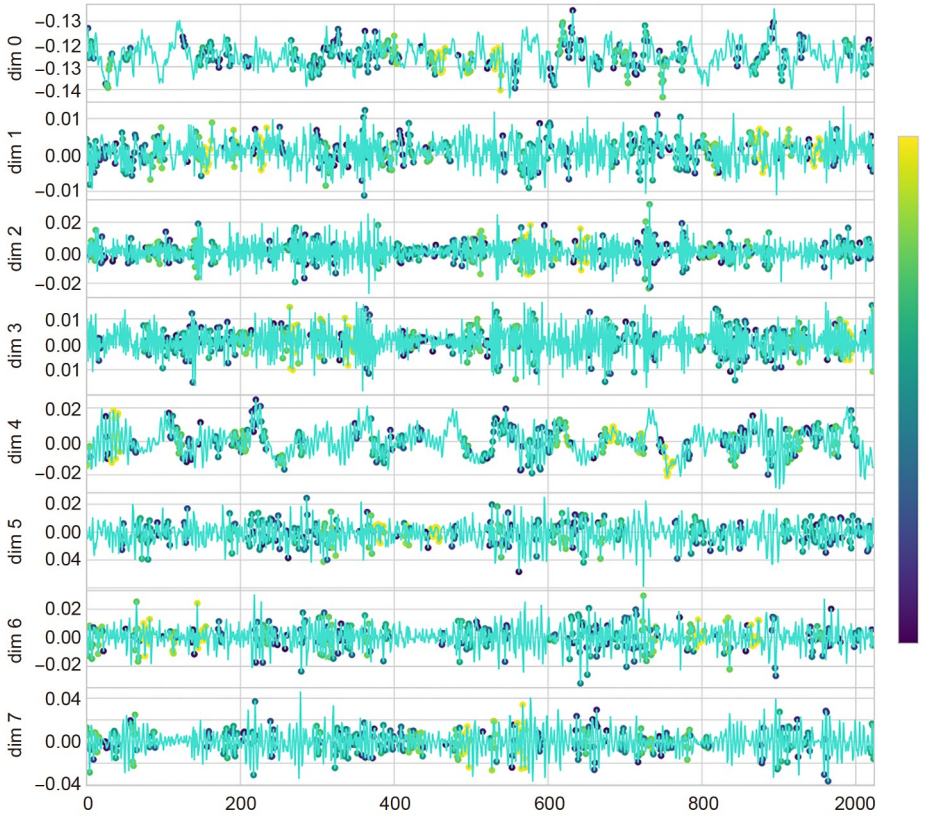


图 14 (网络版彩图) Grad-Cam热力可视化

Figure 14 (Color online) Thermal visualization of Grad-Cam.

参考文献

1 Zhang Y, Sun J, Zhang J, et al. Health state assessment of bearing with feature enhancement and prediction error compensation strategy. Mech Syst Signal Proc, 2023, 182: 109573

2 Zhang Y, Gong Z W, Zheng Y, et al. Joint study on health state assessment and degradation trend prediction of industrial equipment (in Chinese). Sci Sin Tech, 2022, 52: 180–197 [张永, 龚众望, 郑英, 等. 工业设备的健康状态评估和退化趋势预测联合研究. 中国科学: 技术科学, 2022, 52: 180–197]

3 Ji D, Wang C, Li J, et al. A review: Data driven-based fault diagnosis and RUL prediction of petroleum machinery and equipment. Syst Sci Control Eng, 2021, 9: 724–747

4 Zhang Y, Xin Y, Liu Z, et al. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE. Reliability Eng Syst Saf, 2022, 220: 108263

5 Zheng Y, Zhou W, Yang W, et al. Multivariate/minor fault diagnosis with severity level based on Bayesian decision theory and multidimensional RBC. J Process Control, 2021, 101: 68–77

6 Shakiba F M, Shojaee M, Azizi S M, et al. Real-time sensing and fault diagnosis for transmission lines. Inter J Network Dynam and Intel, 2022, 1: 36–47

7 Yuan Y, Ma G, Cheng C, et al. A general end-to-end diagnosis framework for manufacturing systems. Natl Sci Rev, 2020, 7: 418–429

8 Zhang Y, Tu L, Xue Z, et al. Weight optimized unscented Kalman filter for degradation trend prediction of lithium-ion battery with error compensation strategy. Energy, 2022, 251: 123890

9 Ran S, Yang X, Liu M, et al. Homecare-oriented ECG diagnosis with large-scale deep neural network for continuous monitoring on embedded devices. IEEE Trans Instrum Meas, 2022, 71: 1–13

10 Ke L, Zhang Y, Yang B, et al. Fault diagnosis with synchrosqueezing transform and optimized deep convolutional neural network: An application in modular multilevel converters. Neurocomputing, 2021, 430: 24–33

11 Liu Z, Chen Y, Zhang Y, et al. Diagnosis of arrhythmias with few abnormal ECG samples using metric-based meta learning. Comput Biol Med, 2023, 153: 106465

12 Zhu J, Cao J, Saxena D, et al. Blockchain-empowered federated learning: Challenges, solutions, and future directions. ACM Comput Surv, 2023, 55: 1–31

13 Ghimire B, Rawat D B. Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. IEEE Internet Things J, 2022, 9: 8229–8249

14 Rodríguez-Barroso N, Jiménez-López D, Luzón M V, et al. Survey on federated learning threats: Concepts, taxonomy on attacks and defences, experimental study and challenges. Inf Fusion, 2023, 90: 148–173

15 Huo Z, Yang X, Yang T H, et al. A prototypical networks-based multi-task model for few-shot fault diagnosis. In: Proceedings of the IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS). Chengdu, 2022. 996–1001

16 Peng P, Lu J, Xie T, et al. Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation. IEEE Trans Ind Inf, 2023, 19: 2463–2473

17 Zhang T, Chen J, Liu S, et al. Domain discrepancy-guided contrastive feature learning for few-shot industrial fault diagnosis under variable working conditions. IEEE Trans Ind Inf, 2023, 19: 1–11

18 Chen J, Li J, Huang R, et al. Federated transfer learning for bearing fault diagnosis with discrepancy-based weighted federated averaging. IEEE Trans Instrum Meas, 2022, 71: 1–11

19 Shao H D, Xiao Y M, Yan S. Simulation data-driven enhanced unsupervised domain adaptation for bearing fault diagnosis (in Chinese). J Mech Eng, 2023, 59: 76–85 [邵海东, 肖一鸣, 颜深. 仿真数据驱动的改进无监督域适应轴承故障诊断. 机械工程学报, 2023, 59: 76–85]

20 Ma X, Wen C, Wen T. An asynchronous and real-time update paradigm of federated learning for fault diagnosis. IEEE Trans Ind Inf, 2021, 17: 8531–8540

21 Warnat-Herresthal S, Schultze H, Shastry K L, et al. Swarm Learning for decentralized and confidential clinical machine learning. Nature, 2021, 594: 265–270

22 Yuan Y, Liu J, Jin D, et al. DeceFL: A principled fully decentralized federated learning framework. National Science Open, 2023, 2: 20220043

23 Yang D, Lu J, Dong H, et al. Pipeline signal feature extraction method based on multi-feature entropy fusion and local linear embedding. Syst Sci Control Eng, 2022, 10: 407–416

24 Yu H, Wang K, Li Y, et al. Representation learning with class level autoencoder for intelligent fault diagnosis. IEEE Signal Process Lett, 2019, 26: 1476–1480

25 Dou J, Song Y. An improved generative adversarial network with feature filtering for imbalanced data. International Journal of Network Dynamics and Intelligence, 2023, 2: 100017

26 Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). Munich, 2018. 3–19

27 Hou Y, Zhang Y, Lu J, et al. Application of improved multi-strategy MPA-VMD in pipeline leakage detection. Syst Sci Control Eng, 2023, 11: 2177771

28 Xue Y, Yang R, Chen X, et al. A novel local binary temporal convolutional neural network for bearing fault diagnosis. IEEE Trans Instrum Meas, 2023, 72: 1–13

29 Wang C, Wang Z D, Ma L F, et al. Subdomain-alignment data augmentation for pipeline fault diagnosis: An adversarial self-attention network. IEEE Trans Ind Inf, 2024, 20: 1374–1384

30 Hu J, Li W, Zheng X, et al. Prior knowledge-based residuals shrinkage prototype networks for cross-domain fault diagnosis. Meas Sci Technol, 2023, 34: 105011

31 Yuan Ye, Zhang Y, Ding H, Research on key technology of industrial artificial intelligence and its application in predictive maintenance (in Chinese). Acta Autom Sin, 2020, 46: 2013–2030 [袁烨, 张永, 丁汉. 工业人工智能的关键技术及其在预测性维护中的应用现状. 自动化学报, 2020, 46: 2013−2030]

32 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS). 2017. 4080–4090

33 Koch G, Zemel R, Salakhutdinov R, et al. Siamese neural networks for one-shot image recognition. ICML deep learning work shop, 2015

34 Vinyals O, Blundell C, Lillicrap T, et al. Matching net-works for one shot learning. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS). Red Hook, 2016

35 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS). Red Hook, 2017

36 Sung F, Yang Y, Zhang L, et al. Hospedales, learning to compare: Relation network for few-shot, learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, 2018. 1199–1208

37 Chen Z, Xu Y Q, Wang H, et al. Deep STFT-CNN for spectrum sensing in cognitive radio. IEEE Commun Lett, 2021, 25: 864–868

38 Liang R J, Ran W F, Yu C L, et al. Recognition of gearbox operation fault state based on CWT-CNN (in Chinese). J Aerosp Power, 2021, 36: 2465–2473 [梁睿君, 冉文丰, 余传粮, 等. 基于CWT-CNN的齿轮箱运行故障状态识别, 航空动力学报, 2021, 36: 2465–2473]

39 Xu Q F, Cheng Q L, Jiang C X, et al. Fault diagnosis of wind turbine bearing and gearbox based on group-sequence and multi branch CNN-LSTM model (in Chinese). J Mech Electr Eng, 2022, 39: 1050–1060 [许启发, 程启亮, 蒋翠侠, 等. 基于组序列多分支CNN-LSTM的风机轴承和齿轮箱故障诊断研究. 机电工程, 2022, 39: 1050–1060]

40 Du W L, Gao J J, Yang L K, et al. Multi-scale weighted CEEMD-1DCNN rotating machinery fault diagnosis (in Chinese). Mech Tool Hydr, 2023, 51: 202–208 [杜文辽, 高军杰, 杨凌凯, 等. 多尺度加权CEEMD-1DCNN旋转机械故障诊断. 机床与液压, 2023, 51: 202–208]

Semi-supervised small-sample gearbox fault diagnosis with privacy protection

HE Xin[1], DUAN YaQiong[2,4], WANG ZiDong[3] & ZHANG Yong[2]

[1] *School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;*
[2] *School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China;*
[3] *Department of Computer Science, Brunel University London, London UB83PH, United Kingdom;*
[4] *Yuanshi Intelligent Technology Co., Ltd, Nantong 226000, China*

Gearboxes, as a key component of mechanical drive systems, encounter problems of data leakage and low model accuracy when employing traditional fault diagnosis methods in terms of limited labeled data and data privacy protection requirements. Therefore, this study proposes a privacy-preserving framework based on federated learning using a small-sample gearbox fault diagnosis method that combines semi-supervised prototype networks with comparative learning. Initially, the DeceFL federated learning framework is constructed to produce positive and negative sample pairs using a limited number of labeled samples for each client. Meanwhile, the pretraining method of contrast learning is used to provide initialization parameters for the autoencoder. Subsequently, the autoencoder is used as a feature mapping function for the prototype network to compute the category prototype using limited labeled samples. Finally, category prototype is fine-tuned using the prototype refinement method to reduce the interference of abnormal data and obtain a more stable and accurate prototype. After validation with real gearbox data, the results show that the proposed semi-supervised small-sample fault diagnosis method can achieve better fault recognition accuracy with very few samples. This research provides an innovative approach to address data challenges in real industrial applications and promotes further research in the gearbox fault diagnosis field.

gearbox, troubleshooting, privacy protection, small sample, semi-supervised learning

doi: 10.1360/SST-2023-0351