



面向工业设备寿命预测的边缘节点自适应联邦学习方法

罗韬, 郭亮*, 高宏力, 闫问鼎, 胡平, 李俊

西南交通大学机械工程学院, 成都 610031

* E-mail: guoliang@swjtu.edu.cn

收稿日期: 2026-01-11; 接受日期: 2026-02-05; 网络版发表日期: 2026-03-25

国家自然科学基金联合基金(批准号: U2568229)、国家重点研发计划(编号: 2022YFB3402101)和四川省自然科学基金(编号: 2023NSFSC1960)资助项目

摘要 在分布式的工业设备寿命预测任务中, 受数据隐私保护和分布式部署等因素限制, 运行数据难以集中共享, 导致单节点数据训练的寿命预测模型难以跨设备应用. 联邦学习通过在不共享原始数据的前提下实现多节点协同建模, 为解决分布式场景的寿命预测提供了有效途径. 然而, 在实际工业场景中, 受计算能力和通信条件限制, 各节点参与联邦训练的程度往往不一致, 同时各节点在运行工况和数据分布等方面普遍存在差异, 上述因素易引发模型聚合偏移, 进而影响模型在不同节点上的预测性能. 针对上述问题, 本文提出了一种面向工业设备寿命预测的边缘节点自适应联邦学习方法(federated adaptive edge node learning method, FedAEN). 该方法从节点与模型两个层面进行考虑: (1) 在节点层面, 针对不同节点训练参与程度不一致的问题, 通过度量各节点模型更新之间的相似性, 引导全局模型进行自适应加权聚合, 从而抑制低一致性节点对聚合结果的干扰; (2) 在模型层面, 针对节点数据分布差异引起的特征表示偏移问题, 将寿命预测模型拆分为共享表示模块与本地自适应预测模块, 并在训练过程中约束共享表示模块在相邻联邦轮次之间的变化幅度, 以稳定特征表示空间, 同时利用全局模型对本地训练过程进行知识引导, 在保持节点本地预测能力的基础上提升模型的整体泛化性能. 在刀具和轴承两类典型工业设备数据集上的实验结果表明, 所提方法在多节点分布式条件下能够有效提升寿命预测的精度与稳定性, 验证了其在工业场景中的有效性.

关键词 寿命预测, 联邦学习, 自适应学习, 数据分布差异, 工业设备

1 引言

随着工业装备向高可靠性与智能化方向持续发展, 监测关键部件的健康状态、预测剩余使用寿命(remaining useful life, RUL)已成为保障工业设备安全

运行、降低运维成本的重要手段之一^[1-4]. 刀具、轴承等核心部件在长期服役过程中, 由于制造误差、装配条件及服役环境等差异, 不同装备个体之间呈现出明显的性能差异与退化不确定性^[5-7]. 相比依赖精确机理建模的物理方法, 数据驱动的寿命预测技术能够

引用格式: 罗韬, 郭亮, 高宏力, 等. 面向工业设备寿命预测的边缘节点自适应联邦学习方法. 中国科学: 技术科学, 2026, 56: 1148–1159
Luo T, Guo L, Gao H L, et al. Federated adaptive edge node learning method for industrial equipment remaining useful life prediction (in Chinese). Sci Sin Tech, 2026, 56: 1148–1159, doi: 10.1360/SST-2026-0015

直接从多源运行数据中挖掘退化特征, 在复杂工况下表现出更强的建模灵活性和应用潜力^[8,9].

在大数据背景下, 围绕寿命预测模型的精度提升与工程适用性开展了大量研究. 相关工作主要基于振动、电流等运行信号, 结合深度学习模型对设备退化过程进行建模. 例如, Ruiz-Sarmiento等人^[10]通过融合专家知识与产线运行数据, 构建了面向实际产线的退化评估与预测模型; Wang等人^[11]提出的融合循环卷积结构与概率推断的深度模型, 在捕捉退化时序依赖关系与预测不确定性方面取得了良好效果. 针对滚动轴承等关键部件, Cheng等人^[12]将卷积神经网络与双向长短期记忆网络相结合, 有效提升了退化建模精度与寿命预测精度.

这类方法在特定设备或工况条件下表现良好, 但其性能往往依赖于训练数据与测试场景之间较强的一致性. 在实际工业应用中, 同类设备在不同产线上的运行工况、负载水平存在差异, 导致采集数据在统计特性与退化模式上呈现出显著的不一致性. 这使得基于单一数据源得到的模型, 在跨设备或跨工况应用时难以保持稳定性能. 针对上述问题, 已有研究从迁移学习与域自适应角度展开探索, 试图通过对不同数据分布之间关系的建模来提升模型的泛化能力. 例如, Wang等人^[13]通过结合时序卷积网络与迁移学习, 实现了轴承跨工况寿命预测; Gao等人^[14]针对刀具磨损预测中的数据不足与分布差异问题, 引入多粒度特征与迁移机制提升预测精度. 在更一般的跨域场景中, Wu等人^[15]与Dong等人^[16]分别从语义一致性对齐与多约束分布对齐角度出发, 进一步改善了复杂工况下的寿命预测性能.

尽管上述方法在一定程度上缓解了分布差异带来的性能退化, 但大多依赖于集中式数据访问或额外的目标域信息, 在数据隐私受限、难以集中共享的工业场景中应用受到限制. 近年来, 联邦学习(federated learning, FL)通过在不共享原始数据的前提下实现多个边缘节点协同建模, 为工业设备寿命预测提供了一种新的技术路径. 该框架能够有效利用分布在不同企业或设备端的运行数据, 在保障数据安全的同时提升模型的整体性能^[17]. 然而, 现有联邦学习方法多假设各节点能够以相似方式参与训练, 并通过简单的参数平均构建单一全局模型^[18]. 但在实际场景中, 由于数据分布差异, 统一的全局模型难以同时兼顾各节点的

特征表示需求, 其结果往往趋于折中解, 削弱了模型在各个用户上的表现^[19].

为应对上述问题, 近年来研究者开始在联邦学习框架下引入个性化建模机制, 通过参数解耦、分层建模或局部优化等方式, 在共享跨节点知识的同时增强模型对本地数据分布的适应能力. 典型方法包括通过共享表征与本地判别相结合的方式实现个性化建模^[20], 以及通过特征对齐与局部、全局协同优化来缓解模型更新不一致问题^[21-24]. 此外, 也有研究从模型结构与训练策略层面提升个性化性能. 例如, Chen等人^[25]通过设计双预测器结构, 在训练中同时优化通用任务与本地个性化任务. Oh等人^[26]通过解耦模型在通用性与个性化方面的职责, 采用鲁棒损失训练通用预测器并结合轻量自适应模块实现个性化预测的联邦鲁棒解耦方法. 而Zhang等人^[27]采用自适应局部聚合策略, 在每轮迭代前根据本地目标动态融合全局与本地模型参数, 以提升本地模型初始性能的自适应局部聚合方法. 上述个性化技术的演进, 推动了联邦学习在数据异构环境中的实用性, 为实际部署提供了更灵活的解决方案.

然而, 上述方法多集中于通用分类任务, 在工业寿命预测任务中的应用研究较为缺乏. 此外, 节点参与程度不一致与运行工况差异往往同时存在, 易引发模型聚合偏移, 尚难以直接应用于复杂工业场景中的寿命预测需求. 图1对工业寿命预测场景中的两类问题进行了示意说明.

针对上述问题, 本文提出了一种面向工业设备寿命预测的边缘节点自适应联邦学习方法(federated adaptive edge node learning method, FedAEN). 该方法在联邦学习框架下从节点层面与模型层面协同设计自适应策略. 在节点层面, 针对不同边缘节点因计算能力、通信条件差异导致的参与轮次不一致问题, 引入模型信息相似性度量机制, 引导全局模型的加权聚合过程, 以抑制低一致性模型对整体性能的干扰. 在模型层面, FedAEN由共享表示模块与本地自适应预测模块组成, 在联邦训练过程中约束共享表示在相邻轮次之间的变化幅度, 以稳定特征表示空间, 同时利用全局模型对本地训练过程进行知识引导, 在保持节点本地预测能力的基础上提升模型整体泛化性能. 本文的主要工作和贡献可总结如下.

(1) 针对寿命预测中边缘节点训练参与程度不一

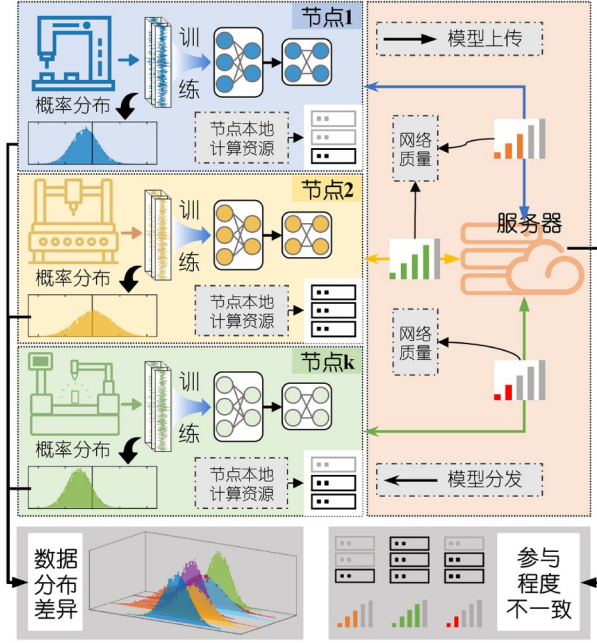


图1 工业场景寿命预测关键问题示意图

Figure 1 Key challenges in remaining useful life prediction for industrial scenarios.

致的问题, 提出了一种基于模型更新相似性的自适应聚合策略, 缓解了不均衡参与带来的聚合偏移。

(2) 构建了一种共享表示与本地自适应预测相结合的联邦学习建模方式, 并通过约束共享表示模块在训练过程中的演化幅度, 提高了特征表示的稳定性。

(3) 将全局模型知识引导机制引入本地训练过程, 在保持节点自适应预测性能的同时, 提升了模型在多节点分布式条件下的整体泛化能力与预测稳定性。

2 边缘节点自适应联邦学习方法

2.1 自适应联邦学习问题描述

对于一个由 K 个边缘节点组成的联邦学习系统, 节点集合记为 $\mathbb{K} = \{1, 2, \dots, K\}$. 每个节点 $k \in \mathbb{K}$ 持有本地私有数据集 $\mathcal{D}_k = \{(x_i^k, y_i^k)\}_{i=1}^{N_k}$, 其中 x_i^k 表示设备运行信号, y_i^k 表示对应的寿命标签. 由于不同设备运行工况、使用策略及数据采集条件存在显著差异, 各节点数据通常服从不同分布, 即 $\mathcal{D}_k \sim P_k(x, y)$, 且 $P_k \neq P_j$, $k \neq j$.

本文在自适应联邦学习框架下, 假设每个节点均采用一种由共享表示模块与本地自适应预测模块构成的模型结构, 用以在共享跨节点知识的同时增强对本地数据分布的适应能力. 为明确区分两部分参数, 共享表示模块的参数记为 Φ , 本地自适应预测模块的参数记为 Ψ_k . 因此, 节点 k 的完整模型参数可表示为

$$\Lambda_k = \Psi_k(\Phi(\cdot)), \quad (1)$$

则节点 k 的自适应模型在其本地数据 $\mathcal{D}_k$ 上的经验风险可表示为

$$\mathcal{L}_k(\Lambda_k) = \mathbb{E}_{(x,y) \sim \mathcal{D}_k} [\ell(\Lambda_k; x, y)], \quad (2)$$

其中, $\ell(\cdot)$ 为整体损失. 自适应联邦学习的总体优化目标是, 在通过全局参数 Φ 促进协作学习的同时, 为每个边缘节点找到最优的本地参数 Ψ_k . 该目标可形式化表述为如下联合优化问题:

$$\min_{\{\Lambda_k\}_{k=1}^K} \left\{ \mathcal{F}(\{\Lambda_k\}_{k=1}^K) := \frac{1}{K} \sum_{k=1}^K \mathcal{L}_k(\Lambda_k) \right\}, \quad (3)$$

其中, $\{\Lambda_k\}_{k=1}^K$ 表示所有节点的本地参数集合. 该目标旨在通过最小化所有节点平均经验风险, 协同优化全局共享表示与本地特定适配器, 从而在差异化数据的环境中实现有效的本地化建模。

为更直观地说明所提方法在联邦学习框架下的整体结构与训练流程, 图2给出了边缘节点自适应联邦学习方法的总体框架. 该框架清晰展示了节点层面的基于相似性的自适应加权聚合机制以及模型层面的稳定性表示与全局知识引导训练过程. 基于该整体框架, 下面将分别从节点层面和模型层面对所提方法进行详细介绍。

2.2 基于节点相似性的自适应加权聚合

在工业联邦学习场景中, 各节点之间不仅存在数据分布差异, 还普遍存在计算能力与通信条件的不一致. 这种差异往往导致不同节点参与联邦训练的轮次和本地更新步数不一致, 从而使得各节点模型在同一通信轮结束时存在显著差别. 若直接对所有节点模型进行等权聚合, 容易引入低质量或低相关模型更新, 对全局模型性能产生负面影响。

为缓解上述问题, 本文从节点层面引入一种基于节点相似性的自适应加权聚合策略. 在服务器聚合阶

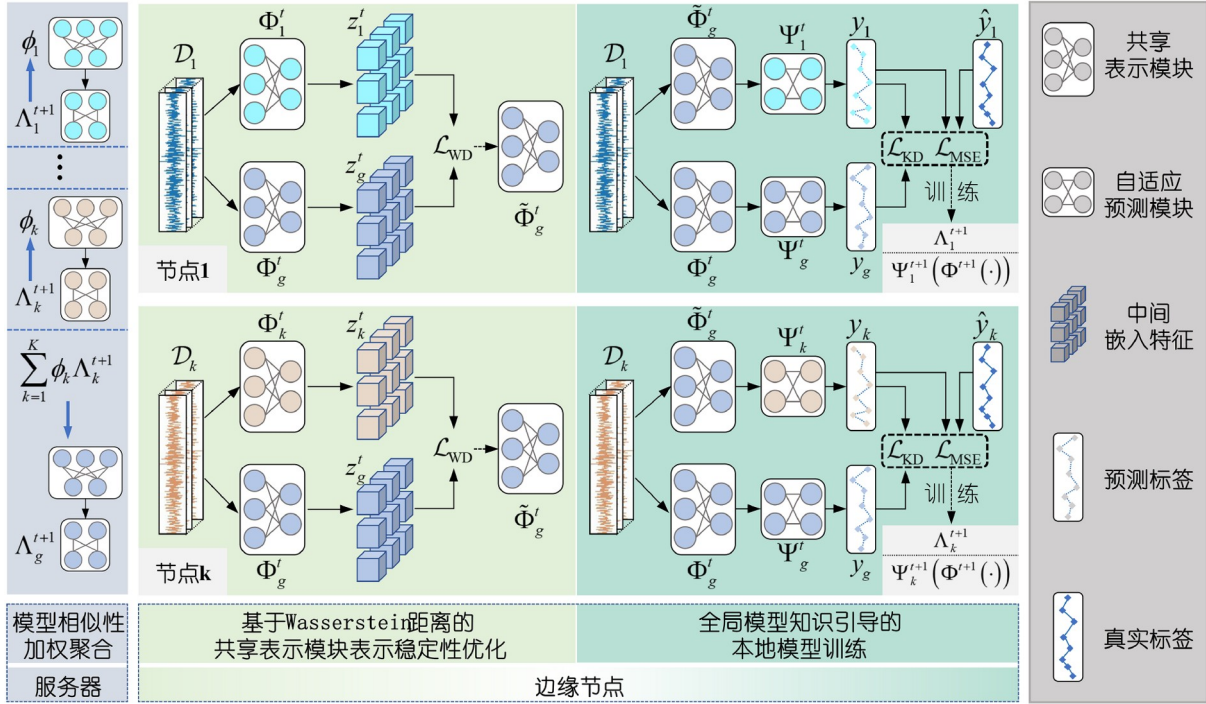


图2 边缘节点自适应联邦学习方法(FedAEN)框架

Figure 2 Framework of the federated adaptive edge node learning method (FedAEN).

段,通过度量不同节点模型之间的相似性,引导全局模型参数的自适应更新.在服务器中,自适应加权聚合操作作用于完整模型参数集合,即共享表示模块与本地预测模块同时参与聚合过程.通过共同聚合,共享表示模块可以在全局层面提取更具泛化性的特征表示,本地预测模块可以集成其他节点信息用于后续的知识引导训练.

具体而言,首先对节点模型参数进行向量化表示.以节点 k 的完整模型参数为例,其参数向量可表示为

$$\theta_k = \text{vec}(\Lambda_k), \quad (4)$$

其中, $\text{vec}(\cdot)$ 表示参数展平操作.随后,采用余弦相似度度量节点 k 与节点 j 模型之间的信息相似性:

$$\text{sim}(k, j) = \frac{\langle \theta_k, \theta_j \rangle}{\|\theta_k\| \|\theta_j\|}. \quad (5)$$

由于服务器需要为每个节点生成一个聚合参考模型,本文进一步定义节点 k 的整体模型相似性权重为与其其余所有节点模型相似度之和:

$$s_k = \sum_{j \in \mathbb{K}, j \neq k} \text{sim}(k, j). \quad (6)$$

为消除尺度差异,对相似性权重进行归一化处理,得到最终聚合权重:

$$\phi_k = \frac{s_k}{\sum_{i \in \mathbb{K}} s_i}. \quad (7)$$

在此基础上,服务器端对共享表示模块与本地预测模块分别执行加权聚合,得到全局模型参数:

$$\Lambda_g = \sum_{k=1}^K \phi_k \Lambda_k. \quad (8)$$

该策略能够在聚合过程中自适应强调与大多数节点一致性较高的模型更新,从而抑制因训练参与不一致带来的负面影响.

2.3 稳定性表示与全局知识引导训练

尽管通过节点层面的加权聚合可以提升全局模型的鲁棒性,但在联邦学习框架下,模型在不同通信轮之间仍需频繁经历“本地更新-全局聚合-模型下载”的交

替过程. 在寿命预测任务中, 该过程容易导致共享表示模块在连续训练轮次中发生剧烈漂移, 使得特征语义不连续, 进而削弱模型对退化模式的刻画能力. 为此, 本文在模型层面设计了一个两阶段的本地训练策略, 分别从表示稳定性与模型泛化性两个角度对本地更新过程进行约束.

(1) 基于Wasserstein距离的共享表示模块表示稳定性优化. 设第 t 轮通信结束后, 节点 k 上一轮本地共享表示模块为 Φ_k^t , 服务器聚合下发的共享表示模块为 Φ_g^t . 在本地训练开始前, 各节点利用上述两个模块对同一批本地样本 $x \sim \mathcal{D}_k$ 提取特征:

$$z_k^t = \Phi_k^t(x), z_g^t = \Phi_g^t(x). \quad (9)$$

为限制共享表示在连续轮次中的剧烈变化, 本文采用Wasserstein距离度量两组特征分布之间的差异:

$$\mathcal{L}_{\text{WD}}^k = W(\mathbb{P}(z_k^t), \mathbb{P}(z_g^t)), \quad (10)$$

其中, $W(\cdot, \cdot)$ 表示一阶Wasserstein距离.

在第一阶段, 本地训练仅针对共享表示模块进行优化, 其目标函数定义为

$$\min_{\Phi_k} \mathcal{L}_{\text{WD}}^k. \quad (11)$$

该过程能够显式约束共享表示模块在联邦聚合前后的特征分布差异, 增强表示空间在跨轮次更新过程中的连续性与稳定性, 为后续预测建模提供一致的特征基础. 需要说明的是, 经过多层卷积与特征提取后, 中间嵌入特征维度显著降低, 且每一轮训练仅于本地节点计算一次Wasserstein距离, 因此额外计算开销相对有限, 不会对整体训练效率产生明显影响.

(2) 全局模型知识引导的本地模型训练. 在完成共享表示模块的稳定化优化后, 本文进一步通过知识蒸馏机制, 将全局模型中所蕴含的跨节点泛化知识融入本地模型训练过程. 为避免蒸馏过程过于关注点值误差, 忽略模型在退化趋势与函数形态层面的整体一致性. 为此, 本文引入带温度系数的软蒸馏策略, 通过对预测输出进行温度缩放, 使蒸馏过程更加关注预测函数的平滑行为与相对变化趋势, 从而更有效地传递全局模型的泛化特性.

在第 t 轮通信中, 全局模型参数 Λ_g^t 对输入样本 x 的预测输出为

$$y_g = \Lambda_g^t(x). \quad (12)$$

节点 k 的本地模型预测输出为

$$y_k = \Lambda_k^t(x). \quad (13)$$

为实现软蒸馏, 引入温度系数 T ($T > 0$), 对全局模型与本地模型的预测输出进行缩放处理, 以拉近全局与本地模型的输出分布, 因此

$$\tilde{y}_g = \frac{y_g}{T}, \tilde{y}_k = \frac{y_k}{T}. \quad (14)$$

当 $T > 1$ 时, 预测输出的数值分布被适当平滑, 从而弱化单点误差的影响, 突出模型在整体预测行为上的一致性. 基于上述缩放后的预测结果, 定义带温度的蒸馏损失为

$$\mathcal{L}_{\text{KD}}^k = \frac{1}{N_k} \sum_{i=1}^{N_k} (\tilde{y}_{g,i} - \tilde{y}_{k,i})^2. \quad (15)$$

与此同时, 寿命预测的监督损失采用均方误差(MSE)定义为

$$\mathcal{L}_{\text{MSE}}^k = \frac{1}{N_k} \sum_{i=1}^{N_k} (y_i^k - \hat{y}_i^k)^2. \quad (16)$$

综合监督学习与全局知识引导, 本地模型在第二阶段的训练目标函数定义为

$$\min_{\Phi_k, \Psi_k} \mathcal{L}_{\text{MSE}}^k + \lambda \mathcal{L}_{\text{KD}}^k, \quad (17)$$

其中, λ 为权衡系数, 用于调节全局知识引导强度. 通过引入带温度的软蒸馏机制, 本地模型不仅能够保持对自身数据分布的良好拟合, 还能够在训练过程中逐步吸收全局模型中蕴含的共性退化规律, 从而提升其在本地数据的泛化能力与训练稳定性.

综合上述节点层面与模型层面的设计, 提出的FedAEN的整体流程可概括为[算法1](#). 在训练初始化阶段, 服务器端采用统一随机初始化方式构建初始模型参数, 并将包含共享表示模块与本地预测模块的完整模型下发至各边缘节点作为初始模型, 以保证不同节点训练起点的一致性.

3 实验与结果

为验证所提FedAEN在工业设备寿命预测任务中的有效性与鲁棒性, 本文在铣刀与轴承全寿命周期数

算法 1: FedAEN

输入: 初始化模型 $\Lambda = \Psi(\Phi(\cdot))$, 训练轮数 T , 节点数量 K

```

1: 服务器将初始化模型  $\Lambda$  分发至各节点
2: for  $t=1,2,\dots,T$  do
3:   for each 节点  $k \in \mathbb{K}$  do
4:     if  $t = 1$  do
5:       各节点通过式(16)训练本地模型, 完成初始化
6:        $\Lambda_k = \Psi_k(\Phi(\cdot))$ 
7:     end
8:     上传模型参数  $\Lambda_k$  至服务器
9:   end
10: 服务器基于式(8)执行加权聚合得到
11:   $\Lambda_g^t = \Psi_k(\Phi(\cdot))$ 
12: 模型下发至各节点
13: for each 节点  $k \in \mathbb{K}$  do
14:   if  $t > 1$  do
15:     基于式(11)优化共享表示模块稳定性, 得到  $\tilde{\Phi}_g^t$ 
16:     调整后的本地模型更新为  $\Lambda_k^t = \Psi_k(\tilde{\Phi}_g^t)$ 
17:     基于式(17)训练本地模型, 得到  $\Lambda_k^{t+1}$ 
18:   end
19:   上传模型参数  $\Lambda_k^{t+1}$  至服务器
20: end
21: 服务器基于式(8)执行加权聚合得到
22:   $\Lambda_g^{t+1} = \Psi_k(\Phi(\cdot))$ 
23: 模型下发至各节点
24: end

```

数据集上开展实验, 从多个角度对FedAEN的预测性能、稳定性及泛化能力进行评估。

3.1 实验数据集与联邦场景构建

(1) 铣刀全寿命周期数据集. 如图3所示, 本数据采集自大恒五轴立式加工中心VMC 850, 通过力传感器、振动加速度传感器和电流传感器各采集3种信号, 共计9通道信号. 通过测量后刀面磨损宽度, 根据GB/T16460-1996标准, 当主切削刃磨损宽度大于0.3 mm时, 认为该刀具达到退化极限, 停止寿命实验。

寿命周期实验中, 通过调整切削速度、进给速度及切削深度, 模拟了多种加工工况. 不同工况组合下, 刀具退化过程及信号分布呈现出较强的差异. 本文将每种工况视为一个独立节点的加工环境, 各节点仅使用其本地数据进行模型训练与更新。

本次实验中, 共设置了三种不同工况, 每种工况均包含4把刀具的完整退化数据. 具体工况参数如表1所示. 其中C1和C2为定工况实验, C3为变工况实验, 在

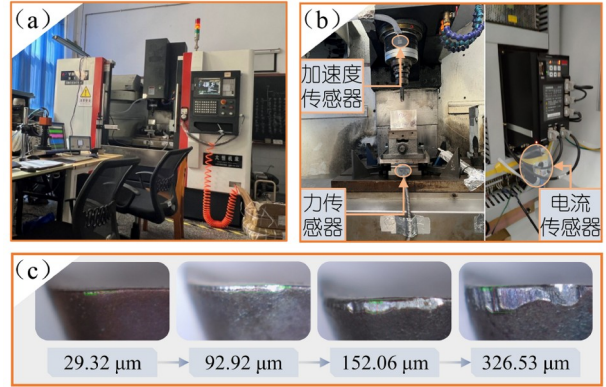


图3 (网络版彩图) 铣刀全寿命周期实验平台. (a) VMC 850 加工中心; (b) 传感器布置图; (c) 铣刀磨损变化过程

Figure 3 (Color online) Full life-cycle experimental platform for milling cutters. (a) VMC 850 machining center; (b) sensor layout; (c) evolution of milling cutter wear.

表1 铣刀磨损实验工况参数设置

Table 1 Parameter settings of operating conditions in the milling cutter wear experiments

工况	转速 (r/min)	进给速度 (mm/min)	切深 (mm)
C1	3000	200	1.0
C2	3000	200	1.5
C3	3000	200	1.0/1.5 (阶段)

主切削刃磨损小于0.15 mm时切深保持在1.0 mm, 当测量到磨损大于0.15 mm时则改变切深为1.5 mm, 直至刀具失效。

为构建实验数据, 从每种工况内选取3把刀具数据作为训练集, 用于模型学习, 剩余1把刀具数据则作为测试集, 用于评估模型在该工况下的泛化预测能力. 具体训练集与测试集划分如表2所示。

全寿命周期实验中采用的采样频率为10 kHz, 单次切削时长为30 s, 为缩减数据大小, 提升训练效率, 仅使用前1 s的数据用于构建数据集. 因此每个用于训练和测试的数据形状为 $1 \times 10000 \times 9$, 其中9表示9通道的信号。

(2) XJTU-SY轴承数据集. 为进一步验证所提方法的性能, 本文采用XJTU-SY轴承数据集进行实验. 该数据集由西安交通大学与长兴昇阳科技有限公司联合搭建的加速寿命试验平台采集得到. 该数据集进行了3种工况下的全寿命周期实验, 每种工况下进行5个轴承试验. 该实验中利用DT9837便携式动态信号采集器收集

表 2 各工况下训练、测试集刀具划分

Table 2 Division of milling cutters into training and testing sets under different operating conditions

节点	训练集	测试集
A	C1-1、C1-2、C1-3	C1-4
B	C2-1、C2-2、C2-3	C2-4
C	C3-1、C3-2、C3-3	C3-4

X和Y两个方向的振动信号。该数据集的更多信息详见文献[28]。

依据实验工况将数据集划分至不同节点，以模拟真实联邦学习场景，具体数据划分详见表3。数据预处理阶段，采用 3σ 原则对原始信号进行判定，识别并提取轴承的初始退化点，该方法已被广泛应用于简单的旋转机械退化阶段划分中^[29]。需说明的是，由于编号Bearing1-4与Bearing1-5的轴承退化阶段数据量不足，本实验中选用编号Bearing1-3的轴承作为节点1的独立测试集。

该数据集中相邻两次采样的间隔为1 min。同样地，为控制数据规模，每个时间点仅采用20000个数据点的数据，用于训练和测试的数据形状为 $1 \times 20000 \times 2$ ，其中2表示两个方向的振动信号。

3.2 模型结构

为保证对比实验的公平性，在实验中对各方法的基础网络规模进行统一设置。在不改变不同联邦学习方法核心训练机制与参数更新策略的前提下，本文为各方法配置了结构规模一致的特征提取网络与预测网络，以避免模型容量差异对实验结果的影响。

其中，基础特征提取部分采用共享表示模块形式，主要负责从原始多通道时序信号中提取退化相关特征，其结构由多层一维卷积网络构成，包含卷积、批归一化、激活函数及池化等操作，用以逐层提取具有退化信息的时序特征。预测部分由全连接网络组成，用于完成剩余使用寿命的回归建模。表4给出了本文采用的基础网络结构配置。

3.3 实验设置与评价指标

模型训练采用Adam优化器，初始学习率设为0.0001，训练轮次设为50，批大小为64。为保证实验公平性，所有对比方法均在相同的数据划分与训练配置

表 3 XJTU-SY轴承数据集划分

Table 3 Data partition of the XJTU-SY bearing dataset

节点	转速 (r/min)/负载 (kN)	训练集	测试集
A	2100/12	Bearing1-1 Bearing1-2 Bearing1-4 Bearing1-5	Bearing1-3
B	2250/11	Bearing2-1 Bearing2-2 Bearing2-3 Bearing2-4	Bearing2-5
C	2400/10	Bearing3-1 Bearing3-2 Bearing3-3 Bearing3-4	Bearing3-5

表 4 模型结构配置

Table 4 Model architecture configuration

模块	训练集	测试集
共享表示模块	Conv1D+BN+ReLU	通道数32，卷积核5
	MaxPooling	池化尺寸2
	Conv1D+BN+ReLU	通道数64，卷积核3
	MaxPooling	池化尺寸2
自适应预测模块	Conv1D+BN+ReLU	卷积核3，输出通道128
	Adaptive Avg Pooling	—
自适应预测模块	全连接层+ReLU	128→64
	全连接层	64→1

下进行。对于所提的FedAEN方法，软蒸馏策略中的温度系数 T 设为2，蒸馏损失权重 λ 设为0.2。考虑到温度系数 T 和蒸馏损失权重 λ 对整体性能表现影响较为平缓，本文选取的参数配置能够在保证训练稳定性的同时有效发挥知识引导作用，因此在后续实验中保持固定设置。

考虑到实际工业边缘环境中节点可能因通信或计算资源受限而出现间歇性离线的情况，在联邦训练过程中引入随机扰动机制以模拟节点不稳定参与的运行状态。具体而言，在每一轮联邦训练开始前，以固定概率随机选取部分节点不参与当轮本地训练与模型上传，其余节点正常执行本地更新并参与服务器端聚合。未参与当轮训练的节点在下一轮中仍有机会重新加入联邦过程。该设置模拟了节点参与的不确定性，从而更贴近实际边缘计算场景中节点的运行特点。

性能评价方面，采用均方根误差(RMSE)和平均绝对误差(MAE)作为主要量化指标，综合衡量模型预测精度。

3.4 对比实验

在铣刀数据集上, 本文将所提FedAEN与多种代表性联邦学习方法进行对比, 包括基础联邦学习方法FedAvg^[18], 以及先进的联邦学习方法FedRoD^[25]、FedBABU^[26]和FedALA^[27]. 各方法在测试集上的定量评价结果如图4所示.

可以看出, 在引入节点随机扰动用以模拟边缘节点间歇性离线的条件下, FedAEN在两个评价指标上均取得了近乎最优的预测性能, 在三个边缘节点上的平均RMSE和MAE分别为0.2957和0.2573, 整体优于其余对比方法. 相比之下, 传统FedAvg方法由于仅通过参数平均获得单一全局模型, 且未考虑节点参与程度不一致对模型聚合过程的影响, 在节点随机缺失的情况下更易受到噪声更新的干扰, 其预测误差最大, 平均RMSE和MAE分别为0.3203和0.2670, 表现出较差的鲁棒性. FedRoD、FedBABU和FedALA作为先进联邦学习方法, 针对节点数据分布差异进行了改进, 但在节点间歇性离线的实验设定下, 上述方法仍然依赖固定或弱约束的聚合策略, 难以有效区分不同节点更新的一致性程度, 其预测性能虽较FedAvg有明显提升, 但整体仍弱于所提出的FedAEN方法.

为直观展现出不同方法在随机扰动条件下的预测差异, 图5给出了各节点上的寿命预测曲线对比结果. 为减小预测的RUL曲线波动, 对预测结果采用高斯滤波进行平滑处理. 可以观察到, FedAvg在训练过程中受到节点参与不连续性的影响, 其预测结果存在较为明显的整体偏移与局部波动. FedRoD、FedBABU和FedALA在一定程度上改善了预测曲线的平滑性, 但

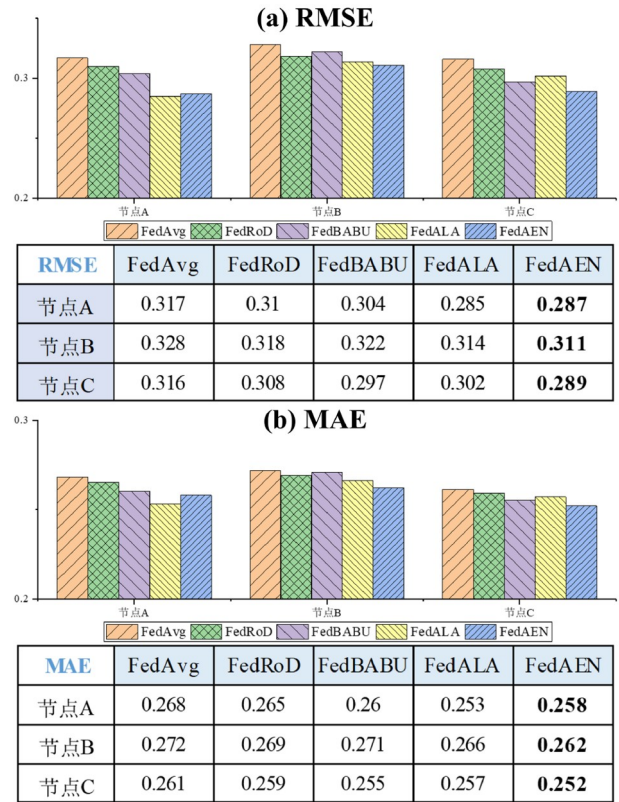


图4 不同方法在铣刀数据集上的预测性能对比

Figure 4 Comparison of prediction performance of different methods on the milling cutter dataset.

在部分退化阶段仍表现出过度预测或滞后响应的问题, 反映出其对不稳定模型更新的敏感性. 相比之下, FedAEN在整个退化过程中均能够较好地跟踪真实寿命变化趋势.

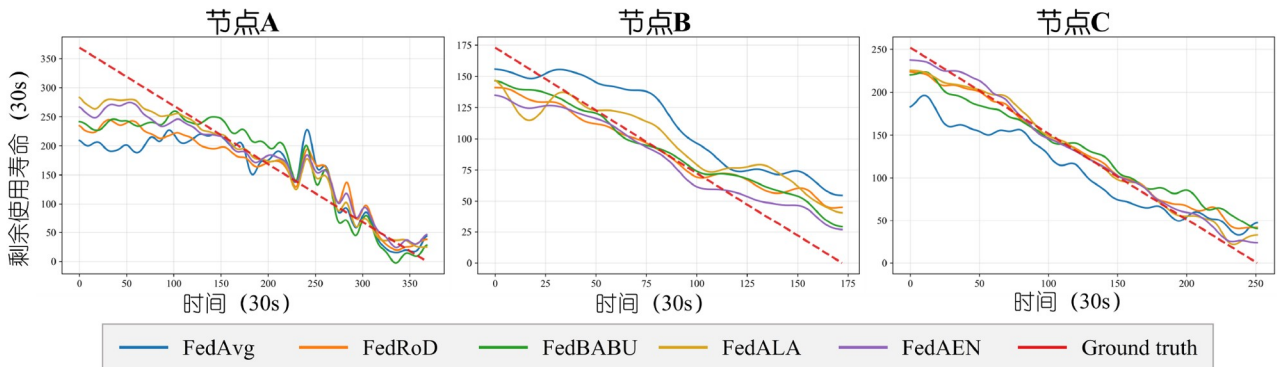


图5 同方法在典型铣刀上的寿命预测结果对比

Figure 5 Comparison of remaining useful life prediction results of different methods on a representative milling cutter.

3.5 消融实验

为验证FedAEN中各关键机制对模型性能的贡献, 在铣刀数据集上进一步开展消融实验. 具体而言, 构建了三种消融变体: M1, 基于模型相似性的加权聚合机制; M2, 基于Wasserstein距离的表示稳定性优化; M3, 全局模型知识引导机制. 各消融方法的定量结果如表5所示, 表中符号“√”表示启用了对应机制或策略, 符号“×”则表示未使用相关部分的功能.

从表5可以看出, 移除任何一个关键模块均会导致模型预测性能出现不同程度的下降. 去掉M1后, 平均RMSE和MAE出现了一定程度的上升, 说明在节点参与不一致和数据差异较大的场景下, 直接进行无区分的模型聚合会放大低一致性模型更新的影响, 削弱联邦训练效果. 去掉M2后的结果表明, 若不对共享表示模块在跨轮次训练过程中的特征分布变化进行约束, 模型容易发生表示漂移, 从而影响后续自适应预测模块的学习稳定性. 其中, M3的性能退化最为明显, 平均RMSE和MAE相较完整模型分别上升了0.0153和

0.0074, 证明了全局模型知识引导机制的有效性, 缺失该机制后, 模型的泛化能力下降明显, 预测精度受到明显影响.

3.6 XJTU-SY轴承数据集泛化验证

为进一步验证所提方法在不同设备类型和数据分布条件下的泛化能力, 本文在XJTU-SY轴承数据集上对FedAEN进行了补充实验验证. 图6展示了不同方法在典型轴承节点上的寿命预测曲线, 图7给出了相应的定量评价结果.

从预测曲线可以看出, FedAEN在引入节点随机扰动的条件下同样能够较好地刻画真实退化趋势, 其预测结果在整体变化趋势和局部响应上均与真实曲线保持较好的一致性. 定量结果进一步表明, FedAEN在RMSE和MAE指标上均优于其他联邦学习方法, 验证了其在不同设备类型和运行数据条件下的稳定性能表现. 上述结果说明, FedAEN并未过度依赖特定数据集或设备特性, 具备较强的跨任务适应能力.

表 5 FedAEN消融实验结果对比

Table 5 Comparison of ablation study results for FedAEN

键部分			节点A		节点B		节点C	
M1	M2	M3	RMSE	MAE	RMSE	MAE	RMSE	MAE
×	×	×	0.317	0.268	0.328	0.272	0.316	0.261
×	√	√	0.302	0.260	0.319	0.267	0.305	0.257
√	×	√	0.298	0.259	0.315	0.264	0.303	0.254
√	√	×	0.308	0.265	0.323	0.270	0.311	0.259
√	√	√	0.296	0.258	0.311	0.262	0.289	0.252

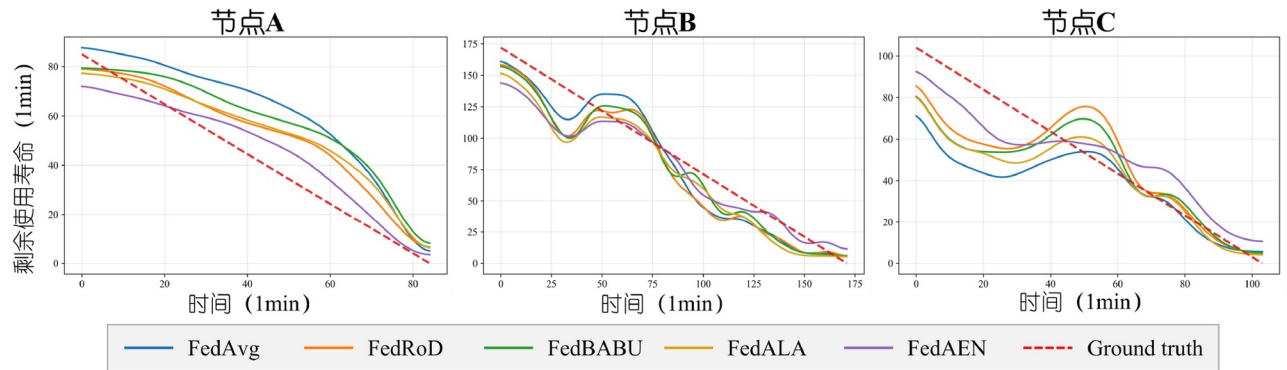


图 6 不同方法在轴承数据集上的预测结果对比

Figure 6 Comparison of prediction results of different methods on the bearing dataset.

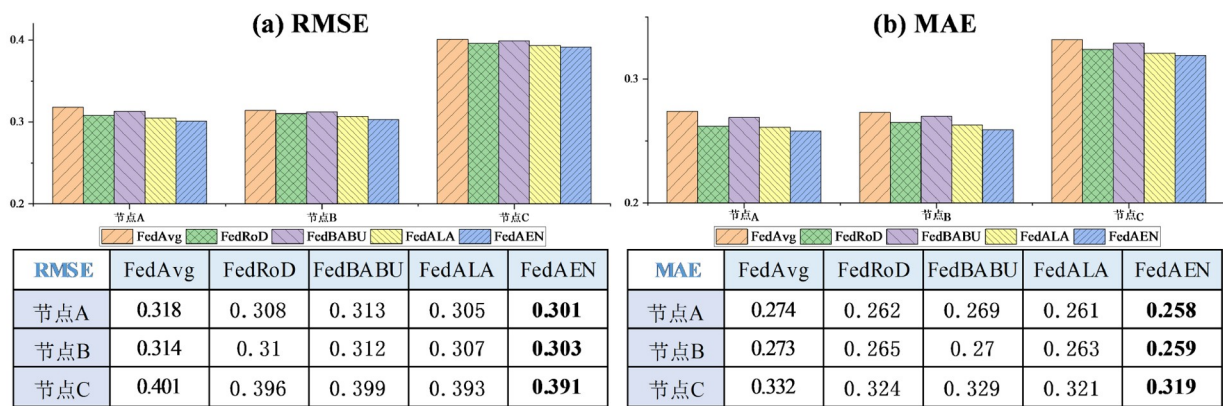


图 7 (网络版彩图)不同方法在轴承数据集上的预测性能对比
Figure 7 (Color online) Comparison of prediction performance of different methods on the bearing dataset.

4 结论

本文针对分布式工业设备寿命预测任务中节点参与不稳定、数据分布差异显著问题,提出了一种面向工业设备寿命预测的边缘节点自适应联邦学习方法 FedAEN。该方法从节点层面与模型层面进行协同设计:在节点层面,通过度量不同节点模型更新之间的相似性,引导全局模型进行自适应加权聚合,有效减弱了因节点间歇性离线导致参与程度不一致所带来的聚合偏移;在模型层面,通过共享表示稳定化约束与全

局知识引导训练机制,在提升模型泛化能力的同时保持了模型在本地的预测适应性。

通过在铣刀和轴承两类典型工业设备数据集上的实验,结果表明在引入节点随机扰动、模拟实际边缘节点运行条件的情况下, FedAEN在寿命预测精度、训练稳定性以及跨设备泛化能力方面均取得了良好效果。消融实验进一步验证了各关键机制在提升模型鲁棒性和预测性能中的有效性。上述研究为联邦学习在复杂工业边缘场景下的寿命预测应用提供了一种可行的技术路径。

参考文献

- Guo L, Yu Y, Qian M, et al. FedRUL: A new federated learning method for edge-cloud collaboration based remaining useful life prediction of machines. *IEEE ASME Trans Mechatron*, 2023, 28: 350–359
- Gebraeel N, Lei Y, Li N, et al. Prognostics and remaining useful life prediction of machinery: Advances, opportunities and challenges. *J Dyn Monitor Diagnost*, 2023, 2: 1–12
- Zio E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliab Eng Syst Saf*, 2022, 218: 108119
- Song K, Hu D, Tong Y, et al. Remaining life prediction of lithium-ion batteries based on health management: A review. *J Energy Storage*, 2023, 57: 106193
- Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech Syst Signal Process*, 2013, 35: 108–126
- Jiang L, Zhang T, Lei W, et al. A new convolutional dual-channel transformer network with time window concatenation for remaining useful life prediction of rolling bearings. *Adv Eng Inf*, 2023, 56: 101966
- Zhang X, Shi B, Feng B, et al. A hybrid method for cutting tool RUL prediction based on CNN and multistage Wiener process using small sample data. *Measurement*, 2023, 213: 112739
- Misbah I, Lee C K M, Keung K L. Fault diagnosis in rotating machines based on transfer learning: Literature review. *Knowl-Based Syst*, 2024,

283: 111158

- 9 Liu X, Chen Y, Xiong L, et al. Intelligent fault diagnosis methods toward gas turbine: A review. *Chin J Aeronaut*, 2024, 37: 93–120
- 10 Ruiz-Sarmiento J R, Monroy J, Moreno F A, et al. A predictive model for the maintenance of industrial machinery in the context of industry 4.0. *Eng Appl Artif Intell*, 2020, 87: 103289
- 11 Wang B, Lei Y, Yan T, et al. Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery. *Neurocomputing*, 2020, 379: 117–129
- 12 Cheng Y, Hu K, Wu J, et al. A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings. *Adv Eng Inf*, 2021, 48: 101247
- 13 Wang Y, Ding H, Sun X. Residual life prediction of bearings based on SENet-TCN and transfer learning. *IEEE Access*, 2022, 10: 123007–123019
- 14 Gao Z, Hu Q, Xu X. Condition monitoring and life prediction of the turning tool based on extreme learning machine and transfer learning. *Neural Comput Applic*, 2022, 34: 3399–3410
- 15 Wu K, Li J, Zuo L, et al. Weighted adversarial domain adaptation for machine remaining useful life prediction. *IEEE Trans Instrum Meas*, 2022, 71: 1–11
- 16 Dong X, Zhang C, Liu H, et al. A multi-constrained domain adaptation network for remaining useful life prediction of bearings. *Mech Syst Signal Process*, 2024, 206: 110900
- 17 Sattler F, Wiedemann S, Muller K R, et al. Robust and communication-efficient federated learning from non-i.i.d. data. *IEEE Trans Neural Netw Learn Syst*, 2020, 31: 3400–3413
- 18 McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Ft. Lauderdale: PMLR, 2017. 1273–1282
- 19 Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks. *Proc Mach Learn Syst*, 2020, 2: 429–450
- 20 Collins L, Hassani H, Mokhtari A, et al. Exploiting shared representations for personalized federated learning. In: *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021. 2089–2099
- 21 Zhou T, Zhang J, Tsang D H K. FedFA: Federated learning with feature anchors to align features and classifiers for heterogeneous data. *IEEE Trans Mobile Comput*, 2024, 23: 6731–6742
- 22 Xu J, Tong X, Huang S L. Personalized federated learning with feature alignment and classifier collaboration. arXiv: [2306.11867](https://arxiv.org/abs/2306.11867)
- 23 Zhang J, Liu Y, Hua Y, et al. FedTGP: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning. *AAAI*, 2024, 38: 16768–16776
- 24 Zhang J, Hua Y, Wang H, et al. FedCP: Separating feature information for personalized federated learning via conditional policy. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 2023. 3249–3261
- 25 Chen H Y, Chao W L. On bridging generic and personalized federated learning for image classification. In: *Proceedings of the Tenth International Conference on Learning Representations*. 2021
- 26 Oh J, Kim S, Yun S Y. FedBABU: Towards enhanced representation for federated image classification. arXiv: [2106.06042](https://arxiv.org/abs/2106.06042)
- 27 Zhang J, Hua Y, Wang H, et al. FedALA: Adaptive local aggregation for personalized federated learning. *AAAI*, 2023, 37: 11237–11244
- 28 Wang B, Lei Y, Li N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans Rel*, 2020, 69: 401–412
- 29 Lei Y, Li N, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process*, 2018, 104: 799–834

Federated adaptive edge node learning method for industrial equipment remaining useful life prediction

LUO Tao, GUO Liang, GAO HongLi, YAN WenDing, HU Ping & LI Jun

School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

In distributed industrial equipment remaining useful life (RUL) prediction tasks, operational data are difficult to centrally share due to constraints such as data privacy protection and distributed deployment, which makes RUL prediction models trained on single-node data challenging to generalize across devices. Federated learning provides an effective solution for distributed RUL prediction by enabling multi-node collaborative modeling without sharing raw data. However, in real industrial settings, limitations in computational capabilities and communication conditions often lead to inconsistent participation levels among nodes in federated training. Moreover, differences in operating conditions and data distributions among nodes commonly exist, which can cause model aggregation bias and thereby degrade prediction performance across different nodes. To address these issues, this paper proposes a federated adaptive edge node learning method (FedAEN) for industrial equipment RUL prediction. The method incorporates coordinated design at both the node level and the model level: (1) At the node level, to handle inconsistent participation in training across nodes, similarity among model updates from different nodes is measured to guide adaptive weighted aggregation of the global model, thereby mitigating interference from low-consistency nodes on aggregation results; (2) At the model level, to address feature representation shift caused by heterogeneous data distributions among nodes, the RUL prediction model is split into a shared representation module and a local adaptive prediction module. During training, the variation of the shared representation module between consecutive federated rounds is constrained to stabilize the feature representation space, while the global model provides knowledge guidance for local training processes to enhance overall generalization performance without compromising node-specific predictive capabilities. Experiments on two typical industrial equipment datasets, namely cutting tools and bearings, demonstrate that the proposed method effectively improves the accuracy and stability of RUL prediction under multi-node distributed settings, which verifies its effectiveness in industrial applications.

remaining useful life prediction, federated learning, adaptive learning, data distribution heterogeneity, industrial equipment

doi: [10.1360/SST-2026-0015](https://doi.org/10.1360/SST-2026-0015)