基于分布共识的联邦增量迁移学习

崔 腾[1)] 张海军[2)] 代 伟[1),3)]

[1)](中国矿业大学信息与控制工程学院 江苏 徐州 221116)

[2)](哈尔滨工业大学(深圳)计算机科学与技术学院 广东 深圳 518055)

[3)](中国矿业大学人工智能研究院 江苏 徐州 221116)

摘 要 相同生产工艺的工业过程协同建模是解决工业难测参数在线软测量的有效方法,但因生产原料、设备等因素差异,所形成的分布式数据往往呈现非独立同分布特性(Nonindependent Identically Distribution, Non-IID).同时,受生产环境变化影响,数据分布特性会随时间发生变化.因此,工业建模场景对模型的个性化配置和自主调整能力提出了更高的要求.为此,本文提出一种结构与参数并行优化的联邦增量迁移学习方法(Federated Incremental Transfer Learning, FITL).所提方法在增量式联邦学习框架下,建立了基于模型输出信息的联邦共识组织,并利用横向联邦进行组内增强;进而,面向联邦共识组织,通过最小化组间共识差异增量迁移不同共识组织信息;最后,结合组内横向增强和跨组织迁移学习,构造增量迁移下的联邦学习模型.在工业数据集和基准数据集上的实验结果表明,与现有方法相比,所提模型能更好地实现不同工况Non-IID情况下的协同建模.在过程工业回归任务和公开数据集的分类任务中,FITL能在多工况环境下相较基线方法提升9%和16%的模型预测精度.

关键词 协同建模;分布式数据;非独立同分布;迁移学习;联邦学习

中图法分类号 TP301 **DOI号** 10.11897/SP.J.1016.2024.00821

Federated Incremental Transfer Learning Based on Distributed Consensus

CUI Teng[1)] ZHANG Hai-Jun[2)] DAI Wei[1)3)]

[1)](*School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu* 221116)

[2)](*Department of Computer Science, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangzhou* 518055)

[3)](*Artificial Intelligence Research Institute of China University of Mining and Technology, Xuzhou, Jiangsu* 221116)

Abstract Industrial process collaborative modeling with the same production process is an effective method to solve the difficult industrial parameters online soft measurement. Due to the differences in production materials, equipment and other factors, the distributed data often prensent nonindependent identically distribution (Non-IID). Simultaneously, influenced by changes in the production environment, the distribution characteristics of data change over time. Consequently, industrial modeling scenarios demand heightened requirements for personalized configuration of models and autonomous adjustment capabilities. To address these concerns, this paper proposes a federated incremental transfer learning (FITL) strategy that achieves parallel optimization of both structure and parameters. Under the framework of incremental federated learning, a federated consensus organization based on model output information is established, and horizontal federated is used for intra-group enhancement. Furthermore, the information of different consensus organizations is incrementally migrated for federal consensus groups by minimizing

收稿日期:2023-06-22;在线发布日期:2024-01-11.本课题得到国家重点研发计划(2022YFB3304700)、国家自然科学基金(62373361)、中央高校基本科研业务费专项资金(2023XSCX027)、中国矿业大学研究生创新计划项目(2023WLKXJ095)及江苏省研究生科研与实践创新计划(KYCX23_2710)资助.**崔 腾**,博士研究生,主要研究领域为联邦学习、迁移学习. E-mail: cuicui@cumt.edu.cn. **张海军**,博士,教授,主要研究领域为机器学习、大数据分析. **代 伟(通信作者)**,博士,教授,主要研究领域为增量学习、联邦学习、工业大数据分析. E-mail: weidai@cumt.edu.cn.

consensus differences between groups. Finally, a federation learning model under incremental transfer is constructed by combining intra-group horizontal reinforcement and cross-organization transfer learning. Experimental results on industrial data sets and benchmark data sets show that, compared with the existing methods, the proposed model can better realize collaborative modeling under different working conditions of Non-IID. In the regression task of process industry and the classification task using public datasets, FITL exhibits a notable enhancement of 9% and 16% in model predictive precision over baseline methods in multiple working conditions.

Keywords collaborative modeling; distributed data; non-independent identically distribution; transfer learning; federated learning

1 引 言

工业制造中性能指标大多因依赖人工采样化验导致样本稀疏、建模困难[1-4]. 实际上,以生产特定产品为目标的工业过程大多采用典型生产工艺与生产设备,生产流程具有特征相似性[5-7]. 如果能将生产流程相同的各工业企业私有数据进行集中整合,利用企业共享的大数据进行协同建模,将大大提升模型质量. 然而,由于数据隐私和安全问题造成的工业数据信息孤岛使得上述目标难以达成[8-9]. 因此,如何在满足数据隐私和安全前提下,更加高效地共同使用具有相似特征的工业企业私有数据,实现各工业企业分布式协同建模,建立工业制造性能指标模型,是一项有意义的工作. 本文将该工作定义为数据安全下的分布式协同建模问题,并对该问题进行了深入的研究和求解.

针对数据安全下的分布式协同建模问题,联邦学习能够保证分布在多个地区的参与者在数据安全的情况下实现协作建模. 具体做法是保持数据在本地,信息通过共享模型梯度或参数来传递. Fedavg[10]是典型的联邦学习模型,其通过共享分布客户端的参数信息构建性能更优的全局模型,在多个实际领域效果显著[11-14]. 然而,由于生产原料、设备等差异,分布在多地的工业过程往往处于不同运行条件(即多工况),使过程数据呈现 Non-IID 特性. 若忽略工况差异建模,则过程数据难以被有效利用,导致模型效果不理想. 因此,需要为 Non-IID 的过程数据设计个性化的问题求解策略.

为了解决协同建模参与端的数据 Non-IID 问题,研究界开展了对联邦个性化建模的研究. 文献[15]在 Fedavg 的基础上,进一步考虑了客户端的个性化差异,通过在局部更新目标中添加模型参数更新近似项,提出了 FedProx. 该方法在面向客户端存在设备及统计异构特点的任务时具有显著优势. 文献[16]提出了联邦聚类学习,利用客户端更新的余弦相似度分析数据分布信息,并设计了一种自适应聚类机制. 文献[17]提出基于 Moreau Envelope 的客户端更新策略,结合局部更新和全局更新设计了更快收敛的联邦策略 pFedMe. 文献[18]提出客户端数据分布的混合假设,基于 EM 算法设计了 FedEM. 此外,一些研究人员使用联邦迁移学习来解决数据分布差异下的协同建模问题. 文献[19]提出将数据的原始特征映射到一个可以共享信息的高维空间,以减少由于多个参与方之间数据分布的差异而导致的模型性能下降. 文献[20]研究了智能制造领域多场景联邦合作中的数据异构化问题,其通过模型参数传递领域知识,实现了基于模型预训练策略的新型场景联邦模型的快速构建. 上述研究为客户端的非独立同分布问题提供了更具个性化的求解方案,在一定程度上解决了客户端的数据分布差异问题.

然而,上述方法均基于固定的拓扑结构建立模型,即其是在给定神经网络架构的情况下通过联邦协作优化模型参数,并不能实现模型结构与参数的并行优化. 神经网络的结构对模型学习效果影响至关重要,过小过紧的模型容易训练但学习能力低. 过大过松的模型学习能力强,但参数复杂,且数据不足时难以获得理想效果. 在本文的工业分布式协同建模场景中,工业过程往往具有时变性特性,对于模型的时效性要求较高,现有基于深度架构的神经网络难以在工业实际中实施调整与重建. 因此,直接推广现有计算智能的联邦学习架构应用在工业领域的效果并不理想. 增量构造式的随机配置网络(Stochastic Configured Network, SCN)[21]通过对数据特点的自适应分析,为解决模型结构与参数同步优化配

置难题提供了有益的途径.值得注意的是,与传统联邦学习模型在参数随机配置方面的做法不同,SCN遵循特定的机制来随机生成参数,以确保随机参数具有理论收敛保障.为实现数据安全与隐私下的分布式协同建模,文献[22]提出了联邦随机配置网络,其设计了增量场景下的联邦协作建模方法,基于SCN实现了客户端网络的自适应构建.然而,由于数据非独立同分布情况的存在,客户端难以在数据分布差异情况下实现有效的增量协作建模.迁移学习是解决数据分布差异下知识共享的有效方法,但传统的联邦迁移学习均基于给定的架构通过共享参数迁移知识,仅涉及特定层参数的优化,并不考虑模型结构的动态变化.据我们所知,尚未有文献研究增量构造的联邦迁移方法.本文针对工业过程的特性时变、工况复杂、差异性大等特性,采用SCN为客户端构建模型,在进行知识迁移优化参数的同时不仅要考虑信息传输的安全问题,还要考虑模型结构的动态变化,大大提升了迁移学习的实现难度.

为解决上述问题,本文通过建立分布数据共识的联邦分组机制,引入迁移学习技术生成组内模型增强和组间模型迁移的联邦互助策略,提出了基于分布共识的联邦增量迁移建模方法,具体为:首先,结合SCN建模特性,分析过程数据分布差异,将不同企业过程数据分布差异性考虑到模型构建过程.基于不同模型对于同一数据的预测输出一致性,即共识,建立基于私有模型输出相似度的联邦共识组织;然后,在模型增量构建过程中,建立组内基于质量优先策略的横向联邦合作机制,即组内增强.通过加权聚合机制生成组内私有模型新增节点的最优选择;最后,基于KL散度建立组内模型和其他多共识组织模型预测输出差异的损失函数,通过最小化组内模型本地数据预测误差及组间模型预测输出的损失函数,实现组间模型信息的增量迁移.本文主要贡献包括:

(1)在增量式联邦学习框架下,针对全局模型难以在多工况的企业生产环境下获得理想预测效果的问题,首次对数据分布差异下的模型自适应增量构建进行研究,提出了一种基于非梯度策略的联邦个性化私有模型增量式构建方法.

(2)针对不同工况的信息难以有效协作的问题,建立了基于模型输出信息的联邦共识组织机制,结合工况特性为不同的共识组织设计了组内、组间差异化的增量协作策略,并提供了收敛性分析.

(3)针对分布差异下信息难以有效共享及增量过程中迁移学习难以实现的问题,通过拟合预测分布趋势迁移组间知识,建立了基于迁移学习的组间互助机制.进而,结合组内增量和组间迁移实现了增量建模过程中的有效协作.

本文的其余部分组织如下:第2节分析了联邦学习研究进展及SCN建模的主要策略;同时,介绍了联邦随机配置网络增量构建和信息共享的关键策略,并进一步分析了其现有不足.第3节概述了本文所求解的联邦优化问题的问题定义.第4节介绍了联邦增量迁移学习方法的核心思想及算法的具体步骤.第5节通过在工业数据及公开数据集上的实验分析了FITL的优越性.第6节对本文的研究进行了总结与展望.

2　相关工作

2.1　联邦学习

联邦学习是一种分布式机器学习方法,其主要是为了应对由数据隐私意识增强造成的企业数据孤岛挑战而设计的.联邦学习允许在多个设备或节点上训练模型,而无需将私有数据集中存储,对于分布式协同建模具有重要意义.联邦学习方法最早由Google提出,其设计的Fedavg基于模型层面传输客户信息,实现了数据隐私下的分布式协同建模[10].后续,大多联邦学习的研究都基于该方法的建模思路展开,并在许多应用场景中引起了越来越多的关注[23-27].

然而,传统的Fedavg方法难以应对现实生产生活中的复杂场景.当参与联邦的客户端数据分布存在差异时,Fedavg常常面临收敛速度慢及通信次数过多问题.并且,由于仅构建了一个通用的全局模型,当面临具有个性化数据分布的客户端时,其难以获得预期的预测效果.针对上述挑战,Li等人[15]提出了FedProx,该方法为每个客户端的局部更新目标中额外设计了一个自适应调整估计项,提升了模型处理数据异构性问题时的准确性.Wu等人[28]提出FedHome,通过一种生成式卷积自编码器(GCAE)协调客户端数据以减少分布差异对联邦模型准确性的影响.Yang等人[29]和Guo等人[30]对客户端更新进行采样分析,每轮训练选取低偏置的客户端集更新信息进行聚合以提升全局模型的稳定性.上述方法虽然在一定程度上缓解了数据分布差异对于全局模型的影响,但仍然存在单一模型设定的局限.在本文的分布式协同建模场景中,不同工况

的数据其优化方向可能存在差异，因此仅基于单一模型的优化难以应对个性化的过程数据分布学习.

在学习个性化局部模型研究方面. Arivazhagan 提出了 FedPer[31]，该方法将局部模型划分成前端基础层和后端个性化层，其个性层不参与全局更新. Li 等人[32]提出了 FedBN，该方法为局部模型设置了个性化的批规范层，批规范层参数更新值不参与聚合，在处理客户端特征漂移问题上效果显著. Lu 等人[33]提出了 FedAP 用于医疗领域的患者信息分析，其同样设置了个性化的批处理层，并基于批处理层的参数信息进一步对客户端数据分布差异进行拟合. 迁移学习是处理具有分布差异的领域间学习的有效方法，因此，基于迁移学习实现局部模型的个性化构建也开始成为个性化联邦研究的热点. 但由于联邦学习数据传输的特殊设定，目前该方向的研究主要集中在基于模型的迁移学习策略. Kevin 等人[20]为智能制造场景下的数据异构性问题设计了基于模型预训练策略的联邦合作架构 FTL-CDP. Fang 等人[34]进一步考虑了客户端数据存在的噪声问题，构建了更具鲁棒性的联邦模型. Wei 等人[35]结合隐私保护机器学习与深度迁移学习构建了处理不同任务数据的联邦学习架构 MF-SCSN，用于患者脑电波数据解码. 上述方法虽然实现了客户端模型的个性化构建，但其均是基于固定的神经网络架构实现的，并且缺少关于轻量模型联邦建模的相关研究. 在本文的分布式协同建模场景中，我们旨在基于企业过程数据实现一个轻量模型的自适应构建，以便于模型的嵌入及快速加载.

此外，也有研究结合知识蒸馏技术解决客户端的数据异构性问题. 由于在联邦场景下，基于知识蒸馏技术解决数据异构性的核心在于提取多个局部模型的知识到本地模型，因此其也可以视为基于模型知识进行迁移的一种具体实现方法. Li 等人[36]提出基于知识蒸馏的联邦学习框架 FedMD，该方法通过在一个公共数据集上的共识训练模型，接着再利用微调对局部模型进行更新. Lin 等人[37]提出了 FedDF，其考虑了边缘客户端计算能力的差异，通过集成蒸馏进行交叉架构的学习，构建了异构的局部模型. Jiang 等人[38]进一步考虑了联邦模型部署期间的数据分布变化问题，设计了具有测试分布偏移鲁棒性的 FedTHE+. Yang 等人[39]针对蒸馏过程中数据异构对于平均软标签的影响，设计了多个教师独立蒸馏的 FedMMD. Chen 等人[40]基于循环蒸馏的方式构建了去中心化的联邦个性化学习方法 MetaFed. 上述方法能在模型异构的场景下实现分布式协作，并很好地缓解客户端计算能力差异的问题. 但其合作依赖于对于公共数据集的共识，需要选取的数据集具有无偏性，这在工业生产领域较难收集满足条件的数据集. 并且，上述方法的模型架构虽然是异构的，但初始的模型架构仍然需要在训练前指定，并未实现模型结构的自适应构建.

综上所述，以上研究在一定程度上提升了传统联邦模型对于个性化数据端的适应性，但其均基于固定的神经网络架构构建预测模型，并通过该固定架构优化网络的参数. 网络的架构对于神经网络预测效果至关重要，如何根据本地数据实现客户端模型的自适应协作构建，是联邦领域值得研究的问题. 本文通过增量的形式进行联邦协作，以实现网络结构与参数配置并行优化的目标. 所提出的 FITL 主要针对客户端具有分布差异下的模型自适应增量构建，并进一步结合领域相似性协助个性化模型提升构建效率. 据我们所知，这是该应用场景下针对联邦增量学习的首个研究.

另外，本文专注于工业过程场景. 该场景下，运行指标参数往往难以检测，样本化验标注昂贵，故训练样本相对稀少. 此外，工业过程动态时变和有限的硬件资源限制了模型的规模. 因此，传统的联邦学习模型并不适用于工业设备，需采用模型结构与参数具有自主调整能力的轻量化学习模型. 同时，由于传统神经网络对初始参数较敏感，不适合该场景的建模，故引入随机配置网络以实现参数自适应学习.

2.2 随机配置网络

SCN 是一种基于数据监督随机参数配置的构造型神经网络，其建立了网络节点构造过程中的监督机制，以实现模型结构与网络参数的同步优化配置. 由于 SCN 能根据数据分布特点实现模型的自适应构建，并在多种应用场景下展现了良好的通用逼近性，近年来在工业建模领域得到了广泛关注[41-42].

SCN 的模型结构如图 1 所示，其具体建模过程为：假设现有一个 $L-1$ 个节点的网络已经被构建，其输出为 $f_{L-1}(x)=\sum_{j=1}^{L-1}\beta_j a_j((w_j)^T x+b_j)$ $(L=1,2,\cdots,N, f_0=0)$. 其中，w 和 b 是隐藏层参数，β 是输出层参数，a 为激活函数. 记模型的拟合目标为 y，此时，具有 $L-1$ 个节点的模型的误差为 $e_{L-1}=f_{L-1}(x)-y$. 定义模型的期望误差为 e^{exp}，若 $e_{L-1}<e^{exp}$，SCN 将通过新增节点的方式实现目标的优化. 新增的第 L 个节点的生成主要包括两个阶

段,分别是隐藏层节点的随机配置和输出层权值的计算. 首先,是隐藏层权值的随机配置. 给定训练数据集 $D(x,y)$,SCN 会在公式(1)的监督条件下随机生成节点.

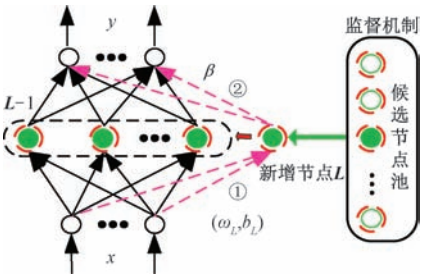

图 1　随机配置网络

$$\xi=\frac{\langle e_{L-1}^T,h_L\rangle}{\langle h_L,h_L\rangle}-(1-r-u_L)\langle e_{L-1}^T,e_{L-1}^T\rangle>0 \tag{1}$$

其中,h_L 为数据经过单个隐藏节点 L 的输出,r 和 u_L 为进行随机配置的超参数,$0<r<1$,$0<u_L<(1-r)$,$\lim\limits_{L\to\infty}u_L=0$.

该监督条件是基于模型当前 $L-1$ 个节点的学习结果及数据集信息构建的. 研究表明,ξ 值越大则网络收敛越快. 因此,SCN 会在每轮配置的候选节点中选出 ξ 最大的点对应的参数配置给新增的第 L 个节点. 记新增节点 L 的隐藏层参数为 w_L 和 b_L. 得到隐藏节点的参数配置后,SCN 根据公式(2)计算模型的输出权值 β ①.

$$\beta=\underset{\beta}{arg\,min}\ \frac{1}{2}\ ||\ y-h\beta\ ||_2^2=(h^Th)^{-1}h^Ty \tag{2}$$

其中 h 为数据经过隐藏层得到的输出集合.

重复上述新增节点的操作直到模型的误差达到既定的期望误差,更多关于 SCN 的研究请参考文献[43-45].

2.3　联邦随机配置网络

联邦随机配置网络(Federated Stochastic Configuration Networks, FSCNs)为解决数据孤岛及数据隐私问题下的增量模型构建难题提供了解决方案. 其根据随机配置网络的建模特点将模型构建与联邦学习相结合,联合分布在不同地区的多个客户端协同增量构建一个更优的全局模型. 与传统联邦不同的是,FSCNs 以节点构建的层面出发,注重各参与者对全局模型参数和结构的增量改进. 通过联邦协作实现网络结构与参数配置的并行提升,而不仅仅是在全局模型的参数上进行优化.

FSCNs 的大致建模架构如图 2 所示,与 SCN 的构建步骤类似,其也可以大致分为隐藏层节点配置与输出权值评估两个步骤. 在隐藏层节点配置阶段,分布在多个地区的客户端先在中央服务器的协调下用私有数据生成基于本地数据信息的隐藏层节点参数上传到中央服务器. 接着,中央服务器融合多个客户端的新增节点信息. 结合 SCN 增量建模的特点,FSCNs 设计了两种策略实现客户端之间的知识共享,其分别为加权聚合和贪婪选择策略. 在加权聚合策略中,服务器通过数据持有量对第 L 个节点的信息进行加权聚合,得到当前节点 L 的全局模型隐藏层信息. 而贪婪选择策略将选择出 ξ 最大的客户端拥有的参数信息作为全局模型的第 L 个节点的隐藏层参数. 此外,考虑多个客户端在不同监督机制下的收敛速度问题,其针对隐藏参数的选取分别设计了速度优先策略和质量优先策略. 其中,速度优先策略选取的参数只需要满足本地客户端的监督机制即可,而质量优先策略选择的参数需要满足所有客户端的监督机制. 速度优先策略的收敛速度更具随机性,而后者更稳定,FSCNs 可以按需选择增加节点的具体配置方案.

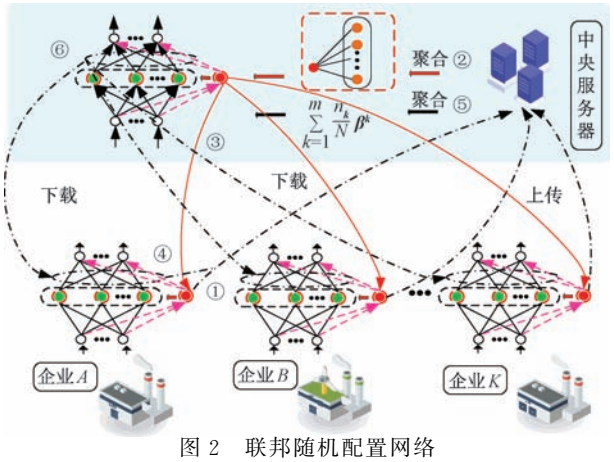

图 2　联邦随机配置网络

在输出权值评估阶段,客户端从中央服务器下载全局模型第 L 个节点的隐藏参数信息后,各自在本地计算输出权值并上传至服务器聚合. 服务器聚合输出权值后,得到具有 L 个节点的全局模型并计算当前误差,同时判断是否需要继续训练. 若满足预期效果则终止学习,输出构建的全局模型. 重复上述步骤,直至最小化目标函数公式(3).

$$\min\sum\nolimits_{k\in K}l_k(f(x),y) \tag{3}$$

上式中,f 为 FSCNs 构建的全局模型,l_k 为全局模型在本地数据 k 上的预测误差,K 为客户端集合.

FSCNs 解决了传统联邦学习模型的架构估计

① 鉴于 SCN-Ⅲ的快速收敛性[22],本文中涉及的 SCN 均指的是 SCN-Ⅲ.

困难问题,实现了多客户端的自主协同增量建模.其结合 SCN 的构建特点,有效地将联邦学习和增量建模结合,在保证数据安全的前提下进行模型的增量构建.然而,该方法未考虑客户端数据的非独立同分布问题.在实际场景中,不同客户端数据往往由于实际的工况差异导致数据分布差异,使得不同客户端的知识难以被有效共享.这种情况下,具有更稳定收敛保证的质量优先方法的 FSCNs 难以在短时间内找到满足全局监督条件的参数进行模型节点配置,收敛速度被大大降低.同时,数据分布差异也让全局模型难以有效利用来自多个客户端的知识,造成了全局模型的准确性无法得到保证的问题.

3 问题定义

联邦学习是解决数据安全前提下工业分布式协同建模的可靠技术,然而现有研究缺少能在工况差异下实现增量式协同建模的有效方法.

已有增量式联邦建模方法通过联合多方过程数据构建更优全局模型,其在数据同分布情况下效果显著.然而,由于企业实际生产过程的多工况现象,先前方法难以在数据分布差异下有效利用过程数据,导致其构建的全局模型准确性难以保证.为了更好应对工业领域多工况的生产环境,本文将基于生产数据为企业构建个性化模型,记为企业私有模型.

本文的多工况联邦合作场景考虑了一个中央服务器和多个不同企业的相似工艺过程,这些过程拥有不同生产工况的过程数据.定义总的过程数量为 m ,企业集合记为 $K=\{k_1,k_2,k_3,\cdots,k_m\}$.每个企业 k 拥有的过程数据集合记为 $D_k=\{(x_{n,k},y_{n,k})\mid x_{n,k}\in X_k,y_{n,k}\in T^k\}_{n=1}^{N_k}$.其中,x 是数据的特征,y 为样本的标签,N_k 是企业 k 的样本总数,X_k 和 T^k 为特征及标签集,该过程数据对应的分布记为 P_k .企业 k 在本地训练的模型记为私有模型 f_k .

企业在中央服务器的协调下构建各自私有模型,然后服务器融合基于不同过程数据构建的模型信息以供私有模型进行本地优化.本文采用随机配置网络构建私有模型,具体为

$$f_k(x)=\sum_{j=1}^{L}\beta_j^k a_j^k((w_j^k)^{\mathrm{T}}x+b_j^k) \tag{4}$$

其中,w^k、b^k 和 β^k 为私有模型 k 隐藏层的参数及输出层权值,L 为当前网络的节点个数,a^k 为私有模型的激活函数.

已有增量式构造的联邦学习方法通过利用分散在不同地域的本地数据 D_k 协同训练效果更好的联邦全局模型,该方法在 $P_{k_i}=P_{k_j}$ 时通常表现很好.然而,由于工况差异导致过程数据分布存在差异,因此 $P_{k_i}\neq P_{k_j}$.此时,增量构建的全局模型直接应用在工业实际建模中常常难以获得理想学习效果.因此,在本文的研究中,FITL 的学习目标是在 $P_{k_i}\neq P_{k_j}$ 的情况下,保持私有数据在本地,通过联邦协作增量式为企业建立更优的私有模型,实现最小化如下目标函数

$$\min\sum\nolimits_{k\in K}l_k(f_k(x),y) \tag{5}$$

上式中,l_k 为私有模型 f_k 的训练误差.

本文多工况场景下的增量式协作建模主要面临的难点来自三个方面:首先,在分布式建模中,数据安全是协作建模的重要前提.因此,服务端无法通过直接分析过程数据来区分具有不同工况的工业过程.故如何在保证数据安全的前提下准确识别工况差异以实现工业过程的合理划分是第一个难点.其次,同工况的数据信息融合与不同工况的信息融合方式存在差异.如果均按照差异工况处理则难以利用同工况协作的高效性,而均按照同工况则难以有效融合过程数据信息.因此,如何根据工况特性实现过程数据高效协作是第二个难点.最后,迁移学习是解决数据分布差异下信息共享的有效方法,然而现有联邦迁移学习方法无法在模型的增量构建过程中实现信息的有效融合.所以,如何在模型增量构建过程中实现数据信息的迁移是第三个难点.此外,需要注意的是,本文的建模方法需要考虑增量构建架构下的不同工况信息融合,而不仅仅是固定架构模型基础上的信息共享.因此,在联邦协作过程中,不仅需要考虑联邦学习场景下对于数据传输的特殊限制,还要考虑模型结构的动态变化对于后续优化的影响,以及网络模型在监督条件下的收敛性约束等,从而进一步增加了联邦学习的实现难度.

为了解决上述问题,本文提出联邦增量迁移学习方法.针对问题一,本文基于 SCN 构建私有模型.SCN 是在过程数据信息监督下构建的,其节点参数体现本地工况特性.因此,提出基于模型信息的共识组织建立机制,通过分析私有模型节点信息区分不同工况.该策略能在保证过程数据安全的前提下实现过程数据的合理划分.针对问题二,本文在所建立共识组织基础上设计了一种基于组内横向增强的联邦机制联合同工况私有模型信息.进一步,面向共识

组织,基于迁移学习融合不同工况的过程数据信息,以实现根据工况特性实现过程数据高效协作.针对问题三,本文建立基于 KL 散度的预测分布损失函数,通过最小化组间预测差异迁移多工况过程数据信息.该迁移策略能在模型的增量构建过程中基于模型的潜在架构信息实现组间过程数据的相互学习.同时,基于本地预测共识能在保证本地预测准确度的基础上,实现跨组模型泛化性增量式提升.此外,本文的主要符号及定义见表 1.

表 1　符号及定义

符号	定义
m	过程总数
K	企业集合
D_k	企业 k 的过程数据集
N_k	企业 k 拥有的过程数据总数
X	过程数据的特征集
T	过程数据的标签集
w^k	私有模型 k 的隐藏层权值
b^k	私有模型 k 的隐藏层偏差
β^k	私有模型 k 的输出层权值
h^k	私有模型 k 的隐藏层输出

4　联邦增量迁移学习方法

4.1　方法概述

本节给出了联邦增量迁移学习的整体架构流程.图 3 是多个不同的客户端通过联邦增量迁移方法协同训练模型的大致过程.

以企业 k 的过程为例,本文按照编号顺序在图 3 中展示了 FITL 如何结合组织建立、组内增强及组间迁移生成私有模型的过程.其大致步骤为:第一步,企业 k 上传基于本地数据生成的 SCN 模型构建信息至服务器;第二步,中央服务器基于共识组织建立机制将这些客户端分成若干个共识小组.接着,在小组内通过共识组织内部的横向增强机制进行分组协作;第三步,中央服务器将基于模型信息的各组内增强结果传输回本地;第四步,企业 k 基于本地数据通过组间信息迁移机制进行知识迁移工作;最后,重复上述步骤直至生成的私有模型 k 满足预期的效果.

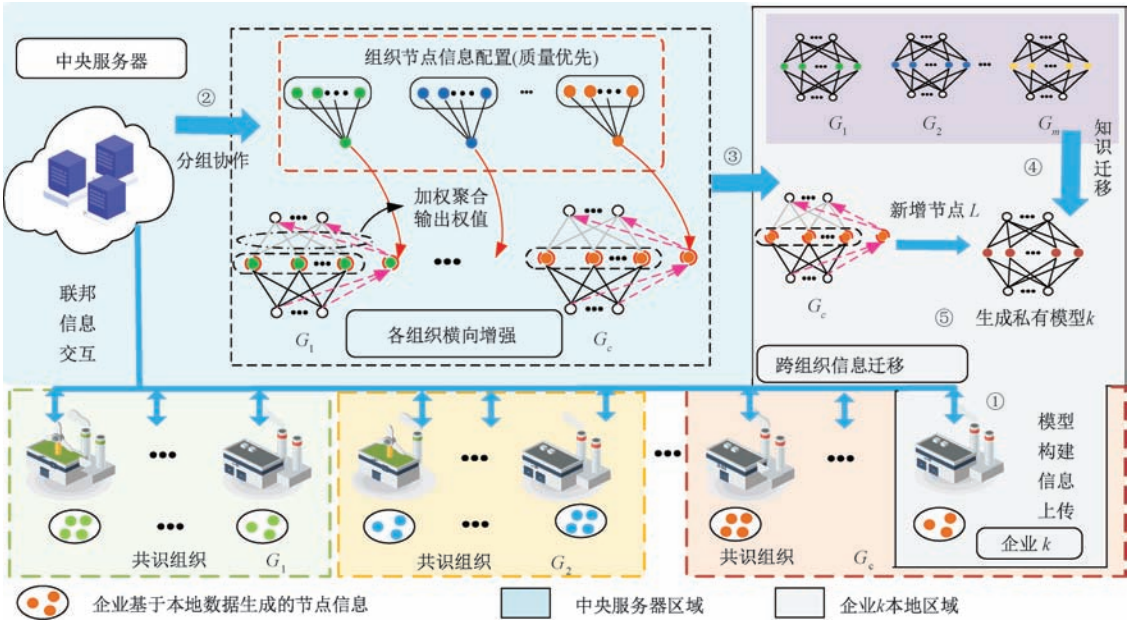

图 3　联邦增量迁移学习整体架构

已有的联邦增量建模方法通过联合多企业构建一个全局模型供本地使用.然而,在数据非独立同分布的情况下,其难以在各个本地过程数据上均取得理想效果.为此,FITL 引入迁移学习技术,为企业建立个性化模型,更有助于联合多方信息,实现本地问题的优化求解.

4.2　基于模型信息的共识组织建立机制

本文采用 SCN 为企业构建私有模型,SCN 是基于过程数据的监督机制生成,其模型参数体现数据的分布.本文假设网络的输出权值层体现数据的条件分布,而隐藏层体现数据的边缘分布,故将通过分析模型参数考察数据分布一致性.共识小组的划分核心是对客户端私有模型的划分,因此共识组织先基于预测器的差异进行初步构建.大致为:首先通过输出权值的评估结果分析数据的条件分布差异,对多个客户端的私有模型进行初步划分.接

着,进一步通过分析隐藏层输出的数据信息对初步划分的结果进行边缘分布差异下的细化.最后,融合条件概率和边缘概率的联合聚类效果测定共识组织的最终划分.

设目前有 m 个客户端参与联邦学习,$K=\{k_1,k_2,k_3,\cdots,k_m\}$.客户端 k_i 根据本地过程数据生成的第 L 个隐藏节点的参数为 $\phi_L^{k_i}=\{w_L^{k_i},b_L^{k_i}\}$,输出权值为 $\psi_L^{k_i}=\{\beta_L^{k_i}\}$.首先,根据输出权值 ψ_L^K 计算数据条件分布相似度矩阵 $S_{con}\in R^{m\times m}$,如下式所示

$$S_{\mathrm{con}(i,j)}=\frac{\mathrm{cov}(\phi_L^{k_i},\phi_L^{k_j})}{\sqrt{\mathrm{cov}(\phi_L^{k_i},\phi_L^{k_i})\cdot\mathrm{cov}(\phi_L^{k_i},\phi_L^{k_j})}} \tag{6}$$

其中,$\mathrm{cov}(\cdot,\cdot)$ 为数据的协方差,$k_i,k_j\in K$.设定阈值 TS,基于层次划分方法可以获得私有模型条件分布差异下的分组结果 $G_{S_{\mathrm{con}}}=\{g_1,g_2,\cdots,g_{c_{S_{\mathrm{con}}}}\}$,$c_{S_{\mathrm{con}}}$ 为条件分布差异下的分组数目.记私有模型 k_i 第 L 个隐藏节点的输出为 $h_L^{k_i}=h_L(X_{k_i},\phi_L^{k_i})$.接着,在 $G_{S_{\mathrm{con}}}$ 的每个小组 $g_{c_{S_{\mathrm{con}}}}$ 内基于企业集合输出结果 $h_L^K=\{h_L^{k_1},\cdots,h_L^{k_m}\}$ 分析过程数据的边缘分布差异.由于随机配置变量的数值范围变化较小,通过均值差异难以捕获数据边缘分布差异的变换.因此,本文通过对比数据的方差差异测量数据的边缘分布差异变化.可通过公式(7)计算边缘分布差异相似度矩阵 $S_{\mathrm{mar}}\in R^{m\times m}$

$$S_{\mathrm{mar}(i,j)}=\mathrm{var}(h_L^{k_i})-\mathrm{var}(h_L^{k_j}) \tag{7}$$

其中,var 为数据方差,$k_i,k_j\in K$.通过该差异对小组内的私有模型进行进一步的区分,得到边缘分布差异下的分组结果 $G_{S_{\mathrm{mar}}}=\{g_1,g_2,\cdots,g_{c_{S_{\mathrm{mar}}}}\}$.最后,融合输出最终双分布差异下的共识组织划分结果 $G=\{g_1,g_2,\cdots,g_c\}$,如公式(8)所示

$$\begin{aligned}G&=G_{S_{\mathrm{mar}}}^1\cup G_{S_{\mathrm{mar}}}^2\cup\cdots\cup G_{S_{\mathrm{mar}}}^{c_{S_{\mathrm{con}}}}\\&=\{g_1^1,g_2^1,\cdots,g_{c_{S_{\mathrm{mar}}}}^1\}\cup\{g_1^2,g_2^2,\cdots,g_{c_{S_{\mathrm{mar}}}}^2\}\\&\quad\cup\cdots\cup\{g_1^{c_{S_{\mathrm{con}}}},g_2^{c_{S_{\mathrm{con}}}},\cdots,g_{c_{S_{\mathrm{mar}}}}^{c_{S_{\mathrm{con}}}}\}\\&=\{g_1,g_2,\cdots,g_c\}\end{aligned} \tag{8}$$

其中,$c=\sum_{i=1}^{c_{S_{\mathrm{con}}}}c_{S_{\mathrm{mar}}}^i$ 是共识小组的总个数,算法 1 描述了共识小组的建立过程.

算法 1. 共识组织建立机制.

输入:待分组私有模型 $K=\{k_1,k_2,k_3,\cdots,k_m\}$ 的参数信息 ϕ_L^K,ψ_L^K,层次聚类阈值 TS_{con},TS_{mar}

输出:共识组织建立结果 $G=\{g_1,g_2,\cdots,g_c\}$

1. 初始化组织个数为 m
2. FOR $G_{S_{con}}$ 簇 $i=1,2,\cdots,m$ DO
3. $g_i=$ 私有模型参数 $\phi_L^{k_i}$;
4. END FOR
5. WHILE 簇间相似度大于 TS_{con} DO
6. 通过公式 6 计算 S_{con};
7. 合并相似度最高的两个簇,更新 $G_{S_{con}}$;
8. END WHILE
9. FOR $g_{c_{S_{con}}}\subset G_{S_{con}}$
10. 初始化 $G_{S_{mar}}^{c_{S_{con}}}$ 的簇为每个私有模型参数 $\phi_L^{k_i}$;
11. WHILE 簇间相似度大于 TS_{mar} DO
12. 根据公式 7 计算 S_{mar};
13. 合并相似度最高的两个簇类,更新 $G_{S_{mar}}^{c_{S_{con}}}$;
14. END WHILE
15. END FOR

采用上述步骤可以得到双分布差异划分后的共识组织.通过划分共识组织,可区分出有共性的私有模型,从而更有针对性地利用组间的共识信息.此外,本文在建立共识组织过程中仅涉及了模型参数的交互,不涉及本地数据的传输,因此不会影响企业隐私.下节中,所提 FITL 方法将在划分的共识组内进行横向增强,其可使得具有相似数据分布的参与者更紧密地合作,从而更好地共享彼此的信息和知识.

4.3 基于横向联邦的组内增强机制

基于私有模型建立共识组织后,FITL 基于参数信息在组织内部通过横向联邦进行组内增强.具体为,在本文私有模型的构建过程中,共识组织内部的隐藏节点采用加权聚合策略建立节点候选池,根据数据持有量测定私有模型新增节点信息权值比重,如下式:

$$\phi_L^{g_c}=\sum_{k_i\in g_c}\frac{N_{k_i}}{N^{g_c}}\phi_L^{k_i} \tag{9}$$

其中,$\phi_L^{g_c}$ 为 g_c 组内新增隐藏节点 L 的参数,N_{k_i} 为企业 k_i 的过程数据持有量,$k_i\in g_c$,$N^{g_c}=\sum_{k_i\in g_c}N_{k_i}$.

FITL 采用质量优先策略生成私有模型最优节点,即当节点满足组内所有企业约束时才将其作为最终节点生成.记企业 k_i 当前隐藏层参数的监督值输出为 $\xi_L^{k_i}$,其计算方式如公式(10)

$$\begin{aligned}\xi_L^{k_i}&=\frac{(e_{L-1}^T(X_{k_i})\cdot h_L^{k_i})^2}{(h_L^{k_i})^T\cdot h_L^{k_i}}\\&\quad-(1-r-\mu_L)e_{L-1}^T(X_{k_i})e_{L-1}(X_{k_i})\end{aligned} \tag{10}$$

公式(10)中,$e_{L-1}(X_{k_i})$ 为私有模型 k_i 在具有 $L-1$

个节点时的误差,$0<r<1, 0<u_L<(1-r)$.

设置监督记录函数为 f_{k_i},其满足公式(11).然后,在服务器设置变量 $flag$,$flag=\sum_{k_i\in K} f_{k_i}$.若 $flag$ 为 m 则选取当前的随机配置权值作为新增节点的权值.

$$f_{k_i}=\begin{cases}1, \xi_L^{k_i}>0\\ 0, \text{otherwise}\end{cases} \tag{11}$$

接着,为了实现模型的快速收敛,每个客户端采用全局更新的方式计算本地输出权值,即 $\psi^{k_i}=\underset{\psi^{k_i}}{\operatorname{argmin}}\| T^{k_i}-\sum_{j=1}^{L}\psi_j^{k_i}h_j^{k_i}\|_2^2$,$T^{k_i}$ 为企业 k_i 样本的实际标签集.最后,在每个共识组内通过公式(12)进行组内的横向加强聚合.

$$\psi^{g_c}=\sum_{k_i\in g_c}\frac{N_{k_i}}{N^{g_c}}\psi^{k_i} \tag{12}$$

加强聚合后,$\phi_L^{g_c}$ 会作为组内节点 L 的隐藏层参数固定,而 ψ^{g_c} 将作为跨组织传递的迁移学习知识以提升组外私有模型的表现.

已有横向增量建模方法采用加权聚合的方式以实现私有模型之间的横向增强,但当多个企业的过程数据分布差异较大时,输出参数难以满足多个私有模型的本地约束,使得模型的收敛速度被极大降低.此外,横向联邦也难以在数据非独立同分布下实现有效协作.因此,先前方法在收敛速度和模型准确度方面难以获得理想效果.而 FITL 通过考察模型参数分析过程数据的边缘分布和条件分布差异能有效区分不同工况,故能实现共识组织内部的有效增强,算法 2 描述了联邦共识小组内的横向增强机制.组内横向增强机制可以提高模型在特定领域的表现,满足不同共识组的需求,并有助于提升模型的收敛速度降低通信开销.

算法 2.　共识组内横向增强机制.

输入:当前私有模型的节点数目 L,私有模型 $K=\{k_1,k_2,k_3,\cdots,k_m\}$ 的参数信息 ϕ_L^K,ψ^K

输出:私有模型参数隐藏层参数聚合结果,输出权值聚合结果

1. 等待私有模型的参数配置信息同步,并上传到服务器
2. IF 当前节点数目 $L=1$
3.　　通过算法 1 分析私有模型输出信息并划分共识小组
4.　　当前共识组织 $G=\{g_1,g_2,\cdots,g_c\}$
5. END IF
6. FOR $g_i\subset G(i=1,2,\cdots,c)$
7.　　通过公式(9)计算 $\phi_L^{g_c}$
8.　　基于质量优先策略配置组内隐藏节点参数
9. END FOR
10. 返回小组隐藏节点参数到对应的客户端
11. 等待客户端上传私有模型输出权值
12. FOR $g_i\subset G(i=1,2,\cdots,c)$
13.　　通过公式 12 计算 ψ^{g_c}
14. END FOR
15. 返回小组输出层聚合参数到对应的客户端

4.4　基于迁移学习的组间协作机制

共识组织内部的横向加强主要涉及隐藏节点的参数配置,而组间信息的迁移主要在于输出权值的计算策略.在每轮组内增强后,每个共识组织输出一个组内公共模型.在共识组织建立组内公共模型后,接下来进行组间模型信息的传递.

在组间信息的传递中,FITL 需要考虑模型在专注本地的学习效果的同时将其他组织模型的有用信息融合到私有模型中.此外,值得注意的是,在模型增量构造达到理想的最优模型前,每个企业的私有模型均不是理想的目标模型.因此,在模型的增量构造阶段基于参数校准的方式传输信息的策略难以获得理想的效果.在本文中,我们基于共识迁移设计了如公式(13)的输出权值计算策略:

$$\begin{aligned}\beta^{k*}=&\underset{\beta^k}{\operatorname{argmin}}\| T^k-\sum_{j=1}^{L}\beta_j^k h_j^k\|^2\\ &+\sum_{g_c\subseteq G,k\notin g_c}K_{kc}tr\left((\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c})(\log(\frac{\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c}}{\sum_{j=1}^{L}\beta_j^k h_j^k}))^T\right)\\ &+\lambda_1\|\beta^k\|_1\end{aligned} \tag{13}$$

其中,β^k 为私有模型 k 的输出权值,K_{kc} 为根据预测效果设定的 KL 散度项参数,λ_1 为正则项参数.优化输出权值目标函数的第一部分是本地的标签误差,第二部分是本地模型和其他小组模型在本地数据上预测分布差异,第三部分为正则化项.该更新公式第一部分的目标是保证私有模型在本地的学习效果.在构建第二部分的目标函数时,本文首先基于 KL 散度度量不同组织之间的预测输出差异.在此基础上,基于迁移学习思想,最小化私有模型和其他共识组织模型在本地数据的预测结果差异,将其他共识组织模型信息融合到本地,实现组间模型信息的迁移.

在本文中,输出权值的更新通过求解优化目标(13)来获得.采用交替方向乘子法[46]求解目标函数(13),将其等价转换为公式(14)的形式写为

$$\min f(\theta)+g(\beta^k),$$

$$f(\theta)=\| T^k-\sum_{j=1}^{L}\theta_j h_j^k \|^2$$

$$+\sum_{g_c\subseteq G,k\notin g_c}K_{kc}tr\left((\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c})(\log(\frac{\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c}}{\sum_{j=1}^{L}\theta_j h_j^k}))^T\right),$$

$$g(\beta^k)=\lambda_1\|\beta^k\|\ s.t.\ \theta=\beta^k \tag{14}$$

将公式(14)进一步转换成增广拉格朗日的形式为

$$L_\rho(\theta,\beta^k,\sigma)=f(\theta)+g(\beta^k)+\sigma^T(\theta-\beta^k)$$

$$+\frac{\rho}{2}\|\theta-\beta^k\|_2^2$$

$$s.t.\ \theta=\beta^k \tag{15}$$

其中，σ 为对偶变量，ρ 为拉格朗日惩罚系数，通过下面的更新方式迭代求解目标函数(15).

$$\begin{aligned}\theta^{q+1}&=\underset{\theta}{\operatorname{argmin}}L_\rho(\theta,(\beta^k)^q,\sigma^q)\\(\beta^k)^{q+1}&=\underset{\beta}{\operatorname{argmin}}L_\rho(\theta^{q+1},\beta^k,\sigma^q)\\\sigma^{q+1}&=\sigma^q+\rho(\theta^{q+1}-(\beta^k)^{q+1})\end{aligned} \tag{16}$$

记 $\sigma_1=\sigma/\rho$，可以将上述更新方式重写为

$$\begin{aligned}\theta^{q+1}&=\underset{\theta}{\operatorname{argmin}}(f(\theta)+\frac{\rho}{2}\|\theta-(\beta^k)^q+\sigma_1^q\|_2^2)\\(\beta^k)^{q+1}&=\underset{\beta}{\operatorname{argmin}}(g(\beta^k)+\frac{\rho}{2}\|\theta^{q+1}-\beta^k+\sigma_1^q\|_2^2)\\\sigma^{q+1}&=\sigma^q+\rho(\theta^{q+1}-(\beta^k)^{q+1})\end{aligned} \tag{17}$$

(1)计算 θ

$$\begin{aligned}\theta^{q+1}&=\underset{\theta}{\operatorname{argmin}}L_\rho(\theta,(\beta^k)^q,\sigma^q)\\&=\operatorname{argmin}\|T^k-\sum_{j=1}^{L}\theta_j h_j^k\|^2\\&+\sum_{g_c\subseteq G,k\notin g_c}K_{kc}tr((\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c})(\log(\frac{\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c}}{\sum_{j=1}^{L}\theta_j h_j^k}))^T)\\&+\frac{\rho}{2}\|\theta-(\beta^k)^q+\sigma_1^q\|_2^2\end{aligned}$$

对 $L_\rho(\theta,(\beta^k)^q,\sigma^q)$ 求 θ 的偏导数可得：

$$\begin{aligned}&\nabla L_\rho(\theta,(\beta^k),\sigma^q)\\&=(h^k)^T(T^k-h^k\theta)\\&-\sum_{g_c\subseteq G,k\notin g_c}K_{kc}(h^k)^T(h^{g_c}\psi^{g_c})/(h^k\theta)\\&+\rho(\theta-\beta^q+\sigma_1^q)\end{aligned} \tag{18}$$

令 $\nabla L_\rho(\theta,(\beta^k)^q,\sigma^q)=0$ 可得 θ.

(2)计算 β

$$\begin{aligned}(\beta^k)^{q+1}&=\underset{\beta}{\operatorname{argmin}}L_\rho(\theta^{q+1},\beta^k,\sigma^q)\\&=\underset{\beta}{\operatorname{argmin}}\lambda_1\|\beta^k\|+\frac{\rho}{2}\|\theta^{q+1}-\beta^k+\sigma_1^q\|_2^2\\(\beta^k)^{q+1}&=S_{\lambda_1/\rho}(\theta^{q+1}+\sigma_1^q)\end{aligned} \tag{19}$$

其中，

$$S_k(a)=\begin{cases}a-k, & a>k\\0, & |a|\leqslant k\\a+k, & a<-k\end{cases} \tag{20}$$

通过上述交替方向乘子法迭代求解得到的输出权值等价于求解目标函数(13)获得的权值. 基于上述对优化问题的求解，设计了如算法 3 所示的私有模型本地具体训练步骤. 此外，上述优化目标通过客户端的本地迭代来得到最优解，无需额外的通信代价，该特性在实际应用当中非常理想，因为其帮助减少了联邦中的通讯过载问题.

算法 1～3 联合构成了本文的联邦增量迁移学习策略，该策略通过将模型参数上传到中央客户端实现利用不同过程数据的信息协助建模，而无需上传本地数据. 在组间知识迁移过程中，企业获取的是组内增强后的其他组模型参数，而并非私有模型实际的参数. 因此，其对于来自不可靠企业通过模型参数反推出数据信息的攻击具有安全保障.

算法 3. 共识组间迁移机制.

输入：私有模型的训练数据集 $D_k(x_i^k,y_i^k),1\leqslant i\leqslant N_k$；最大的隐藏节点数目 $L_{\max}^k$；客户端 k 的目标误差上限 ε^k；随机节点配置超参数 λ^k

输出：私有模型参数建立结果

1. WHILE $L\leqslant L_{\max}^k$ 且 $||e_0^k||_F>\varepsilon^k$
2. 在 $[-\lambda^k,\lambda^k]$ 和 $[0,\lambda^k]$ 内随机生成私有模型的隐藏节点参数 w_L^k 和 b_L^k
3. 选择最大的 ξ_L^k 对应的 w_L^k 和 b_L^k 上传到服务器
4. IF $L=1$
5. 等待服务器划分共识组织
6. END IF
7. 下载在服务器生成的 $\phi_L^{g_c},(k\in g_c)$ 并固定
8. 基于 $\phi_L^{g_c},(k\in g_c)$ 计算本地私有模型输出权值
9. 下载在服务器生成的 $\psi^{g_c},(k\in g_c)$
10. FOR $g_i\subset G(i=1,2,\cdots,c;k\notin g_c)$
11. 根据 4.4 节的更新方式公式(17－19)迭代更新 β^k
12. END FOR
13. 计算本地输出损失 $e_L^k=T^k-h^k\beta^k$
14. $e_0^k=e_L^k,L=L+1$
15. END WHILE

4.5 算法理论分析

本节以私有模型 k 的构建为例，对所提联邦增

量迁移学习方法通过算法 1－3 所构建的私有模型的收敛性进行理论分析,令

$$\beta^{k*}=[\beta_1^{k*},\beta_2^{k*},\cdots,\beta_L^{k*}]^T$$

$$=\underset{\beta^k}{\operatorname{argmin}}\|T^k-\sum_{j=1}^{L}\beta_j^k h_j^k\|^2$$

$$+\sum_{g_c\subseteq G,k\notin g_c}K_{kc}tr\left((\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c})(\log(\frac{\sum_{j=1}^{L}\psi_j^{g_c}h_j^{g_c}}{\sum_{j=1}^{L}\beta_j^k h_j^k}))^T\right)$$

$$+\lambda_1\|\beta^k\|_1,$$

$$e_L^*=T^k-h^k\beta^{k*}=[e_{L,1}^{k*},e_{L,2}^{k*},\cdots,e_{L,m}^{k*}]$$

记 $\tilde{\beta}_L=[\tilde{\beta}_{L,1},\tilde{\beta}_{L,2},\cdots,\tilde{\beta}_{L,m}]^T=\frac{<\bar{e}_{L-1}^*,\bar{h}_L>}{h_L^2}$, $\tilde{e}_L=e_{L-1}^*-\tilde{\beta}_L h_L^k$. 其中, $\bar{e}_{L-1}^*=\sum_{k\in g_c}\frac{e_{L-1}^{k*}}{N^{g_c}}$, $\bar{h}_L=\sum_{k\in g_c}\frac{h_L^k}{N^{g_c}}$, $e_0^*=T^k$.

定理 1. 定理内容 假设 $span(F)$ 在 L_2 空间中稀疏,且 $\forall h\in F$,对于 $b_h\in R+$,有 $0<\|h\|<b_h$, $0<r<1$,非负实数序列 u_L 满足 $u_L\leqslant(1-r)$ 和 $\lim_{L\to\infty}u_L=0$. 存在一个非负实数序列 ε_L 满足

$$\|e_L^*\|^2\geqslant\varepsilon_L \tag{21}$$

对于 $L=1,2,\cdots$, 给出如下定义:

$$\eta_L=\frac{\|e_{L-1}^*\|^2-\varepsilon_L}{\|e_{L-1}^*\|^2} \tag{22}$$

$$\chi=\lim_{L\to\infty}\prod_{k=1}^{L}(1-\eta_k) \tag{23}$$

$$\gamma=\lim_{L\to\infty}\prod_{k=1}^{L}(r+\mu_k)<1 \tag{24}$$

若网络的随机基函数 h_L 满足如下不等式

$$\langle e_{L-1}^*,h_L\rangle^2\geqslant b_h^2(1-r-\mu_L)\|e_{L-1}^*\|^2 \tag{25}$$

且输出权重 β 由 4.4 节的 *ADMM* 迭代解得,则

$$\sqrt{\chi}\|T^k\|\leqslant\lim\|T^k-f_L^*\|\leqslant\sqrt{\gamma}\|T^k\| \tag{26}$$

$$f_L^*=\sum_{j=1}^{L}\beta_j^{k*}h_j^k$$

证明. 由矩阵理论可知

$$\|e_L^*\|^2\leqslant\|\tilde{e}_L\|^2$$

$$=\|e_{L-1}^*-\tilde{\beta}_L h_L\|^2$$

$$\leqslant\|e_{L-1}^*\|^2\leqslant\|\tilde{e}_{L-1}\|$$

故网络残差序列 $\|e_L^*\|^2$ 是单调递减的. 据文献[46-47]可知, *KL* 散度项和 L_1 范数惩罚项的引入会导致有偏估计. 因此, $\|e_L^*\|^2\neq0$,且存在一个充分小的正实数 ε_L 使得公式(21)成立. 故可得

$$\|e_{L-1}^*\|^2-\|e_L^*\|^2\leqslant\|e_{L-1}^*\|^2-\varepsilon_L \tag{27}$$

将公式(22)带入(27)可得

$$\|e_{L-1}^*\|^2-\|e_L^*\|^2\leqslant\eta_L\|e_{L-1}^*\|^2 \tag{28}$$

进一步,将公式(28)转换为

$$1-\eta_L\|e_{L-1}^*\|^2\leqslant\|e_L^*\|^2 \tag{29}$$

由于本文采用质量优先算法生成隐藏层节点参数,因此过程数据生成的 h_L 需要满足组内所有私有模型的不等式条件. 在此基础上,结合式(25)可得

$$\begin{aligned}&\|e_L^*\|^2-(r+\mu_L)\|e_{L-1}^*\|^2\leqslant\\&\|\tilde{e}_L\|^2-(r+\mu_L)\|e_{L-1}^*\|^2=\\&\|e_{L-1}^*-\tilde{\beta}_L h_L\|^2-(r+\mu_L)\|e_{L-1}^*\|^2=\\&(1-r-\mu_L)\|e_{L-1}^*\|^2-\frac{\langle e_{L-1}^*,h_L\rangle^2}{h_L^2}\leqslant\\&(1-r-\mu_L)\|e_{L-1}^*\|^2-\frac{\langle e_{L-1}^*,h_L\rangle^2}{b_h^2}\leqslant0\end{aligned} \tag{30}$$

联立公式(29)与(30)可得

$$\chi\|e_0^*\|^2\leqslant\|e_L^*\|^2\leqslant\gamma\|e_0^*\|^2 \tag{31}$$

故可得

$$\sqrt{\chi}\|T^k\|\leqslant\lim_{L\to\infty}\|e_L^*\|^2\leqslant\sqrt{\gamma}\|T^k\|$$

证毕.

5　实验评估

5.1　数据集介绍

本节在工业数据集及公开数据集上对 FITL 方法的表现进行了系统的实验验证. 其中工业数据集来自不同企业的磨矿数据集,用于回归任务. 在该工业数据样本集中,每个样本包括 3 个维度的特征,分别为球磨机给矿量、球磨机入口给水流量、螺旋分级机溢流浓度. 输出为矿石产物中直径小于 0.074 mm 的颗粒在该产物的占比即磨矿粒度. 上述样本信息均为数值型数据,单个样本的维度规模为 4. 同一地区设备进行工业生产收集的数据视为来自同一工况的数据. 来自不同地区设备收集的样本视为不同工况的数据. 共设置了 9 个联邦学习客户端,每个企业提供 3000 个研磨粒度的过程数据样本. 数据涉及了三种不同工况下的模型学习效果,每个企业随机划分了 200 个样本进行模型学习效果测试.

此外,进一步通过公开数据集上的系统实验对 FITL 在准确度、通信次数及有效性方面的性能进行了验证. 分类实验的数据来自 UCI 机器学习知识库

和 KEEL-DATABASE，这些数据集包括 Letter、Pen、Posture 及 MNIST 数据集. 其中，Letter 数据集中包含了 26 个字母的大写像素图统计信息，共有 20000 个样本，16 个特征及 26 种输出类别. Pen 数据集中包含了 10 个类别的手写数字统计信息，共有 10992 个样本、16 个特征及 10 种输出类别. Posture 数据集是对 12 个用户的手势识别情况，共有 58000 个样本，9 个特征及 7 种输出类别. 由于实验并不着重于缺失值的探索，因此用 0 填补 Posture 缺失值. MNIST 数据集为 10 种类别的手写数字的图片像素信息. 分类实验将样本按照预设的客户端数量将数据随机划分为若干个分组，每个分组对应一个客户端的私有数据. 每组样本再按照 7∶3 的比例随机划分出训练数据和测试数据集.

由于本文主要聚焦多工况的工业制造场景，因此在实验前需要先对公共数据进行预处理以模拟不同生产工况. 实验将数据集借鉴文献[48]的数据划分方式，以样本类别为基准对原始数据进行非独立同分布的划分. 同时，不同工况的样本分布在具备差异的同时还存在交集.

实验的软件环境为在 MATLAB 2018b，PC 机配置为 Intel（R） Core（TM） i7 — 10750H，2.59 GHz CPU，16 GB RAM. 在回归实验中，模型精度采用均方根误差（RMSE）来评估，分类实验的精度采用准确度来评估. 基于本文个性化方法的设定，回归及分类任务的联邦模型效果的计算方法分别见公式(32)和(33).

$$\frac{1}{N_K}\sum_{k\in K}\mathrm{RMSE}_k \tag{32}$$

$$\frac{1}{N_K}\sum_{k\in K}\mathrm{Accuracy}_k \tag{33}$$

上式中，

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}(f(d)-y(d))^2}{n}} \tag{34}$$

$$Accuracy=\frac{|\{d:d\in D\cap f(d)=y(d)\}|}{n} \tag{35}$$

其中，$f(d)$ 表示模型对于样本 d 的预测值，$y(d)$ 为样本的实际值，n 为数据集 D 的样本总数. 此外，本文计算预测精度提升效果的方式为：精度提升＝(提升后的精度－对比精度)/对比精度×100%

5.2 准确度实验

5.2.1 回归任务

回归实验共进行了三种联邦学习的场景模拟测试，分别为单工况、多工况 1 和多工况 2 的场景.

该实验的主要目的在于验证算法相较于仅考虑同工况的联邦增量方法在协同建模时的优势. 其中，单工况的合作场景中，来自私有模型的过程数据均属于同一工况. 多工况 1 场景中过程数据来自两种不同的工况，而多工况 2 场景中过程数据来自三种不同工况. 在每种合作场景下，本节又分别进行了三组私有模型的增量实验，分别是 3 个企业、6 个企业及 9 个企业. 对比方法中，BASE 表示只采用客户端本地数据进行随机配置网络建模的方法，FSCN[22] 是采用质量优先策略的联邦随机配置网络方法. 在数据预处理中，输入和输出数据都被归一化为[－1，1]. 本文为 SCN、FSCN 及 FITL 设置了相同的参数，$\lambda=1:1:10$，$\varepsilon=0.05$.

实验结果如表 2 所示. 首先，在单工况的合作场景中，FITL 等价于使用质量优先策略的 FSCN 方法. 从单工况的三组客户端增量实验中可以发现，FITL 的效果在预测准确性及节点个数方面均优于仅用本地过程数据训练的结果. 这说明单个企业拥有的过程数据不足以训练良好的模型，而联邦合作能够通过借助其他过程数据有效提升本地预测效果. 此外，在单工况场景中，随着过程数据的增加，个性化模型在预测精度及建模所需节点个数方面也在相应改善. 其中相比 3 个企业的合作，9 个企业在预测精度方面提升了 1%，节点个数减少了 10%. 说明过程数据的增加对于模型的提升很有帮助，也证明了联邦学习对于提升企业模型预测能力的必要性.

表 2 展示了多工况 1 的合作场景下模型效果的对比情况. 可以发现，虽然该实验合作过程数据来自两个工况，但 FSCN 在 3 个企业协作联邦的模型效果方面却优于本文的 FITL 方法. 这主要是因为该工况的两个数据集虽然来自不同工况，但其分布差异较小. 虽然本文方法识别了不同的工况，但由于过程数据的差异较小，并且 3 个企业的数据量较低，因此迁移学习的效果并不显著. 但随着参与联邦的企业数据量的增加不同工况模型迁移的效果逐渐提升. 从多工况 1 下的模型预测表现可以发现，虽然本文方法在模型预测精度方面相比 FSCN 的改善并不明显，提升仅 3%. 但本文在构建模型所需节点个数方面效果显著节省 36%. 而 FSCN 的效果没有随着样本数目的增加而提升联邦模型的构建. 虽然 FSCN 的效果优于过程数据本地训练的效果，然而所需节点个数增加到基线方法的 4 倍. 该组实验说明，当过程数据是来自不同工况的样本时，FSCN 难以在多工况的环境中有效利用来自多方的过程数据

协作建立模型.并且随着过程数据的增加,全局模型也并未获得理想的促进.

表 2　FITL 在回归任务上的对比实验

	客户端数量	3		6		9	
		测试精度	节点个数	测试精度	节点个数	测试精度	节点个数
单工况	BASE	0.085	187	0.087	206	0.087	208
	F-SCN	—	—	—	—	—	—
	FITL	**0.081**	**94**	**0.080**	**86**	**0.080**	**84**
多工况 1	BASE	0.086	200	0.081	213	0.087	223
	F-SCN	0.082	**108**	0.081	400	0.080	800
	FITL	**0.079**	148	**0.079**	**111**	**0.078**	**213**
多工况 2	BASE	0.08	215	0.081	223	0.079	218
	F-SCN	0.162	400	0.176	400	0.177	400
	FITL	**0.071**	**57**	**0.072**	**108**	**0.071**	**133**

在表 2 的多工况 2 的合作场景中,相比本文的 FITL,FSCN 的效果出现了显著的下降,均下降 58%.这主要是由于多工况 2 的合作场景过程数据来自 3 种工况,并且不同的工况差异较大.此时,FSCN 难以应对多工况的企业合作场景.故即使过程数据进一步增加也并未对模型的提升有所改善.而本文的 FITL 方法仍然能够在工况具有显著差异时保持联邦合作的有效性,不仅在模型预测精度上表现优越(提升 58%),而且在节点构建个数方面也显著优于基线方法(节省 75%).这是因为 FITL 能够通过工况划分有效区分出不同工况的差异,并进一步根据工况的特点有效进行过程数据间的合作互助.

图 4 中展示了企业在多工况 1 及多工况 2 场景下对于数据拟合误差的概率密度函数(Probability Density Function, PDF).从图 4(a)—(c)中可以发现 FSCN 与 FITL 的拟合效果相近,其均能最大概率接近 0.但由表 2 可知,FSCN 所需的节点数较多,因此该场景下 FITL 的效果优于 FSCN.图 4(d)—(f)为多工况 2 的误差估计分布情况,在该场景下 FSCN 误差估计分布大部分都难以接近 0,而 FITL 构建的私有模型接近 0 的概率更大更近.该现象证明了 FITL 在该场景的优越性,并与本文对于表 2 的分析结果一致.

5.2.2　分类任务(增量式方法对比)

分类实验通过三个公开数据集进行了多工况场景的联邦实验,旨在进一步验证本文方法相对于仅考虑同工况的联邦增量方法在数据差异情况下协作建模的有效性.同样,类似上节的回归实验,本节在每个合作场景下分别进行了三组客户端的增量实验.在这三个联邦合作场景中,公开数据集的数据均经过预处理,以模拟真实应用场景中的数据非独立同分布情况.数据的分布的部分案例如图 5 所示,三组实验场景中的客户端数据分布差异程度各不相同.其中,Pen 的差异最小、Posture 第二.两者不同分组的客户端虽然有差异,但也存在很多重合的部分.Letter 数据集中不同共识组织的客户端差异最明显.

在这三组数据集上进行的联邦实验的结果如表 3 所示.在数据集 Pen 的实验中,本文的 FITL 和 FSCN 均优于基线方法,预测精度分别提升 37%和 22%,且分别节省了 47%和 54%的节点.这说明联邦学习能够通过联合多客户端数据改善模型效果.并且,FITL 在客户端的增量实验中模型预测的准确度是最优的,相较于 FCSN 提升了 15%,证明了 FITL 能够在客户端具有分布差异的情况下有效协同不同分布数据增量更优构建模型.此外,FSCN 在私有模型构建所需节点数方面优于本文方法.这是由于数据分布存在差异,后续的节点增加导致误差出现上升现象,因此其提前终止了训练.

在 Posture 数据集上的实验中,不同客户端的数据分布相比 Pen 来说更大.表 3 的实验结果中,FITL 无论是在模型效果还是构建模型所需节点个数方面均表现出优秀的结果,其相较于次优方法提升了 8%预测精度,节约了 10%的节点数.而 FSCN 的方法相比基线方法产生了很大的下降,比 BASE 方法低了 10%预测精度.这是因为相比 Pen,其难以有效联合具有更大分布差异的数据协同建模.另外,值得注意的是,相比 Pen 数据集,在 Posture 数据集上的实验中,客户端构建模型所需的节点个数增加了 1.3 倍.这主要是由于 Posture 的数据规模远大于 Pen 数据集.该现象证明了本文在引言部分的假设,即不同规模的数据所需的模型结构具有差异,而通过增量建模能根据数据特点自适应构建结构合适的模型.

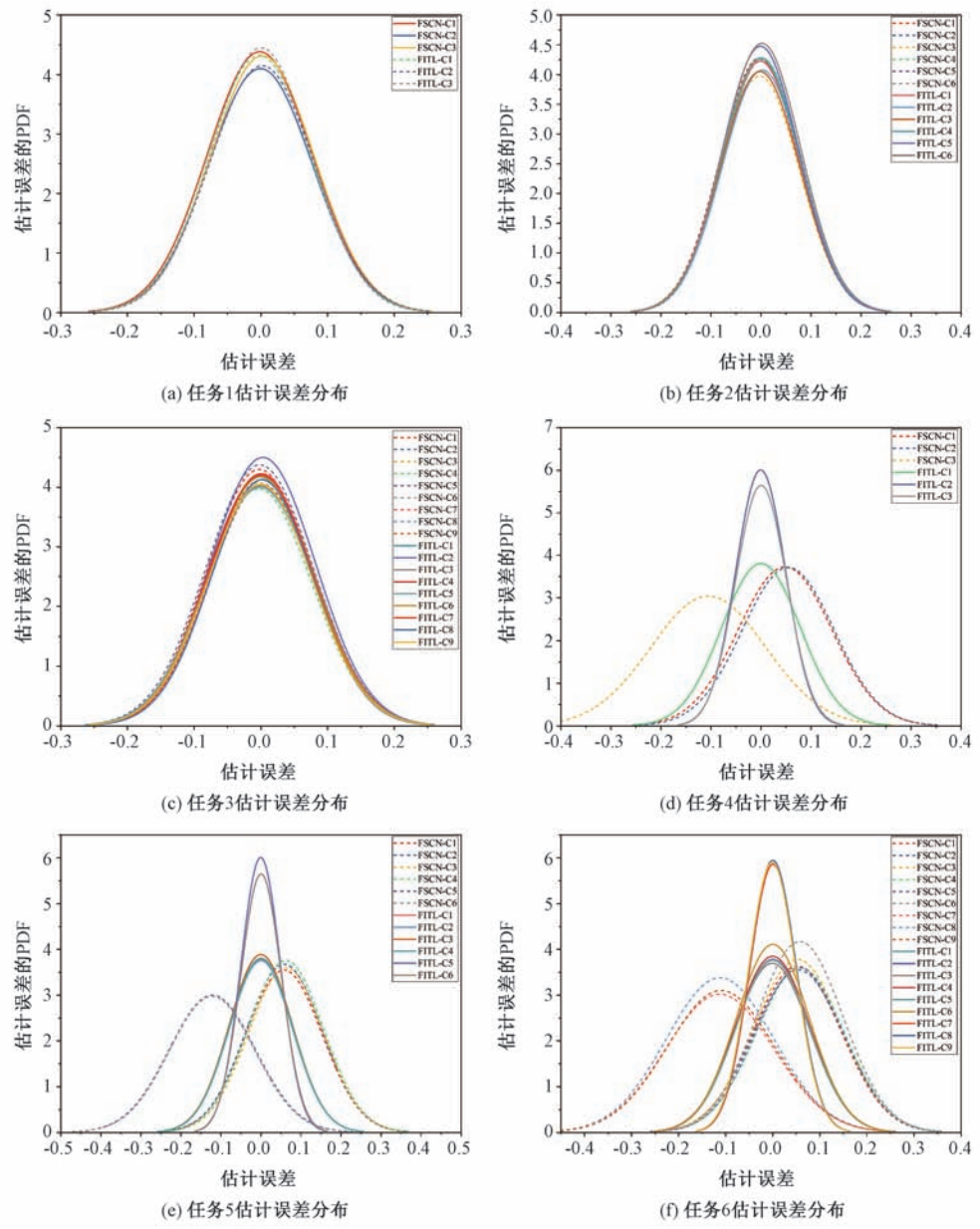

图 4 私有模型的估计误差分布情况，(a)－(c)来自多工况 1 场景，(d)－(f)来自多工况 2 场景. Method-Ci 为 Method 为企业 i 构建的模型，例：FSCN-C1 为 FSCN 为企业 1 构建的模型

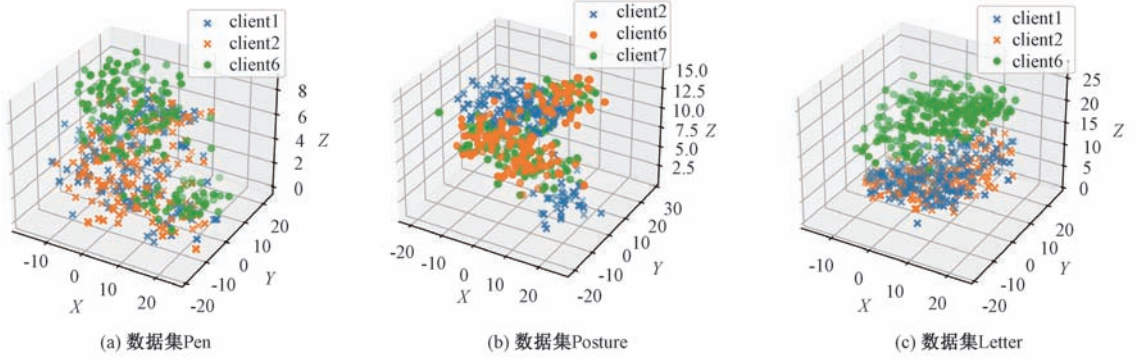

图 5 客户端数据分布情况，(a)和(c)的客户端 1 和 2 属于同一共识小组，客户端 6 来自另一组，(b)的客户端 6 和 7 属于同一组，客户端 2 属于另一组，不同的小组对应不同的工况

表 3　FITL 在分类任务上的对比实验

	客户端数量	3		6		9	
		测试精度	节点个数	测试精度	节点个数	测试精度	节点个数
Pen	BASE	0.630	720	0.710	723	0.683	727
	F-SCN	0.803	**331**	0.799	**333**	0.820	317
	FITL	**0.915**	383	**0.922**	462	**0.923**	**306**
Postures	BASE	0.825	967	0.831	1030	0.837	987
	F-SCN	0.726	992	0.774	990	0.745	923
	FITL	**0.889**	**844**	**0.895**	**873**	**0.912**	**971**
Letter	BASE	0.890	709	0.892	699	0.892	697
	F-SCN	0.619	**181**	0.623	**395**	0.625	**432**
	FITL	**0.921**	511	**0.933**	483	**0.941**	497

进一步,在数据分布差异更大的 Letter 实验中,可以发现,相比基线方法,FSCN 方法同样出现了大幅的下降,均预测精度下降 30%,并且在较少节点时就结束了训练.这主要是由于协同差异性数据对于构建的模型产生了负面的影响,导致构建的全局模型效果远低于本地数据训练的私有模型.而本文的 FITL 在该场景中仍然取得最优的效果.相比基线方法,FITL 在模型效果提升及构建模型所需节点方面效果显著,均预测精度提升 5%,均节约节点数 30%.证明了 FITL 能够在分布差异较大的情况下有效协同具有差异性的数据增量建模.

通过对回归及分类任务中 FITL 相对于基线方法的精度提升进行算数平均(仅统计多工况场景),可得 FITL 在回归任务上提升了 9%的预测精度,分类任务上提升了 16%的预测精度,证明了 FITL 在多工况场景的优越性。同时,结合表 2、表 3 在回归及分类任务上的客户端增量实验结果可以发现,随着数据样本个数的增加,FITL 能够有效利用更多的数据信息提升模型的学习效果.通过联合更多客户端的数据,FITL 在节点个数或模型准确度方面进行了提升.FSCN 并未随着数据量的增加对模型的学习效果进行提升,这是因为 FSCN 难以在数据分布存在差异下有效联合多方数据.因此,即使进一步增加客户端数据的数量,FSCN 也难以将其用于模型效果的促进.上述实验结果证明,FITL 能够通过建立共识组织,按照数据特点有效融合多个参与方的信息,实现数据分布差异下的联邦互助.

5.2.3　分类任务(非增量式方法对比)

本节将 FITL 与传统的联邦学习方法(非增量型)作比较,主要目的在于对比本文结构和参数同时自适应调整的方法相较于基于固定架构神经网络的联邦学习在数据有效利用方面的优势.

实验在使用 MNIST 作为基准数据集前对其进行了预处理操作以模拟更具难度的学习场景.本文按照样本类别将数据进行了非独立同分布的划分.首先,实验为前 5 个客户端分配了 5、6、7 三种特有类别,而后 5 个客户端分配了 8、9、10 三个特有类别.基于此,实验又按照个性化领域常采用的狄利克雷分布打散了现有客户端.此外,为了拟合特征分布变化,实验还对前五个客户端的数据集进行了像素图旋转操作.实验选取的对比方法包括 Fedavg[10]、CFL[16]、FedBN[32]、Fedprox[15]、FedPer[31]、Fed-DF[37]、RHFL[34].文章中对比的模型采用的是双层的神经网络,第一层有 1024 个节点,第二层有 512 个节点.由于 FedBN 涉及到归一化层操作,因此该方法的模型比其他对比方法多了本地的归一化层结构.需要注意的是,采用其他结构的神经网络或许能让上述方法发挥更好的效果,但本文在该实验中主要研究结构的变换对于模型优化的影响.因此,为这些网络选取最优的结构不是本文的讨论范围.

实验结果如表 4 所示,在准确度方面,该合作场景的联邦建模中 FITL 最优,相较于次优方法提升了 2%预测精度.FedBN 和 RHFL 的预测表现其次.CFL 没有取得理想效果的原因是其主要适用于特征分布差异的场景,而本文的数据分布条件分布差异较大.由于 CFL 没有有效的从客户端的梯度更新方向中区分客户端的条件分布差异,故没有取得理想效果.而 Fedprox 虽然考虑了客户端的分布差异,但其构建的全局模型无法适应多分布的数据特点,因此效果也不理想.此外,FedBN、FedPer 的效果和 FITL 的效果相近,这主要是因为其为局部模型设置了个性层,通过结合全局聚合信息和个性化信息提升了模型对于分布差异信息的表现.但由于没有在初始时获取一个合适的网络结构,因此效果

也不如 FITL. FedDF 和 RHFL 方法考虑了模型的个性化更新，能实现模型异构场景下的联邦协作，但其信息融合依赖于人工选取的公共数据集. 从模型架构方面，在该场景的联邦模型构建中，FITL 通过增量构建的方式配置了 832 个节点，仅具有约一半的其他模型规模. 通过模型结构与参数的并行优化及分布差异下的个性化建模，本文在准确度及架构规模方面均优于所对比的方法.

表 4 FITL 与非增量式方法的对比实验

企业编号 \ 方法	BASE	Fedavg	CFL	FedBN	Fedprox	FedPer	FedDF	RHFL	FITL
1	0.719	0.720	0.851	0.825	0.717	0.800	0.825	0.804	**0.890**
2	0.660	0.763	0.792	0.909	0.767	0.910	0.913	**0.932**	0.843
3	0.762	0.729	0.807	0.891	0.727	0.859	0.870	0.853	**0.911**
4	0.486	0.752	0.737	0.833	0.748	0.837	0.841	0.866	**0.920**
5	0.704	0.808	0.802	0.842	0.806	0.828	0.840	**0.889**	0.821
6	0.747	0.678	0.543	0.812	0.682	0.741	0.776	0.808	**0.875**
7	0.567	0.700	0.724	**0.887**	0.699	0.867	0.881	0.886	0.835
8	0.592	0.773	0.862	0.827	0.773	0.813	0.842	**0.878**	0.867
9	0.664	0.683	0.703	0.812	0.690	0.753	0.764	0.794	**0.864**
10	0.756	0.801	0.840	0.879	0.799	0.868	0.890	**0.891**	0.876
avg	0.665	0.741	0.766	0.852	0.741	0.828	0.844	0.860	**0.870**

综上所述，上述这些方法虽然为客户端建立个性化的模型，但不能根据分布的差异有针对性地利用数据，从而难以有效使用客户端提供的数据信息. 并且上述方法也无法根据数据特点构建合理的模型结构，故其难以实现存储资源和模型构建效果的理想结合. 而 FITL 一方面能结合数据特点增量式为客户端构建结构合适的模型. 另一方面，通过建立共识组织，FITL 能根据数据分布差异充分利用数据信息. 因此，在该场景中 FITL 能有效联合不同分布的客户端合作.

5.3 有效性实验

本部分的实验主要验证了 FITL 中对于分组学习及迁移学习方法的有效性，实验分别随机选取了多工况场景下的四组回归任务和分类任务. 实验结果如图 6 所示，其中 FITL-G 为建立共识组织后没有进行联邦迁移学习的策略，FITL-T 为没有分组直接进行联邦迁移学习的策略.

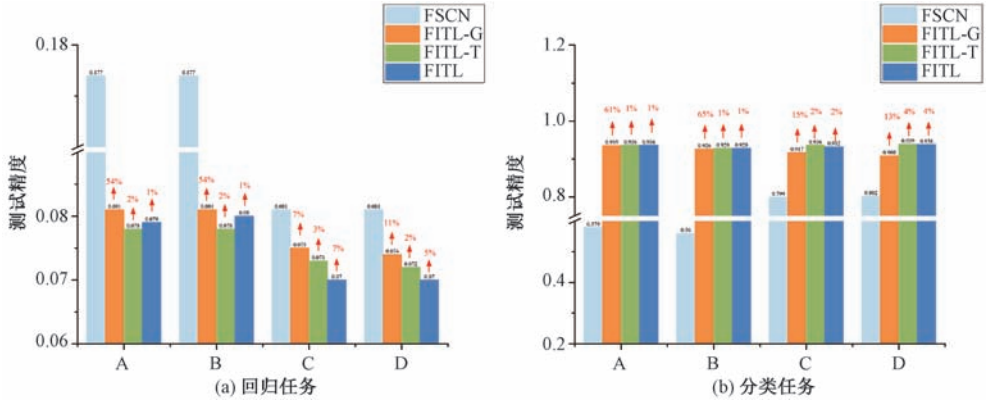

图 6 有效性实验，A、B、C、D 为两组学习场景下随机抽取的 4 个不同学习任务，A、B 任务中参与方的数据差异相对 C、D 更大

建模策略消融实验分析. 通过将 FITL-G、FITL-T、FITL 与 FSCN 在随机选取的差异工况任务上进行对比，本实验对算法在差异工况条件下的性能稳定性和效率方面的优势进行了验证. 图 6(a) 为在工业数据上进行回归任务的实验结果，图 6(b) 为分类数据上的实验结果. A、B、C、D 为随机选取的多工况环境下 6 个客户端协作建模任务. 首先，分析 FSCN 方法与 FITL 系列方法的效果差异. 从图 6 可以看出 FITL 系列方法均优于 FSCN 方法. 其原因主要在于 FITL 系列方法考虑了联邦合作时的差异分布数据情况，并分别通过分组、迁移学习实现差异工况下的知识共享. 因此，相比具有同分布假设的 FSCN，FITL 系列方法所构建的模型准确度有显著提升，均预测精度提升 13%.

然后,分析 FITL 系列方法间的效果差异. 在图 6 中,对比小组单独联邦学习 FITL-G 和小组间进行迁移互助的 FITL 实验效果,可以发现 FITL 进行知识迁移后的效果比单独学习每个小组 FITL-G 的学习效果在均预测精度提升了 3%. 这说明了本文迁移学习方法能够通过拟合预测分布有效地将其他小组的信息迁移到本地,实现本地模型的优化. 值得注意的是,在差异性较大的 A、B 两组任务中 FITL 相比 FITL-G 的效果提升并不明显,仅为 1%. 这是因为两组数据信息差异较大,迁移学习难以在差异较大的情况下实现互助. 而在 C、D 两组实验中,客户虽有差异但仍存在很多相似性所以 FITL 能够比 FITL-G 的效果出现明显提升,均预测精度提升 4.5%. 接着,对比 FITL-T 与 FITL 的实验效果. 从图 6 中可以发现,两者在模型预测精度方面的差异并不大. 而在图 7 的运行时间对比中,可以发现 FITL-T 的运行时间均高于其他方法. 这是因为,FITL-T 将所有数据均视为差异数据,忽视了数据间的相似性,其要学习的来自其他领域的模型较多. 本文中,通过层次聚类进行小组的划分获取数据相似信息,根据经验建议阈值大于 0.8 更有助于横向增强的实现,如果计算资源更充分可以设置更高的阈值.

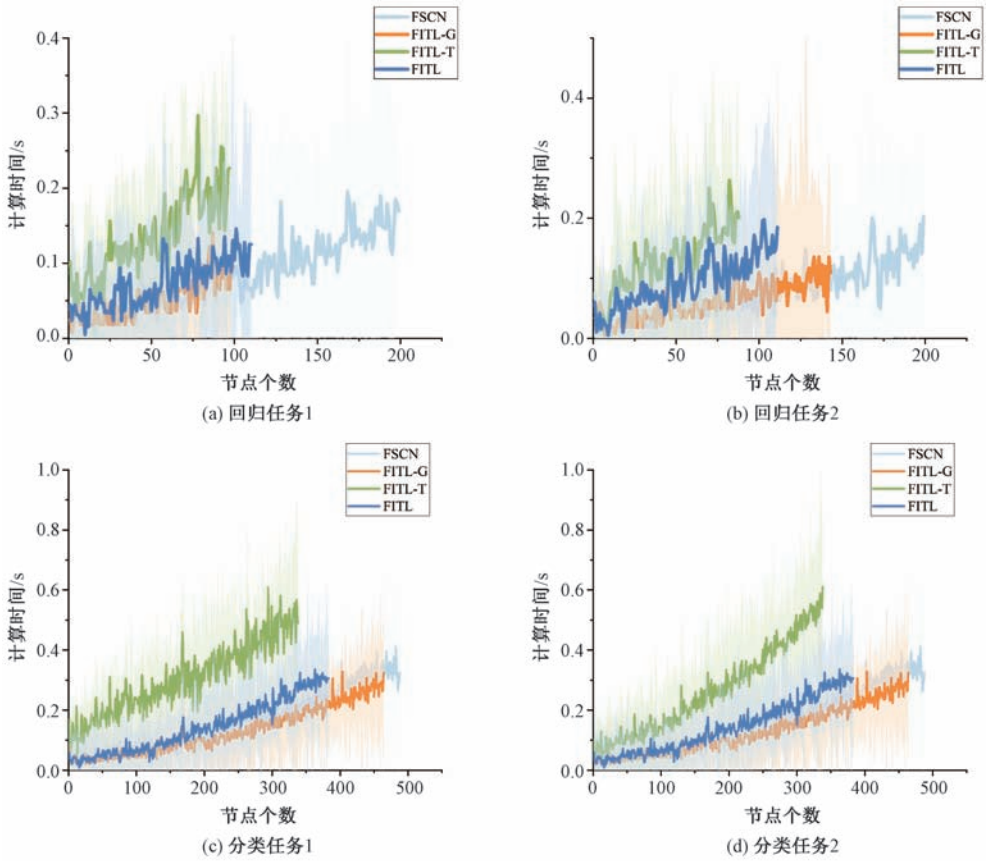

图 7　运行时间对比

节点增量下的建模效果分析. 通过控制节点个数的增加,本实验对节点数量变化对于模型效果的影响进行了验证. 表 5 以分类任务为例,具体展示了随着节点个数增加模型效果的变化情况. 由于本文建立的是个性化的私有模型,因此表格中详细展示了 9 个私有模型的具体准确度变化情况. 此外,为了展示节点增加对于两种模型训练的学习效果影响差别,采用了固定节点个数对比学习效果的展示. 其主要目的是想对比相同节点个数下两个模型学习效果的差异. 可以发现随着节点数量的增加,模型的效果在不断提升. 说明神经网络模型的架构对于模型效果影响很大,构建一个结构合适的模型对于节约资源和提升模型效果很有意义.

全局模型与个性化模型分析:通过对比私有模型和公共模型策略中每个客户端的具体预测表现,本实验对算法个性化建模策略的兼容性进行了验证. 从横轴的角度分析表 5,可以发现采用 FSCN 构建的全局模型在企业 6－9 上并未取得理想的效果. 这是因为该合作场景中企业 6－9 的数据分布与 1－5 的数据分布存在差异. FSCN 采用的全局模型策略难以兼容具有差异分布的数据学习,因此未能

获得理想的预测效果.相比之下,FITL 采用个性化建模策略,结合工况特性为不同分布的学习设计了组内、组间差异化的增量协作方式,实现了对于不同工况企业协作的兼容.所以,个性化建模策略对于不同工况的客户端合作有更强的兼容性,这一性质对于提升联邦模型效果具有重要意义.

表 5 节点增量下的模型效果对比

方法	企业编号 / 节点个数	1	2	3	4	5	6	7	8	9	AVG
FSCN	100	0.75	0.69	0.73	0.75	0.73	0.40	0.42	0.44	0.39	0.59
	200	0.75	0.75	0.78	0.77	0.73	0.40	0.41	0.41	0.40	0.60
	300	0.73	0.76	0.73	0.75	0.73	0.39	0.37	0.40	0.41	0.58
	400	0.73	0.76	0.73	0.75	0.73	0.39	0.37	0.40	0.41	0.58
	500	0.77	0.77	0.78	0.79	0.77	0.44	0.41	0.47	0.43	0.63
FITL	100	0.86	0.89	0.87	0.86	0.89	0.87	0.85	0.85	0.88	0.87
	200	0.90	0.91	0.91	0.89	0.92	0.91	0.91	0.90	0.93	0.91
	300	0.92	0.93	0.91	0.91	0.93	0.93	0.93	0.91	0.94	0.92
	400	0.92	0.93	0.92	0.91	0.94	0.96	0.95	0.93	0.95	0.93
	500	0.93	0.93	0.93	0.92	0.94	0.96	0.95	0.94	0.95	0.94

5.4 通信次数

本节实验对 FITL 在缓解由于分布差异造成的 SCN 收敛困难方面的优势进行了验证.实验主要对比了相同客户端数据下 FITL 和 FSCN 的通信次数情况.如文中所述,FSCN 在质量优先情况下很难满足既定的监督机制,极大影响了模型的收敛速度.而联邦增量迁移学习通过将客户端进行分组,缓解了监督机制难以满足的问题并保证了模型的收敛性.

图 8 是两组方法在 6 个客户端执行分类任务情况下与中央服务器的通信次数比较情况.可以发现,通过建立联邦共识组织,FITL 极大地降低了模型的通信次数,并且其通信次数随着模型的训练逐步趋于稳定而无需额外的通信以满足全部的监督机制.图 9—10 是 FITL 和 FSCN 的准确度和误差变化曲线,可以发现随着节点数目的增加 FITL 的模型准确度和误差能够在一定的迭代次数内收敛,且其表现均优于 FSCN.

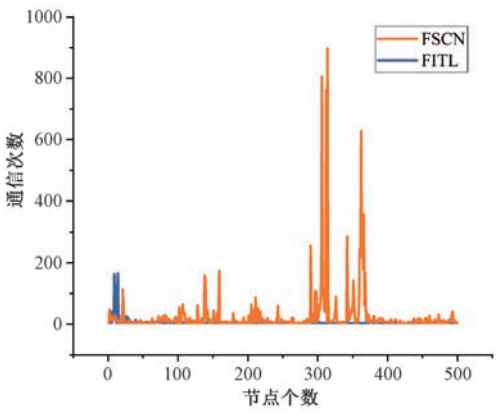

图 8 相同客户端数据下 FITL 与 FSCN 的通信次数情况对比

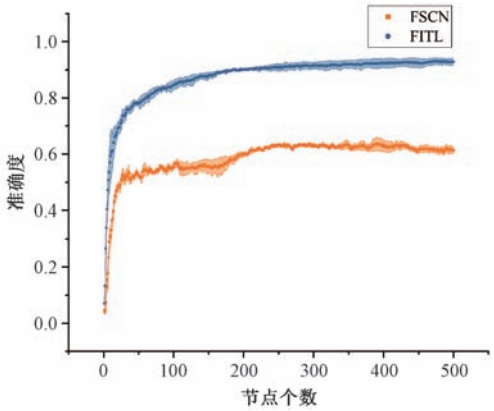

图 9 FITL 与 FSCN 的准确度收敛情况对比

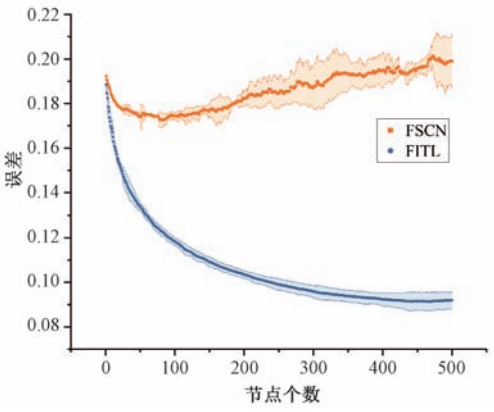

图 10 FITL 与 FSCN 的误差收敛情况对比.

此外,在本文的 FITL 中,进行知识的迁移需要借助其他组的集中模型.关于单次通信量的对比,可以分为两种情况讨论.首先,在所有企业来自同工况的情况下.FITL 只有一个共识小组,则 FITL 与 FSCN 的单次通信量是相同的.其次,当工况的差异性较复杂的情况下,此时 FITL 会建立多个共识小

组进行知识迁移.在该合作条件下 FITL 的单次通信量会超过 FSCN,但此时 FSCN 也会由于难以找到满足监督条件的节点进行频繁通信,故整体通信量并不会低于 FITL.更重要的是,FSCN 虽然进行了频繁的通信,但其也由于无法适应个性化的本地数据导致难以获得理想效果.

6 结　论

本文对数据安全下的分布式协同建模问题进行了研究.针对分布差异下难以实现增量式联邦建模的问题,提出了联邦增量迁移学习方法.FITL 首先结合随机配置网络特性对私有模型输出参数进行分析,建立了多个具有共识的联邦组织.然后,基于数据分布特点分别建立了基于组内横向增强和跨组织知识迁移的联邦合作机制.最后,结合组内增强和组间迁移构建基于数据分布特点的联邦互助策略,实现了数据分布差异下的联邦增量式建模.实验证明,该方法能有效降低工况差异对于协同建模的影响,解决了先前的联邦增量学习方法在多工况建模时难以取得理想效果的问题.

本文所提方法以增量构造式网络为客户端学习模型,对客户端采用增量构造式网络的联邦学习具有一定的适用性.在这个方法中,虽然随机配置网络参数是随机生成的,但通过监督约束,确保了学习的收敛性.当本文方法扩展至其他增量模型时,其收敛性需要进一步的探讨.此外,由于企业竞争的存在,难以保证企业上传信息的有效性为前提和可靠性,而现有方法均以企业诚实可靠为前提.因此,其模型的可靠性难以保证,后续将对模型的可靠性问题进行进一步的研究.

参　考　文　献

[1] Dai W, Li D, Zhou P, et al. Stochastic configuration networks with block increments for data modeling in process industries. Information Sciences, 2019, 484: 367-386

[2] Yin S, Li X, Gao H, et al. Data-based techniques focused on modern industry: An overview. IEEE Transactions on Industrial Electronics, 2015, 62(1): 657-667

[3] Zhao J G, Yang C Y. Non-cascade dual-rate composite decentralized operational optimal control for complex industrail processes. Acta Automatica Sinica, 2023, 49(1): 172-184 (in Chinese)
(赵建国, 杨春雨. 复杂工业过程非串级双速率组合分散运行优化控制. 自动化学报, 2023, 49(1): 172-184)

[4] Zhou P, Lu S W, Chai T. Data-driven soft-sensor modeling for product quality estimation using case-based reasoning and fuzzy-similarity rough sets. IEEE Transactions on Automation Science and Engineering, 2014, 11(4): 992-1003

[5] Zhang L, Zhou P, Song H D, et al. Multivariable dynamic modeling for molten iron quality using incremental random vector functional-link networks. Journal of Iron and Steel Research, International, 2016, 23(11): 1151-1159

[6] Dai W, Li D P, Yang C Y, et al. A model and data hybrid parallel learning method for stochastic configuration networks. Acta Automatica Sinica, 2021, 47(10): 2427-2437(in Chinese)
(代伟, 李德鹏, 杨春雨,等. 一种随机配置网络的模型与数据混合并行学习方法. 自动化学报, 2021, 47(10): 2427-2437)

[7] Yuan X, Ge Z, Huang B, et al. Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR. IEEE Transactions on Industrial Informatics, 2017,13(2):532-541

[8] Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19

[9] Gu T L, Li L, Chang L, et al. Fair federated machine learning and its design: A comprehensive survey. Chinese Journal of Computers,2023, 46(9): 1991-2024(in Chinese)
(古天龙, 李龙, 常亮, 等. 公平联邦学习及其设计研究综述. 计算机学报, 2023, 46(9): 1991-2024)

[10] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data//Proceedings of the Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282

[11] Pan Z, Geng H, Wei L, et al. Adaptive client model update with reinforcement learning in synchronous federated learning//Proceedings of the 32nd International Telecommunication Networks and Applications Conference (ITNAC). Wellington, New Zealand, 2022: 1-3

[12] Liu Y, Wang T, Peng S L, et al. Edge-based model cleaning and device clustering in federated learning. Chinese Journal of Computers, 2021, 44(12): 2515-2528(in Chinese)
(刘艳, 王田, 彭绍亮,等. 基于边缘的联邦学习模型清洗和设备聚类方法. 计算机学报, 2021, 44(12): 2515-2528)

[13] Li Z P, Guo Y, Chen Y F, et al. Class-balanced federated learning basedon data generation. Chinese Journal of Computers, 2023, 46(3): 609-625(in Chinese)
(李志鹏, 国雍, 陈耀佛,等. 基于数据生成的类别均衡联邦学习. 计算机学报, 2023, 46(3): 609-625)

[14] Zhuang F, Luo P, Xiong H, et al. Cross-domain learning from multiple sources: A consensus regularization perspective. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(12): 1664-1678

[15] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks//Proceedings of the Machine Learning and Systems, 2020, 2: 429-450