doi:10.20149/j.cnki.issn1008-1739.2024.04.002
引用格式:陈家良,冯金顺,董少然,等.利用联邦学习技术促进大数据中心的集成应用[J].计算机与网络,2024,50(4):289-295.[CHEN Jialiang, FENG Jinshun, DONG Shaoran, et al. Integrated Application of Big Data Center Using the Technology of Federated Learning[J]. Computer and Network,2024,50(4):289-295.]

利用联邦学习技术促进大数据中心的集成应用

陈家良，冯金顺，董少然，郭新苍，范烁晨，朱光耀，高　静，马胤垚
（中国电科网络通信研究院，河北 石家庄 050081）

摘　要：基于开源联邦学习框架，构建支持横向联邦学习、纵向联邦学习以及联邦迁移学习的联邦学习平台，提供从数据接入到模型部署的一站式开发能力，简化联邦学习模型的开发和上线工作，降低使用的门槛。联邦学习平台能够有机地融入现有大数据中心，结合现有的数据资源分布特点，灵活运用平台的角色分配机制，创建有效的联邦学习模型。联邦学习平台及其部署方案能够打破现在大数据中心独立建设带来的“数据孤岛”局面，全方位提升大数据平台的使用效率。

关键词：联邦学习；大数据中心；横向联邦学习；纵向联邦学习；联邦迁移学习
中图分类号：TP399　　**文献标志码**：A　　**文章编号**：1008-1739(2024)04-0289-07

Integrated Application of Big Data Center Using the Technology of Federated Learning

CHEN Jialiang, FENG Jinshun, DONG Shaoran, GUO Xincang, FAN Shuochen, ZHU Guangyao, GAO Jing, MA Yinyao
(*Academy for Network & Communications of CETC*, *Shijiazhuang* 050081, *China*)

Abstract: Based on the open-source federated learning framework, the federated learning platform is built to support horizontal federated learning, vertical federated learning and federated transfer learning, providing one-step development capabilities from data access to model deployment, simplifying the development and deployment of federated learning models and reducing the threshold for use. The federated learning platform can be organically integrated into the existing big data center. With the distribution characteristics of the existing data resources combined, the role allocation mechanism is flexibly used, and the effective federated learning model is created. The federated learning platform and the deployment solution are able to break the "data island" situation brought by the independent construction of the big data center, comprehensively improving the efficiency of the big data platform.

Keywords: federated learning; big data center; horizontal federated learning; vertical federated learning; federated transfer learning

0　引言

21 世纪以来，云计算、大数据、人工智能等信息技术在民用领域发展迅猛，得到了广泛应用。大数据中心的建设正如火如荼地展开，业务中台数据采集量、存储量巨大，为人工智能技术在实际业务中的应用提供了数据基础。但是各数据中心之间又存在着“数据孤岛”的问题，打破数据隔离壁垒，建立跨专业、跨领域的紧密协作、集中管理的数据共享机制成为急需解决的问题。

联邦学习是一种机器学习方法，最早在 2016 年由谷歌公司提出，使用移动设备上的数据，不转移、交换数据，实现机器学习模型的训练。香港科技大学杨强教授团队将联邦学习概念扩展为隐私保护的分布式协作机器学习技术，提出了机器学习的基本分类，主要包括横向联邦学习、纵向联邦学习、联邦迁移学习等[1]。

针对现有部分数据中心“数据孤岛”问题，基于 Federated AI Technology Enabler（FATE）联邦学习框架构建的平台，充分运用横向联邦学习、纵向联邦学习和联邦迁移学习，探索联邦学习技术在大数据集成应用中的模式，促进数据的集成应用，提高数据的利用率。

1　联邦学习

联邦学习是一种机器学习范式[2]，可以在一个中央服务器的协调下让多个客户端互相合作，在原

收稿日期:2024-07-11

始数据不离开各客户端的情况下，得到一个完整的机器学习模型。联邦学习的 2 个核心主题是中央服务器和客户端。其中，中央服务器负责宏观上协调各客户端，客户端则负责在本地训练机器学习模型。

依据数据的分布特点和数据拥有者之间的差异，联邦学习主要分为横向联邦学习、纵向联邦学习和联邦迁移学习[1]。下文分别介绍这 3 种联邦学习方法。

1.1 横向联邦学习

横向联邦学习也称按样本划分的联邦学习[3]，它的每个参与者之间数据源特征相同，数据分布不同，适用于参与者间业态相同、但客户不同的情形，即特征重叠多、用户重叠少的场景。

以卫星样本为例，不同用户之间通信卫星不同，但转发器、信道等特征相同，即样本重叠少、特征重叠多。横向学习的本质是样本的联合，可以应用于联邦学习的各个参与方的数据集有相同的特征空间和不同样本空间的场景，其特征维度空间如图 1 所示。

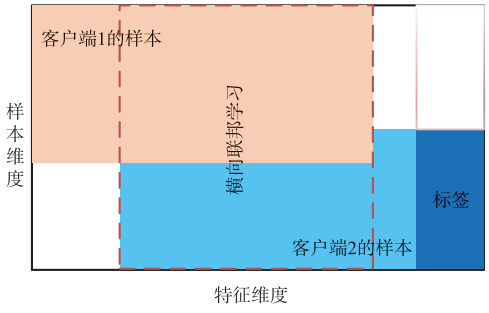

图 1 横向联邦学习样本特征

以基于深度训练的神经网络为例，横向联邦学习过程如图 2 所示。

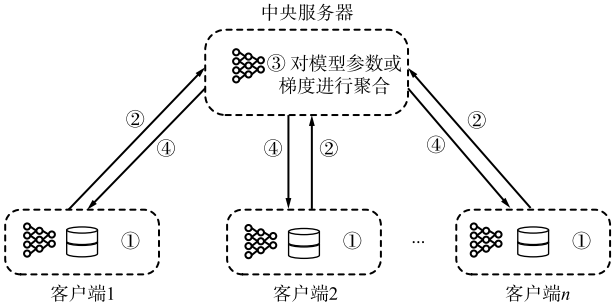

图 2 横向联邦学习过程

① 所有客户端分别独立地基于本地数据进行训练；

② 客户端对模型参数或者模型参数的梯度进行加密，上传到服务器；

③ 服务器对搜集到的客户端模型参数或者梯度进行聚合；

④ 服务器将模型下发到各个客户端；

⑤ 重复步骤①~④，直至训练结束。

在横向联邦学习的过程中，每个客户端机器都是相同且完整的模型，相互之间不会产生交流和依赖，可以把这个过程看作基于样本的分布式模型训练，最终训练结果模型均能够完成独立的分类预测。

1.2 纵向联邦学习

纵向联邦学习适用于参与方数据集具有相同的样本空间和不同的特征空间的联邦学习场景，也可以理解为按特征划分的联邦学习[4]。纵向联邦学习的本质是交叉用户在不同业态下的特征联合。

以通信卫星数据为例，大数据中心甲和乙有相同卫星的数据，但中心甲只有转发器数据，中心乙只有信道数据，可通过纵向联邦学习来训练模型，其特征维度空间如图 3 所示。

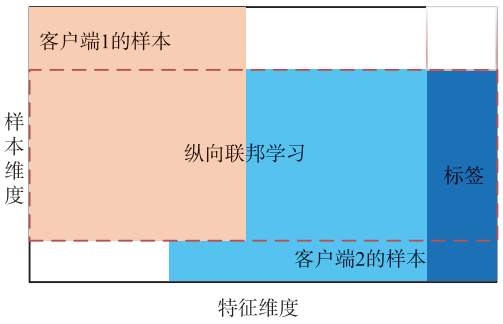

图 3 纵向联邦学习样本特征

纵向联邦学习过程包括加密样本对齐，选择同一批次数据并对齐样本进行模型加密训练，如图 4 所示。

① 第三方中央服务器向客户端 1 和 2 发送公钥，用来加密需要传输的数据；

② 客户端 1 和 2 分别计算和自己相关的特征中间结果，并加密交互，用来求得各自梯度和损失；

③ 客户端 1 和 2 分别计算各自加密后的梯度并添加掩码发送给中央服务器，同时客户端 2 计算加密后的损失发送给中央服务器；

④ 中央服务器解密梯度和损失后回传给客户端 1 和 2，它们去除掩码后更新模型；

⑤ 重复步骤①~④，直至训练结束。

由于各参与方只能得到与自己相关的参数模型，在进行预测时，需要双方协作，共同完成。

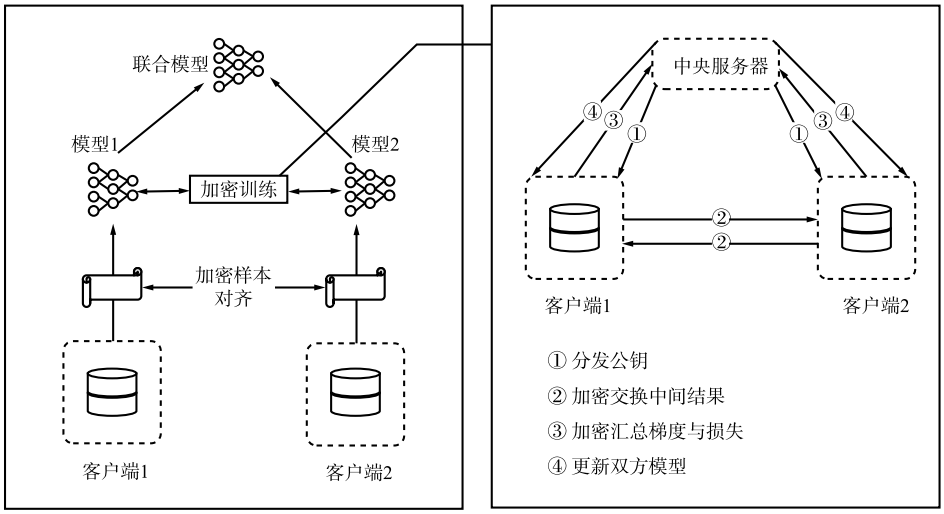

图 4　纵向联邦学习过程

1.3　联邦迁移学习

联邦迁移学习是一种特殊的迁移学习，适用于 2 个数据集之间用户特征向量和用户都很少重叠的情况[5]。它的本质是发现资源丰富的源域和资源稀缺的目标域之间不变性和相似性，并利用该不变性在 2 个领域之间传输知识。

以卫星数据为例，数据在某些业务场景下并没有本质的区别，可以通过迁移联邦学习把样本的特征迁移到其他领域，得到迁移模型，用于结果预测，其特征维度空间如图 5 所示。

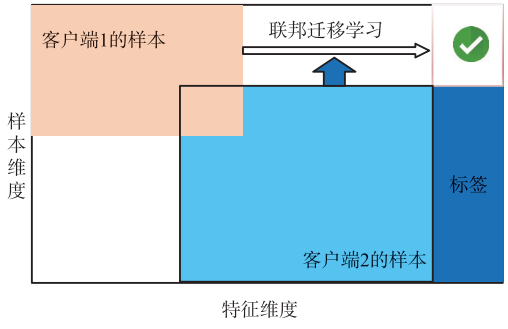

图 5　联邦迁移学习样本特征

联邦迁移学习是纵向联邦学习的一种特例，与纵向联邦学习的相同点在于数据特征维度重叠部分较少，而联邦迁移学习面临的情况更加苛刻。由于用户特征维度重叠部分少，2 份数据只是在某种程度上处于相同类别，联邦迁移学习就是为了寻找这种共同点而提出的。

联邦迁移学习的源域 A，用 $\boldsymbol{D}_{\mathrm{A}}$ 表示：

$$\boldsymbol{D}_{\mathrm{A}}=\{(x_i^{\mathrm{A}},y_i^{\mathrm{A}})\}_{i=1}^{N_{\mathrm{A}}},\tag{1}$$

式中：x_i^{A}、y_i^{A} 为源域 A 的样本标签，N_{A} 为源域 A 的样本数量。

目标域 B，用 $\boldsymbol{D}_{\mathrm{B}}$ 表示：

$$\boldsymbol{D}_{\mathrm{B}}=\{(x_j^{\mathrm{B}})\}_{j=1}^{N_{\mathrm{B}}},\tag{2}$$

式中：x_j^{B} 为目标域 B 的样本标签，N_{B} 为目标域 B 的样本数量。

假设源域 A 和目标域 B 之间存在共同样本 $\boldsymbol{D}_{\mathrm{AB}}$：

$$\boldsymbol{D}_{\mathrm{AB}}=\{(x_i^{\mathrm{A}},x_i^{\mathrm{B}})\}_{i=1}^{N_{\mathrm{AB}}},\tag{3}$$

式中：N_{AB} 为共同样本 $\boldsymbol{D}_{\mathrm{AB}}$ 的数量，源域和目标域具有相同的样本标签 x_i^{A}、x_i^{B}。

对于其共同样本存在 $\boldsymbol{D}_{\mathrm{C}}$：

$$\boldsymbol{D}_{\mathrm{C}}=\{(x_i^{\mathrm{B}},y_i^{\mathrm{A}})\}_{i=1}^{N_{\mathrm{C}}},\tag{4}$$

式中：x_i^{B}、y_i^{A} 为样本标签，N_{C} 为共同样本 $\boldsymbol{D}_{\mathrm{C}}$ 的数量。

$\boldsymbol{u}_{\mathrm{A}}$、$\boldsymbol{u}_{\mathrm{B}}$ 分别为源域和目标域间的隐层特征不变量，定义对目标域的分类函数 $\varphi(\boldsymbol{u}_j^{\mathrm{B}})$ 为：

$$\varphi(\boldsymbol{u}_j^{\mathrm{B}})=\frac{1}{N_{\mathrm{A}}}\sum_i^{N_{\mathrm{A}}}y_i^{\mathrm{A}}\boldsymbol{u}_i^{\mathrm{A}}(\boldsymbol{u}_j^{\mathrm{B}})'=\boldsymbol{\Phi}^{\mathrm{A}}\boldsymbol{\Omega}(\boldsymbol{u}_j^{\mathrm{B}}),\tag{5}$$

式中：$\boldsymbol{\Phi}^{\mathrm{A}}=\frac{1}{N_{\mathrm{A}}}\sum_i^{N_{\mathrm{A}}}y_i^{\mathrm{A}}\boldsymbol{u}_i^{\mathrm{A}}$，$\boldsymbol{\Omega}(\boldsymbol{u}_j^{\mathrm{B}})=(\boldsymbol{u}_j^{\mathrm{B}})'$。

目标函数 L_1、L_2 分别为：

$$\underset{\boldsymbol{\Theta}^{\mathrm{A}},\boldsymbol{\Theta}^{\mathrm{B}}}{\operatorname{argmin}}L_1=\sum_i^{N_{\mathrm{C}}}l_1(y_i^{\mathrm{A}},\varphi(\boldsymbol{u}_i^{\mathrm{B}})),\tag{6}$$

$$\underset{\boldsymbol{\Theta}^{\mathrm{A}},\boldsymbol{\Theta}^{\mathrm{B}}}{\operatorname{argmin}}L_2=\sum_i^{N_{\mathrm{AB}}}l_2(\boldsymbol{u}_i^{\mathrm{A}},\boldsymbol{u}_i^{\mathrm{B}}),\tag{7}$$

式中：$\boldsymbol{\Theta}^{\mathrm{A}}$、$\boldsymbol{\Theta}^{\mathrm{B}}$ 分别表示 A、B 域的模型参数，l_1、l_2 为损失函数。

整体目标函数 L 为：

$$\underset{\boldsymbol{\Theta}^{\mathrm{A}},\boldsymbol{\Theta}^{\mathrm{B}}}{\operatorname{argmin}} L = L_1 + \gamma L_2 + \frac{\lambda}{2}(\|\boldsymbol{\Theta}^{\mathrm{A}}\|^2 + \|\boldsymbol{\Theta}^{\mathrm{B}}\|^2) \quad (8)$$

式中：γ 和 λ 为权重参数。

使用 BP 算法，根据目标函数 L 分别对 $\boldsymbol{\Theta}^{\mathrm{A}}$、$\boldsymbol{\Theta}^{\mathrm{B}}$ 求梯度，双方交互计算梯度和损失需要用到的中间结果重复迭代，直至收敛。整个学习过程利用源域 A 和目标域 B 之间共同样本来学习二者各自的特征不变量 $\boldsymbol{u}_{\mathrm{A}}$、$\boldsymbol{u}_{\mathrm{B}}$，同时利用源域 A 的所有样本标签 $\boldsymbol{y}_{\mathrm{A}}$ 和 A 的不变量特征 $\boldsymbol{u}_{\mathrm{A}}$ 学习分类器。

在这个过程中，联邦体现在源域 A 和目标域 B 通过安全交互中间结果共同学习一个模型，迁移体现在目标域 B 迁移了源域 A 的分类能力。

预测时，$\boldsymbol{u}_{\mathrm{B}}$ 依赖于 $\boldsymbol{u}_{\mathrm{A}}$、$\boldsymbol{y}_{\mathrm{A}}$ 组成的分类器，和纵向联邦学习相同，需要源域 A 和目标域 B 所在的客户端共同完成预测过程。

1.4 联邦学习的问题

联邦学习涉及 2 个问题：通信开销和隐私保护问题。其中，联邦学习的通信开销问题主要是由客户端和中央服务器之间经过网络连接和传输模型、参数数据造成的。隐私保护问题主要是指经由网络传输时用户信息、模型信息的隐私和安全保护问题。

1.4.1 通信开销

对于通信开销问题，最直接的方案是降低模型的准确度，在联邦学习的框架中，仅训练占用信道容量较小的低开销模型。Google 的 Konečný 等[6]发表了通过降低上行通信成本来解决网络开销问题的方法：客户端只将本地计算得到的模型更新传递到中央服务器，而不是完整的本地模型。这种方法虽然能有效降低网络开销，但无法满足复杂的业务场景。在文献[6]的研究基础上，Caldas 等[7]提出了一种能够有效降低下行成本的方法，具体包括在服务器到客户端的全局模型上使用有损压缩，以及允许用户在全局模型的较小集上高效完成本地训练，减少客户端到服务器的通信成本。

与二者思路类似，Rothchild 等[8]提出了使用 Count Sketch 对客户端模型更新进行压缩处理。由于 Count Sketch 是线性的，可以通过 Sketch 计算动量和误差累积，从而将动量和误差累积的计算任务从客户端转移到中央服务器，保持了高压缩率和良好的收敛性。Reisizadeh 等[9]提出了一种周期平均和量化的处理方法 FedPAQ，量化处理本身也是压缩的一种方式。FedPAQ 允许网络中的客户端在与中央服务器同步之前执行本地训练，仅将活跃客户端的更新发送回中央服务器，且发回的仅为本地信息的量化版本。

与从压缩角度出发的方法不同的是，Hamer 等[10]提出了一种主要解决下行通信成本问题的集成方法——FedBoost。集成方法是机器学习中的一种通用技术，用于组合多个基本预测因子或专家来创建一个更精确的模型。FedBoost 主要通过学习一组预先训练好的基本预测因子实现联邦集成。

目前，大多数文献都从压缩传输数据的方式来解决通信开销问题，压缩后通信的上行和下行数据量都会减少。上述文献中提到的压缩方法包括有损压缩、提取、量化等，这些算法能够针对不同业务场景，达到更好减少网络开销的目的。

1.4.2 隐私安全性

联邦学习让原始数据在不离开本地的情况共同完成机器学习的模型训练，能够有效地在保护用户隐私的情况下打破数据孤岛，广泛应用于金融、通信等领域。但是模型逆向[11]、模型提取攻击[12]手段的出现，可以让攻击者通过梯度恢复部分原始数据，带来了安全隐患。现行的通用手段是通过差分隐私和安全多方计算来增强联邦学习的隐私安全性。

1.5 总结

联邦学习的 3 种分类：横向、纵向联邦学习和联邦迁移学习，能够提供不同业务场景、数据分布不同的解决方案。在解决了数据传输和隐私安全性问题的前提下，联邦学习可以充分发挥其数据保留在本地、降低泄露风险的优势。联邦学习的建模效果和将整个数据集放在一处建模的效果相同或相似。在保证数据安全的基础上，联邦学习可以达到或接近完整数据集的建模效果。

2 联邦学习框架

目前主流的联邦学习开源框架主要包括微众银行牵头提出的 FATE 框架、百度牵头提出的 Paddle-FL(Paddle Federated Learning)框架、谷歌牵头提出的 TFF(TensorFlow Federated)框架等。从产品能力和支持模型类型方面给出了以上 3 个联邦学习开源框架的对比，如表 1 所示。

表 1　开源框架对比

算法框架	产品能力	模型类型
FATE	联邦建模、部署、任务控制、可视化、联邦特征工程	横向联邦学习、纵向联邦学习、联邦迁移学习
PaddleFL	联邦建模、部署	横向联邦学习、纵向联邦学习
TFF	联邦建模	横向联邦学习

FATE 框架作为目前国内优秀的联邦学习开源框架，支持横向联邦学习、纵向联邦学习以及联邦迁移学习，同时提供了其他框架所没有的联邦特征工程算法、Kubernetes 容器化应用和联邦在线推理。同时，FATE 框架也是开源社区热度最高的联邦学习框架。

相对而言，PaddleFL 仅支持横向和纵向的联邦学习，TFF 框架仅实现了横向联邦学习的支持。目前二者均缺乏联邦树模型算法的实现，如梯度提升决策树（Gradient Boosting Decision Tree，GBDT）和 SecureBoost。

3　大数据平台联邦学习解决方案

通过第 1 节的联邦学习介绍可知，针对不同的业务场景以及数据分布状态，灵活选择合适的联邦学习手段可以达到更好的建模效果。工业级开源架构 FATE，同时支持联邦机器学习、联邦深度学习和联邦迁移学习，可以满足不同的业务需求。下面介绍基于 FATE 构建的联邦学习平台。

3.1　平台功能架构

基于 FATE 架构的联邦学习平台，支持横向、纵向联邦学习和联邦迁移学习，拥有联邦模型在线、离线推理能力，能够灵活满足绝大多数的联邦学习应用场景。该平台支持容器化部署，可在线扩展联邦参与方组织，使单个数据中心的算法、算力以及数据存储能力都可以方便地横向扩展。提供从数据接入到部署模型一站式的开发能力，简化联邦学习模型的开发、上线工作，降低使用的门槛。

平台功能架构如图 6 所示。

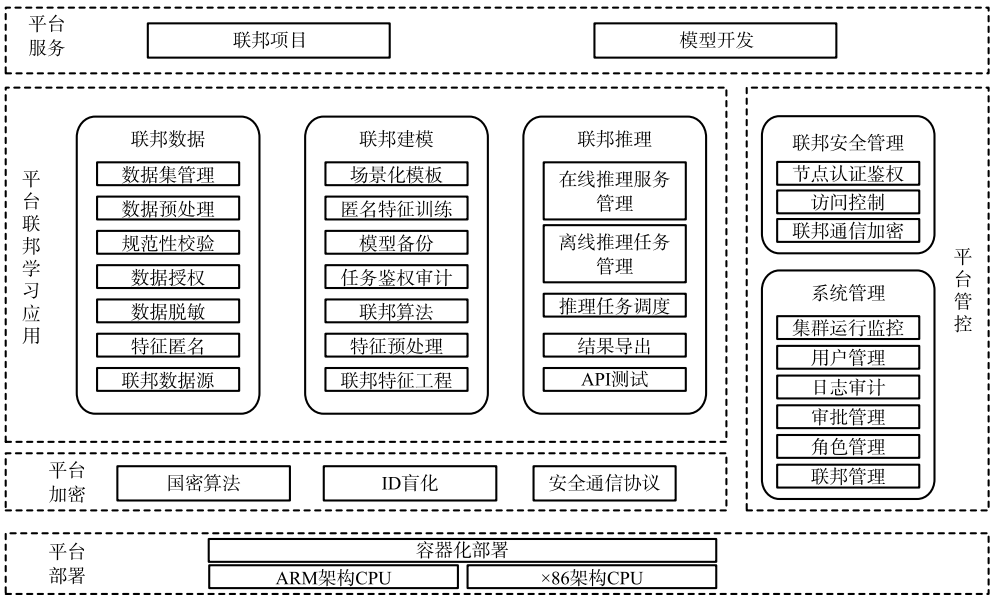

图 6　平台功能架构

平台架构包括平台部署、平台加密、平台管控、平台联邦学习应用和平台服务功能。目前平台已经适配了基于×86 架构和 ARM 架构的部分厂家的 CPU 型号，并支持容器化部署。针对联邦学习对安全保密的特殊要求，提供了国密算法、安全通信协议和 ID 盲化等加密功能。针对日常的系统和安全管理，提供了用户管理、日志审计等系统管理功能和访问控制、节点认证鉴权等联邦安全管理功能。针对联邦学习的建模和推理过程，提供了从联邦数据接入、预处理等联邦数据功能到方便建模的场景化模板、联邦算法等联邦建模功能，以及在线推理服务和离线推理任务管理等联邦推理功能。针对任务发起和多样化展示等日常应用需求，提供了联邦项目和模型开发等对外服务功能。

3.2　基于联邦学习平台的大数据中心

3.2.1　建设方案

基于现有的大数据中心部署联邦学习平台，建立新的数据整合方式，解决多源数据的整合问题。

首先,明确大数据中心的建设需求,明确专业领域,根据目标在联邦学习平台构建模型库和算法。然后,数据所在区域大数据中心根据分析需要,对该区域掌握的数据进行汇总,通过联邦平台,配合中央服务器进行挖掘分析。最后,各区域大数据中心根据需求,对挖掘分析后的结果进行多样化的展示。

3.2.2 部署方案

在各大数据中心同时部署联邦学习平台,各方可通过平台的组织类型功能设置控制。联邦学习平台根据所处中心拥有的数据,在不同任务中扮演不同角色。例如,某个大数据中心可在T任务中作为任务发起方,在P任务中作为数据提供方。中央服务器和客户端之间的关系如图7所示。

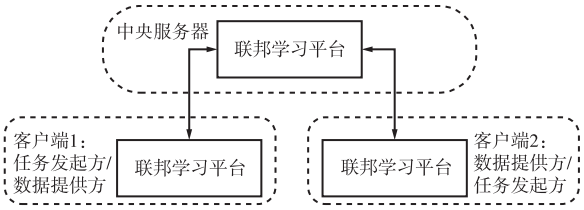

图7 联邦学习部署关系

3.2.3 应用场景分析

通信卫星在轨长期运行,会出现老化问题。太空环境较为复杂,如果环境恶化,其老化过程将会加速。随着设备寿命的减少,通信风险会增大。可以通过直接分析卫星锂电池来预测卫星的寿命,为卫星升级更新、规避可能的风险提供依据。

大数据中心引接外部数据流程较为复杂。假设部署了联邦学习平台的大数据中心S需要预测现有通信卫星的使用寿命,与该数据中心部署在同一网络上有区域大数据中心R和T。其中,部署了联邦学习平台的区域大数据中心R刚建成不久,缺少相应引接手段,只有少量通信卫星的锂电池使用数据。为了提高预测的准确性,需要更多卫星锂电池的使用数据作为预测模型的样本输入。而部署了联邦学习平台的区域大数据中心T拥有大量卫星(非通信卫星)电池使用情况的测量数据集。

针对这种样本维度和特征维度重叠部分很少的情况,结合平台的联邦迁移学习方法设计算法能够更好地得到理想的模型。

数据处理流程关系如图8所示,数据中心R和T参与目标模型的训练,其中数据中心S作为中央服务器,负责聚合全局模型。由于R通信卫星充放电数据和T的卫星充放电数据不存在样本交集,故没有样本对齐步骤。整个模型的构建过程步骤如下:

① 数据中心R和T分别计算和自己相关的特征中间结果并加密交互,用来求得各自梯度和损失;

② 数据中心R和T分别计算梯度和损失,发送给中央服务器;

③ 中央服务器把梯度和损失回传给数据中心R和T,然后分别更新各自的模型;

④ 重复①~③直至收敛;

⑤ 数据中心T在接收到中央服务器传回的模型后,使用通信卫星充放电数据进行重新训练,得到适合通信卫星数据的预测模型M_S。

中央服务器运用适用于通信卫星的模型M_S,预测通信卫星的寿命。

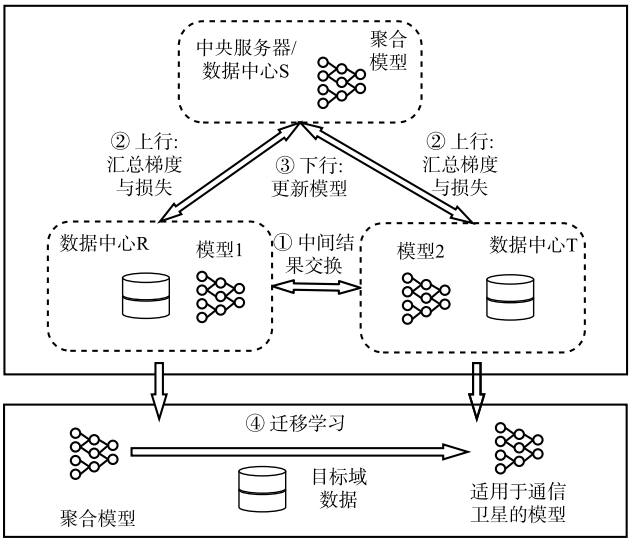

图8 数据处理流程关系

4 结束语

通过对比,FATE框架相比于PaddleFL框架和TFF框架支持更多的联邦学习算法,在开源社区也拥有最高的热度。以FATE框架为基础,联邦学习平台支持容器化部署,能够满足大部分的联邦学习应用场景,提供从数据接入到模型部署一站式的开发能力,简化联邦学习模型的开发、上线工作,降低使用的门槛。

该平台具备的横向、纵向联邦学习和联邦迁移学习能力,能够针对特征重叠多用户重叠少、用户重叠多特征重叠少、用户重叠少特征重叠少等场景,依

托现有的大数据中心,在保证数据传输和隐私安全的前提下,解决现有数据中心之间数据共享不畅导致的模型训练困难的问题,有效提升数据的使用效率。

此外,联邦学习平台提供了平台部署、加密、管控、联邦学习应用、服务等功能,能够有效屏蔽各种类型的大数据中心硬件差异,提供一站式的开发能力,简化联邦学习模型的开发和上线工作,降低使用的门槛。

参考文献

[1] KONEČNÝ J,MCMAHAN H B,RAMAGE D,et al. Federated Optimization:Distributed Machine Learning for On-device Intelligence[EB/OL]. (2016-10-08)[2024-06-10]. https://arxiv.org/abs/1610.02527.

[2] LI Q,WEN Z,WU Z,et al. A Survey on Federated Learning Systems:Vision,Hype and Reality for Data Privacy and Protection[J]. IEEE Transactions on Knowledge and Data Engineering,2019,35(4):3347-3366.

[3] LIU Y,YUAN X L,XIONG Z H,et al. Federated Learning for 6G Communications:Challenges,Methods,and Future Directions[J]. China Communications,2020,17(9):105-118.

[4] FENG S W,YU H. Multi-participant Multi-class Vertical Federated Learning[EB/OL]. (2020-01-30)[2024-06-10]. https://arxiv.org/abs/2001.11154.

[5] YANG H W,HE H,ZHANG W Z,et al. FedSteg:A Federated Transfer Learning Framework for Secure Image Steganalysis[J]. IEEE Transactions on Network Science and Engineering,2021,8(2):1084-1094.

[6] KONEČNÝ J,MCMAHAN H B,YU F X,et al. Federated Learning:Strategies for Improving Communication Efficiency[EB/OL]. (2016-10-18)[2024-06-10]. https://arxiv.org/abs/1610.05492.

[7] CALDAS S,KONEČNY J,MCMAHAN H B,et al. Expanding the Reach of Federated Learning by Reducing Client Resource Requirements[EB/OL]. (2018-12-18)[2024-06-10]. https://arxiv.org/abs/1812.07210.

[8] ROTHCHILD D,PANDA A,ULLAH E,et al. FetchSGD:Communication-efficient Federated Learning with Sketching[EB/OL]. (2020-07-15)[2024-06-10]. https://arxiv.org/abs/2007.07682.

[9] REISIZADEH A,MOKHTARI A,HASSANI H,et al. FedPAQ:A Communication-efficient Federated Learning Method with Periodic Averaging and Quantization[C]//International Conference on Artificial Intelligence and Statistics. Palermo:PMLR,2020:2021-2031.

[10] HAMER J,MOHRI M,SURESH A T. FedBoost:A Communication-efficient Algorithm for Federated Learning[C]//International Conference on Machine Learning.[S.l.]:PMLR,2020:3973-3983.

[11] FREDRIKSON M,JHA S,RISTENPART T. Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver:ACM,2015:1322-1333.

[12] TRAMÈR F,ZHANG F,JUELS A,et al. Stealing Machine Learning Models via Prediction APIs[C]//25th USENIX Security Symposium (USENIX Security 16). Austin:USENIX,2016:601-618.

作者简介

陈家良　男,(1991—),硕士,工程师。
冯金顺　男,(1981—),硕士,高级工程师。
董少然　男,(1985—),硕士,高级工程师。
郭新苍　男,(1975—),硕士,高级工程师。
范烁晨　男,(1997—),硕士,助理工程师。
朱光耀　男,(1997—),硕士,助理工程师。
高　静　女,(1997—),硕士,助理工程师。
马胤垚　男,(1996—),硕士,助理工程师。