联邦学习算法综述

王健宗[1]，孔令炜[1]，黄章成[1]，陈霖捷[1]，刘懿[1]，何安珣[1]，肖京[2]
1. 平安科技（深圳）有限公司，广东 深圳 518063
2. 中国平安保险（集团）股份有限公司，广东 深圳 518031

摘要
近年来，联邦学习作为解决数据孤岛问题的技术被广泛关注，已经开始被应用于金融、医疗健康以及智慧城市等领域。从3个层面系统阐述联邦学习算法。首先通过联邦学习的定义、架构、分类以及与传统分布式学习的对比来阐述联邦学习的概念；然后基于机器学习和深度学习对目前各类联邦学习算法进行分类比较和深入分析；最后分别从通信成本、客户端选择、聚合方式优化的角度对联邦学习优化算法进行分类，总结了联邦学习的研究现状，并提出了联邦学习面临的通信、系统异构、数据异构三大难题和解决方案，以及对未来的期望。

关键词
联邦学习；算法优化；大数据；数据隐私

中图分类号：TP311 文献标识码：A doi: 10.11959/j.issn.2096-0271.2020055

Research review of federated learning algorithms

WANG Jianzong[1], KONG Lingwei[1], HUANG Zhangcheng[1], CHEN Linjie[1], LIU Yi[1], HE Anxun[1], XIAO Jing[2]
1. Ping An Technology (Shenzhen) Co., Ltd., Shenzhen 518063, China
2. Ping An Insurance (Group) Company of China, Ltd., Shenzhen 518031, China

Abstract
In recent years, federated learning has been proposed and received widespread attention to overcome data isolated island challenge. Federated learning related researches were adopted in areas such as financial field, healthcare domain and smart city related application. Federated learning concept was introduced into three different layers. The first layer introduced the definition, architecture, classification of federated learning and compared the federated learning with traditional distributed learning. The second layer presented comparison and analysis of federated learning algorithms from machine learning and deep learning aspects. The third layer separated federated learning optimization algorithms into three aspects to optimize federated learning algorithm through reducing communication cost, selecting proper clients and different aggregation method. Finally, the current research status and three main challenges on communication, heterogeneity of system and data to be solved were concluded, and the future prospects in federated learning domain were proposed.

Key words
federated learning, algorithm optimization, big data, data privacy

1 引言

随着数字化技术进入高速发展期，大数据和人工智能等技术迎来爆发式发展[1-2]，这一方面为传统业态带来了升级变革的新机遇[3-5]，另一方面不可避免地给数据和网络安全带来了全新的挑战，而数据孤岛问题[6-7]是关键挑战之一。纵向来看，行业顶尖的巨头公司垄断了大量的数据信息，小公司往往很难得到这些数据，导致企业间的层级和差距不断拉大；横向来看，同一层级不同行业的公司，由于系统和业务的闭塞性与阻隔性，很难实现数据信息的交流与整合，联合建模需要跨越重重壁垒。

针对上述人工智能行业目前面临的痛点，联邦学习给出了答案。联邦学习是由谷歌研究院在2016年率先提出的概念[8-10]。该技术可在数据不共享的情况下完成联合建模。具体来讲，各个数据拥有者（个人/企业/机构）的自有数据不会离开本地，通过联邦系统中加密机制下的参数交换方式（即在不违反数据隐私法规的情况下）联合建立一个全局的共享模型，建好的模型在各自的区域只为本地的目标服务[11]。尽管联邦学习[12-14]和分布式机器学习[15-19]有部分相似的地方，但是在应用领域、系统设计、优化算法方面，联邦学习有自己的特征。在数据量庞大、所需计算资源较高时，分布式机器学习（如参数服务器）有明显的优势，它将独立同分布（independently identically distribution，IID）的数据或模型参数存储在各个分布式节点上，中心服务器调动数据和计算资源，联合训练模型。因客户端的地理、时间等分布差异，联邦学习经常要处理非独立同分布（non-IID）的数据。本文结合联邦学习的现状，对联邦学习系统进行分层，按模块整理联邦学习目前取得的相关成果。

联邦学习算法结构如**图1**所示。

为了整合多个来源的数据，当前比较普遍的做法是通过数据预处理ETL（extract-transform-load）工具将不同源的数据移动到关系数据库中，将具有庞大计算量的任务部署到多台机器上，以提升计算效率，减少任务耗能。

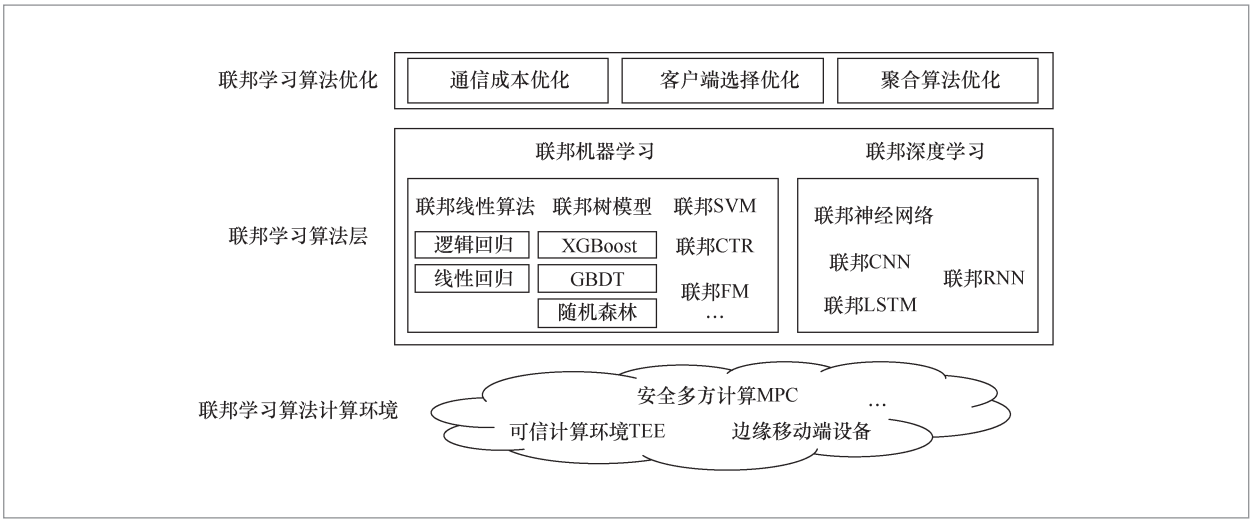

图1 联邦学习算法结构

2 联邦学习概述

2.1 联邦学习的定义

2016年，谷歌研究院在解决面向用户个体的键盘输入法优化问题时，提出了联邦学习这一全新的人工智能解决方案。联邦学习面向的场景是分散式多用户 $\{F_1,\cdots,F_N\}$，每个用户客户端拥有当前用户的数据集 $\{D_1,\cdots,D_N\}$。传统的深度学习将这些数据收集在一起，得到汇总数据集 $D=U_1\cup\cdots\cup U_N$，训练得到模型 M_{SUM}。联邦学习方法则是由参与的用户共同训练一个模型 M_{FED}，同时用户数据 D_i 保留在本地，不对外传输。如果存在一个非负实数 δ，使得 M_{FED} 的模型精度 V_{FED} 与 M_{SUM} 的模型精度 V_{SUM} 满足如下不等式：

$$\left|V_{\mathrm{FED}}-V_{\mathrm{SUM}}\right|<\delta \qquad (1)$$

则称该联邦学习算法达到 δ-精度损失[4]。联邦学习允许训练模型存在一定程度的性能偏差，但是为所有的参与方提供了数据的安全性和隐私保护。联邦学习常用的框架有两种，一种是客户端-服务器架构[8]，另一种是对等网络架构[20]。在客户端-服务器架构中，联邦学习的训练方式是让各个数据持有方根据自己的条件和规则在本地训练模型，然后将脱敏参数汇总到中央服务器进行计算，之后再下发回各个数据持有方更新自己本地的模型，直至全局模型稳健为止。在对等网络架构中进行联邦学习训练时，参与方之间可以直接通信，不需要借助第三方，安全性得到了进一步提高，但是需要更多的计算操作进行加密和解密[21-24]。目前的研究更多的是基于第三方服务器的框架。因此本文着重介绍客户端-服务器架构的联邦学习流程。

2.2 客户端-服务器架构的联邦学习流程

在物理层面上，联邦学习系统一般由数据持有方和中心服务器组成。各数据持有方的本地数据的数量或特征数可能并不足以支持一次成功的模型训练，因此需要其他数据持有方的支持。而联邦学习中心服务器的工作类似于分布式机器学习的服务器，其收集各数据持有方的梯度，并在服务器内进行聚合操作后返回新的梯度[25]。在一次联邦学习的合作建模过程中，数据持有方对本地数据的训练仅发生在本地，以保护数据隐私，迭代产生的梯度在脱敏后被作为交互信息，代替本地数据上传给第三方受信任的服务器，等待服务器返回聚合后的参数，对模型进行更新[8]。**图2**展示了客户端-服务器架构的联邦学习流程。

步骤1：系统初始化。首先由中心服务器发送建模任务，寻求参与客户端。客户端数据持有方根据自身需求，提出联合建模设想。在与其他合作数据持有方达成协议后，联合建模设想被确立，各数据持有方进入联合建模过程。由中心服务器向各数据持有方发布初始参数。

步骤2：局部计算。联合建模任务开启并初始化系统参数后，各数据持有方将被要求首先在本地根据己方数据进行局部计算，计算完成后，将本地局部计算所得梯度脱敏后进行上传，以用于全局模型的一次更新。

步骤3：中心聚合。在收到来自多个数据持有方的计算结果后，中心服务器对这些计算值进行聚合操作，在聚合的过程中需要同时考虑效率、安全、隐私等多方面的问题。比如，有时因为系统的异构性，中心服务器可能不会等待所有数据持有方的上传，而是选择一个合适的数据持有方子集作为收集目标，或者为了安全地对参数进

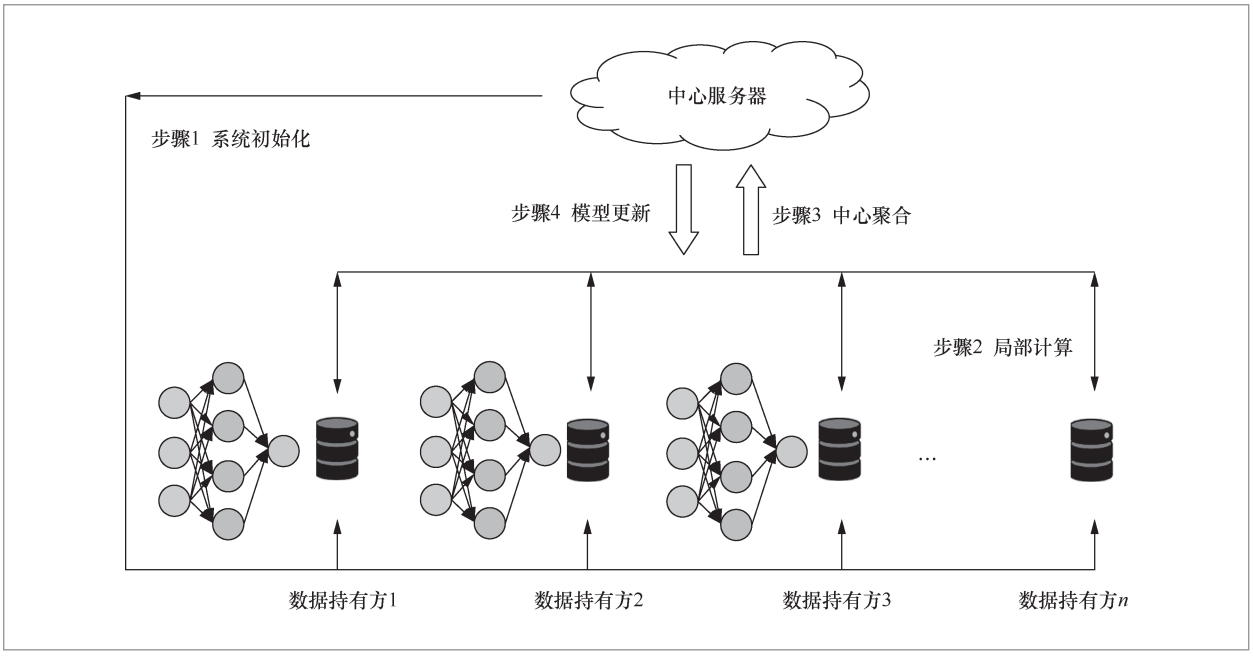

图 2 客户端 – 服务器架构的联邦学习流程

行聚合，使用一定的加密技术对参数进行加密，这些方法将会在后面的章节中详细讨论。

步骤4：模型更新。中心服务器根据聚合后的结果对全局模型进行一次更新，并将更新后的模型返回给参与建模的数据持有方。数据持有方更新本地模型，并开启下一步局部计算，同时评估更新后的模型性能，当性能足够好时，训练终止，联合建模结束。建立好的全局模型将会被保留在中心服务器端，以进行后续的预测或分类工作。

上述过程是一个典型的基于客户端–服务器架构的联邦学习过程。但并不是每个联邦学习任务都一定要严格按照这样的流程进行操作，有时可能会针对不同场景对流程做出改动，例如，适当地减少通信频率来保证学习效率[26-31]，或者在聚合后增加一个逻辑判断，判断接收到的本地计算结果的质量[32-35]，以提升联邦学习系统的鲁棒性[36]。

2.3 联邦学习与传统分布式学习的区别

基于客户端–服务器架构的联邦学习和分布式机器学习[37]都是用来处理分布式数据的，但在应用领域、数据属性和系统构成等方面，其与分布式机器学习存在差异，主要如下。

（1）应用领域

大量的数据或者较大的模型往往对计算资源有较高的要求。单一的计算节点已经不能满足需求。分布式机器学习将训练数据或模型参数分布在各个计算或存储节点，利用中心服务器对节点进行调度，加速模型的训练。而当数据具有隐私敏感属性时，分布式机器学习的中心调度将会给用户数据带来极大的隐私泄露风险[38]。联邦学习始终将数据存储在本地，相比需要将数据上传到服务器的方式，可以最大限度地保障数据隐私。

（2）数据属性

机器学习的主要目的是寻找数据的概

率分布，这在数据集满足独立同分布的情况下相对比较容易。分布式机器学习与经典机器学习处理的数据往往是独立同分布的，联邦学习则有所不同。由于客户端的地理位置、时间等分布的差异性，联邦学习系统的原始数据往往是非独立同分布的。同时，横向联邦学习和纵向联邦学习也是根据客户端数据的不同属性来进行分类的。客户端之间的数据特征和分类标签差异较大，在进行训练时需要进行对齐工作。

（3）系统构成

在物理组成上，联邦学习系统和分布式系统较为相似，都由中心服务器和多个分布式节点构成。在分布式系统中，数据计算和模型更新统一由中心服务器进行调度，节点和中心服务器之间的数据时延较小，模型训练时间主要由计算时间决定。而在联邦系统中，各个参与方地位平等，可以自主决定是否参与模型训练。且由于分布式节点多为计算能力差异较大、网络环境不同以及所处状态不可控的客户端，在系统设计上，需要考虑数据传递时延、数据非独立同分布以及隐私安全等众多因素，这就要求系统对联邦学习算法做出适应性的改变[39-41]。联邦聚合是联邦学习系统中不同于分布式机器学习的优化算法，为解决数据非独立同分布和减轻数据异构提供了新的思路。同时，由于联邦学习具有极好的隐私保护能力，在系统的各个环节都要注意加密算法的应用。加密数据的传递、目标函数损失的计算、梯度的计算与传递模型参数的传递等都对传统的算法提出了新的要求。

2.4 联邦学习分类

联邦学习的孤岛数据有不同的分布特征。对于每一个参与方来说，自己所拥有的数据可以用一个矩阵来表示。矩阵的每一行表示每一个用户或者一个独立的研究对象，每一列表示用户或者研究对象的一种特征。同时，每一行数据都会有一个标签。对于每一个用户来说，人们希望通过他的特征X，学习一个模型来预测他的标签Y。在现实中，不同的参与方可能是不同的公司或者机构，人们不希望自己的数据被别人知道，但是人们希望可以联合训练一个更强大的模型来预测标签Y。

根据联邦学习的数据特点（即不同参与方之间的数据重叠程度），联邦学习可被分为横向联邦学习[42]、纵向联邦学习[43]、迁移联邦学习[44]。

当两个参与方的用户重叠部分很少，但是两个数据集的用户特征重叠部分比较多时，这种场景下的联邦学习叫作横向联邦学习。比如一个银行系统在深圳和上海的分部为参与方，两边业务类似，收集的用户数据特征比较类似，但是两个分部的用户大部分是本地居民，用户重叠比较少，当两个分部需要做联邦模型对用户进行分类的时候，就属于横向联邦学习。

当两个参与方的用户重叠部分很多，但是两个数据集的用户特征重叠部分比较少时，这种场景下的联邦学习叫作纵向联邦学习。比如同一个地区的两个机构，一个机构有用户的消费记录，另一个机构有用户的银行记录，两个机构有很多重叠用户，但是记录的数据特征是不同的，两个机构想通过加密聚合用户的不同特征来联合训练一个更强大的联邦学习模型，这种类型的机器学习模型就属于纵向联邦学习。

当两个参与方的用户重叠部分很少，两个数据集的用户特征重叠部分也比较少，且有的数据还存在标签缺失时，这种场景下的联邦学习叫作迁移联邦学习。比如两个不同地区的机构，一个机构拥有所在地区的用户消费记录，另一个机构拥有

所在地区的银行记录，两个机构具有不同的用户，同时数据特征也各不相同，在这种情况下联合训练的机器学习模型就是迁移联邦学习。

目前大部分的研究是基于横向联邦学习和纵向联邦学习的，迁移联邦学习领域的研究暂时还不多。因此，本文将重点讨论横向联邦学习和纵向联邦学习的算法类型。横向联邦学习中数据特征重叠维度较多，根据重合维度进行对齐，取出参与方数据中特征相同而用户不完全相同的部分进行联合训练；纵向联邦学习用户重合较多，根据用户ID进行匹配，取出参与方数据中用户相同而特征不完全相同的部分进行联合训练。

2.5 联邦学习算法的特点

基于上述对联邦学习的介绍，总结出以下几点联邦学习算法的特点。

- 支持非独立同分布数据：这是联邦学习算法的一个很重要的特性。联邦学习算法必须在非独立同分布数据中有良好的表现。在联邦学习的实际使用中，数据持有方的数据质量和分布是不可控的，无法要求数据持有方的数据满足独立同分布[45]，因此联邦学习算法需要支持非独立同分布数据。
- 通信高效：联邦学习算法需要考虑数据持有方的系统异构性，并在不损失准确率或损失很小的情况下提高通信效率，降低通信损耗。
- 快速收敛：在联合建模过程中，首先需要保证模型收敛，同时需要提高收敛速度。
- 安全性和隐私性：数据隐私安全是联邦学习的重要特点，因此安全性和隐私性是对联邦梯度更新的必要要求。安全性和隐私性可以通过加密等方式在聚合过程中进行，也可以反映在单机优化的过程中。
- 支持复杂用户：复杂用户指用户本身数量大，且用户数据存在不均衡性或偏移。这在联邦学习的实际应用中是非常可能的，联邦优化算法需要对这种情况具有很好的兼容效果。

3 联邦学习算法分类

联邦学习系统是面向多客户端的模型训练系统，各个客户端在参与训练时，数据保留在本地，不会被发送给其他客户端或中心服务器。中心服务器通过对客户端发送的本地模型更新进行整合，最终完成共享模型的训练。客户端在进行本地模型训练时，由于设备间计算能力的差异，各个客户端完成计算的时间不同。同时由于本地数据和全局数据之间的分布差异，部分异常数据会对共享模型造成破坏，导致模型精度降低。联邦学习算法针对以上问题，在机器学习算法和深度学习的基础上做出修改，满足非独立同分布数据、网络时延以及隐私保护的需求。

3.1 基于机器学习的联邦学习算法

联邦机器学习算法指在联邦学习框架下的经典机器学习算法实现。联邦机器学习，尤其是横向联邦学习，在整体模式上与分布式机器学习类似[16]。但是，相较于传统的机器学习算法，由于联邦学习特有的迭代模式和特点，即需要在数据不出本地的基础上双方交换训练参数以完成联合建模，联邦学习框架下的机器学习算法实现更加复杂。联邦机器学习算法的实现往往基于上述联邦优化算法的框架，但因为机器学习算法之间的差异性，有时又需要做一些针对性的修改，同时也需要考虑实际过程中的安全性等因素。下面介绍几种目

前常见的联邦机器学习算法。

3.1.1 联邦线性算法

Yang K等人[30]提出了一种中心联邦学习框架下的纵向联邦逻辑回归实现方法，这种方法实现了纵向联邦学习中的逻辑回归，其目标函数是：

$$\min\left\{\frac{1}{N}\sum_{n}^{N}\mathcal{L}\left(\omega;x_n;y_n\right)\right\} \quad (2)$$

其中，ω为模型的参数，x_n为模型的特征，y_n为模型的标签，$n\in\{1,N\}$为数据的数量，$\mathcal{L}(\omega;x_n;y_n)$为模型损失函数。在纵向联邦学习中，通常假设数据持有方分为有标签数据持有方和无标签数据持有方。这种算法在联邦优化算法的框架下结合了同态加密的思想，训练过程通过同态加密的方法对双方的数据和梯度进行加密。假设无标签数据持有方α的数据为$d_\alpha=\omega^{\alpha^\tau}x$，其中$\omega^{\alpha^\tau}$表示第$\tau$轮状态下的无标签数据持有方的模型参数。用$[d_\alpha]$表示对$d_\alpha$的同态加密，整个训练过程可以描述如下。

无标签数据持有方α首先向有标签数据持有方β发送$[d_\alpha]$、$[d_a^2]$及$[\Delta\overline{d}_a]$，β计算梯度与损失，加密后回传。中心服务器收集来自α、β的加密梯度后，辅助α、β进行模型更新。为减少通信次数，降低通信损耗，这种方法引入了一个向量$\boldsymbol{s}$来体现模型的变化，辅助更新，并且使用了周期性梯度更新。

Yang S W等人[41]提出了一种去中心联邦学习框架下的纵向联邦逻辑回归实现方法。他们认为在现实生活中，找到合作双方共同信任的第三方辅助方是很难的，并且这也在无形中提高了数据泄露的风险和系统的整体复杂性，因此他们认为取消第三方的参与对整个过程有很大的积极意义。

在这种方法中，有标签数据持有方在训练过程中起主导作用。从某种意义上讲，有标签数据持有方承担了被取消的中心服务器的责任。假设有标签数据持有方α和无标签数据持有方β协定合作建模，α首先向β发送建模密钥，α、β分别初始化参数ω^1、ω^2，并计算ω^ix^i，其中$i\in\{1,2\}$。计算完毕后β将计算结果发送给α，α对双方计算结果取和，并利用逻辑回归方程求取最终输出，在对相同标签值计算损失结果后，加密损失并返回。之后双方分别计算梯度（对于β来说是加密后的梯度）。β将加密后的梯度添加噪声后交由α解密返回，双方分别进行梯度更新。在整个过程中，双方彼此之间始终对数据进行保密，传输通道中也均为保密信息，这就使得数据的隐私性不止针对合作方，也拥有了一定的对抗外部异常攻击的能力。

3.1.2 联邦树模型

Liu Y等人[46]提出了一种基于中心纵向联邦学习框架的随机森林实现方法——联邦森林。在建模过程中，每棵树都实行联合建模，其结构被存储在中心服务器及各个数据持有方，但是每个数据持有方仅持有与己方特征匹配的分散节点信息，无法获得来自其他数据持有方的有效信息，以保障数据的隐私性。最终整个随机森林模型的结构被打散存储，中心服务器中保留完整的结构信息，节点信息被分散在各数据持有方。在使用模型进行预测时，首先获取本地存储的节点信息，然后通过中心节点联合调用树结构中其他客户端的节点信息。这种方法减少了预测时每棵树的通信频率，对提高通信效率有一定的帮助。

SecureBoost[47]是一种基于梯度提升决策树（gradient boosting decision tree，GBDT）的去中心纵向联邦学习框架，同样包含有标签数据持有方和无标签数据持有方。梯度提升决策树算法中联邦

学习需要交换的参数与联邦线性算法有很大区别，涉及二阶导数项。根据一般的梯度提升决策树算法，目标函数为：

$$\min\left\{\mathcal{L}^{\tau} \simeq \sum_{n}^{N}\left[j\left(y_n, \hat{y}_i^{(\tau-1)}\right)+F\left(x_i\right)\right]\right\} \quad (3)$$

其中，τ为回归树的第τ次迭代，$\mathcal{L}^{\tau}$为目标函数的最小化损失值，$j(\cdot,\cdot)$为每个叶子节点上损失的计算函数，$F(x)$为预测残差的一阶、二阶导数之和，即泰勒二次展开式。为防止过拟合，在损失函数中添加正则项：

$$\varphi\left(f_{\tau}\right)=\gamma T+\frac{1}{2}\lambda\|\omega\|^2 \quad (4)$$

其中，γ和λ为超参数，分别控制树和特征的数量，ω为权重值，T为原始损失函数。

在一般分布式机器学习中，可以通过向参与方发送$F(x)$实现联合建模。但是由于使用$F(x)$可以反推出数据标签，这样的方法显然不适用于联邦学习框架，因此，SecureBoost采用一种在保护数据隐私的同时，保证训练性能的联合建模方法。有标签数据持有方α首先计算$F(x)$，并将结果加密后发送给无标签数据持有方β。β根据同态加密求和方法进行局部求和，并将结果回传。收到计算结果后，α将数据按照特征分桶，并进行聚合操作，将加密结果发送给β。最终由α将从β中收集的局部最优解进行聚合，产生最优解，并下发回β，完成联合建模的过程。需要说明的是，SecureBoost支持多方合作，即无标签数据持有方β表示所有无标签数据持有方的集合，但是有标签数据持有方仅为一方。与分布式XGBoost相比，SecureBoost在保障模型准确率的情况下，保护了数据的隐私，成功地将纵向GBDT应用到联邦学习框架中。

Li Q B等人[48]提出了一种实现多方GBDT建模的去中心横向联邦学习框架——基于相似度的联邦学习（similarity-based federated learning，SimFL）。这种方法总体分为两个步骤。首先，在预训练时，各个数据持有方在本地对数据进行哈希分类，分类依据为局部敏感哈希（locality sensitive hashing，LSH）；之后对各个本地哈希表进行聚合，生成全局哈希表，并向所有数据持有方发布。因此各个数据持有方在训练阶段可以基于全局哈希表进行建模，而不会直接接触到其他数据持有方的数据。LSH还可以用于获得不同数据持有方之间数据的相似性，数据的相似度越高，在哈希表中表现相同值的可能性就越大。

当某个数据持有方表现出与多个数据持有方有高度的数据相似性时，可以认为这个数据持有方的数据是很重要的，因此SimFL使用一种加权梯度上升（weighted gradient boosting）的方法进行单棵树建模，具体思想表现为将相似程度与梯度权值关联，相似程度越大，梯度权值越高，聚合时产生的表现力就越强。

这种通过哈希表加密的方法单从隐私保护性能上来讲，无法超越差分隐私等方法，但是在牺牲小部分隐私保护强度的情况下，该方法在通信效率方面得到了补偿，是一种联邦学习框架下树类算法实现的新方向。

3.1.3 联邦支持向量机

Hartmann V等人[49]提出了一种将支持向量机（support vector machine，SVM）安全部署在联邦学习中的方法，主要通过特征哈希、更新分块等方式对数据隐私性进行保障。其目标函数如下：

$$\min\left\{\mathcal{L}(\omega)=\frac{1}{N}\sum_{i}^{N}L\left(\omega, x_i, y_i\right)+\lambda R(\omega)\right\} \quad (5)$$

其中，N为训练数据，ω为模型参数，$L(\omega, x_i, y_i)$为在点(x_i, y_i)的损失，$\lambda R(\omega)$为

损失函数的正则项，超参数λ控制惩罚力度。在支持向量机中，其损失函数为：$L(\omega, x_i, y_i)=\max\{0, 1-\omega^T x_i y_i\}$。类似于SimFL，这里也对特征值进行降维哈希处理，以隐藏实际的特征值。除此之外，由于在线性支持向量机中，中心服务器有可能根据更新梯度反推出数据标签，为了保护数据的隐私性，这里采用次梯度更新的更新方式。在实际表现中，这种支持向量机在联邦框架下的应用具有不亚于单机支持向量机的性能。

3.2 基于深度学习的联邦学习算法

为了保障数据隐私安全，联邦学习客户端在进行数据通信时，往往会对传输的信息进行编码和加密，同时因为原始用户数据对中心服务器不可见，所以在模型搭建时训练样本对中心服务器以及模型设计人员不可观测。之前用于经典深度学习的相关模型在联邦学习系统中不一定是最优设计。为了避免网络模型的冗余，需要对经典深度学习模型进行相应的修改，如神经网络（neural network，NN）、卷积神经网络（convolutional neural networks，CNN）、长短期记忆网络（long short-term memory，LSTM）等。同时，为了适应联邦学习的流程，提高训练效果，学习训练的一些环节（如参数初始化、损失计算以及梯度更新等）也需要相应的调整。

3.2.1 联邦神经网络

McMahan H B等人[10]分别用NN和CNN在MNIST数据集上进行了测试。对于NN，模型的具体结构为含有两个隐藏层的神经网络，每个隐藏层包含200个神经元，且隐藏层用ReLU激活函数进行激活。然后将MNIST数据集分配到两个计算节点，每个计算节点含有样本量大小为600且无交集的子数据集。在进行联邦训练时，为了验证模型参数初始化和聚合比例带来的影响，实验分为具有不同初始化方式的两组：一组使用相同的随机种子初始化分配在两个计算节点的模型参数，另外一组使用不同的随机种子初始化模型参数。每组实验对来自不同节点的模型参数采用不同的权重比例进行加权整合，获取最终的联邦共享模型，即：

$$\omega_{FL}=\theta\omega+(1-\theta)\omega' \tag{6}$$

其中，ω_{FL}为联邦模型参数，ω和ω'为分布在不同节点的模型参数，θ用来调整两个模型参数之间的比例。实验发现，在达到相同的精度时，相比于单一数据本地训练，使用模型平均方法的联邦学习模型需要的训练回合更少，训练效率更高。在都使用联邦学习时，使用相同的随机初始化种子的联邦模型具有较好的效果，同时在模型参数比例为1:1时，达到最优损失。

3.2.2 联邦卷积神经网络

Zhu X H等人[50]使用简单的CNN训练隐私场景中的中文字体识别模型来测试现有的联邦学习框架（TensorFlow federated（TFF）和PySyft）以及数据集和客户端数量对联邦模型的影响。虽然对于文本识别问题常采用递归网络，但过于复杂的网络结构往往会影响联邦学习的收敛效率，于是采用含有4个卷积层和2个全连接层的简单CNN来训练模型。然后根据样本ID将数据集随机分配到不同的客户端，形成不同子集来模拟分布式数据。在进行训练时，客户端先在本地数据集上进行梯度计算和参数更新。在每个训练迭代结束后，汇总每个客户端累积的参数更新，用来更新最终的联邦模型。

作为对比，首先采用非联邦学习模式

进行训练，将所有数据放在TensorFlow上进行模型训练，获得的基础对比模型的准确率为42.65%。当客户端数量固定，改变每个客户端拥有的数据子集大小时，模型精度基本上随着数据集的增大而上升。不过在PySyft上，最佳精度始终无法达到基线（baseline）精度，且网络迭代次数多于基线模型的迭代次数。但TFF的模型收敛效果要优于PySyft，且在客户端样本数量达到一定程度时，联邦模型表现出优于基线模型的成绩，迭代次数也显著减少。两个框架下不同模型的效果差异可能是由于采用了不同的优化算法。针对联邦深度学习模型的框架还有很多限制，很多技术问题需要进一步解决，如TFF上对GPU卷积和池化计算的支持、PySyft上对更多优化器的支持。

影响卷积网络效果的因素还有很多，例如，客户端和服务器之间进行参数传递时，为了减轻对带宽的占用，往往对卷积网络模型的参数进行压缩。Sattler F等人[36]利用视觉几何组网络11（visual geometry group 11，VGG11）[51]发现，具有参数压缩的联邦聚合算法受non-IID数据的影响比较大，而在IID数据上则表现出几乎与非压缩聚合算法相同的收敛速度。用于联邦学习系统的稀疏三元压缩（sparse ternary compression，STC）[52]证明，在联邦学习环境中，该编码技术的通信协议优于联邦平均算法FedAvg（federated averaging）。

3.2.3 联邦LSTM

也有许多学者将LSTM运用到联邦语言模型中，用于预测字符[53-54]。他们将数据集人工分割为分配在多个客户端的联邦学习数据集，在合适的超参数设置下，这些模型在non-IID数据集上均达到了常规情况下的模型精度[8]。Sahu A K等人[39]在联邦数据集中训练LSTM分类器，提出了解决统计异质性的联邦学习框架FedProx，用于情感分析和字符预测。实验表明，相比于FedAvg，FedProx具有更快的收敛速度。Sattler F等人[36]也在卷积网络的基础上研究了优化模型参数压缩在non-IID数据集上的应用。在客户端与中心服务器通信时，相较于无压缩基线的2 422 MB网络参数量，使用基于STC编码通信协议的联邦学习系统可以在保证模型收敛效果的同时，将上行通信参数量压缩至10 MB左右，将下行参数量压缩到100 MB左右。

联邦学习算法目前的主要研究方向和瓶颈是如何提升联邦聚合的优化效率和性能，以达成模型的快速收敛和精准训练。因此，目前关于联邦深度学习模型的研究主要是如何优化联邦聚合，而针对联邦深度学习模型的研究还相对较少。

表1从算法、框架和特点等角度，对比了联邦机器学习和联邦深度学习的算法。

4 联邦学习算法的优化分类方法

相对于分布式学习，联邦学习有一些独特的属性，具体如下：

- 联邦学习的通信是比较慢速且不稳定的；
- 联邦学习的参与方设备异构，不同设备有不同的运算能力；
- 联邦学习更关注隐私和安全，目前大部分的研究假设参与方和服务器方是可信的，然而在现实生活中，其可能是不可信的。

在实现联邦学习的过程中，需要考虑如何优化联邦学习的算法，从而解决存在的现实问题。本文将从通信成本、客户端选择、异步聚合的角度介绍优化联邦学习的算法。在介绍优化算法之前，先介绍最

表1 联邦学习算法对比

类型	基础	算法	框架	特点
联邦机器学习	联邦线性算法	逻辑回归[40]	中心	同态加密，观察模型变化，周期性梯度更新
		逻辑回归[41]	去中心	取消第三方参与，有标签数据持有方主导，差分隐私
	联邦树模型	联邦森林[46]	中心	模型分散存储，中心服务器储存结构
		梯度上升树SecureBoost[47]	去中心	同态加密，特征分桶聚合，保障准确率
		梯度上升树SimFL[48]	去中心	哈希表加密，加权梯度上升，通信效率高
	联邦支持向量机	支持向量机Valentin[49]	中心	哈希表加密，次梯度更新，隐私性较好
联邦深度学习	联邦神经网络	NN[8]	中心	比传统神经网络收敛更快，参数联合初始化时具有更好的收敛效果
	联邦卷积神经网络	CNN[55]	中心	网络结构比RNN简单，收敛速度更快
		VGG11[51]	中心	non-IID数据上，参数压缩的优化算法收敛效果较差；不压缩的收敛效果较好，但参数量较大
	联邦LSTM	LSTM[8,39,51]	中心	受数据分布影响较大，不同的参数聚合方式效果不同

传统的联邦学习算法——FedAvg算法。

FedAvg算法[8]是目前最常用的联邦学习优化算法。与常规的优化算法[37-38]不同，其本质思想是对数据持有方采用局部随机梯度下降的方法进行本地模型优化，在中心服务器方[39]进行聚合操作。目标函数定义如下：

$$f(\omega^*)=\min\left\{f(\omega):=\frac{1}{M}\sum_{n=1}^{M}\mathbb{E}\left[f(\omega:x;x\in n)\right]\right\} \quad (7)$$

其中，M表示参与联合建模的数据持有方的数量，ω表示模型当前的参数，$\mathbb{E}$表示均方差函数。FedAvg算法是一种比较基础的联邦优化算法，其部署相对来说比较简单，应用领域很广泛[36]。

4.1 从通信成本角度优化的联邦学习算法

机器学习算法，特别是复杂的深度学习算法，在训练的过程中需要训练大量的参数，比如CNN可能需要训练上百万个参数，每一次更新过程需要更新上百万个参数；其次，网络通信的状态也可能导致很高的通信成本，比如不稳定的网络情况、参数上传和下载的过程中速度不一致都会导致整个算法的模型训练成本过大。因此需要根据这些特性来考虑如何从通信成本的角度优化联邦学习算法[8]。可以从以下角度考虑减少通信成本。

4.1.1 增加客户端计算成本

在联邦学习体系中，有时终端节点只会在有Wi-Fi时参与联邦学习训练，或者有时网络状况不佳，在这些情况下，更多的计算可以在本地进行，从而减少通信的次数。很多算法是从这个角度来优化通信成本的。比如Konečný J[21]考虑了优化FedAvg算法，增加每一轮迭代在每个客户端的本地更新参数的计算次数，并且与每一轮服务器参数更新只需要一次客户端本地更新的FedSGD算法进行了对比，实验通过MINSTCNN模型测试表明，当数据为IID时，算法可以明显减少通信成本，当数据为non-IID时，算法只能轻微地减少通信成本。Sahu A K等人[39]提出了一种更通用的FedProx算法，这种算法在数据为non-IID时优化效果更明显，因为联合训练的

终端参与方的数据、运算能力都是不均衡的，因此每一次参数更新时，不同的参与方要参与的运算次数都统一的话，会导致客户端的计算资源不能充分利用。为了避免这种情况，优化通信效率，FedProx算法可以动态地更新不同客户端每一轮需要本地计算的次数，使得算法更适合非独立同分布的联合建模场景。Liu Y等人[54]使用同样的优化思路优化联邦优化算法，并且在纵向联邦学习的框架下进行学习。LI X等人[56]则分析了FedAvg算法的收敛性，并证明了数据的异质性会导致联邦学习收敛速度降低。

4.1.2 模型压缩

有的优化算法目的是减少每一轮通信的参数量，例如通过模型压缩的技术（比如量化、二次抽样的方式）来减少每一次参数更新要传递的参数总量。Konečný J等人[9]提出了一种结构化的模型更新方式来更新服务器参数，在每一轮的参数通信过程中，减小参与方传递给服务器的模型更新参数的大小，从而减少通信。结构化更新是指通过提前定义上传模型参数的矩阵结构来上传模型，轮廓更新是指每次更新的参数需要在参与方进行压缩编码；模型最后通过CIFAR-10图像算法进行验证，实验表明，参与方越多，压缩效果越好；Caldas S等人[57]考虑的是从服务器到参与方的模型参数传递优化，通过有损压缩以及联邦参数筛选（federated dropout）的方式来减少从服务器到客户端需要传递的参数数量，降低通信成本的代价是在一定程度上降低模型的准确率。

在实现联邦学习时，通信是一个瓶颈。降低通信成本是非常重要的一个优化环节。有的优化以增加参与方的本地计算为代价，有的优化以降低整个模型的准确性为代价。在实际优化的过程中，可以根据实际情况和需求决定采用何种方式降低通信成本。

4.2 从客户端选择角度优化的联邦学习算法

联邦学习的客户端设备具有异构性的特征，并且不同的客户端的资源是有限的。通常，客户端随机选择参与联邦学习的模型训练过程。因此，在联邦学习训练的过程中，有的算法会考虑从客户端选择的角度进行优化。

不同的客户端的网络速度、运算能力等不同，每个客户端拥有的数据分布也是不平衡的，如果让所有的客户端都参与联邦学习的训练过程，将会有迭代落后的参与方出现，某些客户端长时间没有响应可能会导致整个系统无法完成联合训练。因此，需要考虑如何选择参与训练的客户端。FedAvg算法随机选择参与训练的客户端。但在网络结构复杂以及数据非独立同分布的情况下，FedAvg算法模型并不一定有好的表现。下面两篇参考文献介绍了一些优化方案。

Nishio T等人[58]提出了一种FedCS算法，设计了一种贪心算法的协议机制，以达到在联合训练的每一次更新中都选择模型迭代效率最高的客户端进行聚合更新的目的，从而优化整个联邦学习算法的收敛效率。实验表明，FedCS算法可以达到更高的准确性，但缺点是只有在模型比较基础的情况下，如基础的动态神经网络，才有好的表现，对于网络结构或参数数量较为复杂的情况来说，FedCS选择最优的聚合客户端的效率会降低，造成通信次数的增多和时间效率的降低。

Yoshida N等人[59]提出了一种Hybrid-FL的协议算法，该协议可以处理数据集为

non-IID的客户端数据，解决基于non-IID数据在FedAvg算法上性能不好的问题。Hybrid-FL协议使得服务器通过资源请求的步骤来选择部分客户端，从而在本地建立一种近似独立同分布的数据集用于联邦学习的训练和迭代。他们通过实验表明，对于non-IID数据类型的联邦学习分类算法来说，Hybrid-FL有较好的准确率表现。

4.3 从异步聚合角度优化的联邦学习算法

在FedAvg的算法中，聚合是与模型的更新保持同步的。每一次更新，服务器都同步聚合模型参数，然后将聚合参数发送给每一个客户端。在同步聚合中，服务器需要在接收到所有参与训练的客户端的参数之后才可以开始聚合，但是有的客户端运算传输快，有的客户端运算传输慢，为了避免出现通信迟滞现象，有研究者考虑用异步的方式进行聚合，从而优化联邦学习算法。

Sprague M R等人[60]提出了一种在联邦训练的过程中加入客户端的异步聚合方法，并且通过实例证明了这种方法的鲁棒性。当服务器接收到任何客户端的更新参数时，就进行一次聚合。但是这种算法的弊端是当模型数据为non-IID的时候，模型的收敛会出现很大的问题。

Xie C等人[61]为了解决异步同步的算法在non-IID数据上的适用性的问题，提出了另一种FedAsync算法，加入加权聚合的方法，使得服务器在接收到客户端的参数后，会通过当前训练的更新次数来设计加权聚合，从而解决non-IID数据的异步聚合的算法收敛问题。该参考文献理论上证明了在非凸性问题上FedAsync算法具有更好的收敛性。

联邦学习算法的优化分类方法见**表2**。

5 结束语

本文讨论了联邦学习目前的发展状况，从联邦学习算法的角度出发，将联邦学习相关算法分为联邦优化算法和联邦机器

表 2 联邦学习算法的优化分类方法

优化角度	文献方法	优化方法	优缺点
通信成本	FedAvg[8]	IID 数据；增加参与方本地计算	增加计算成本；non-IID数据优化效果差
	FedProx[39]	non-IID数据；增加本地计算	增加计算成本，可优化non-IID数据，代价是准确性降低
	VFL[54]	纵向联邦算法；增加本地计算	增加计算成本，代价是降低准确性
	结构和轮廓更新机制[9]	压缩传输模型，提升参与方到服务器的通信效率	参与方到服务器参数压缩，代价是复杂的模型结构可能出现收敛问题
	服务器-客户端更新[57]	压缩传输模型，提升服务器到参与方的通信效率	服务器到参与方参数压缩，代价是准确性降低，可能有收敛问题
客户端选择	FedCS[58]	选择迭代效率最优的模型训练参与方	比FedAvg更准确，但是只能被应用于简单的NN模型，不适合复杂模型
	Hybrid-FL[59]	服务器选择客户端数据组成近似IID的数据集	non-IID数据收敛有问题
异步聚合	AsyncFedAvg[60]	服务器接收到客户端参数更新就立刻聚合	存在non-IID数据收敛问题
	FedAsync[61]	服务端通过加权聚合的方式获取客户端的模型参数	难调参数，存在收敛问题

学习算法，对适合中心和去中心两种联邦学习结构的相关算法进行了论述，同时将联邦学习框架下的机器学习算法和联邦深度学习模型分别进行总结讨论。在联邦算法优化的过程中，从降低通信成本、最优客户端选择以及优化模型聚合方式的角度讨论了现有的联邦优化算法之间的差异和优缺点。

联邦学习目前依然处于快速发展的阶段，关于联邦学习在实际中的应用有大量的研究与讨论，但是在实现联邦学习的过程中，还有很多难题和挑战，本文给出了以下3类主要难题，即通信难题、系统异构难题以及数据异构难题。

● 通信难题。在联邦学习系统中，联邦网络可能由大量的设备组成。因此网络中的通信效率会对整体速度产生较大的影响。因此，开发通信效率高的方法就显得尤为重要。通常可以从降低传输频率和减少每轮传输的信息量着手。降低传输频率主要依靠减少客户端与中心服务器梯度的交换次数，为此可以适当提高一次全局迭代中客户端本地优化的次数。而减少信息量则主要依靠降低客户端与中心服务器的交换次数来实现。为此可以进行适当的梯度压缩或者量化，以减少通信占用的带宽。

● 系统异构难题。在联邦学习系统中，另一大问题就是众多客户端设备之间的异构性，包括存储、CPU计算能力、网络传输等多个方面的差异。这些异构性使得设备的计算时间不同，甚至导致个别设备直接掉线。异步通信解决了设备完成一次本地更新的时间不同、中心服务器等待过久的问题。此前分布式机器学习的研究已经充分应用了异步通信的方式。此外，提升系统的鲁棒性同样也能减轻系统异构对联邦学习产生的影响。在众多设备参与的情况下，需要提高系统的容错能力，提升系统的冗余度。

● 数据异构难题。联邦学习中设备经常以非独立同分布的方式在网络中生成和收集数据，例如，移动端的用户在进行输入法下一单词预测的任务时，使用不同的语言会导致数据异构问题。此外，跨设备的数据持有方持有的数据数量很可能分布不均匀。因此，许多常见的针对独立同分布数据假设的优化算法对于联邦学习来说都是不适用的。因此，如何使优化算法更加兼容联邦学习实际使用中复杂的数据结构，成为联邦学习未来发展的一个研究方向。元学习和多任务学习的思想都支持个性化或基于特定设备的建模，是处理数据统计异质性的一种有效的方法。元学习通过使参与联邦学习的各客户端本地模型学习独立但相关的模型，实现各个参与方本地模型的个性化，是一种应对联邦学习数据异构性的可行方案。

最后，笔者针对联邦学习的未来发展提出以下展望。

● 增加算法的联邦部署。本文讨论了目前存在的联邦学习算法，但是有关机器学习、深度学习算法在联邦框架下的部署研究问题还处于发展阶段。使用联邦学习框架进行机器学习、深度学习算法实现是人工智能领域落地的一个可行方案，也是更高效、更全面的边缘数据利用方法。

● 联邦学习的隐私性保证。数据隐私性的保证是联邦学习理念的关键点之一。尽管目前有许多与联邦学习隐私性相关的研究，但是在联邦学习实际应用的过程中，依然会面临许多复杂的隐私性挑战。联邦学习系统需要时刻提升对各类不良攻击的防御能力，保障用户数据的隐私性。

● 联邦学习的多领域协同发展。联邦学习的系统发展与多个领域有所关联，如边缘计算[62-63]、区块链[64-65]、网络安全[66-67]等。多领域的协同发展可以提升联邦学习

的性能，同时更好地发挥联邦学习的便捷性、隐私性等优势。

参考文献:

[1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[2] 王健宗, 黄章成, 肖京. 人工智能赋能金融科技[J]. 大数据, 2018, 4(3): 114-119.
WANG J Z, HUANG Z C, XIAO J. Artificial intelligence energize Fintech[J]. Big Data Research, 2018, 4(3): 114-119.

[3] KAIROUZ P, MCMAHAN H B, AVENT B, et al. Advances and open problems in federated learning[J]. arXiv preprint, 2019, arXiv:1912.04977.

[4] YANG Q, LIU Y, CHEN T, et al. Federated machine learning: concept and applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

[5] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

[6] MEHMOOD A, NATGUNANATHAN I, XIANG Y, et al. Protection of big data privacy[J]. IEEE Access, 2016, 4: 1821-1834.

[7] 方滨兴, 贾焰, 李爱平, 等. 大数据隐私保护技术综述[J]. 大数据, 2016, 2(1): 1-18.
FANG B X, JIA Y, LI A P, et al. Privacy preservation in big data: a survey[J]. Big Data Research, 2016, 2(1): 1-18.

[8] KONEČNÝ J, MCMAHAN H B, RAMAGE D, et al. Federated optimization: distributed machine learning for on-device intelligence[J]. arXiv preprint, 2016, arXiv:1610.02527.

[9] KONEČNÝ J, MCMAHAN H B, YU F X, et al. Federated learning: strategies for improving communication efficiency[J]. arXiv preprint, 2016, arXiv:1610.05492.

[10] MCMAHAN H B, MOORE E, RAMAGE D, et al. Federated learning of deep networks using model averaging[J]. arXiv preprint, 2016, arXiv:1602.05629.

[11] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Conference on Artificial Intelligence and Statistics. [S.l.:s.n.], 2017.

[12] LI T, SANJABI M, BEIRAMI A, et al. Fair resource allocation in federated learning[J]. arXiv preprint, 2019, arXiv:1905.10497.

[13] CHEN Y, SUN X Y, JIN Y C. Communication-efficient federated deep learning with layer wise asynchronous model update and temporally weighted aggregation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019: Accepted.

[14] REHAK D R, DODDS P, LANNOM L. A model and infrastructure for federated learning content repositories[C]// Interoperability of Web-Based Educational Systems Workshop. [S.l.:s.n.], 2005.

[15] LI M, ANDERSEN D G, PARK J W, et al. Scaling distributed machine learning with the parameter server[C]//The 11th USENIX Symposium on Operating Systems Design and Implementation. [S.l.:s.n.], 2014: 583-598.

[16] LIN Y J, HAN S, MAO H Z, et al. Deep gradient compression: reducing the communication bandwidth for distributed training[J]. arXiv preprint, 2017, arXiv:1712.01887.

[17] DAI W, KUMAR A, WEI J, et al. High-performance distributed ML at scale through parameter server consistency models[C]//AAAI Conference on Artificial Intelligence. New York: ACM Press, 2015.

[18] RECHT B, RE C, WRIGHT S, et al. Hogwild: a lock-free approach to parallelizing stochastic gradient descent[C]//Advances in Neural Information Processing Systems. [S.l.:s.n.], 2011: 693-701.

[19] HO Q, CIPAR J, CUI H G, et al. More effective distributed ml via a stale synchronous parallel parameter server[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2013: 1223-1231.

[20] FENG S W, YU H. Multi-participant multi-class vertical federated learning[J]. arXiv preprint, 2020, arXiv:2001.11154.

[21] KONEČNÝ J. Stochastic, distributed and federated optimization for machine learning[J]. arXiv preprint, 2017, arXiv:1707.01155.

[22] LIU X Y, LI H W, XU G W, et al. Adaptive privacy-preserving federated learning[J]. Peer-to-Peer Networking and Applications, 2020.

[23] HU R, GONG Y M, GUO Y X. CPFed: communication-efficient and privacy-preserving federated learning[J]. arXiv preprint, 2020, arXiv:2003.13761.

[24] RYFFEL T, TRASK A, DAHL M, et al. A generic framework for privacy preserving deep learning[J]. arXiv preprint, 2018, arXiv:1811.04017.

[25] ANTONIOUS M, DEEPESH D, SUHAS D, et al. Shuffled model of federated learning: privacy, communication and accuracy trade-offs[J]. arXiv preprint, 2020, arXiv:008.07180.

[26] SMITH V, CHIANG C K, SANJABI M, et al. Federated multi-task learning[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2017: 4424-4434.

[27] CORINZIA L, BUHMANN J M. Variational federated multi-task learning[J]. arXiv preprint, 2019, arXiv:1906.06268.

[28] CALDAS S, SMITH V, TALWALKAR A. Federated kernelized multi-task learning[C]//SysML Conference 2018. [S.l.:s.n.], 2018.

[29] KALLMAN R, KIMURA H, NATKINS J, et al. H-store: a high-performance, distributed main memory transaction processing system[J]. Proceedings of the VLDB Endowment, 2008, 1(2): 1496-1499.

[30] YANG K, JIANG T, SHI Y M, et al. Federated learning via over-the-air computation[J]. IEEE Transactions on Wireless Communications, 2020, 19(3): 2022-2035.

[31] NISHIO T, YONETANI R. Client selection for federated learning with heterogeneous resources in mobile edge[C]//2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1-7.

[32] WANG J Y, SAHU A K, YANG Z Y, et al. MATCHA: speeding up decentralized SGD via matching decomposition sampling[J]. arXiv preprint, 2019, arXiv:1905.09435.

[33] REISIZADEH A, MOKHTARI A, HASSANI H, et al. Fedpaq: a communication-efficient federated learning method with periodic averaging and quantization[J]. arXiv preprint, 2019, arXiv:1909.13014.

[34] KHALED A, MISHCHENKO K, RICHTÁRIK P. Better communication complexity for local SGD[J]. arXiv preprint, 2019, arXiv:1909.04746.

[35] LI S Y, CHENG Y, LIU Y, et al. Abnormal client behavior detection in federated learning[J]. arXiv preprint, 2019, arXiv:1910.09933.

[36] SATTLER F, WIEDEMANN S, MÜLLER K R, et al. Robust and communication-efficient federated learning from non-IID data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019.

[37] CROTTY A, GALAKATOS A, KRASKA T. Tupleware: distributed machine learning on small clusters[J]. IEEE Data Engineering Bulletin, 2014, 37(3): 63-76.

[38] JOLFAEI A, OSTOVARI P, ALAZAB M, et al. Guest editorial special issue on privacy and security in distributed edge computing and evolving IoT[J]. IEEE Internet of Things Journal, 2020, 7(4): 2496-2500.

[39] SAHU A K, LI T, SANJABI M, et al. Federated optimization for heterogeneous networks[J]. arXiv preprint, 2018, arXiv:1812.06127.

[40] YANG K, FAN T, CHEN T J, et al. A quasi-newton method based vertical federated learning framework for logistic regression[J]. arXiv preprint, 2019, arXiv:1912.00513.

[41] YANG S W, REN B, ZHOU X H, et al. Parallel distributed logistic regression for vertical federated learning without third-party coordinator[J]. arXiv preprint, 2019, arXiv:1911.09824.

[42] GAO D S, JU C, WEI X G, et al. HHHFL: hierarchical heterogeneous horizontal federated learning for electroencephalography[J]. arXiv preprint, 2019, arXiv:1909.05784.

[43] LIU Y, KANG Y, ZHANG X W, et al. A communication efficient vertical federated learning framework[J]. arXiv preprint, 2019, arXiv:1912.11187.

[44] SHARMA S, XING C P, LIU Y, et al. Secure and efficient federated transfer learning[J]. arXiv preprint, 2019, arXiv:1910.13271.

[45] ZHAO Y, LI M, LAI L Z, et al. Federated learning with non-IID data[J]. arXiv preprint, 2018, arXiv:1806.00582.

[46] LIU Y, LIU Y T, LIU Z J, et al. Federated forest[J]. IEEE Transactions on Big Data, 2020: Accepted.

[47] CHENG K W, FAN T, JIN Y L, et al. SecureBoost: a lossless federated learning framework[J]. arXiv preprint, 2019, arXiv:1901.08755.

[48] LI Q B, WEN Z Y, HE B S. Practical federated gradient boosting decision trees[J]. arXiv preprint, 2019, arXiv:1911.04206.

[49] HARTMANN V, MODI K, PUJOL J M, et al. Privacy-preserving classification with secret vector machines[J]. arXiv preprint, 2019, arXiv:1907.03373.

[50] ZHU X H, WANG J, HONG Z, et al. Federated learning of unsegmented Chinese text recognition model[C]//2019 IEEE 31st International Conference on Tools with Artificial Intelligence. Piscataway: IEEE Press, 2019: 1341-1345.

[51] BHOWMICK A, DUCHI J, FREUDIGER J, et al. Protection against reconstruction and its applications in private federated learning[J]. arXiv preprint, 2018, arXiv:1812.00984.

[52] DUCHI J, JORDAN M I, MCMAHAN B. Estimation, optimization, and parallelism when data is sparse[C]//In Advances in Neural Information Processing Systems. New York: ACM Press, 2013.

[53] CHILIMBI T, SUZUE Y, APACIBLE J, et al. Project adam: building an efficient and scalable deep learning training system[C]//The 11th USENIX Symposium on Operating Systems Design and Implementation. New York: ACM Press, 2014: 571-582.

[54] LIU Y, MUPPALA J K, VEERARAGHAVAN M, et al. Data center networks: topologies, architectures and fault-tolerance characteristics[M]. Heidelberg: Springer Science & Business Media, 2013.

[55] BONAWITZ K, EICHNER H, GRIESKAMP W, et al. Towards federated learning at scale: system design[J]. arXiv preprint, 2019, arXiv:1902.01046.

[56] LI X, HUANG K, YANG W, et al. On the convergence of FedAvg on non-IID data[J]. arXiv preprint, 2019, arXiv:1907.02189.

[57] CALDAS S, KONEČNY J, MCMAHAN H B, et al. Expanding the reach of federated learning by reducing client resource requirements[J]. arXiv preprint, 2018, arXiv:1812.07210.

[58] NISHIO T, YONETANI R. Client selection for federated learning with heterogeneous resources in mobile edge[C]//ICC 2019-2019 IEEE International Conference on Communications. Piscataway: IEEE Press, 2019: 1-7.

[59] YOSHIDA N, NISHIO T, MORIKURA M, et al. Hybrid-FL for wireless networks: cooperative learning mechanism using non-IID data[J]. arXiv preprint, 2019, arXiv:1905.07210.

[60] SPRAGUE M R, JALALIRAD A, SCAVUZZO M, et al. Asynchronous federated learning for geospatial applications[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Heidelberg: Springer, 2018: 21-28.
[61] XIE C, KOYEJO S, GUPTA I. Asynchronous federated optimization[J]. arXiv preprint, 2019, arXiv:1903.03934.
[62] YANG J L, DUAN Y X, QIAO T, et al. Prototyping federated learning on edge computing systems[J]. Frontiers of Computer Science, 2020, 14: 1-3.
[63] WANG S Q, TUOR T , SALONIDIS T, et al. Adaptive federated learning in resource constrained edge computing systems[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1205-1221.
[64] ZHAO Y, ZHAO J, JIANG L S, et al. Mobile edge computing, blockchain and reputation-based crowd-sourcing IoT federated learning: a secure, decentralized and privacy-preserving system[J]. arXiv preprint, 2019, arXiv:1906.10893.
[65] LI Z Y, LIU J, HAO J L, et al. CrowdSFL: a secure crowd computing framework based on blockchain and federated learning[J]. Electronics, 2020, 9(5): 773.
[66] KANG J W, XIONG Z H, NIYATO D, et al. Incentive design for efficient federated learning in mobile networks: a contract theory approach[C]//2019 IEEE VTS Asia Pacific Wireless Communications Symposium. Piscataway: IEEE Press, 2019: 1-5.
[67] ISAKSSON M, NORRMAN K. Secure federated learning in 5G mobile networks[J]. arXiv preprint, 2020, arXiv: 2004.06700.

作者简介

王健宗（1983- ），男，博士，平安科技（深圳）有限公司副总工程师，资深人工智能总监，联邦学习技术部总经理。美国佛罗里达大学人工智能博士后，中国计算机学会（CCF）高级会员，CCF大数据专家委员会委员，曾任美国莱斯大学电子与计算机工程系研究员，主要研究方向为联邦学习和人工智能等。

孔令炜（1995- ），男，平安科技（深圳）有限公司联邦学习团队算法工程师，CCF会员，主要研究方向为联邦学习系统和安全通信等。

黄章成（1990- ），男，平安科技（深圳）有限公司联邦学习团队资深算法工程师，人工智能专家，CCF会员，主要研究方向为联邦学习、分布式计算及系统和加密通信等。

陈霖捷（1994- ），男，平安科技（深圳）有限公司联邦学习团队算法工程师，主要研究方向为联邦学习与隐私保护、机器翻译等。

刘懿（1994- ），女，平安科技（深圳）有限公司联邦学习团队算法工程师，主要研究方向为联邦学习系统等。

何安珣（1990- ），女，平安科技（深圳）有限公司联邦学习团队高级算法工程师，CCF会员，主要研究方向为联邦学习技术在金融领域的落地应用、联邦学习框架搭建、加密算法研究和模型融合技术。

肖京（1972- ），男，博士，中国平安保险（集团）股份有限公司首席科学家。2019年吴文俊人工智能科学技术奖杰出贡献奖获得者，CCF深圳会员活动中心副主席，主要研究方向为计算机图形学学科、自动驾驶、3D显示、医疗诊断、联邦学习等。

收稿日期： 2020-04-21
基金项目： 国家重点研发计划基金资助项目（No.2018YFB1003503，No.2018YFB0204400，No.2017YFB1401202）
Foundation Items: The National Key Research and Development Program of China (No.2018YFB1003503, No.2018YFB0204400, No.2017YFB1401202)