February | 02 2024 278

联邦学习法助力隐私计算新应用

■ 秦鹏 上官丽丽 | 文

隐私计算前景广阔

随着科技的飞速发展，人们通过智能设备和互联网交互产生的海量数据已经成为重要的生产要素，数据流通和数据开放逐渐成为企业业务创新和实现增长的必选项。但随之频发的数据泄露事件也引起了世界各地政府的重视，相继就信息保护出台相关政策和法规。我国于 2021 年 6 月 10 日通过《中华人民共和国数据安全法》，2023 年 3 月中共中央、国务院印发的《党和国家机构改革方案》中规划组建国家数据局……数据监管和建设力度逐年加大。

隐私计算作为平衡隐私和创新发展的有效手段，近年来得到快速发展。根据 IDC 发布的《IDC MarketShare：中国隐私计算平台市场份额，2022》市场调研显示：2022 年，中国隐私计算平台市场以 92.9% 的市场增速实现 1.2 亿美元的市场规模，市场前景极为广阔。

隐私计算的原理

隐私计算（Privacy Computing）是在保护数据本身不对外泄露的前提下实现数据分析计算的技术集合。

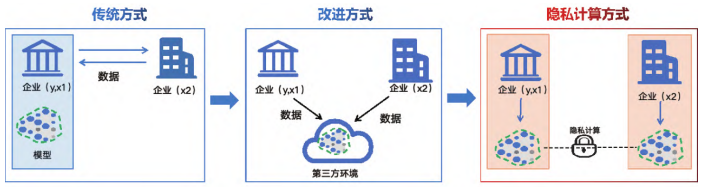

图 1 隐私计算实现逻辑

隐私计算主要包括数据加密、数据切分、安全计算、安全通信四个方面，通过技术切割，保证在计算过程中的数据隐私和计算结果的安全性。

隐私计算主要技术方向

隐私计算技术有三个主要方向：第一个方向是基于密码学的隐私计算技术，其中以多方安全计算（Secure Multi-party Computation）为代表；第二个方向是人工智能与隐私保护技术的结合，以联邦学习（Federated Learning）为代表；第三个方向是基于可信硬件的隐私计算技术，其中以可信执行环境（Trusted Execution Environment）为代表。这些不同技术经常组合使用，既保证了原始数据的安全和隐私性，又能够完成对数据的计算和分析任务。

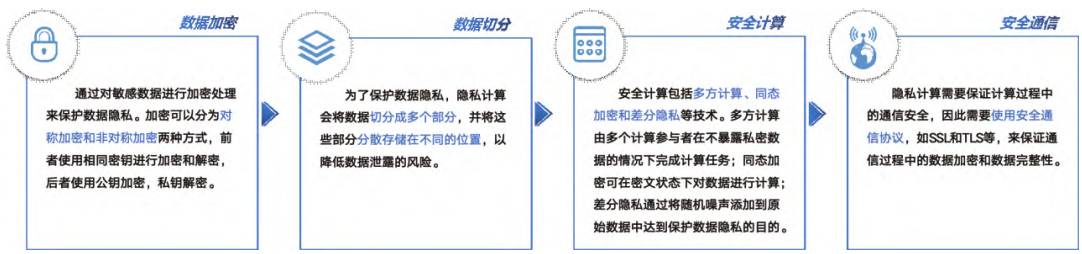

图 2 隐私计算原理

图 3 隐私计算三大技术方向

中国隐私计算平台市场主要特征

随着近几年隐私计算的迅猛发展，我国隐私计算平台市场呈现出以下三个主要特征。

硬件加速和互联互通成为厂商技术升级的重点任务。为解决隐私计算分布式技术架构计算效率问题，技术提供方结合算法协议改造和硬件加速等多种能力提高产品性能。同时，采用通用接口解决平台底座和组件服务之间的兼容性问题，打破数据孤岛，实现数据流通。

金融与营销成为典型的应用行业和场景。从行业角度来看，金融行业的营销场景和风控场景、政务数据共享场景以及通信行业运营商数据价值转化场景仍是隐私计算项目落地的重点场景。其中营销类场景通常使用联邦学习进行联合建模，以提高模型预测的准确性，提升营销拉新效果；在数据开放等场景中，多方企业对匿踪查询类应用的需求逐渐攀升。

在传统行业的企业数据交换流程中实现规模发展仍需时日。隐私计算通常由企业的技术部门引入进行试点探索，但是结合业务的上线运营就需要业务部门与合规部门进行双重合规性审核，其间，需要面对的额外风险成为部分企业以技术创新推动业务发展的较大掣肘因素。

联邦学习的发展和应用

联邦学习（FL）作为隐私计算行之有效的重要技术之一，其本质是一种分布式机器学习技术，其核心思想是在保障底层数据隐私安全及合法合规的前提下，通过交换加密的机器学习中间结果实现联合建模。

联邦学习兼顾人工智能应用与隐私保护，促进跨机构的数据共享融合，实现多个机构间构建统一的安全、高效、合规的多源数据应用生态系统。因其开放合作、协同性高，能够为金融、消费、互联网等行业的业务创新场景提供更丰富、更高质量的大数据服务，使得其在多个行业得到广泛应用。

联邦学习的三种类型

联邦学习涉及两个不同的概念：第一种是谷歌提出的联邦学习，旨在解决“云”和“端”训练过程中隐私保护问题，适用于面向消费者的数据（To C），通常为数据水平切分的场景。第二种是我国提出的联邦学习，主要用于解决企业间合作中各方隐私保护的问题，适用于面向企业的数据（To B），既可应用于数据

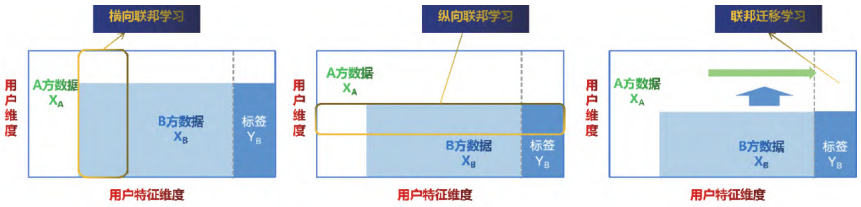

图 4 联邦学习的类型

的水平切分场景，也可应用于数据的垂直切分场景。这两种联邦学习的概念主要体现在数据维度不同。根据数据维度不同，联邦学习分为横向联邦学习、纵向联邦学习、联邦迁移学习三种。

基于实例的横向联邦学习：本质是样本的联合，适用于参与者间业态相同但触达客户不同的场景，即两个数据集的用户特征重叠较多而用户重叠较少。数据集按照用户维度切分后，取出双方用户特征相同而用户不完全相同的部分数据进行训练。

基于特征的纵向联邦学习：本质是特征的联合，适用于参与者间样本相同但业态不同的场景，即两个数据集的用户重叠较多而用户特征重叠较少。数据集按照特征维度切分后，取出双方用户相同而用户特征不完全相同的部分数据进行训练。

基于模型的联邦迁移学习：核心是找到源领域和目标领域之间的相似性，适用于两个数据集的用户与用户特征重叠都较少的情况。不对数据进行切分，利用数据、任务或模型之间的相似性，将在源领域学习过的模型应用于目标领域，例如人类学会了打羽毛球，也可以尝试学会打网球的迁移学习能力。

联邦学习的两类框架

联邦学习的架构分为两种：中心化联邦（客户端 / 服务器）架构和去中心化联邦（对等计算）架构。在联合多方用户的联邦学习场景中，通常采用中心化联邦架构，其中企业作为服务器，负责协调全局模型。而在面对联合多家数据孤岛企业进行模型训练的场景中，由于较难从多家企业中选出协调服务器方，通常采用去中心化联邦架构。

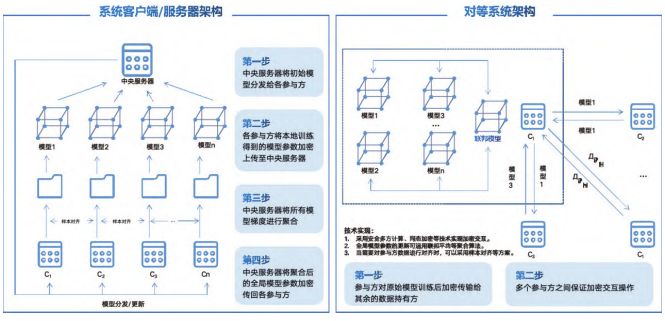

图 5 联邦学习的系统框架对比

联邦学习的发展应用

近年来，国内外积极开展联邦学习的应用研究。美国为联邦学习的发源地，在信用卡反欺诈和跨行反洗钱领域，美国已有金融机构摩根大通（JP Morgan）、科技公司 IBM 以及学术机构开展联邦学习的理论探讨和建模测试。我国在联邦学习方面的应用起步稍晚，但后发优势逐渐显现。目前已有的联邦学习平台包括腾讯“神盾 – 联邦计算平台”、微众银行 FATE、百度“百度金融安全计算平台”、京东科技 Fedlearn、平安科技“蜂巢联邦智能平台”和华控清交“PrivPy 多方计算平台”等。

另外，在联邦学习 + 区块链深度融合方面，我国也取得了一些阶段性进展。目前已建平台有：多方大数据隐私计算平台 WeDPR-PPC（微众银行）、区块链网络平台 FAIR（蚂蚁链）、金融业数据共享平台（趣链科技）、政府税务数据平台（八分量）等。

在业务应用方面，除联合营销、政务数据安全开放、联合风控、反欺诈等场景外，联邦学习在计算机视觉领域、自然语言处理和推荐系统领域也有突出进展，应用于汽车自动驾驶、智能家居系统、智慧城市等场景。

山西移动在联邦学习的应用研究

在政府政策支持快速推出的大背景下，山西移动积极跟进隐私计算技术，开展深入研究，与山西晋商银行达成深度合作，通过联邦学习解决通信数据在金融营销的应用问题。

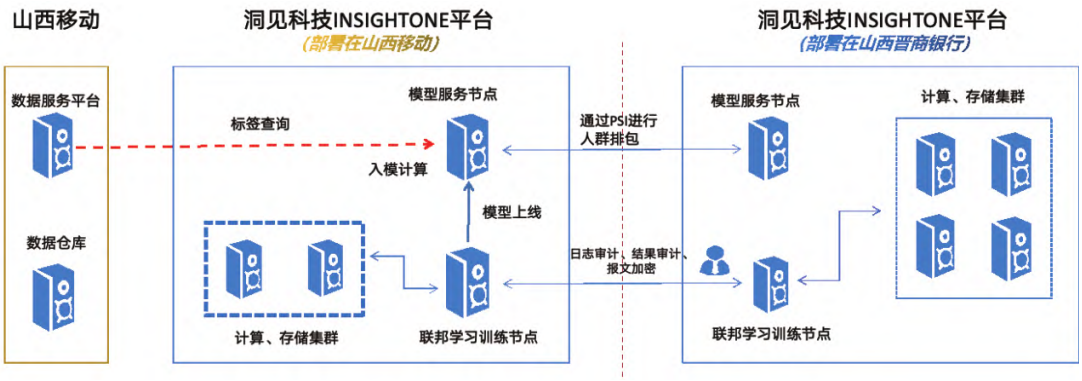

图 6 山西移动—晋商银行隐私计算架构图

应用实施方案

在明确山西晋商银行的分析目标后，山西移动就平台资源的性能、节点安全、算法安全进行测试，评估项目的可行性与合规性，双方按照约定的加密方式进行数据加密，基于山西移动洞见数智联邦平台（INSIGHTONE）建立联合模型，在进行参数优化和模型评估后输出可供银行侧营销的数据。在得出模型后，洞见数智联邦平台将数据进行销毁，保证晋商银行和山西移动双方的样本数据安全。

应用成效

山西移动通过联合模型为银行补充客户在运营商侧所表现出来的金融属性，以便银行建立更全面的客户洞察。晋商银行通过联合模型识别信用卡客户的分期意愿，选出具有高分期意愿的客户进行重点营销，提升信用卡客户的分期比例及提高用户的分期额度。

目前，山西移动向晋商银行共计推送 338 万用户，银行侧平均成功收单率达 5.03%，较常规营销提升 6 倍，创造大数据收入 138 万元。在保证数据价值安全释放的基础上，实现了数据价值的商业变现，达到了双方共赢的良性发展目标，助力数字经济快速发展。

更多应用展望

依托山西晋商银行异业合作项目的流程跑通，山西移动就信用卡分期意愿识别、贷中监测等场景将在银行侧、证券侧等其他公司进行纵向应用推广。另外，还将就政府数据安全开放、公安部反欺诈、互联网金融信用评分、数字网关智能选址等场景进行合作研究及横向应用推广。 CTT

作者单位：中国移动山西分公司、中国信息通信研究院