



智谱·AI ×



清华大学
Tsinghua University

清华大学-人工智能研究院知识智能研究中心
清华-中国工程院知识智能联合研究中心



2021联邦学习全球研究 与应用趋势报告

2021 Global Report of Federated Learning Research and
Application Trend

2021.09

数据支持：AMiner.cn

主要发现

联邦学习是一种新兴的人工智能基础技术。本报告从技术研究、行业应用、学者画像以及发展趋势等维度，较为全面深入地分析了联邦学习自2016年诞生至2020年的重要发展成就。

联邦学习“中美双雄”格局显现

- 中美两国有关联邦学习的论文发布量遥遥领先于其他国家；在论文发布量TOP 10机构中，中美各占4席和3席；中美两国论文合作数量也全球最多，且半数以上的高被引论文来自中美两国，但美国的论文引用量显著领先，中国位居第二。
- 全球专利受理数量以中国地区最多，共1514项；美国位居第二，共579项。在专利申请数量TOP10机构中，中国占7席，美国占3席。
- 开源框架主要来自中美，其中OpenMined 推出的Pysyft、微众银行的FATE和谷歌的TFF框架的热度居于全球前三位。
- 联邦学习领域的全球学者共计2,764名，中美分别拥有816和817名，各占全球总量的30%。

未来联邦学习研究趋势将与算法模型和安全隐私技术相关

- 目前联邦学习研究热点主要聚焦在机器学习方法、模型训练、隐私保护三方面。
- 未来几年研究趋势将与算法模型和安全隐私技术相关，如Edge Computing（边缘计算）、Data Heterogeneity（数据异质性）、Internet Of Things（物联网）、Blockchain（区块链）、Wireless Communication（无线通信）、Communication Efficiency（沟通效率）等。
- 行业应用研究方向呈现出不断与区块链、物联网、车辆交互、5G等技术融合的态势。

人工智能之联邦学习—— 《2021联邦学习全球研究与应用趋势报告》

编写团队

顾问

李涓子 清华大学人工智能研究院知识智能中心
唐杰 清华大学人工智能研究院知识智能中心

编写团队

张淼 张建伟 张淳

数据

仇瑜 赵慧军

版式设计

边云风

目录

报告说明

- 数据范围
- 联邦学习知识树

引言

- 人工智能可持续发展面临的困境
- 联邦学习概念的介绍

联邦学习技术研究与应用现状

- 技术研究现状
- 联邦学习框架与系统现状
- 联邦学习行业应用现状

联邦学习发展趋势

- 研究趋势
- 技术成熟度
- 市场化与商业化趋势
- 推行联邦学习的国内外标准
- 建立联邦学习生态

结语

联邦学习 (Federated Learning) 是在进行分布式机器学习的过程中, 各参与方可借助其他参与方数据进行联合建模和使用模型。参与各方无需传递和共享原始数据资源, 同时保护模型参数, 即在数据不出本地的情况下, 进行数据联合训练、联合应用, 建立合法合规的机器学习模型^[1]。

联邦学习是一种新兴的人工智能基础技术, 其概念于 2016 年由谷歌公司 H. Brendan McMahan 在论文 *Federated Learning of Deep Networks using Model Averaging* ^[2] 中最先提出, 原本用于解决安卓手机终端用户在本地更新模型的问题, 后经香港科技大学与微众银行杨强教授所领导团队在 2018 年将其扩展为机构间 B2B 分布式联合建模架构, 包括按样本、特征分割以及异构多方建模, 同时可以建立去中心协调器的 Peer-to-Peer 架构形式, 其设计目标是在保障大数据交换时的信息安全、保护终端数据和个人数据隐私、保证合法合规的前提下, 在多参与方或多计算结点之间开展高效率, 安全、可靠的机器学习。联邦学习同时包括鼓励多方持续参与合作生态的激励机制, 建立正向激励的数据价值交易市场机制。当下, 联邦学习已经被大量应用于金融^[3]、安防^[4]、医疗^[5]、在线推荐系统^[6]等领域。联邦学习有望成为下一代人工智能协同算法,

隐私计算和协作网络的基础。

《2021 联邦学习全球研究与应用趋势报告》主要从技术研究、学者画像、主流框架、行业应用, 以及发展趋势几大方面, 较为全面深入地介绍联邦学习自 2016 年诞生以来到 2020 年的技术研究和应用进展, 并展望该技术的未来发展方向与前景。

(一) 数据范围

本报告研究数据范围是科技情报大数据挖掘与服务系统平台 AMiner 数据库所收录的 2016-2020 年期间与联邦学习研究主题强相关的论文数据、专利数据以及公开数据等。论文的引用量数据统计截止日期为 2021 年 5 月 31 日。

(二) 联邦学习知识树

本报告根据联邦学习的关键技术和相关技术, 利用 AMiner 数据库中近年来该领域的高水平学术论文, 挖掘出了全球活跃的联邦学习的重要技术点, 并表征为知识树结构, 如图 1 所示。

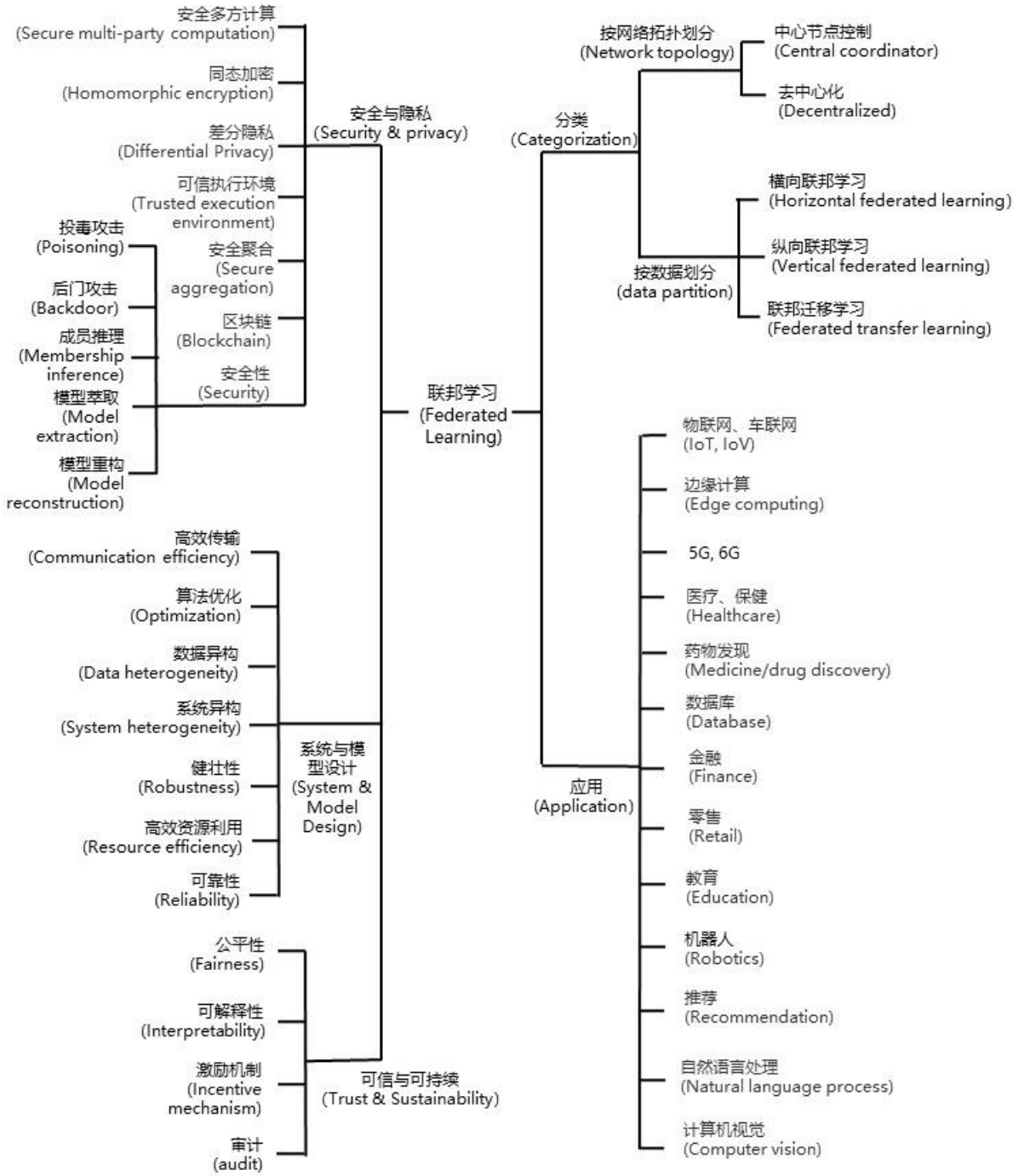
^[1]杨强、刘洋、陈天健等:《联邦学习》, 载《中国计算机学会通讯》, 2018 年版, 第 49-55 页。

^[2]McMahan, H. B., Moore, E., Ramage, D., & y Arcas, B. A. (2016). Federated learning of deep networks using model averaging. arXiv preprint arXiv:1602.05629.

^[3] <https://www.fedai.org/cases/utilization-of-fate-in-anti-money-laundering-through-multiple-banks/>

^[4] Liu, Y., Huang, A., Luo, Y., Huang, H., Liu, Y., Chen, Y., Feng, L., Chen, T., Yu, H., & Yang, Q. (2020). "FedVision: An Online Visual Object Detection Platform Powered by Federated Learning," Proceedings of the AAAI Conference on Artificial Intelligence, 34(08), 13172-13179.

图 1 联邦学习知识树



^[5] Li W. et al. "Privacy-Preserving Federated Brain Tumour Segmentation," In: Suk HI., Liu M., Yan P, Lian C. (eds) Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol 11861. Springer, Cham.

^[6] Ben Tan, Bo Liu, Vincent Zheng, and Qiang Yang. 2020. A Federated Recommender System for Online Services. In Fourteenth ACM Conference on Recommender Systems (RecSys '20). Association for Computing Machinery, New York, NY, USA, 579–581. DOI:<https://doi.org/10.1145/3383313.3411528>

人工智能未来能否可持续发展面临两大困境。

一是数据困境。人工智能和机器学习算法具有对数据强依赖的特性。现实中，多数行业领域存在着数据有限且质量较差的问题，并且以碎片化的形式分散存在，不足以支撑人工智能技术的实现。同时，数据源之间存在着难以打破的壁垒。由于行业竞争、隐私安全、行政手续复杂等问题，数据还是以孤岛形式存在的。此外，研究界和企业界目前的情况是收集数据的一方通常不是使用数据的一方。因此，将分散在各地、各机构的数据进行整合用于机器学习所需的成本非常巨大。

二是法律挑战。当前，重视数据隐私和安全已经成为世界性的趋势，各国都在不断地推出和加强完善对数据安全和隐私保护的相关法规。欧盟2018年正式施行《通用数据保护条例》（General Data Protection Regulation, GDPR）。在中国，全国信息安全标准委员会先后于2017年12月和2020年3月发布了两版《信息安全技术个人信息安全规范》（GB/T 35273-2017、GB/T 35273-2020），对个人信息收集、储存、使用做出了明确规定。此外，在2017年起实施的《中华人民共和国网络安全法》^[7]和《中华人民共和国民法总则》^[8]中也指出网络运营者不得泄露、篡改、毁坏其收集的个人信息，并且与第三方进行数据交易时需确保在合同中明确约定拟交易数据的范围和数据保护义务。

针对以上困境，“狭义”联邦机器学习的概念于2016年由谷歌研究人员首先提出，随后成为一个解决数据孤岛问题、满足隐私保护和数据安全的一个可行性解决方案^[9]。联邦学习的特征是数据不出本地、各个参与者的身份和地位平等、它能够实现多个参与方在保护数据隐私、满足合法合规要求的前提下进行机器学习，协同地进行模型训练与结果预测，并且建模效果和将整个数据集放在一处建模的效果相同或相差不大（在各个数据的用户对齐（user alignment）或特征对齐（feature alignment）的条件下），从而实现企业间的数据融合建模，解决数据孤岛问题。

“广义”联邦学习的概念，由香港科技大学杨强教授所领导的微众银行AI团队在2018年提出，将联邦学习扩展为机构和个人间的B2C模式和不同机构间B2B分布式联合建模架构，包括按样本、按特征分割以及异构多方建模，同时可以建立去中心协调器的Peer-to-Peer架构形式，其设计目标是在保障大数据交换时的信息安全、保护终端数据和个人数据隐私、保证合法合规的前提下，在多参与方或多计算结点之间开展高效率，安全、可靠的机器学习和模型使用。联邦学习同时包括鼓励多方持续参与合作生态的激励机制，建立正向激励的数据价值交易市场机制。

^[7] 《中华人民共和国网络安全法》，中共中央网络安全和信息化委员会办公室、中华人民共和国国家互联网信息办公室，http://www.cac.gov.cn/2016-11/07/c_1119867116.htm

^[8] 《中华人民共和国民法总则》，中华人民共和国中央人民政府，http://www.gov.cn/xinwen/2017-03/18/content_5178585.htm#1

^[9] 杨强、刘洋、陈天健等：《联邦学习》，载《中国计算机学会通讯》，2018年版，第49-55页。

如上所述，根据孤岛数据的分布特点（用户与用户特征的重叠情况），联邦学习可以分为横向联邦学习、纵向联邦学习与联邦迁移学习^[10]。

联邦学习能够成功的一个重要根基，在于与激励机制、隐私保护等技术的融合。联邦学习激励机制研究的是如何量化每个参与方对数据联邦带来的收益，公平地与参与者分享部分收益以此作为激励，从而实现数据联邦长期的可持续经营^[11]。为了防止恶意攻击者通过模型反演等攻击手段复现原始数据，联邦学习通过与安全多方计算（Secure Multi-Party Computation, MPC）、同态加密（Homomorphic Encryption, HE）、差分隐私（Differential Privacy, DP）和可信执行环境（Trusted Execution Environment, TEE）等隐私计算技术相融合，进一步提升对数据的隐私保护。联邦学习与隐私计算技术的融合通常需要在模型精度、模型训练效率和数据安全性这三个维度之间进

行权衡和取舍。如何能够在这三个维度上得到综合性的提升，是联邦学习的一个热点研究方向。

联邦学习作为未来AI发展的底层技术，它依靠安全可信的数据保护措施下连接数据孤岛的模式，将不断推动全球AI技术的创新与飞跃。随着联邦学习在更大范围和更多行业场景中的渗透及应用，它不仅能辅助人类的工作及生活，也将逐步改变人类的认知模式，促进全社会智能化水平提升，并以“合作共赢”的模式带动跨领域的企业级数据合作，有效降低技术应用的成本和门槛，催生基于联合建模的新业态，进而推动社会经济及发展^[12]。

由于目前没有关于联邦学习技术发展的权威统计，本报告将主要回顾其从2016年诞生至2020年的技术发展趋势，作为学者们了解该技术进展的重要渠道。未来我们将定期进行该技术的阶段性回顾。



^[10] Liu Y, Chen T, Yang Q. Secure Federated Transfer Learning Framework[J]. IEEE Intelligent Systems, vol. 35, no. 4, pp. 70-82, 1 July-Aug. 2020.

^[11]杨强，刘洋，程勇，康焱，陈天健：《联邦学习》，电子工业出版社：北京，2020年:99-99.

^[12]微众银行人工智能部、鹏城实验室、腾讯研究院、中国信通院云大所、平安科技、招商局金融科技、电子商务与电子支付国家工程实验室(中国银联)：《联邦学习白皮书V2.0》，深圳，2020年，第5-7页。

联邦学习技术研究与应用现状

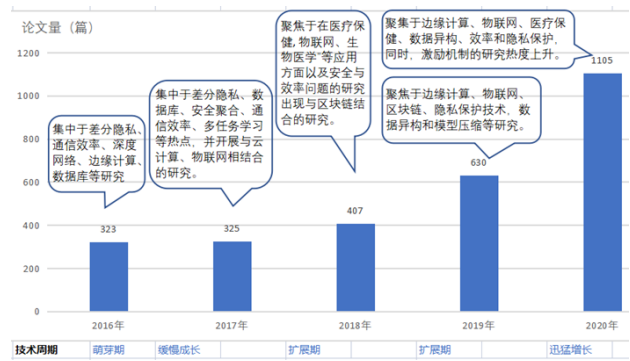
（一）技术研究现状

3.1.1 科研论文成果现状

（1）论文年度发表量不断增长

基于AMiner 系统，通过关键词组^[13]在标题和摘要中检索2016年至2020年论文数据。结果显示，研究时段内联邦学习相关论文共计2790篇，自2016年被提出以来，研究论文数量逐年增多，于2020年达到顶峰1105篇，相关论文趋势如图 2 所示。

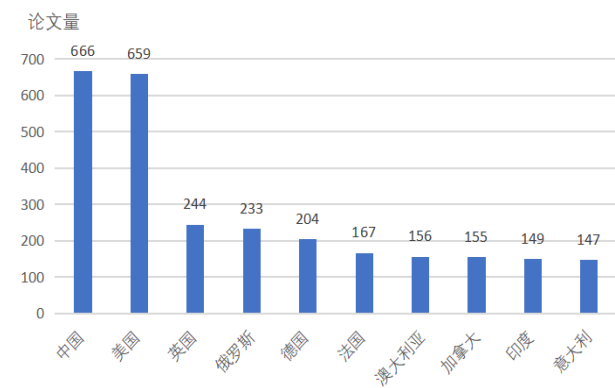
图 2 联邦学习研究论文趋势（2016-2020年）



（2）论文发布量以中美两国为引领

根据论文作者所在机构所属国家进行排序分析，发现近五年来，联邦学习论文发布量TOP 10国家是中国、美国、英国、俄罗斯、德国、法国、澳大利亚、加拿大、印度和意大利。相关论文量较突出的国家是中国（666篇）和美国（659篇），详细信息如图 3 所示。

图 3 联邦学习论文发表量TOP 10 国家（2016-2020年）



^[13]联邦学习关键词检索式：Federated Machine Learning OR Federated optimization OR federated learning OR federation learning OR (Privacy AND Distributed AND data mining) OR (Secure AND Distributed AND data mining) OR (Secure AND Multiparty) OR (Secure AND Multi-party) OR (privacy AND Multi-party) OR (privacy AND Multiparty) OR (Privacy AND Distributed AND machine learning) OR (Secure AND Distributed AND machine learning) OR (Privacy and joint learning) OR (Secure and joint learning) OR (Privacy AND Distributed AND deep learning) OR (Secure AND Distributed AND deep learning)

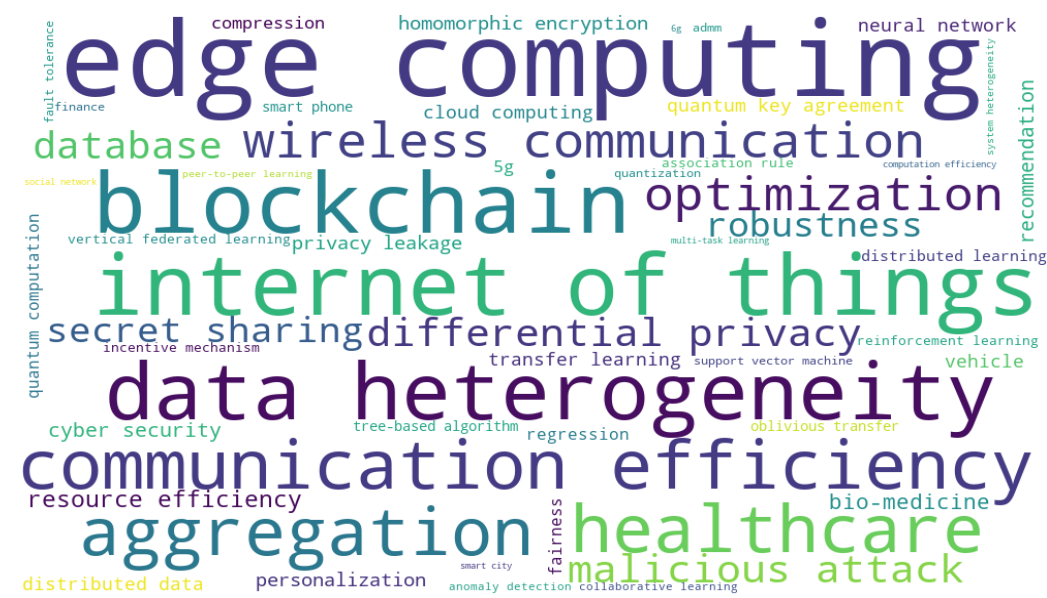
(3) 研究热点涵盖应用、系统和模型设计、安全隐私三个领域

① 总体研究热点

总体来看，基于AMiner 系统论文的热词分析，发现2016-2020年联邦学习领域的研究热点TOP 10 按热度递减依次包括：edge computing（边缘计算）、blockchain（区块链）、Internet of things（物联网）、data heterogeneity（数据异质性）、

communication efficiency（沟通效率）、healthcare（医疗保健）、aggregation（聚合）、wireless communication（无线通信）、optimization（优化）、differential privacy（差分隐私）等，如图4 所示。可见，在研究时段内，机器学习技术及相关算法模型等是联邦学习技术研究领域的主要热点，而当前较热门的激励机制在当时的热度还处于累积阶段。

图 4 联邦学习领域研究热点词云图（2016-2020年）



② 年度研究热点

分年度来看，联邦学习研究热点从机器学习到优化、从信息统计到量子密码、从数据隐私到行业应用，学者们不断探索落地联邦学习的方法，一方面是利

用交替方向乘法（ADMM）、量化、压缩等方式进行联邦学习算法优化，另一方面是引入区块链、密码学、物联网等技术建立全局共享的数据集，并对



2016

communication efficiency

differential privacy

biology medicine

deep network

graph computation

vertical federated learning

quantum signature

secret sharing

edge computing

quantum key agreement

quantum cryptography

association rule

incentive mechanism

fault tolerant

horizontal federated learning

database

logistic regression

support vector machine

cloud computing

online learning

homomorphic encryption

secure aggregation

byzantine problem

model averaging

information-theoretic secure

quantum computation

smartphone

此外，当时热点还包括 Support vector machine, graph computation, vertical federated learning 等。

2017

differential privacy

Quantum Key Agreement

secret sharing

bayesian learning

secure aggregation

collusion attack

anomaly detection

game theory

social network

Multi-task learning

reinforcement learning

communication efficiency

byzantine problem

association rule

expectation maximization

database

cloud computing

healthcare

obfuscation

ADMM

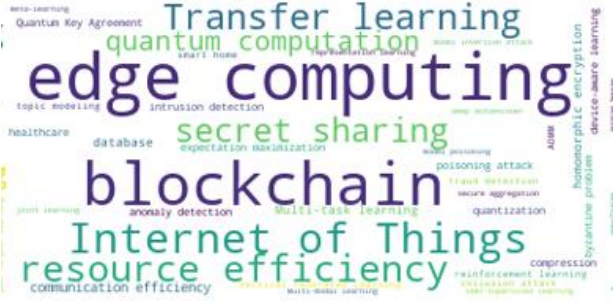
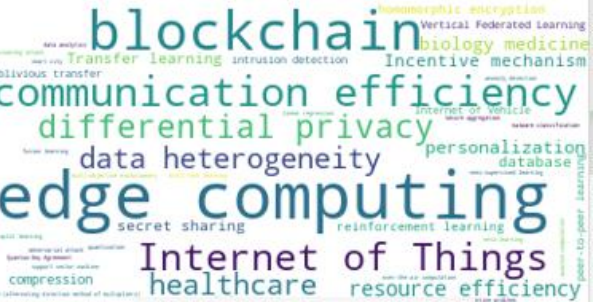
Internet of Things

quantum machine

2018 年联邦学习应用相关研究热度增加并居于前列，如 healthcare, Internet of Things, biology medicine, edging computing。同时，学者们依旧较关注 differential privacy, secret sharing, homomorphic encryption, Quantum Key Agreement, communication efficiency 等联邦学习安全与效率问题的研究。在这一阶段区块链“blockchain”技术成为热点，为联邦学习提供了保障用户隐私的新方法。

[illegible]

2018 年联邦学习应用相关研究热度增加并居于前列，如 healthcare, Internet of Things, biology medicine, edging computing。同时，学者们依旧较关注 differential privacy, secret sharing, homomorphic encryption, Quantum Key Agreement, communication efficiency 等联邦学习安全与效率问题的研究。在这一阶段区块链“blockchain”技术成为热点，为联邦学习提供了保障用户隐私的新方法。

2019 	2019 年，edge computing, Internet of Things, blockchain 成为热门研究领域，homomorphic encryption, secret sharing, secure aggregation 等隐私保护技术依旧受到高度关注。2019 年加大了对数据异构和模型压缩等技术的研究力度，transfer learning, multi-task learning, quantization, compression 成为热点研究方向。如何在联邦学习中有效地利用资源也成为 2019 年研究的热点之一。
2020 	2020 年，edge computing, Internet of things, healthcare 依旧是联邦学习的应用热点。联邦学习中的效率和隐私保护技术依旧是研究热点。2020 年加大了对联邦学习数据异构方向的研究，热门的研究点包括 personalization 和 transfer learning。同时，联邦学习激励机制 “Incentive mechanism” 的研究开始增加，以此激励更多机构加入数据联邦。

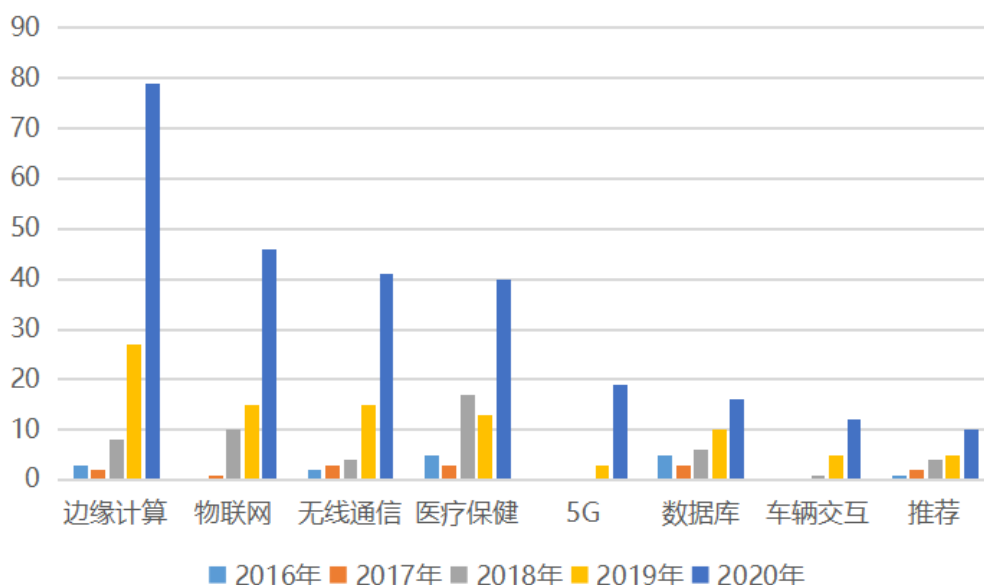
③ 主题热点趋势

通过TF-IDF算法对所研究时段内每一年的联邦学习主题相关论文数量进行计算，获取论文数量TOP 30 的热点词，然后聚合成联邦学习的应用（application）、系统和模型设计（system and model design）和安全隐私（secure and privacy）三个主题领域的研究热点集。这三个细分主题的研究趋势呈现出如下特征。

在应用研究领域，联邦学习的研究热点按照总热度由高到低依次包括边缘计算（edge computing）、物联网（Internet of things）、无线通信（wireless communication）、医疗保健（healthcare）、5G（第5代移动网络）、数据库

（database）、车辆交互（vehicle）以及推荐（recommendation），详细信息如图 5 所示。其中，在边缘计算、无线通信、医疗保健、数据库以及推荐方面的应用研究热度近五年来呈现出逐年上升的趋势；同时，边缘计算自2019年起成为该技术应用研究热度之榜首并保持至今。相比而言，数据库、医疗保健的研究热度在2016年与2017年的研究热度相对较高且不相上下，近三年则被其他主题的研究热度所超过，2018年联邦学习相关的医疗保健应用研究热度明显超出其他的应用研究热度。联邦学习在物联网方面应用研究热度于2017年开始出现，在车辆交互方面应用研究热度于2018年开始出现，在5G方面应用研究热度则是从2019年开始出现的。

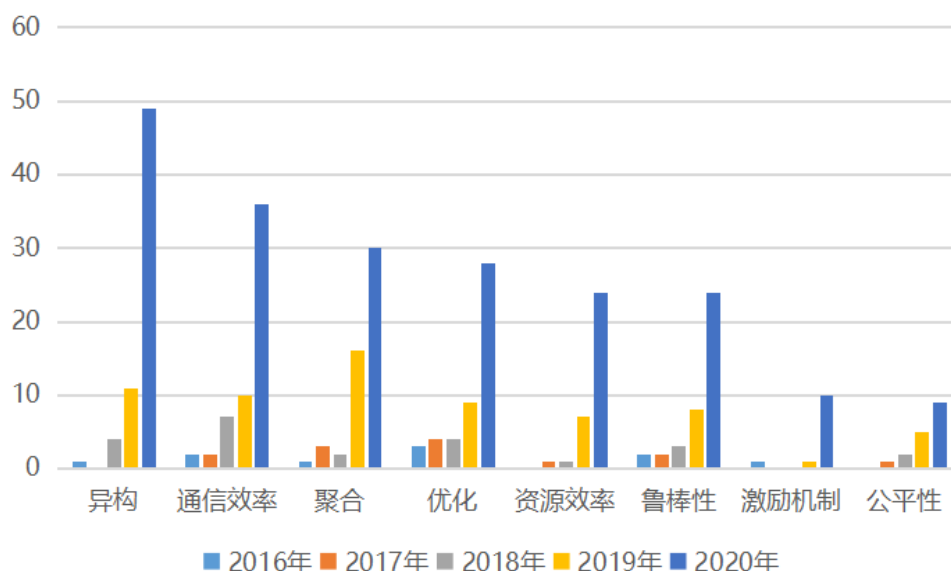
图 5 联邦学习在应用方面的研究热点趋势 (2016-2020年)



关于联邦学习在系统和模型设计方面的研究热点趋势情况如图 6 所示。由图可见，2016至2020年期间，在系统和模型设计方面研究热度领先的热点词分别是异构（heterogeneity）、通信效率（communication efficiency）、聚合（aggregation）、优化（optimization）、资源效率（resource efficiency）、鲁棒性（robustness）、激励机制（incentive mechanism）和公平性（fairness）。2016和2017年热度最高的研究主题是优化。2017年，资

源效率和公平性相关主题研究开始崭露头角；2018年通信效率相关研究占据热度榜第一；2019年热度最高的是和安全聚合相关研究。同时，对联邦学习（数据和系统）异构的研究大幅提升；2020年与异构相关研究上升为最热门，和激励机制相关的研究数量大幅提升。从热度持续性看，通信效率、聚合、优化、鲁棒性的相关研究在研究时段内一直保持着不同程度的热度上扬。

图 6 联邦学习系统和模型设计方面的研究热点趋势 (2016-2020年)



在安全隐私方面，联邦学习研究主题依据热度递减依次包括区块链（blockchain）、差分隐私（differential privacy）、安全多方计算（multiparty computation）、恶意攻击（malicious attack）、网络安全（cyber security）、隐私泄露（privacy leakage）、同态加密（homomorphic encryption）、以及容错（fault tolerance），具体热度趋势情况如图 7 所示。这五年间，差分隐私、多方计算、恶意攻击、同态加密和容错一直是研究热点并且其热度总体逐年上涨。2016 年研究最热的是对联邦学习中恶意攻击的研究，2017 年研究最热的是差分隐私，2018 年研究最热的是安全多方计算所涉及数据安全和隐私保护技术，而与区块链结合的相关研究于 2018 年出现并成为 2019 年和 2020 年最热的研究主题。

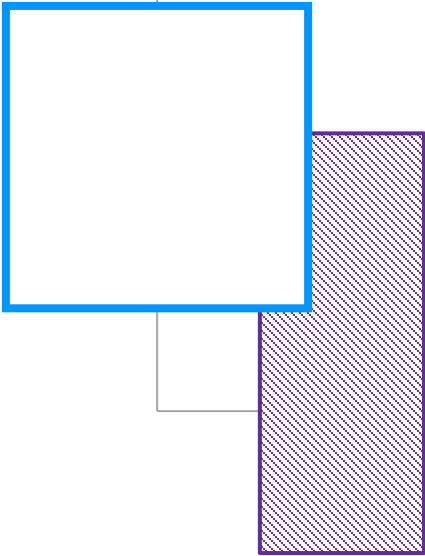
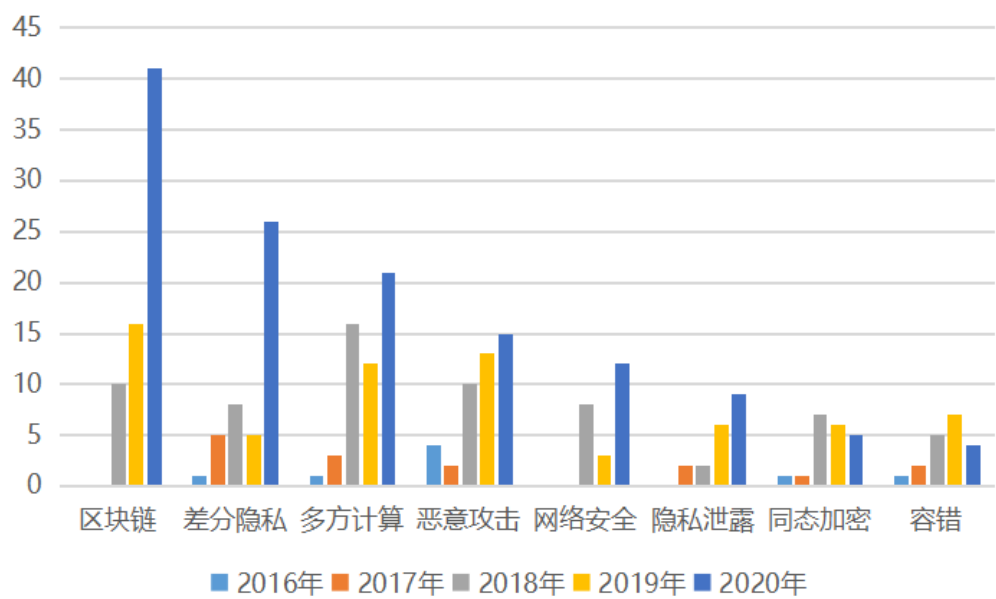


图 7 联邦学习安全隐私方面的研究热点趋势（2016-2020 年）



3.1.2 高被引论文分析

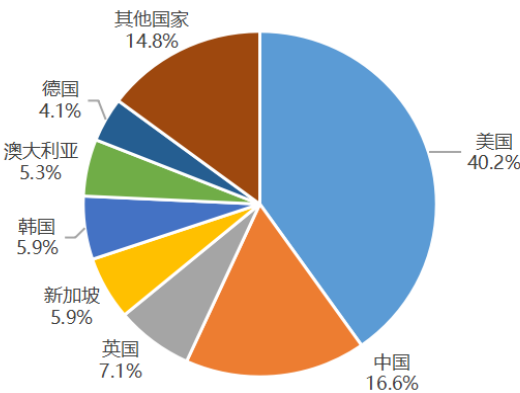
论文的被引用次数是文献计量学中测量论文的影响力或者质量的基本指标。高被引论文可以被视为具有重大学术影响的成果。根据联邦学习领域论文的发表量和引用量特点，本报告根据论文被引用量进行排序并选取了排名前3%的论文作为高被引论文进行分析，从而首先排除了那些被引总量在40次以下的论文。然后，分析论文作者所隶属机构、机构所属国家的特征。同时考虑到，在科研实践中，一篇论文通常由来自不同国家或不同机构的几名作者共同合作完成，本报告采取累计权重法统计计算国家

和机构的总被引用量，给予主要作者如第一作者的所属国家和机构100%权重，给予其他作者的所属国家和机构按一定次序给定权重，如50%、25%等。基于以上方法，统计分析得到以下的相关发现。

(1) 半数以上高被引论文来自中美两国

根据论文作者所在机构的所属国家进行统计分析，发现联邦学习的近五年来高被引论文发表主要是来自于美国和中国。其中，美国发表的论文占40.2%，全球最多；中国发表的论文占16.6%；其余国家所发表论文的占比均低于10%，详细信息如图 8 所示。

图 8 联邦学习高被论文国家分布 (2016-2020年)



(2) 美国的论文引用量全球显著领先

联邦学习相关论文总引用量TOP 10国家是美国、中国、英国、德语、新加坡、沙特阿拉伯、澳大利亚、韩国、日本和瑞士，具体信息如图 9 所示。其中，美国的论文总引用量明显高于其他国家。从领先国家来看，美国联邦学习被引用量最高的论文是谷歌公司研究科学家H. Brendan McMahan作为一作发表的论文 Communication-efficient learning of deep networks from decentralized data [14]，该论文于2016年发表于ArXiv e-prints

(2016): arXiv-1602，并在2017年收录于AISTATS (International Conference on Artificial Intelligence and Statistics)，目前其被引用2334次[15]。中国联邦学习总体论文引用量居于第二，其中被引用最高的论文是香港科技大学计算机科学与工程学系教授杨强为第一作者、与微众银行AI部门、北京航空航天大学计算机学院的研究人员联合发表的Federated Machine Learning: Concept and Applications[16]，该文被引用量986次[17]。

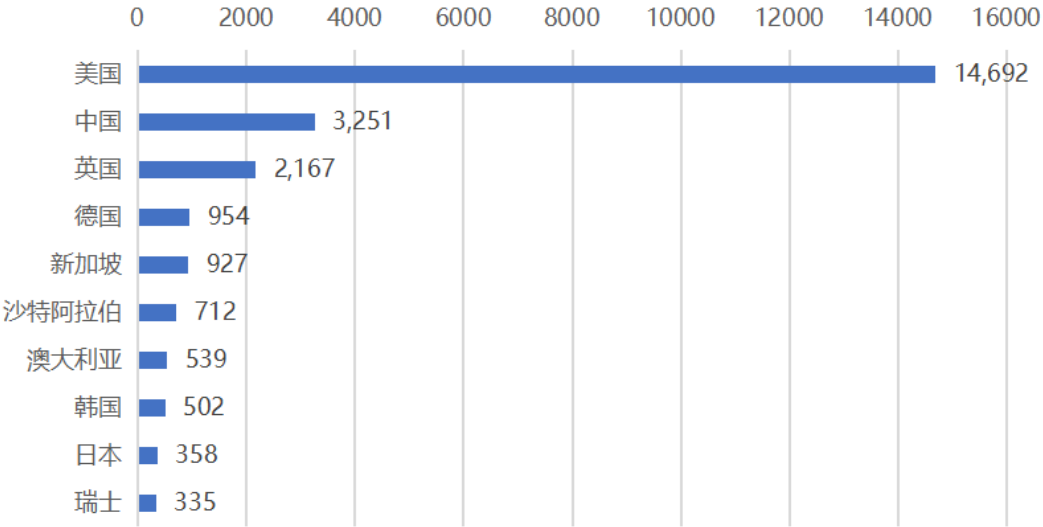
[14] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics (pp. 1273-1282). PMLR..

[15]引用量数据统计截止到2021年5月31日。

[16] Yang, Q. , Liu, Y. , Chen, T. , & Tong, Y. . (2019). Federated Machine Learning: Concept and Applications. ACM Trans. Intell. Syst. Technol. 10, 2, Article 12, February, 2019. DOI:https://doi.org/10.1145/3298981

[17]论文的被引用量数据统计截止到2021年5月31日。

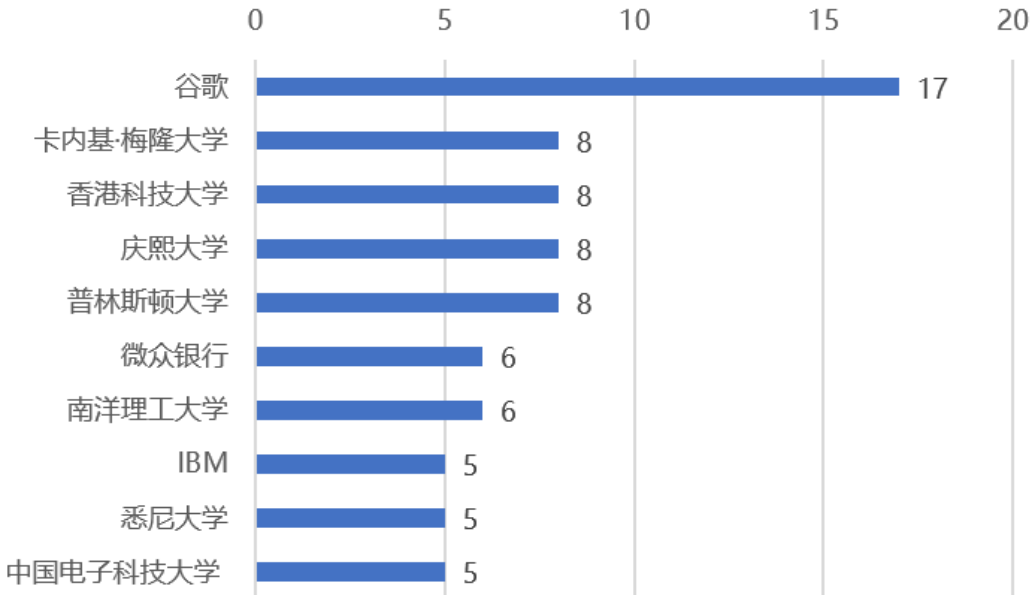
图 9 联邦学习论文引用量TOP 10 国家 (2016-2020年)



(3) 论文发布量TOP 10 机构来自美中韩澳新五国
根据论文作者所属机构进行排序分析，发现从全球范围来看，联邦学习领域近五年（2016-2020年）论文发表量TOP10机构是谷歌、卡内基·梅隆大学、香港科技大学、庆熙大学、普林斯顿大学、微众银

行、南洋理工大学、IBM、悉尼大学和中国电子科技大学。在论文量TOP10机构之中，有三家企业、七家大学；美国机构占据四席，中国机构占据三席，另外三家分别来自韩国、澳大利亚和新加坡。相关机构详细分布情况如图 10 所示。

图 10 联邦学习高被论文发表量TOP 10 机构 (2016-2020年)



(4) 联邦学习十大算法

选出了引用量大于100的前十大算法相关论文，包

通过对2016年至2020年底所发表的涉及联邦学习

括8篇横向、2篇纵向的联邦学习场景。这些算法及

算法的论文进行引用量排序（去除高引综述论文），

具体信息按照相关论文引用量排序显示如表 1 所示。

表 1 联邦学习十大算法

算法名	主要研究问题	联邦学习场景	论文	引用量
Federated Averaging (FedAvg)	Aggregation	横向联邦学习	Communication-Efficient Learning of Deep Networks from Decentralized Data	2334
Secure Aggregation	Security, Aggregation	横向联邦学习	Practical Secure Aggregation for Privacy-preserving Machine Learning	773
Federated Stochastic Variance Reduced Gradient (FedSVRG)	Communication-efficient	横向联邦学习	Federated Optimization: Distributed Machine Learning for On-device Intelligence	595
	Data heterogeneity			
MOCHA	Communication-efficient	横向联邦学习	Federated Multi-Task Learning	526
	Data heterogeneity			
FedProx	Data heterogeneity	横向联邦学习	Federated Optimization in Heterogeneous Networks	353
	System heterogeneity			
Federated Learning with Client Selection (FedCS)	System heterogeneity	横向联邦学习	Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge	308
Agnostic Federated Learning (AFL)	Data heterogeneity	横向联邦学习	Agnostic Federated Learning	188

续上表

Secure Logistic Regression	Security, Aggregation	纵向联邦学习	Private Federated Learning on Vertically Partitioned Data via Entity Resolution and Additively Homomorphic Encryption	166
SCAFFOLD	Data heterogeneity	横向联邦学习	SCAFFOLD: Stochastic Controlled Averaging for Federated Learning	150
Lossless Privacy-preserving Tree-boosting Algorithm (SecureBoost)	Security	纵向联邦学习	SecureBoost: A Lossless Federated Learning Framework	110
	Aggregation			

(5) 高引用量论文TOP10解读
通过对2016年至2020年底所发表论文的引用量进

行统计和排序，得到联邦学习领域高引论文TOP10，如表 2 所示。其中，论文的被引用量数据统计截止到2021年5月31日。本部分将对这些论文进行解读。

表 2 联邦学习领域高引论文TOP 10 (2016-2020年)

排名	论文标题	作者	发表年份	引用量
1	<i>Communication-Efficient Learning of Deep Networks from Decentralized Data</i>	McMahan, H. Brendan; Moore, Eider; Ramage, Daniel; ...	2016 ^[18]	2334
2	<i>Federated learning: Strategies for improving communication efficiency</i>	J Konečný, HB McMahan, FX Yu, P Richtárik, AT Suresh, D Bacon	2016	1334
3	<i>Federated Machine Learning: Concept and Applications</i>	Yang, Qiang; Liu, Yang; Chen, Tianjian; ...	2019	986
4	<i>Practical Secure Aggregation for Privacy-Preserving Machine Learning</i>	Bonawitz, Keith; Ivanov, Vladimir; Kreuter, Ben; ...	2017	773
5	<i>Advances and open problems in federated learning</i>	P Kairouz, HB McMahan, B Avent, A Bellet, M Bennis, AN Bhagoji, ...	2019	622
6	<i>Towards federated learning at scale: System design</i>	K Bonawitz, H Eichner, W Grieskamp, D Huba, A Ingerman, V Ivanov, ...	2019	661
7	<i>Federated optimization: Distributed machine learning for on-device intelligence</i>	J Konečný, HB McMahan, D Ramage, P Richtárik	2016	547
8	<i>Federated Learning: Challenges, Methods, and Future Directions</i>	Li, Tian; Sahu, Anit Kumar; Talwalkar, Ameet; ...	2020	542
9	<i>Federated Multi-Task Learning</i>	Virginia Smith , Chao-Kai Chiang , Maziar Sanjabi , Ameet Talwalkar	2017	526
10	<i>Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning</i>	Hitaj, Briland; Ateniese, Giuseppe; Perez-Cruz, Fernando	2017	488

^[18]该文最早发表在ArXiv e-prints (2016): arXiv-1602, 后于2017年被International Conference on Artificial Intelligence and Statistics (AISTATS)收录。

- 论文标题: Communication-Efficient Learning of Deep Networks from Decentralized Data

作者: McMahan, H. Brendan; Moore, Eider; Ramage, Daniel; Seth Hampson; Blaise Agüera y Arcas

发表期刊: ArXiv e-prints (2016): arXiv-1602; International Conference on Artificial Intelligence and Statistics (AISTATS), 2017

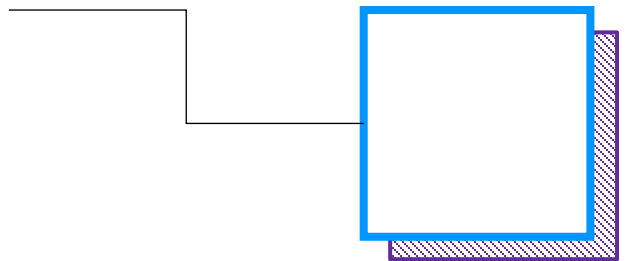
论文引用量: 2334

论文地址: <https://www.aminer.cn/pub/599c7cc1601a182cd27d4688/>

论文摘要:

现代移动设备可以访问大量适合学习模型的数据,这反过来又可以大大改善设备上的用户体验。例如,语言模型可以改进语音识别和文本输入,图像模型可以自动选择好的照片。然而,这些丰富的数据通常是隐私敏感的、数量庞大的,或者两者兼而有之,这可能会妨碍使用传统方法登录到数据中心并在那里进行训练。由此,学者们提出一种替代方案,将训练数据分布在移动设备上,并通过聚合本地计算

的更新来学习共享模型,并将这种分布式方法称为联邦学习。本文提出了一种基于迭代模型平均的深度网络联邦学习的实用方法,并进行了广泛的实证评估,考虑五种不同的模型架构和四个数据集。实验表明,该方法对不平衡和非IID数据分布具有鲁棒性,这是该设置的一个定义特征。通信成本是主要限制因素,与同步随机梯度下降相比,该方法显示所需的通信轮次减少10-100倍。



- 论文标题: *Federated Learning: Strategies for Improving Communication Efficiency*

作者: J Konečný; HB McMahan; FX Yu; P Richtárik; AT Suresh; D Bacon

发表期刊: arXiv: Machine Learning (cs.LG), 2018

论文引用量: 1334

论文地址: <https://www.aminer.cn/pub/58437725ac44360f1082f72b/>

论文摘要:

联邦学习是一种机器学习设置,其目标是训练高质量的集中式模型,同时训练数据仍然分布在具有不可靠且相对较慢的网络连接的大量客户端上。本文考虑针对此设定的学习算法,在每一轮中,每个客户端根据其本地数据独立计算当前模型的更新,并将此更新传达给中央服务器,在那里客户端更新被聚合以计算新的全局模型。此设定中的典型客户端是手机,通信效率是最重要的。本文提出了两种降

低上行链路通信成本的方法:一个是结构化更新,直接从使用较少数量变量参数化的受限空间中学习更新,例如低秩或随机掩码;另一个是草图更新,学习完整的模型更新,然后在将其发送到服务器之前使用量化、随机旋转和子采样的组合对其进行压缩。在卷积网络和循环网络上的实验表明,本文所提出的方法可以将通信成本降低两个数量级。

- 论文标题: Federated Machine Learning: Concept and Applications

作者: Yang, Qiang; Liu, Yang; Chen, Tianjian; Yongxin Tong

发表期刊: ACM Transactions on Intelligent Systems and Technology, Article No.: 12pp 1–19, 2019

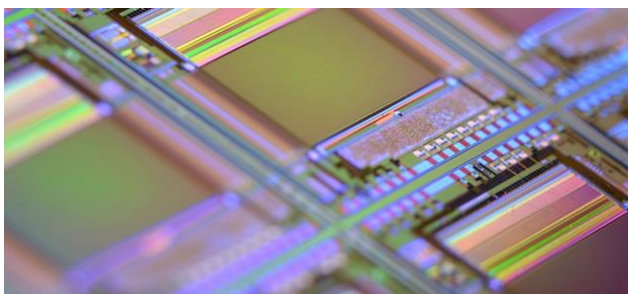
论文引用量: 986

论文地址: <https://www.aminer.cn/pub/5c9f688c6558b90bfa34dcb4>

论文摘要:

今天的人工智能仍然面临两大挑战。一是在大多数行业中, 数据以孤岛的形式存在; 另一个是加强数据隐私和安全。本文为这些挑战提出了一个可能的解决方案: 安全联邦学习。除了谷歌在2016年首次提出的联邦学习框架之外, 本文还引入了一个全面

的安全联邦学习框架, 其中包括横向联邦学习、纵向联邦学习和联邦迁移学习。本文提供了联邦学习框架的定义、体系结构和应用程序, 并提供了关于这个主题的现有工作全面调查。此外, 还提出了在组织间建立基于联邦机制的数据网络, 作为在不损害用户隐私的前提下实现知识共享的有效解决方案。



- 论文标题: *Practical Secure Aggregation for Privacy-Preserving Machine Learning*

作者: Bonawitz, Keith; Ivanov, Vladimir; Kreuter, Ben; Antonio Marcedone; H. Brendan McMahan; Sarvar Patel; Daniel Ramage; Aaron Segal; Karn Seth

发表期刊: Computer and Communications Security pp: 1175-1191, 2017

论文引用量: 773

论文地址: <https://www.aminer.cn/pub/5a260c2817c44a4ba8a23355/>

论文摘要:

本论文设计了一种新颖、通信高效、故障稳健的协议, 用于高维数据的安全聚合。该协议允许服务器以安全的方式 (即无需了解每个用户的个人贡献) 计算来自移动设备的大型用户持有数据向量的总和, 并且可以用于 (例如, 在联邦学习设定中) 聚合用户提供的深度神经网络模型更新。本文在诚实但好奇且活跃的对手设置中证明了该协议的安全性, 并

表明即使任意选择的用户子集随时退出, 也能保持安全性。本文评估了该协议的效率, 并通过复杂性分析和具体实现表明, 即使在大型数据集和客户端池上, 其运行时和通信开销仍然很低。对于 16 位输入值, 本文的协议以明文形式发送数据, 为 210 个用户和 220 维向量提供 1.73 倍的通信扩展, 并为 214 个用户和 224 维向量提供 1.98 倍扩展。

- 论文标题: Advances and Open Problems in Federated Learning

作者: Kairouz Peter; McMahan H. Brendan; Avent Brendan; Bellet Aurélien; Bennis Mehdi; Bhagoji Arjun Nitin; Bonawitz Keith; Charles Zachary; Cormode Graham; Cummings Rachel; D'Oliveira Rafael G. L.; Rouayheb Salim El

发表期刊: Foundations and Trends® in Machine Learning, no. 1 , 2019

论文引用量: 622

论文地址: <https://www.aminer.cn/pub/5df20fc53a55acbe6bfcc74f/>

论文摘要:

联邦学习是一种机器学习设定, 许多客户端 (例如移动设备或整个组织) 在中央服务器 (例如服务提供商) 的编排下协同训练一个模型, 同时保持训练数据的分散。联邦学习体现了集中数据收集和最小化的原则, 可以减轻许多由传统的、集中的机器学习

习和数据科学方法造成的系统性隐私风险和成本。

本文讨论了最近的联邦学习研究进展, 并提出了广泛的开放式问题和挑战。

- 论文标题: *Towards Federated Learning at Scale: System Design*

作者: Keith Bonawitz; Hubert Eichner; Wolfgang Grieskamp; Dzmitry Huba; Alex Ingerman; Vladimir Ivanov; Chloe Kiddon; Jakub Konečný; Stefano Mazzocchi; H. Brendan McMahan; Timon Van Overveldt; David Petrou

发表期刊: Proceedings of Machine Learning and Systems Volume: 1, pp: 374-388, 2019

论文引用量: 661

论文地址: <https://www.aminer.cn/pub/5fae6f4dd4150a363cef042f/>

论文摘要:

联邦学习是一种分布式机器学习方法, 可以在大量分散数据的语料库上进行模型训练。本文基于 TensorFlow 为移动设备领域的联邦学习构建了一个可扩展的生产系统, 描述了由此产生的高级设计,

勾勒出一些挑战及其解决方案, 并涉及未解决的问题和未来的方向。

- 论文标题: Federated Optimization: Distributed Machine Learning for On-Device Intelligence

作者: J Konečný; HB McMahan; D Ramage; P Richtárik

发表期刊: arXiv preprint arXiv:1610.02527 (2016).

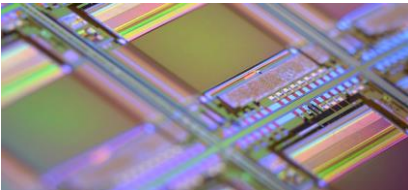
论文引用量: 547

论文地址: <https://www.aminer.cn/pub/58437725ac44360f1082ff8b/>

论文摘要:

本文为机器学习中的分布式优化引入一个新的、相关性越来越强的设置,其定义优化的数据不均匀地分布在大量节点上。其目标是培养一个高质量的称为联邦优化的集中模型。在这种情况下,通信效率是最重要的,而最小化通信轮数是主要目标。当将培训数据保存在usersu0027移动设备本地,而不是将其记录到数据中心进行培训时,就出现了一个激励的示例。在联合优化中,这些设备被用作计算节点,对本地数据执行计算,以更新全局模型。假设

在网络中有非常多的设备——与给定服务的用户数量一样多,每个用户只拥有一小部分可用数据的。特别是,本文预计本地可用的数据点数量要比设备数量少得多。此外,由于不同的用户使用不同的模式生成数据,可以合理地假设没有任何设备具有总体分布的代表性样本。本文证明了现有的算法不适合这种设定,并提出了一种新的算法,它显示了稀疏凸问题,出现了令人鼓舞的实验结果。这项工作还为联邦优化方面的未来研究奠定了基础。



- 论文标题: *Federated Learning: Challenges, Methods, and Future Directions*

作者: Li, Tian; Sahu, Anit Kumar; Talwalkar, Ameet; Smith Virginia

发表期刊: IEEE Signal Processing Magazine, no. 3, pp: 50-60, 2020

论文引用量: 542

论文地址: <https://www.aminer.cn/pub/5d5e6b9a3a55acfce79a16af/>

论文摘要:

联邦学习涉及在远程设备或孤岛数据中心(如移动电话或医院)上训练统计模型,同时保持数据本地化。在异构和潜在的大规模网络中进行训练会带来新的挑战,需要从根本上背离大规模机器学习、分

布式优化和隐私保护数据分析的标准方法。本文讨论了联邦学习的独特特征和挑战,提供了当前方法的广泛概述,并概述了与广泛的研究社区相关的几个未来工作方向。

• 论文标题: *Federated Multi-Task Learning*

作者: Virginia Smith; Chao-Kai Chiang; Maziar Sanjabi; Ameet Talwalkar

发表期刊: Advances in Neural Information Processing Systems 30 (NIPS), 2017

论文引用量: 526

论文地址: <https://www.aminer.cn/pub/599c797f601a182cd264476f/>

论文摘要:

联邦学习在通过分布式设备网络训练机器学习模型方面带来了新的统计和系统挑战。在这项实验中, 本文展示了自然适合处理这种设置统计挑战的多任务学习, 并提出了一种新颖的系统感知优化方法 MOCHA, 它对实际系统问题具有鲁棒性。本文的

方法和理论首次考虑了分布式多任务学习的高通信成本、滞后性和容错性问题。与联合设置中的替代方法相比, 所得到的方法实现了显著加速, 正如作者通过模拟真实世界联合数据集所证明的那样。

• 论文标题: *Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning*

作者: Hitaj, Briland; Ateniese, Giuseppe; Perez-Cruz, Fernando

发表期刊: CCS '17: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security Pages 603–618, 2017

论文引用量: 488

论文地址: <https://www.aminer.cn/pub/58d82fced649053542fd6955/>

论文摘要:

深度学习最近在机器学习中非常流行, 因为它能够解决端到端的学习系统, 其中特征和分类器是同时学习的, 在高度结构化和大型数据库的情况下显著提高了分类准确性。它的成功归功于最近的算法突破、日益强大的计算机以及对大量数据的访问。为此, 研究人员还考虑了深度学习对隐私的影响。模型通常以集中方式训练, 所有数据都由相同的训练算法处理。如果数据是用户隐私数据的集合, 包括习惯、个人图片、地理位置、兴趣等, 中央服务器将可以访问可能被错误处理的敏感信息。为了解决这个问题, 最近提出了协作深度学习模型, 其中各方在本地训练他们的深度学习结构, 并且只共享参

数的一个子集, 以保密他们各自的训练集。正如 Shokri和Shmatikov 在 CCS'15 上提出的那样, 还可以通过差分隐私 (DP) 来混淆参数, 从而使信息提取更具挑战性。不幸的是, 任何保护隐私的协作深度学习都容易受到在本文中设计的强大攻击。特别是, 分布式、联合或分散的深度学习方法从根本上被打破, 并且不能保护诚实参与者的训练集。本文开发的攻击利用了学习过程的实时性, 允许对手训练生成对抗网络 (GAN), 该网络生成旨在保密的目标训练集的原型样本 (由 GAN 生成的样本是旨在与来自训练数据相同的分布)。正如之前工作所建议的那样, 应用于模型共享参数的记录级差分隐私是无效的 (即记录级DP不是为了解决本文的攻击而设计的)。

(6) 中美两国论文合作数量全球最多

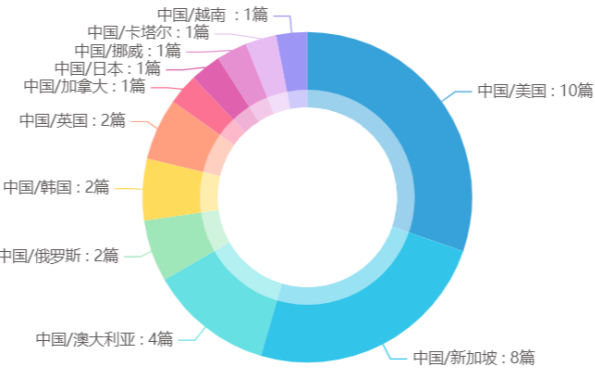
AMiner发现，约四成的高被引论文研究都发生过国际之间科研合作。如图 11 所示，中国和美国合作的论文数量最多，高达10篇；其次是美国 and 英国、中国和新加坡、美国和新加坡，两者之间各分别有9篇、8篇和5篇的合作论文；之后、澳大利亚和韩国、澳大利亚和中国之间都各有4篇合作论文。其他各国家之间虽有合作但大部分为3篇及以下，此外，还有20多个国家之间开展过1篇的论文合作。

图 11 联邦学习高被引论文的国际合作情况 (2016-2020年)

国家 国家	美国	新加坡	澳大利亚	英国	韩国	俄罗斯	越南	加拿大	卡塔尔	挪威	日本	法国	德国	瑞士
中国	10	8	4	2	2	2	1	1	1	1	1	0	0	0
美国	/	5	1	9	2	1	0	1	0	0	1	0	1	0
澳大利亚	1	2	/	0	4	2	2	0	0	0	0	0	0	0
芬兰	1	0	1	0	3	0	/	0	0	0	0	0	0	0
俄罗斯	1	2	2	0	0	/	0	0	0	0	0	0	0	0
德国	1	0	0	2	1	0	0	0	0	0	0	1	/	1
荷兰	1	0	0	1	0	0	0	0	0	0	0	0	0	0
加拿大	1	0	0	0	0	0	0	/	0	0	0	0	0	0
日本	1	0	0	0	0	0	0	0	0	0	/	0	0	0
沙特阿拉伯	1	0	0	1	0	0	0	0	0	0	0	0	1	1
印度	1	0	0	0	0	0	0	0	0	0	0	0	0	0
法国	0	0	0	1	0	0	0	0	0	0	0	/	1	0
土耳其	0	0	0	1	0	0	0	0	0	0	0	0	0	0
越南	0	1	2	0	1	0	/	0	0	0	0	0	0	0
瑞士	0	0	0	0	0	0	0	0	0	0	0	0	1	/
韩国	2	1	4	0	/	0	1	0	0	0	0	0	0	0
新加坡	5	/	0	0	1	0	0	0	1	0	0	0	0	0

在中国的高被引论文之中，有67.9%存在国际之间科研合作。其中，有两篇论文的合作国家数量各多达4个。从中国在联邦学习领域所开展的国际合作情况看，美国是中国科研论文合作最多的国家，新加坡和澳大利亚也与中国开展了较多的合作，此外，中国还与俄罗斯、韩国、英国、加拿大、日本等国进行过论文合作。

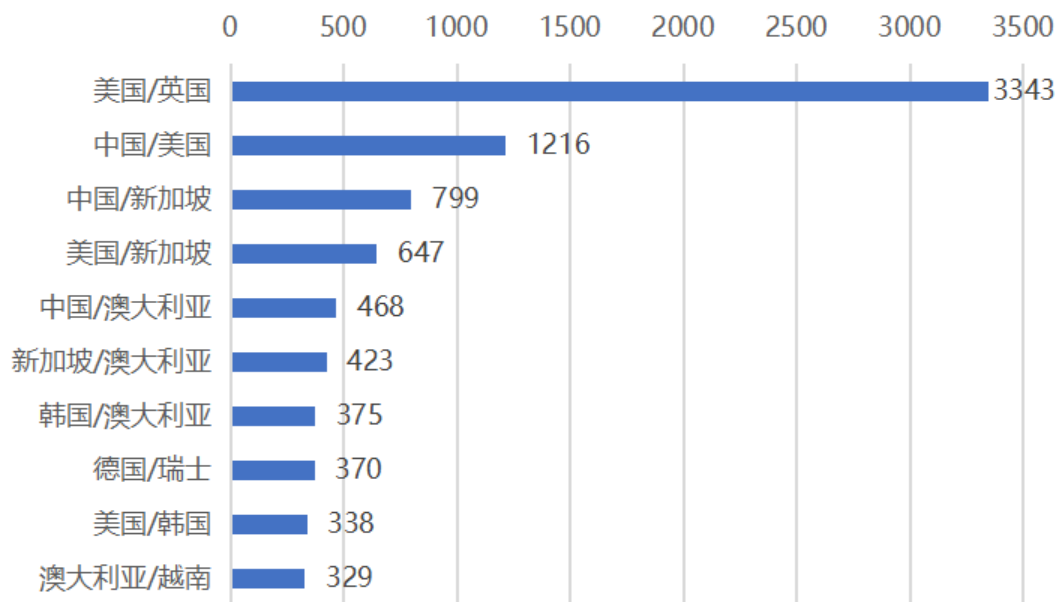
图 12 联邦学习高被引论文中外合作情况 (2016-2020年)



(7) 美英两国合作论文被引量全球领先
在各个国家之间合作发表的高被引论文之中，美国与英国、美国与中国，以及新加坡与中国的合作论文引用量居于前三，详细情况如图 13 所示。由图可见，美国和英国合作论文的引用量最高，影响力

明显高于其他国家之间合作论文的影响力。其中，美国谷歌研究人员与英国爱丁堡大学（苏格兰）学者等合作发表的论文 Federated learning: Strategies for improving communication efficiency^[19]引用量最高，达1334次^[20]。

图 13 联邦学习国际合作论文的引用量TOP10国家组合（2016-2020年）

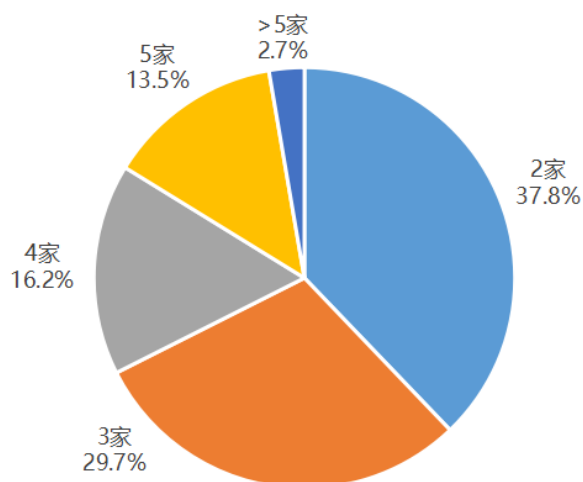


(8) 七成论文存在跨机构合作现象
国内外机构之间开展联邦学习论文合作较为常见。高被引论文中有71.8%是通过机构之间合作发表的。在机构之间合作的论文之中，一篇论文合作机构数量少则两家、多则十几家，具体分布情况如图 14 所示。由图可见，由2家机构合作完成的论文占比最多，其次是由3家机构合作的论文占比。值得一提的是，合作机构数量最多的论文是 *The future of*

digital health with federated learning ^[21]，该论文合作机构涵盖了来自德国的慕尼黑工业大学、德国癌症研究中心、海德堡大学医院，美国的宾夕法尼亚大学、范德比尔特大学、英特尔、国立卫生研究院，英国的伦敦帝国理工学院、伦敦国王学院、牛津大学、人工智能治理中心、OpenMined和法国的奥金以及英伟达在各国的公司等共计16家机构。

^[19] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.
^[20] 论文的被引用量数据统计截至到2021年5月31日。
^[21] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H., Albarqouni, S., ... Maier-Hein, K. H. (2020). The future of digital health with federated learning. Npj Digital Medicine, 3(1), 119–119.

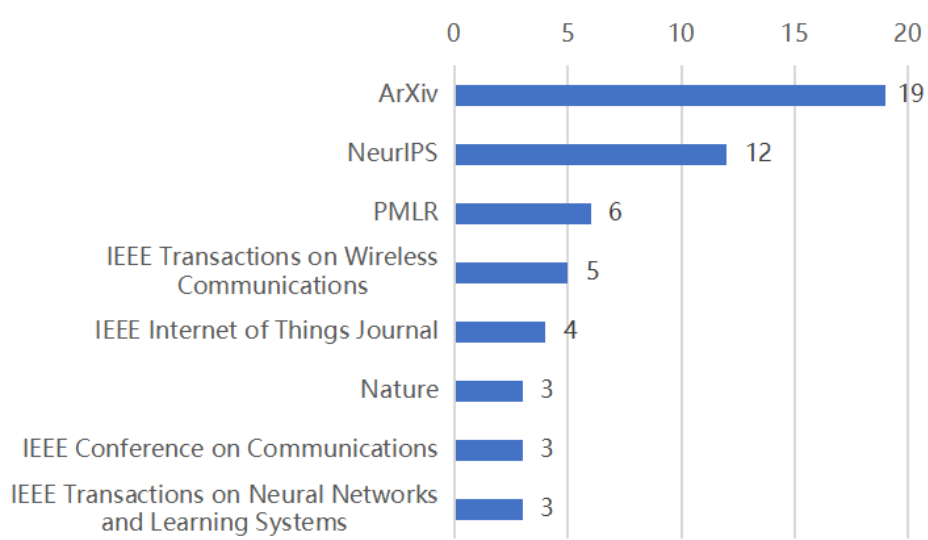
图 14 联邦学习合作论文的机构数量分布



(9) ArXiv是高被引论文的最多发布渠道
从发布渠道看，2016-2020年期间联邦学习的高被引论文发表在共计41个期刊会议等渠道上。由图15可知，高被引论文最多发布在ArXiv渠道（由美

国康奈尔大学运营维护的一个非盈利的数据库），有19篇，其次是人工智能领域国际学术会议——神经信息处理系统大会 NeurIPS（包括workshop），有12篇，以及机器学习顶级期刊PMLR有6篇。

图 15 高被引论文的较多发布来源



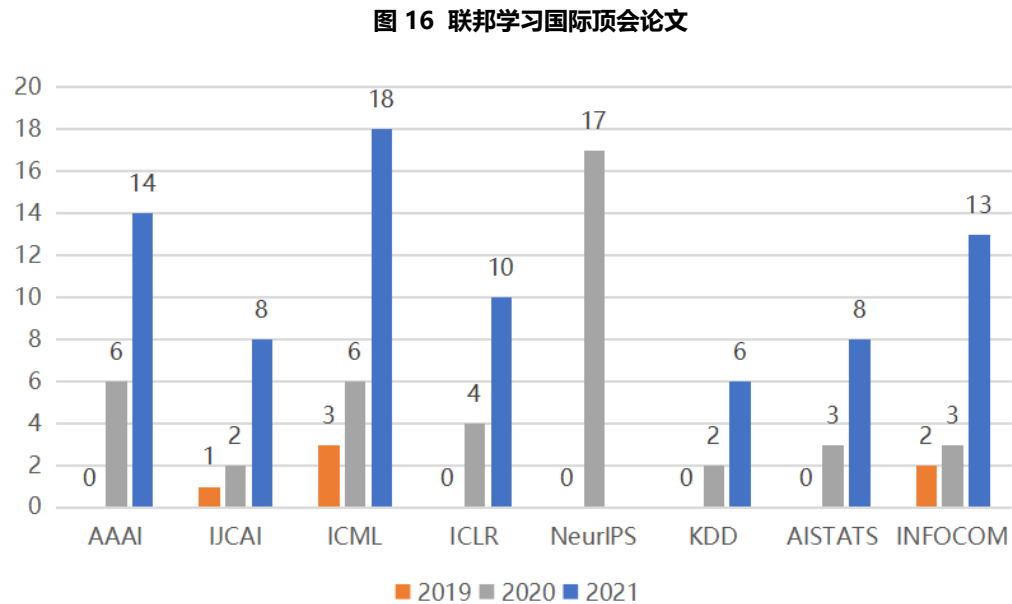
ArXiv上发表过的联邦学习最高引用论文是2016年的 *Federated Optimization: Distributed Machine Learning for On-Device Intelligence*，该论文提出一种联合优化的新算法用于设备智能的分布式机器学习，其研究目标是训练一个高质量的

中心化模型，提高通信效率。发表在NeurIPS 上的最高被引论文是 *Federated Learning: Strategies for Improving Communication Efficiency*，该论文也发表于2016年，提出了结构化更新和草图更新这两种降低上行链路通信成本的方法，目标是利用

联邦学习提高通信效率。发表在PMLR 上的最高被引论文是2018年的*How to Backdoor Federated Learning*, 该文评估标准联邦学习任务在不同假设和攻击场景下的攻击, 还展示了如何在训练攻击模型时通过将逃避纳入损失函数来逃避基于异常检测的防御。

(10) 国际顶会相关论文收录量逐年增加
人工智能国际顶会（主会）所收录的联邦学习相关

论文数量自2019年起呈现成倍增长趋势，如图 16 所示。2019年仅ICML、INFOCOM、IJCAI三个会议收录了相关论文，共计6篇。2020年，这些会议收录联邦学习的论文量达43篇，其中，联邦学习在NeurIPS 会议中论文收录量高达17篇。而2021年会议截止目前收录联邦学习的论文量已超过80篇，其中，ICML 2021收录了18篇联邦学习论文。



3.1.3 联邦学习的特刊、书籍和综述

(1) 特刊
截至2020年国内外关于联邦学习主题的特刊有一份，即美国出版的双月刊IEEE INTELLIGENT SYSTEMS [22] (2020年影响因子/JCR分区：3.210/Q2)。该

联邦学习特刊 [23] 具体文章（按照刊文顺序）情况如表 3 所示。

[22] <https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=9670>

[23] 论文的被引用量数据统计截至到2021年5月31日。

表 3 IEEE INTELLIGENT SYSTEMS联邦学习特刊 (2020年第 35 卷, 第 4 期) 文章一览

序号	论文标题及链接	作者	引用量	亮点
1	Introduction to the Special Issue on Federated Machine Learning	Yang Liu, Han Yu, and Qiang Yang	1	展示所刊文章的主要内容亮点
2	Preserving User Privacy for Machine Learning: Local Differential Privacy or Federated Machine Learning?	Huadi Zheng, Haibo Hu, and Ziyang Han	4	比较了在物联网应用中 LDP 和 FL 的可实现效率和隐私保护属性
3	Joint Intelligence Ranking by Federated Multiplicative Update	Chi Zhang, Yu Liu, Le Wang, Yuehu Liu, Li Li, and Nanning Zheng	0	提出了一种隐私保护矩阵分解方法, 该方法在自动驾驶等许多智能系统中具有潜在的适用性
4	Distributed Privacy Preserving Iterative Summation Protocols	Yang Liu, Qingchen Liu, Xiong Zhang, Shuqi Qin, and Xiaoping Lei	0	开发了一种用于隐私保护的分布式迭代协议, 该协议对节点的动态加入和离开具有弹性, 可以成为增强动态 FL 系统中隐私保护的有用技术
5	SMSS: Secure Member Selection Strategy in Federated Learning	Kun Zhao, Wei Xi, Zhi Wang, Jizhong Zhao, Ruimeng Wang, and Zhiping Jiang	0	寻求通过选择那些具有更多共同实体的数据所有者加入 FL 模型训练来解决来自不同数据所有者的不同数据质量问题
6	Federated Generative Privacy	Aleksei Triastcyn and Boi Faltings	10	关注隐私保护数据共享问题, 提出基于 GAN 的方法来生成人工数据样本以支持联合平均操作, 而无需公开敏感的本地信息
7	A Sustainable Incentive Scheme for Federated Learning	Han Yu, Zelei Liu, Yang Liu, Tianjian Chen, Mingshu Cong, Xi Weng, Dusit Niyato, and Qiang Yang	9	着眼于 FL 设置中的激励机制设计重要问题, 开发了一个公平意识的利润分享计划, 以激励数据所有者参与联邦学习

续上表

8	A Secure Federated Transfer Learning Framework	Yang Liu, Yan Kang, Chaoping Xing, Tianjian Chen, and Qiang Yang	109	提出了第一个联邦迁移学习方法，帮助 FL 应用程序处理那些样本空间和特征空间重叠的都很罕见的具有挑战性的情况
9	FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare	Yiqiang Chen, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao	80	报告了在医疗保健应用领域应用 FTL 的经验
10	Proxy Experience Replay: Federated Distillation for Distributed Reinforcement Learning	Han Cha, Jihong Park, Hyesung Kim, Mehdi Bennis, and Seong-Lyun Kim	7	提出了一种在分布式深度强化学习中提高通信效率和保护私人信息的方法

此外，目前正在组稿且待发表的联邦学习主题特刊 它们的具体征文信息（call for papers）如表 4 所示。有八份，预计将于2021年~2022年陆续正式发表。

表 4 待发表的联邦学习特刊一览

序号	特刊名称及链接	期刊	截稿日期
1	Special Issue on Robust Federated Learning over Future Wireless Networks	Internet of Things and Cyber-Physical Systems	8 月 30 日 2021
2	Special Issue "Federated Learning: Challenges, Applications and Future"	Electronics (ISSN 2079-9292)	12 月 31 日 2021
3	Special Issue: AI-Based Federated Learning for 6G Mobile Networks	Wireless Communications and Mobile Computing	8 月 27 日 2021
4	Special Issue: FMLDH-CMC 2021 : Federated Machine Learning on Digital Health	CMC -Computers, Materials & Continua	9 月 15 日 2021
5	Special Issue on Federated Learning for Decentralized Cybersecurity	Computers & Security	3 月 1 日 2021
6	Special Issue on Federated Learning and Blockchain Supported Smart Networking in Beyond 5G (B5G) Wireless Communication	Computer Networks	8 月 15 日 2021
7	Special Issue on Federated Learning: Algorithms, Systems, and Applications	ACM Transactions on Intelligent Systems and Technology	8 月 30 日 2021
8	Special Issue on Trustable, Verifiable, and Auditable Federated Learning	IEEE Transactions on Big Data, Special Issue	1 月 15 日 2022 30

(2) 书籍

在2020年出版，其余三本于2021年出版。相关介

联邦学习主要书籍截至目前发现有四本，其中一本 绍如下。

书名	联邦学习=Federated Learning
作者	杨强，刘洋，程勇，康焱，陈天健，于涵
出版社	电子工业出版社
出版时间	2020-04-01 第 1 版
正文语种	中文
ISBN	9787121385223
该书是首部全面和系统论述联邦学习的中文著作。该书阐述了联邦学习的定义、分类和发展历程，并且介绍了与联邦学习紧密相关的基础知识，比如分布式机器学习和隐私保护技术。该书对联邦学习的每一分类，即横向联邦学习、纵向联邦学习和联邦迁移学习，所涉及的架构和算法进行了详尽的介绍。同时，该书也讨论了联邦强化学习，联邦学习的激励机制和应用实例。该书适合作为读者入门和探究联邦学习的第一本书。	

书名	联邦学习技术及实战
作者	彭南博，王虎 等
出版社	电子工业出版社
出版时间	2021-03-01 第 1 版
正文语种	中文
ISBN	9787121405976
该书由京东科技集团有着多年联邦学习实战经验的工程人员合作编写，内容包括联邦学习基础、具体的联邦学习算法、联邦学习的产业应用和展望三个大部分，并给出较多案例。该书针对产业界在智能化过程中普遍面临的数据不足问题，详细地阐述了联邦学习如何帮助企业引入更多数据、提升机器学习模型效果。该书广泛介绍了联邦学习技术的实战经验，主要内容包括隐私保护、机器学习等基础知识，联邦求交、联邦特征工程算法，以及工程架构、产业案例、数据资产定价等。	

书名	联邦学习实战
作者	杨强, 黄安埠, 刘洋, 陈天健
出版社	电子工业出版社
出版时间	2021-05-01 第 1 版
正文语种	中文
ISBN	9787121407925
该书是微众银行联邦学习团队在该领域的第二本专著。相较于第一本以理论和概述为主, 该书以实战为主, 兼顾对理论知识的系统总结。该书在联邦学习的理论知识基础上, 主要介绍如何使用 Python 和 FATE 进行联邦学习建模, 包括大量联邦学习的案例分析, 筛选了经典案例进行讲解, 部分案例用 Python 代码实现, 部分案例采用 FATE 实现。此外, 介绍了联邦学习相关的高级知识点, 包括联邦学习的架构和训练的加速方法等。该书适合对联邦学习和隐私保护感兴趣的高校研究者和企业研发人员阅读。	

书名	深入浅出联邦学习：原理与实践
作者	王健宗, 李泽远, 何安珣
出版社	机械工业出版社
出版时间	2021-05-01
正文语种	中文
ISBN	9787111679592
该书从理论与实践的双重维度对联邦学习进行了阐述, 提供了可动手实践的源码案例, 也分享了作者对联邦学习发展趋势的洞察和思考。全书分为四个部分。第一部分主要介绍了联邦学习的概念、由来、发展历史、架构思想、应用场景、优势、规范与标准、社区与生态等基础内容。第二部分详细讲解了联邦学习的工作原理、算法、加密机制、激励机制等核心技术。第三部分主要讲解了 PySyft、TFF、CrypTen 等主流联邦学习开源框架的部署实践, 并给出了联邦学习在智慧金融、智慧医疗、智慧城市、物联网等领域的具体解决方案。第四部分概述了联邦学习的形态、联邦学习系统架构、当前面临的挑战等, 并探讨了联邦学习的发展前景和趋势。	

(3) 综述

联邦学习自2016年提出以来，就吸引了学界和工业界的广泛兴趣。在联邦学习的各个领域如基础理论、系统设计方法、实施应用，面临的挑战和范式创新等都涌现了大量研究，相应地也产生了许多综述文章。这里我们基于综述的引用量和关注范围的多样性，选取了9篇综述进行介绍。详细信息如表 5 所示。

表 5 联邦学习综述性文章一览

序号	文章 Paper	范围 Scoping
1	<i>Federated Machine Learning: Concept and Application</i> ^[24]	General overview
	是联邦学习领域最早的综述，介绍了联邦学习的概念，分类，系统架构和涉及的主要技术方法。基于数据分布特点，该综述将联邦学习分为横向联邦学习，纵向联邦学习和联邦迁移学习，并列举了相关应用场景。此外，通过总结相关领域的论文，讨论了联邦学习与其它学习范式，如分布式学习，边缘计算和联邦数据库系统的关联和区别。	
2	<i>Advances and Open Problems in Federated Learning</i> ^[25]	General overview
	对联邦学习的理论和应用进行了系统和全面的介绍，涵盖了联邦学习的各个方面，包括定义，分类，效率和效能，数据隐私保护，攻击及故障的鲁棒性，参与方的公平性等，并重点探讨了联邦学习待解决的问题和面临的挑战，给研究员总结了联邦学习的研究方向。	
3	<i>Federated Learning: Challenges, Methods, and Future Directions</i> ^[26]	General overview
	主要讨论了联邦学习的特点及其相较于传统分布式计算面临的挑战，包括节点间的通信效率，系统的异构性，数据的不均匀性和隐私保护能力。通过深入分析这些问题提出了解决思路和未来研究方向。	

^[24] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ArXiv190204885 Cs, Feb. 2019, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1902.04885>

^[25] P. Kairouz et al., "Advances and Open Problems in Federated Learning," ArXiv191204977 Cs Stat, Dec. 2019, Accessed: Aug. 10, 2020. [Online]. Available: <http://arxiv.org/abs/1912.04977>

^[26] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, May 2020, doi: 10.1109/MSP.2020.2975749.

续上表

4	<i>A Survey on Federated Learning System: Vision, Hype and Reality for Data Privacy and Protection</i> ^[27]	System review
	作者主要从系统的角度对于联邦学习进行了归纳，分析和总结。首先，介绍了联邦学习系统的定义和系统组件。基于数据分布、机器学习模型、隐私保护技术、通信架构，系统规模和联邦的动机六个维度对现有联邦学习系统和方法进行了分类和研究总结，此外还探讨了联邦学习系统的设计方法，典型案例和未来的研究方向。	
5	<i>Federated Learning in Mobile Edge Networks: A Comprehensive Survey</i> ^[28]	mobile edge networks
	聚焦将联邦学习应用于移动端边缘计算。首先介绍了边缘计算的动机和如何与联邦学习结合进行联合模型训练。然后重点分析了基于联邦学习的边缘计算在通信成本、计算资源分配、数据隐私和数据安全方面所面临的挑战及未来研究方向。此外，介绍了联邦学习与边缘计算结合的一些应用和实现。	
6	<i>Threats to Federated Learning: A survey</i> ^[29]	Security and privacy
	从联邦学习系统的威胁模型及可能受到的攻击方式的角度进行了总结，主要聚焦会影响模型期望行为的“投毒”和“推断”攻击。	
7	<i>A Survey on Security and Privacy of Federated Learning</i> ^[30]	Security and privacy
	在为研究员在联邦学习安全和隐私保护领域提供一个清晰的研究方向。该综述对联邦学习中所涉及的安全威胁和隐私隐患进行的全面的阐述，并且给出了可能降低这些安全威胁和隐私隐患的基本方法和可能带来的成本。	

^[27] Q. Li et al., “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection,” ArXiv190709693 Cs Stat, Jan. 2021, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1907.09693>

^[28] W. Y. B. Lim et al., “Federated Learning in Mobile Edge Networks: A Comprehensive Survey,” IEEE Commun. Surv. Tutor., vol. 22, no. 3, pp. 2031–2063, thirdquarter 2020, doi: 10.1109/COMST.2020.2986024.

^[29] L. Lyu, H. Yu, and Q. Yang, “Threats to Federated Learning: A Survey,” ArXiv200302133 Cs Stat, Mar. 2020, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/2003.02133>

^[30] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, “A survey on security and privacy of federated learning,” Future Gener. Comput. Syst., vol. 115, pp. 619–640, Feb. 2021, doi: 10.1016/j.future.2020.10.007.

续上表

4	<i>A Survey on Federated Learning System: Vision, Hype and Reality for Data Privacy and Protection</i> ^[27]	System review
	作者主要从系统的角度对于联邦学习进行了归纳，分析和总结。首先，介绍了联邦学习系统的定义和系统组件。基于数据分布、机器学习模型、隐私保护技术、通信架构，系统规模和联邦的动机六个维度对现有联邦学习系统和方法进行了分类和研究总结，此外还探讨了联邦学习系统的设计方法，典型案例和未来的研究方向。	
5	<i>Federated Learning in Mobile Edge Networks: A Comprehensive Survey</i> ^[28]	mobile edge networks
	聚焦将联邦学习应用于移动端边缘计算。首先介绍了边缘计算的动机和如何与联邦学习结合进行联合模型训练。然后重点分析了基于联邦学习的边缘计算在通信成本、计算资源分配、数据隐私和数据安全方面所面临的挑战及未来研究方向。此外，介绍了联邦学习与边缘计算结合的一些应用和实现。	
6	<i>Threats to Federated Learning: A survey</i> ^[29]	Security and privacy
	从联邦学习系统的威胁模型及可能受到的攻击方式的角度进行了总结，主要聚焦会影响模型期望行为的“投毒”和“推断”攻击。	
7	<i>A Survey on Security and Privacy of Federated Learning</i> ^[30]	Security and privacy
	在为研究员在联邦学习安全和隐私保护领域提供一个清晰的研究方向。该综述对联邦学习中所涉及的安全威胁和隐私隐患进行的全面的阐述，并且给出了可能降低这些安全威胁和隐私隐患的基本方法和可能带来的成本。	

^[27] Q. Li et al., “A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection,” ArXiv190709693 Cs Stat, Jan. 2021, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1907.09693>

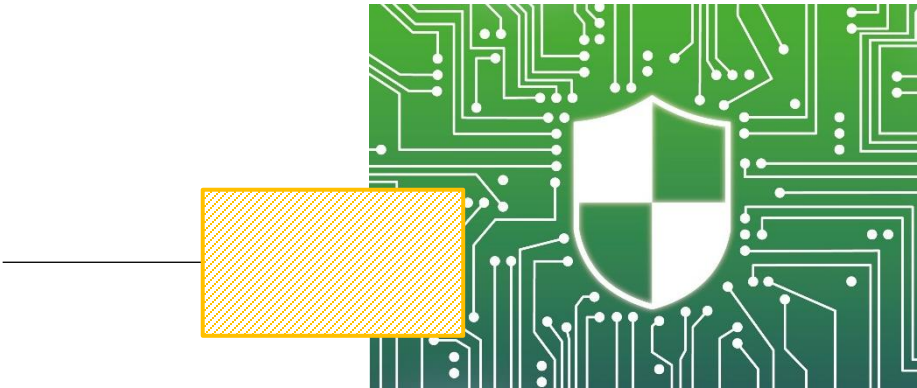
^[28] W. Y. B. Lim et al., “Federated Learning in Mobile Edge Networks: A Comprehensive Survey,” IEEE Commun. Surv. Tutor., vol. 22, no. 3, pp. 2031–2063, thirdquarter 2020, doi: 10.1109/COMST.2020.2986024.

^[29] L. Lyu, H. Yu, and Q. Yang, “Threats to Federated Learning: A Survey,” ArXiv200302133 Cs Stat, Mar. 2020, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/2003.02133>

^[30] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, “A survey on security and privacy of federated learning,” Future Gener. Comput. Syst., vol. 115, pp. 619–640, Feb. 2021, doi: 10.1016/j.future.2020.10.007.

续上表

8	<i>A Systematic Literature Review on Federated Machine Learning – From a Software Engineering Perspective</i> ^[31]	Software engineering perspective
	从软件工程的角度对联邦学习的研究进行了系统的分析和总结。该综述详细阐述了软件开发生命周期中的需求分析，背景理解，架构设计，系统实现和性能评估等各个环节所对应的联邦学习研究问题。	
9	<i>Federated Learning for Healthcare Informatics</i> ^[32]	Healthcare
	分析了联邦学习技术应用于医疗领域所面临的困难与挑战，并总结了现有的解决方案。同时分享了联邦学习在医疗领域的应用场景。	



^[31] S. K. Lo, Q. Lu, C. Wang, H.-Y. Paik, and L. Zhu, “A Systematic Literature Review on Federated Machine Learning: From A Software Engineering Perspective,” *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–39, Jun. 2021, doi: 10.1145/3450288.

^[32] Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, “Federated Learning for Healthcare Informatics,” *ArXiv191106270 Cs*, Aug. 2020, Accessed: Jun. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1911.06270>

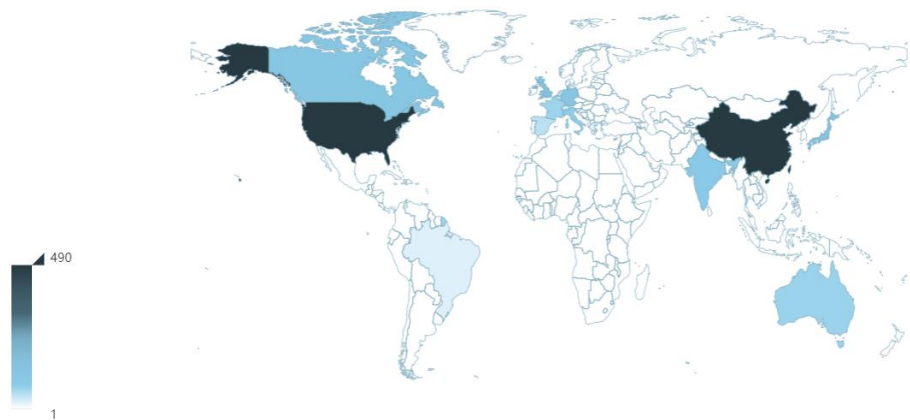
3.1.4 学者人才地图与画像

(1) 全球学者主要聚集在美国和中国

基于AMiner 系统，通过关键词组^[33]在标题和摘要中检索2016年至2020年联邦学习相关论文数据，然后对这些论文数据进行挖掘分析，获取论文作者信息，通过命名消歧和信息抽取等大数据分析和挖掘技术，进行作者画像和人才相关分析。此外，还抽取论文作者的供职机构和国家信息，对不同国家和机构的研究者和论文数量进行统计和分析。

结果显示，在研究时段内，联邦学习领域的全球学者共计2764名，分布在亚洲、北美洲、欧洲以及大洋洲的18个国家之中，所在国家分布如图 17 所示，从分布密度来看，这些学者主要聚集在东亚的中国（816名）、日本（146名），北美洲的美国（817名）和欧洲的英国（151名）、德国（121名）等国家。

图 17 联邦学习全球学者位置分布（2016-2020年）

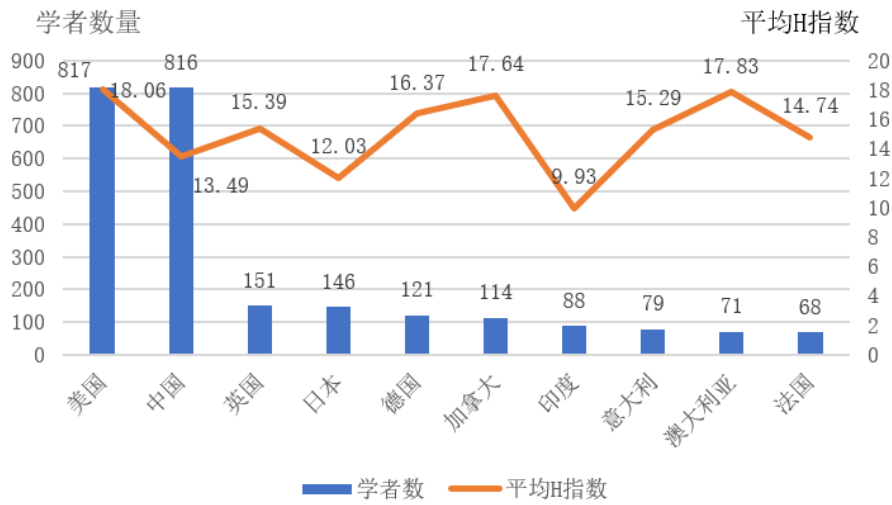


(2) 学者量TOP 10国家之间无明显学术差距
联邦学习学者主要聚集的国家是美国和中国，其拥有学者数量分别为817名和816名，明显多于其他国家的学者数量，如图 18 所示。从学术水平来看，美国和加拿大、澳大利亚等国家的学者水平相对较

高，平均H指数高于17；总体来看，除了印度，学者量前十的其他国家学者的平均H指数差距并不显著。

^[33] 联邦学习关键词检索式: Federated Machine Learning OR Federated optimization OR federated learning OR federation learning OR (Privacy AND Distributed AND data mining) OR (Secure AND Distributed AND data mining) OR (Secure AND Multiparty) OR (Secure AND Multi-party) OR (privacy AND Multi-party) OR (privacy AND Multiparty) OR (Privacy AND Distributed AND machine learning) OR (Secure AND Distributed AND machine learning) OR (Privacy and joint learning) OR (Secure and joint learning) OR (Privacy AND Distributed AND deep learning) OR (Secure AND Distributed AND deep learning)

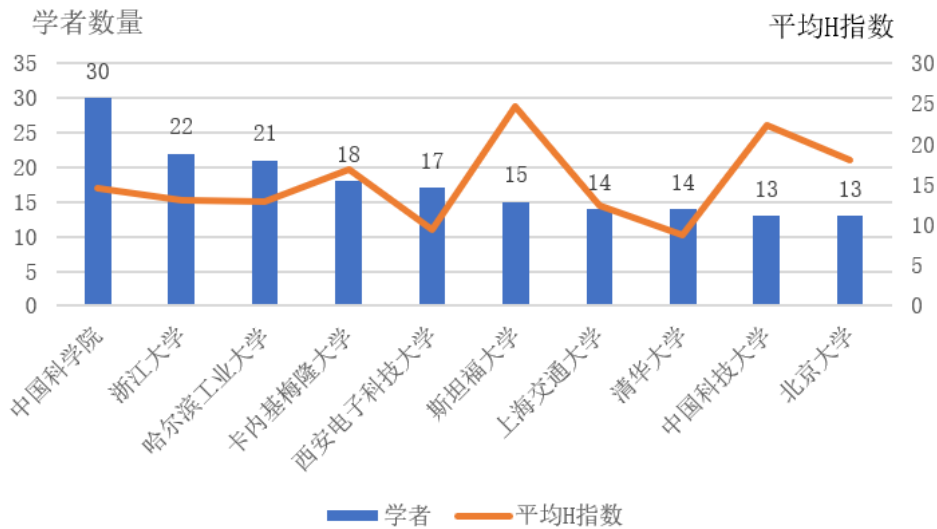
图 18 联邦学习领域学者数量TOP 10 国家 (2016-2020年)



(3) 学者量TOP 10机构较多为中国占据
基于对研究时段内相关论文作者所供职机构信息的抽取分析，发现从全球范围来看，联邦学习领域学者总量TOP 10机构之中，八成席位被中国机构占据，如图 19 所示。总体来看，各家机构的联邦学习领

域研究学者数量在10~30位。其中，中国科学院拥有该领域学者数量最多，为30名。从学者学术水平来看，斯坦福大学学者平均H指数最高，中国机构之中中国科技大学学者的平均H指数最高。

图 19 联邦学习领域学者数量TOP 10 机构 (2016-2020年)

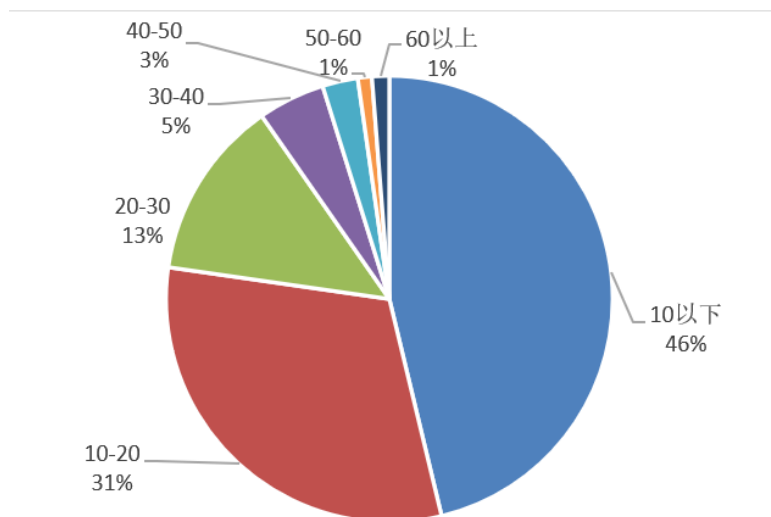


(4) 高学术水平学者占比不足5%

H指数是衡量学者学术水平的重要指标。研究联邦学习的学者H指数在10以下的为1819名，占比46%；H指数在10-20的学者量，占比31%；H指数在50

以上和在60以上的学者数量较少，占比各自均为1%。具体的领域学者H指数分布情况如图 20 所示。

图 20 联邦学习领域学者学术水平H指数分布



上述这种现象反映出，尽管联邦学习是一个有活力、有前途的热门发展领域，但是目前联邦学习领域的高端研究人才比较稀缺。究其原因，一方面可能是由于联邦学习是一个起源于工业界且已落地于医疗、金融等应用场景的新技术，而工业界研究者相对学术界研究者较少发布论文，所以基于论文发表量和被引用量的H指数未能充分反映出这些工业界研究者的实际水平；另一方面则可能是由于目前学术界的联邦学习高水平研究者通常是从在其他研究领域的顶尖研究者迁移过来的，例如，曾在机器学习领域发表多篇高水平论文且特别关注联邦学习的香港科技大学研究者杨强教授，或者是隐私计算领域曾发表多篇高水平论文并且特别关注联邦学习的加州大学伯克利分校研究者宋晓东教授。不过，联邦学习是一个未来可期的研究方向，该领域的高学术水平学者数量将不断增长。

(5) 不同研究方向的代表学者画像

在AMiner 学术搜索服务平台上，根据相关算法，通过对AAAI、CCS、ICLR、ICML、IJCAI、NIPS、SP等联邦学习领域顶尖学术会议近年来收录论文的挖掘，并结合热心网友的推荐和整理，筛选出了“联邦学习”主题领域 100 篇经典必读论文（简称 Topic 必读论文）。可以帮助用户快速了解该领域知识，从而提高学习效率。用户只需在检索框输入“Federated Learning”或中文“联邦学习”，就能看到联邦学习 TOPIC 页面（<https://www.aminer.cn/topic/600e890992c7f9be21d74695>），该页面中包括相关的简要概述和精选必读论文。每一篇必读论文由程序自动计算出了一句话内容概括作为“推荐理由”；必读论文列表支持按照发表年份、引用数、点赞数等进行排序，并在页面右侧，列出了相关作者的论文发表情况。

本部分简要介绍了联邦学习领域的代表性学者及其代表论文。其中，代表学者是指该学者作为某篇论文前两位作者或者通讯作者出现；学者的代表作论文则是指该学者在 2016-2020 年发表的引用量大于30^[34]的论文。这里，代表性学者的排名不分先后。限于报告篇幅，我们不能对所有学者逐一罗列，如要获得更多学者信息，请查看网址<https://www.aminer.cn/>获取更多领域学者资料。

① 联邦学习方向

Qiang Yang (杨强)

香港科技大学 教授; 微众银行 首席人工智能官

最高学位毕业院校: 美国马里兰大学 博士

曾经任职: 香港科技大学计算机与工程系主任、第四范式有限公司联合创始人、华为诺亚方舟研究实验室创始主任、加拿大BC省西蒙弗雷泽大学副教授/正教授、加拿大滑铁卢大学计算机科学助理/副教授等。

研究兴趣: 人工智能、迁移学习、联邦学习、机器学习、数据挖掘



杨强 (Qiang Yang)

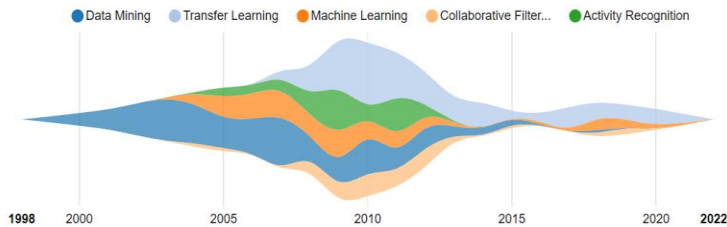
h-index: 107 | 论文数: 733 | 引用数: 58458

教授

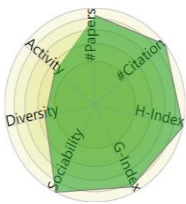
香港科技大学计算机科学与工程学系

Data Mining Transfer Learning Machine Learning Collaborative Filtering Activity Recognition Case Base Reasoning

研究兴趣



作者统计



#Papers: 733
#Citation: 58458
H-Index: 107
G-Index: 231
Sociability: 7
Diversity: 0
Activity: 7

³⁴ 论文引用量数据统计截至到2021年5月31日。

相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	<i>FedVision: An Online Visual Object Detection Platform Powered by Federated Learning</i>	https://www.aminer.cn/pub/5e257a973a55acdfeeb9ed85	AAAI, no. 08 (2020): 13172-13179
2	<i>A Fairness-aware Incentive Scheme for Federated Learning</i>	https://www.aminer.cn/pub/5e3e887a3a55ac6b075ba5a4	AIES, pp.393-399, (2020)
3	<i>A Communication Efficient Collaborative Learning Framework for Distributed Features</i>	https://www.aminer.cn/pub/5f8d43449e795ea21af78f0a	arXiv preprint arXiv:1912.11187 (2019) / FL-NeurIPS 2019
4	<i>A Secure Federated Transfer Learning Framework</i>	https://www.aminer.cn/pub/5ecbc8ab9fcd0a24b522f2e/	IEEE Intelligent Systems, 35(4), 70-82.
5	<i>SecureBoost: A Lossless Federated Learning Framework</i>	https://www.aminer.cn/pub/5ced0ffda562983788df3f4/	IEEE Intelligent Systems (2021)
6	<i>Secure Federated Matrix Factorization</i>	https://www.aminer.cn/pub/5d06e46cda562926acc32de9/	IEEE Intelligent Systems (2020)
7	<i>Federated Machine Learning: Concept and Applications</i>	https://www.aminer.cn/pub/5c9f688c6558b90bfa34dcb4/	ACM Transactions on Intelligent Systems and Technology (TIST) 10.2 (2019): 1-19.
8	<i>Privacy-preserving Heterogeneous Federated Transfer Learning</i>	https://www.aminer.cn/pub/5dea0ee0930831239d97ec64/	2019 IEEE International Conference on Big Data (Big Data)

H. Brendan McMahan

谷歌公司 研究科学家

最高学位毕业院校：美国卡耐基梅隆大学 计算机科学博士

研究兴趣：机器学习、联邦学习、分布式优化、差异隐私、深度学习



H. Brendan McMahan

h-index: 35 | 论文数: 74 | 引用数: 8523

Research Scientist

Google

Markov Decision Process

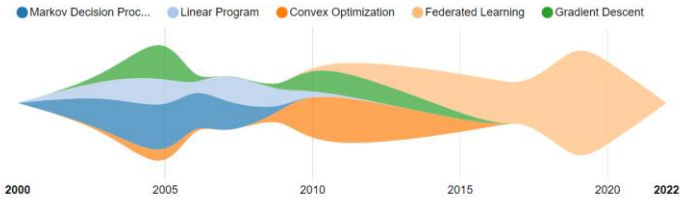
Linear Program

Convex Optimization

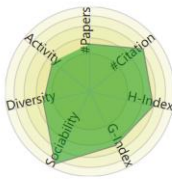
Federated Learning

Gradient Descent

研究兴趣



作者统计



#Papers: 74
#Citation: 8523
H-Index: 35
G-Index: 74
Sociability: 6
Diversity: 0
Activity: 1

相关论文代表作：

序号	论文名称	论文地址	发表期刊/年份
1	<i>Advances and Open Problems in Federated Learning</i>	https://www.aminer.cn/pub/5cde109da562983788e9865	Foundations and Trends® in Machine Learning, no. 1 (2019)
2	<i>Generative Models for Effective ML on Private, Decentralized Datasets</i>	https://www.aminer.cn/pub/5e5e18ca93d709897ce3198a/	ICLR (2020)
3	<i>Communication-efficient learning of deep networks from decentralized data</i>	https://www.aminer.cn/pub/599c7cc1601a182cd27d4688/	AISTATS, pp.1273-1282, (2017)
4	<i>Federated Optimization: Distributed Optimization for On-Device Intelligence</i>	https://www.aminer.cn/pub/58437725ac44360f1082ff8b/	arXiv preprint arXiv:1610.02527 (2016)

5	<i>Federated Learning: Strategies for Improving Communication Efficiency</i>	https://www.aminer.cn/pub/58437725ac44360f1082f72b/	arXiv preprint arXiv:1610.05492 (2016)
6	<i>Can You Really Backdoor Federated Learning?</i>	https://www.aminer.cn/pub/5dd50ed43a55ac51376177ec/	arXiv preprint arXiv:1911.07963 (2019)

Jakub Konečný

谷歌公司 研究科学家

最高学位毕业院校：英国爱丁堡大学 博士

研究兴趣：联邦学习



Jakub Konečný
Research Scientist, [Google](#)

相关论文代表作：

序号	论文名称	论文地址	发表期刊/年份
1	<i>Federated learning: Strategies for improving communication efficiency</i>	https://www.aminer.cn/pub/58437725ac44360f1082f72b/	arXiv:1610.05492 (2016).
2	<i>Federated optimization: Distributed machine learning for on-device intelligence</i>	https://www.aminer.cn/pub/58437725ac44360f1082ff8b/	arXiv:1610.02527 (2016).
3	<i>Expanding the Reach of Federated Learning by Reducing Client Resource Requirements</i>	https://www.aminer.cn/pub/5c2c7a9217c44a4e7cf3168e/	arXiv:1812.07210 (2018).
4	<i>Improving federated learning personalization via model agnostic meta learning</i>	https://www.aminer.cn/pub/5d91d22d3a55acb3c9c57b35/	arXiv:1909.12488 (2019).
5	<i>AIDE: Fast and communication efficient distributed optimization</i>	https://www.aminer.cn/pub/58437725ac44360f1082ffb2/	arXiv:1608.06879 (2016).

Peter Kairouz

谷歌研究院 研究员

最高学位毕业院校：美国伊利诺伊大学厄巴纳-香槟分校 博士

曾经任职：斯坦福大学 博士后

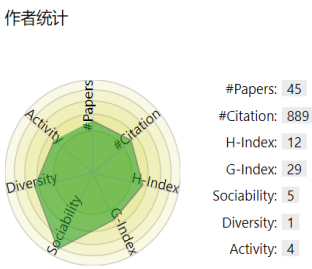
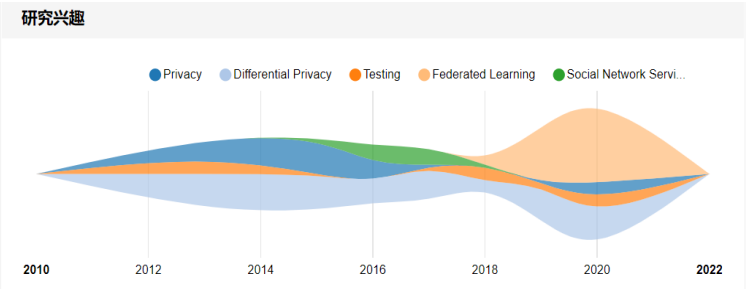
研究兴趣：差分隐私、联邦学习、人工智能、机器学习、信息理论



Peter Kairouz

h-index: 12 | 论文数: 45 | 引用数: 889

Google



相关论文代表作：

序号	论文名称	论文地址	发表期刊/年份
1	<i>Advances and Open Problems in Federated Learning</i>	https://www.aminer.cn/pub/5df20fc53a55acbe6bfcc74f/	Foundations and Trends® in Machine Learning, no. 1 (2019)
2	<i>Can You Really Backdoor Federated Learning?</i>	https://www.aminer.cn/pub/5dd50ed43a55ac51376177ec/	arXiv:1911.07963 (2019)
3	<i>DP-cgan: Differentially Private Synthetic Data and Label Generation</i>	https://www.aminer.cn/pub/5da6ed2c3a55ac45df741fcf/	Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019
4	<i>Context-aware Generative Adversarial Privacy</i>	https://www.aminer.cn/pub/5a260c8b17c44a4ba8a32913/	Entropy 19.12 (2017): 656.

Ananda Theertha Suresh

谷歌公司 高级研究科学家
最高学位毕业院校: 美国加州大学圣地亚哥分校 博士
研究兴趣: 联邦学习、统计分析、信息理论



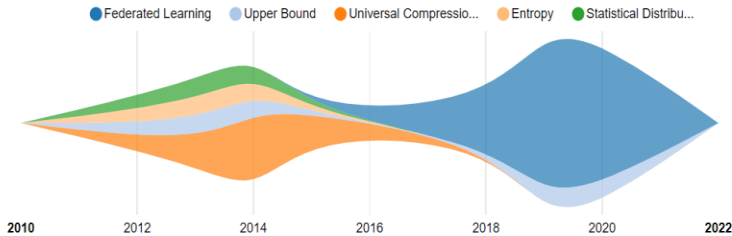
Ananda Theertha Suresh

h-index: 22 | 论文数: 72 | 引用数: 2570

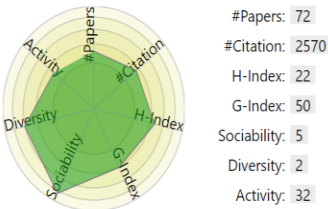
Google

Federated Learning Upper Bound Universal Compression Entropy Statistical Distributions Discrete Distributions

研究兴趣



作者统计



相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	Three Approaches for Personalization with Applications to Federated Learning	https://www.aminer.cn/pub/5e5644103a55ac122e36c3f9/	arXiv:2002.10619 (2020)
2	SCAFFOLD: Stochastic Controlled Averaging for Federated Learning	https://www.aminer.cn/pub/5df20fc53a55acbe6bfcc74f	Foundations and Trends® in Machine Learning, 2019.
3	Agnostic Federated Learning	https://www.aminer.cn/pub/5ced0f7da562983788d5f5f/	CoRR, pp. 4615-4625, 2019./ International Conference on Machine Learning. PMLR, 2019.

4	<i>cpSGD: Communication-efficient and Differentially-private Distributed SGD</i>	https://www.aminer.cn/pub/5b3d98cc17c44a510f8021d3/	arXiv:1805.10559 (2018)
5	<i>Distributed Mean Estimation with limited Communication</i>	https://www.aminer.cn/pub/58d82fcbd649053542fd6790/	International Conference on Machine Learning. PMLR, 2017

Tian Li

最高学位毕业院校: 美国卡内基梅隆大学 博士

研究兴趣: 大规模机器学习、分布式优化、数据密集型系统



Tian Li

[Carnegie Mellon University](#)

相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	<i>Federated Learning: Chanlleges, Methods, and Future Directions</i>	https://www.aminer.cn/pub/5d5e6b9a3a55acfc e79a16af/	IEEE Signal Processing Magazine, no. 3 (2019): 50-60
2	<i>Federated Learning in Heterogeneous Networks</i>	https://arxiv.org/abs/1812.06127	MLSys (2020).
3	<i>Fair Resource Allocation In Federated Learning</i>	https://www.aminer.cn/pub/5e5e18c493d709897ce2ee58/	ICLR, (2020)
4	<i>Feddane: A Federated Newton-type method</i>	https://www.aminer.cn/pub/5e15a8503a55ac40c85f6acd/	ACSSC, pp.1227-1231, (2020)

Yang Liu (刘洋)

清华大学智能产业研究院 副研究员/副教授

最高学位毕业院校：美国普林斯顿大学 博士

曾经任职：深圳前海微众银行股份有限公司 资深研究员、AI部门研究团队负责人、美国 Dataminr Inc公司 数据科学家、美国空气产品公司（Air Products）高级研究工程师等

研究兴趣：机器学习、联邦学习、迁移学习、多代理系统、统计力学以及这些技术在行业中的应用。



Yang Liu (刘洋) 绑定

副研究员
 Institute of AI Industry Research, Tsinghua University
 清华大学智能产业研究院

相关论文代表作：

序号	论文名称	论文地址	发表期刊/年份
1	<i>Secure Federated Transfer Learning Framework</i>	https://www.aminer.cn/pub/5ef4765b91e01165a63ba6a9/	IEEE Intelligent Systems, vol. 35, no. 4, pp. 70-82, 1 July-Aug. 2020
2	<i>FedVision: Visual Object Detection Powered by Federated Learning</i>	https://www.aminer.cn/pub/5e257a973a55acdfeeb9ed85/	Thirty-Second Annual Conference on Innovative Applications of Artificial Intelligence, AAAI, no. 08 (2020): 13172-13179
3	<i>Federated Machine Learning: Concept and Applications</i>	https://www.aminer.cn/pub/5c9f688c6558b90bfa34dcb4/	ACM Transactions on Intelligent Systems and Technology (TIST) 10.2 (2019): 1-19.

4	<i>A Communication Efficient Collaborative Learning Framework for Distributed Features</i>	https://www.aminer.cn/pub/5f8d43449e795ea21af78f0a	arXiv preprint arXiv:1912.11187 (2019) / FL-NeurIPS 2019
5	<i>BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning</i>	https://www.aminer.cn/pub/5f229e6d91e01155225f9e8e/	USENIX Annual Technical Conference 2020

② 算法安全（隐私保护）方向

Reza Shokri

新加坡国立大学 教授

最高学位毕业院校: 瑞士洛桑联邦理工学院EPFL 博士

研究兴趣: 计算机安全和隐私、机器学习



Reza Shokri

h-index: 27 | 论文数: 53 | 引用数: 4307

Professor

Department Of Computer Science, National University of Singapore

Location Privacy

Privacy

Ad Hoc Networks

Data Privacy

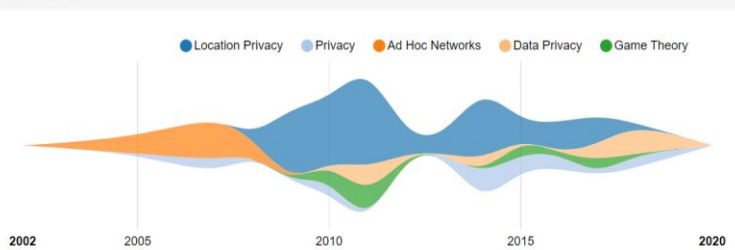
Game Theory

Mobile Computing

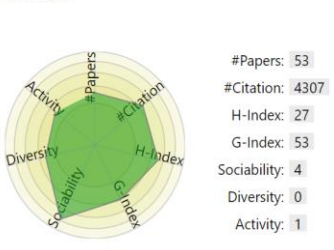
Location Based Services

Wireless Sensor Network

研究兴趣



作者统计



相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	<i>Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning</i>	https://www.aminer.cn/pub/5ce3a8a0ced107d4c655642b	IEEE symposium on security and privacy, pp.739-753, (2019)
2	<i>Machine Learning with Membership Privacy using Adversarial Regularization</i>	https://www.aminer.cn/pub/5b67b4b417c44aac1c8672c6/	Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security.
3	<i>Privacy Risks of Securing Machine Learning Models against Adversarial Examples</i>	https://www.aminer.cn/pub/5cf48a3dda56291d582a0290/	CCS, pp. 241-257, 2019.
4	<i>Synthesizing Plausible Privacy-preserving Location Traces</i>	https://www.aminer.cn/pub/57d063feac44367354297ed2/	IEEE Symposium on Security and Privacy, pp.546-563, (2016)
5	<i>Membership Inference Attacks against Machine Learning Models</i>	https://www.aminer.cn/pub/58437725ac44360f1083029c/	IEEE Symposium on Security and Privacy, (2017)

Dawn song (宋晓东)

加州大学伯克利分校电气工程与计算机科学系 教授

最高学位毕业院校：美国加州大学伯克利分校博士

曾经任职：卡内基梅隆大学助理教授

研究兴趣：深度学习、区块链和去中心化系统，计算机安全、隐私和应用密码学，使用程序分析、算法设计和机器学习来确保安全和隐私。



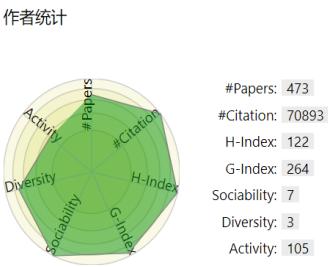
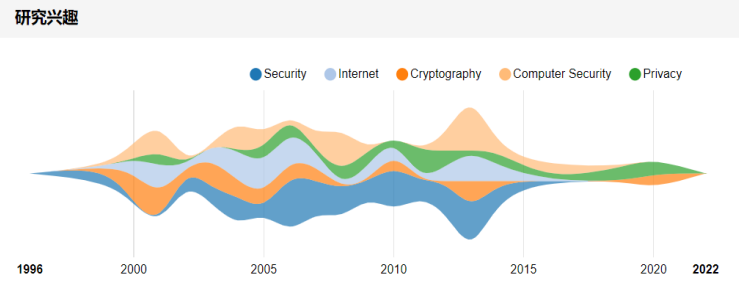
宋晓东 (Dawn Song)

h-index: 114 | 论文数: 473 | 引用数: 62130

教授

Department of Electrical Engineering and Computer Science, University of California Berkeley

Security Internet Cryptography Computer Security Privacy



相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	<i>Epione: Lightweight Contact Tracing with Strong Privacy</i>	https://www.aminer.cn/pub/5ea9503e91e0118eb1e19f59/	IEEE Data Eng. Bull., no. 2 (2020): 95-107
2	<i>Keystone: An Open Framework for Architecting Trusted Execution Environments</i>	https://www.aminer.cn/pub/5e9c27d49fced0a24b1f07fe/	EuroSys '20: Fifteenth EuroSys Conference 2020, Heraklion Greece, April, 2020, pp.1-16, (2020)

3	<i>The Secret Revealer: Generative Model-Inversion Attacks Against Deep Neural Network</i>	https://www.aminer.cn/pub/5dd3bf513a55ac1bdd46d7e4/	CVPR, pp.250-258, (2019)
4	<i>The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Network</i>	https://www.aminer.cn/pub/5d70dfa83a55acf39f3e7f83/	USENIX Security Symposium, pp.267-284, (2019)
5	<i>Towards Practical Differential Privacy for SQL Queries</i>	https://www.aminer.cn/pub/5a9cb64017c44a376ffb683a/	Proceedings of the VLDB Endowment, no. 5 (2018): 526-539
6	<i>Ekiden: A Platform for Confidentiality-preserving, Trustworthy, and Performant Smart Contracts</i>	https://www.aminer.cn/pub/5d67a2a73a55ac09fb007f8a/	EuroS&P, pp.185-200, (2019)
7	<i>Targeted Backdoor Attacks on Deep Learning System using Data Poisoning</i>	https://www.aminer.cn/pub/5a73cbc317c44a0b3035f0b8/	arXiv: Cryptography and Security, (2017)

Kallista.Bonawitz

谷歌公司

最高学位毕业院校：美国麻省理工学院 博士

研究兴趣：人工智能、隐私保护技术（差分隐私、安全多方计算）



Kallista Bonawitz

Google Inc. | Google · Research Department
Doctor of Philosophy

相关论文代表作:

序号	论文名称	论文地址	发表期刊/年份
1	<i>Practical Secure Aggregation for Privacy-preserving Machine Learning</i>	https://www.aminer.cn/pub/5df20fc53a55acbe6bfcc74f/	Proceedings of the2017 ACM SIGSAC Conference on Computer and Communications Security
2	<i>Secure Single-Server Aggregation with (Poly)Logarithmic Overhead</i>	https://www.aminer.cn/pub/5fae6f35d4150a363ceed232/	Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. 2020.
3	<i>Towards Federated Learning at Scale: System Design</i>	https://www.aminer.cn/pub/5df20fc53a55acbe6bfcc74f/	Found. Trends Mach. Learn., no. 1-2 (2021): 1-210
4	<i>Federated Learning with Autotuned Communication-Efficient Secure Aggregation</i>	https://www.aminer.cn/pub/5de632503a55ac4f55c25467/	ACSSC, pp.1222-1226, (2019)
5	<i>Practical Secure Aggregation for Federated Learning on User-held Data</i>	https://www.aminer.cn/pub/58d82fced649053542fd6b2d/	arXiv preprint arXiv:1611.04482 (2016)

3.1.5 专利申请现状

基于AMiner和智慧芽专利数据库，通过联邦学习相关关键词检索式^[35]，在“标题/摘要/权利要求”中进行相关专利搜索。数据结果显示，2016年至2020年五年期间，共计得到3088条联邦学习技术相关专利申请记录。

(1) 全球专利申请趋势逐年攀升

专利在一定程度上能够反映出某项技术的发展方向和潜在前景。从专利申请趋势来看，联邦学习全球专利申请量自2016年被提出以来，其热度就不断攀升，并在2020年达到高峰1111条专利，具体情况如图21所示。

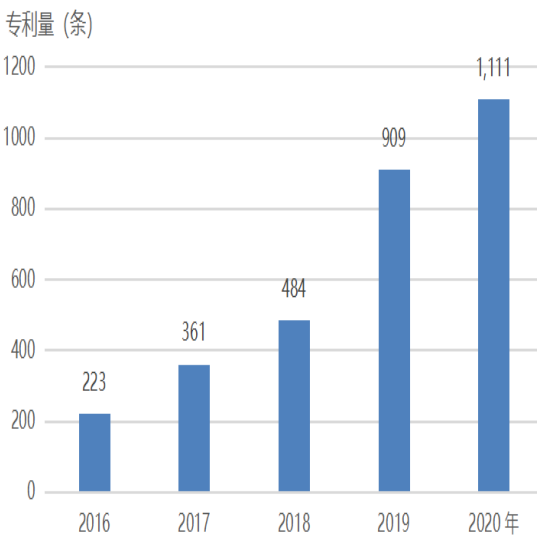


图 21 2016-2020年联邦学习专利申请趋势

(2) 全球专利受理情况以中国地区最多

全球范围内，联邦学习近五年的相关专利申请受理量领先国家和地区是中国、美国等。这些国家的联邦学习相关专利申请情况如图22所示。由图可见，联邦学习专利申请量最多的国家是中国，共1514条相关专利申请，数量明显大幅度高于其他国家。这反映出联邦学习技术创新和推广应用在中国更加热门。

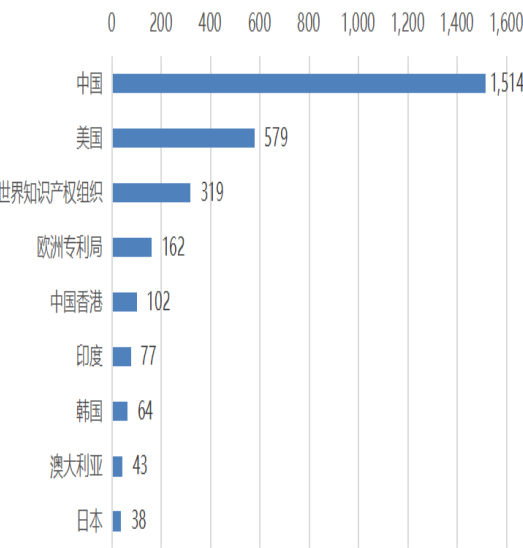


图 22 联邦学习专利申请全球受理局分布 (2016-2020年)

³⁵ 关键词检索式：TAC_ALL:(Federated Machine Learning OR Federated optimization OR federated learning OR federation learning OR (Privacy AND Distributed AND data mining) OR (Secure AND Distributed AND data mining) OR (Secure AND Multiparty) OR (Secure AND Multi-party) OR (privacy AND Multi-party) OR (privacy AND Multiparty) OR (Privacy AND Distributed AND machine learning) OR (Secure AND Distributed AND machine learning) OR (Privacy AND joint learning) OR (Secure AND joint learning) OR (Privacy AND Distributed AND deep learning) OR (Secure AND Distributed AND deep learning))

从专利申请受理量的年度趋势看，2016-2017年各国家之间联邦学习专利申请量相差不大，但是自2019年开始，中国研究人员的专利申请量出现了大幅度增长，2019年相比2018年实现翻番，并在2020年专利申请量达819条，约是2016年专利申请的12倍，也是美国当年申请量的近六倍。如图23所示。

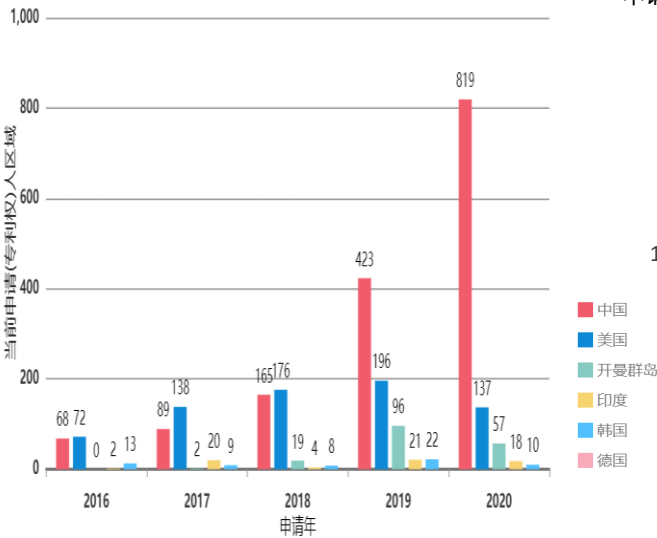


图 23 联邦学习专利申请受理国家趋势

(3) 专利申请人类型七成以上为企业

从专利第一申请人的所属机构类型来看，联邦学习专利第一申请人涵盖了企业、高校、研究所等不同类型，具体分布如图24所示，其中，企业第一申请人占比最多，达73.2%；高校申请人的数量紧随其后，占比为15.05%；以研究所为第一申请人的相关专利申请数量最少。可见，联邦学习相关专利申请主体主要是企业。

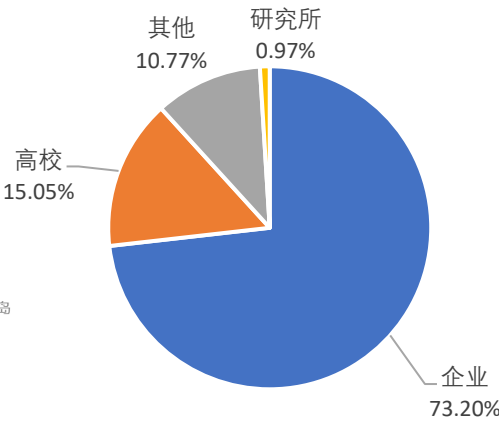


图 24 2016-2020年联邦学习专利第一申请人类型分布

(4) 专利申请量前十机构全部为企业

从专利人来看，联邦学习专利申请总量TOP10机构全部为企业，且中国企业占据七席；从同族专利申请TOP10机构来看，包括八家企业、

两家高校或科研机构，且中国企业占据两席，美国机构占据两席，具体情况如图25所示。值得注意的是，中国的深圳前海微众银行股份有限公司、支付宝信息技术有限公司的相关专利申请量全球领先。

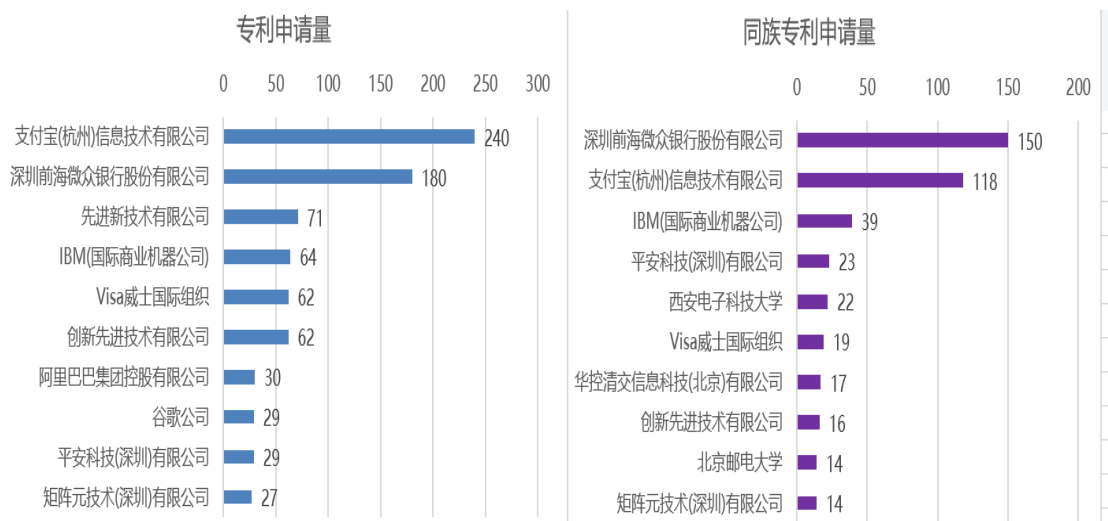


图 25 联邦学习专利申请量TOP 10 机构 (2016-2020年)

(5) 国内专利申请以广东、浙江和北京领先

在国内，近五年联邦学习专利申请量TOP10 省市分别是广东、浙江、北京、江苏、上海、四川、

陕西、湖北、安徽和山东，详细申请情况如图26所示。其中，广东、浙江和北京属于该领域第一梯队，专利申请量均高于100，明显超过其他省市。

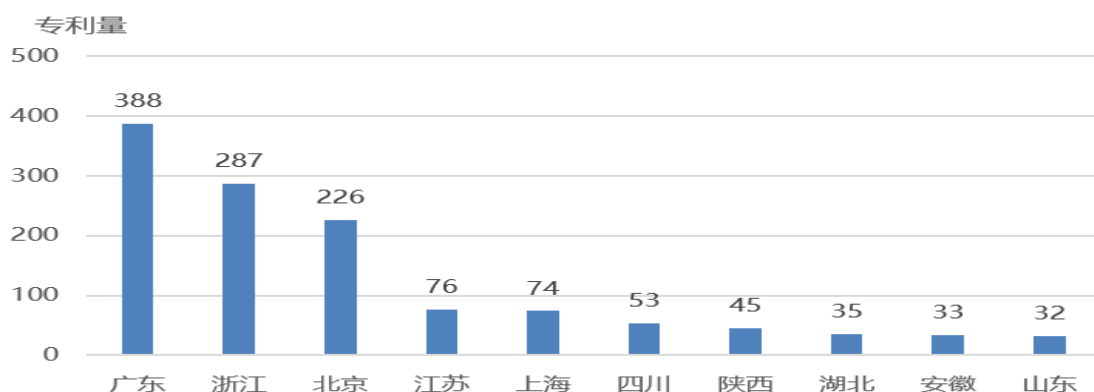


图 26 联邦学习专利量TOP 10 国内省市分布（2016-2020年）

(6) 专利申请关键词最多聚焦于机器学习

基于前文所述的关键词检索式，获得联邦学习相关专利，再通过算法对联邦学习相关专利进行统计分析和文本聚类，提取该领域排名靠前的关键词并制作词云图，如图27所示。最热门的联邦学习技术主题词包括机器学习、分布式、模型参数、模型训练、电子设备、数据处理、训练方法、隐私保护、全局模型等。这反映出联邦学习目前的专利布局主要聚焦机器学习方法、模型训练、隐私保护三大方面。同时，对学习效率、安全等方面的专利布局相对较少。



图 27 联邦学习相关专利申请涉及的关键词云

(7) 专利申请最多布局在数据存取访问平台保护与机器学习两个IPC分类

在联邦学习专利之中，申请数量最热门的专利IPC分类是G06F21/62（通过一个平台保护数

据存取访问，例如使用密钥或访问控制规则〔2013.01〕〔2013.01〕），专利量是562条；其次是G06N20 机器学习〔2019.01〕，机器学习也是该技术的重要技术研究分支，相应的专利量为559条。详细信息如图28所示。

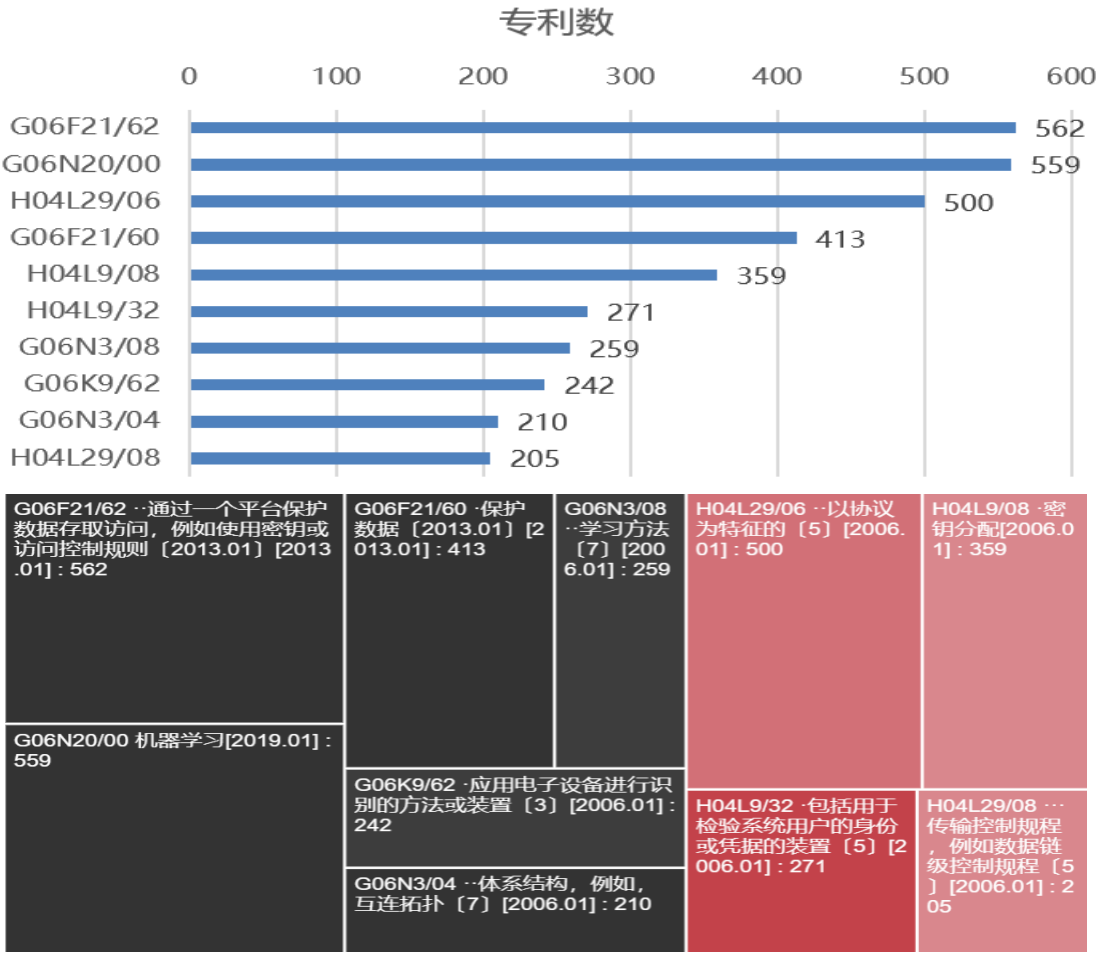


图 28 联邦学习专利申请量TOP 10 的IPC分类

在数据存取访问平台保护、机器学习两个最热门的IPC分类下，TOP 5的专利申请机构主要来自中国和美国，详细情况如图29所示。其中，支付宝公司在数据存取访问平台保护方面进行了最多数

量的联邦学习专利布局，微众银行在机器学习方面进行了最多数量的联邦学习专利布局，此外，IBM、创新先进技术公司和平安科技等公司也在相应类别下进行了不同数量的专利布局。

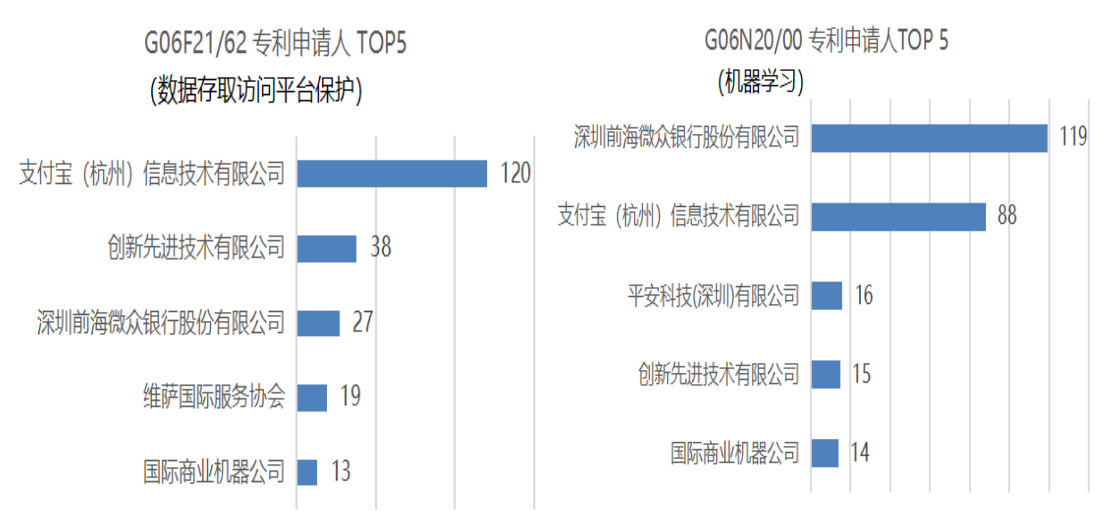


图 29 数据存取访问平台保护、机器学习两个IPC分类下联邦学习专利领先申请人

(二) 联邦学习框架与系统现状

近年来，联邦学习算法框架和系统的开发和部署正在蓬勃发展。目前，市面上有许多来自于科研机构或企业的关于联邦学习的开源工程。本部分通过AMiner 数据库中的新闻数据，分析梳理

了国内外知名高校、科研机构、科技企业巨头、金融科技公司，以及初创公司等推出的主要联邦学习相关系统框架。

本报告将这些联邦学习框架分为开源与非开源两种，具体信息如下。

3.2.1 开源框架

开源的联邦学习框架多数是由国内外企业推出的，高校科研机构发布的相对较少。根据其在Github（代码托管服务平台）上的热度排序（数

据统计日期截至到2021年5月31日），其中，OpenMined 推出的Pysyft、微众银行的FATE和谷歌的TFF框架的热度均过千，且居于前三。联邦学习相关开源系统框架的详细信息如表6所示。

表 6 开源的联邦学习框架

Github 热度	发布方	系统名称	开源时间	系统特点
7300	OpenMined	PySyft	2017.7	● 一个用于安全和私有深度学习的 Python 库 ● 基于 PyTorch，使用 Unity Game Engine ● 安全多方计算 ● 联合学习、差异隐私
3100	微众银行	FATE	2019.2	● 工业级框架，采用Python开发，底层计算存储基于EGGROLL、Spark等高性能计算引擎 ● 提供一站式的联邦模型企业级服务解决方案。提供多插件支持联邦学习企业和科研应用 ● 支持主流的分类、回归、聚类 and 迁移学习的联邦化算法 ● 提供多种安全计算协议支撑上层应用，支持同态加密协议、秘密共享协议、不经意传输协议和DH密钥交换算法等 ● 提供20多个联邦算法组件
1500	谷歌	TensorFlow Federated , TFF	2019.3	● 可以选择 ML 模型架构 ● 模型设计理念以数据为主
872	DropoutLabs, OpenMined, 阿里巴	TF-Encrypted	2018.3	● 安全多方计算
799	Facebook	CrypTen	2019.10	● 安全多方计算

Github 热度	发布方	系统名称	开源时间	系统特点
624	美国南加州大学	FedML	2020.7	支持三种计算范例：分布式训练、移动设备训练、独立仿真
586	字节跳动	Fedlearner	2020.1.20	- 代码里有大量的JS、HTML模块 - 强调联邦学习在推荐、广告等业务中的落地 - 可输出性
302	矩阵元	Rosetta	2020.8	- 安全多方计算 - 基于TensorFlow
250	百度	PaddleFL	2020.2.19	- 可信计算 - 基于飞桨（PaddlePaddle）和 Kubernetes - 面向深度学习设计，提供在计算机视觉、自然语言处理、推荐算法等领域的联邦学习策略及应用场景 - 简化大规模分布式集群部署 - 二次开发接口允许各方定义私有化的数据读取器 - 提供了基础编程框架，并封装了一些公开的联邦学习数据集
74	京东	9NFL 九数联邦学习	2020初	- 支持百亿级规模样本、百T级容量数据的超大规模的样本匹配、联合训练 - 在电商推荐领域可实现线上业务落地 - 实现分布式异步框架、Failover、拥塞控制等机制 - 针对跨域与跨公网的复杂环境，设计了一系列的可用性与容灾的机制与策略
34	同盾科技	FLEX	2020.2	一套标准化的联邦协议：约定了联邦过程中参与方之间的数据交换顺序，以及在交换前后采用的数据加解密方法
7	台湾人工智能实验室	Harmonia	2020.6	去中心化的信息分享算法

以上部分的联邦学习系统框架的详细介绍信息如下。

(1) OpenMined——PySyft

PySyft 是开源社区OpenMined推出的一个用于安全和私有深度学习的 Python 库。它使用联邦学习、差分隐私和加密计算来解耦私人人和敏感数据，可以在主要的深度学习框架中使用，例如TensorFlow 和 PyTorch。PySyft 代表了在深度学习程序中启用可靠的隐私模型的首批尝试之一。

PySyft 的核心组件是称为 SyftTensor 的抽象。SyftTensors旨在表示数据的状态或转换，并且可以链接在一起。链结构始终在其头部具有PyTorch张量，并且使用child属性向下访问由SyftTensor体现的变换或状态，而使用parent属性向上访问由SyftTensor体现的变换或状态。

开源地址：
<https://github.com/OpenMined/PySyft>
PySyft 的系统框架如图30所示。

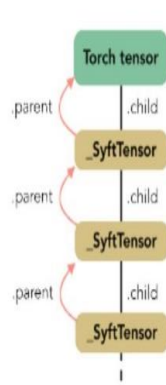


Figure 1: General structure of a tensor chain

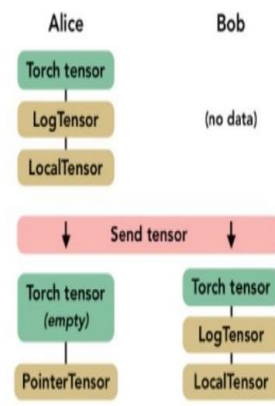


Figure 2: Impact of sending a tensor on the local and remote chains

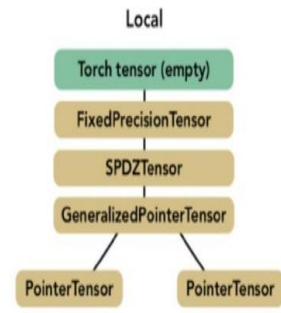


Figure 3: Chain structure of a SPDZ tensor

来源：<https://arxiv.org/pdf/1811.04017.pdf>

图 30 OpenMined PySyft系统框架

(2) 微众银行——FATE

微众银行AI部门研发了FATE（Federated AI Technology Enabler）联邦学习开源项目，是首个开源的联邦学习工业级框架。目前FATE开源社区已汇聚了700多家企业、300余所高校等科研机构的开发者，是国内最大的联邦学习开源社区。FATE项目使用多方安全计算（MPC）以及同态加密（HE）技术构建底层安全计算协议，以此支持不同种类的机器学习的安全计算，包括逻辑回归、树算法、深度学习（人工神经网络）和迁移学习等。FATE目前支持三种类型联邦学习算法：横向

联邦学习、纵向联邦学习以及迁移学习。开源地址：<https://github.com/FederatedAI/>

FATE整体架构如图31所示。FATE主仓库包含FederatedML核心联邦算法库和多方联邦建模Pipeline调度模块FATE-Flow，FATE拥抱大数据生态圈，底层引擎支持使用微众银行自主研发的EGGROLL或者Spark进行高性能的计算。围绕FATE联邦学习生态，FATE还提供了完整的联邦学习生态链，如联邦可视化模块FATE-Board、联邦在线推理模块FATE-Serving、联邦多云管理FATE-Cloud等。

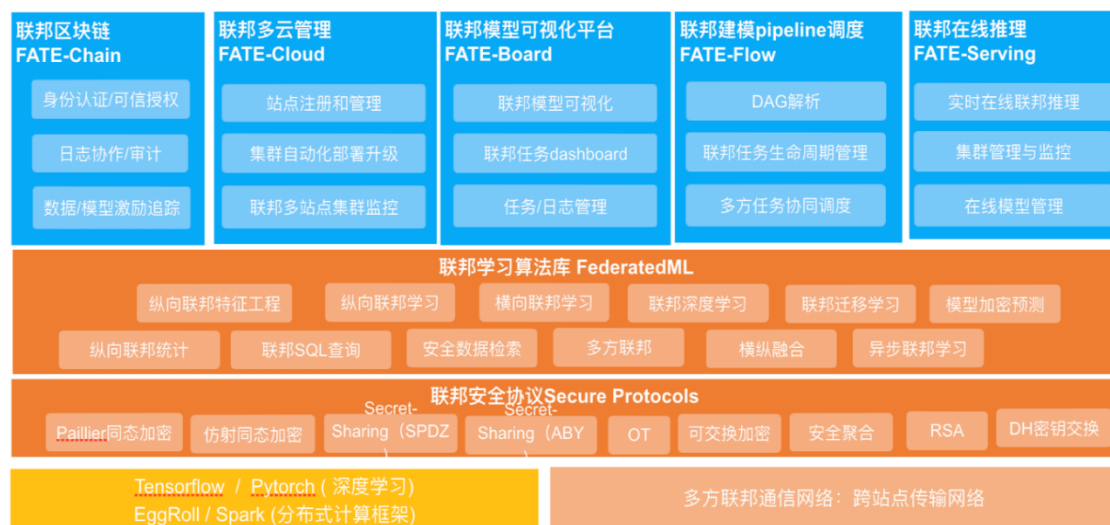


图 31微众银行 FATE 系统架构



FederatedML是FATE的联邦学习算法库模块，提供了20+种联邦学习算法，支持纵向联邦学习、横向联邦学习、联邦迁移学习三种联邦建模场景，覆盖了工业建模的数据处理、特征变换、训练、预测、评估的全建模流程，如图32所示。另外，封装了众多的多方安全计算协议以提供给上层算法的调度和支持联邦学习开发者的联邦算法开发。

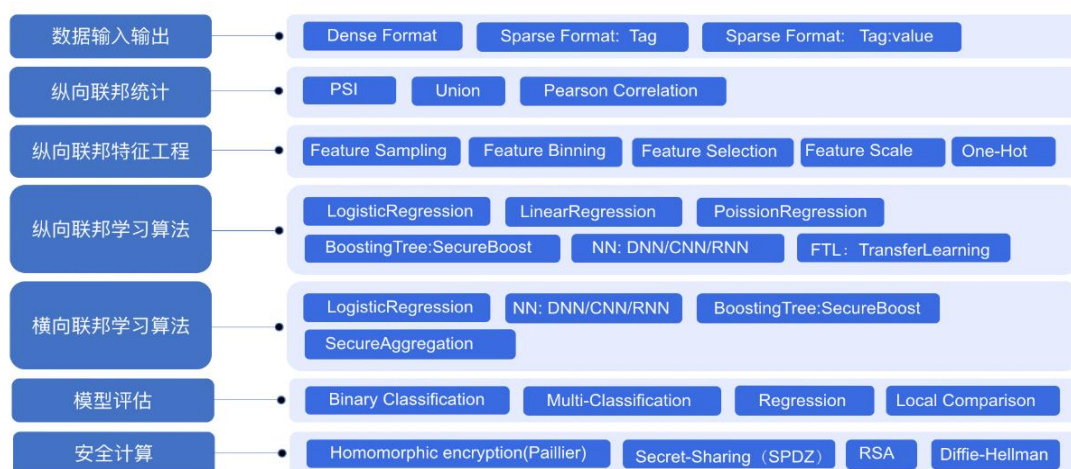


图 31微众银行 FATE-FederatedML

FATE-Flow为FATE提供了端到端联邦建模Pipeline调度和管理，主要包括DAG定义联邦建模pipeline、联邦任务生命周期管理、联邦任务协同调度、联邦任务追踪、联邦模型管理等功能，实现了联邦建模到生产服务一体化。

FATE-Board联邦学习建模的可视化工具，为终端用户提供可视化和度量模型训练的全过程。FATE-Board由任务仪表盘、任务可视化、任务管理与日志管理等模块组成，支持模型训练过程全

流程的跟踪、统计和监控等。

FATE-Serving为FATE提供联邦在线推理服务，主要包含实时在线预测、集群管理与监控、在线模型管理与监控、服务治理等功能。

FATE-Cloud是构建和管理联邦数据合作网络的基础设施，为跨机构间、机构内部不同组织间提供了安全可靠、合规的数据合作网络构建解决方案，实现多客户端的云端管理，

FATE Chain是联邦学习区块链网络框架，在满足多方数据隐私、安全和监管要求的前提下，将联邦学习与区块链融合，提供去中心化的应用，通过分布式的存储和运行，保证极高的透明度和安全性，使数据管控满足不可篡改、可追溯、可审计等要求，实现联邦多方对等合作与合规有序的发展。

(3) 谷歌——TensorFlow Federated, TFF

TensorFlow Federated project (TFF) 由谷歌公司开发和维护，是一个为联邦机器学习和其他计算方法在去中心化数据集上进行实验的开源框架。TFF 让开发者能在自己的模型和数据上模拟实验现有的联邦学习算法，以及其他新颖的算法。TFF 提供的建造块也能够应用于去中心化数据集上，来实现非学习化的计算，例如聚合分析。TFF 的接口有两层构成：联邦层（FL）应用程序接口（API）和联邦核心（FC）API。TFF 使得开发者能够声明和表达联邦计算，从而能够将其部署于各类运行环境。TFF 中包含的是一个单机的实验运行过程模拟器。

该联邦学习的训练流程，如图33所示。训练的三个步骤A、B、C是不断循环迭代的，具体含义如下。

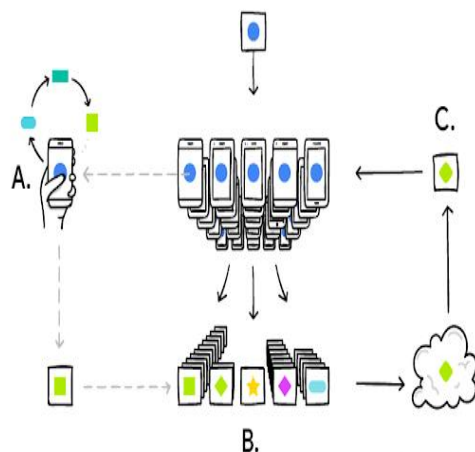


图 33 谷歌TFF联邦学习训练流程图

步骤A：clients从中心节点获取当前模型参数（蓝色圆），在本地设备上利用自己的local data对模型进行训练，得到新的模型参数（绿色方块）。

步骤B：多个clients分别训练了不同的模型参数（各种颜色的各种形状），融合为一个模型。注意，图33中各种形状的数量是不等的，一方面是由于local data的分布不均，导致训练出来的模型也分布不均，另一方面是由于不是所有的clients都参与了训练。

步骤C：中心服务器server将新一轮的模型分发给clients，在这个过程中，server可以选择只让一部分clients参与训练。

在实现方面，Tensorflow专门为联邦学习推出了一个学习框架（TensorFlow Federated，简称TFF），现有的TensorFlow（简称TF）或Keras模型代码通过一些转换后就可以变为联邦学习模型，甚至可以加载单机版的预训练模型，以迁移学习的模式应用到分散式数据的机器学习中。

不同于分布式训练理念，**TFF框架设计理念是以数据为主**，而不是代码分离上。在编写模型、训练代码的时候，将clients和server看作一个整体，同一个文件里不需要分割开Server端（S端）和Clients端（C端）的代码，C端和S端的区分是在代码逻辑层面的。也就是说，用户在编写TFF代码时，不需要指明某段代码是应该运行在C端还是S端）仅需要指出每个数据是储存在C端/S端、是全局唯一的还是有多份拷贝的即可。类似TF的non-eager模式，当用户编写完模型代码和训练代码后，TFF会自动地将代码分别放置到clients和server设备上。用户只要关注模型架构、C&S端交互的数据格式、聚合多clients模型的方式即可。

TFF通过Python代码来编写运算逻辑，实际运行则是编译成另一种语言去执行，以便让模型能运行在真实分布式场景下。

开源地址：

<https://github.com/tensorflow/federated>

(4) 字节跳动——Fedlearner

字节跳动联邦学习平台 Fedlearner 基于字节跳动在推荐和广告领域积累的机器学习建模技术和个性化推荐算法，可以支持多类联邦学习模式，已经在电商、金融、教育等行业多个落地场景实际应用。该平台已经于2020年初开源并持续更新，

开 源 地 址：

<https://github.com/bytedance/fedlearner>。

Fedlearner联邦学习平台整个系统包括控制台、训练器、数据处理、数据存储等模块，各模块对称部署在参与联邦的双方的集群上，透过代理互相通信，实现训练。

Fedlearner 双方在发起训练之前，必须要基于双方的数据进行求交，找出交集从而实现模型训练。训练数据求交的方式主要分为两种：流式数据求交、PSI 数据求交。

(5) 百度——PaddleFL

PaddleFL 是一个基于百度飞桨（PaddlePaddle）的开源联邦学习框架。PaddleFL提供很多联邦学习策略及其在计算机视觉、自然语言处理、推荐算法等领域的应用，例如，横向联邦学习（联邦平均、差分隐私、安全聚合）和纵向联邦学习（带privc的逻辑回归，带ABY3的神经网络）。研究人员可以用PaddleFL复制和比较不同的联邦学习算法。

此外，PaddleFL还提供传统机器学习训练策略的应用，例如多任务学习、联邦学习环境下的迁移学习、主动学习。依靠PaddlePaddle的大规模分布式训练和Kubernetes对训练任务的弹性调度能力，PaddleFL可以基于全栈开源软件轻松地部署。PaddlePaddle背靠百度的信息库，提供的预训练模型的准确率较高。

开源地址：

<https://github.com/PaddlePaddle/PaddleFL>。

整体架构如图34所示。



图 34 百度PaddleFL整体架构

PaddleFL 中主要提供两种解决方案：Data Parallel 以及 Federated Learning with MPC (PFM)。通过Data Parallel，各数据方可以基于经典的横向联邦学习策略（如 FedAvg, DPSGD 等）完成模型训练。此外，PFM是基于多方安全计算（MPC）实现的联邦学习方案。作为 PaddleFL的一个重要组成部分，PFM可以很好地支持联邦学习，包括横向、纵向及联邦迁移学习等多个场景。

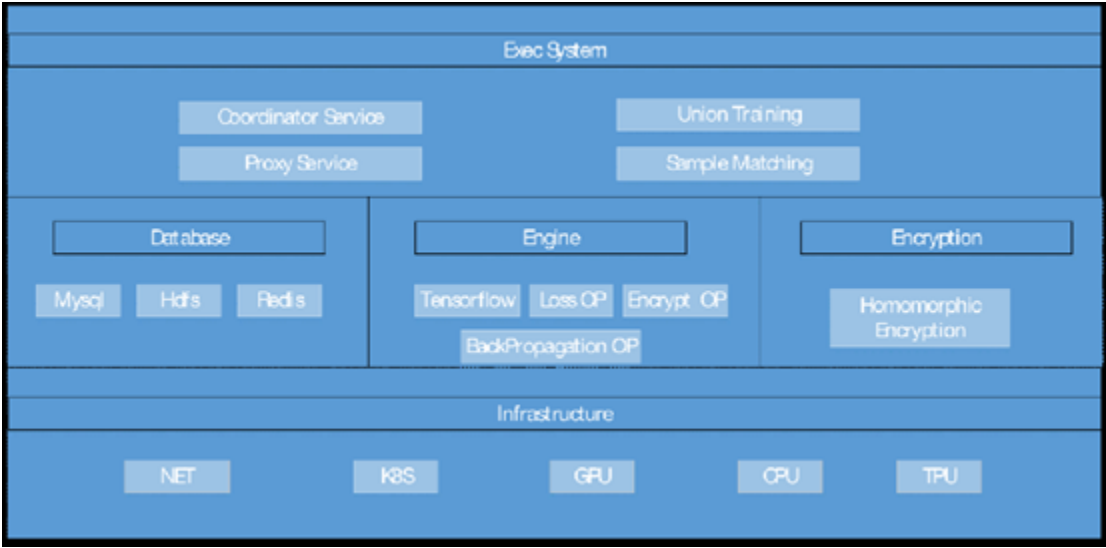
(6) 京东——九数联邦学习9NFL

京东自研的九数联邦学习平台（9NFL）于2020年初正式上线。9NFL平台基于京东商业提升事业部9N机器学习平台进行开发，在9N平台

离线训练、离线预估、线上推断（inference）、模型的发版等功能的基础上，增加了多任务跨域调度、跨域高性能网络、大规模样本匹配、大规模跨域联合训练、模型分层级加密等功能。整个平台可以支持百亿级/百T级超大规模的样本匹配、联合训练，并且针对跨域与跨公网的复杂环境，对可用性与容灾设计了一系列的机制与策略,保障整个系统的高吞吐、高可用、高性能。

开源地址：<https://github.com/jd-9n/9nfl>

9NFL整体系统架构分为四大模块：整体调度与转发模块、资源管理与调度模块、数据求交模块、训练器模块。如图35所示。



来源：新浪VR^[36]

图 35 九数联邦学习平台 (9NFL)

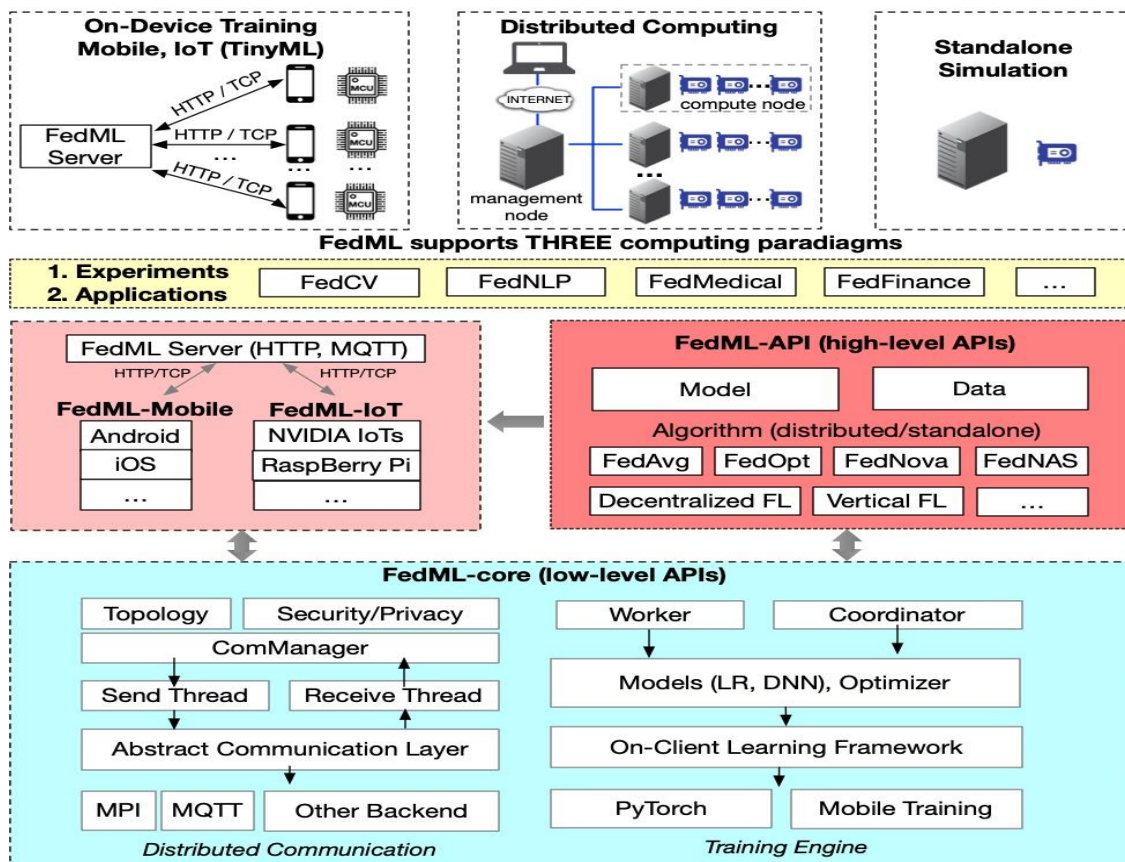
(7) 美国南加州大学——FedML

美国南加州大学USC联合MIT、Stanford、MSU、UW-Madison、UIUC以及腾讯、微众银行等众多高校与公司联合发布了FedML联邦学习开源框架。FedML是一个开放的研究库和基准，支持分布式训练、移动设备训练、独立仿真三种计算范例，可促进新的联合学习算法的开发和公平的性能比较。其系统架构如图36所示。

FedML还通过灵活且通用的API设计和参考基准实现和促进了各种算法研究。针对非I.I.D设置的精选且全面的基准数据集旨在进行公平比较。FedML可以为联合学习研究社区提供开发和评估算法的有效且可重复的手段。

开 源 地 址：<https://github.com/FedML-AI/FedML>

³⁶ 京东开源超大规模联邦学习平台，2020-09-15 来源：新浪VR，<http://vr.sina.com.cn/news/hz/2020-09-15/doc-iivhvpwy6836041.shtml>



来源: FedML-AI/FedML, <https://github.com/FedML-AI/FedML>

图36 美国加州大学USC FedML系统架构

(8) 台湾人工智能实验室——Harmonia

台湾人工智能实验室 (AI Labs) 开发了一个开源项目Harmonia，旨在开发系统/基础设施和图书馆，以简化联合学习的研究和生产用途。Harmonia 使用工程师熟悉的环境和语言，比如热门的开源工具Kubernetes、Git Large File Storage和GitOps等。Harmonia 利用 Git 进行访问控制、模型版本控制和服务器和联合培训 (FL) 运行参与者之间的同步。FL 训练策略、全局模型和本地模型/渐变保存在 Git 存储库中。这些 Git repositories 的更新会触发 FL 系统状态转换。这将自动化 FL 培训过程。

K8S 吊舱。操作容器负责维护 FL 系统状态，并通过 gRPC 与应用程序容器通信。本地训练和聚合函数封装在应用程序容器中。此设计可在 Kubernetes 群集环境中轻松部署，并快速插件现有机器学习 (ML) workflow。

开 源 地 址 : <https://github.com/ailabstw/harmonia>

Harmonia系统架构如图37所示。

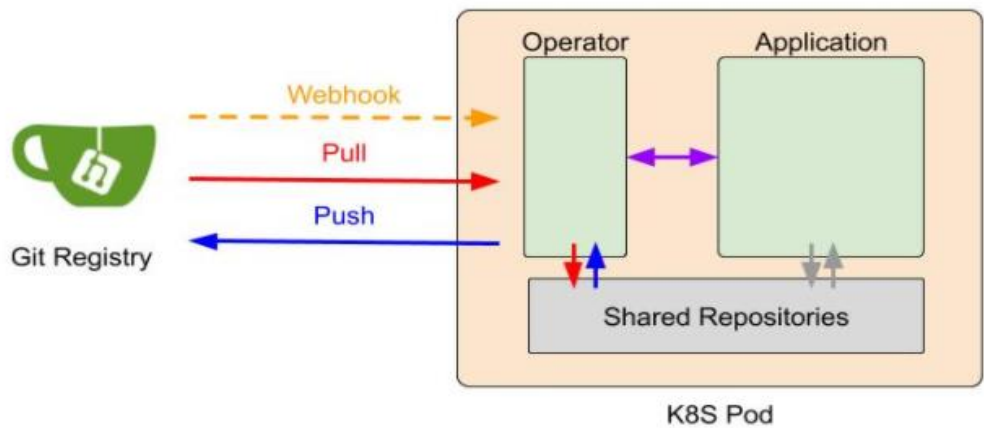


图 37 Harmonia系统架构

3.2.2 非开源框架与系统

非开源的联邦学习框架基本上都是由企业推出的。根据其正式发布时间进行排序，发现这些联邦学习框架主要集中发布于2019至2020年期间。其中，翼方健数的联邦学习框架发布时间最早。非开源联邦学习系统框架的详细信息如表7所示。

图 37 Harmonia系统架构

发布时间	发布方	系统名称	系统特点
2019年4月15日	翼方健数	翼数坊XDP	● 基于隐私计算的原理和应用 ● 通过多方安全计算MPC/同态加密、联邦学习、安全沙箱计算/TEE等技术实现 ● 通过自主研发的DaaS服务进行数据治理和清洗以达到数据可用
2019年9月5日	星云Clustar	AIOS	● 以联邦学习和区块链作为基础设施 ● 采用 FATE 联邦学习软件框架
2019年9月19日	华为	NAIE	目前以横向联邦为基础，内置了众多联邦学习能力，包括联邦汇聚、梯度分叉、多方计算、压缩算法等。
2020年3月23日	腾讯	AngelFL Angel PowerFL	● 支持超大规模数据量的多方联合建模 ● 有高容错性 ● 不依赖于可信第三方
2020年4月23日	上海富数科技	FMPC	● 密文训练联邦学习误差小于 1% ● 安全计算支持的算法包括：普通多方计算、统计分析、机器学习（LR、DT、RF、LightGBM等） ● 机器学习训练收敛速度提高了3倍；匿踪查询100亿条+记录秒级响应 ● 支持本地私有化、对等网络链接的部署
2020年5月27日	光之树科技	天机可信计算框架、云间联邦学习平台	● 基于芯片TEE技术和其他加密技术的可信计算体系 ● 基于机器学习、深度学习算法和加密协议的安全计算框架

发布时间	发布方	系统名称	系统特点
2020年8月28日	平安科技	蜂巢平台	● 定位是服务于营销、获客、定价、风控、智慧城市和智慧医疗 ● 支持传统的统计学习以及深度学习的模型，比如逻辑回归、线性回归、树模型等 ● 提供加密方式，支持同态加密等多方安全计算机制。在模型训练中，对梯度进行非对称加密，整合梯度和参数优化、更新模型；最后加密原始传输数据，实现推理结果 ● 支持单机和多机训练 ● 可使用CPU和GPU训练 ● 支持多种深度学习框架，如TensorFlow，Keras，Pytorch，Mxnet
2020年10月12日	京东数科	Fedlearn	● 提出了并行加密算法、异步计算框架、创新联邦学习等技术架构，达到融合亿级规模数据的能力 ● 在通讯方面，引入中心化数据交换的概念，使得数据交换独立于参与方 ● 采用异步计算框架，提高了模型训练速度，并推动异步联邦学习的发展 ● 应用于信贷风控、智能营销等方向

来源：根据公开资料整理

以上部分非开源的联邦学习系统平台的介绍信息如下。

(1) 腾讯——AngelFL

AngelFL联邦学习平台构建在Angel智能学习平台的基础之上，是一种“无可信第三方”的联邦学习框架。整个系统以 Angel 的高维稀疏训练平台作为底层，抽象出“算法协议”层，供实现各种常见机器学习算法。

AngelFL目前没有开源。腾讯于2021年1月22日申请公开“联邦学习方法、装置、计算机设备及介质”专利信息，公开号为CN112257876A。

AngelFL联邦学习系统架构如图38所示。该架构图在技术上，以两个业务方 A 和 B 为例，也

可以扩展到多个参与方的场景。A、B 双方分别拥有各自的与用户相关的数据，存储在本地集群，在整个联合建模的训练过程中，A 和 B 双方的原始数据均不出本地。在上述系统框架的基础上，再抽象出一层算法协议层，利用平台提供的计算、加密、存储、状态同步等基本操作接口，实现各种联邦机器学习算法。在训练任务执行时，通常拥有标签的一方作为训练的驱动方，算法协议层会控制本地训练步骤，例如梯度计算、残差计算、模型更新、消息发送等，同时与 AngelFL 流程调度模块交互同步执行状态，并按照协议触发对方进行下一步动作。

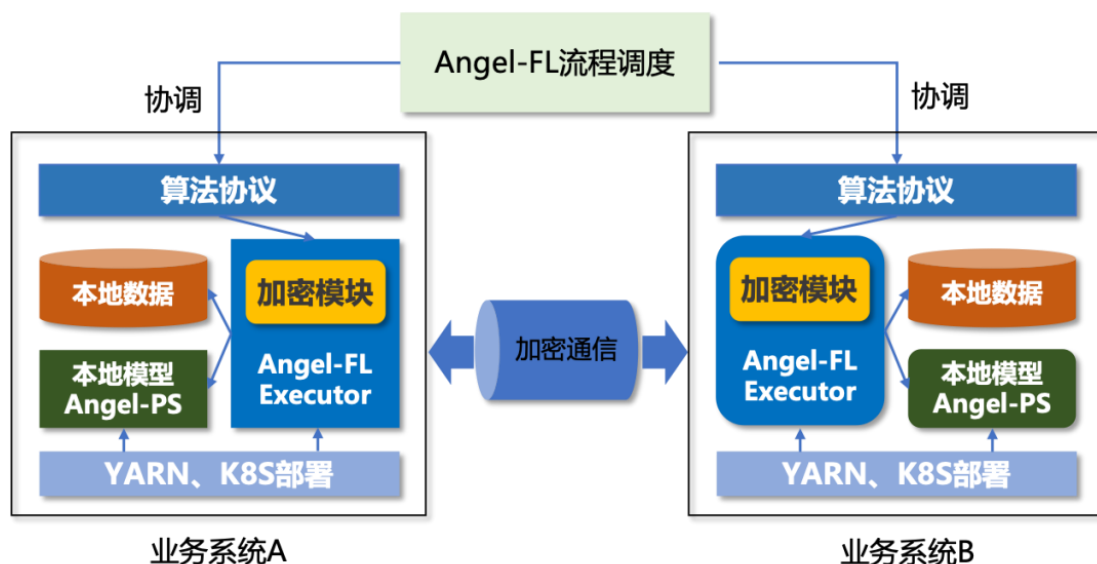


图 38 腾讯AngelFL联邦学习系统架构图

AngelFL架构具有以下特点：

(1) A、B 两方独立部署 Angel 联邦学习的本地框架，支持 YARN、K8S 多种资源申请方式，与业务现有系统完全兼容，灵活易用；

(2) 本地训练框架 AngelFL Executor 的计算采用 Spark，充分利用其内存优先和分布式并行的优点，效率高且易于和现有大数据生态（如 HDFS 等）打通；

(3) 本地模型保存在 Angel-PS 参数服务器中，支持大规模数据量训练；同时，PS 写有 checkpoint，意外失败的任务可以从上次保存的进度继续执行，具有很好的容错性；

(4) 模型训练相关的数据经过加密模块加密后，在 A、B 两方之间直接通信而不依赖第三方参与“转发”，实现了“去中心化”，整个训练流程仅需要协调双方的进度即可，能够增强实际应用中的安全性。

(2) 京东数科——Fedlearn

京东数字科技集团（简称：京东数科，现名：京东科技）于2020年10月推出自主研发的联邦学习平台Fedlearn。Fedlearn平台具有“六位一体”核心能力：多自研联邦学习算法、多方同态加密、轻量级分布式架构、区块链与联邦学习融合、数据安全容器、一站式操作平台。

京东数科Fedlearn平台具有三大特点：

第一，在数据和模型隐私方面，不同参与方之间没有直接交换本地数据和模型参数，而是交换更新参数所需的中间数值。为了避免从这些中间数值中恢复数据信息，采用增加扰动对这些数值进行保护，确保了数据和模型的隐私安全；

第二，在通讯方面，引入中心化数据交换的概念，使得数据的交换独立于参与方；

第三，采用异步计算框架，提高了模型训练的速度。

Fedlearn平台融合了密码学、机器学习、区块链等联邦学习算法，搭建出一套安全、智能、高效的链接平台，在各机构数据不用向外传输的前提下，通过联合多方机构数据，实现共同构建模型等多方数据联合使用场景，获得加成效应。相较于传统的数据共享交换方法，Fedlearn平台创新性地提出了并行加密算法、异步计算框架、创新联邦学习等技术架构，在保证数据安全的前提下提升学习效率，并逐步达到融合亿级规模数据的能力。

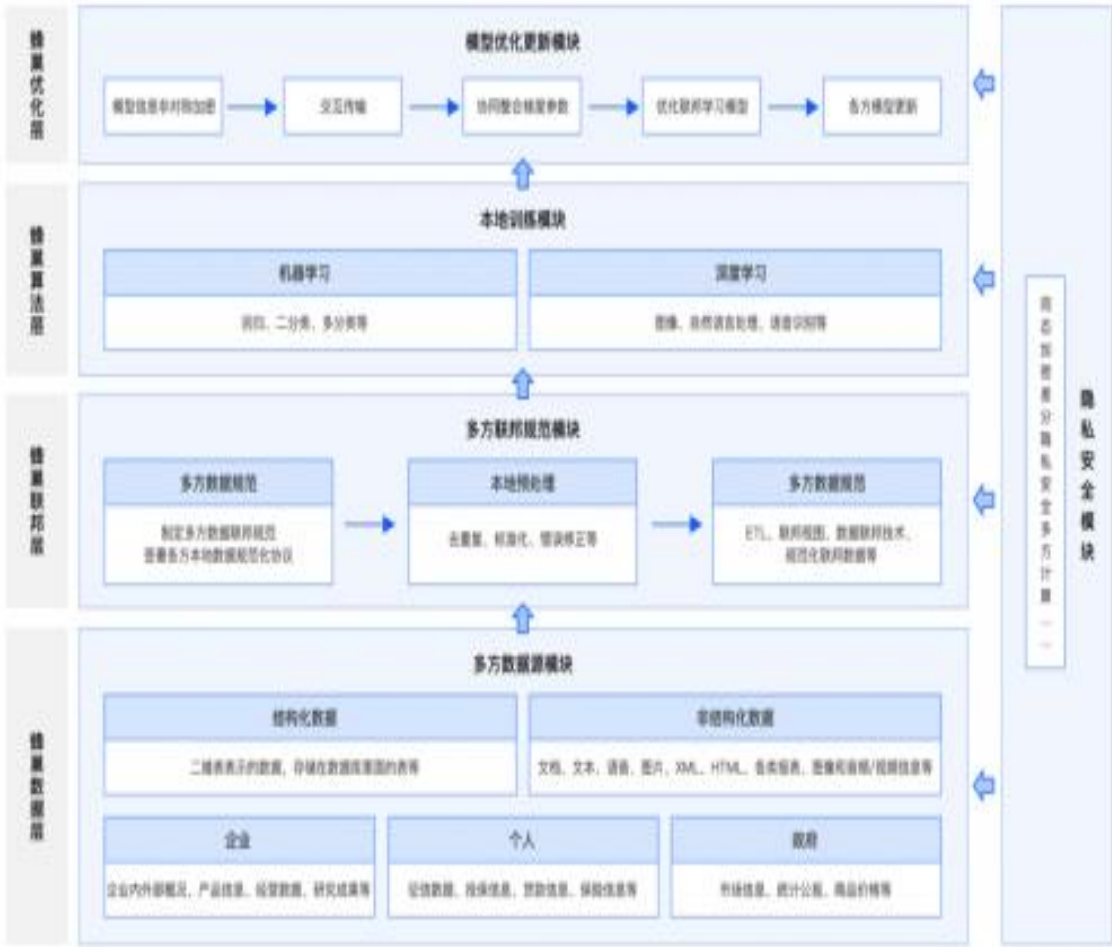
京东数科Fedlearn平台实现了“基于核的非线性联邦学习算法”。这一方法不传输原始样本及梯度信息，充分保护数据隐私；并使用首创的双随机梯度下降，大大提高计算速度，充分利用计算资源，通过增加扰动提高数据的安全保护。

产 品 地 址：

<https://www.jddgglobal.com/products/digitalgateway>

(3) 平安科技——蜂巢

平安科技研发的蜂巢联邦智能平台，是数据安全保护、企业数据孤岛、数据垄断、数据壁垒等问题的商用级解决方案。它能够让参与方在不共享原始数据的基础上联合建模，从技术上打破数据孤岛，从而综合化标签数据，丰富用户画像维度，从整体上提升模型的效果，实现 AI 协作。蜂巢平台的功能框架如图39所示。



来源：平安官网链接<https://tech.pingan.com/>

图 39 蜂巢平台功能结构

平安科技联邦智能平台蜂巢的建模是在保护用户隐私的前提下进行。原始数据不离开用户，建模所交换的是模型的中间参数和梯度。此外，采用GPU等异构计算芯片来加速联邦学习的加密和通信过程，从而达到效率升级的效果。

(4) 富数科技——FMPC

富数多方安全计算平台（FMPC）是上海富数科技旗下产品，目前未开源，主要通过体验或者服务购买方式使用。产品官网地址：<https://www.fudata.cn/>

FMPC目前公开的技术架构如图40所示。

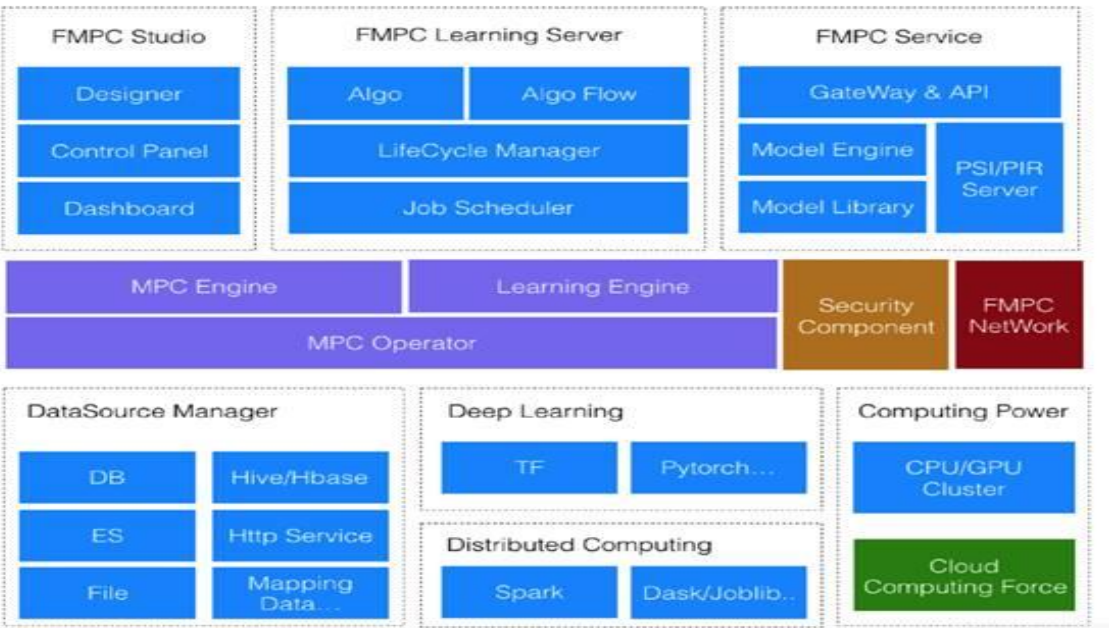


图 40 富数科技 FMPC 系统架构

- FMPC架构具有以下特点：
- ① 联邦学习：原始数据不出门，参与各方本地建模；没有敏感数据流通，只交互中间计算结果；整个模型被保护，参与各方只有自己模型参数；私有化部署；开放API快速开发；支持主流机器学习算法，如LR, DT, RF, Xgboost等；建模速度快3倍；密文训练精度误差<1%。
 - ② 多方安全计算：落地应用计算量1.1万+次/天；支持多方数据安全求交；支持一次多项式；

- 支持多方归因统计分析；支持多方多维数据钻取分析；私有化部署。
- ③ 匿踪查询：支持100亿+条记录；秒级响应时间；查询授权存证；甲方查询信息不泄露；加密隧道避免中间留存；私有化部署。
- ④ 联盟区块链：联盟节点30+；高性能扩展1万TPS；合约调用20万次/天；电子存证和智能合约；隐私保护协议；快捷部署场景应用；开源开发社区。

(5) 星云cluster ——AIOS

星云AIOS (AI Operating System) 是一款具备高性能、高可靠、高灵活及高扩展特性的人工智能操作系统，由高性能AI加速中间件、深

度学习训练平台及数据推理平台三个子系统构成，为用户提供数据处理、模型训练、推理服务及AI应用等完整的AI解决方案。总体框架如图41所示。



图 41 星云AIOS系统框架

来源：星云Cluster官网

AIOS产品矩阵^[38]

① 星云联邦数据网络（数据）：通过API提供服务，隐私保护的大数据安全连接平台，以联邦学习和区块链作为基础设施，拼接多方数据源，建立企业间数据合作的安全桥梁，实现企业效能和数据价值的最大化。

② 星云联邦计算平台（框架）

FATE联邦学习软件框架，由多个主要功能模块构成：联邦算法仓库、联邦训练服务、联邦推理服务、可视化面板。企业可以轻松的通过可视化面板直接对各类联邦算法模型进行调用与实验，可大幅降低联邦学习的使用门槛。

³⁸ 来源：星云 Cluster 官网
<https://www.clustarai.com/productService/guardianDock/privacyDataSystem>



星云FATE企业版，为基于数据隐私保护的安全建模过程提供丰富的可视化呈现，为终端用户可视化和度量模型训练的全过程，支持模型训练过程全流程的跟踪、统计和监控等，帮助模型开发人员快速搭建联邦学习任务，可根据客户需求深度定制开发。



来源：星云Cluster官网

图 42 星云FATE企业版联邦架构层

③ 星云隐私计算一体机（算力）

针对数据使用方和数据提供方提供不同产品方案：一体机完美融合CPU/GPU/FPGA服务器、FATE和FDN，开箱即用，大大降低了企业使用联邦学习的门槛；密态计算效率提升400%、降低延迟300%、降低功耗70%，强大算力推动各方数据协作，实现数据资产变现。

(6) 光之树科技——天机、云间

光之树科技旗下有天机可信计算框架和云间联邦学习平台两个隐私计算产品，提供从共享模型训练即“云间”联邦学习到基于芯片TEE技术的“天机”机密计算在内的全流程、多场景安全多方计算框架，保护数据资产权益，安全发挥数据价值。

① 天机可信计算框架

天机可信计算框架于2019年8月发布。它是一个基于芯片中的可信执行环境（TEE：Trusted Execution Environment）和其他加密技术的可信计算体系，主要通过将数据从共享到联合计算在硬件创建的可信执行环境中进行的方式，从而做到数据可用不可见，确保了数据隐私、安全和合规。它具有的安全机制可同时保护模型和计算

过程中的数据，可直接运行机器学习级别的高复杂度计算 / 模型，兼容当前主流的大数据和机器学习框架包括xgboost、scikit-learn（支持逻辑回归等算法）、tensorflow等。用户无需二次开发，可快速部署于公有云、私有或线下环境，并兼容主流数据库以及数据服务。它搭配区块链用于数据存证和权限控制，做到数据使用全程可追溯可审计。



图 43 天机可信计算框架总体框架图^[39]

³⁹ 来源：光之树官网<https://www.guangzhishu.com/>

② 云间联邦学习平台

云间联邦学习平台是基于机器学习、深度学习算法和加密协议的安全计算框架。数据无需离开本地，主要通过将模型下发到数据联盟本地服务器训练的模式，以最小的数据交互对模型进行更新和迭代的计算方法，从而达到保证数据安全性的前提下多方联合计算的目的。应用于普惠金融、贸易金融、保险反欺诈、供应链金融等场景。具有以下优势：

- a. 安全性：通过联邦学习特有的算法保证数据不出本地，并通过加密协议确保数据交互的安全性。
- b. 一键式训练和模型部署：拥有自动建模功能，支持多种机器学习和深度学习的联邦学习训练和模型部署。
- c. 可视化：对训练状态和训练效果进行全方位监控。
- d. 快速部署：支持多种数据库的接入，快速进行私有化部署。
- e. 场景多样性：支持多种场景，包括横向和纵向学习。

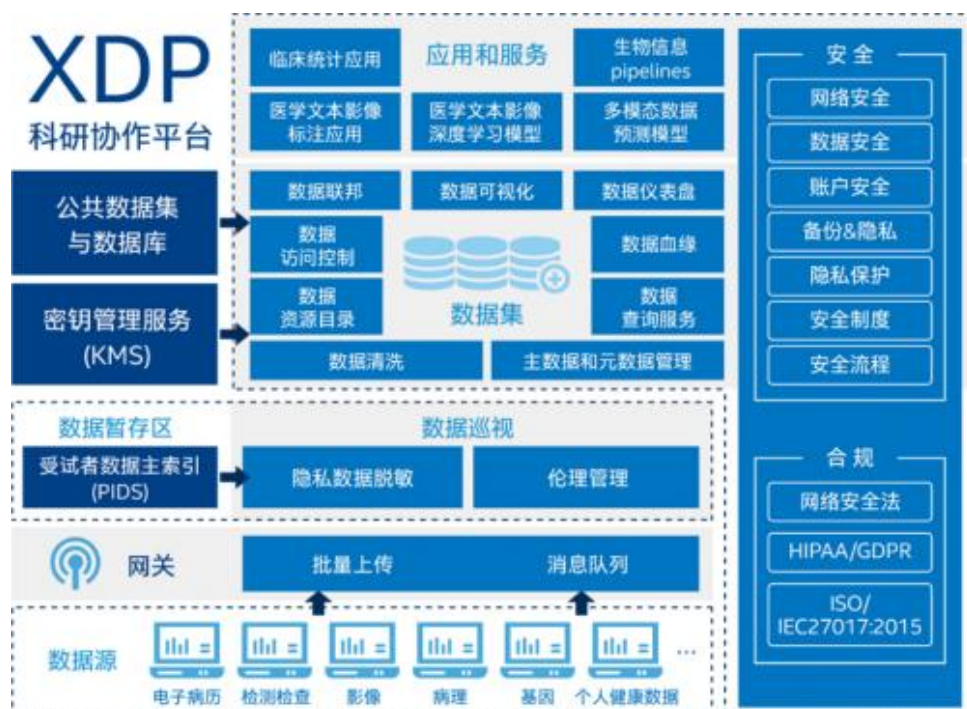
(7) 翼方健数——翼数坊XDP

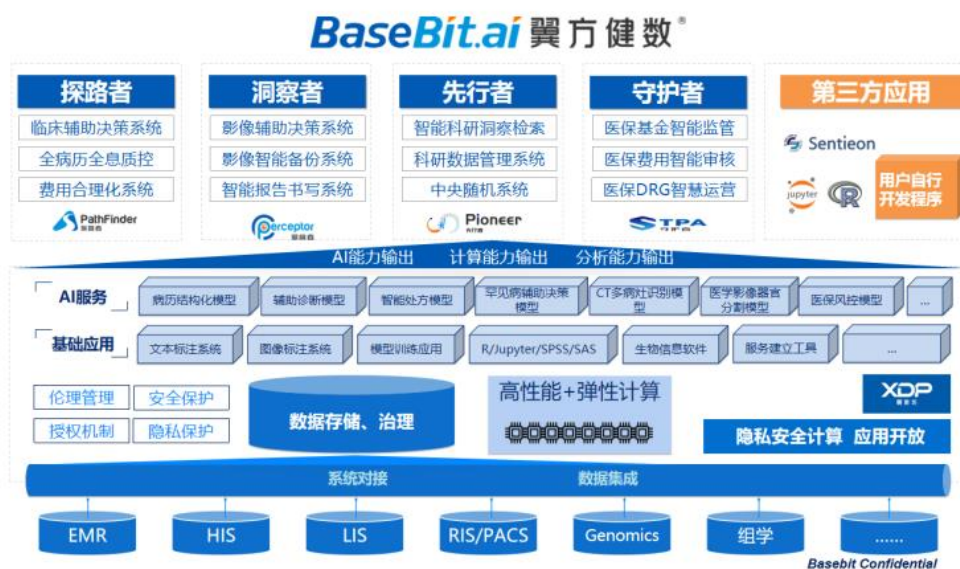
翼方健数通过多方安全计算MPC/同态加密、联邦学习、安全沙箱计算/TEE等前沿技术，实现数据“可用而不可见”，提出“数据和计算互联网”（IoDC）的概念并付诸实践。在技术运用层面，翼方健数自主研发的DaaS服务，可以对多组学数据、表型数据、临床数据进行数据治理和清洗，达到数据可用的状态，从而实现不分享原始数据、数据在平台内授权使用、通过计算来分享数据的价值这一目的。

2019年4月13日，医疗数据隐私计算平台XDP 翼数坊 v1.0 发布。翼数坊XDP利用隐私安全计算技术，实现合理的、授权下的数据价值共享，创造数据流通性，降低数据科学的门槛。翼数坊XDP平台的整体设计从最底层开始，完全基于隐私计算的原理和应用。采用了一系列新型技术，包括多方安全计算、同态加密、联邦学习、可信执行环境、零知识验证等，具有开放、安全、整合、高效、智能五大性能。

XDP平台可基于智能合约技术追溯源数据集，建立“数据血缘”。此外，XDP构筑出的封闭的数据存储和计算环境，将从各医疗机构采集到的数据进行清洗、脱敏、归一，形成DaaS 数据集后进行加密，杜绝数据的泄露。形成的数据权限管理系统，可以确保平台用户所有者授权后才能使用数据，数据所有者的权益也可以得到保障。

平台数据仅限于在平台内使用，即使被授权的数据也不能离开平台，从而进一步保护数据所有者的权益。XDP平台上可以关联、集成并融合各个医疗机构、检验检查以及健康数据；数据应用方面，XDP平台拥有分层可扩展的技术架构，能够实现高密度存储、快速访问和迅速分析计算，并且支持多种人工智能模型的建立，从而多角度直观分析和展示数据。





隐私安全计算技术 Privacy Computing Technology, PCT

BaseBit.ai 翼方健数®

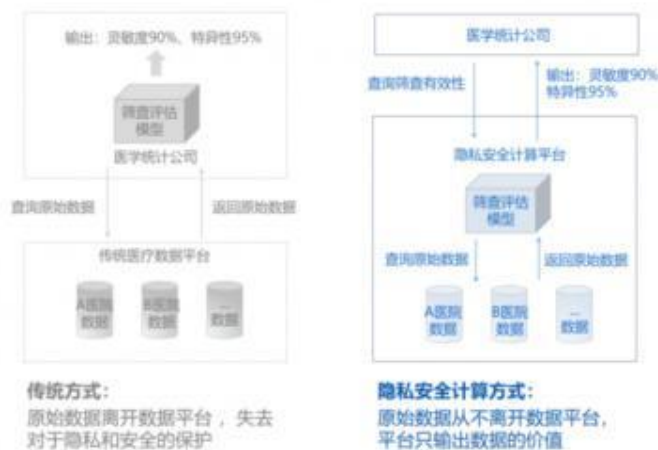


图 44 翼数坊XDP平台总体架构 [40]

⁴⁰来源：翼方健数官网<https://www.basebit.me/>

(三) 联邦学习行业应用现状

联邦学习目前在各行业各领域都开始了广泛的落地探索，获得了较广泛关注。例如，国际性科技期刊 Nature《自然》近期曾发表有几篇关于联邦学习在医疗领域应用的文章，如下文表 8所示。

表 8 《自然》关于联邦学习技术在医疗业应用相关文章

应用场景	论文	简介	来源
精准医疗、 医疗数据 隐私保护	Swarm Learning for Decentralized and Confidential Clinical Machine Learning	引入分散式机器学习方法 Swarm Learning来整合各地医疗数据，它结合了边缘计算、基于区块链的点对点网络和协调，无需中央协调器即可保持机密性。	Nature, no. 7862 (2021): 265-270
医疗成像 及潜在攻 击向量和 未来	Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging	为了促进旨在改善患者护理的大型数据集科研并保护患者隐私，必须实施技术方案以同时满足数据保护和利用的需求。该文概述了当前和下一代联合、安全和隐私保护人工智能的方法，重点是医学成像应用，以及医学成像及其他领域的潜在攻击向量和未来前景。	Nature Machine Intelligence, no. 6 (2020): 305-311
医疗数据 集分析； 医疗用药 诊断；精 准 / 个 性化医疗	Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations Without Sharing Patient Data	表明通过多个数据私有机构合作而增加的数据访问可以更多地有益于训练模型质量。联邦学习的临床采用有望对精准/个性化医学产生催化影响。	Scientific reports, no. 1 (2020): 12598
数字健康	The Future of Digital Health with Federated Learning	如果无法获得足够的数据，机器学习将无法充分发挥其潜力，并最终无法从研究过渡到临床实践。本本文考虑了导致该问题的关键因素，探讨了联邦学习如何为数字健康的未来提供解决方案，并强调需要解决的挑战和注意事项。	NPJ DIGITAL MEDICINE, no. 1.0 (2020): 119

本部分通过新闻事件分析挖掘和搜索系统NewsMiner数据库中的新闻数据，分析得到2016-2020年度已经开始应用联邦学习技术的主要行业和公司，主要应用动态如表 9所示。新闻数据显示，联邦学习技术的行业应用最早出现在2018年，主要应用于包括IT科技、安全防护、金融、智慧城市、医疗健康、智慧零售、电信、教育等领域。

1. 在IT 行业应用

表 9 2016-2020年度联邦学习技术在IT行业应用动态

IT行业应用场景	标题	年-月	来源
用户数据保护	腾讯云发布数据安全解决方案数盾	2018-05	腾讯
隐私数据安全流转	ARPA测试网1.0 版本ASTRAEA正式发布	2019-03	金色财经
可扩展分布式数据协作	趣链科技自主研发BitXMesh正式发布	2019-05	太平洋电脑
联合学习、联合计算、数据共享、模型训练	光之树发布天机可信计算框架和云间联邦学习平台	2019-08	搜狐
跨行业数据融合、隐私保护	富数科技结合联邦学习和安全多方计算技术推出了富数安全计算平台	2019-08	凤凰网
面向产业应用的工具组件	百度发布3项深度学习前沿技术工具组件：联邦学习PaddleFL、图神经网络PGL和多任务学习PALM 等	2019-11	钱江晚报
提出知识联邦框架	同盾科技人工智能研究院深度学习实验室发布成果：“面向联邦学习的加密神经网络”	2019-09	极客网
扩大光大联邦学习生态圈	光大科技加入FATE联邦学习社区技术指导委员会（TSC）并贡献关键算法源码	2020-01	新华网
数据脱敏及去标识化、加密算法支持、DMZ区建设	同盾科技联邦学习技术加持 让数据“可用不可见”	2020-03	网易
大数据安全	平安科技联邦智能平台“蜂巢”落地	2020-09	搜狐

来源：根据公开资料整理

2. 在电信业应用

表 10 2016-2020年度联邦学习技术在电信行业应用动态

电信业应用场景	标题	年-月	来源
车联网通信	华为数字算法实验室利用联邦学习原理解解决车联网中可靠低延迟通信的联合功率和资源分配问题	2018-07	arXiv.org
智能手机	谷歌发布全球首个移动端分布式机器学习系统，数千万手机同步训练	2019-02	亿欧
联邦节点管理、边缘节点管理、联邦实例运行	华为NAIE联邦学习服务助力华为CloudMSE基于业务感知 (Service Awareness, SA) 技术的业务管理	2019-09	知乎
数据采集、模型训练、推理判断及智能预测	中国移动在3GPP标准引入基于联邦学习的分布式智能架构	2020-07	通信世界
识别业务流量后的带宽控制、阻塞控制、业务保障，用户信用评估、用户满意度提升	华为CloudMSE的业务感知 (Service Awareness, SA) 技术	2020-10	知乎

来源：根据公开资料整理

3. 在金融业应用

表 11 2016-2020年度联邦学习技术在金融业应用动态

金融业应用场景	标题	年-月	来源
金融风险管理	建设银行创新合作伙伴揭晓 京东数科、科大讯飞、同盾科技等企业入选	2018-06	CSDN
数据安全、隐私保护	蚂蚁金服推出“摩斯MORSE”多方安全计算平台	2018-08	CSDN
小微信贷	微众银行开源FATE	2019-02	新华网
深度联合建信用模型、客服、侦测欺诈	同盾科技与招联金融共建AI创新实验室 联邦学习为主攻方向之一	2019-06	搜狐
高性能分布式异构计算技术、软硬件解决方案	星云和微众达成合作，推动AI新技术联邦学习的发展	2019-08	科学中国
提升金融服务质量、安全深入地挖掘数据价值	微众银行和腾讯云合作升级 联邦学习携手神盾沙箱共建行业标杆	2019-09	搜狐
数据价值共享、加速金融行业转型进化	AI的最后一公里 英特尔助力平安科技联邦学习落地	2019-09	新浪
多方联合建模	蚂蚁金服基于 MPC 的共享学习	2019-09	ITPUB
支持多方纵向联邦建模、支持spark引擎、支持FATEServing服务治理、支持secureboost在线预测、支持公有云和私有云部署和使用	微众银行发布FATE v1.1，联合VMware中国研发开放创新中心云原生实验室的团队发布KubeFATE项目。FATEBoard：简单高效，联邦学习建模过程可视化	2019-11	贤集网
打造大规模AI协作通用方案	微众银行与蒙特利尔学习算法研究所合作打造安全金融AI实践	2019-12	腾讯
智能化信用卡	江苏银行与腾讯安全举行联邦学习线上发布会，将联合共建“智能化信用卡管理联合实验室”，围绕联邦学习开展合作	2020-04	CSDN
金融数据保密、信贷业务综合评估、控制企业技术升级成本	编织联邦学习的产业路径，腾讯向金融智能化的更远处进发	2020-04	搜狐
金融产品管理、营销、安全风险、客户服务、运营管理	百度金融安全计算平台（度信）建设与实际应用	2020-06	腾讯安全
帮助银行解决数字化转型的风险	腾讯安全天御凭借其在信贷风控场景的落地实践，荣获首个CCF-GAIR“联邦学习应用奖”	2020-08	搜狐
反诈骗技术、普惠金融	反诈骗、管控金融风险，腾讯安全发力联邦学习技术	2020-09	新浪
金融服务、风险识别能力、数字营销	京东数科自研联邦学习平台Fedlearn，助力数据安全保护并大幅提升学习效率	2020-10	机器之心
电商营销、广告投放、个性化内容推荐、广告推荐	字节跳动破局联邦学习：开源Fedlearner框架，广告投放增效209%	2020-10	CSDN

4. 在医疗业应用

表 12 2016-2020年度联邦学习技术在医疗业应用动态

医疗业应用场景	标题	年-月	来源
解决信息孤岛，提供数据安全和授权使用机制	医疗数据隐私计算平台 XDP 翼数坊 v1.0 全球首发	2019-04	搜狐
医疗成像	英伟达在MICCAI 2019上发布首个面向医学影像的隐私保护型联邦学习系统	2019-10	摩尔芯闻
医疗服务患者数据保护	英伟达推出了NVIDIA Clara联邦学习	2019-12	极客公园
生物医药、健康管理、养老旅游、医疗设备、健康保险、保健食品等	Hitacea（医图亚）打造成为基于区块链+联邦学习等新兴技术的亚洲首家全链条大健康科技产业平台	2020-04	科学中国
疾病预测	腾讯天行实验室联合微众银行研发医疗联邦学习 AI利器让脑卒中预测准确率达80%	2020-04	CSDN
医疗诊断	英特尔和宾夕法尼亚大学佩雷尔曼医学院组建医疗联盟研发用以识别脑肿瘤的人工智能模型	2020-05	中电网
AI影像辅助诊断、高精度疾病检测、多维分析以及3D术前规划与模拟	商汤科技SenseCare®智慧诊疗平台推出包含胸部CT、胸部X线、心脏冠脉、病理、骨肿瘤等多款产品解决方案	2020-07	趣味科技
保护用户隐私建模、医保基金控费、个人与机构拒付识别、医学影像辅助诊断、医院运营、临床医疗、健康管理、科研教学	腾讯医疗健康携手微众银行成立联合实验室	2020-08	TechWeb
医学统计分析、临床试验模拟、药物研发	中科院上海药物所联合华为云发布基于ModelArts平台的药物联邦学习服务	2020-09	飞象网
药物隐私数据保护 药物研发	同济大学与微众银行AI团队协同提出了一种基于联邦学习的协同药物定量构效原型系统FL-QSAR	2020-12	科学中国
临床验证评估、医学影像辅助诊断	德国癌症研究中心、伦敦国王学院、麻省总医院、NVIDIA、斯坦福大学和范德堡大学推出MONAI（Medical Open Network for AI）	2020-12	电子发烧友

来源：根据公开资料整理

5. 在其他行业应用

表 13 2016-2020年度联邦学习技术在其他行业应用动态

行业	应用场景	标题	年-月	来源
安全防护	城市管理、公安、社区安防	微众银行与特斯联在北京宣布成立“AIoT联合实验室”	2019-12	贤集网
	大数据安全与隐私保护	华控清交CEO张旭东：数据“可用不可见”和“规定用途用量”，让数据真正成为生产要素	2020-06	搜狐
教育	教育客户广告跑量、课程客户获课续费	字节跳动与教育行业结合，基于 Fedlearner，提升客户的续课率	2020-10	CSDN
人工智能	数据安全和隐私	启智平台发布联邦学习开源数据协作项目 OpenI 纵横	2019-06	开源中国
	降低模型训练成本，提升效率，加快AI应用落地	星云Clustar发布星云AIOS系统	2019-08	星云Clustar
	资源管理和调度等深度优化与加速	星云Clustar和AMAX联合打造的星云智能一体机	2019-09	星云Clustar
	数据安全、数据跨区域通信	星云AIOS人工智能操作系统及联邦学习一体机亮相“英伟达GTC2019”商展览区	2019-12	网易
	人脸识别	京东数科公开联邦学习战略全布局-人脸识别已落地	2020-06	CSDN
	数据中心	星云Clustar与赛灵思公司达成深度合作推出联邦学习加速卡	2020-08	金融界

行业	应用场景	标题	年-月	来源
智慧城市	智慧城市政务、安全、交通、医疗、物流，跨部门、跨领域、跨区域的即时数据处理和数据融合	京东城市基于城市计算和联邦学习技术打造的产品“数字网关”	2019-10	技术前线
	公共安全、智能交通、智能能源	京东城市发布了城市操作系统升级版本“智能城市操作系统2.0”	2019-12	链财经
	重大灾难中的人群疏散；零售、物流业的仓库选址	微众银行AI团队可视化再获新里程碑，两篇论文获EuroVis 2020收录	2020-03	CSDN
	城市交通监测	星云Clustar打造智慧城市领域的数据集CityNet	2020-09	腾讯
	信用城市、市域治理现代化、智能商业等	京东数科联邦数字网关、区块链技术获工信部网络安全应用试点示范项目殊荣	2020-12	央广网
智慧零售	居民消费	苏宁控股与科大讯飞联合推进数字经济发展，提高AI普惠能力	2020-11	新浪
	智慧零售、风险评估和满意度预测	天津移动打造基于“联邦学习+区块链”的多方安全计算引擎系统“珍珠”	2020-12	C114通信网
自动驾驶	共享数据、云计算	英伟达发布了用于自动驾驶和机器人的软件定义平台——NVIDIA DRIVE AGX Orin	2019-12	镁客网

来源：根据公开资料整理

四、联邦学习发展趋势

（一）研究趋势

根据关键词，从AMiner数据库中查找出联邦学习相关论文，其中包含论文所在领域的分支术语和年份，统计含有这些术语的论文数量，给出论文量排名前十的术语，再统计这些术语的起止年份，划分时间窗格，生成大数据智能的发展趋势河流图，如图45所示。图中的每个色带表示一个分支术语，其宽度表示该术语在当年的热度，与当年该分支的论文数量呈正相关；各分支在每一年份中按照其热度进行排序，越热的技术主题则其位置排在越上方。

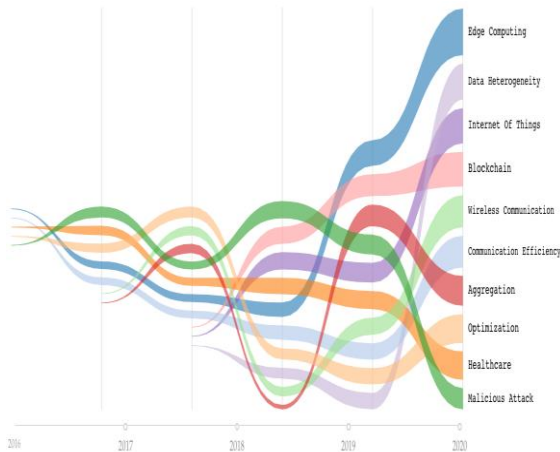


图 45 联邦学习技术发展趋势

由图 45 可见，自从2016年以来，除了 Aggregation（聚合）的研究热度波动较大、Malicious Attack（恶意攻击）研究热点明显下降之外，其余相关研究基本都呈现平稳上升的发展趋势。基于边缘计算、数据异质性的联邦学习研究以及在物联网应用方面的研究热度在2019年左右上升明显，并且之后一直居于领先地位。2020年研究热度前十的主题是与算法模型或安全隐私技术相关，依次分别是：Edge Computing（边缘计算）、Data Heterogeneity（数据异质性）、Internet Of Things（物联网）、Blockchain（区块链）、Wireless Communication（无线通信）、Communication Efficiency（沟通效率）、Aggregation（聚合）、Optimization（优化）、Healthcare（医疗保健）、Malicious Attack（恶意攻击）。



（二）技术成熟度

技术成熟度指单项技术或技术系统在研发过程中所达到的一般性可用程度^[41]。研究机构Gartner发布的技术成熟度曲线（Hype Cycle）因模型较成熟，已被广泛用来评估新科技的可见度，目前已成为是科技产业界技术预测的风向标。

基于Gartner 近年发布的相关技术成熟度曲线，本报告发现，联邦学习（Federated Machine Learning）于2019年首次出现在Gartner数据科学与机器学习技术成熟度曲线（Hype Cycle for Data Science and Machine Learning）之中，并且被视为“在分布环境下的训练机器学习算法的重要创新”^[42]。这表明联邦学习技术应用趋势发展较快，自诞生后仅用了三年时间就吸引了投资者、企业家和消费者的关注，也吸引到Gartner 对该技术应用影响的研究。

此后两年，联邦学习相继出现在其他三个Gartner的技术成熟度曲线里面，分别是2020年发布的数据科学与机器学习技术成熟度曲线、以及2021年的隐私技术成熟度曲线（Hype Cycle for Privacy）与公用事业行业 IT 技术成熟度曲线（Hype Cycle for Utility Industry IT），详细情况如表 14所示。

由表 14可见，在这些技术成熟度曲线之中，联邦学习都是处于“创新触发期”（Innovation Trigger），效益评级均为“高”，都属于“新兴”技术，到达生产高峰期（the Plateau of Productivity）的时间都预计为5~10年，且市场渗透率（Market Penetration）都低于1%。

⁴¹ 朱毅麟.技术成熟度对航天器研制进度的影响[J]. 航天器工程, 2009, 18(2): 9.

⁴² Hype Cycle for Data Science and Machine Learning, 2019, ARCHIVEDPublished 6 August 2019 - ID G00369766 -By Shubhangi Vashisth, Alexander Linden, et al, <https://www.gartner.com/document/3955984?ref=solrAll&refval=295245018>

表 14 联邦学习相关Gartner 技术成熟度曲线

Hype Cycle	Time	Benefit Rating	Maturity	Time to Plateau	Market Penetration
Hype Cycle for Data Science and Machine Learning, 2019 ^[43]	Innovation Trigger	High	Emerging	5~10年	Less than 1% of target audience
Hype Cycle for Data Science and Machine Learning, 2020 ^[44]	Innovation Trigger	High	Emerging	5~10年	Less than 1% of target audience
Hype Cycle for Privacy, 2021 ^[45]	Innovation Trigger-	High	Emerging	5~10年	Less than 1% of target audience
Hype Cycle for Utility Industry IT, 2021 ^[46]	Innovation Trigger	High	Emerging	5~10年	Less than 1% of target audience

来源：Gartner 公司

值得关注的是，在2019年“数据科学与机器学习”技术成熟度曲线之中，由于首轮风投刚开始以及边缘数据收集问题等因素影响，当年Gartner预计联邦学习技术按照当时进行中的研究进展“不太可能在5到10年内”达到“生产高峰期”（the Plateau of Productivity）。随着隐私法规的激增、对数据隐私保护的需求增加，以及集中收集和存储大数据难度的增加等多个驱动因素影响，联邦学习被采用的范围和程度逐年增加。在2020年之后的技术成熟度曲线之中，虽然联邦学习技术仍然都处于“创新触发期”（Innovation Trigger），但相比2019年，联邦学习在2020年距离“期望膨胀期”（Peak of Inflated Expectations）又更进一步，

已经度过了公司初创和第一轮风投的发展阶段，正处于“第一代产品期、价格高、大量客户化定制”（First-generation products, high price, lots of customization needed）的阶段^[47]。

而在隐私技术成熟度曲线（Hype Cycle for Privacy）与公用事业行业IT技术成熟度曲线（Hype Cycle for Utility Industry IT）中，联邦学习则是于2021年开始才占有一席之地的。这主要是由于联邦学习的采用在过去一年加速发展，特别是因为它在新冠流行期间已成功用于医疗保健，以及该技术特别适用于例如物联网、网络安全、隐私、数据货币化和数据共享等受监管行业。

⁴⁷ Gartner Hype-Cycle: Everything You Need To Know, <https://www.wowso.me/blog/gartner-hype-cycle>

(三) 市场化与商业化趋势

联邦学习技术在国内外发展快速。有公开资料可查的联邦学习研究或应用单位已超过百家 [1]。联邦学习可以被看成是一种连接联邦成员的大数据资产“连接”工具，具有非常广泛的市场应用价值，适用于医学研究、金融风控、医疗、智慧城市、移动互联网等多个实际场景。一些大型企业也开展了联邦学习技术的战略布局和应用，推出了相关的行业解决方案和项目，这反映出联邦学习的市场需求较热。

随着国内外相关标准和法规的完善和实施，以及解决方案和开源项目的不断迭代，联邦学习技术的未来应用场景将持续增加。未来能否出现大规模联邦学习商业化应用，将主要与网络带宽问题密切相关。这是因为联邦学习需要非常大量的中间结果交互，在某些场景下需要超过100Mb/s的网络带宽才能在有效的时间内完成建模，而某些银行仅支持2Mb/s的网络带宽，在样本量较大的情况下，这可能导致建模时间长达数月，无法满足业务的需求。5G技术的发展和信息高速公路的建设，将会促进联邦学习大规模商业化应用的实现。

此外，联邦学习未来市场与商业化的实际落地将出现更多的异构场景下的应用。应用场景可分为同构场景和异构场景。同构场景指的是两个企业属于相同或相近的领域，所拥有的数据性质相似、特征相近，但是样本不同。如在银行和金融机构间

的合作，双方拥有的不同的用户样本，但是样本属性同质，这种场景下使用横向联邦学习，可达到将双方样本放到一起的建模效果。异构场景指的是两个企业分属不同的领域，所拥有的数据性质不同、特征不同，但是有重叠的样本ID。比如银行与互联网公司之间的合作，双方有重叠的用户ID，但是企业间各自拥有用户不同的特征，如银行有用户的收入和交易行为，互联网公司有用户的社交或出行行为，这种场景下使用纵向联邦学习建模，可达到特征增加的建模效果。在当前的联邦学习市场化应用中，同构场景下的探索更为成熟。未来将出现更多的联邦学习在行业垂直领域的应用尤其是异构场景下的应用。

(四) 推行联邦学习的国内外标准

技术标准化建立与实施是联邦学习技术落地应用的重要依据。通过研制和建立联邦学习的国内标准（如团体标准和国家标准）与国际标准（如IEEE企业标准），制定联邦学习的算法框架规范、使用模式和使用规范，可帮助更多行业和海内外不同类别的实体在保证用户隐私和数据安全的情况下，合作共赢、建立更准确的数据模型，同时，也给人工智能在不同产业中的实际落地提供可行性依据。

截至目前，联邦学习领域已经由企业或行业联盟协会发起并建立了初步的企业级国际标准和国家级团体规范。部分标准信息如表 15所示。

⁴⁸—文读懂联邦学习的前世今生，东科技技术说，2020-11-17，
<https://blog.csdn.net/JDDTechTalk/article/details/109738346>

表 15 联邦学习相关国内外标准

领域	类别	标准名称	发布方	发布时间
人工智能	团体规范标准	《信息技术服务联邦学习参考架构》 ^[49]	中国人工智能开源软件发展联盟 (AIOSS)	2019年6月
	国际标准	IEEE P3652.1《联邦学习架构和应用规范》(Guide for Architectural Framework and Application of Federated Machine Learning)	电气与电子工程师协会 (IEEE) 标准委员会 (SASB)	2021年3月
5G通信	国际标准	NWDAF (Network Data Analytics Function-5G网络AI) 的联邦学习技术标准 ^[50]	3GPP 通过, 由亚信科技与中国移动共同提交	2020年7月
	团体标准	《基于联邦学习的数据流通产品技术要求与测试方法》 ^[51]	中国通信标准化协会	2020年7月
金融	行业标准	《多方安全计算金融应用技术规范》(JR/T 0196-2020) ^[52]	中国人民银行	2020年11月

⁴⁹ 国内首个联邦学习标准正式出台,微众银行AI团队领衔, 2019-07-01, https://www.sohu.com/a/323923758_99974896

⁵⁰ 国内首个联邦学习标准正式出台,微众银行AI团队领衔, 2019-07-01, https://m.sohu.com/a/323923758_99974896

⁵¹ 中国信通院解读“隐私计算系列标准与测试方法” 2021-01-25, https://www.sohu.com/a/446614289_735021

⁵² 央行发布《多方安全计算金融应用技术规范》 确保数据安全, 2020-12-24, <https://www.cebnet.com.cn/20201224/102711761.html>



随着国际与国内联邦学习标准的相继出台，在未来发展中，相关标准的实施与执行将是联邦学习领域的发展重点，影响着该技术作为下一代人工智能协作网络基础的能力。能够有效推行标准化的联邦学习技术规范，不仅有利于来自不同行业、不同业务类别的企业在开展业务或进行合作的过程中合法合规地共同使用数据、保护用户隐私和数据安全，而且有助于建立更为准确的数据模型，进而促进该技术走向成熟化和开启大规模工业化应用。

(五) 建立联邦学习生态

随着国际与国内联邦学习标准的相继出台，未来将有更多行业的更多企业机构加入和布局该技术的

应用。由此需要建立一个联邦学习生态联盟。未来在联邦联盟中，所有成员的数据在合法合规下可以带来真正的价值流动，为自身带来收益，同时各个行业还可以建立各自的联邦数据网络，不同行业的网络间还将有所交甚至连接紧密^[53]，从而促进各自行业良性发展。在良好的联邦学习生态联盟中，联邦学习参与方，不仅可以获得相关的技术支持等服务与产品，快速便捷地完成相关应用的开发部署工作，而且可以在良好的开源环境下，更加高效、准确地自建模型、联合建模、共享模型、共建联邦学习生态。联邦学习生态的建立，需要学术界和产业界的共同推动^[54]，使之将成为参与各方机构之间数据合作的桥梁，挖掘数据背后的真正的知识和价值。

⁵³ 微众银行人工智能部、鹏城实验室、腾讯研究院、中国信通院云大所、平安科技、招商局金融科技、电子商务与电子支付国家工程实验室(中国银联):《联邦学习白皮书V2.0》，深圳，2020年，第28-30页。

⁵⁴ 微众银行首席AI官杨强：建立联邦学习生态需学术和产业界共同推动 [N] TechWeb，2020年11月16. 日 <http://www.myzaker.com/article/5fb1f78a8e9f0945d855fd1d/>



五、结语

针对近年来在工业界和学术界都大火的联邦学习技术，本报告从科研论文、专利、书籍、行业应用、学者地图与画像、技术发展趋势等多个角度，全景展示和分析了联邦学习技术自从2016年被提出以来至2020年的重要进展，并展望了该技术的未来发展方向与前景。

本报告发现，联邦学习研究论文数量和专利申请量都在逐年增多，反映出其研究热度逐年上升。中国在联邦学习领域的论文量全球领先、论文总引用量仅次于美国。高被引论文半数以上被中美两国占据，这两国合作的论文数量也多于其他国家之间。同时，中国的该领域专利申请量也居于全球首位。在国内，广东、浙江和北京的相关专利申请量超过其他省市。这反映出，相比其他国家而言，我国学术界和产业界对联邦学习科研和推广应用更为热衷。从技术研究热点看，较多聚焦于机器学习方法、模型训练、隐私保护等主题。

报告还发现，中国和美国是该领域学者主要聚集的国家。从学者学术水平来看，学者数量前十国家的学者平均H指数差距并不显著。基于AMiner推荐的联邦学习必读论文和学者库，本报告还详细展示了联邦学习、激励机制、算法安全及隐私保护等不同研究方向上的代表学者学术画像。

此外，本报告还梳理了市面上主要的联邦学习系统框架，以及在IT科技、安全防护、金融、智慧城市、医疗健康、智慧零售、电信、教育等行业落地应用场景，并探讨了该技术的市场化与商业化趋势，以及推行的国内外标准与建立联邦学习生态等问题。

联邦学习从技术维度上解决了人工智能发展过程中的安全问题，被学术界和产业界寄予厚望。中国已经成为联邦学习技术的深度参与方，国内企业和科研机构积极参与联邦学习的技术研发和应用，以及标准制定。未来，随着人工智能技术和应用的不断升级，联邦学习的技术研发和落地应用还将进一步扩大和深入。

附录一 联邦学习领域顶级国际期刊会议列表

以《CCF推荐国际学术期刊和会议目录》为数据来源，并征求领域顾问专家意见而确定。

序号	期刊/会议名称	简称
1	ACM Conference on Computer and Communications Security	CCS
2	The Network and Distributed System Security Symposium	NDSS
3	USENIX Security Symposium	USENIX Security
4	IEEE Symposium on Security and Privacy	SP
5	International Conference on Learning Representations	ICLR
6	Neural Information Processing Systems	NIPS
7	Machine Learning and Systems	MLSys
8	Distributed AI	DAI
9	IEEE International Conference on Distributed Computing Systems	ICDCS
10	International Conference on Machine Learning	ICML
11	AAAI Conference on Artificial Intelligence	AAAI
12	International Joint Conference on Artificial Intelligence	IJCAI
13	ACM Transactions on Intelligent Systems and Technology	— —
14	IEEE International Conference on Big Data	— —
15	Nature	NATURE
16	IEEE Internet of Things Journal	— —
17	IEEE Transactions on Industrial Informatics	IINF
18	IEEE Transactions on Parallel and Distributed Systems	TPDS
19	IEEE Transactions on Big Data	— —
20	Future Generation Computer Systems	— —
21	Procedia Computer Science	— —
22	Journal of Network and Computer Applications	— —
23	Computer Networks	— —
24	Computers & Security	— —
25	Network and System Security	NSS
26	IEEE International Conference on Communications	ICC
27	International Conference on Machine Learning and Intelligent Communications	MLICOM



附录二 《联邦学习架构和应用规范》简介

IEEE P3652.1《联邦学习架构和应用规范》(Guide for Architectural Framework and Application of Federated Machine Learning) 相关信息如下。

1. 目标 (Purpose)

本规范的目的是为AI工业应用提供可行的解决方案，即集体使用数据而无需直接交换数据。在隐私和数据保护问题变得越来越重要的情况下，本规范有望促进协作，将促进并允许使用分布式数据源来开发AI，而不会违反法规或道德考量。(The purpose of this guide is to provide a feasible solution for industrial application of AI -- using data collectively without exchanging data directly. This guide is expected to promote and facilitate collaborations where privacy and data protection issues have become increasingly important. This guide will promote and enable to use of distributed data sources for the purpose of developing AI without violating regulations or ethical considerations.)

2. 范围 (Scope)

联合学习定义了一种机器学习框架，该框架允许从分布在数据所有者之间的数据构建一个集体模型。本规范提供了跨组织的数据使用和模型构建的蓝图，同时满足了所适用的隐私，安全和法规要求。它定义

了联合机器学习的体系结构框架和应用程序准则，包括：1) 联合学习的描述和定义，2) 联合学习的类型和每种类型适用的应用场景，3) 联合学习的性能评估，以及 4) 相关法规要求。(Federated learning defines a machine learning framework that allows a collective model to be constructed from data that is distributed across data owners. This guide provides a blueprint for data usage and model building across organizations while meeting applicable privacy, security and regulatory requirements. It defines the architectural framework and application guidelines for federated machine learning, including: 1) description and definition of federated learning, 2) the types of federated learning and the application scenarios to which each type applies, 3) performance evaluation of federated learning, and 4) associated regulatory requirements.)

参考文献

- [1] 《中华人民共和国民法总则》，中华人民共和国中央人民政府，http://www.gov.cn/xinwen/2017-03/18/content_5178585.htm#1
- [2] 《中华人民共和国网络安全法》，中共中央网络安全和信息化委员会办公室、中华人民共和国国家互联网信息办公室，http://www.cac.gov.cn/2016-11/07/c_1119867116.htm
- [3] 2019 年我国数字经济规模达 35.8 万亿元 [N]，2020-11-16，人民网 - 强国论坛，<http://www.people.com.cn/big5/n1/2020/1116/c32306-31932847.html>
- [4] Ammad-Ud-Din, M., Ivannikova, E., Khan, S. A., Oyomno, W., Fu, Q., Tan, K. E., & Flanagan, A. (2019). Federated collaborative filtering for privacy-preserving personalized recommendation system. arXiv preprint arXiv:1901.09888.
- [5] B Gu, Z Dang, X Li, H Huang. Federated Doubly Stochastic Kernel Learning for Vertically Partitioned Data[J].2020.
- [6] Chen, M., Mathews, R., Ouyang, T., & Beaufays, F. (2019). Federated learning of out-of-vocabulary words. arXiv preprint arXiv:1903.10635.
- [7] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. Jama, 2016, 316(22): 2402-2410.
- [8] IEEE 联邦学习国际标准 (IEEE 3652.1-2020) ,
https://www.techstreet.com/ieee/standards/ieee-p3652-1?gateway_code=ieee&vendor_id=7453&product_id=2183131
- [9] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2019). Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977.
- [10] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency[J]. NIPS Workshop on Private Multi-Party Machine Learning, 2016.

参考文献

- [11] Li J , Huang H . Faster Secure Data Mining via Distributed Homomorphic Encryption[J]. 2020.
- [12] Liu Y, Chen T, Yang Q. Secure Federated Transfer Learning Framework[J]. IEEE Intelligent Systems, vol. 35, no. 4, pp. 70-82, 1 July-Aug. 2020.
- [13] Liu, D., Miller, T., Sayeed, R., & Mandl, K. D. (2018). Fadl: Federated-autonomous deep learning for distributed electronic health record. arXiv preprint arXiv:1811.11400.
- [14] McMahan H B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[J]. arXiv preprint arXiv:1602.05629, 2016.
- [15] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data[J]. In Artificial Intelligence and Statistics (pp. 1273-1282). PMLR.
- [16] McMahan, H. B., Moore, E., Ramage, D., & y Arcas, B. A. (2016). Federated learning of deep networks using model averaging. arXiv preprint arXiv:1602.05629.
- [17] Pandey, Shashi Raj, et al. A crowdsourcing framework for on-device federated learning. IEEE Transactions on Wireless Communications 19.5 (2020): 3241-3256.
- [18] Waymo. Waymo官网[OL]. [2020-02-17]. <https://waymo.com>
- [19] Xu, Guowen, et al. VerifyNet: Secure and Verifiable Federated Learning , IEEE Transactions on Information Forensics and Security, pp. 911-926, 2020.
- [20] Yang, Q. , Liu, Y. , Chen, T. , & Tong, Y. . (2019). Federated Machine Learning: Concept and Applications. ACM Trans. Intell. Syst. Technol. 10, 2, Article 12, February, 2019. DOI:<https://doi.org/10.1145/3298981>

参考文献

- [21] 百度.Apollo自动驾驶解决方案[OL].[2020-02-17]. <http://apollo.auto/>
- [22] 国内首个联邦学习标准正式出台,微众银行AI团队领衔 [N], 2019-07-01, https://www.sohu.com/a/323923758_99974896
- [23] 获 IEEE 全票通过, 首个联邦学习国际标准将正式推行 [N], 2020-09-29, https://blog.csdn.net/m0_46317295/article/details/108870325
- [24] 京东数科首度公开联邦学习战略全布局 [N], 2020-06-03, 京东数科, <https://www.asmag.com.cn/news/202006/103870.html>
- [25] 联邦学习最新医疗场景发布, 同济大学刘琦教授团队与微众银行杨强教授AI团队合作打破药物数据共享壁垒[N], 机器之心, 2020-12-17, <https://www.jiqizhixin.com/articles/2020-12-17-7>
- [26] 商汤科技SenseCare创“心”升级,探索“联邦学习”入选欧洲计算机视觉国际会议(ECCV) [N], 2020-07-20, 慧聪网, <https://med.hc360.com/26/268213.html>
- [27] 特斯拉.Autopilot系统介绍[OL].[2020-02-17]. <https://www.tesla.cn/autopilot>
- [28] 腾讯天衍实验室联合微众银行研发医疗联邦学习AI利器让脑卒中预测准确率达80% [N], 搜狐, 2020-04-13, https://www.sohu.com/a/387647468_120230267
- [29] 腾讯医疗健康携手微众银行成立联合实验室, 联邦学习破解隐私难题 [N], 2020-08-22, 维科网, <https://www.ofweek.com/medical/2020-08/ART-11106-8450-30454114.html>
- [30] 微众银行人工智能部, 鹏城实验室, 腾讯研究院, 中国信通院云大所, 平安科技, 招商局金融科技, 电子商务与电子支付国家工程实验室(中国银联): 《联邦学习白皮书V2.0》, 深圳, 2020年.

参考文献

- [31] 微众银行首席AI官杨强：建立联邦学习生态需学术和产业界共同推动 [N] TechWeb, 2020-11-16, <http://www.myzaker.com/article/5fb1f78a8e9f0945d855fd1d/>
- [32] 央行发布《多方安全计算金融应用技术规范》 确保数据安全 [N], 2020-12-24, <https://www.cebnet.com.cn/20201224/102711761.html>
- [33] 杨强, 刘洋, 程勇, 康焱, 陈天健:《联邦学习》, 电子工业出版社: 北京, 2020年:8-10.
- [34] 杨强、刘洋、陈天健等:《联邦学习》, 载《中国计算机学会通讯》, 2018年版, 第49-55页。
- [35] 一文读懂联邦学习的前世今生 [N], 东科技技术说, 2020-11-17, <https://blog.csdn.net/JDDTechTalk/article/details/109738346>
- [36] 英特尔联手宾夕法尼亚大学 采用“联邦学习”技术的AI识别脑肿瘤 [N], TechWeb, 2020.05.26, <http://www.techweb.com.cn/ucweb/news/id/2791494>
- [37] 拥抱“新基建” 京东数科成立产业AI中心 [N], 三言财经, 2020-03-19 https://www.sohu.com/a/381337566_100117963
- [38] 这家银行运用“联邦学习”, 为金融与科技融合“上保险” [N]. 财经头条.2020-03-24, <https://t.cj.sina.com.cn/articles/view/5675440730/152485a5a02000sg9m?from=tech>
- [39] 中科院上海药物所蒋华良院士团队联合华为云, 发布AI药物联邦学习服务 [N], 极客公园, 2020-09-30, <http://www.geekpark.net/news/267104>
- [40] 字节跳动在联邦学习领域的探索及实践 [N]. 字节跳动 .2021-01-14, <https://segmentfault.com/a/1190000038984381>



致谢

感谢以下人员为《2021联邦学习全球研究与应用趋势报告》提供数据、分析、技术、建议和专家评论等支持。

杨强 香港科技大学 教授； 微众银行 首席人工智能官	在报告架构创新设计、重要定义和概念核验、技术完整性与专业性审核等多个方面，做出特别重要贡献
康焱 微众银行 高级研究员	在联邦学习知识树、研究热点分析、经典算法、核心数据核查、创新性分析等多个章节，做出特别重要贡献
刘洋 清华大学智能产业研究院 副研究员 /副教授	在报告内容的同行评价和建议方面，做出重要贡献
周柚池 微众银行 高级研究员	在选题策划、联邦学习专利与应用等章节，做出重要贡献
蔡杭 微众银行 高级研究员	在联邦学习专利等章节，做出重要贡献
范涛 微众银行 高级研究员	在联邦学习系统和框架等章节，做出重要贡献
何芸 智谱·AI TIME 事业部总监	在本报告选题策划与撰写流程中，做出重要贡献



版权说明

AMiner研究报告版权为AMiner团队独家所有，拥有唯一著作权。AMiner咨询产品是AMiner团队的研究与统计成果，其性质是供用户内部参考的资料。

AMiner研究报告提供给订阅用户使用，仅限于用户内部使用。未获得AMiner团队授权，任何人和单位不得以任何方式在任何媒体上（包括互联网）公开发布、复制，且不得以任何方式将研究报告的内容提供给其他单位或个人使用。如引用、刊发，需注明出处为“报告名称（AMiner.org）”，且不得对本报告进行有悖原意的删节与修改。

AMiner研究报告是基于AMiner团队及其研究员认可的研究资料，所有资料源自AMiner后台程序对大数据的自动分析得到，本研究报告仅作为参考，AMiner团队不保证所分析得到的准确性和完整性，也不承担任何投资者因使用本产品与服务而产生的任何责任