科技论文写作文献检索

联邦学习中的通信优化

叶茂青　王珺

Friday 12th June, 2020

总览

联邦学习的定义

- 联邦学习的数学模型被 McMahan et al.[1] 定义为

$$\min_w F(w), \text{ where } F(w) := \sum_{k=1}^{m} p_k F_k(w)$$

- 通俗来说，就是多个客户端通过中心服务器的协调合作解决一个机器学习问题

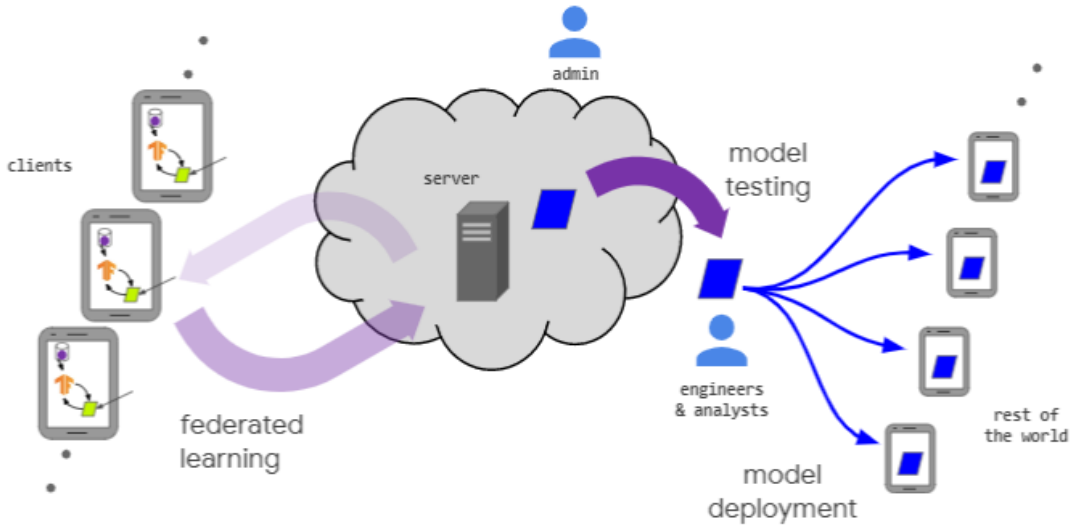


Figure: The lifecycle of an FL-trained model.[2]

联邦学习的特点

- 客户端数据保存在本地，不会上传给服务器

联邦学习的特点

- 客户端数据保存在本地，不会上传给服务器
- 数据不满足 IID(独立同分布) 假设

联邦学习的特点

- 客户端数据保存在本地，不会上传给服务器
- 数据不满足 IID(独立同分布) 假设
- 客户端的处理能力，带宽是有限的，且不一定可靠

联邦学习的分类

- 横向联邦学习
- 纵向联邦学习
- 联邦迁移学习

横向联邦学习

适用于特征重叠多，但用户重叠少的情况，比如不同地区之间的银行、医院，参与的设备数一般较少，且客户端基本可保持稳定

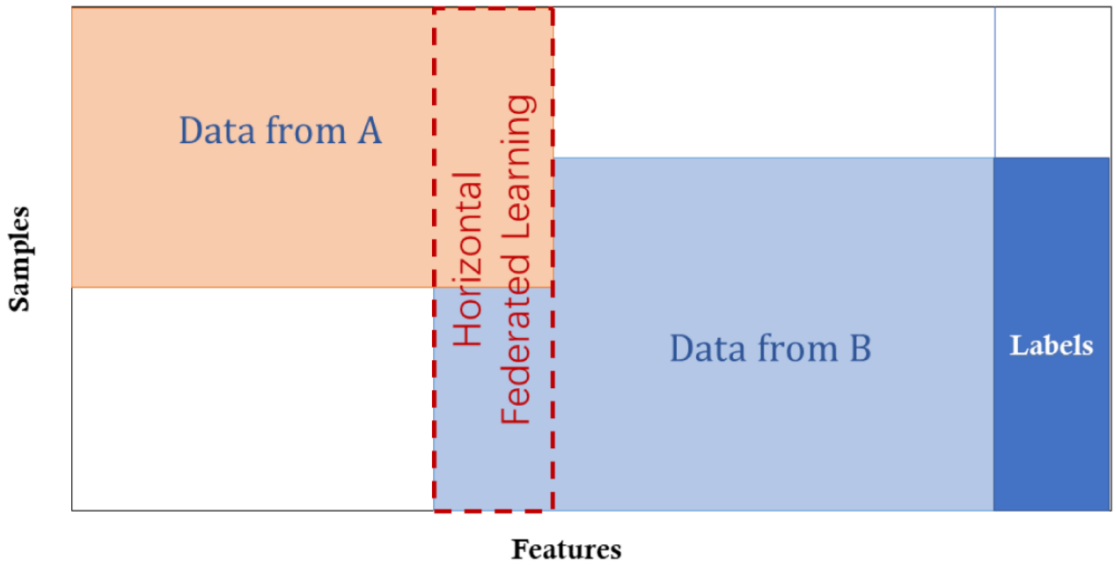


Figure: Horizontal Federated Learning[3]

纵向联邦学习

适用于特征重叠少，但用户重叠多的情况，比如使用同一个 app 的用户，参与的设备数数量级庞大，客户端不可靠，不能保证稳定在线，且有可能存在攻击者

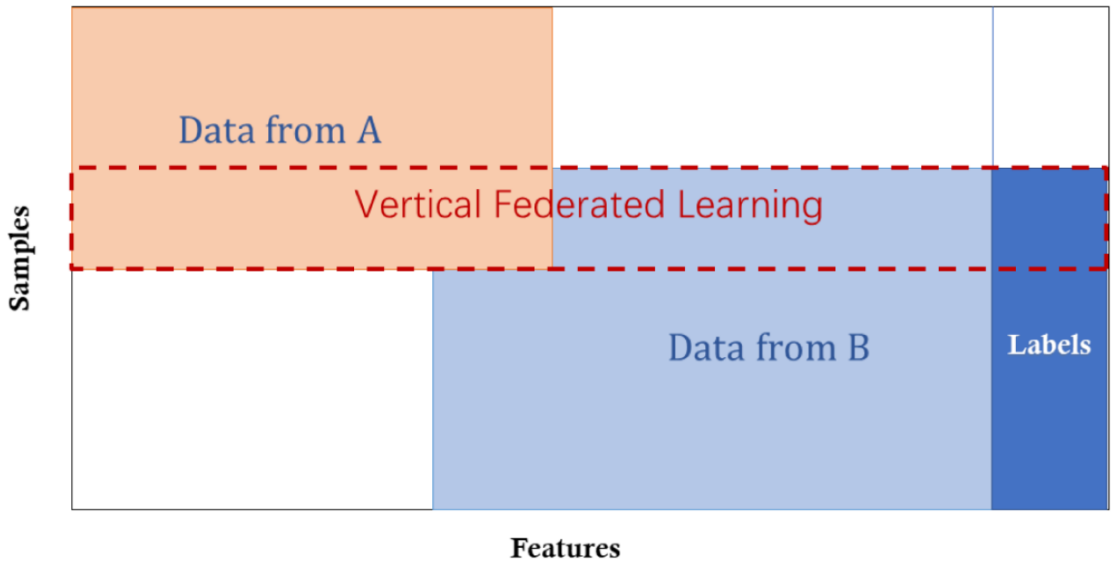


Figure: Vertical Federated Learning[3]

联邦迁移学习

对于特征和用户重叠都较少的状况，比如不同领域的不同公司，需要利用迁移学习来提升模型的效果

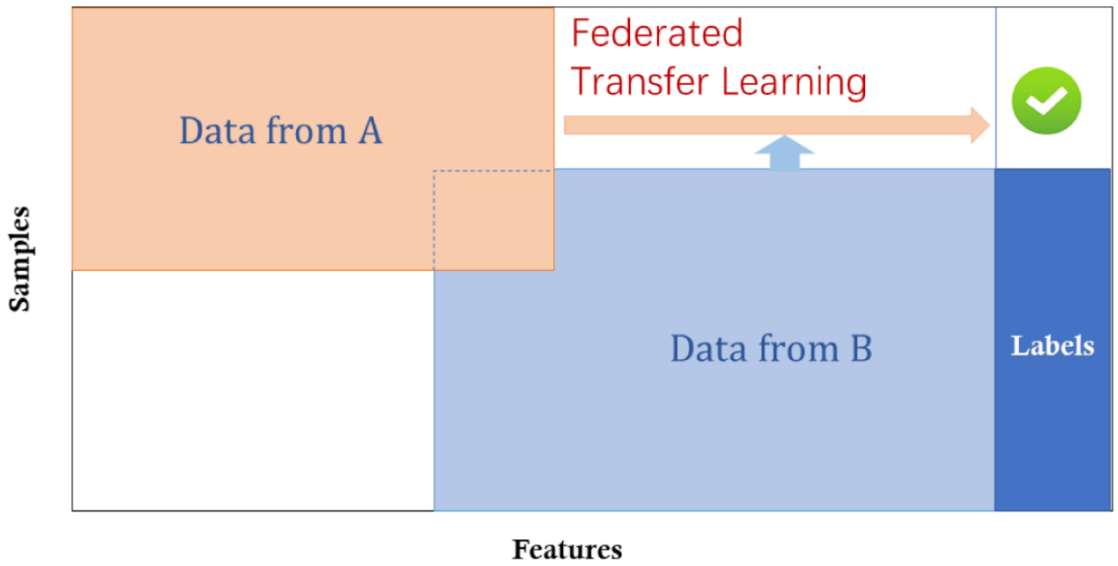


Figure: Federated Transfer Learning[3]

寻找综述论文

federated learning review

找到约 66,100 条结果 (用时0.04秒)

Federated learning: Challenges, methods, and future directions [PDF] arxiv.org

T Li, AK Sahu, A Talwalkar, V Smith - arXiv preprint arXiv:1908.07873, 2019 - arxiv.org

... decentralized training has been demonstrated to be faster than centralized training when operating on networks with low bandwidth or high latency; we defer readers to [47, 67] for a more comprehensive review. Similarly, in federated learning, decentralized algorithms can in ...

☆ 99 被引用次数: 95 相关文章 所有 3 个版本 »

Mitigating sybils in federated learning poisoning [PDF] arxiv.org

C Fung, CJM Yoon, I Beschastnikh - arXiv preprint arXiv:1808.04866, 2018 - arxiv.org

Page 1. Mitigating Sybils in Federated Learning Poisoning ... Existing approaches, such as federated learning, collect the outputs computed by a group of devices at a central aggregator and run iterative algorithms to train a globally shared model ...

☆ 99 被引用次数: 39 相关文章 所有 3 个版本 »

Federated machine learning: Concept and applications [PDF] acm.org

Q Yang, Y Liu, T Chen, Y Tong - ACM Transactions on Intelligent ..., 2019 - dl.acm.org

... This requires security models and analysis to provide meaningful privacy guarantees. In this section, we briefly review and compare different privacy techniques for federated learning. We also identify approaches and potential chal- lenges for preventing indirect leakage ...

☆ 99 被引用次数: 216 相关文章 所有 7 个版本

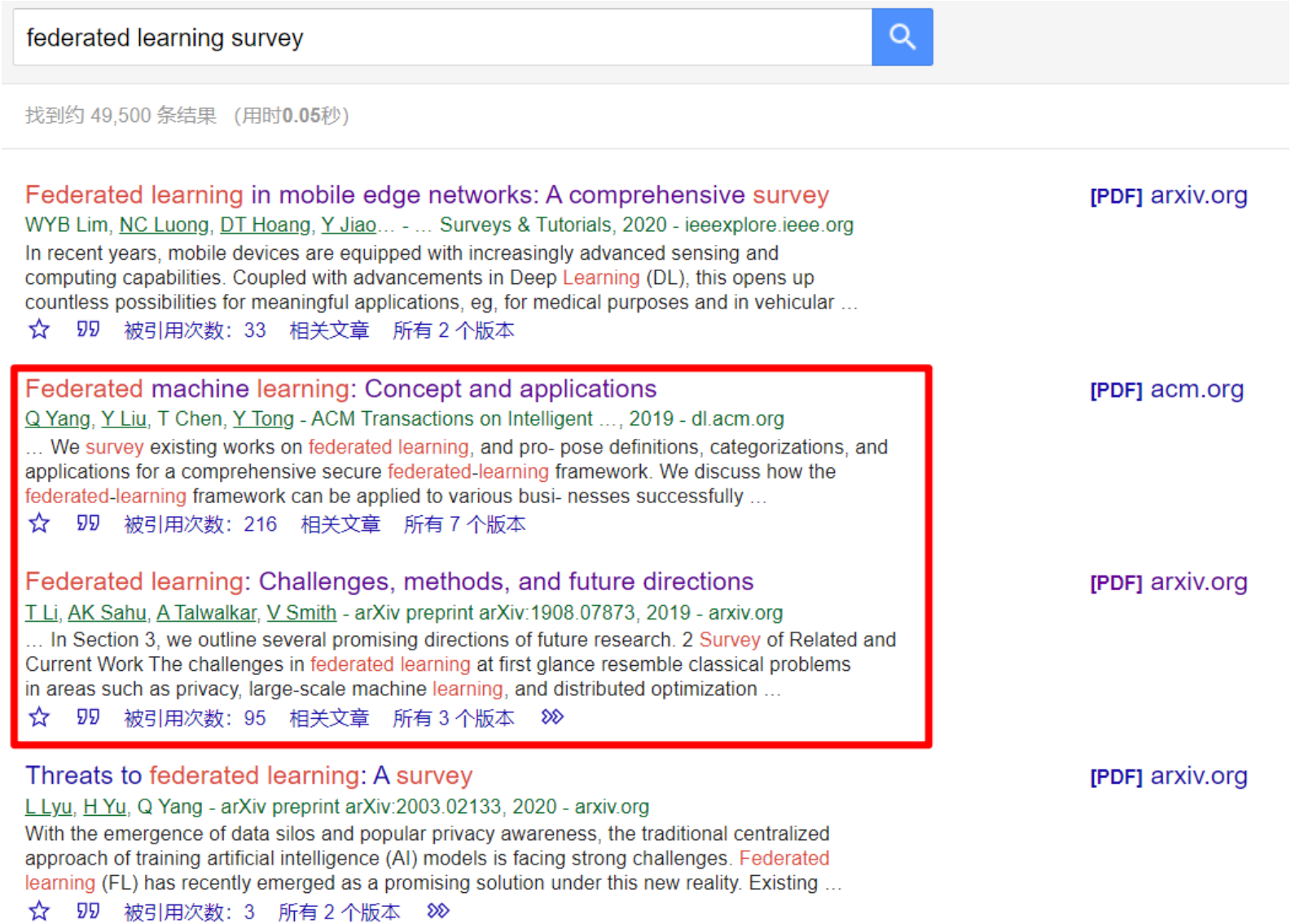
Federated learning in mobile edge networks: A comprehensive survey [PDF] arxiv.org

WYB Lim, NC Luong, DT Hoang, Y Jiao... - ... Surveys & Tutorials, 2020 - ieeexplore.ieee.org

... Recently, in light of increasingly stringent data privacy legislations and growing privacy concerns, the concept of Federated Learning (FL) has been introduced ... Then, we highlight the aforementioned challenges of FL implementation and review existing solutions ...

☆ 99 被引用次数: 33 相关文章 所有 2 个版本

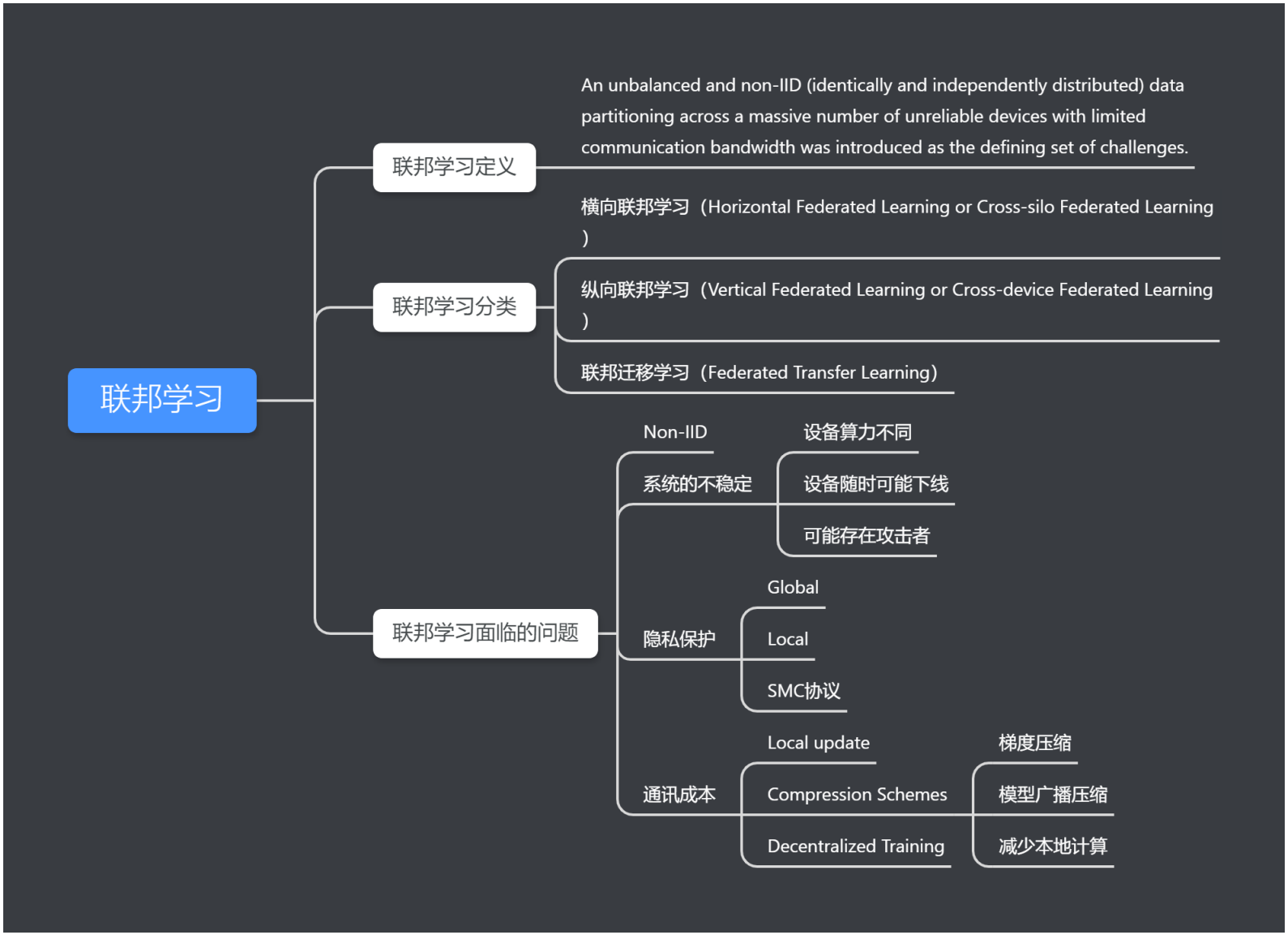
寻找综述论文


federated learning survey
找到约 49,500 条结果 （用时0.05秒）
Federated learning in mobile edge networks: A comprehensive survey
[PDF] arxiv.org
WYB Lim, NC Luong, DT Hoang, Y Jiao... - ... Surveys & Tutorials, 2020 - ieeexplore.ieee.org
In recent years, mobile devices are equipped with increasingly advanced sensing and computing capabilities. Coupled with advancements in Deep Learning (DL), this opens up countless possibilities for meaningful applications, eg, for medical purposes and in vehicular ...
被引用次数：33 相关文章 所有 2 个版本
Federated machine learning: Concept and applications
[PDF] acm.org
Q Yang, Y Liu, T Chen, Y Tong - ACM Transactions on Intelligent ..., 2019 - dl.acm.org
... We survey existing works on federated learning, and pro- pose definitions, categorizations, and applications for a comprehensive secure federated-learning framework. We discuss how the federated-learning framework can be applied to various busi- nesses successfully ...
被引用次数：216 相关文章 所有 7 个版本
Federated learning: Challenges, methods, and future directions
[PDF] arxiv.org
T Li, AK Sahu, A Talwalkar, V Smith - arXiv preprint arXiv:1908.07873, 2019 - arxiv.org
... In Section 3, we outline several promising directions of future research. 2 Survey of Related and Current Work The challenges in federated learning at first glance resemble classical problems in areas such as privacy, large-scale machine learning, and distributed optimization ...
被引用次数：95 相关文章 所有 3 个版本
Threats to federated learning: A survey
[PDF] arxiv.org
L Lyu, H Yu, Q Yang - arXiv preprint arXiv:2003.02133, 2020 - arxiv.org
With the emergence of data silos and popular privacy awareness, the traditional centralized approach of training artificial intelligence (AI) models is facing strong challenges. Federated learning (FL) has recently emerged as a promising solution under this new reality. Existing ...
被引用次数：3 所有 2 个版本

了解定义及相关问题

- 什么是联邦学习?
- 联邦学习有哪些核心问题?
- 对于这些核心问题，解决方案有哪些？研究方向是什么?

整理脉络


联邦学习
联邦学习定义
An unbalanced and non-IID (identically and independently distributed) data partitioning across a massive number of unreliable devices with limited communication bandwidth was introduced as the defining set of challenges.
联邦学习分类
横向联邦学习（Horizontal Federated Learning or Cross-silo Federated Learning）
纵向联邦学习（Vertical Federated Learning or Cross-device Federated Learning）
联邦迁移学习（Federated Transfer Learning）
联邦学习面临的问题
Non-IID
设备算力不同
系统的不稳定
设备随时可能下线
可能存在攻击者
隐私保护
Global
Local
SMC协议
通讯成本
Local update
梯度压缩
Compression Schemes
模型广播压缩
Decentralized Training
减少本地计算

细化领域

联邦学习中的通信优化方法主要有三类，核心思想在于减少通信的次数或降低通信所耗的带宽

- Local update
- Compression Schemes
- Decentralized Training

细化领域

通过相关论文的 Related Work 再寻找细化领域的相关研究

DEEP GRADIENT COMPRESSION: REDUCING THE COMMUNICATION BANDWIDTH FOR DISTRIBUTED TRAINING

Yujun Lin *
Tsinghua University
linyy14@mails.tsinghua.edu.cn

Song Han †
Stanford University
Google Brain
songhan@stanford.edu

Huizi Mao
Stanford University
huizi@stanford.edu

Yu Wang
Tsinghua University
yu-wang@mail.tsinghua.edu.cn

William J. Dally
Stanford University
NVIDIA
dally@stanford.edu

细化领域

通过相关论文的 Related Work 再寻找细化领域的相关研究

2 Related Work

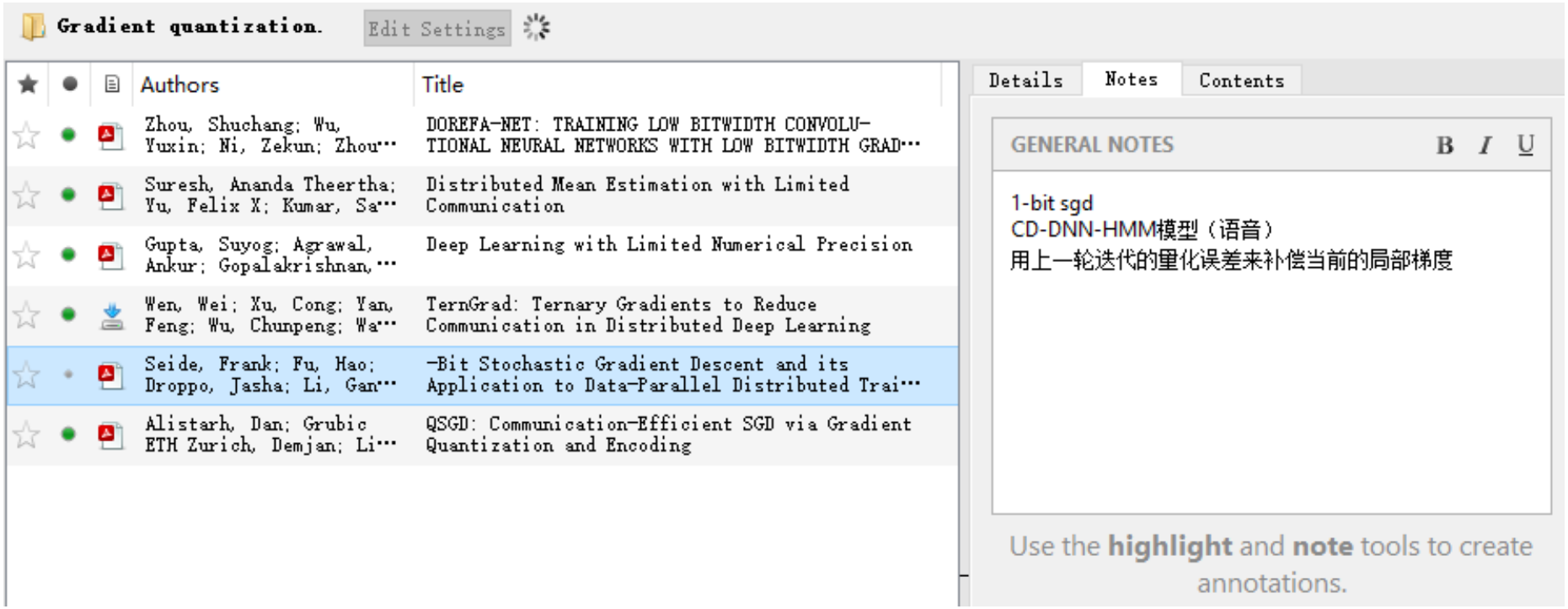
Researchers have proposed many approaches to overcome the communication bottleneck in distributed training. For instance, asynchronous SGD accelerates the training by removing gradient synchronization and updating parameters immediately once a node has completed back-propagation (Dean et al., 2012; Recht et al., 2011; Li et al., 2014). Gradient quantization and sparsification to reduce communication data size are also extensively studied.

Gradient Quantization Quantizing the gradients to low-precision values can reduce the communication bandwidth. Seide et al. (2014) proposed 1-bit SGD to reduce gradients transfer data size and achieved 10× speedup in traditional speech applications. Alistarh et al. (2016) proposed another approach called QSGD which balance the trade-off between accuracy and gradient precision. Similar to QSGD, Wen et al. (2017) developed TernGrad which uses 3-level gradients. Both of these works demonstrate the convergence of quantized training, although TernGrad only examined CNNs and QSGD only examined the training loss of RNNs. There are also attempts to quantize the entire model, including gradients. DoReFa-Net (Zhou et al., 2016) uses 1-bit weights with 2-bit gradients.

Gradient Sparsification Strom (2015) proposed threshold quantization to only send gradients larger than a predefined constant threshold. However, the threshold is hard to choose in practice. Therefore, Dryden et al. (2016) chose a fixed proportion of positive and negative gradient updates separately, and Aji & Heafield (2017) proposed Gradient Dropping to sparsify the gradients by a single threshold based on the absolute value. To keep the convergence speed, Gradient Dropping requires adding the layer normalization(Lei Ba et al., 2016). Gradient Dropping saves 99% of gradient exchange while incurring 0.3% loss of BLEU score on a machine translation task. Concurrently, Chen et al. (2017) proposed to automatically tunes the compression rate depending on local gradient activity, and gained compression ratio around 200× for fully-connected layers and 40× for convolutional layers with negligible degradation of top-1 accuracy on ImageNet dataset.

文献阅读

按使用的方法整理各类文献，记下文献提出的算法，使用的模型



References I

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*, feb 2016. [Online]. Available: http://arxiv.org/abs/1602.05629

[2] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. G. L. D'Oliveira, S. E. Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, P. B. Gibbons, M. Gruteser, Z. Harchaoui, C. He, L. He, Z. Huo, B. Hutchinson, J. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konečný, A. Korolova, F. Koushanfar, S. Koyejo, T. Lepoint, Y. Liu, P. Mittal, M. Mohri, R. Nock, A. Özgür, R. Pagh, M. Raykova, H. Qi, D. Ramage, R. Raskar, D. Song, W. Song, S. U. Stich, Z. Sun, A. T. Suresh, F. Tramèr, P. Vepakomma, J. Wang, L. Xiong, Z. Xu, Q. Yang, F. X. Yu, H. Yu, and S. Zhao, "Advances

References II

and Open Problems in Federated Learning," dec 2019. [Online]. Available: http://arxiv.org/abs/1912.04977

[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, jan 2019. [Online]. Available: https://doi.org/10.1145/3298981

Thank You!