華中科技大學
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

2021 年春季《物联网》课程报告

联邦学习学习笔记综述

姓　名　吴 叶 赛

班　级　电磁 1802

学　号　U201813405

日　期　2021.4.17

写在前面

笔记综述主要参考王琛老师的联邦学习相关 PPT 和《Federated Machine Learning Concept and Applications》和《联邦学习》，再在一些自己比较感兴趣的地方比如 FL 隐私方面的知识查阅相关资料加以拓展。这是一个入门总结笔记，当然也是结合了各种资料的内容，希望可以作为进一步学习的总体框架。

摘要

随着大数据的进一步发展，重视数据隐私和安全已经成为了世界性的趋势，同时，大多数行业数据呈现数据孤岛现象，如何在满足用户隐私保护、数据安全和政府法规的前提下，进行跨组织的数据合作是困扰人工智能从业者的一大难题。而“联邦学习”将成为解决这一行业性难题的关键技术。

联邦学习旨在建立一个基于分布数据集的联邦学习模型。

两个过程：模型训练和模型推理。

在模型训练中模型相关的信息可以在各方交换（或者以加密形式交换）

联邦学习是具有以下特征的用来建立机器学习模型的算法框架

有两个或以上的联邦学习参与方协作构建一个共享的机器学习模型。每一个参与方都拥有若干能够用来训练模型的训练数据

在联邦学习模型的训练过程中，每一个参与方拥有的数据都不会离开参与方，即数据不离开数据拥有者

联邦学习模型相关的信息能够以加密方式在各方之间进行传输和交换，并且需要保证任何一个参与方都不能推测出其他方的原始数据

联邦学习模型的性能要能够充分逼近理想模型(指通过所有训练数据集中在一起并训练获得的机器学习模型）的性能。

一．联邦学习总览

1.联邦学习背景介绍

当今，在几乎每种工业领域正在展现它的强大之处。然而，回顾 AI 的发展，不可避免地是它经历了几次高潮与低谷。AI 将会有下一次衰落吗？什么时候出现？什么原因？当前 大数据的可得性 是驱动 AI 上的 public interest 的部分原因：2016 年 AlphaGo 使用 20 万个游戏作为训练数据取得了极好的结果。

然而，真实世界的情况有时是令人失望的：除了一部分工业外，大多领域只有有限的数据 或者 低质量数据，这使得 AI 技术的应用困难性超出我们的想象。有可能通过组织者间转移数据把数据融合在一个公共的地方吗？ 事实上，非常困难， 如果可能的话，很多情况下要打破数据源之间的屏障。由于工业竞争、隐私安全和复杂的行政程序，即使在 同一公司的不同部分间的数据整合 都面临着严重的限制。几乎不可能整合遍布全国和机构的数据，否则成本很高。

同时，意识到大公司在数据安全和用户隐私上的危害，重视数据隐私和安全已经成为世界范围的重大事件。 公众数据泄露的新闻正在引起公众媒体和政府的高度关注。 例如，Facebook 最近的数据泄露引起了广泛抗议。作为回应，世界各个州正在加强法律保护数据安全和隐私。例如 欧盟在 2018.5.25 实施的《通用数据保护条例》General Data Protection Regulation（GDPR），旨在保护用户个人隐私和提供数据安全。它要求企业使用清晰明了的语言来达成用户协议，并授予用户“被遗忘的权力”，也就是说，用户可以删除或撤回其个人数据。违

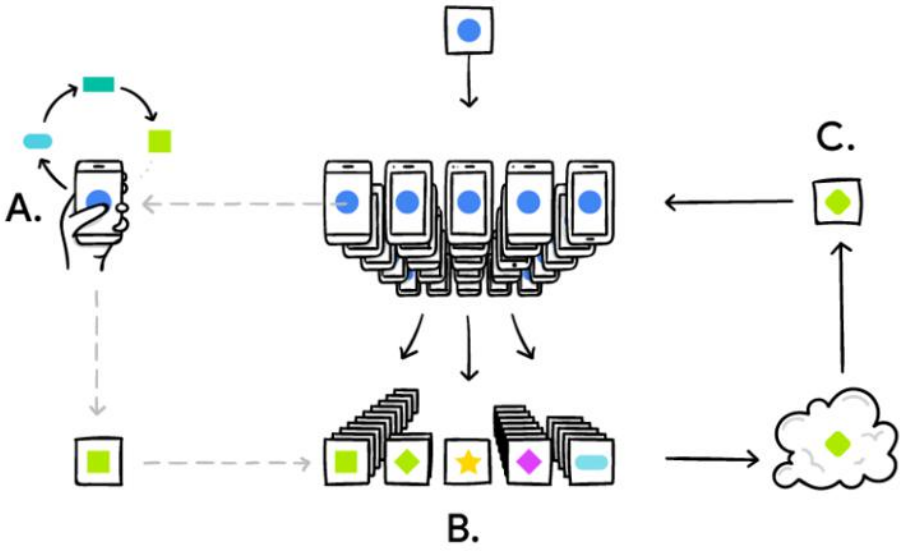
反该法案的公司将面临严厉罚款。 2017 年实施的《中国网络安全法》和《民法通则》要求互联网业务在与第三方做数据交易时禁止在手记个人信息时进行泄露和篡改，并且他们需要确保提出的合同遵守法律数据保护业务。 这些法规的建立显然将有助于建立一个更加文明的社会，但也将给当今 AI 中普遍使用的数据交易程序带来新的挑战。

更具体地，AI 传统的数据处理模型通常参与简单的数据交易模型， 一方参与者收集和转移数据到另一方，其他参与者负责清洗和融合数据，最终一个第三方机构 将拿走整合的数据 建立模型 models 给其他参与者使用。 这些模型 models 通常是最终产品作为服务售卖。 这种传统的程序面临以上新数据法规的挑战。并且，因为用户可能不清楚模型的未来用途，所以交易违反了 GDPR 等法律。

因此，我们面临着一个难题，即我们的数据是孤立的孤岛形式，但是在许多情况下，我们被禁止在不同地方收集、融合和使用数据进行 AI 处理。如何合法地解决数据碎片和隔离问题是当今 AI 研究人员和从业人员面临地主要挑战。

2．什么是联邦学习

联邦学习（Federated Learning）是一种新兴的人工智能基础技术，在 2016 年由谷歌最先提出，原本用于解决安卓手机终端用户在本地更新模型的问题，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的机器学习。其中，联邦学习可使用的机器学习算法不局限于神经网络，还包括随机森林等重要算法。联邦学习有望成为下一代人工智能协同算法和协作网络的基础。



3.设计目标与框架

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的机器学习

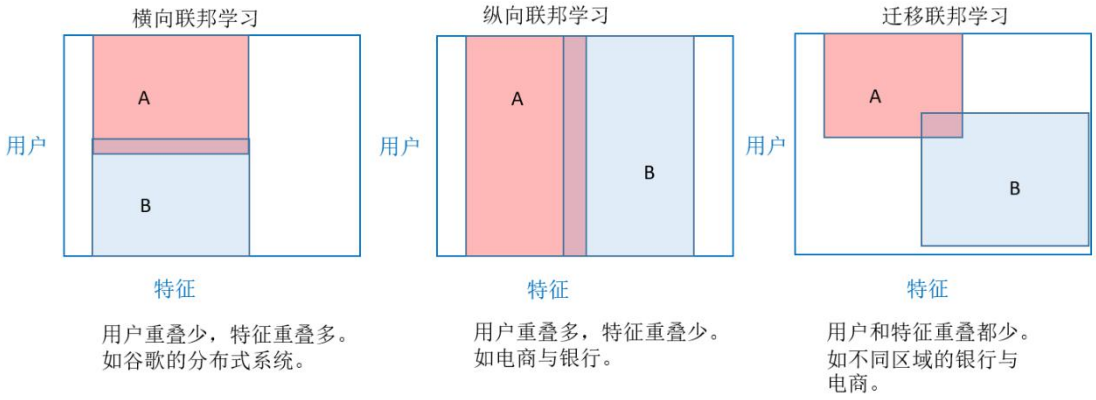
例如两公司要联合训练一个 AI 框架，但是又无法直接进行数据交换，可以使用联邦学习来建立模型。更广泛的模型就是通过众多的用户端得到的非平衡且非独立同分布的数据来训练。

更形象点：正常的机器学习模型是把数据收集到服务器端，将数据带到代码面前来；而联邦学习则是把代码发送到数据端，从而使用户的数据无需上传就可以参与模型的训练。

不过在用户端需要上传训练的梯度给服务器，服务器负责把梯度进行聚合得到最终的模型，同时在刚开始训练的时候也会去拉去最新的模型。

4．联邦学习分类

联邦学习根据数据和数据持有者的性质可以分为：横向联邦学习，纵向联邦学习和迁移联邦学习。



联邦学习的分类体系，如上图中显示的包括：

纵向联邦学习，两个数据集的用户（U1, U2, …）重叠部分较大，而用户特征（X1, X2, …）重叠部分较小；

横向联邦学习，两个数据集的用户特征（X1, X2, …）重叠部分较大，而用户（U1, U2, …）重叠部分较小；

联邦迁移学习，通过联邦学习和迁移学习，解决两个数据集的用户（U1, U2, …）与用户特征重叠（X1, X2, …）部分都比较小的问题。

稍微白话一点来说的话

1、横向联邦学习：两个数据集的用户特征重叠较多，而用户重叠较少的情况，把数据集横向（按用户维度）切分，取出双方：用户特征相同而用户不完全相同的数据训练。例如：对一个产品会有多个不同的店面，（不同店面保护自己的数据）使得各地特征相同而用户不同

2、纵向联邦学习：两个数据集的用户重叠较多而用户特征重叠较少的情况，把数据集纵向（按特征维度）切分，取出双方：用户相同而用户特征不完全相同的数据训练。例如：对同一个产业链，上下游具有相同的客户但又不同的特征，他们都要保护用户数据。这个时候把这些特征在加密状态聚合来增强模型的能力。

3、联邦迁移学习：两个数据集的用户和用户的特征都重叠较少的情况，这个时候不能切分，而使用迁移学习来克服数据或标签不足的情况。例如：一个在华科的物联网课程组，另一个是在武大的保安团队。。。。。（随便举的例子）

4．联邦学习框架

客户端-服务器架构（用的多）

各数据持有方根据自己的条件和规则在本地训练模型，将脱敏参数汇总到中央服务器。其次中央服务器将新参数下发回数据持有方，接着各参与者更新自己的模型。重复迭代上述模型，直到稳健。

对等网络架构（都是客户端参与）

不需要第三方服务器，安全性高但是增加计算开销。

5．联邦学习和分布式学习的区别

其实联邦学习就是分布式学习的一种，可以参考美国的联邦政体。各州都有很大的自治权，即每一个参与者都可以看做一个邦，有很大的自主权利。

主要区别在一下几个方面：

① FL 的用户可以控制自己的设备或者数据（控制权程度不同）

② 终端节点不稳定，具有很强的异构性（随时参与或退出，网络环境也不同）

③ 通信代价＞计算代价（带宽低，延迟高）

④ 各终端数据集不平衡

二．联邦学习隐私

1.存在的安全问题

存在的安全问题可以分为攻击，安全和可信性三个方面

云服务器在数据训练过程没有设置参与者的权限，这就导致恶意参与者会上传不正确的模型，并达到破坏全局模型的能力

将本地模型更新和全剧模型参数结合可以得到训练数据中的隐含知识从而造成信息泄露

不可信的云服务器和恶意参与者的合谋攻击下会造成用户数据的泄露（从内部实体发起攻击具有更强大的威胁性）

1.1 攻击者主要破坏方面

机密性：机密性主要体现在两个方面。①窃取训练数据中的敏感信息②暴露目标模型信息及预测结果

完整性：攻击者会诱导模型行为，使得模型输出为自己指定的分类标签

可用性：阻止用户获得模型的正确输出，干预用户获取模型的某些特征，使得模型不具备可靠性

1.2 存在的一些问题：

（1）服务器的半诚实模型

（2）云服务器和端之间的不信任问题

（3）被共享的全局参数

2.联邦学习隐私

2.1 安全多方计算(Secure Multiparty Computation (SMC)）

SMC 安全模型涉及多方，并在定义明确的仿真框架中提供安全性证明，以确保完全零知识，也就是说，除了输入和输出外，每一方都不知道。零知识是高度期望的，但是这种期望的属性通常需要复杂的计算协议，并且可能无法有效实现。在某些情况下，如果提供安全保证，则可以认为部分知识公开是可以接受的。可以在较低的安全性要求下用 SMC 建立安全性模型以换取效率《Privacy-preserving multivariate statistical analysis:Linear regression and classification，2004》。最近，一项研究使用 SMC 框架来训练带有两个服务器和半诚实假设的机器学习模型。最先进的 SMC 框架之一是 Sharemind 《 A framework for fast privacy-preserving com-putations，2008》。《 mixed protocol framework for machine learning，2018》的作者以诚实的多数提出了 3PC 模型，并**考虑了半诚实和恶意假设中的安全性。这些作品要求参与者的数据在非竞争服务器之间秘密共享。

2.2 差分隐私（Differential Privacy）

另一工作线使用差异隐私技术《Differential privacy: A survey of results. 2008》或 k-匿名用户技术《K-anonymity: A model for protecting privacy.2002》进行数据隐私保护。差异隐私，k 匿名和多样化的方法《Privacy-preserving data mining.2000》

涉及给数据添加噪声，或使用归纳法掩盖某些敏感属性，直到第三方无法区分个人为止，从而使数据无法恢复保护用户隐私。但是，这些方法的根本仍然要求将数据传输到其他地方，这通常需要在准确性和隐私之间进行权衡。在《Differentially private federated learning: A client level perspec-tive.2017》中，作者介绍了一种用于联合学习的差分隐私方法，目的是通过隐藏客户在培训期间的贡献来为客户数据提供保护。

2.3 同态加密（Homomorphic Encryption）

在机器学习过程中，也采用同态加密《 On data banks and privacy homomorphisms.1978》来通过加密机制下的参数交换来保护用户数据隐私。与差异隐私保护不同，数据和模型本身不会被传输，也不会被另一方的数据猜中。因此，在原始数据级别泄漏的可能性很小。最近的工作采用同态加密来集中和训练云上的数据。在实践中，加性同态加密《A survey on homomorphic encryptionschemes: Theory and implementation.2018》被广泛使用，并且需要采用多项式近似来评估机器学习算法中的非线性函数，从而在准确性和隐私性之间进行权衡。

3.一些类型攻击防御

结合 2 中的三种方法，从差分隐私（Differential privacy，DP）、同态密码系统（Homomorphic cryptosystem，HC）和安全多方聚合（Secure multi-party aggregation，SMA）3 个角度，简要介绍几种具有可行性的策略，

3.1 投毒攻击防御

数据投毒防御： 防御方法应从保护数据的角度出发。一方面，在训练模型之前应当保证数据来源的真实性与可靠性。另一方面，在使用不能保证安全性的数据之前，应当进行相应的检测以保证数据完整性不受篡改

模型投毒防御： 假定服务器是可信的，那么防御的重点在于对恶意参与方的识别以及对错误更新参数的检测。

3.2 对抗攻击防御

对抗训练： 将真实的样本和对抗样本一起作为训练集，来训练出最后的模型，它可以使得模型在训练过程中就学习到对抗样本的特征，提高模型的健壮性。

数据增强： 数据增强是对抗攻击的一种扩充。在训练过程中不可能穷举所有对抗样本，但通过对原始数据集中的数据进行随机化处理可以增强模型的泛化能力。

数据处理： 数据处理技术是指对样本进行降噪处理，以减小对抗样本的干扰。

数据压缩： 数据压缩是一种特殊的数据处理方法，专门针对图像训练过程，即使用压缩后的图片进行训练。

防御蒸馏： 主要思想是先利用训练集得到一个模型，然后再通过模型提取，从原来的模型“蒸馏”提纯出另外一个模型，从而降低模型的复杂度。

梯度正则化： 模型训练中常使用正则化来防止过拟合，即过度学习样本特征。若模型过拟合程度越高，其泛化能力越弱，越容易遭受到对抗样本的攻击。梯度正则化是指在训练模型的目标函数上对输人与输出的变化进行惩罚，从而限制了输人的扰动对于预测结果的影响。

对抗样本检测： 若能区分出对抗样本与正常样本的不同之处，然后以较高精度检测出对抗样本，就能较好地防范对抗攻击

基于 GAN 的防御： 生成式对抗网络（generative adversarial net，GAN）是一种机器学习模型，由两个模块组成。一个是生成模块G，利用接收到的随机噪声生成虚假样本，另一个是判别模块D，用以判断出某样本是否为G生成的虚假样本。

3.3 隐私泄露防御

联邦学习中的隐私保护主要从两大主体—参与方与服务器的角度进行保证。同时对于训练完成的模型也要防止模型提取攻击和模型逆向攻击。

差分隐私

秘密分享

同态加密

混合防御机制

三．结论与展望

众所周知，随着数据隐私和数据安全的重要性日益提高，以及公司利润与其数据之间的紧密关系，云计算模型受到了挑战。但是，联邦学习的业务模型为大数据的应用提供了新的范例。当每个机构占用的孤立数据无法产生理想的模型时，联邦学习的机制使机构和企业无需数据交换就可以共享一个统一的模型。此外，借助区块链技术的共识机制，联邦学习可以为利润分配制定公平的规则。数据拥有者，无论他们拥有的数据规模如何，都将被激励加入数据联盟并获得自己的利润。

我们认为应该将数据联盟的业务模型的建立与联邦学习的技术机制一起进行。我们还将制定各个领域的联合学习标准，以尽快将其投入使用。

参考文献

Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2): 1-19.

《联邦学习》 杨强等著

2016 年谷歌研究院论文网址 Communication-Efficient Learning of Deep Networks from Decentralized Data

王琛老师物联网教学 PPT