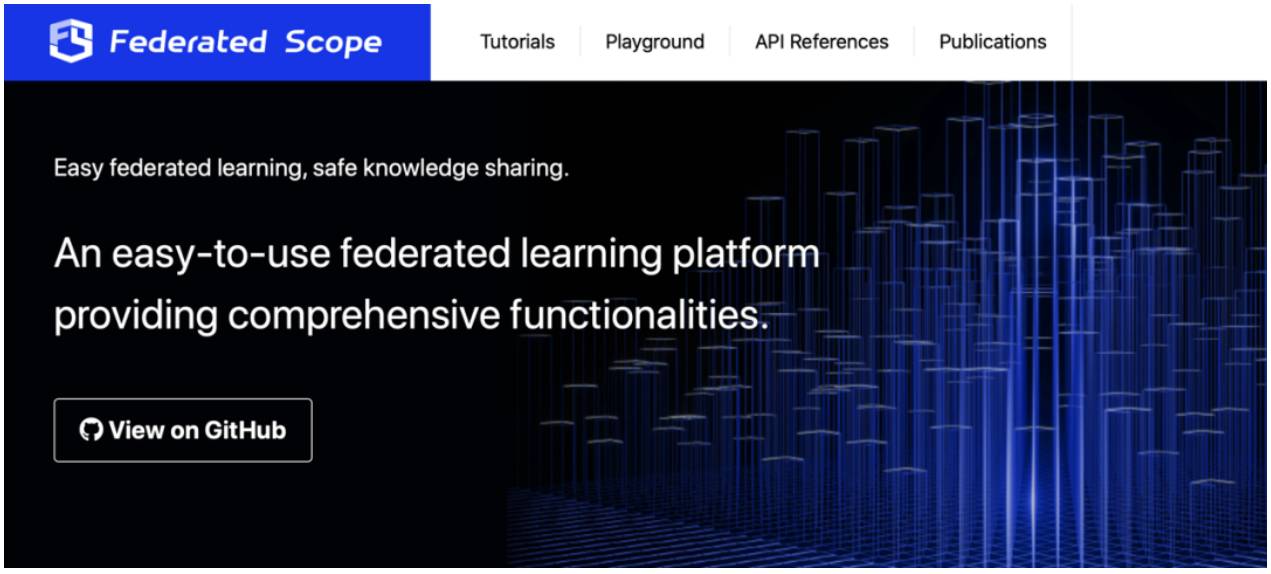
今日开源：阿里达摩院最新框架FederatedScope来了！让联邦学习从可用到好用

机器之心 2022-05-05 13:05

机器之心发布

机器之心编辑部

> 刚刚，阿里巴巴达摩院发布新型联邦学习框架 FederatedScope，该框架支持大规模、高效率的联邦学习异步训练，能兼容不同设备运行环境，且提供丰富功能模块，大幅降低了隐私保护计算技术开发与部署难度。该框架现已面向全球开发者开源。



隐私保护是数字经济的安全底座，如何在保障用户数据隐私的同时提供高质量连通服务，成为数字经济时代的重要技术课题。为破解隐私保护与数据应用的两难，以 “数据不动模型动” 为理念的联邦学习框架应运而生，并成为隐私保护计算近年最主流的解决方案之一。

具体而言，联邦学习框架成功实现了“数据可用不可见”。用户自身的数据从始至终都停留在用户自己的手机或汽车等终端内，不会“出域”；同时，训练机器学习模型需要的信息，譬如梯度，会以不同的方式被保护(加密、加噪声或拆分)，然后在云端的服务器进行聚合，从而进行模型训练；此后云端再将更新的模型推送给手机端或者车端。通过这样的交互和迭代过程，服务提供商既能够训练高性能的模型为用户提供服务，同时也能保护好用户的数据隐私。

目前开源的联邦学习框架包括TensorFlow Federated（TFF）、FATE等。这些框架提供了联邦学习相关基础组件及实现方式，如联邦聚合、差分隐私、同态加密等，为联邦学习相关社区研究和工业应用都提供了一定支持。

然而，现实生活中日益多样化的应用场景，以及联邦学习任务中存在的异构特点（如数据异构，系统资源异构，行为异构等），给联邦学习框架带来了新的挑战。**目前已有的联邦学习框架难以灵活高效地满足现实中越来越复杂的计算需要，需从注重“可用”向注重“好用”转变。**

首先，联邦学习参与方之间传递的信息形式会更加丰富，不再局限于模型参数或者梯度这一类的同质信息。例如在图数据上的联邦学习，参与方之间还会传递节点的嵌入式表示等信息；在垂直联邦学习的场景下，参与方之间还会传递公钥和一些加密过的中间结果信息。丰富的信息种类要求联邦学习框架能灵活支持不同类型的信息传递。

其次，联邦学习参与方的行为种类更加多变，不再拘泥于传统的 “服务器端负责聚合，用户端负责本地训练” 的模式。例如在跨设备的联邦学习场景中，往往需要对服务器端的模型做压缩处理，来满足终端设备的运行要求；而在终端设备上，往往会对收到的模型进行微调来取得更好的效果。多样化的参与方的行为要求联邦学习框架能够灵活地支持多种自定义行为。

同时，联邦学习参与方的响应速度和可靠性参差不齐，采用传统的同步训练的方式容易造成训练效率差，系统利用率低等问题。这要求联邦学习框架能够允许开发者根据应用场景采用不同的异步训练策略，在保证训练效果的同时提升训练的效率。

再者，现实应用中联邦学习参与方可能搭载不同的模型训练环境，例如有些设备后端环境使用的是 PyTorch，而另外一些则使用 TensorFlow。这要求联邦学习框架需要有更好的兼容性，能支持跨平台组建联邦学习，而避免要求使用者费时费力地对所有参与方进行环境的适配。

最后，联邦学习框架应该为单机仿真和分布式部署提供统一的算法描述和接口，以满足研究者和开发人员不同的应用需求，并降低从仿真到部署的迁移难度，缩小联邦学习从学术研究到工业应用的鸿沟。

为解决上述挑战，达摩院智能计算实验室研发了联邦学习框架 FederatedScope，该框架于 5 月 5 日正式对外发布并开源。

FederatedScope 采用事件驱动的编程范式，用于支持现实场景中联邦学习应用的异步训练，并借鉴分布式机器学习的相关研究成果，集成了异步训练策略来提升训练效率。具体而言，FederatedScope 将联邦学习看成是参与方之间收发消息的过程，通过定义消息类型以及处理消息的行为来描述联邦学习过程。

FederatedScope 通过把联邦过程（例如协调不同的参与方）和模型训练行为（例如训练数据采样、优化等）解耦开，使开发者能够专注于定制参与方处理收到消息的行为，而不需要从顺序执行的角度考虑如何串联不同参与方。例如在经典的 FedAvg 算法实现中，用户只需定义聚合端收到用户端发送的模型参数信息后的聚合行为，以及用户端收到聚合端广播新一轮模型参数之后的本地训练行为。

对于包含异质消息传递和丰富消息处理行为的联邦学习任务，FederatedScope 支持用户通过添加额外的消息类型和处理行为进行定制化。同时，FederatedScope 内置了大量的消息类型和相应的消息处理行为，能够很好地服务不同场景下的联邦任务，很大程度地降低了开发者和使用者的上手门槛。

同时，达摩院团队对 FederatedScope 训练模块进行抽象，使其不依赖特定的深度学习后端，能兼容 PyTorch、Tensorflow 等不同设备运行环境，大幅降低了联邦学习在科研与实际应用中的开发难度和成本。

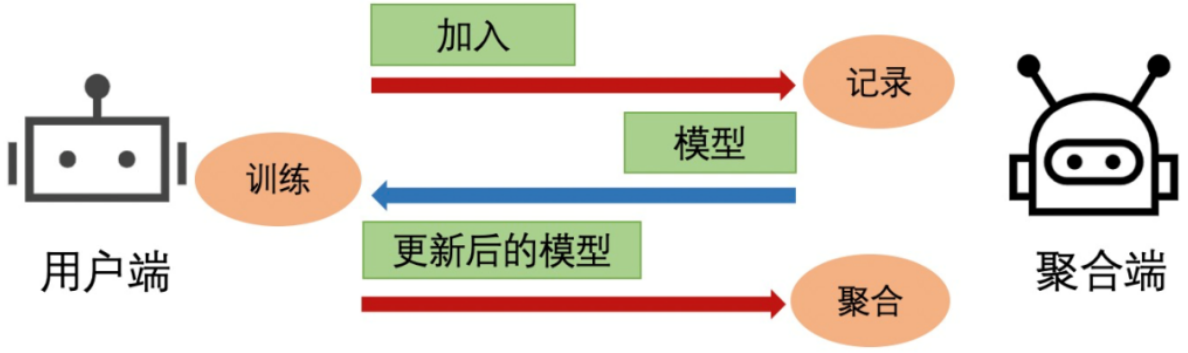


图 1. 经典联邦学习

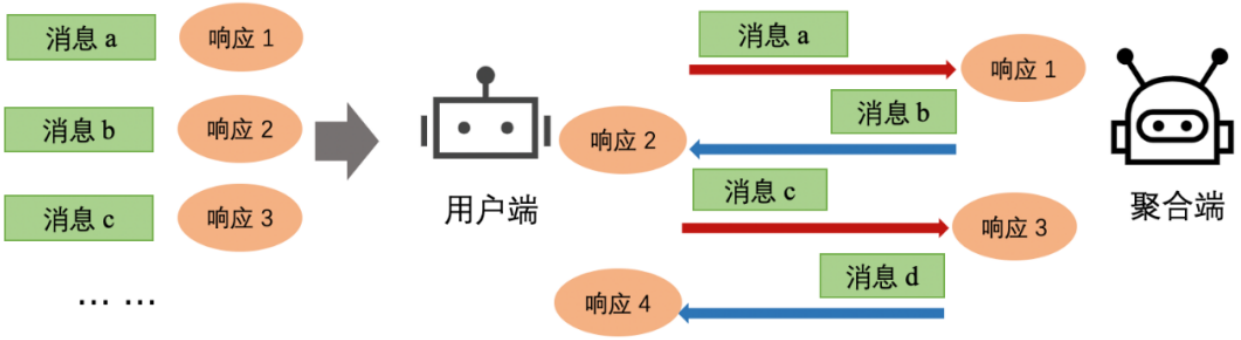


图 2. 事件驱动

为进一步适应不同的应用场景，FederatedScope 还集成了多种功能模块，包括自动调参、隐私保护、性能监控、端模型个性化。FederatedScope 支持开发者通过配置文件便捷地调用集成模块，也允许通过注册的方式为这些模块添加新的算法实现并调用。具体而言：

- （1）自动调参能大幅降低搜索最优超参的时间和资源消耗。FederatedScope 提供了最新的联邦学习自动调参算法方便开发者直接使用。同时，自动调参模块也抽象了自动调参算法框架，从而方便研究人员开发新的调参算法。

- （2）隐私保护是所有场景的通用需求，FederatedScope 的隐私保护模块提供了主流的隐私保护机制，包括多方安全计算、同态加密和差分隐私。除此以外，隐私保护模块额外提供了主流的隐私评估算法，方便开发者验证隐私保护的强度。

- （3）性能监控能够帮助开发者随时了解训练进展，及时发现训练异常。FederatedScope 的性能模块能以友好的界面展示训练过程的多种中间信息，包括每一个用户端的训练结果和聚合端的评价等。

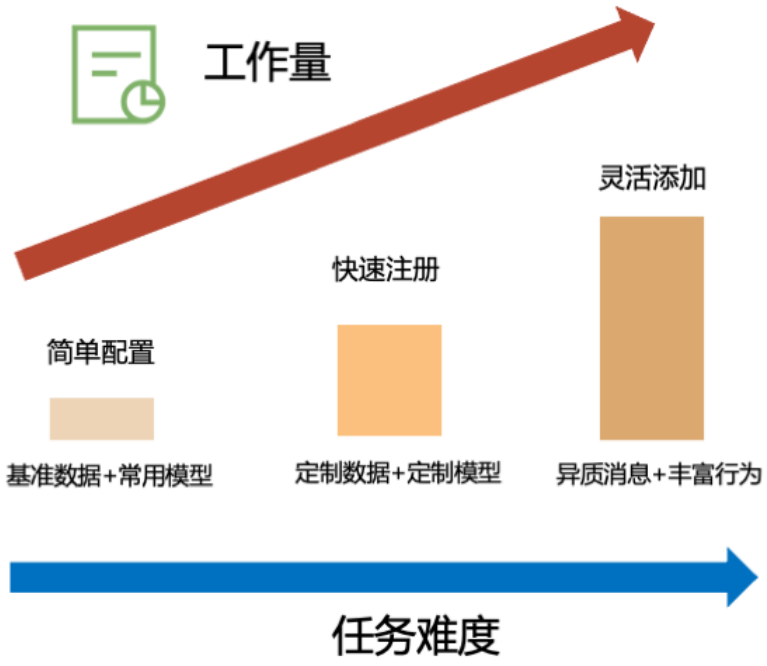
- （4）由于联邦学习参与方的数据分布和设备性能可能存在较大的差异性，端模型个性化是应用场景中的强需求。FederatedScope 的端模型个性化模块实现了差异化训练配置、定制训练模块、个性化参与方的训练行为、维护全局和个性化的本地模型等功能，从而达成端云协同。同时端模型个性化模块提供了丰富的个性化算法方便开发者调用。

相比传统的联邦学习框架，FederatedScope 易用性尤为突出，以下几个例子可具体说明：

- (1)对于初次接触联邦学习的使用者来说，FederatedScope 提供了详尽的教程、文档和运行脚本，能够引导用户快速入门上手联邦学习。FederatedScope 也包含了常用的模型架构实现，对一些基准数据集也做了统一的预处理和封装，以帮助用户便捷地开展实验。

- (2)对于希望将经典联邦学习应用在不同下游任务的开发者，如使用不同的数据和模型架构，FederatedScope 允许通过注册的方式使用准备好的新数据集和模型架构，而不需要修改其他的细节。另外，FederatedScope 也支持根据任务类型定制不同的性能监控和评价指标。

- (3)对于希望深入研究和开发联邦学习算法的用户，需要足够的自由度在联邦学习中添加异质信息交换和多样的处理行为，在FederatedScope中只需定义消息的类型和相应的处理函数。相比现有的联邦学习框架，FederatedScope的优点在于不需要开发者将联邦学习的过程用顺序执行的视角来完整描述，而只需采用事件驱动的方式增加新的消息类型和消息处理行为，系统协助完成自动调参和高效异步训练，降低了所需的开发量以及复杂度。



总体而言，通过采用事件驱动的编程范式，将联邦学习抽象成异构消息的传输和处理，同时集成丰富多样的算法策略和功能模块，FederatedScope 能够很好的应对联邦学习应用中存在的异构特点，灵活地支持不同联邦学习应用场景的多样化需求，且易于使用和二次开发。与现有的联邦学习框架相比，FederatedScope 大幅降低了开发者应用的难度。

达摩院智能计算实验室隐私保护计算团队负责人丁博麟表示，“数据已成为重要的生产要素，而隐私保护计算是保障这一要素发挥作用的关键技术。通过开源最新联邦学习框架，我们希望促进隐私保护计算在研究和生产中的广泛应用，让医药研发、政务互通、人机交互等数据密集领域更

安全、更顺畅地发展。”

Gartner 相关报告显示，到 2025 年之前，约 60% 的大型企业预计将应用至少一种隐私保护计算技术。达摩院 2022 十大科技趋势同样将隐私保护计算列为重要趋势，认为该技术将从覆盖少量数据的场景走向全域保护，从而激发数字时代的新生产力。

更多信息可访问 FederatedScope

介绍网站：https://federatedscope.io/；

开源地址：https://github.com/alibaba/FederatedScope

喜欢此内容的人还喜欢

「铜三铁四」裁员潮，大厂 AI 青年生存也遇问题?

AI科技评论

阿里李飞飞：在云计算时代，云原生数据库变得越来越重要

AI科技评论