



移动边缘计算中的资源高效联邦学习

联邦学习中资源优

移动边缘计算中联邦学习的瓶颈是移动客户端在计算、带宽、能量和数据等方面的密集资源。

本文提出了一种**基于模块联合学习的神经结构感知资源管理方法**，该方法根据移动客户端局部资源的状态分配全局模型的不同子网。

实验证明了该方法在资源弹性高效利用方面的优越性。

联邦学习中资源优化的方法大致分为两类:黑盒方法和白盒方法。

黑盒方法：

(1) 训练技巧:

使用训练技巧是提高联合学习性能的一种简单而有效的方法。通过经验调整训练的超参数，如梯度下降策略、批大小、时间、学习率和衰减率，以减少通信轮数。

(2) 客户端选择:

移动边缘计算采用同步模型聚合进行联邦学习，每一轮通信的持续时间由计算和通信时间最长的客户端决定。因此，选择一组具有良好无线信道条件和足够计算能力的合适的客户端是至关重要的。为了避免延迟参数聚合，“弱”客户端将被较少地涉及。为了保证调度的

公平性和数据分布的均衡，当客户端的无线信道条件和计算资源恢复到平均状态时，客户端的调度需要对弱调度进行补偿。

(3)数据补偿:

以往的理论分析表明，数据分布在客户端之间的异质性是制约联合学习收敛速度的关键因素。为了平衡数据的异构性，每个客户端的一小部分本地数据被全局共享，以缓解非iid的程度，从而可能提高模型的准确性。

(4)分层聚合:

将分层网络拓扑结合到联合学习中，可以采用client-edge-cloud多阶段方法进行模型聚合，即先在边缘服务器上进行一次到两个阶段的局部聚合，然后在中心服务器上进行一次全局聚合。这种多阶段聚合的基本原理是，网络边缘的早期模型聚合具有较低的通信成本，能够有效地缓解由于局部数据的随机性而导致的模型更新的不确定性。一旦中央服务器更新了精细化的模型，全局聚合的收敛速度就会相当快。

白盒方法：

(1):模型压缩:

模型压缩由一组优化技术组成，用于缩小模型尺寸，便于在硬件环境中部署，因此，它被广泛用于减少分布式学习中的通信开销。

(2)知识蒸馏:

基于知识蒸馏的联合学习选择在客户机之间共享类分数，并进行聚合，以获得模型更新的一致意见。

(3)特征融合:

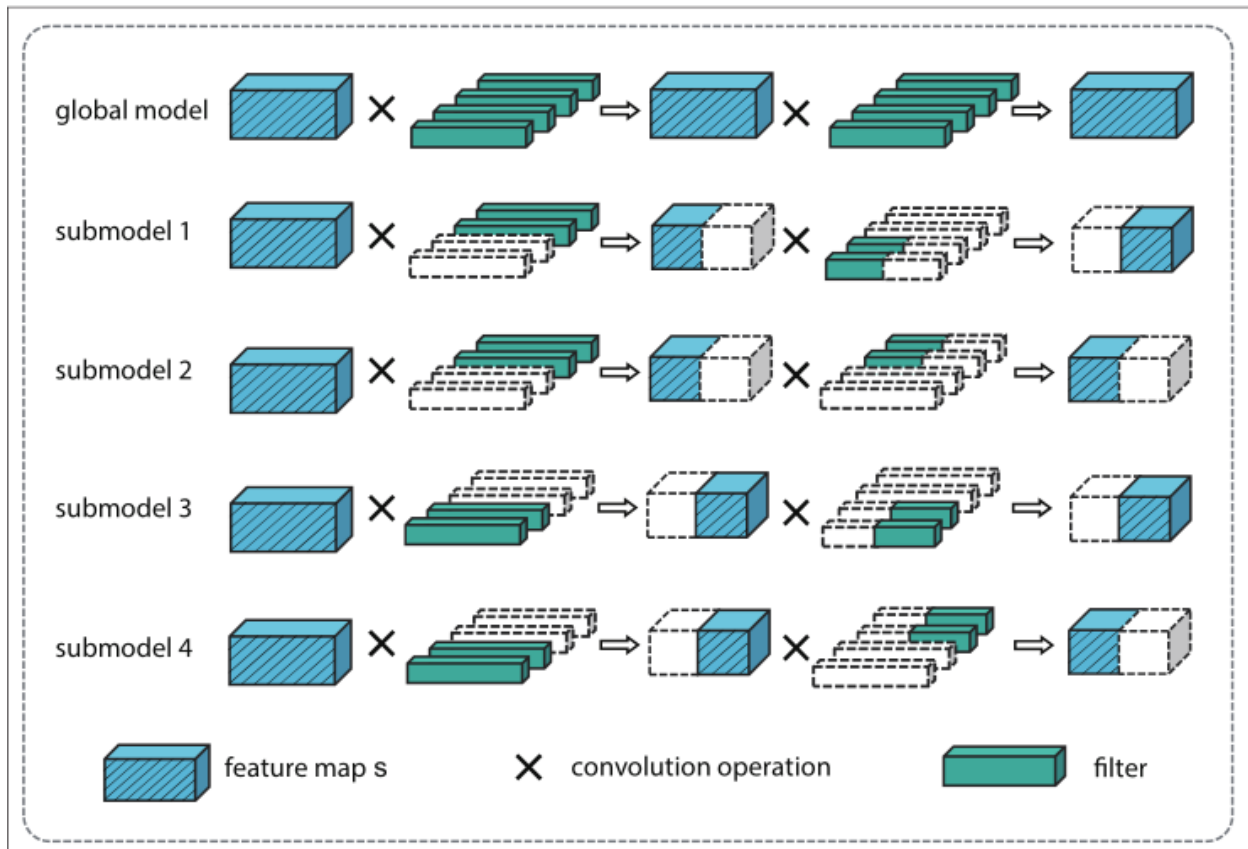
每个客户端从中央服务器接收到全局模型更新后，复制全局特征提取器来替换本地特征提取器。全局提取器和局部提取器的两流输出通过特征融合模块进行融合，然后传递给分类器进行损失评估。

(4)异步更新:

利用异步学习的框架，允许中央服务器在不收集所有本地模型的情况下进行模型聚合。与时间持续时间相对应的时间系数与模型权重相乘，以获得更好的聚合。

在本节中，提出了一种基于模块的白盒方法，将全局模型划分为多个子模型进行训练。

通过模型划分，可以分配不同的客户端来联合训练相同的子模型，或者使用本地资源分别训练不同的子模型。该联合学习框架同时集成了数据并行和模型并行两种并行计算模式。因此，可以有效地协调移动客户端的资源来完成子模型的培训。



模型划分与聚合

考虑按宽度(即卷积层中过滤器的数量)划分模型。

通过将滤波器分组，将全局模型划分为多个子模型。将一层的分区因子定义为该层的输出通道组数。分区因子对于不同的层可能是不同的，通常以2的幂给出值。子模型由来自所有卷积层的任意数量的过滤器块，以及输入、输出和完全连接的层组成。

中央服务器定期交换子模型之间的过滤器块，以确保全局模型的所有过滤器都能得到所有客户端的同等训练。

对于模型聚合，中心服务器应该重新组装所有分区的子模型，并重构全局模型。全局模型的多个副本可以同时被划分和训练。这些全局模型的副本可能以不同的方式进行划分，并由不同的客户进行培训。将多路复用系数定义为全局模型副本的数量。当聚合全局模型时，中央服务器将首先重构全局模型的所有副本，然后在副本之间进行梯度聚合。模型梯度的聚集系数与局部数据集的大小成正比。

能效高的模型训练问题

为了为模型训练分配资源，评估不同子模型结构的能量和时间成本。包括计算能量消耗、传输能量消耗、计算时间消耗、传输时间消耗、无线信道路径损耗。

通过将所有客户端通信和计算的相应能量成本相加，计算出总能量成本 E_k 。总时间代价 T_k 是所有客户端时间代价的最大值，对于每个客户端，其总时间代价 T_k 是通信和计算的时间代价之和。

提高对不同的客户端的在每次迭代中模型学习效率

启发式搜索算法：

在给定客户端、功率预算以及特定的CNN模型，通过搜索算法，实现最佳模型学习效率。

(1)子模型结构搜索:

对于每个模型分区，根据子模型的大小从大到小排序，而客户端根据其通道条件从好到坏排序。按顺序将客户端与子模型匹配。调用程序得到次优的能量和时间成本，然后得到次优的效用。比较所有次优效用，得到最优效用。

(2)时间持续搜索:

给定搜索时间代价的范围，采用二分搜索技术。在每一轮搜索中，调用程序，得到次优计算频率和发射速率。计算能源成本，并与预算进行比较。二分搜索一直进行到能量消耗足够接近预算为止。回报次优的精力和时间成本。

(3)资源配置搜索:

通过输入时间开销和模型划分，对所有客户端以最小的能量开销计算出次优的计算频率和传输速率。返回次优计算频率、传输速率和能量成本。

结论：

本文研究了联邦学习中关于资源优化的现有工作，并将其大致分为黑盒方法和白盒方法。提出了基于模块的联合学习框架和感知神经结构的资源管理方法。实验表明，它们在提高资源利用效率方面具有一定的优势。