联邦学习聚合架构 MOON 与 FedAvg 对比实验

徐子荣

摘要

联邦学习作为一种协作式机器学习方法，允许用户通过共享模型而不是原始数据的方式进行多方模型训练，在实现隐私保护的同时能够充分利用用户数据。然而，由于不能实现数据的互通有无，整体模型的准确度相比于全局训练会有一定程度的下降。这就需要我们对模型聚合阶段提出新的方法或架构来弥补这种准确度的损失。MOON 框架便是最新提出的模型聚合架构，不仅能够在一定程度上弥补准确度损失，还能弥补异质训练数据造成的模型偏移影响。本文对 MOON 架构[1] 和 FedAvg 架构[2]进行了对比实验，证明了 MOON 框架的有效性。

关键词：联邦学习；数据异质性

1 引言

深度学习需要大量的训练数据来进行训练，模型能够从大量有代表性的数据集中学到很多“知识”。然而，在实际运用时，数据通常散布在不同的区域（例如，不同的终端设备中）。由于隐私问题和数据安全愈发受到人们的重视，各方逐渐希望能够避免将私人数据上传至集中服务器训练的这种模型训练方式。

为了解决上述问题，联邦学习作为一种能够联合多方共同训练学习的机器学习架构被提出了。在联邦学习的诸多架构中，普遍使用的是 FedAvg 架构。在 FedAvg 的每一个通讯轮次中，每个分部更新的局部模型都被传输到服务器，在服务器中进行聚合以更新全局模型。在这种学习过程中，原始的训练数据不会被交换，能够避免私人数据的泄露。

然而，联邦学习的一个关键挑战在于数据在不同分部的分布是不同的，是异质的。这会使得每个分部在更新本地模型时，使得本地目标远离全局目标，从而导致聚合出来的全局模型会远离全局最优值。目前已有一些方法例如 FedProx[3]通过二范数距离来限制局部更新，然而通过实验研究发现其效果只是延缓了偏离的步伐，没有最终改变其偏离的方向。

在 MOON 这项工作中，作者基于整个数据集上训练的全局模型会比在有偏子集上训练的局部模型更好这一角度，来解决 Non-IID 的问题。具体而言，MOON 架构通过最大化本地模型表征与全局模型表征的一致性来纠正本地目标的偏离。通过比较不同模型学到的模型表征进行模型对比学习，使得局部模型的更新方向与全局更加一致，由此解决 Non-IID 的问题。

2 相关工作

2.1 联邦学习

联邦学习是一种直观的聚合模型的框架。最为常用的 FedAvg 已经是一个在实际使用的框架。其具体流程如图 1 所示。

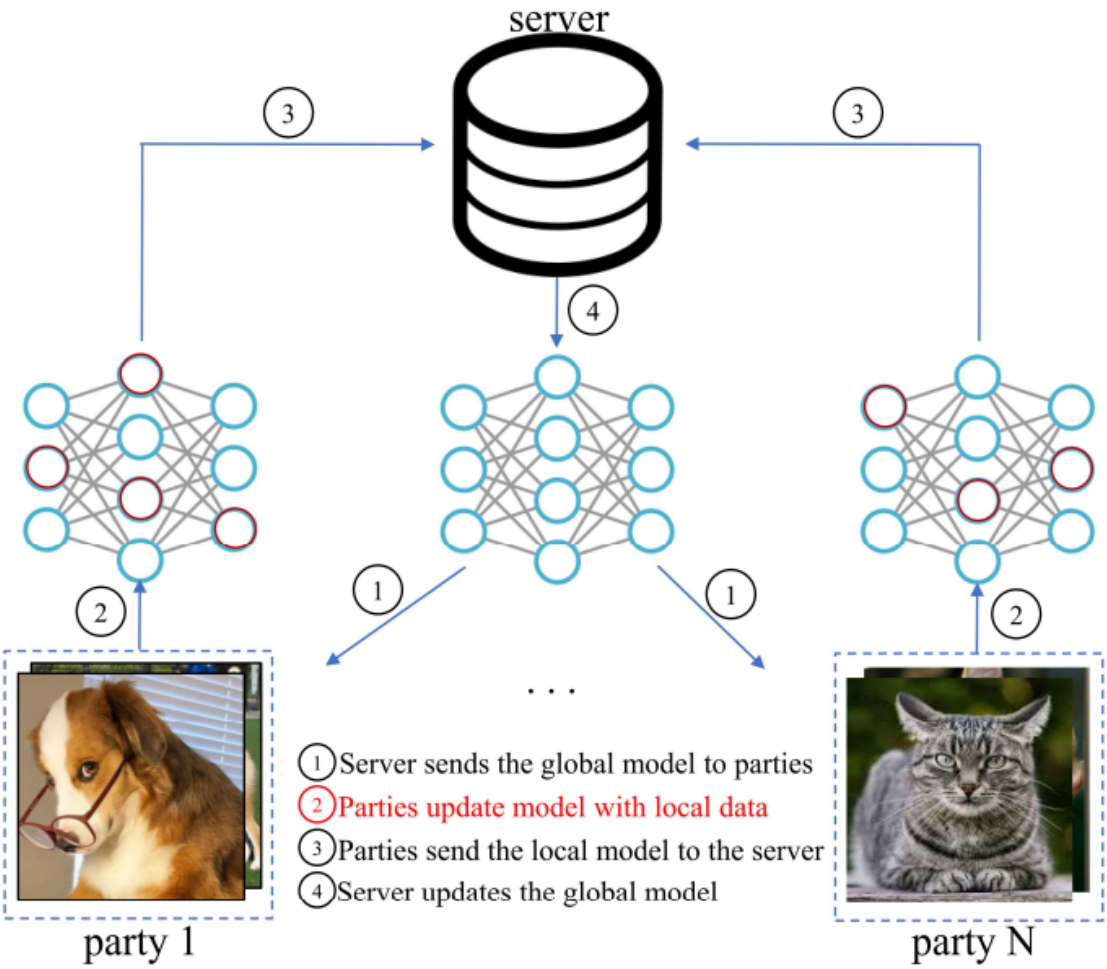

图 1: FedAvg 流程示意图

每一轮的 FedAvg 有四个步骤。首先，服务器向各方发送一个全局模型。第二，各方进行随机梯度下降用本地数据来更新本地模型。第三，本地模型发送到中央服务器。最后，服务器对模型的权重进行平均产生一个全局模型，用于下一轮的训练。

已经有很多的研究试图在 Non-IID 数据上改进 FedAvg。这些研究可以分为两类：改进本地训练过程和改进聚合方式。MOON 框架属于第一类。

对于改进本地训练过程的研究，FedProx 在本地训练时将根据全局模型和局部模型之间的二范数距离计算本地参数更新方向和距离。因此，在局部训练过程中，局部模型的更新会受到限制。SCAFFOLD[4]通过引入控制变量来纠正本地模型更新，这个控制变量在本地训练期间动态更新与设置。然而，PedProx 和 SCAFFOLD 在图像数据集上的深度学习模型有效性还没有彻底的研究。MOON 对这两种框架进行实验，发现这些框架相比于 FedAvg 没有本质的提升。

对于改进聚合过程的研究，FedMA[5]使用贝叶斯参数方法，进行逐层的匹配和权重平均。FedAvgM[6]在更新全局模型时引入动量。FedNova[7]在平均化之前对局部更新进行归一化处理。

2.2 对比学习

自监督学习是近期热门的研究方向，其试图从未标记的数据中学到较好的数据表示。在这些研究中，对比学习方法在学习视觉表征方面取得了最好的成果。对比学习的关键思想是减少同一图像不同增强试图表示之间的距离，并且增加不同图像增强试图表示之间的距离。

一个典型的对比学习框架是 SimCLR[8]。给定一个图像 x，SimCLR 首先使用不同的数据增强方法创建该图像的两个相关视图，分别表示为 xi 和 xj。而后利用对比性损失用于图像表征向量，试图使得

同一图像的不同增强视图之间的一致性最大。

除了 SimCLR，还有其他对比学习框架，如 CPC[9]，CMC[10]和 MoCo[11]。这些框架的对比学习基本思想是相似的：从不同图像中获得的表征应该彼此远离，而从同一图像中得到的表征应该彼此接近。

在 MOON 文章中，作者提出使用模型对比学习来比较不同本地模型学习的表征，从而制约局部模型的更新方向。

3 模型对比联邦学习

3.1 本文方法概述

MOON 框架主要在于将本地训练的损失函数进行改造，从而引导模型的更新往全局聚合方向靠近，从而提升局部模型聚合的效果。主要的框架如下图 2 所示：

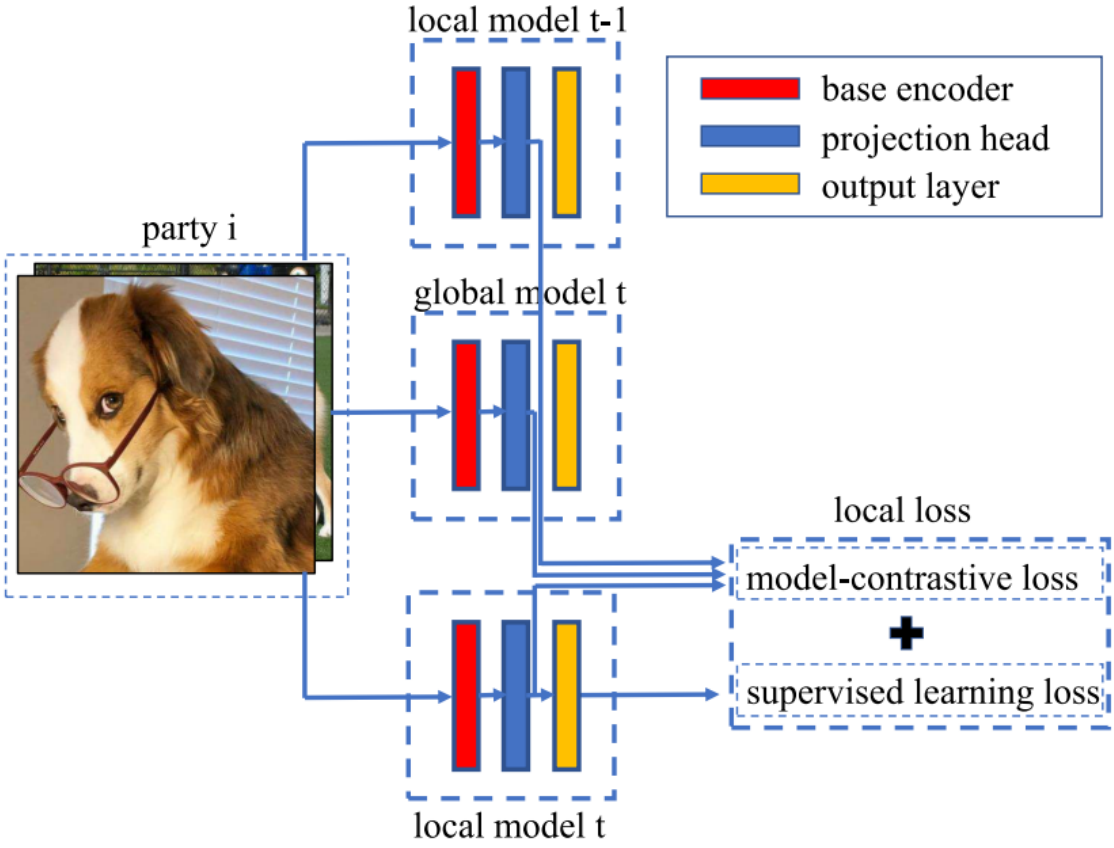

图 2: MOON 框架示意图

如图所示，可以看出在每个分部 party i 中，在局部训练模型可以从形式化分成三个部分，分别是基础编码器层、预测头层和输出层。MOON 认为，局部模型预测头层输出的向量表示能够表示当前局部模型的表征能力，而全局模型预测头层输出的向量能够表示当前全局模型的表征能力。根据对比学习的启发，MOON 定义了两个损失函数，一个是模型对比损失和监督学习损失。

此处的模型可以是很多种模型，可以是简单的全连接网络，可以是简单的卷积神经网络等等，都可以形式化的将使用的模型划分成上述的三个部分，从而进行模型对比学习。

3.2 损失函数定义

MOON 的局部模型训练损失函数分成模型对比损失和监督学习损失。具体定义如下 3 所示：

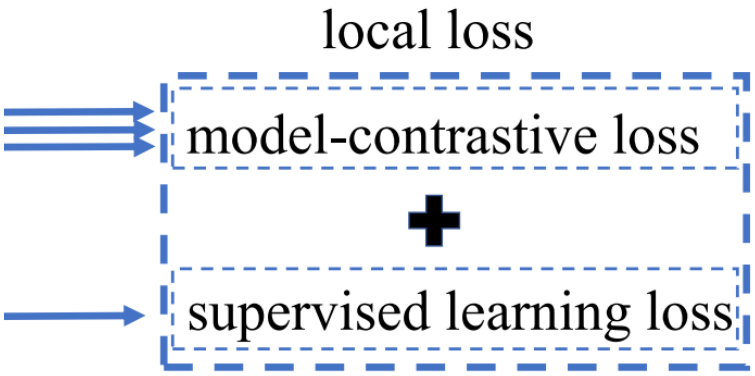

图 3: 损失函数示意图

监督学习损失是根据本地数据标签和预测标签的差别算出的损失，是常规的机器学习训练损失，此处不再赘述。

重点在于模型对比损失，模型对比损失又分为两个部分，一个是与全局模型的差别损失，一个是与上一轮局部模型的近似损失。MOON 的目的是，让本地训练模型与全局聚合模型愈发接近，与上一轮局部模型愈发疏远。

4 复现细节

4.1 核心代码

在复现过程中，借鉴了一部分原论文提供的代码，借鉴了源代码中在一台电脑中实现多个分部分别训练的部分。其他的神经网络模型部分是自己改进和实现的。

整个框架的伪代码流程如下：

Procedure 1 The MOON framework

Input: number of communication rounds T, number of parties N, number of local epochs E, learning rate η, hyper-parameter μ,

Output: The final model w^T

Server executes:

initialize w^0

for $t = 1 \rightarrow T$ **do**

  for $i = 1 \rightarrow N$ **do**

    send the global model w^t *to* P_i

    $\mathrm{w}_i^t \leftarrow PartyLocalTraining(i, w^t)$

  end

  $\mathrm{w}^{t+1} \leftarrow \sum_{k=1}^{N} \frac{|D^i|}{|D|} w_k^t$

end

return w^T

PartyLocalTraining(i, w^t) :

$\mathrm{w}_i^t \leftarrow w^t$

for *epoch* $i = 1 \rightarrow E$ **do**

  for *each batch* $b = \{x, y\}$ *of* D^i **do**

    $\mathrm{l}_{sup} \leftarrow CrossEntropyLoss(F_{w_i^t}(x), y)$

    $\mathrm{z} \leftarrow R_{w_i^t}(x)$

    $\mathrm{z}_{glob} \leftarrow R_{w^t}(x)$

    $\mathrm{z}_{prev} \leftarrow R_{w_i^{t-1}}(x)$

    $\mathrm{l}_{con} \leftarrow -log\frac{exp(sim(z, z_{glob}))}{exp(sim(z, z_{glob})) + exp(sim(z, z_{prev}))}$

    $\mathrm{l} \leftarrow l_{sup} + \mu l_{con}$

    $\mathrm{w}_i^t \leftarrow w_i^t - \eta \nabla l$

  end

end

return w_i^t *to server*

训练的关键代码如下图 4 所示，我们将模型的表征输出和预测输出分别存放到变量中，而后将当前全局模型的表征输出也保存下来。对于当前模型的表征和全局模型表征计算余弦相似度，而后计算当前的模型表征和上一轮训练模型的模型表征之间的余弦相似度，这一步是为了计算模型对比损失。由于二者都是张量形式的数据，所以二者损失使用交叉熵损失的方式来计算。

同时，由于当前模型的预测输出也获取了，我们同样能够计算监督损失。

最后，根据两个损失共同训练局部模型，能够使得局部模型优化方向不会受到数据偏移的严重影响，使得模型训练更加稳定，

```python
# 在每一个训练轮中，计算损失，进行梯度回传
for epoch in range(epochs):
    epoch_loss_collector = []
    epoch_loss1_collector = []
    epoch_loss2_collector = []
    for batch_idx, (x, target) in enumerate(train_dataloader):
        x, target = x.cuda(), target.cuda()

        optimizer.zero_grad()
        x.requires_grad = False
        target.requires_grad = False
        target = target.long()

        _, pro1, out = net(x) # pro1是模型的特征表示、out是模型的输出
        _, pro2, _ = global_net(x) # pro2是当前全局模型的特征表示

        posi = cos(pro1, pro2) # 计算当前模型和全局模型的余弦相似度
        logits = posi.reshape(-1,1)
        # 对于上一轮模型
        for previous_net in previous_nets:
            previous_net.cuda()
            _, pro3, _ = previous_net(x) # 得到一个上一轮模型的特征表示
            nega = cos(pro1, pro3) # 同样计算当前模型和上一轮模型的余弦相似度
            logits = torch.cat((logits, nega.reshape(-1,1)), dim=1)

            previous_net.to('cpu')

        logits /= temperature
        labels = torch.zeros(x.size(0)).cuda().long()

        loss2 = mu * criterion(logits, labels) # loss2是模型对比损失，需要尽可能的labels即0接近


        loss1 = criterion(out, target) # loss1就是普通的分类损失
        loss = loss1 + loss2

        loss.backward()
        optimizer.step()

        cnt += 1
        epoch_loss_collector.append(loss.item())
        epoch_loss1_collector.append(loss1.item())
        epoch_loss2_collector.append(loss2.item())

    epoch_loss = sum(epoch_loss_collector) / len(epoch_loss_collector)
    epoch_loss1 = sum(epoch_loss1_collector) / len(epoch_loss1_collector)
    epoch_loss2 = sum(epoch_loss2_collector) / len(epoch_loss2_collector)
    logger.info('Epoch: %d Loss: %f Loss1: %f Loss2: %f' % (epoch, epoch_loss, epoch_loss1, epoch_loss2))
```

图 4: 训练部分核心代码实现

4.2 实验环境搭建

本文的实验环境搭建在租用的云服务器上, 使用的是一张 3090 显卡进行计算，使用 Anaconda 虚拟环境进行 python 环境配置。安装了 PyTorch、torchvision 和 scikit-learn 等一些机器学习常用的包。

4.3 创新点

非独立同分布是联邦学习有效性的重要挑战。本文的创新点在于，创新性地引入了模型对比学习的思想，MOON 架构通过最大化本地模型表征与全局模型表征的一致性来纠正本地目标的偏离。通过比较不同模型学到的模型表征进行模型对比学习，使得局部模型的更新方向与全局更加一致，由此缓解 Non-IID 引发的问题。

此外，由于这种 MOON 框架对数据的格式并无特殊要求，所以 MOON 框架可以延拓到其他领域，例如语音识别、文字处理等领域。

5 实验结果分析

以下两个表分别是使用 FedAvg 和 MOON 两种联邦学习框架在 CIFAR100 和 CIFAR10 两个数据集上进行实验的结果。

表 1: CIFAR-100 实验结果

Accuracy(%)	FedAvg	MOON
IID	70.68	69.47
Non-IID	64.80	**67.71**

表 2: CIFAR-10 实验结果

Accuracy(%)	FedAvg	MOON
IID	71.78	71.15
Non-IID	66.30	**68.54**

从表中可以看出，在独立同分布的情况下，MOON 的效果与 FedAvg 相差无几。而在非独立同分布的情况时，MOON 的准确度要比 FedAvg 高，说明 MOON 这种联邦学习框架能够有效地克服非独立同分布带来的训练效果差的问题。

而后对不同分部数量对准确度的影响进行了实验。

表 3: 不同分部数量实验结果

Accuracy(%)	4	8	16	32
Non-IID	**70.68**	69.47	67.55	63.24

可以发现，分布数量过多时会对整体性能产生影响，分布数量过多会将模型引向更复杂的分部情况，更难从零散的数据分布中学习到我们希望学到的模型知识。

6 总结与展望

联邦学习作为当前解决许多领域中数据孤岛问题的前沿方法，对于非独立同分布的数据分布时的模型训练不尽如人意。为了提高联邦深度学习模型在非独立同分布数据集上的性能，文章提出了联邦模型对比学习的一种简单而有效的 MOON 框架。MOON 框架中引入了一种新的学习概念，即模型级的对比学习。通过广泛的实验，可以发现 MOON 在非独立同分布的数据下，能够取得比当前最先进方法的明显改进。

但本次复现仅局限于一些图像分类任务，从本质上而言，本文提出的框架能应用于更多领域，对于非视觉问题也能够有良好的效果，未来希望能够在 MOON 框架的指引下，在更多不同的机器学习领域上使用 MOON 框架进行稳定性和准确性的提升。

参考文献

[1] LI Q, HE B, SONG D. Model-Contrastive Federated Learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[2] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-Efficient Learning of Deep Networks from Decentralized Data[C]//. 2016.

[3] LI T, SAHU A K, ZAHEER M, et al. Federated Optimization in Heterogeneous Networks[J]., 2018.

[4] KARIMIREDDY S P, KALE S, MOHRI M, et al. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning[Z]. 2019.

[5] WANG H, YUROCHKIN M, SUN Y, et al. Federated Learning with Matched Averaging[J]., 2020.

[6] HSU T, QI H, BROWN M. Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification[Z]. 2019.

[7] WANG J, LIU Q, LIANG H, et al. Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization[Z]. 2020.

[8] CHEN T, KORNBLITH S, NOROUZI M, et al. A Simple Framework for Contrastive Learning of Visual Representations[J]., 2020.

[9] OORD A, LI Y, VINYALS O. Representation Learning with Contrastive Predictive Coding[J]., 2018.

[10] TIAN Y, KRISHNAN D, ISOLA P. Contrastive Multiview Coding[J]., 2019.

[11] HE K, FAN H, WU Y, et al. Momentum Contrast for Unsupervised Visual Representation Learning[J]., 2019.