

移动边缘技术与联邦学习

Koorye

2024 年 3 月 27 日

目录

1	介绍	2
2	背景	3
2.1	深度学习的发展历史	3
2.2	移动边缘技术的发展历史	4
3	方法	4
3.1	权重平均的联邦学习方法	4
3.1.1	Federated Averaging (FedAvg)	4
3.1.2	Federated Proximal (FedProx)	5
3.1.3	Federated Matched Averaging (FedMA)	5
3.1.4	总结	6
3.2	个性化的联邦学习方法	6
3.2.1	Federated Learning with Personalization Layers (Fed- Per)	6
3.2.2	Federated Multi-Task Learning (MOCHA)	7
3.2.3	Personalized Federated HyperNetworks (pFedHN)	7
3.2.4	其他个性化的联邦学习方法	8
3.2.5	总结	8
4	应用	9
4.1	联邦学习在移动设备上的应用	9
4.2	联邦学习在车辆驾驶领域的应用	10
4.3	联邦学习在其他领域的应用	11
5	讨论	12
5.1	联邦学习的优缺点	12
5.2	联邦学习的发展方向	13
6	结论	13

摘要

随着深度学习的发展，大量的深度学习大模型被提出，随时随地方便了人们的生产和生活。然而，这些模型在计算资源和存储资源上的需求也在不断增加。移动边缘技术是一种新兴的技术，它可以将计算资源和存储资源从云端迁移到边缘设备上，从而减少网络延迟和能耗，这为大模型的训练和部署提供了新的思路。本文介绍了一种深度学习中的移动边缘技术——“联邦学习”，这种技术是一种分布式学习技术，通过在不同设备上独立训练一个局部模型，并只进行少量的信息交换，达到训练全局模型的目的。联邦学习技术可以提高模型的训练效率，同时减少计算和存储成本、保护用户隐私，这种技术被广泛应用于移动设备、物联网设备等边缘设备上，如智能手机、智能车辆等，为人们提供更好的服务。本文希望通过移动边缘技术中联邦学习的介绍，为大家提供一个新的思路，帮助大家更好地理解深度学习和移动边缘技术之间的关系，为未来的研究提供一些参考。

关键词：深度学习、移动边缘技术、联邦学习。

1 介绍

随着深度学习的发展，近年来大量模型被提出和应用，如 ResNet[9]、BERT[5] 等。这些模型在计算机视觉、自然语言处理等领域取得了巨大的成功。同时，大模型的出现也使得深度学习的应用范围更加广泛。如 ChatGPT、GPT-4 等模型在对话生成、文本生成等任务上取得了很好的效果。同时，研究者还利用这些大模型来辅助解决复杂任务，如自动驾驶 [8]、医疗诊断 [13] 等，大模型的出现使得深度学习的应用范围更加广泛。

然而，大模型的出现也带来了一些问题，如计算和存储成本问题。大模型的训练和推理需要大量的计算资源，例如，GPT-4 的训练中使用了约 25,000 个 A100 GPU，耗费 90 到 100 天，训练一次的成本就达到 6300 万美元。另外，大模型的存储也是一个问题。大模型的存储和加载需要大量的内存，而这些内存也是非常昂贵的。此外，大模型也涉及到安全、隐私问题，如用户的数据可能会被泄露。因此，大模型训练和部署的成本和安全性问题亟待解决。

移动边缘技术 [17, 1] 是一种新兴的技术，它将计算资源和存储资源放在网络边缘，使得用户可以更快地访问数据和服务。移动边缘技术可以提高用户体验，减少网络延迟，降低网络拥塞。同时，移动边缘技术还可以提高

网络的安全性，保护用户的隐私。因此，移动边缘技术在智能手机、物联网、自动驾驶等领域有着广泛的应用。

目前，研究者们已经开始研究如何将深度学习模型部署到移动边缘设备上。联邦学习 [27, 16] 是一种将深度学习模型部署到移动边缘设备上的方法，本质上是通过多个用户设备共同训练一个代表所有用户设备的全局模型，而训练的过程不需要用户数据的交换。这种方法可以在不泄露用户数据的情况下，训练深度学习模型，并减少数据传输的开销、计算和存储成本等问题，联邦学习的优势使得它在移动边缘设备上有着广泛的应用前景。

截至目前，联邦学习作为一种隐私保护的重要解决方案，已在全球范围内得到广泛关注和应用。在联邦学习领域，中国和美国是联邦学习论文发布量最多的两个国家，高被引论文的六成以上来自这两国。此外，各国研究者提出了各种联邦学习开源框架，其中 OpenMined 推出的 Pysyft[29]、微众银行的 FATE[11] 和谷歌的 TFF 框架热度居于全球前三位，这些框架为联邦学习的研究和应用提供了重要的支持。此外，各国政府和企业也在积极推动联邦学习的发展，如美国的联邦数据战略、中国的数据安全法等，这些政策和法规为联邦学习的发展提供了政策支持。

本文将介绍联邦学习技术的背景，一些联邦学习的方法，以及联邦学习在一些应用场景中的应用，最后讨论联邦学习的优缺点和未来发展方向。

2 背景

这一部分将介绍联邦学习技术的一些背景知识，包括深度学习和移动边缘技术的发展历史。

2.1 深度学习的发展历史

深度学习是机器学习的一个分支，它模仿人类大脑的神经网络，通过多层神经元进行信息处理。深度学习的发展可以追溯到上世纪 50 年代，但直到 2012 年，AlexNet 的出现才使得深度学习开始受到广泛关注。AlexNet[4] 是一种卷积神经网络，它在 ImageNet[4] 图像识别比赛中取得了巨大的成功。2016 年，ResNet 模型的出现进一步推动了深度学习的发展。ResNet 是一种残差神经网络，它通过残差块的设计，解决了深度神经网络训练过程中的梯度消失和梯度爆炸问题。2018 年，BERT 模型的出现进一步推动了自然语言处理领域的发展。BERT 是一种预训练的语言模型，它在多项自然语

言处理任务上取得了很好的效果。2019 年，GPT-2 模型的出现进一步推动了文本生成领域的发展。GPT-2[20] 是一种基于 Transformer 的语言模型，它在文本生成任务上取得了很好的效果。这些模型的出现使得深度学习在计算机视觉、自然语言处理等领域取得了巨大的成功。

2.2 移动边缘技术的发展历史

移动边缘技术是一种新兴的技术，它将计算资源和存储资源放在网络边缘，使得用户可以更快地访问数据和服务。移动边缘技术最早可以追溯到 MEC (Mobile Edge Computing) [24] 技术，它是一种将计算资源和存储资源放在无线接入网的边缘，使得用户可以更快地访问数据和服务。MEC 技术可以提高用户体验，减少网络延迟，降低网络拥塞。另外，Fog Computing[3] 技术也是一种移动边缘技术，它是一种将计算资源和存储资源放在网络边缘的技术，使得用户可以更快地访问数据和服务。Fog Computing 技术可以提高用户体验，减少网络延迟，降低网络拥塞。移动边缘技术在智能手机、物联网、自动驾驶等领域有着广泛的应用。

3 方法

这一部分将介绍联邦学习技术一些方法的分类，包括权重平均的联邦学习方法和个性化的联邦学习方法。

3.1 权重平均的联邦学习方法

权重平均的联邦学习方法是指在联邦学习过程中，中心节点维护一个全局模型，每个参与方在本地训练模型，然后将本地模型上传到中心节点，中心节点对本地模型进行聚合，得到全局模型。下面介绍一些典型的权重平均的联邦学习方法。

3.1.1 Federated Averaging (FedAvg)

FedAvg[18] 是一个经典的联邦学习方法，它的基本思想是将本地模型的参数上传到服务器，服务器计算所有模型参数的平均值，然后将这个平均值广播回所有本地设备。这个过程可以迭代多次，直到收敛。FedAvg 的过程可以表示为算法 1。

Algorithm 1 FedAvg

```

1: 初始化  $W$ 
2: for  $t = 1, 2, \dots, T$  do
3:    $W_k \leftarrow \arg \min_W \mathcal{L}(W; D_k)$  for  $k = 1, 2, \dots, K$ 
4:    $W \leftarrow \frac{1}{K} \sum_{k=1}^K W_k$ 
5: end for

```

其中, W_k 表示第 k 个参与方的本地模型, $\mathcal{L}(W; D_k)$ 表示本地模型在本地数据集 D_k 上的损失函数。

3.1.2 Federated Proximal (FedProx)

当数据分布不均匀时, FedAvg 可能会导致一些参与方的模型性能下降。为了解决这个问题, 研究者提出了 FedProx[12] 方法。FedProx 的基本思想是在 FedAvg 的基础上增加一个正则项, 用于约束本地模型和全局模型的差异。FedProx 的过程可以表示为算法 2。

Algorithm 2 FedProx

```

1: 初始化  $W_0$ 
2: for  $t = 1, 2, \dots, T$  do
3:    $W_t^k \leftarrow \arg \min_w \mathcal{L}(W_t; D_k) + \frac{\lambda}{2} \|W - W_{t-1}\|_2^2$  for  $k = 1, 2, \dots, K$ 
4:    $W_t \leftarrow \frac{1}{K} \sum_{k=1}^K W_t^k$ 
5: end for

```

其中, W_t^k 表示第 t 轮训练的第 k 个参与方的权重。 λ 是一个超参数, 用于控制正则项的权重。正则项 $\frac{\lambda}{2} \|w - w_{t-1}\|_2^2$ 用于约束本地模型和全局模型的差异, 从而减小模型性能的下降。

3.1.3 Federated Matched Averaging (FedMA)

FedMA[26] 是一个基于匹配的联邦学习方法, 它的基本思想是逐层匹配本地模型和全局模型的参数。首先, 服务器收集所有本地模型的第一层权重, 聚合后广播回所有本地设备。然后, 本地设备冻结第一层权重, 训练第二层权重, 上传到服务器, 服务器聚合后广播回所有本地设备。这一过程不断重复, 直到所有层被训练。FedMA 的过程可以表示为算法 3。

Algorithm 3 FedMA

```

1: 初始化  $W = \{w^{(1)}, \dots, W^{(N)}\}$ 
2:  $n = 1$ 
3: while  $n \leq N$  do
4:    $\{\Pi_k\}_{k=1}^K = \text{BBP-MAP}(\{W_k^{(n)}\}_{j=1}^J)$ 
5:    $W^{(n)} \leftarrow \frac{1}{K} \sum_{k=1}^K W_k^{(n)} \Pi_j^T$ 
6:   for  $k = 1, 2, \dots, K$  do
7:      $W_k^{(n+1)} \leftarrow \Pi_k W_k^{(n+1)}$ 
8:      $W_k^{(n+1)} \leftarrow \arg \min_W \mathcal{L}(W; D_k)$  with  $W^{(n)}$  frozen
9:   end for
10: end while

```

其中, $W^{(n)}$ 表示全局模型第 n 层的权重, $W_k^{(n)}$ 表示第 k 个参与方的第 n 层权重。BBP-MAP 是一个用于匹配本地模型和全局模型的函数。 Π_k 是一个匹配矩阵, 用于匹配本地模型和全局模型的参数。

3.1.4 总结

总的来说, 权重平均的联邦学习方法是一种简单有效的联邦学习方法, 它的基本思想是在本地训练模型, 然后上传到服务器, 服务器聚合所有模型的参数, 得到全局模型。这种方法的优点是简单, 然而它也存在许多问题。例如, 这种方法的同步频率和性能难以平衡, 同步频率越高时, 性能越好, 但通信开销也越大。

3.2 个性化的联邦学习方法

个性化的联邦学习方法是指在联邦学习过程中, 为每个参与方训练一个个性化的模型, 从而适应不同的数据和需求。下面介绍一些典型的个性化的联邦学习方法。

3.2.1 Federated Learning with Personalization Layers (FedPer)

FedPer[2] 是一个基于个性化的联邦学习方法, 它的基本思想是在全局模型中增加一个个性化层, 用于适应不同的参与方。FedPer 的过程可以表示为算法 4。

Algorithm 4 FedPer

```

1: 初始化  $W, W_{P_k}$ 
2: for  $t = 1, 2, \dots, T$  do
3:    $(W_k, W_{P_k}) \leftarrow \arg \min_W \mathcal{L}(W, W_{P_k}; D_k)$  for  $k = 1, 2, \dots, K$ 
4:    $W \leftarrow \frac{1}{K} \sum_{k=1}^K W_k$ 
5: end for

```

其中, W_{P_k} 表示第 k 个参与方的个性化层。 $\mathcal{L}(W, W_{P_k}; D_k)$ 表示基本层和个性化层组合后的第 k 个参与方模型在本地数据集 D_k 上的损失函数。

3.2.2 Federated Multi-Task Learning (MOCHA)

MOCHA[25] 是一个基于多任务学习的联邦学习方法, 它的基本思想是在本地训练模型时, 同时训练多个任务, 从而提高模型的泛化能力。MOCHA 的过程可以表示为算法 5。

Algorithm 5 MOCHA

```

1: 初始化  $W$ 
2: for  $t = 1, 2, \dots, T$  do
3:    $W_k \leftarrow \arg \min_W \mathcal{L}(W; D_k) + \mathcal{R}(W, \Omega)$  for  $k = 1, 2, \dots, K$ 
4:    $W \leftarrow \frac{1}{K} \sum_{k=1}^K W_k$ 
5: end for

```

其中, $\mathcal{R}(W, \Omega) = \lambda_1 \text{tr}(W\Omega W^T) + \lambda_2 \|W\|^2$ 。 Ω 是一个可学习矩阵, λ_1, λ_2 是 2 个正则化系数的权重。正则项 $\mathcal{R}(W, \Omega)$ 起到约束不相关任务之间共享信息的作用。

3.2.3 Personalized Federated HyperNetworks (pFedHN)

pFedHN[22] 是一个基于超网络的个性化联邦学习方法, 它的基本思想是在全局模型中增加一个超网络, 用于生成个性化模型的参数。通过超网络, 可以为每个参与方生成独立且多样的个性化的模型。pFedHN 的过程可以表示为算法 6。

其中, $h(v_k; \phi)$ 是一个超网络, 用于生成个性化模型的参数。 ϕ 是超网络的参数, v_k 是超网络中每个参与者的可学习特征。

Algorithm 6 pFedHN

```

1: 初始化  $W, v_k, k = 1, 2, \dots, K$ 
2: for  $t = 1, 2, \dots, T$  do
3:   sample  $k \in [1, 2, \dots, K]$ 
4:    $W_k \leftarrow h(v_k; \phi)$  and  $\tilde{W}_k \leftarrow W_k$ 
5:    $\tilde{W}_k \leftarrow \arg \min_W \mathcal{L}(W; D_k)$ 
6:    $\Delta W_k = \tilde{W}_k - W_k$ 
7:    $\phi \leftarrow \phi - \alpha \nabla_{\phi} W_k^T \Delta W_k$ 
8:    $v_k \leftarrow v_k - \alpha \nabla_{v_k} \phi^T \nabla_{\phi} W_k^T \Delta W_k$ 
9: end for

```

3.2.4 其他个性化的联邦学习方法

除了上述方法外，还有一些其他个性化的联邦学习方法，例如基于元学习的联邦学习方法 [7]、基于迁移学习的联邦学习方法 [15]、基于知识蒸馏的联邦学习方法 [28] 等。基于元学习的联邦学习方法通过学习一个元学习器，用于生成个性化的模型。具体来说，通过元学习找到一个初始共享模型，之后各个参与方可以通过对其自己的数据执行一步或几步梯度下降来轻松地适应其本地数据集。基于迁移学习的联邦学习方法通过迁移学习的思想，将全局模型迁移到参与方。基于知识蒸馏的联邦学习方法通过集合各个参与方的局部预测值，通过博弈理论为各个参与方分配软预测标签，从而提高模型的泛化能力。

3.2.5 总结

总的来说，个性化的联邦学习方法是一种有效的联邦学习方法，它的基本思想是为每个参与方训练一个个性化的模型，在相互协作，共同训练一个全局模型的同时，适应各个参与方不同的数据和需求。这种方法的优点是可以提高模型的泛化能力，适应不同的数据和需求，然而它也存在一些问题。例如，这种方法的计算复杂度较高，需要更多的计算资源。

4 应用

本章将介绍联邦学习技术在一些应用场景中的应用，包括移动设备上的应用、车辆驾驶领域的应用等。

4.1 联邦学习在移动设备上的应用

移动设备上的语音助手，如 Siri、Google Assistant 等，已经成为人们日常生活中的重要组成部分。这些语音助手可以帮助用户完成一些简单的任务，如查询天气、播放音乐等。然而，由于模型需要处理大量的用户数据，用户的隐私安全问题成为了一个严重的问题。传统的移动模型的训练需要大量用户数据，这一过程通常会将用户的数据上传到云端进行处理，这样会导致用户的隐私数据泄露的风险。

为了解决这个问题，研究者提出了使用联邦学习技术部署在移动设备上方法 [10, 14]。这种方法的基本思想是在移动设备上训练一个本地模型，然后将本地模型上传到服务器，服务器对本地模型进行聚合，得到全局模型。这样可以避免用户的隐私数据泄露。联邦学习在移动设备上的应用可以提高用户的隐私安全性，同时提高语音助手的性能。

如图 1 所示，传统的模型学习方法需要将用户的数据上传到云端进行处理，这样会导致用户的隐私数据泄露的风险。而联邦学习技术可以在移动设备上进行模型训练，避免了用户的隐私数据泄露的风险。联邦学习在移动设备上的应用可以分为以下几个步骤：

1. 移动设备上的本地模型训练：移动设备上的本地模型使用用户的语音数据进行训练。
2. 本地模型上传到服务器：本地模型将训练好的模型上传到服务器。
3. 服务器对本地模型进行聚合：服务器对本地模型进行聚合，得到全局模型。
4. 全局模型下发到移动设备：服务器将全局模型下发到移动设备，移动设备使用全局模型进行推理。
5. 移动设备上的本地模型更新：移动设备上的本地模型使用全局模型进行推理，然后将推理结果上传到服务器，服务器使用这些结果更新全局模型。

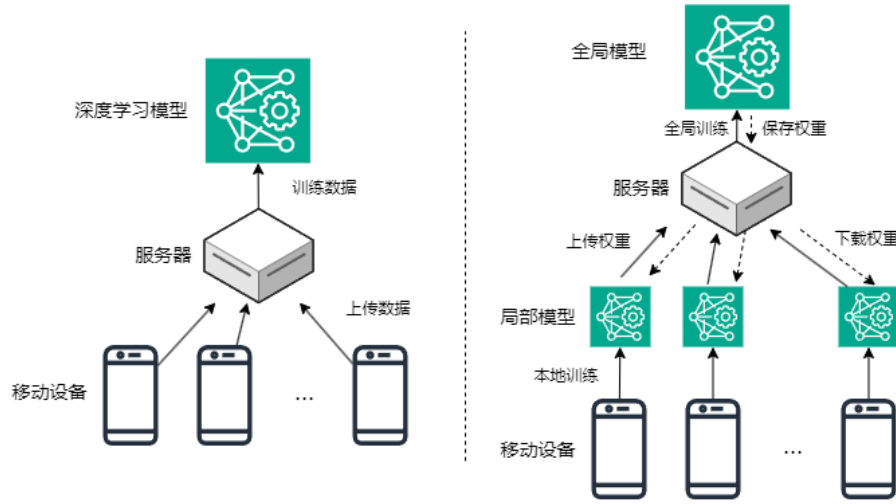


图 1: 联邦学习在移动设备上的应用

6. 重复步骤 2-5，直到模型收敛。

4.2 联邦学习在车辆驾驶领域的应用

车辆驾驶是一个复杂的任务，需要驾驶员不断地观察周围的环境，做出相应的决策。传统的车辆驾驶系统通常使用固定的规则来控制车辆的行驶，这样会导致车辆的行驶效果不佳。为了提高车辆的行驶效果，传统的想法是收集大量的车辆数据，然后使用这些数据训练一个模型，来控制车辆的行驶。然而，由于车辆驾驶环境的复杂性和实时性，传统的方法很难满足实际的需求。

为了解决这个问题，研究者提出了使用联邦学习技术在车辆驾驶领域的应用 [19, 6]。联邦学习技术可以在车辆之间共享模型，同时为每辆车提供个性化的模型。这样可以提高车辆的行驶效果，同时减少车辆之间的通信开销。

如图 2 所示，传统的模型学习方法需要将车辆的数据上传到云端进行处理，这样会导致通信开销较大，且网络传输延迟时间长。而联邦学习技术可以为每辆车提供个性化的模型，同时直接与本地局部模型通信，大量减少延迟，满足行驶环境下的实时性需求。联邦学习在车辆驾驶领域的应用可以分为以下几个步骤：

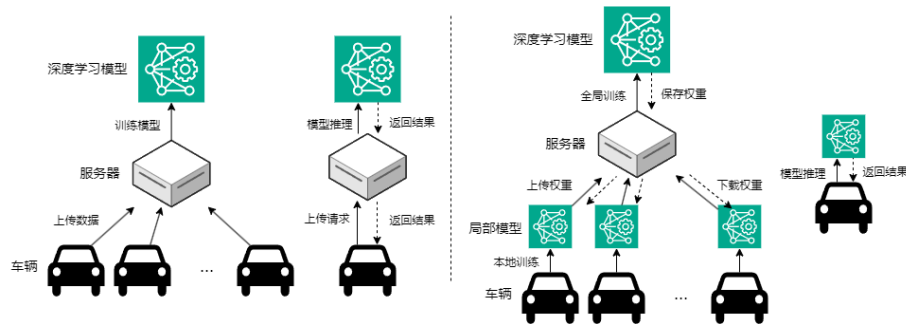


图 2: 联邦学习在车辆驾驶领域的应用

1. 车辆上的本地模型训练：每辆车使用自己的数据训练本地模型。
2. 本地模型上传到服务器：每辆车将训练好的模型上传到服务器。
3. 服务器对本地模型进行聚合：服务器对本地模型进行聚合，得到全局模型。
4. 全局模型下发到车辆：服务器将全局模型下发到车辆，车辆使用全局模型进行推理。
5. 车辆上的本地模型更新：车辆使用全局模型进行推理，然后将推理结果上传到服务器，服务器使用这些结果更新全局模型。
6. 重复步骤 2-5，直到模型收敛。
7. 车辆上的本地模型推理：车辆使用本地局部模型进行推理，快速得到控制车辆行驶的决策。

4.3 联邦学习在其他领域的应用

除了在移动设备和车辆驾驶领域的应用，联邦学习技术还可以在其他领域进行应用。例如，在金融领域，联邦学习技术可以用于风险评估、信用评分等任务。考虑到各个金融机构之间的数据隐私性，联邦学习技术可以在不泄露用户隐私的情况下，共享模型，提高金融机构的风险评估能力 [21]。此外，在医疗领域，联邦学习技术可以用于医疗影像诊断、疾病预测等任务。医疗数据的隐私性是医疗领域的一个重要问题，联邦学习技术可以在不泄露患者隐私的情况下，共享模型，提高医疗诊断的准确性 [23]。

5 讨论

本章节将讨论联邦学习的优缺点和未来发展方向。

5.1 联邦学习的优缺点

联邦学习是一种新兴的机器学习技术，它与移动边缘技术相结合，可以在不泄露用户隐私的情况下共享模型，通过分布式训练提高模型的性能。联邦学习技术有以下优点：

1. 隐私保护：联邦学习技术可以在不泄露用户隐私的情况下共享模型，保护用户的隐私数据。
2. 分布式训练：联邦学习技术可以在多个设备上同时进行模型训练，提高模型的性能。
3. 降低通信开销：联邦学习技术可以在本地设备上同时进行模型训练，减少了数据传输的开销。
4. 个性化模型：联邦学习技术可以为每个设备提供个性化的模型，提高模型的性能。
5. 实时性：联邦学习技术可以在本地设备上同时进行模型训练和推理，满足实时性需求。
6. 扩展性：联邦学习技术可以扩展到大规模的设备上，提高模型的性能。

然而，联邦学习技术也存在一些缺点：

1. 模型收敛速度慢：由于联邦学习技术需要在多个设备上同时进行模型训练，模型的收敛速度较慢。
2. 模型性能下降：由于联邦学习技术需要在本地设备上同时进行模型训练，模型的性能可能会下降。
3. 数据不平衡：由于联邦学习技术需要在多个设备上同时进行模型训练，数据的分布可能会不平衡。
4. 安全性问题：由于联邦学习技术需要在多个设备上同时进行模型训练，模型的安全性可能会受到威胁。

5. 通信开销：由于联邦学习技术需要在多个设备上进行了模型训练，通信开销较大。
6. 难以调试：由于联邦学习技术需要在多个设备上进行了模型训练，模型的调试较为困难。

5.2 联邦学习的发展方向

作为一种新兴的技术，联邦学习尚不够成熟，还有很多问题需要解决。未来，联邦学习技术可能会朝着以下几个方向发展：

1. 模型优化：联邦学习方法目前的性能并不足以与传统的中心化学习方法相媲美，未来，联邦学习技术可能会进一步优化模型，提高模型的性能。
2. 隐私保护：联邦学习方法目前可能还存在一定的隐私问题，如通信传输中的隐私泄露等。未来，联邦学习技术可能会进一步提高隐私保护的水平。
3. 通信优化：联邦学习方法目前通信成本和时间开销不一，有的方法需要大量的通信开销。未来，联邦学习技术可能会进一步优化通信开销。
4. 安全性保障：联邦学习方法目前并不能保证安全性，可能会受到一些攻击，如网络攻击、对抗攻击等。未来，联邦学习技术可能会进一步提高安全性保障。

总的来说，联邦学习技术是一种新型机器学习技术。未来，联邦学习技术可能会进一步发展，为解决用户数据的分布式训练和隐私问题提供更好的解决方案。

6 结论

在大模型大规模普及的今天，用户数据的分布式训练和隐私问题变得尤为重要。联邦学习技术的出现，为解决这一问题提供了新的思路。联邦学习技术是一种与移动边缘技术结合的新型机器学习技术，这种技术可以在不泄露用户隐私的情况下共享模型，通过分布式训练提高模型的性能。本文介绍了联邦学习的背景、方法和应用，并讨论了联邦学习技术的优缺点与未来可能的发展方向，希望可以为联邦学习技术的研究和应用提供一些参考。

参考文献

- [1] Nasir Abbas, Yan Zhang, Amir Taherkordi, and Tor Skeie. Mobile edge computing: A survey. *IEEE Internet of Things Journal*, 5(1):450–465, 2017.
- [2] Manoj Ghuhane Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.
- [3] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. Fog computing and its role in the internet of things. *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pages 13–16, 2012.
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [6] Zhaoyang Du, Celimuge Wu, Tsutomu Yoshinaga, Kok-Lim Alvin Yau, Yusheng Ji, and Jie Li. Federated learning for vehicular internet of things: Recent advances and open issues. *IEEE Open Journal of the Computer Society*, 1:45–61, 2020.
- [7] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020.
- [8] Daocheng Fu, Xin Li, Licheng Wen, Min Dou, Pinlong Cai, Botian Shi, and Yu Qiao. Drive like a human: Rethinking autonomous driving with large language models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 910–919, 2024.

- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [10] Jiawen Kang, Zehui Xiong, Dusit Niyato, Yuze Zou, Yang Zhang, and Mohsen Guizani. Reliable federated learning for mobile networks. *IEEE Wireless Communications*, 27(2):72–80, 2020.
- [11] Ivan Kholod, Evgeny Yanaki, Dmitry Fomichev, Evgeniy Shalugin, Evgenia Novikova, Evgeny Filippov, and Mats Nordlund. Open-source federated learning frameworks for iot: A comparative review and analysis. *Sensors*, 21(1):167, 2020.
- [12] Tianyi Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks: A proximal federated averaging approach. *arXiv preprint arXiv:1812.06127*, 2020.
- [13] Valentin Liévin, Christoffer Egeberg Hother, Andreas Geert Motzfeldt, and Ole Winther. Can large language models reason about medical questions? *Patterns*, 2023.
- [14] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3):2031–2063, 2020.
- [15] Yang Liu, Yan Kang, Chaoping Xing, Tianjian Chen, and Qiang Yang. A secure federated transfer learning framework. *IEEE Intelligent Systems*, 35(4):70–82, 2020.
- [16] Priyanka Mary Mammen. Federated learning: Opportunities and challenges. *arXiv preprint arXiv:2101.05428*, 2021.
- [17] Yuyi Mao, Changsheng You, Jun Zhang, Kaibin Huang, and Khaled B Letaief. A survey on mobile edge computing: The communication per-

- spective. *IEEE communications surveys & tutorials*, 19(4):2322–2358, 2017.
- [18] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629*, 2017.
- [19] Jason Posner, Lewis Tseng, Moayad Aloqaily, and Yaser Jararweh. Federated learning in vehicular networks: Opportunities and solutions. *IEEE Network*, 35(2):152–159, 2021.
- [20] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
- [21] Yuris Mulya Saputra, Diep N Nguyen, Dinh Thai Hoang, Thang X Vu, Eryk Dutkiewicz, and Symeon Chatzinotas. Federated learning meets contract theory: Economic-efficiency framework for electric vehicle networks. *IEEE Transactions on Mobile Computing*, 21(8):2803–2817, 2020.
- [22] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *International Conference on Machine Learning*, pages 9489–9502. PMLR, 2021.
- [23] Micah J Sheller, Brandon Edwards, G Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R Colen, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1):12598, 2020.
- [24] Wenjia Shi, Jian Cao, Quan Zhang, and You Li. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, 2016.
- [25] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. *Advances in neural information processing systems*, 30, 2017.

- [26] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. *arXiv preprint arXiv:2002.06440*, 2020.
- [27] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, 2021.
- [28] Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-free knowledge distillation for heterogeneous federated learning. In *International conference on machine learning*, pages 12878–12889. PMLR, 2021.
- [29] Alexander Ziller, Andrew Trask, Antonio Lopardo, Benjamin Szymkow, Bobby Wagner, Emma Bluemke, Jean-Mickael Nounahon, Jonathan Passerat-Palmbach, Kritika Prakash, Nick Rose, et al. Pysyft: A library for easy federated learning. *Federated learning systems: Towards next-generation AI*, pages 111–139, 2021.