(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119005299 A
(43)申请公布日 2024.11.22

(21)申请号 202411066474.6

(22)申请日 2024.08.05

(71)申请人 河海大学
地址 211100 江苏省南京市江宁区佛城西路8号

(72)发明人 齐广飞 屈志昊 叶保留 谢在鹏

(74)专利代理机构 南京泉为知识产权代理事务所(特殊普通合伙) 32408
专利代理师 陈风平

(51)Int.Cl.
G06N 3/098(2023.01)
G06N 3/096(2023.01)
G06F 18/241(2023.01)

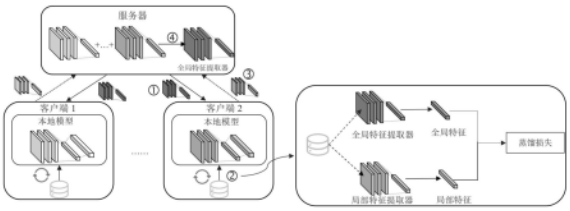
权利要求书3页 说明书9页 附图2页

(54)发明名称

一种基于知识蒸馏实现特征对齐的个性化联邦学习方法及系统

(57)摘要

本发明公开了一种基于知识蒸馏实现特征对齐的个性化联邦学习方法及系统,所述方法包括:客户端利用本地的数据进行蒸馏训练,将样本同时输入到全局特征提取器和局部特征提取器中得到样本的全局全局特征和局部特征;使用均方误差损失衡量全局特征和局部特征之间的差距;通过同时最小化分类损失和蒸馏损失,局部特征提取器同时学习到全局特征信息和局部特征信息,分类器学习本地信息;本地训练完成后,各个客户端将局部特征提取器上传到服务器,而分类器头保持在本地;服务器将各个客户端上传的局部特征提取器进行聚合,得到新一轮的全局特征提取器;此过程反复进行,直至模型收敛或者达到指定的模型精度。本发明提高了客户端个性化模型的精度。



CN 119005299 A

1.一种基于知识蒸馏实现特征对齐的个性化联邦学习方法,其特征在于,包括以下步骤:

服务器向客户端发送全局特征提取器参数;

客户端收到全局特征提取器参数后,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

客户端利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

本地训练完成后,各个客户端向服务器上传自己的局部特征提取器参数;

服务器收到客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

迭代以上过程,直至模型收敛或达到指定的模型精度。

2.根据权利要求1所述的方法,其特征在于,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,包括:

对于单个训练样本x_m,客户端利用保存的全局特征提取器副本对其进行处理,得到全局特征表示$f_{\phi_g}(x_m)$,其中,φ_g表示全局特征提取器,f_{ϕ_g}表示是由φ_g参数化的函数,$f_{\phi_g}(x_m)$是经过函数处理后的一个向量,包含了样本x_m的全局特征信息;

利用局部模型获得局部特征和预测输出,包括:

对于单个训练样本x_m,客户端利用局部模型θ_i对样本进行处理,得到局部特征$f_{\phi_i}(x_m)$和预测输出$\hat{y}_i$,其中,局部模型θ_i由局部特征提取器φ_i和局部分类器x_i组成,i表示第i个客户端,f_{ϕ_i}是由φ_i参数化的函数,$f_{\phi_i}(x_m)$表示经过函数处理后的一个向量,包含了样本x_m的局部特征信息,$\hat{y}_i$表示局部分类器χ_i对样本x_m的预测结果。

3.根据权利要求2所述的方法,其特征在于,蒸馏损失计算方法如下:

$$\ell_i^d(\phi_i) = \frac{1}{n_i}\sum_{m=1}^{n_i} \left\| f_{\phi_i}(x_m) - f_{\phi_g}(x_m) \right\|_2^2$$

其中,$\ell_i^d(\phi_i)$表示第i个客户端的蒸馏损失,n_i为第i个客户端的训练样本的数量。

4.根据权利要求2所述的方法,其特征在于,分类损失计算方法如下:

$$\ell_i^{ce}(\theta_i) = -\sum_j y_{ij} \log(\hat{y}_{ij})$$

其中,$\ell_i^{ce}(\theta_i)$表示第i个客户端的分类损失,y_{ij}是样本x_m属于类别j的真实标签的概率,$\hat{y}_{ij}$是局部模型预测的样本x_m属于类别j的标签的概率。

5.根据权利要求1所述的方法,其特征在于,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型,表示如下:

$$\min \ell_i(\theta_i) = \ell_i^{ce}(\theta_i) + \beta \cdot \ell_i^d(\phi_i)$$

其中,β为平衡分类损失ℓ_i^{ce}和蒸馏损失ℓ_i^{d}的超参数,用以控制全局特征提取器向局部特征提取器转移的知识程度。

6.根据权利要求1所述的方法,其特征在于,服务器聚合各客户端的局部特征提取器的方法如下:

$$\phi_g^{(t+1)}=\sum_{i=1}^{k}\alpha_i\phi_i^{(t)}$$

其中,$\phi_g^{(t+1)}$为第t+1个全局轮次的全局特征提取器,$\phi_i^{(t)}$为第i个客户端在第t个全局轮次的局部特征提取器,α_i为聚合权重,$\alpha_i=\frac{n_i}{\sum_{i=1}^{k}n_i}$,k为上传本地局部特征提取器的客户端总数目,$n_i$为第i个客户端拥有的数据量。

7.一种基于知识蒸馏实现特征对齐的个性化联邦学习方法系统,包括服务器和若干客户端,其特征在于,所述服务器和客户端被配置为迭代执行以下过程,直至模型收敛或达到指定的模型精度:

服务器向客户端发送全局特征提取器参数;

客户端收到全局特征提取器参数后,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

客户端利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

本地训练完成后,各个客户端向服务器上传自己的局部特征提取器参数;

服务器收到客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器。

8.一种计算机设备,其特征在于,包括:一个或多个处理器;存储器;以及一个或多个程序,其中所述一个或多个程序被存储在所述存储器中,并且被配置为由所述一个或多个处理器执行,所述程序被处理器执行时实现如下步骤:

向客户端发送本轮全局特征提取器参数;以及

接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

或者,所述程序被处理器执行时实现如下步骤:

接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

本地训练完成后,向服务器上传自己的局部特征提取器参数。

9.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被

处理器执行时实现如下步骤：

向客户端发送本轮全局特征提取器参数;以及

接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

或者,所述计算机程序被处理器执行时实现如下步骤：

接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

本地训练完成后,向服务器上传自己的局部特征提取器参数。

一种基于知识蒸馏实现特征对齐的个性化联邦学习方法及系统

技术领域

[0001] 本发明涉及分布式计算、深度学习技术领域,具体涉及一种基于知识蒸馏实现特征对齐的个性化联邦学习方法及系统。

背景技术

[0002] 传统的集中式机器学习中,数据通常集中在一个地方进行训练,这可能涉及将数据集上传到云端或者中心服务器。然而,这种方式存在着隐私泄露和数据安全的风险,尤其是当数据包含个人身份信息或敏感商业数据时。此外,数据集过大时,传输和处理数据也会变得非常昂贵和低效,并且传输数据还会带来非常大的通信开销。

[0003] 近年来,随着移动设备的普及和边缘计算的兴起,联邦学习受到了越来越多的关注。联邦学习通过在本地设备上进行模型训练,并仅共享模型更新的梯度或参数,以实现在多个设备上学习全局模型的目标。联邦学习允许数据始终保存在用户本地设备中,不必将数据传输到云端或服务器,这种数据存储和处理方式可以有效的保护数据隐私,降低了数据泄露的风险。目前,联邦学习已被应用到许多现实场景,例如推荐系统、医疗保健、金融等。

[0004] 尽管联邦学习在解决数据隐私和安全性方面取得了显著的成就,但它仍然面临着一些挑战,其中最要的挑战之一就是各参与方的数据异质性。由于联邦学习独特的训练模式,数据往往由端侧产生,这些参与方的数据往往是由其用户、场景、偏好等因素影响,造成各个参与方的数据分布往往不相同。因此,非独立同分布数据是联邦学习中天然存在的一个问题。数据异构性会导致“客户端漂移”现象,即客户端的本地更新方向偏离全局的更新方向,这是由客户端局部优化目标和全局优化目标不一致造成的,这会造成模型收敛速度减慢和性能下降。由于全局和局部数据分布不相同,使用经典的联邦平均算法(FedAvg)等联邦学习算法得到的单一全局模型并不适用于各个客户端。

[0005] 因此,为了解决训练单一全局模型的困难,个性化联邦学习被提出。这种方法致力于为每个客户端构建符合其数据分布的个性化模型。常见的方法包括模型正则化、数据增强和局部微调等。除此之外,模型解耦也是一个重要的研究方向。多任务学习和表示学习的成功表明,将模型解耦为负责提取低维特征的特征提取器和处理与任务强相关的分类器是一种有效的策略。在个性化联邦学习中,特征提取器通过所有客户端共同训练以学习通用的表示,而分类器则进行私有训练,以完成本地分类任务,图1展示了模型解耦方法的训练过程,首先服务器向客户端发送全局特征提取器的参数,客户端收到服后,将其应用到局部模型中,然后客户端利用自己本地的数据训练局部模型,训练完成后客户端仅上传本地的特征提取器的参数,待所有客户端均完成上传后,服务器聚合所有客户端的特征提取器参数,得到新一轮的全局特征提取器参数,反复迭代这一过程直至模型收敛。然而,特征提取器的本地训练通常只能学习到本地的个性化特征信息,而忽视了全局特征信息,导致局部特征提取器可能偏离全局特征标准,从而失去通用性,进而影响全局模型的聚合效果。此

外,仅从参数级别共享特征提取器并不足以从异构数据中获得通用的特征。最近的一些研究提出,在共享特征提取器的基础上,将本地样本的局部特征与全局特征进行特征对齐,从特征级别额外学习通用特征,例如FedPAC(Personalized Federated Learning with Feature Alignment and Classifier Collaboration)提出将本地特征与全局特征质心进行对齐。GPFL(GPFL:Simultaneously Learning Global and Personalized Feature Information for Personalized Federated Learning)则提出将本地特征与全局类别嵌入进行对齐,以引入全局特征信息到本地训练中。然而,这两种方法都需要与服务器通信额外的全局特征信息(全局特征质心或者全局类别嵌入),这会带来额外的隐私和通信开销问题,并且无法为客户端的特征提取器提供细粒度的特征指导。

发明内容

[0006] 发明目的:本发明提出一种基于知识蒸馏实现特征对齐的个性化联邦学习方法,该方法通过知识蒸馏技术实现特征信息的有效传递,使得客户端同时学习个性化和全局特征信息,这在一定程度上提升了局部特征提取器的泛化能力,而且客户端与服务器之间不需要传输额外的特征信息,仅需要传输特征提取器参数,这可以避免额外的通信开销和隐私泄露问题。此外,该方法还能实现细粒度的特征对齐,有效地限制了局部特征提取器的多样性并促进了全局聚合,这使得客户端可以运行更多的本地更新,以通信高效的方式学习通用表示。

[0007] 本发明还提供一种基于知识蒸馏实现特征对齐的个性化联邦学习系统。

[0008] 技术方案:为了实现以上发明目的,本发明的技术方案如下:

[0009] 一种基于知识蒸馏实现特征对齐的个性化联邦学习方法,包括以下步骤:

[0010] 服务器向客户端发送全局特征提取器参数;

[0011] 客户端收到全局特征提取器参数后,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0012] 客户端利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0013] 本地训练完成后,各个客户端向服务器上传自己的局部特征提取器参数;

[0014] 服务器收到客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

[0015] 迭代以上过程,直至模型收敛或达到指定的模型精度。

[0016] 优选的,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,包括:

[0017] 对于单个训练样本x_m,客户端利用保存的全局特征提取器副本对其进行处理,得到全局特征表示$f_{\phi_g}(x_m)$,其中,φ_g表示全局特征提取器,f_{ϕ_g}表示是由φ_g参数化的函数,$f_{\phi_g}(x_m)$是经过函数处理后的一个向量,包含了样本x_m的全局特征信息;

[0018] 利用局部模型获得局部特征和预测输出，包括：

[0019] 对于单个训练样本x_m，客户端利用局部模型θ_i对样本进行处理，得到局部特征$f_{\varphi i}(x_m)$和预测输出$\hat{y}_i$，其中，局部模型θ_i由局部特征提取器φ_i和局部分类器X_i组成，i表示第i个客户端，f_{ϕ_i}是由φ_i参数化的函数，$f_{\phi_i}(x_m)$表示经过函数处理后的一个向量，包含了样本x_m的局部特征信息，$\hat{y}_i$表示局部分类器X_i对样本x_m的预测结果。

[0020] 优选的，蒸馏损失计算方法如下：

[0021] $$\ell_i^d(\phi_i) = \frac{1}{n_i}\sum_{m=1}^{n_i}\left\|f_{\phi_i}(x_m) - f_{\phi_g}(x_m)\right\|_2^2$$

[0022] 其中，$\ell_i^d(\phi_i)$表示第i个客户端的蒸馏损失，n_i为第i个客户端的训练样本的数量。

[0023] 优选的，分类损失计算方法如下：

[0024] $$\ell_i^{ce}(\theta_i) = -\sum_j y_{ij}\log(\hat{y}_{ij})$$

[0025] 其中，$\ell_i^{ce}(\theta_i)$表示第i个客户端的分类损失，y_{ij}是样本x_m属于类别j的真实标签的概率，$\hat{y}_{ij}$是局部模型预测的样本x_m属于类别j的标签的概率。

[0026] 优选的，结合蒸馏损失和分类损失得到总损失，通过最小化总损失来优化局部模型，表示如下：

[0027] $$\min \ell_i(\theta_i) = \ell_i^{ce}(\theta_i) + \beta \cdot \ell_i^d(\phi_i)$$

[0028] 其中，β为平衡分类损失ℓ_i^{ce}和蒸馏损失ℓ_i^d的超参数，用以控制全局特征提取器向局部特征提取器转移的知识程度。

[0029] 优选的，服务器聚合各客户端的局部特征提取器的方法如下：

[0030] $$\phi_g^{(t+1)} = \sum_{i=1}^{k}\alpha_i\phi_i^{(t)}$$

[0031] 其中，$\phi_g^{(t+1)}$为第t+1个全局轮次的全局特征提取器，$\phi_i^{(t)}$为第i个客户端在第t个全局轮次的局部特征提取器，α_i为聚合权重，$\alpha_i = \frac{n_i}{\sum_{i=1}^{k} n_i}$，k为上传本地局部特征提取器的客户端总数目，$n_i$为第i个客户端拥有的数据量。

[0032] 本发明还提供一种基于知识蒸馏实现特征对齐的个性化联邦学习方法系统，包括服务器和若干客户端，其特征在于，所述服务器和客户端被配置为迭代执行以下过程，直至模型收敛或达到指定的模型精度：

[0033] 服务器向客户端发送全局特征提取器参数；

[0034] 客户端收到全局特征提取器参数后，保存一份全局特征提取器参数副本，并使用该参数覆盖掉本地的局部特征提取器参数；

[0035] 客户端利用本地数据进行蒸馏训练，对于每个训练样本，利用保存的全局特征提取器副本获取全局特征，利用局部模型获得局部特征和预测输出，根据全局特征和局部特征的差异得到蒸馏损失，根据预测输出和实际值的差异得到分类损失，结合蒸馏损失和分

类损失得到总损失,通过最小化总损失来优化局部模型;

[0036] 本地训练完成后,各个客户端向服务器上传自己的局部特征提取器参数;

[0037] 服务器收到客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器。

[0038] 本发明还提供一种计算机设备,包括:一个或多个处理器;存储器;以及一个或多个程序,其中所述一个或多个程序被存储在所述存储器中,并且被配置为由所述一个或多个处理器执行,所述程序被处理器执行时实现如下步骤:

[0039] 向客户端发送本轮全局特征提取器参数;以及

[0040] 接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

[0041] 或者,所述程序被处理器执行时实现如下步骤:

[0042] 接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0043] 利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0044] 本地训练完成后,向服务器上传自己的局部特征提取器参数。

[0045] 本发明还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如下步骤:

[0046] 向客户端发送本轮全局特征提取器参数;以及

[0047] 接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

[0048] 或者,所述计算机程序被处理器执行时实现如下步骤:

[0049] 接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0050] 利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0051] 本地训练完成后,向服务器上传自己的局部特征提取器参数。

[0052] 本发明与现有技术相比,具有如下优点和有益效果:

[0053] (1)本发明提出一种新颖的个性化联邦学习新方法,该方法将知识蒸馏技术引入到个性化联邦学习中,以实现对特征信息的有效传递。通过知识蒸馏,客户端能够从全局特征提取器中获取到丰富的特征知识,然后将这些知识应用到局部模型的训练中。这样一来,客户端不仅能够根据本地数据训练出个性化的模型,还能够借助全局特征信息提升特征提取器的泛化能力。允许客户端在训练过程中同时学习个性化和全局特征信息。通过这种方式,不仅能够解决数据异质性带来的问题,还能够有效地提升局部特征提取器的泛化能力,

从而进一步改善联邦学习的性能。(2)该发明实现了细粒度的特征对齐,能够使得局部特征提取器在训练过程中更加一致,减少了模型之间的差异性。这样一来,当客户端进行本地更新时,所学习到的特征更加一致,更具有代表性,能够更好地反映全局数据的特征。同时,通过促进全局聚合,能够将各个客户端学习到的特征信息有效地整合起来,形成更加完整和准确的全局表示。该发明允许客户端以通信高效的方式进行更多的本地更新,能够更充分地利用分布式环境下的资源,加速模型的收敛速度。这不仅提高了联邦学习的效率,也降低了通信开销,使得该发明在实际应用中更加具有可行性和可扩展性。(3)本发明为解决联邦学习中的非独立同分布数据开辟了新的机会。

附图说明

[0054] 图1是一般的个性化联邦学习中模型解耦方法的训练流程图;

[0055] 图2是一般的知识蒸馏示意图;

[0056] 图3是根据本发明的客户端知识蒸馏示意图;

[0057] 图4是根据本发明的基于知识蒸馏的实现特征对齐的个性化联邦学习方法训练流程图。

具体实施方式

[0058] 下面结合附图对本发明的技术方案作进一步说明。

[0059] 知识蒸馏是一种模型压缩技术,旨在通过传输一个模型(通常是较大、复杂的模型)的“知识”给另一个模型(通常是较小、简化的模型),从而使后者能够学习到前者的“知识”,同时保持其性能或提升性能。该方法最初由Hinton等人提出,并在深度学习领域得到广泛应用。在知识蒸馏中,通常有两个模型参与:教师模型和学生模型。知识蒸馏的目标是通过学生模型模仿教师模型的输出或者中间层特征来引导学生模型学习。在知识蒸馏的训练过程中,学生模型在优化过程中既要尽量减小真实标签损失,又要尽量减小蒸馏损失以此来学习教师模型的“知识”,图2为知识蒸馏的示意图。

[0060] 本发明在联邦学习客户端上,将本地模型划分为特征提取器和分类器。根据联邦学习机制,客户端每一轮训练得到客户端本地的特征提取器参数,上送给服务器,由服务器进行聚合形成全局特征提取器,并在下一轮迭代下发给客户端。本发明上下文中,全局特征提取器是服务器聚合后分发给所有客户端的共享特征提取器参数,局部特征提取器是每个客户端基于全局特征提取器参数在本地数据上进行个性化训练后的特征提取器参数。本发明结合蒸馏技术,将全局特征提取器视作教师模型,局部特征提取器视作学生模型,在图3中展示了本地蒸馏的示意图。通过全局特征提取器向局部特征提取器转移知识,能够让客户端在训练过程中同时学习个性化和全局特征信息。这不仅能够解决数据异质性带来的问题,还能够有效地提升局部特征提取器的泛化能力,从而进一步改善联邦学习的性能。

[0061] 参照图4,本发明提出的基于知识蒸馏实现特征对齐的个性化联邦学习方法,具体包括以下步骤:

[0062] 步骤1,服务器向选定的客户端发送全局特征提取器参数。

[0063] 在每一轮训练的开始,服务器向选定的客户端发送全局特征提取器的参数,该全局特征提取器的参数来自于上一轮客户端训练得到的局部特征提取器的聚合。如果是第一

轮全局迭代,发送的是服务器初始化的所有模型参数,包括特征提取器参数和分类器参数,此轮之后仅向客户端发送特征提取器参数。

[0064] 步骤2,客户端将全局特征提取器参数应用到本地。

[0065] 对于收到全局特征提取器参数的客户端,使用全局特征提取器参数覆盖掉本地的特征提取器参数,以确保所有客户端从相同的特征提取器模型基础开始个性化训练。应确保数据在传输过程中没有被损坏,可以将收到的全局特征提取器的架构与本地的特征提取器模型的架构相对比,如果相同,则视为数据没有被损坏,否则视为损坏,如果有损坏,客户端则需要向服务器请求获取全局特征提取器参数。如果没有损坏,客户端则在本地保存一份全局特征提取器参数副本,然后解析全局特征提取器,包括特征提取器的架构信息和具体的参数数据,然后读取本地的特征提取器参数,将全局特征提取器的参数逐一赋值给本地特征提取器对应的参数。

[0066] 步骤3,客户端利用本地的数据进行本地训练。

[0067] 本发明实施例中是对图像识别模型的训练,将图像训练集送入网络,根据网络的实际输出与期望输出间的差别来调整参数。训练模型的步骤如下:

[0068] a、客户端将本地模型(θ)划分为特征提取器(φ)和分类器(χ),其中χ是最后一层全连接层。特征提取器负责从输入数据中提取高维特征,由多层卷积层或全连接层组成,分类器利用特征提取器提取的特征进行分类,一般由若干全连接层组成,最后一层输出类别概率。在这里我们将最后一层全连接层称为分类器,分类器之外的所有层称为特征提取器。

[0069] 本地模型(θ)是一个广义的概念,指从输入数据中提取特征的模型部分,它由特征提取器(φ)和分类器(χ)组成。在模型训练过程中,特征提取器主要分为全局特征提取器(φ_g)和局部特征提取器(φ_i),如上所述,全局特征提取器是服务器聚合后分发给所有客户端的共享特征提取器参数,客户端会在本地保存一份全局特征提取器参数副本;局部特征提取器是每个客户端基于全局特征提取器参数在本地数据上进行个性化训练后的特征提取器参数。客户端i的特征提取器表示为φ_i,全局特征提取器表示为φ_g。局部模型(θ_i)由局部特征提取器(φ_i)和局部分类器(χ_i)组成。

[0070] 记f_φ是一个由φ参数化的函数,将数据点由d维射到k维的特征空间,$\varphi: \mathbb{R}^d \to \mathbb{R}^k$。$f_\chi$是一个由χ参数化的函数,将K维特征映射到标签空间$\chi: \mathbb{R}^k \to y$。因此,客户端的本地损失函数可表示为:$\ell(\phi,\chi) = \ell(\phi) \circ \ell(\chi)$。

[0071] b、选择训练集合中的一个图像样本(x_m),x_n为图像数据,$\hat{y}_m$为标签,即图像所属的类别;

[0072] c、将图像样本x_n输入到本地保存的全局特征提取器φ_g的副本中,计算特征提取器的实际输出$f_{\phi_g}(x_m)$,即样本的全局特征;

[0073] d、将图像样本x_m输入到局部模型θ_i中,经过局部特征提取器φ_i得到样本的局部特征$f_{\phi_i}(x_m)$,经过局部分类器χ_i得到对样本x_m预测概率$\hat{y}_{ij}$;

[0074] e、计算样本x_m的全局特征与局部特征之间的差异,即蒸馏损失。这里的蒸馏损失使用均方误差损失。蒸馏损失可表示为:

[0075] $$\ell_i^d(\phi_i) = \frac{1}{n_i}\sum_{m=1}^{n_i} \left\| f_{\phi_i}(x_m) - f_{\phi_g}(x_m) \right\|_2^2$$

[0076] f、计算预测值与实际值之间的误差。使用交叉熵损失函数衡量模型预测输出与真实标签之间的误差,分类损失可表示为:

[0077] $$\ell_i^{ce}(\theta_i) = -\sum_j y_{ij} \log(\hat{y}_{ij})$$

[0078] 其中y_{ij}是第i个客户端上样本x_m的真实标签的概率,$\hat{y}_{ij}$是第i个客户端上局部模型预测的标签的概率;

[0079] g、计算总体损失。总体损失是蒸馏损失和分类损失的加权和:

[0080] $$\ell_i(\theta_i) = \ell_i^{ce}(\theta_i) + \beta \cdot \ell_i^d(\phi_i)$$

[0081] 其中β为平衡本地交叉熵损失ℓ_i^{ce}和蒸馏损失ℓ_i^d的超参数,用以控制全局特征提取器向本地特征提取器转移的知识程度;通过最小化本地损失ℓ_i,客户端可利用本地数据学习个性化的头部,同时也能显式的将本地特征与全局特征进行对齐,本地特征提取器能够同时学习到局部和全局的特征信息。

[0082] h、利用反向传播根据损失$\ell_i(\theta_i)$计算关于模型参数的梯度信息$\nabla_{\theta_k^{t,j}}\ell_i$;

[0083] i、优化模型参数。使用随机梯度下降优化算法更新模型参数:

[0084] $$\theta_k^{t,j+1} \leftarrow \theta_k^{t,j} - \eta\nabla_{\theta_k^{t,j}}\ell_i$$

[0085] 其中η为超参数,控制模型参数的更新幅度,$\theta_k^{t,j}$表示在训练的第t个全局轮次中第k个客户端的第j次本地更新后的局部模型参数。

[0086] j、对每个图像样本重复上述a-i过程,直至遍历完整个图像样本集,完成本地数据集一次训练迭代;

[0087] k、重复上述a-j过程,客户端完成本地数据集多轮训练迭代。

[0088] 步骤4,客户端上传自己的特征提取器参数。

[0089] 在本地训练完成后,各个客户端将训练后的特征提取器参数上传到中央服务器。此时,各客户端的分类器部分保持在本地,不进行上传。

[0090] 步骤5,服务器聚合客户端的特征提取器参数。

[0091] 服务器接收到所有客户端上传的特征提取器参数后,根据各个客户端的数据量来确定聚合权重。数据量越大的客户端,其模型参数在聚合过程中所占的权重也越大。服务器使用这些权重对各客户端的特征提取器参数进行加权平均,得到新一轮的全局特征提取器。聚合方式如下:

[0092] $$\phi_g^{(t+1)} = \sum_{i=1}^{k} \alpha_i \phi_i^{(t)}, \alpha_i = \frac{n_i}{\sum_{i=1}^{k} n_i}$$

[0093] 其中n_i为第i个客户端拥有的数据量,k为上传特征提取器的客户端总数目,$\phi_g^{(t+1)}$为第t+1个全局轮次的全局特征提取器。

[0094] 步骤6、重复以上步骤1-步骤5,直至模型收敛或者达到指定的模型精度。

[0095] 本发明还提供一种基于知识蒸馏实现特征对齐的个性化联邦学习方法系统,包括服务器和若干客户端,其特征在于,所述服务器和客户端被配置为迭代执行以下过程,直至模型收敛或达到指定的模型精度:

[0096] 服务器向客户端发送全局特征提取器参数;

[0097] 客户端收到全局特征提取器参数后,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0098] 客户端利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0099] 本地训练完成后,各个客户端向服务器上传自己的局部特征提取器参数;

[0100] 服务器收到客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器。

[0101] 本发明还提供一种计算机设备,包括:一个或多个处理器;存储器;以及一个或多个程序,其中所述一个或多个程序被存储在所述存储器中,并且被配置为由所述一个或多个处理器执行,所述程序被处理器执行时实现如下步骤:

[0102] 向客户端发送本轮全局特征提取器参数;以及

[0103] 接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

[0104] 或者,所述程序被处理器执行时实现如下步骤:

[0105] 接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0106] 利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0107] 本地训练完成后,向服务器上传自己的局部特征提取器参数。

[0108] 本发明还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如下步骤:

[0109] 向客户端发送本轮全局特征提取器参数;以及

[0110] 接收客户端的局部特征提取器参数后,根据各个客户端拥有的数据量确定聚合权重,基于聚合权重对各客户端的局部特征提取器进行聚合,得到新一轮全局特征提取器;

[0111] 或者,所述计算机程序被处理器执行时实现如下步骤:

[0112] 接收服务器下发的全局特征提取器参数,保存一份全局特征提取器参数副本,并使用该参数覆盖掉本地的局部特征提取器参数;

[0113] 利用本地数据进行蒸馏训练,对于每个训练样本,利用保存的全局特征提取器副本获取全局特征,利用局部模型获得局部特征和预测输出,根据全局特征和局部特征的差

异得到蒸馏损失,根据预测输出和实际值的差异得到分类损失,结合蒸馏损失和分类损失得到总损失,通过最小化总损失来优化局部模型;

[0114] 本地训练完成后,向服务器上传自己的局部特征提取器参数。

[0115] 本发明的实施例可作为方法、装置、计算机设备或计算机程序产品提供,能够采用完全硬件、完全软件或软硬件结合的形式实现。本发明可包含在计算机可用存储介质(如磁盘存储器、CD-ROM、光学存储器)上的计算机程序产品。这些程序指令可以指引通用计算机、专用计算机、嵌入式处理器或其他可编程设备的处理器,生成用于执行流程图中指定功能的装置。计算机程序指令可存储在计算机可读存储器中,或装载到计算机或其他可编程设备上,以实现指定功能的操作步骤。

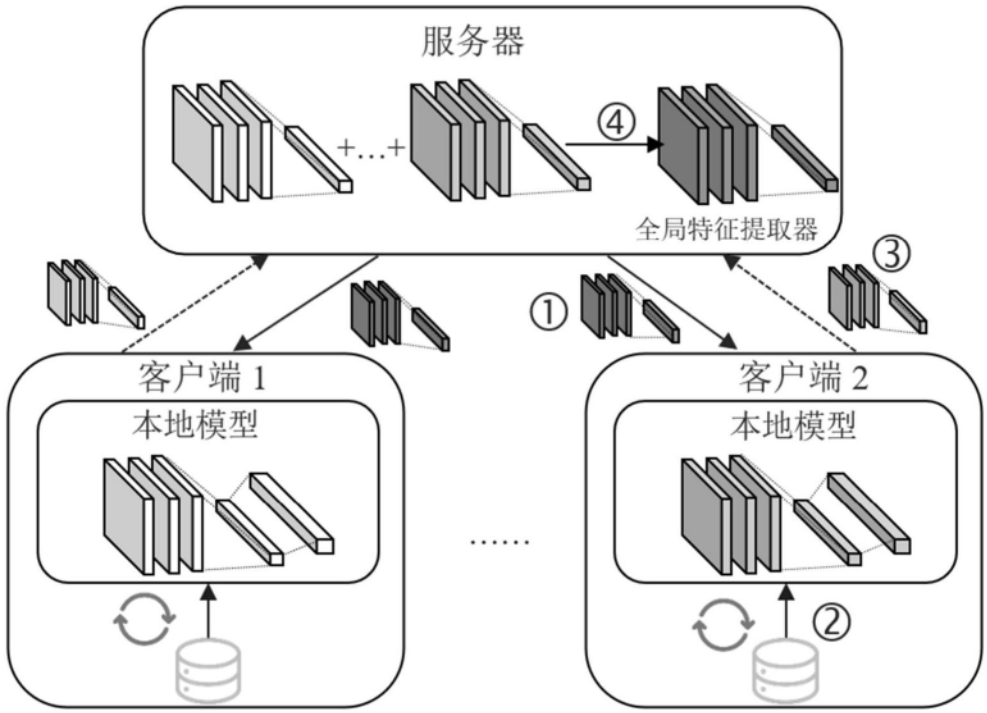

服务器
+…+
④
全局特征提取器
③
①
客户端 1
本地模型
客户端 2
本地模型
……
②

图1

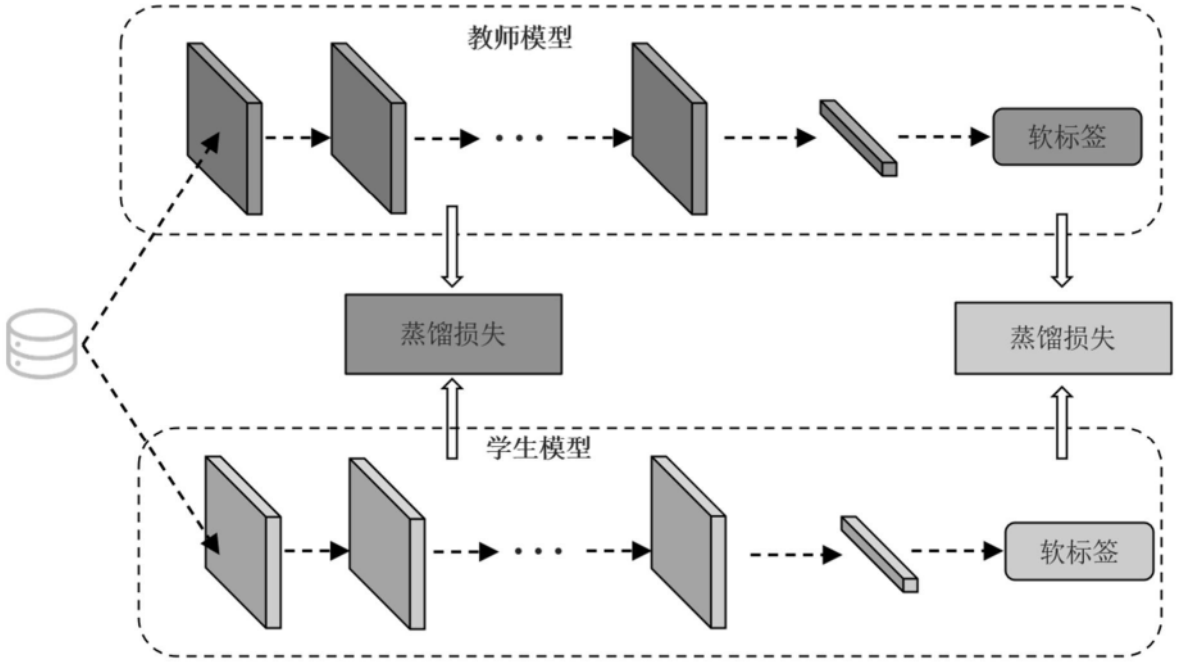

教师模型
软标签
蒸馏损失
蒸馏损失
学生模型
软标签

图2

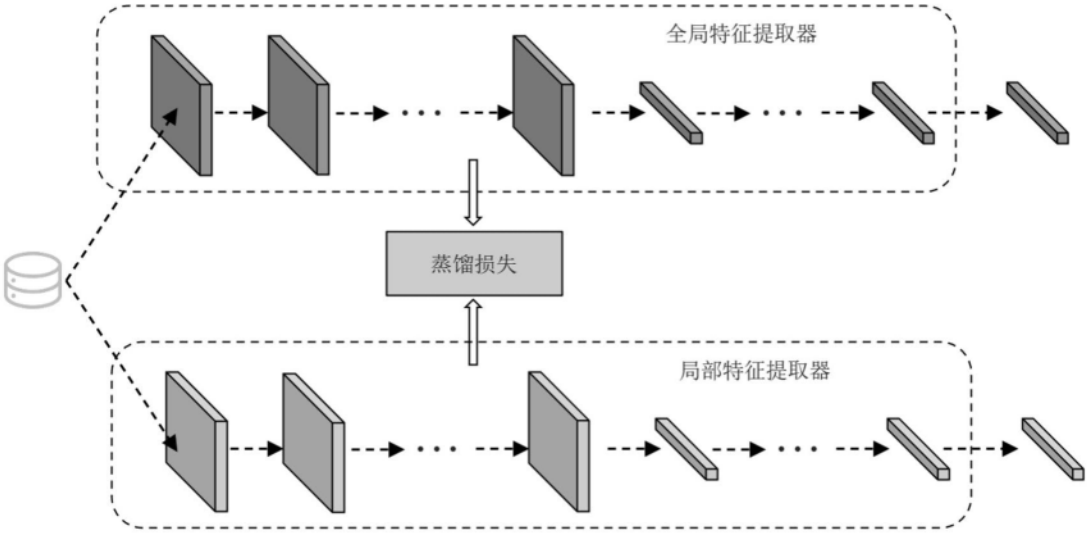

全局特征提取器
蒸馏损失
局部特征提取器

图3

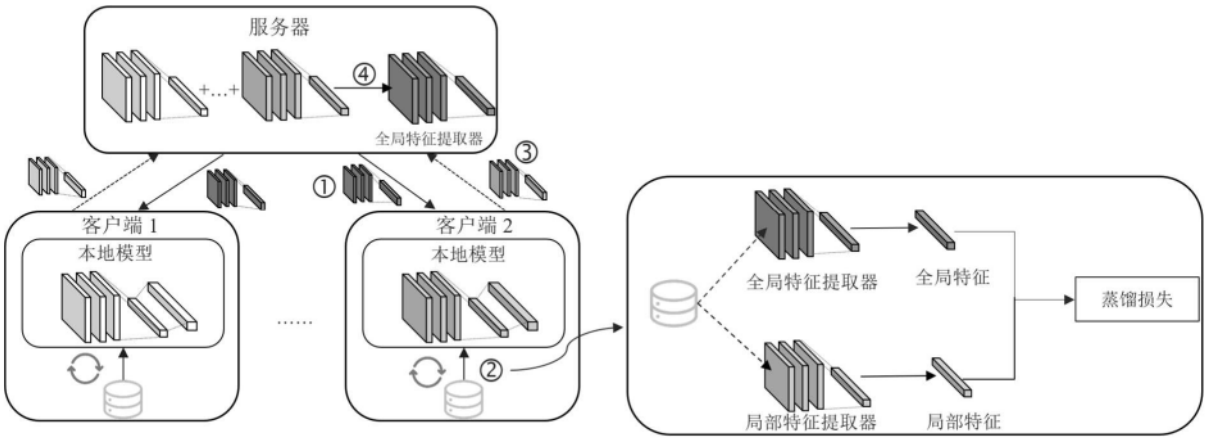

服务器
④
全局特征提取器
③
①
客户端 1
本地模型
客户端 2
本地模型
②
全局特征提取器
全局特征
蒸馏损失
局部特征提取器
局部特征

图4